Metabolomics and Genomics-Driven Genome-Mining in *Streptomyces*

by

Richenda McFarlane

Thesis submitted to the
University of Ottawa
in partial fulfillment of the requirements for the
Master's degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Table of Contents

Table of Contents ii
List of Figures iii
List of Tables v
Abstract vi
Acknowledgements vii

Chapter 1. Natural Product Discovery by Activation of Cryptic Biosynthetic Gene Clusters 1
1.1 Introduction 1
1.2 References 13
Chapter 2. Induction of Secondary Metabolism in Two *Streptomyces* Species Using Elicitors 17
2.1 Overview 17
2.2 Results and Discussion 18
2.3 Methods 48
2.4 References 57
Chapter 3. Genomics-Drive Genome Mining and Assembly in *Streptomyces* 60
3.1 Introduction 60
3.2 Results and Discussion 64
3.3 Conclusion 78
3.4 Methods 78
3.5 References 82
Chapter 4. Implications and Future Work 85
4.1 Elicitors for Activation of Silent Gene Clusters 85
4.2 A Secondary Metabolite-Based Approach to Genome Assembly 87
4.3 References 89
Appendix I. *Streptomyces* sp. PBH53 and Trimethoprim 91
Appendix II. *Streptomyces* sp. PBH53 and Atenolol 98
Appendix III. *Streptomyces armeniacus* and Trimethoprim 103
Appendix IV. *Streptomyces armeniacus* and Atenolol 110
Appendix V. Genome Sequencing and Assembly 113

Table of Figures

Figure 1.1	**Examples of bacterial natural products with clinical relevance**	**2**
Figure 1.2	**Overview of the main BGC activation methods**	**3**
Figure 1.3	**The basis for hormesis**	**7**
Figure 1.4	**The role of dihydrofolate reductase (DHFR) in primary metabolism in *Burkholderia thailandensis***	**9**
Figure 1.5	**Structures of trimethoprim (TMP) and atenolol (AT)**	**11**
Figure 1.6	**The *Streptomyces* life cycle**	**12**
Figure 2.1	**Loadings plot of the first two PCs of the PCA for *S.* sp. PBH53 with and without TMP**	**19**
Figure 2.2	**PCA scores plot comparing secondary metabolite fingerprints of S. sp. PBH53 without TMP**	**20**
Figure 2.3	**Volcano plot of features produced by *S.* sp. PBH53 with TMP**	**21**
Figure 2.4	**Total Ion Chromatogram (TIC) overlay of *S.* sp. PBH53 with TMP and without TMP**	**22**
Figure 2.5	**Mass spectrum of *S.* sp. PBH53 with TMP at a retention time of 5.12 mins**	**23**
Figure 2.6	**Loadings plot of the first two PCs of the PCA for *S.* sp. PBH53 with and without AT**	**25**
Figure 2.7	**PCA scores plot comparing secondary metabolite fingerprints of *S.* sp. PBH53 without AT and with AT**	**26**
Figure 2.8	**Volcano plot of features produced by *S.* sp. PBH53 with AT**	**27**
Figure 2.9	**TIC overlay of *S.* sp. PBH53 with AT and without AT**	**28**
Figure 2.10	**Mass spectrum of *S.* sp. PBH53 with AT at a retention time of 4.92 mins**	**29**
Figure 2.11	**The structure of salinomycin**	**29**
Figure 2.12	**Extracted ion chromatogram of *m/z* 750.99-751.99 in the *S* sp. PBH53 sample and in the salinomycin standard**	**30**
Figure 2.13	**PCA scores plot comparing secondary metabolite fingerprints of *S. armeniacus* without TMP and with TMP**	**32**
Figure 2.14	**Loadings plot of the first two PCs of the PCA for S. *armeniacus* with and without TMP**	**33**
Figure 2.15	**Volcano plot of features produced by *S. armeniacus* with TMP**	**34**
Figure 2.16	**TIC overlay of *S. armeniacus* with TMP and without TMP**	**35**
Figure 2.17	**Mass spectrum of *S. armeniacus* with TMP at a retention time of 4.15 mins**	**36**
Figure 2.18	**PCA scores plot comparing secondary metabolite fingerprints of *S. armeniacus* without AT and with AT**	**37**
Figure 2.19	**Loadings plot of the first two PCs of the PCA for *S. armeniacus* with and without AT**	**38**
Figure 2.20	**Volcano plot of features produced by *S. armeniacus* with AT**	**39**
Figure 2.21	**TIC overlay of *S. armeniacus* with AT and without AT**	**40**
Figure 2.22	**Mass spectrum of *S. armeniacus* with AT at 3.80 mins**	**41**
Figure 2.23	**Mass spectrum of *S. armeniacus* with AT at 3.8 min showing b or y ions**	**42**
Figure 3.1	**A genome sequence of ABBBCAC showing one Eulerian walk through the De Bruijn assembly graph.**	**62**
Figure 3.2	**A genome sequence of ABBCDABCABD with more than one Eulerian walk through the De Bruijn assembly graph**	**62**
Figure 3.3	**Comparison of the Naphthomycin A gene cluster in the published S. sp. PBH53 genome to the reference Naphthomycin A gene cluster**	**68**
Figure 3.4	**Simulated raw reads for the Naphthomycin A BGC assembled using SPAdes**	**72**

Figure 3.5 The Naphthomycin A cluster resulting from the assembly of the resequenced *S.* sp. PBH53 genome, compared to the literature Naphthomycin A sequence **73**

Figure 3.6 An example of one of the proposed Naphthomycin A pathways from the "gene clusters" file of the BiosyntheticSPAdes genome assembly **74**

Table of Tables

Table 2.1	**Summary of culture conditions and treatments for each *Streptomyces* strain-elicitor Interaction**	**49**
Table 3.1	**The results of the assembly of simulated Illumina reads of large PKS pathways by SPAdes and Biosynthetic SPAdes**	**65**
Table 3.2	**Genomes and pathways from the NCBI database found to contain potential assembly errors in PKS gene clusters**	**67**
Table 3.3	**Results of assembling the Naphthomycin A cluster using SPAdes at 30X coverage with varying read lengths**	**69**
Table 3.4	**Results of assembling the Naphthomycin A cluster using SPAdes with 150 bp reads and varying depth of coverage**	**70**

Abstract

Bacterial natural products are important sources of novel therapeutics. However, not all secondary metabolites are expressed concurrently, increasing rediscovery of compounds and limiting identification of novel molecules. Trimethoprim (TMP) and atenolol (AT) were used as elicitors of secondary metabolism and evaluated by metabolomics experiments for two *Streptomyces* strains: *S.* sp. PBH53 and *S. armeniacus* ATCC 15676. TMP acted as a pathway-specific activator in *S.* sp. PBH53, but as a global regulator in *S. armeniacus*, while AT acted as a pathway-specific regulator in both strains. One peak of interest was identified from each treatment, with a total of four molecules chosen for further characterization and structural elucidation. Three of the four are likely uncharacterized molecules, with one having a potential match to salinomycin. TMP and AT prove to be effective elicitors in *Streptomyces* strains, enabling discovery of uncharacterized natural products.

Advancements in whole genome assembly techniques have provided opportunities for natural product discovery to be based on genomic information rather than compound isolation based on biological activity. However, methods for assembling short read sequence data often results in assembly errors in polyketide synthase (PKS) pathways due to repetitive sequences in the acyl-transferase (AT) domains of the pathway. These errors exist in publicly available genomes in the NCBI database, as well as in the assembly of simulated reads of long PKS pathways. Similar errors exist in the *S.* sp. PBH53 genome, where there are assembly errors in the pathway for Naphthomycin A, but the strain is able to produce the molecule indicating a complete biosynthetic gene cluster (BGC) that doesn't appear in the assembly. Evaluations of SPAdes and Biosynthetic SPAdes did not show reliable solutions to the assembly errors occurring in genomes. Future work will explore the possibilities of using literature BGC sequences as scaffolds for more accurate assemblies of genomes.

Acknowledgements

First I would like to thank my mom and dad, and my sister for their love and support.

I would also like to thank Chris Boddy for letting a biology student into the Chemistry Department, and for encouraging me to learn skills outside of my comfort zone.

I would also like to thank Dr. Dave Overy for sharing his expertise in metabolomics, and to him and Amanda Sproule for running our samples and helping develop this project.

A thank you to the members of the Boddy Lab who helped me in my research. To Mike, for being my chemist counterpart in metabolomics; to Jordan, for being the friend I needed in the lab; to Graham for helping run samples and always letting me talk through my ideas; and to everyone else who contributed. This research was not possible without all of you.

Chapter 1: Natural Product Discovery by Activation of Cryptic Biosynthetic Gene Clusters

1.1 Introduction

Natural products are a key source of new leads for drug development, with varied biological activities and applications across all therapeutic areas including as antimicrobial, antifungal, anticancer agents **(Figure 1.1).**[1],[2] Over 40% of drugs approved between 1981 and 2019 are from biological sources or are derivatives of natural products.[2] Bacteria are at the forefront of natural product research, as they exhibit the ability to synthesize secondary metabolites of diverse chemical structures and biological targets.[3] In bacteria, the genes responsible for synthesizing natural products are arranged into biosynthetic gene clusters (BGCs). With the development of genome sequencing technologies and bioinformatics tools such as antiSMASH[4], it is becoming more apparent that the potential for bacteria to make natural products is far greater than what is observed. Many BGCs known as cryptic or silent BGCs are not expressed under normal laboratory conditions, making it difficult to identify their corresponding natural products, and requiring alternative methods to find new secondary metabolites.

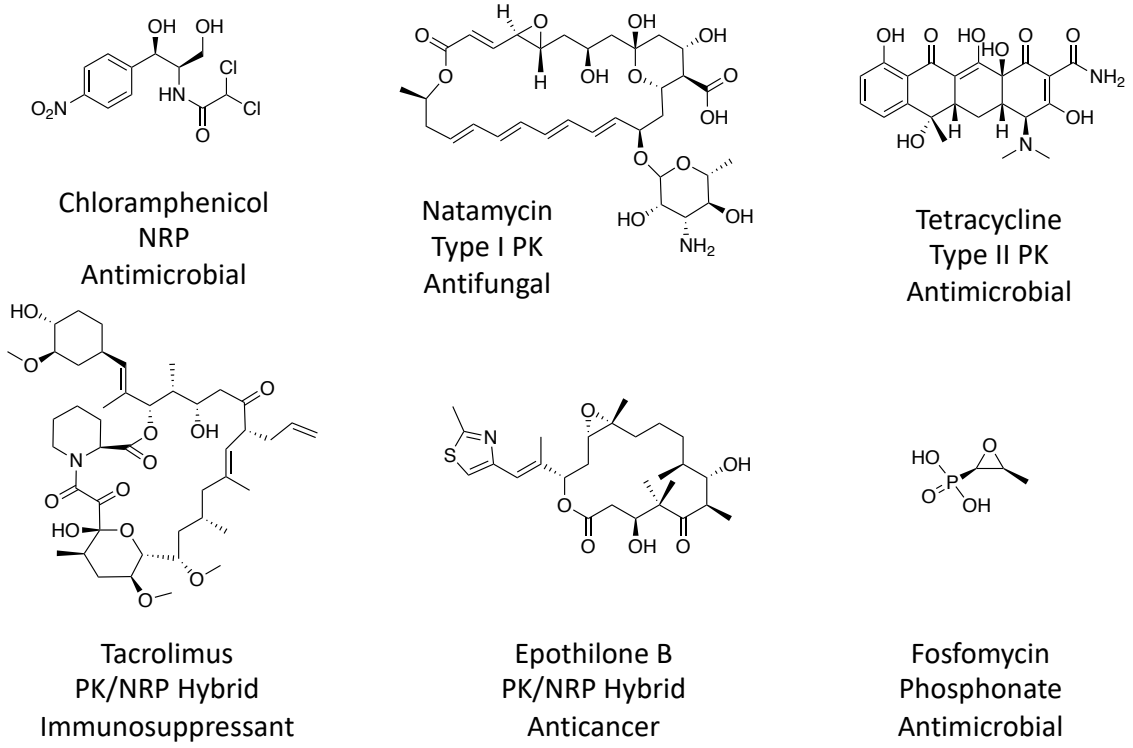

Figure 1.1. Examples of bacterial natural products with clinical relevance (NRP = nonribosomal peptide; PK = polyketide)

Significant effort over the last two decades has been directed towards the development of methods for the activation of these cryptic BGCs for the discovery of new molecules. **(Figure 1.2).**[3],[5] Some BGC activation methods rely on altering the physical growth conditions of the bacteria. One such method is the One Strain-Many Compounds (OSMaC) approach, first used in 2002. This method requires exposing the bacteria of interest to many different culture conditions by varying parameters such as media composition, pH, oxygen levels, and temperature, resulting in different secondary metabolites being produced in different conditions. This approach to BGC activation is broadly applicable to many different organisms including streptomycetes, myxobacteria, and fungi, and in one example led to the isolation of up to 20 novel manumycins and manumycin analogues from *Streptomyces parvulus* Tü 64 when

subject to different levels of oxygenation.[6] Another way of altering the physical growth conditions of the bacteria is with the addition of solid supports such as cotton scaffolds, which facilitates the production of biofilms and can result in the activation of silent clusters, as seen in studies of the marine bacteria *Pseudoalteromonas.*[7]

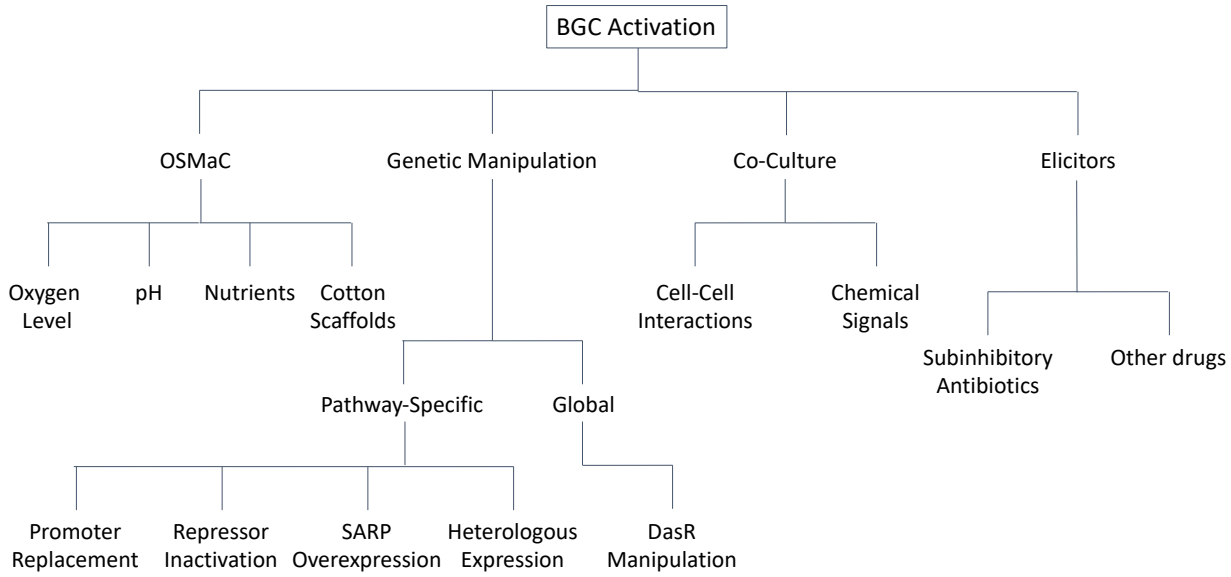

Figure 1.2. Overview of the main BGC activation methods falling under the categories of OSMaC, with select examples provided, genetic manipulation, co-culture with an inducing strain of bacteria, and the addition of elicitors (OSMaC = One Strain-Many Compounds; SARP = *Streptomyces* Antibiotic Regulator Proteins).

Because most cryptic BGCs are not expressed rather than regulated allosterically or covalently, activating their transcription with pathway-specific or global regulators has been a successful strategy. Pathway-specific regulation is often accomplished by replacing the native promoters present in a BGC with constitutive promoters for continuous expression of the product, also known as refactoring.[8], [9] Inducible promoters have also been successfully used to activate silent gene clusters in the presence of the given inducer. For example, exchange of

native promoters for arabinose-inducible promoters in *Photorhabdus luminescens* resulted in the discovery of the novel mevalagmapeptides.[10] Other research uses the understanding of native transcriptional regulators to activate pathways and discover new natural products. Decoy oligonucleotides designed to mimic the binding sites of repressors are effectively used to interfere with binding of the target DNA sequence to prevent repression of a BGC. This technique has been successfully used to stimulate production of actinorhodin in *Streptomyces coelicolor* A3(2).[11] Regulation of gene expression can also be controlled with the constitutive or over-expression of pathway-specific regulators called Streptomyces Antibiotic Regulator Proteins (SARPs), resulting in the discovery of ishigamide from recombinant *Streptomyces* strains.[12] These methods require manipulation of the native host, which presents practical difficulties, including unestablished methods transformation of DNA in many environmental bacteria and ineffective approaches for editing the genomes of the native hosts.

To circumvent the challenges associated with genetic manipulation of a native host strain, heterologous expression offers an attractive alternative. However, it presents new barriers, with the lack of specific transcriptional regulators and biosynthetic precursors in the chosen host causing potential problems in expression. Heterologous expression is best utilized when screening environmental DNA for BGCs of interest, as there is no identified organism associated with the DNA and therefore requires a host organism.[13] This method has proven to be most successful when coupled with other approaches, where the combination of heterologous expression with over-expression of the associated pathway-specific SARP led to the discovery of Tetarimycin A from environmental DNA. The SARP was able to appropriately

activate expression of the foreign gene cluster in the *Streptomyces albus* heterologous host for detection of the novel Type II polyketide (PK).[14]

Global regulation of transcription often involves the DasR regulator, especially in actinomycete bacteria. It plays a role in metabolism of N-acetylglucosamine (GlcNAc), and serves as a pleiotropic regulator in many bacterial species with complex regulatory pathways. Genetic manipulation of *dasR* has revealed its role in activating or repressing expression of different genes. Overexpression of DasR in *Streptomyces cinnamonensis* resulted in upregulation of genes associated with the production of the polyether ionophore monensin.[15] However, in *S. coelicolor*, DasR was shown to repress biosynthesis of the pigmented antibiotics undecylprodigiosin and actinorhodin. When supplemented with GlcNAc, the DNA-binding activity of DasR was inhibited by glucosamine-6-phosphate, the direct metabolic intermediate from GlcNAc and a linchpin in regulating metabolic flux from central metabolism to cell wall biosynthesis, resulting in the upregulation of antibiotic production.[16] The role of DasR also seems to change depending on nutrient availability. In rich media, the addition of GlcNAc resulted in decreased secondary metabolite production[17], while in minimal media it often increases production of antibiotics.[16] Depending on nutrient availability and the role of DasR in the regulation of individual secondary metabolite pathways, the manipulation of DasR can result in the expression of cryptic BGCs and the identification of novel compounds.

Co-culture techniques, which are more similar to the natural environment the bacteria live in than monoculture, are also well-established for activating cryptic gene clusters. This can occur either by cell-to-cell interactions between neighbouring bacterial strains or by signalling via the secreted metabolome. The mycolic acid-containing bacteria *Tsukamurella pulmonis*

grown in co-culture with *Streptomyces endus* S-522 was successfully used to elicit expression of silent gene clusters in *S. endus* S-522, resulting in the discovery of the novel antibiotic alchivemycin A. The production of this antibiotic was dependent both on cell-to-cell interactions and the presence of mycolic acid, as neither exposure to *T. pulmonis* supernatant nor a mycolic acid-deficient mutant resulted in production of alchivemycin A.[18] Silent gene clusters in *Streptomyces* can also be activated by co-culture with other actinomycetes, which led to the discovery of 12 new analogues of the acyl-desferrioxamine siderophores in *S. coelicolor* due to the presence of other siderophores produced by the co-culture strain, *Amycolatopsis* sp. AA4, independently of cell-to-cell interactions.[19] In this case, *S. coelicolor* responded to chemical signals produced by the co-culture strain, a concept that can be applied with the use of chemical elicitors.

Chemical elicitors can originate from molecules produced by other bacteria, similarly to co-culture. Subinhibitory concentrations of antibiotics are known to have wide and highly variable effects throughout the bacterial domain, presenting the idea that antibiotic compounds may serve more than one purpose in nature. At low concentrations, they could act as signalling molecules between bacterial populations, and at high concentrations they might serve to kill off competing bacteria.[20],[21] Where inhibitory concentrations prevent bacterial growth, subinhibitory concentrations can modulate transcription, morphology, and natural product biosynthesis. A 2001 study screened *Streptomyces lividans* treated with extracts of 405 actinomycetes strains for sporulation and the production of actinorhodin. This screen identified a new compound called goadsporin, which had the ability to induce antibiotic production in the target strain and displayed antimicrobial activity against *Bacillus subtilis*. The effects of

antibiotics on transcription were studied in 2002, where a variety of antibiotic compounds of different classes were used at doses below the minimum inhibitory concentration (MIC) to alter gene expression in *Salmonella typhimurium*, with about 5% of transcripts being affected. Different genes responded to antibiotics that varied in mechanism of action. Genes involved in transport, virulence, DNA repair, and other unidentified functions were activated by exposure to various antibiotics, such as rifampicin, erythromycin, and trimethoprim, while other pathways were repressed.[22] This introduces the concept of hormesis, where at subinhibitory concentrations the antibiotic modulates gene expression, and at the MIC has an inhibitory effect on the bacteria **(Figure 1.3)**.[23]

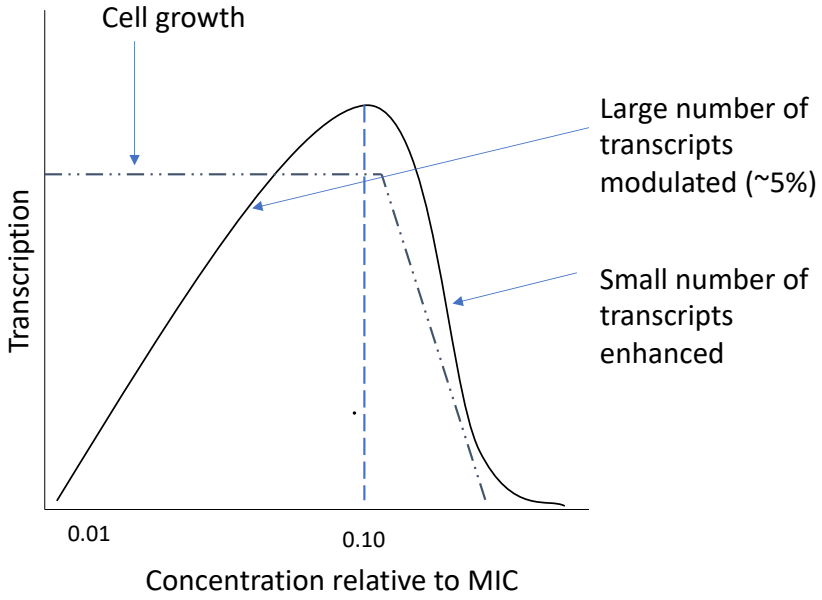

Figure 1.3. The basis for hormesis, where the growth of a microbe is inhibited at high concentrations, but at subinhibitory concentrations the bacteria undergoes changes in transcriptional regulation (MIC= minimum inhibitory concentration).

Selection of elicitors can also be approached rationally. For example, when targeting expression of polyketide synthase (PKS) pathways, primary metabolism of fatty acids can be targeted as they share malonyl-CoA and acetyl-CoA monomers with PK synthesis. Partial inhibition of fatty acid synthesis by ARC2, a triclosan analogue, resulted in activation of PKS pathways due to the increased availability of malonyl-CoA and acetyl-CoA.[24] The same logic was also applied to nonribosomal peptide synthetase (NRPS) pathways, where chloramphenicol was used at 1/30 of its MIC to target ribosomal protein synthesis and increase the availability of amino acids for increased production of the NRPs actinomycin and piperidamycin.[25]

Screening for effective elicitors for a specific organism or pathway is a popular method for activation of cryptic BGCs, as neither the mechanism of action for induction nor the structure or activity of the induced molecules has to be known. These attributes make it highly effective when used in studies aimed at discovering new natural products. To evaluate expression of a specific pathway, reporter genes such as *lacZ* or *eGFP* have been used as an indicator of successful BGC expression.[26],[27] However, genetic manipulation is not necessary to make use of elicitors, avoiding the limitations of genetics-based approaches. Elicitor screens carried out against *Burkholderia thailandensis* by the Seyedsayamdost group revealed that trimethoprim (TMP) not only activated the cryptic virulence factor malleilactone[26], it also triggered the production of over 100 compounds that weren't synthesized under antibiotic-free conditions.[28] Analysis by mass spectrometry and NMR spectroscopy allowed the annotation of over 40 of these molecules, including known compounds and a new group of secondary metabolites named acybolins.[28] Further studies allowed more of the mechanism by which TMP regulates malleilactone biosynthesis and global metabolism in *B. thailandensis* to be revealed.

As TMP inhibits dihydrofolate reductase (DHFR), which acts in primary metabolism to catalyze the formation of tetrahydrofolate (THF) from dihydrofolate (DHF) leading to synthesis of DNA, RNA, and proteins by installation of one-carbon units from downstream THF metabolites, subinhibitory concentrations of TMP could have widespread regulatory and metabolic effects **(Figure 1.4)**. In the case of malleilactone, the addition of TMP causes inhibition of DHFR leading to accumulation of homoserine, which activates expression of the gene cluster, independently of cellular stress responses but requiring decreased DHFR activity **(Figure 1.4)**. Beyond regulation of a single cryptic pathway, TMP has regulatory effects on about 500 genes in the *B. thailandensis* genome, leading to significant changes in its transcriptome and metabolome.[29]

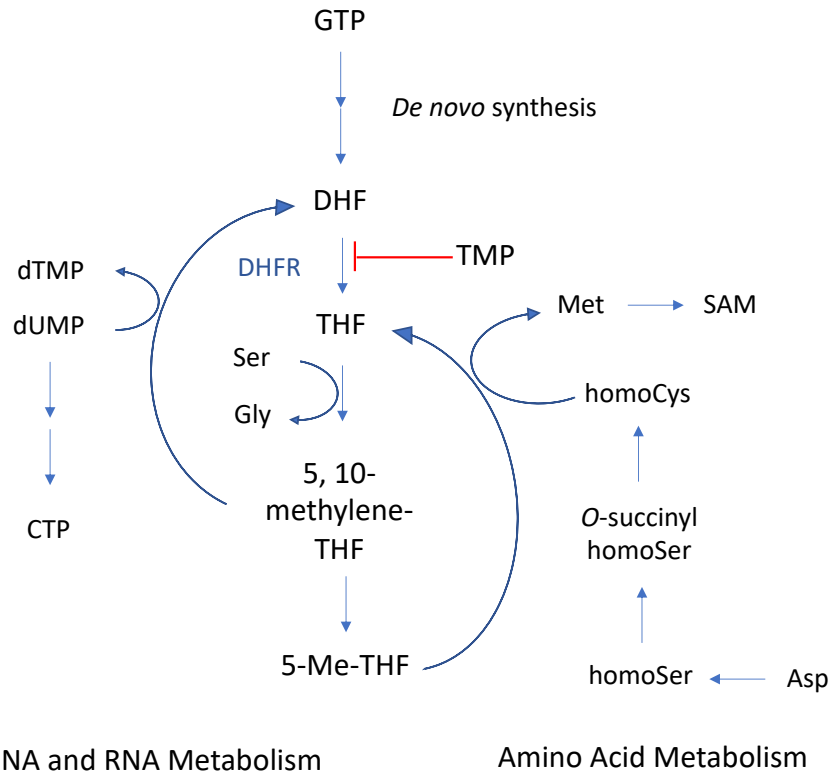

Figure 1.4. The role of dihydrofolate reductase (DHFR) in primary metabolism in *Burkholderia thailandensis*. When trimethoprim (TMP) is added, homoserine accumulates and causes changes in

transcriptional regulation (GTP = guanosine triphosphate; dTMP = deoxythymidine monophosphate; dUMP = deoxyuridine monophosphate; CTP = cytidine triphosphate; DHF = dihydrofolate; THF = tetrahydrofolate; Ser = serine; Gly = glycine; Met = methionine; Asp = aspartate; SAM = *S*-adenosylmethionine).

Recent studies by the Seyedsayamdost group focus on identifying biological activities of interest induced by the presence of an elicitor, with screening of more diverse compound libraries to capitalize on unpredictable microbe-drug interactions. Target strains were cultured in the presence of elicitors and the supernatants were screened for activity against *Escherichia coli* and *Bacillus subtilis*. Biological activity against *E. coli* was observed by growing *Streptomyces hiroshimensis* with the beta-blocker atenolol (AT), resulting in induction of global metabolism and the characterization of pyridindolol, hiroshidine, two isocoumarins, and two taylorflavins belonging to the toxoflavin family of molecules. With AT showing no growth inhibition of the target *Streptomyces* strain and therefore not following principles of hormesis, the mechanism by which it regulates metabolism is more difficult to ascertain.[30]

Metabolomics offers a method for analyzing microbe-elicitor interactions with many advantages over currently used methods. Metabolomics enables identification of changes seen in both abundant and trace components of the metabolome and provides a statistical significance for these individual changes in abundance. Other methods such as the common practise of overlaying High-Performance Liquid Chromatography (HPLC) UV traces rely on the researcher to identify points of interest visually, this both biasing the analysis and lacking statistical rigour. For example, some peaks that are significantly upregulated create a very small UV peak compared to highly abundant highly UV active molecules, and are difficult to pinpoint with UV trace overlays. UV trace overlays also depend on the studied molecules having a strong

UV signal, which is not guaranteed. In addition, the use of biological replicates ensures that the changes seen in the bacterial metabolome are as a result of the presence of the elicitor rather than contamination with another molecule or small differences in culture conditions. To induce expression of interesting and possibly unknown natural products, we designed a metabolomics study based off of the work on TMP and AT **(Figure 1.5)** by the Seyedsayamdost group. These two compounds were used as elicitors against two *Streptomyces* strains, *S* .sp. PBH53 and *S. armeniacus*.

Trimethoprim Atenolol

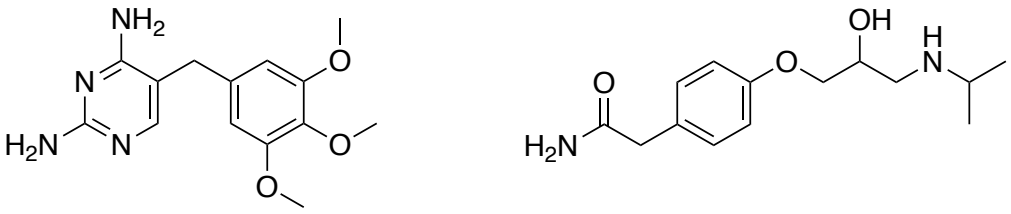

Figure 1.5. Structures of trimethoprim (TMP) and atenolol (AT) used as elicitors of secondary metabolism in bacteria.

Streptomyces are Gram-positive bacteria with high G+C content genomes. They are unique among bacteria as they have a complex lifestyle resembling that of fungi **(Figure 1.6)**. Vegetative hyphae form below the surface of the substrate where DNA replication occurs without significant cell division, then aerial hyphae branch off and extend above the surface. The aerial hyphae begin to septate for sporulation to occur and the spores are dispersed from the hyphae for germination and development of vegetative hyphae again.[31],[32]

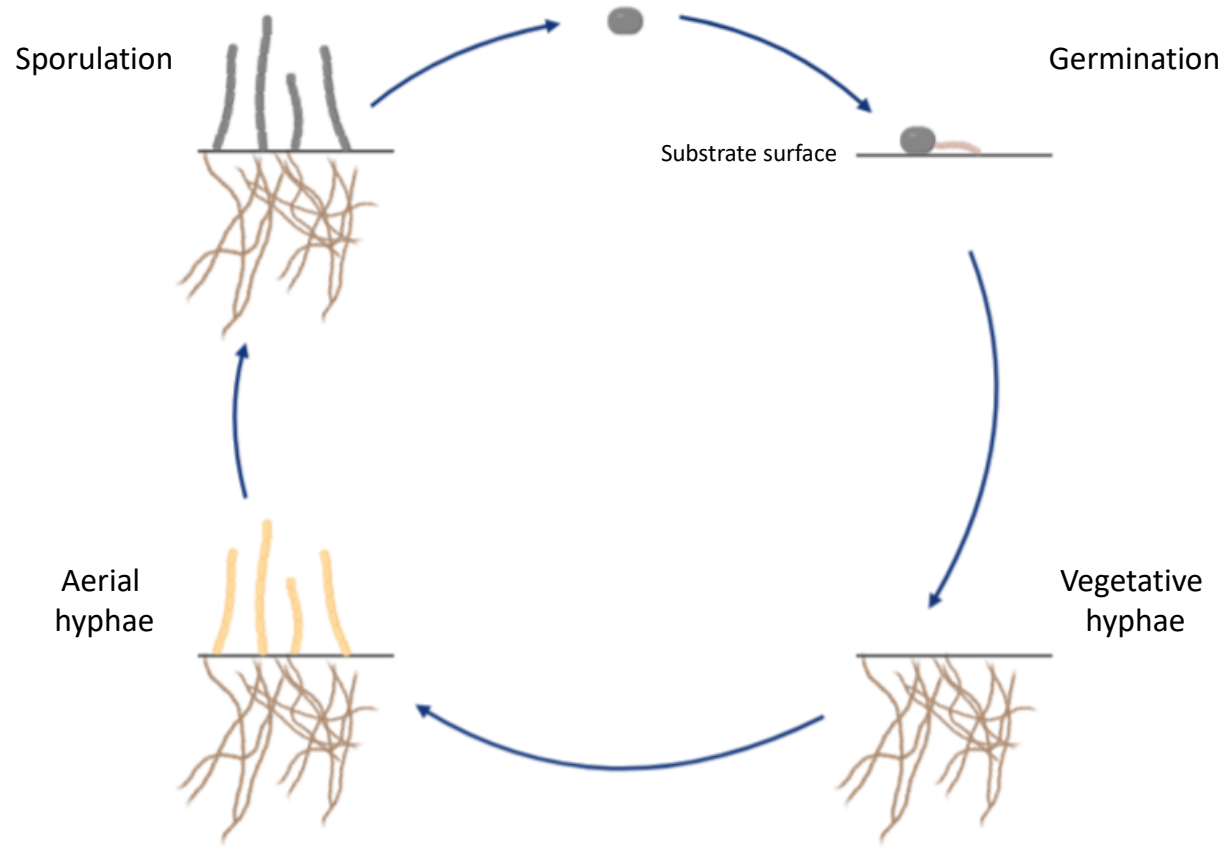

Figure 1.6. The *Streptomyces* life cycle. A spore germinates on the surface of the substrate and develops vegetative hyphae beneath the surface of the substrate. Aerial hyphae are then produced, and septate. Sporulation occurs at the aerial hyphae and the spores are dispersed for germination again.

Streptomyces represent an important source of natural products, and are known to be prolific producers of antibiotic compounds.[33] *S.* sp. PBH53 is an urban strain that was isolated from the campus transit station in Ottawa, Ontario. Its genome was sequenced, which displayed its capacity for secondary metabolite production with the identification of 47 putative BGCs by antiSMASH.[34] *S. armeniacus* was isolated from soil in Armenia, and is most known for its ability to produce the previously unknown antibiotic armeniaspirol.[35] Genome sequencing revealed 36 BGCs predicted by antiSMASH, many of which have no known associated metabolite.[36] These two *Streptomyces* strains have significant potential as producers of natural

products, and were therefore studied for changes in their secondary metabolome in response to the addition of TMP and AT elicitors for the purpose of activating cryptic gene clusters and discovering novel natural products.

1.2 References

[1] Newman, D.J., and Cragg, G.M. (2016). Natural Products as Sources of New Drugs from 1981 to 2014. J. Nat. Prod. *79,* 629-661.

[2] Newman D.J., and Cragg, G.M. (2020). Natural Products as Sources of New Drugs Over Nearly Four Decades from 01/1981 to 09/2019. J. Nat Prod. *83,* 770-803.

[3] Rutledge, P.J., and Challis, G.L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. Nat. Rev. Microbiol. *13*, 509–523.

[4] Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M.H., and Weber, T. (2019). antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res. *47*, W81–W87.

[5] Zarins-Tutt, J.S., Barberi, T.T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D.J., and Goss, R.J.M. (2016). Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. Nat. Prod. Rep. *33*, 54–72.

[6] Bode, H.B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. Chem BioChem. *3*, 619–627.

[7] Timmermans, M.L., Picott, K.J., Ucciferri, L., and Ross, A.C. (2019). Culturing marine bacteria from the genus *Pseudoalteromonas* on a cotton scaffold alters secondary metabolite production. MicrobiologyOpen *8*, e00724.

[8] Ji, C.H., Kim, J.P., and Kang, H.S. (2018). Library of Synthetic Streptomyces Regulatory Sequences for Use in Promoter Engineering of Natural Product Biosynthetic Gene Clusters. ACS Synth. Biol. *7*, 1946–1955.

[9] Ren, H., Hu, P., and Zhao, H. (2017). A plug-and-play pathway refactoring workflow for natural product research *in Escherichia coli* and *Saccharomyces cerevisiae*. Biotechnol. Bioeng. *114*, 1847–1854.

[10] Bode, E., Brachmann, A.O., Kegler, C., Simsek, R., Dauth, C., Zhou, Q., Kaiser, M., Klemmt, P., and Bode, H.B. (2015). Simple “on-demand” production of bioactive natural products. Chembiochem *16*, 1115–1119.

[11] McArthur, M., and Bibb, M.J. (2008). Manipulating and understanding antibiotic production in *Streptomyces coelicolor* A3(2) with decoy oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. *105*, 1020–1025.

[12] Du, D., Katsuyama, Y., Onaka, H., Fujie, M., Satoh, N., Shin-ya, K., and Ohnishi, Y. (2016). Production of a Novel Amide-Containing Polyene by Activating a Cryptic Biosynthetic Gene Cluster in *Streptomyces* sp. MSC090213JE08. ChemBioChem *17*, 1464–1471.

[13] Feng, Z., Kim, J.H., and Brady, S.F. (2010). Fluostatins Produced by the Heterologous Expression of a TAR Reassembled Environmental DNA Derived Type II PKS Gene Cluster. J. Am. Chem. Soc. *132*, 11902–11903.

[14] Kallifidas, D., Kang, H.-S., and Brady, S.F. (2012). Tetarimycin A, an MRSA-active antibiotic identified through induced expression of environmental DNA gene clusters. J. Am. Chem. Soc. *134*, 19552–19555.

[15] Zhang, Y., Lin, C.-Y., Li, X.-M., Tang, Z.-K., Qiao, J., and Zhao, G.-R. (2016). DasR positively controls monensin production at two-level regulation in *Streptomyces cinnamonensis*. J. Ind. Microbiol. Biotechnol. *43*, 1681–1692.

[16] Rigali, S., Titgemeyer, F., Barends, S., Mulder, S., Thomae, A.W., Hopwood, D.A., and van Wezel, G.P. (2008). Feast or famine: the global regulator DasR links nutrient stress to antibiotic production by *Streptomyces*. EMBO Rep. *9*, 670–675.

[17] Liao, C.H., Xu, Y., Rigali, S., and Ye, B.C. (2015). DasR is a pleiotropic regulator required for antibiotic production, pigment biosynthesis, and morphological development in *Saccharopolyspora erythraea*. Appl. Microbiol. Biotechnol. *99*, 10215–10224.

[18] Onaka, H., Mori, Y., Igarashi, Y., and Furumai, T. (2011). Mycolic acid-containing bacteria induce natural-product biosynthesis in *Streptomyces* species. Appl. Environ. Microbiol. *77*, 400–406.

[19] Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., and Kolter, R. (2013). Interspecies Interactions Stimulate Diversification of the *Streptomyces coelicolor* Secreted Metabolome. MBio *4*.

[20] Fajardo, A., and Martínez, J.L. (2008). Antibiotics as signals that trigger specific bacterial responses. Curr. Opin. Microbiol. *11*, 161–167.

[21] Vaz Jauri, P., Bakker, M.G., Salomon, C.E., and Kinkel, L.L. (2013). Subinhibitory antibiotic concentrations mediate nutrient use and competition among soil streptomyces. PLoS One *8*, e81064.

[22] Goh, E.B., Yim, G., Tsui, W., McClure, J., Surette, M.G., and Davies, J. (2002). Transcriptional modulation of bacterial gene expression by subinhibitory concentrations of antibiotics. PNAS *99*, 17025–17030.

[23] Davies, J., Spiegelman, G.B., and Yim, G. (2006). The world of subinhibitory antibiotic concentrations. Curr. Opin. Microbiol. *9*, 445–453.

[24] Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012). Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. Chem. Biol. *19*, 1020–1027.

[25] Tanaka, Y., Izawa, M., Hiraga, Y., Misaki, Y., Watanabe, T., and Ochi, K. (2017). Metabolic perturbation to enhance polyketide and nonribosomal peptide antibiotic production using triclosan and ribosome-targeting drugs. Appl. Microbiol. Biotechnol. *101*, 4417–4431.

[26] Seyedsayamdost, M.R. (2014). High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. PNAS *111*, 7266–7271.

[27] Xu, F., Nazari, B., Moon, K., Bushin, L.B., and Seyedsayamdost, M.R. (2017). Discovery of a Cryptic Antifungal Compound from *Streptomyces albus* J1074 Using High-Throughput Elicitor Screens. J. Am. Chem. Soc. *139*, 9203–9212.

[28] Okada, B.K., Wu, Y., Mao, D., Bushin, L.B., and Seyedsayamdost, M.R. (2016). Mapping the Trimethoprim-Induced Secondary Metabolome of *Burkholderia thailandensis*. ACS Chem. Biol. *11*, 2124–2130.

[29] Li, A., Mao, D., Yoshimura, A., Rosen, P.C., Martin, W.L., Gallant, É., Wühr, M., and Seyedsayamdost, M.R. (2020). Multi-Omic Analyses Provide Links between Low-Dose Antibiotic Treatment and Induction of Secondary Metabolism in *Burkholderia thailandensis*. MBio *11*.

[30] Moon, K., Xu, F., Zhang, C., and Seyedsayamdost, M.R. (2019). Bioactivity-HiTES Unveils Cryptic Antibiotics Encoded in Actinomycete Bacteria. ACS Chem. Biol. *14*, 767–774.

[31] Klieneberger-Nobel, E. (1947). The life cycle of sporing Actinomyces as revealed by a study of their structure and septation. J. Gen. Microbiol. *1*, 22–32.

[32] Erikson, D. (1955). Loss of Aerial Mycelium and other Changes in Streptomycete Development due to Physical Variations of Cultural Conditions. Microbiology, *13*, 136–148.

[33] Procópio, R.E. de L., Silva, I.R. da, Martins, M.K., Azevedo, J.L. de, and Araújo, J.M. de (2012). Antibiotics produced by *Streptomyces*. Braz. J. Infect. Dis. *16*, 466–471.

[34] Gosse, J.T., Hill, P., Dowd, S.E., and Boddy, C.N. (2015). Draft Genome Sequence of *Streptomyces* sp. Strain PBH53, Isolated from an Urban Environment. Gen Announc. *3*.

[35] Dufour, C., Wink, J., Kurz, M., Kogler, H., Olivan, H., Sablé, S., Heyse, W., Gerlitz, M., Toti, L., Nußer, A., et al. (2012). Isolation and structural elucidation of armeniaspirols A-C: potent antibiotics against gram-positive pathogens. J. Chem. *18*, 16123–16128.

[36] Labana, P., Gosse, J.T., and Boddy, C.N. (2018). Draft Genome Sequence of the Type Strain *Streptomyces armeniacus* ATCC 15676. Microbiol Resour Announc *7*.

Chapter 2: Induction of Secondary Metabolism in Two *Streptomyces* Species Using Elicitors

2.1 Overview

To evaluate the impact of elicitors on *Streptomyces* strains, we treated two strains, *S. armeniacus* ATCC 15676 and *S.* sp. PBH53 with either trimethoprim (TMP) or atenolol (AT), using literature concentrations.[1],[2] Trimethoprim, which exhibits antibiotic activity by inhibition dihydrofolate reductase (DHFR), was selected as an elicitor based on previous studies by the Seyedsayamdost group[1] and due to the fact that it was already present in our chemical inventory. Atenolol, which is non-antibiotic, was also selected due work by the Seyedsayamdost group showing its efficacy as an elicitor in a *Streptomyces* species.[2] *S.* sp. PBH53 was cultured in ISP2 media and treated with either TMP or AT at 33 µM. No significant growth inhibition was observed when exposing the strain to 33 µM TMP or AT. *S. armeniacus* was cultured in Czapek Dox media with 2 g/L of yeast extract added, which proved to be a better media for supporting growth of this bacteria. The addition of yeast extract was required for reliable growth in this media. TMP or AT was added at 30 µM, which did not significantly inhibit growth of *S. armeniacus*. Treatments were carried with 6 replicates of each to ensure good statistical evaluation of the metabolome. Each strain with added elicitor was incubated for a week at 27°C before metabolites were extracted using a hydrophobic capture resin, and organic extracts were eluted and dissolved in methanol for analysis.

The metabolome obtained for each treatment group was analyzed by ultra-high-performance liquid chromatography – high resolution mass spectrometry (UPLC-HRMS). The resulting data were statistically analyzed using Principal Component Analysis (PCA) to visualize the variance between treatments, and to see how individual variables or mass features affect the overall variance. Volcano plots were then used to choose mass features of interest based on fold increase and statistical significance between samples. Presented below is the data analysis and interpretation for the metabolomic datasets for each strain with each elicitor compared to vehicle control, as well as the steps taken to identify compounds of interest from each experiment.

2.2 Results and Discussion

2.2.1 *Streptomyces* sp. PBH53 plus trimethoprim

After the extracts were analyzed by UPLC-HRMS, the noise threshold for the *S.* sp. PBH53 Plus 33 µM trimethoprim (TMP) dataset was set at 9.0E5 during data preprocessing, and the data were normalized by median and pareto scaled. PCA was used to visualize the variance between the treatment groups by collapsing three-dimensional data (retention time, *m/z*, and peak intensity) into principal components (PCs) representing total variance in the dataset **(Figure S1.2, 2.1)**. The Loadings plot shows each individual variable in relation to the established PCs **(Figure 2.1)**. PC 1 accounted for 58.4% of the variation seen in the samples, while PC2 accounted for 16.5% of the variation with a cumulative variation of 74.9% in the first two PCs **(Figure 2.2)**. However, greater separation between Plus TMP and Minus TMP

treatments were seen in PC2 based on the scores plot, and only samples Plus TMP A, B, and C appear to be noticeably different from the Minus TMP treatments **(Figure 2.2)**.

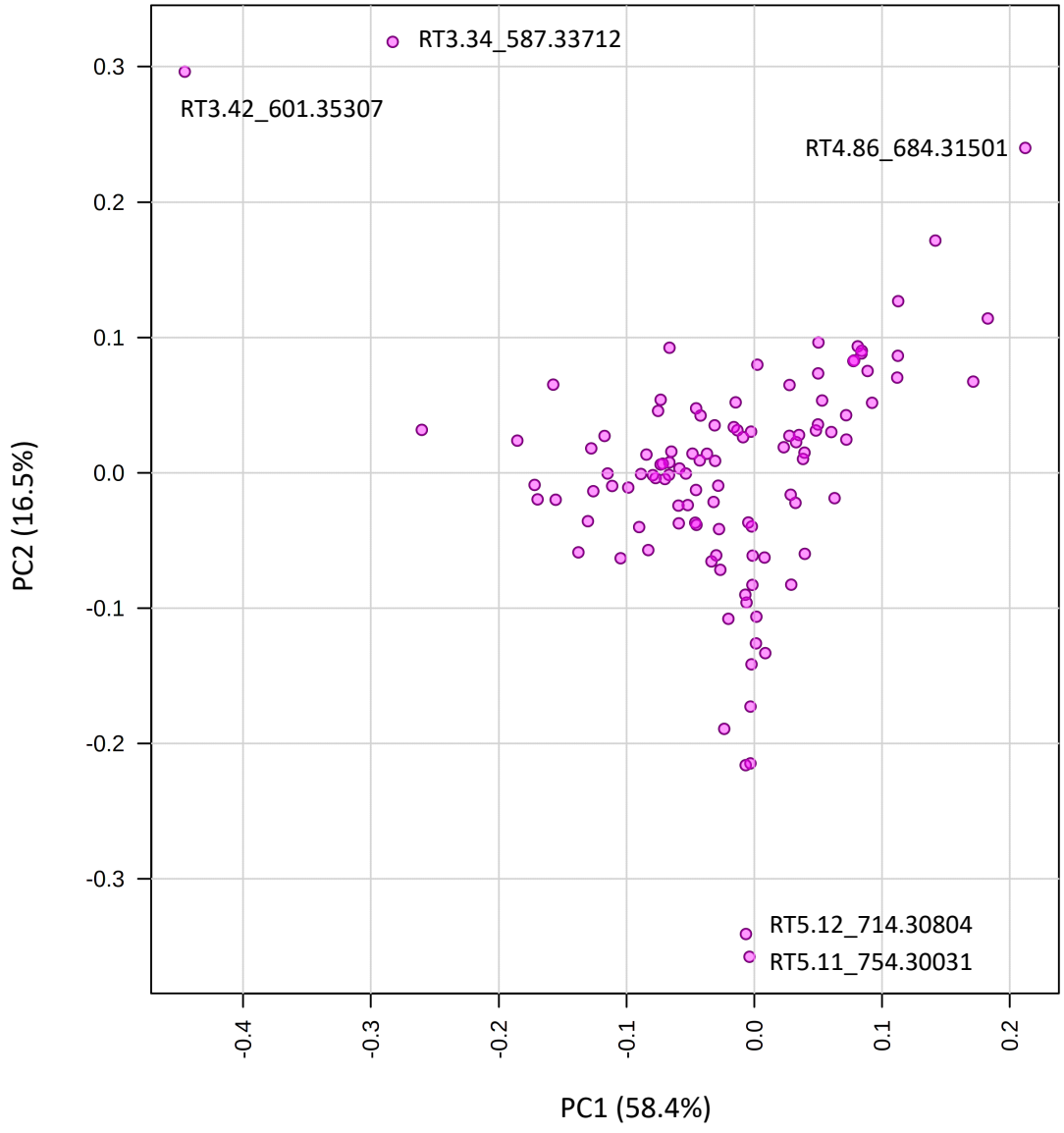

Figure 2.1. Loadings plot of the first two PCs of the PCA for *S.* sp. PBH53 cultured with and without 33 µM TMP. Select variables that contribute to the variance seen in the PCs are labelled.

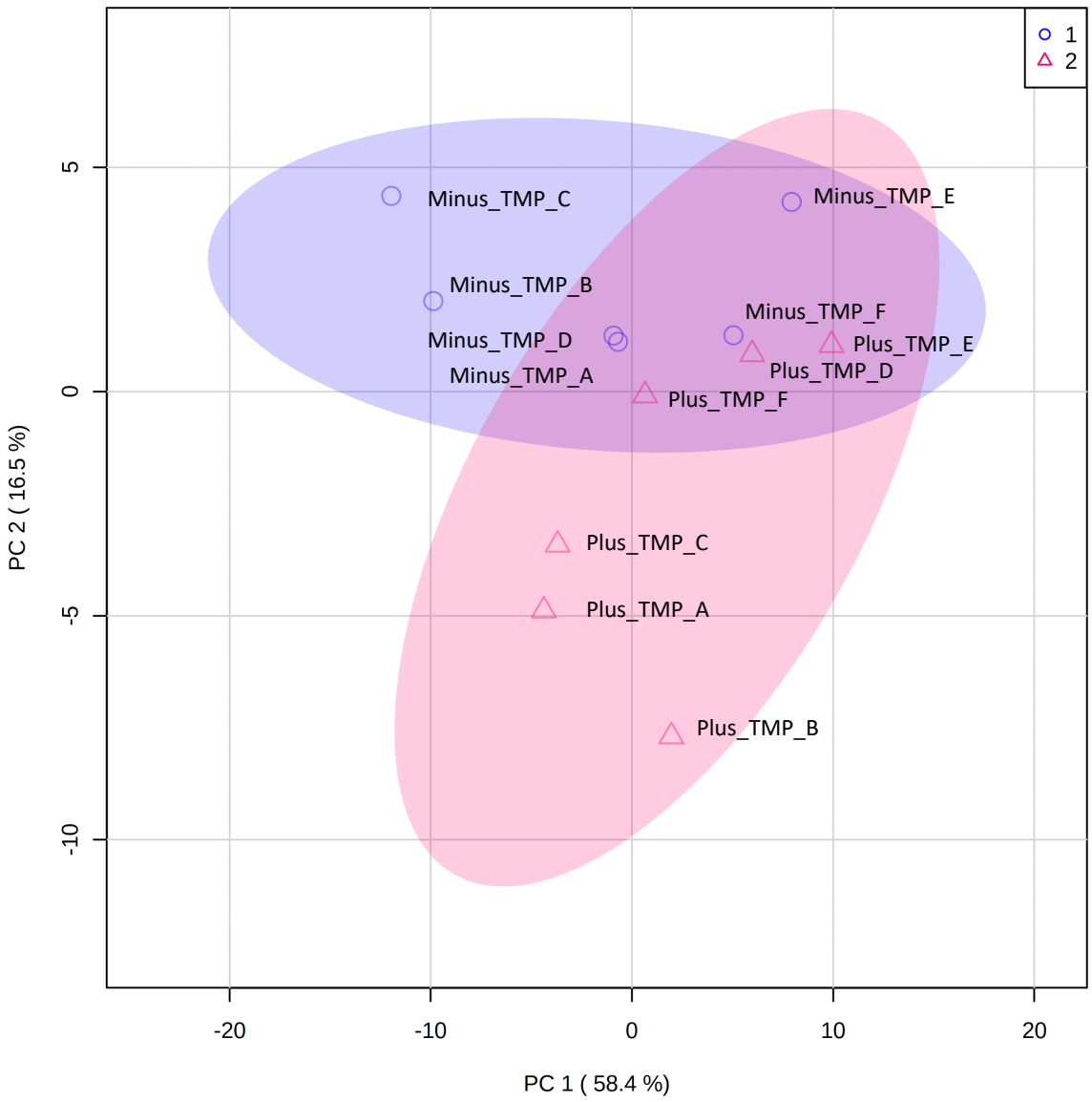

Figure 2.2. PCA scores plot comparing secondary metabolite fingerprints of S. sp. PBH53 cultured without TMP (Blue, Class 1; n=6) and with 33 µM TMP (Red, Class 2; n=6). Shaded areas represent the 95% confidence interval for each treatment.

To choose peaks of interest, a volcano plot was used to identify peaks that were significantly higher ($p<0.05$) in the Plus TMP treatments compared to the Minus TMP treatments and were upregulated at least 2-fold **(Figure 2.3)**. The volcano plot was chosen as it provides both statistical significance in the form of a t-test, and biological significance in the

form of fold change. Six mass features appeared to be significantly higher in the Plus TMP treatments compared to the Minus TMP treatments, all at a retention time of 5.11 or 5.12 minutes **(Figure 2.4)**. Three of these features corresponded to a parent peak with a corrected mass of $[M+H]^+$ 732.3203, representing $[M+H-2H_2O]^+$, $[M+H-H_2O]^+$, and $[M+Na]^+$ **(Figure 2.5).** Two other mass features appeared to be related to each other, with an $[M+H]^+$ of 341.2116 and an $[M+H-H_2O]^+$ of 323.2018. The remaining mass was $[M+H]^+$ 215.1442. $[M+H]^+$ 732 was selected as our compound of interest as the mass of the molecule allows the possibility for interesting structures.

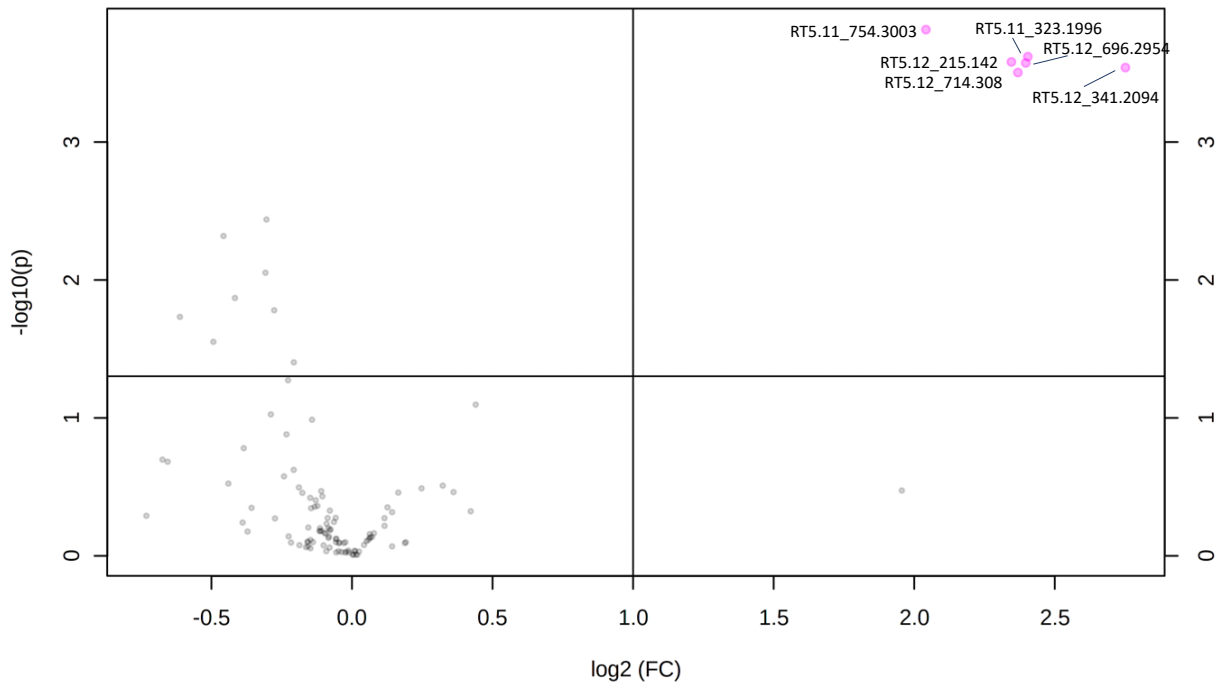

Figure 2.3. Volcano plot of mass features produced by *S*. sp. PBH53 appearing 2-fold greater in the Plus TMP (n=6) treatment than the Minus TMP treatment (n=6), with p-values generated from a t-test. Cut-offs for mass features of interest were set at a 2-fold change (vertical line) and $p<0.05$ (horizontal line). Points of interest are highlighted and labelled (FC = Fold Change).

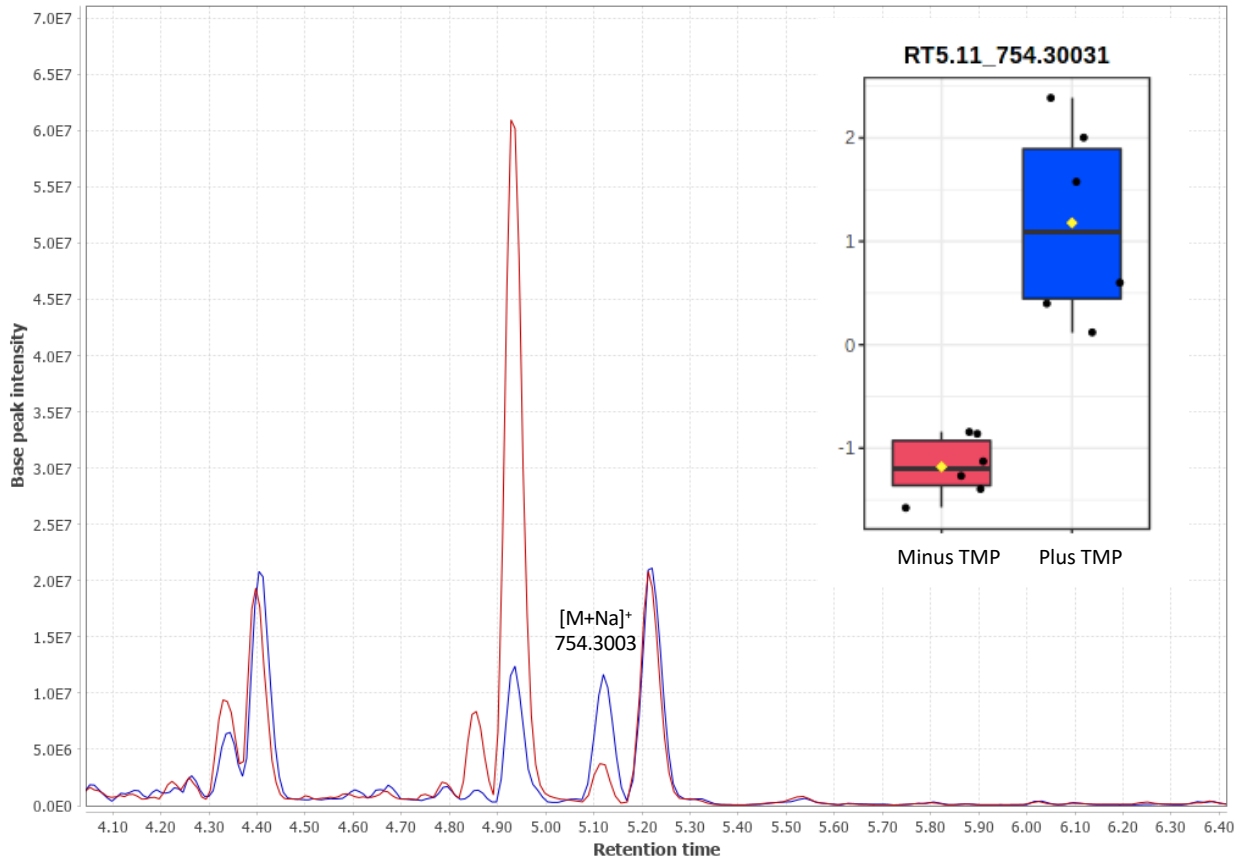

Figure 2.4. Total Ion Chromatogram (TIC) overlay of *S*. sp. PBH53 cultured with 33 µM TMP (blue) and without TMP (red). Data were collected from *m/z* 100-2000. Box plot shows the indicated mass feature in all samples.

.

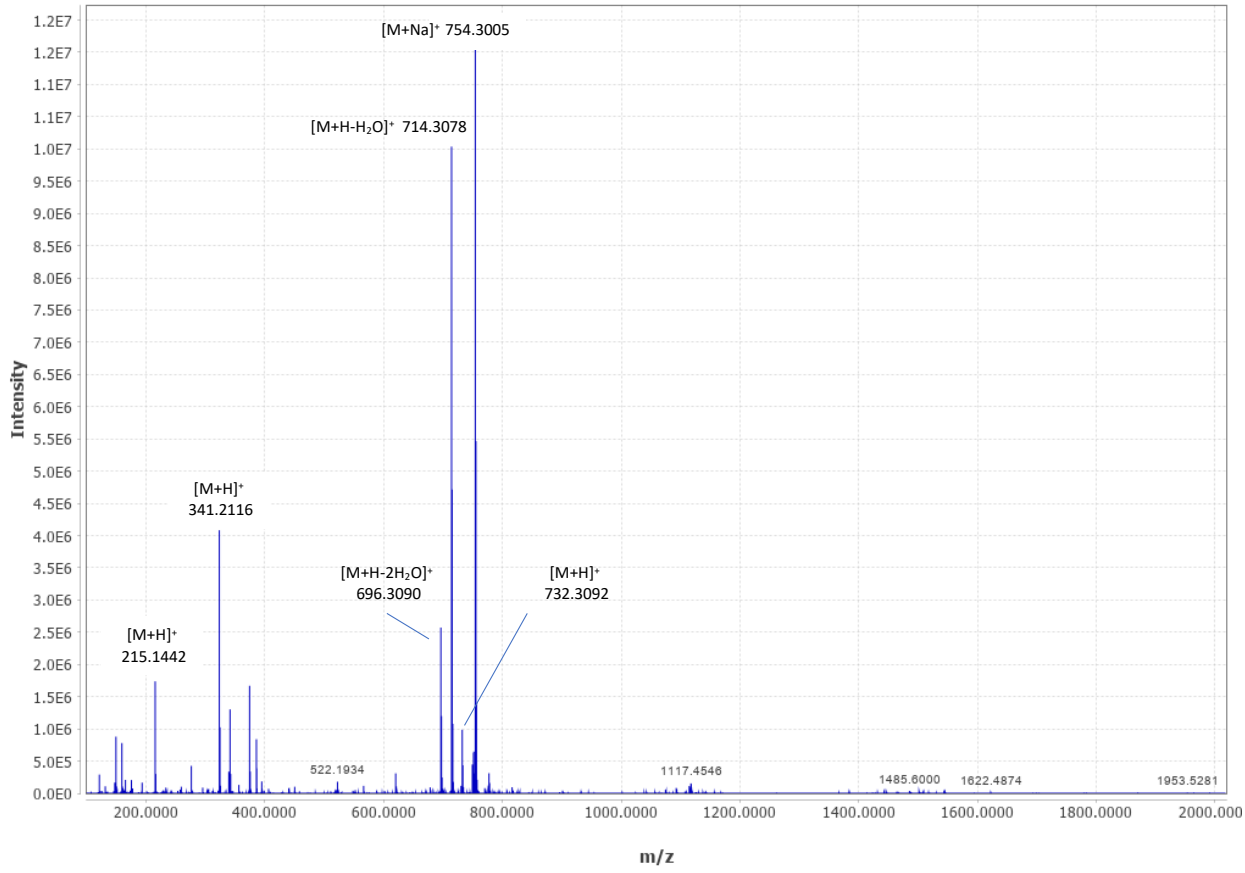

Figure 2.5. Mass spectrum of *S*. sp. PBH53 cultured with 33 µM TMP at a retention time of 5.12 mins. Data were collected between *m/z* 100-2000.

Attempts were made to scale up production of $[M+H]^+$ 732 in larger volumes, but its production was dependent on proper oxygenation and thus no compound was detected in large volume cultures. However, successful scale-ups were accomplished by reducing the volume of liquid in the flask and inoculating a colony into 1.2 L of ISP2 media in a 4 L flask. About 300 mg of crude extract was obtained from 2L of scaled-up cultures. Purification of the scaled-up extracts by HPLC resulted in approximately 300 µg of purified compound which corresponds to a titre of $[M+H]^+$ 732 in the cell culture of 150 µg/L.

From the HRMS we can determine that the molecule likely has at least 2 -OH groups due to the adducts showing 2 neutral losses of water (**Figure 2.5)**. The even value of the $[M+H]^+$ mass also indicates the presence of an odd number of nitrogen atoms. Tandem mass spectrometry (HRMS/MS) experiments were performed on $[M+H]^+$ 732, revealing fragments at *m/z* 536.2251 and 292.0370 **(Figure S1.3)**. The fragmentation pattern is similar to that of the naphthomycin group of compounds, but $[M+H]^+$ 732 could not be identified from searching the MS-Finder database.[3] Attempts were made to elucidate the structure of $[M+H]^+$ 732 by NMR spectroscopy **(Figure S1.4)** but there wasn't enough compound purified to allow for strong enough NMR signals for a structure to be determined. Efforts will be focused on producing larger quantities of $[M+H]^+$ 732 and purifying it for structure elucidation by NMR spectroscopy.

2.2.2 *Streptomyces* sp. PBH53 plus atenolol

The noise threshold for *S.* sp. PBH53 with 33 µM atenolol (AT) was set to 2.5E5 during data preprocessing, and the data were normalized by median and pareto scaled. Sample E of the Plus AT treatment group was identified as an outlier and removed from analysis due to drastic differences in its secondary metabolite profile. PCA was used to visualize the variance in the data. PC1 accounted for 47.4% of the variance seen in the samples, while PC2 accounted for 27.3% of the variation with a cumulative variance of 74.7% in the first two PCs **(Figure S2.2, 2.6)**. There isn't much separation between the Plus AT and Minus AT treatments in these two PCs, indicating that any changes induced by the elicitor are small and are not the main variables driving the variance seen in the first two PCs **(Figure 2.7)**. The Loadings plot of the PCA shows that variables are distributed close to the axes, with a few points having a larger impact on the

variance **(Figure 2.6)**. Based on the PCA, there doesn't seem to be big differences between the Plus and Minus treatments.

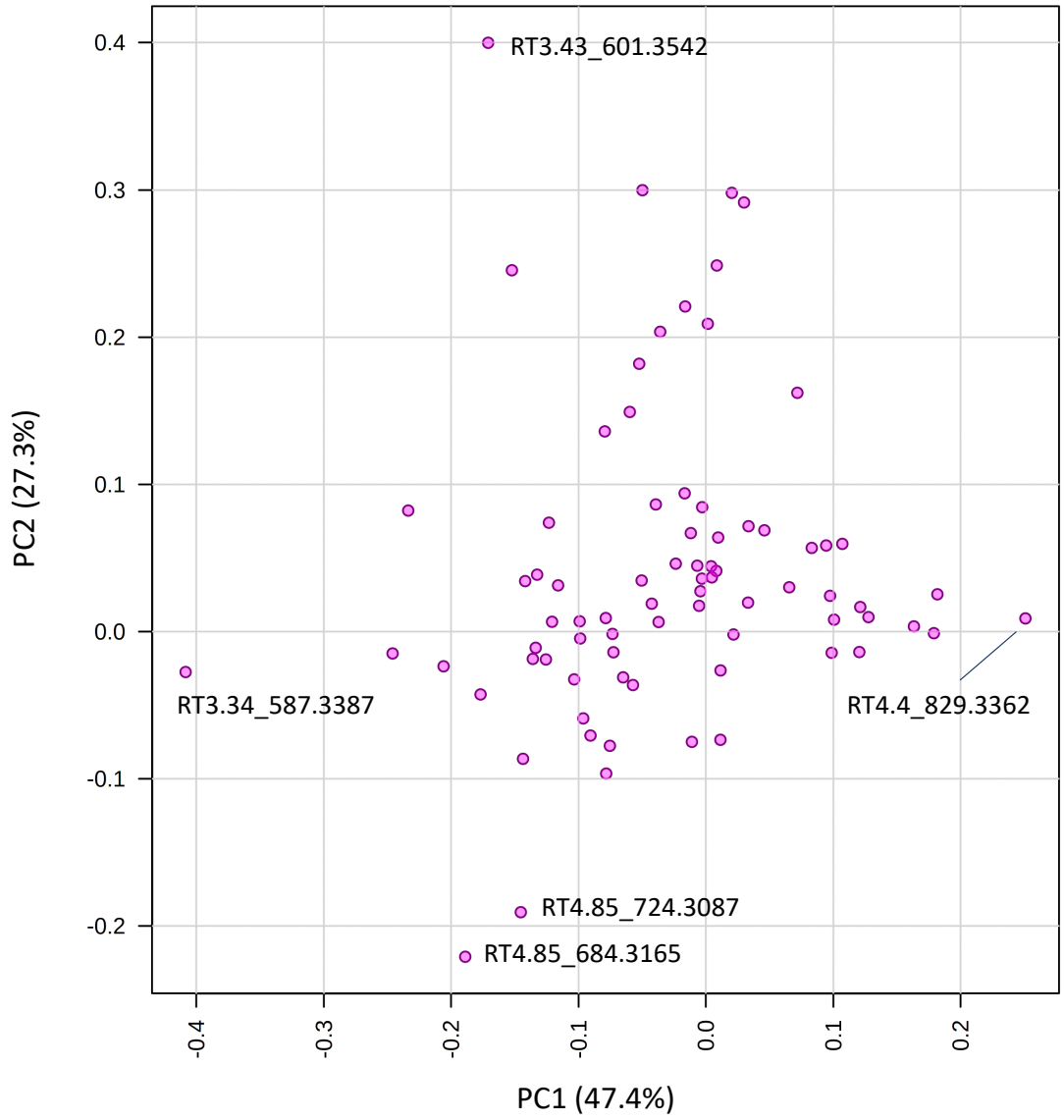

Figure 2.6. Loadings plot of the first two PCs of the PCA for *S*. sp. PBH53 cultured with and without 33 µM AT. Select variables that contribute to the variance seen in the PCs are labelled.

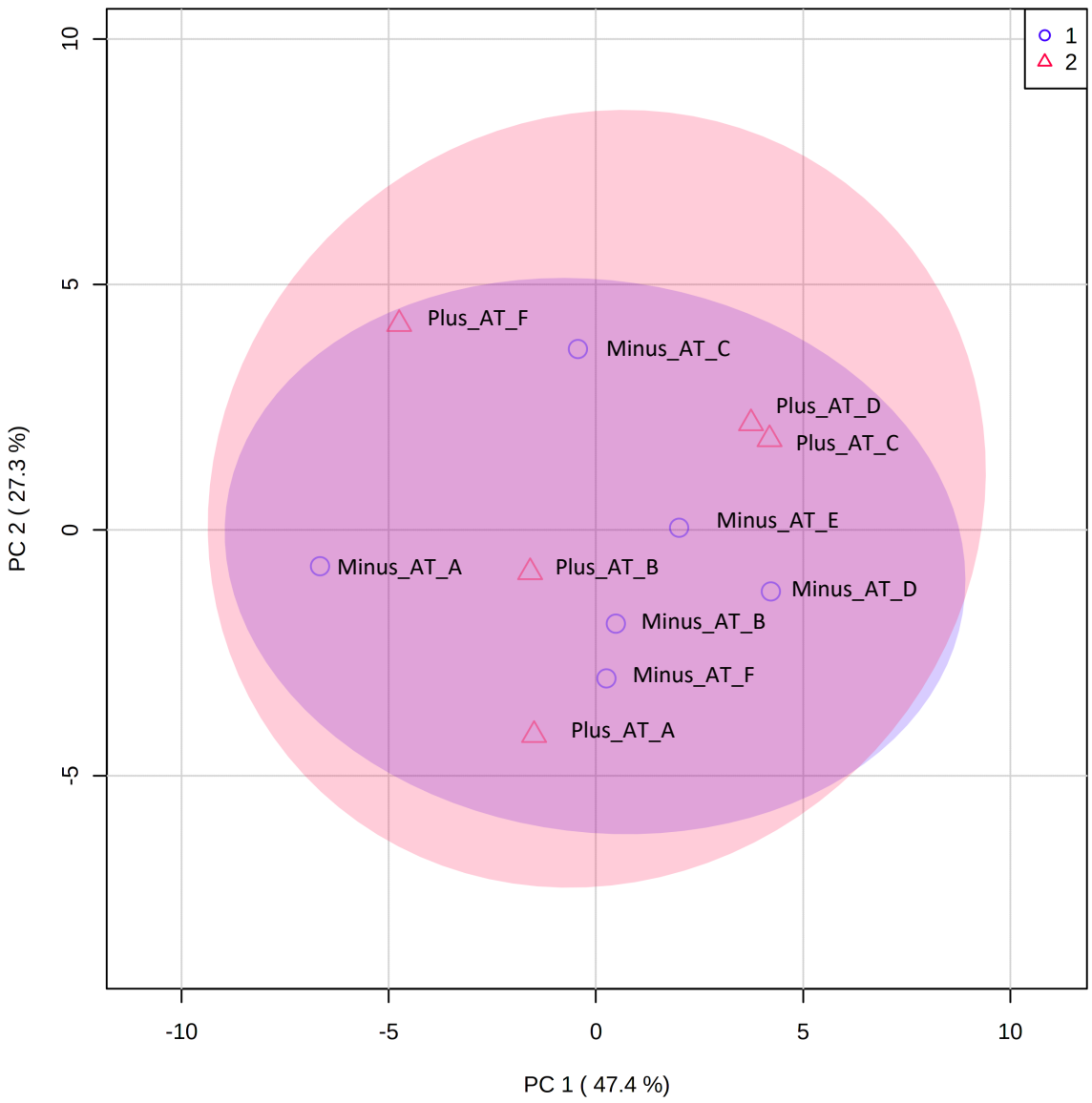

Figure 2.7. PCA scores plot comparing secondary metabolite fingerprints of *S*. sp. PBH53 cultured without AT (Blue, Class 1; n=6) and with 33 µM AT (Red, Class 2; n=5). Shaded areas represent the 95% confidence interval for each treatment.

Two mass features appeared as points of interest from the volcano plot with increased expression in the plus atenolol treatments **(Figure 2.8)**. Both features occurred at a retention time of 4.92 minutes, indicating that the two masses could be related. The first was an $[M+Na]^+$ of 773.4953 **(Figure 2.9)**. The second peak was an $[M+H]^+$ of 393.288. We focused our efforts on $[M+Na]^+$ 773, which had an associated $[M+H]^+$ adduct with a mass of 751.5139 **(Figure 2.10)**.

These masses were queried in the Global Natural Products Social Networking (GNPS) Public Spectral Libraries[4] and the Natural Products Atlas (NP Atlas)[5], which revealed a potential match of salinomycin **(Figure 2.11)**. Salinomycin is an ionophore produced by *Streptomyces* that exhibits anticancer activity by sensitizing tumor cells and inhibiting the anticancer drug efflux protein, P-glycoprotein.[6] Salinomycin also has antibacterial activity against Gram-positive bacteria, including methicillin-resistant *Staphylococcus aureus.*[7]

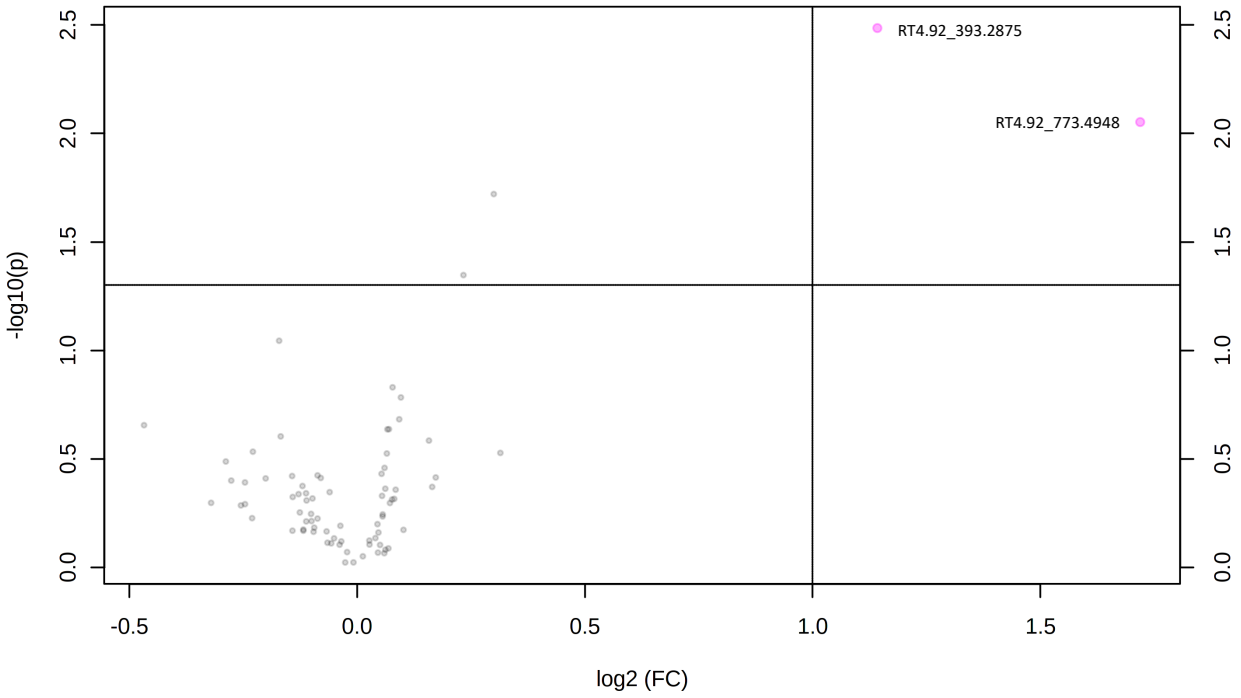

Figure 2.8. Volcano plot of features produced by *S*. sp. PBH53 appearing 2-fold greater in the Plus AT treatment (n=5) than the Minus AT treatment (n=6), with p-values generated from a t-test. Cut-offs for points of interest were set at a 2-fold change (vertical line) and p<0.05 (horizontal line). Points of interest are highlighted and labelled (FC = Fold Change).

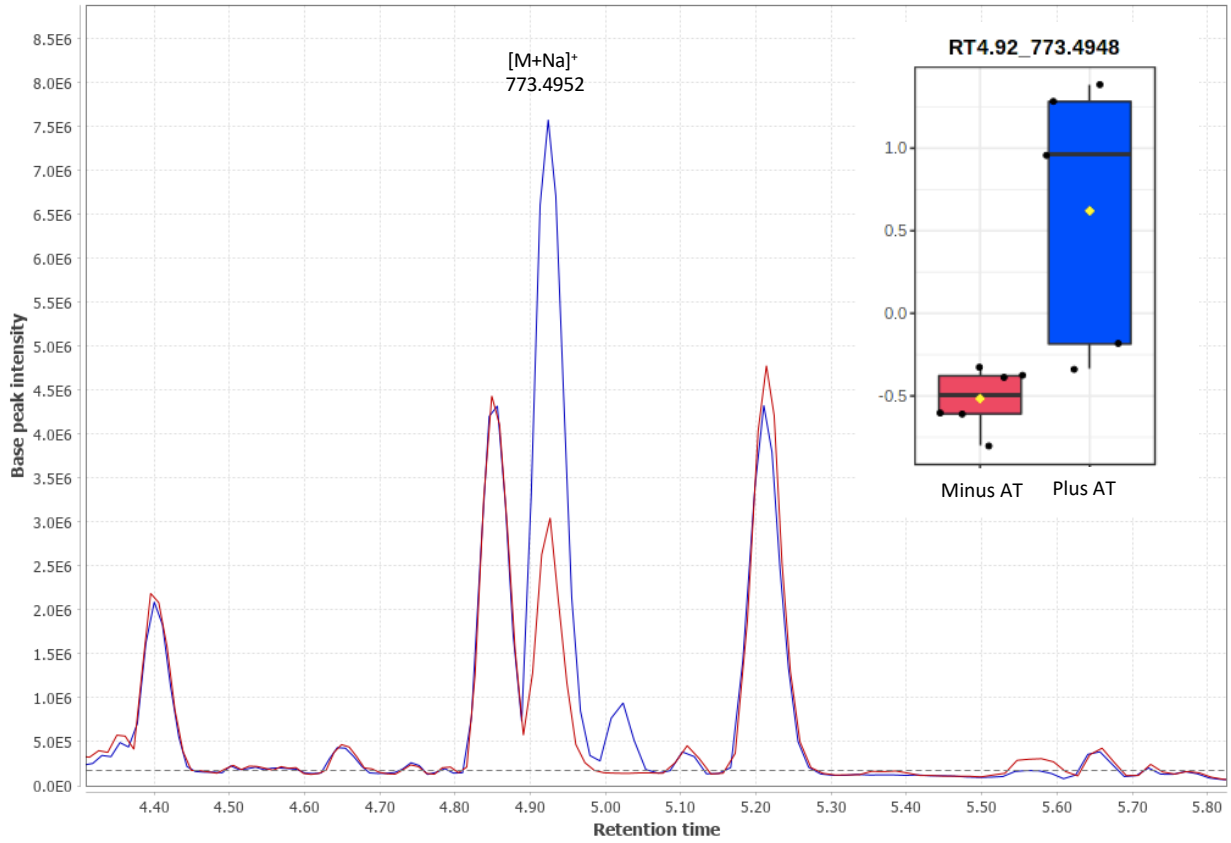

Figure 2.9. TIC overlay of *S.* sp. PBH53 cultured with 33 µM AT (red) and without AT (blue). Data were collected between *m/z* 100-2000. Box plot shows the indicated mass feature in all samples.

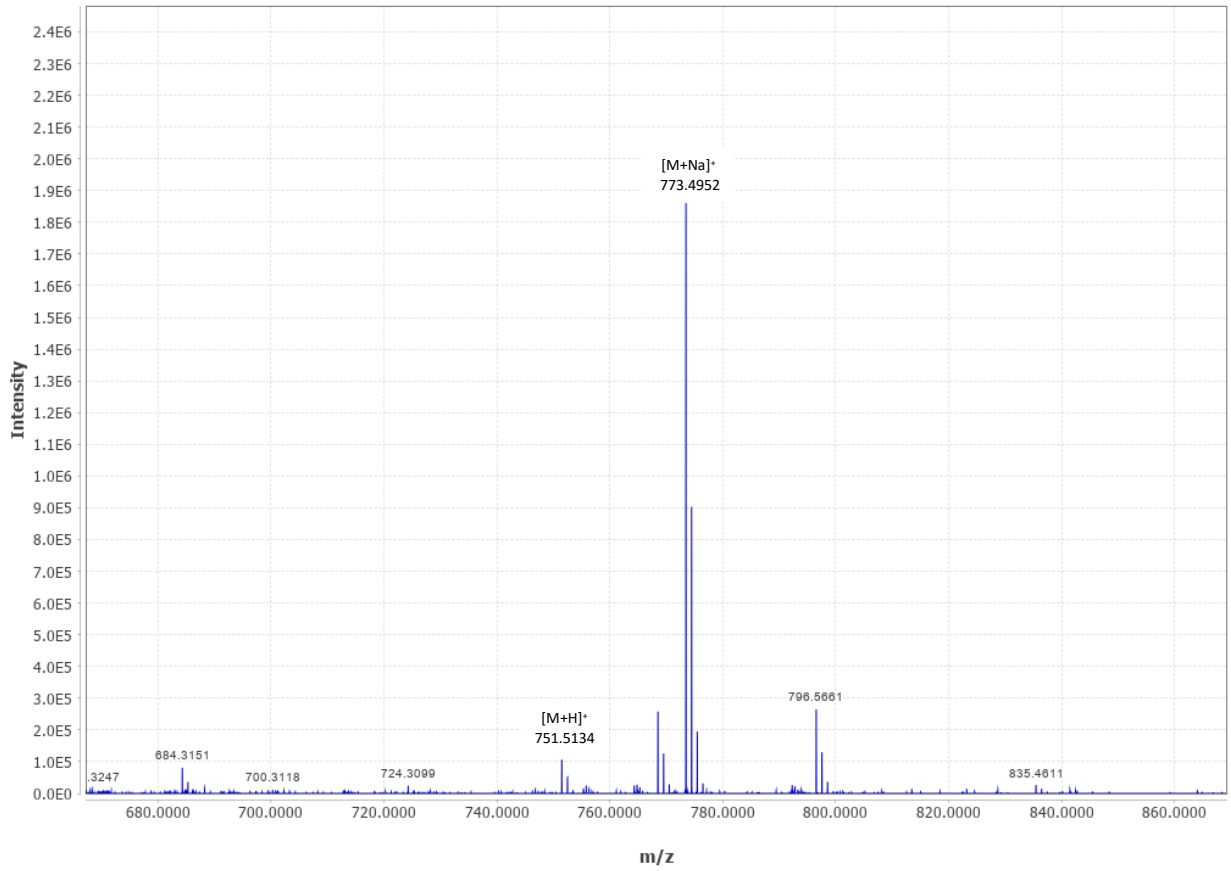

Figure 2.10. Mass spectrum of *S*. sp. PBH53 cultured with 33 µM AT at a retention time of 4.92 mins.

Data were collected between *m/z* 100-2000.

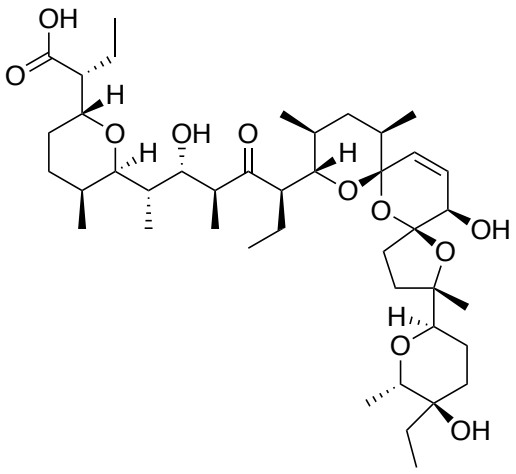

Figure 2.11. The structure of salinomycin, the proposed identity of $[M+Na]^+$ 773 identified in extracts of *S.* sp. PBH53 with AT.

To confirm that the compound produced by *S.* sp. PBH53 is salinomycin, we used LC-MS to compare its retention time and mass to a salinomycin standard. We found the elution of the salinomycin standard occurred at 4.93 mins, which matched the retention time of $[M+H]^+$ 751 in our sample **(Figure 2.12)**. We also saw the same retention times for $[M+Na]^+$ 773 in both the sample and the standard. In addition, we saw the other mass from the volcano plot, *m/z* 393, also present in the standard, indicating that it is likely a fragment of salinomycin. These results led us to confirm the production of salinomycin by *S.* sp. PBH53, completing our analysis of this experiment.

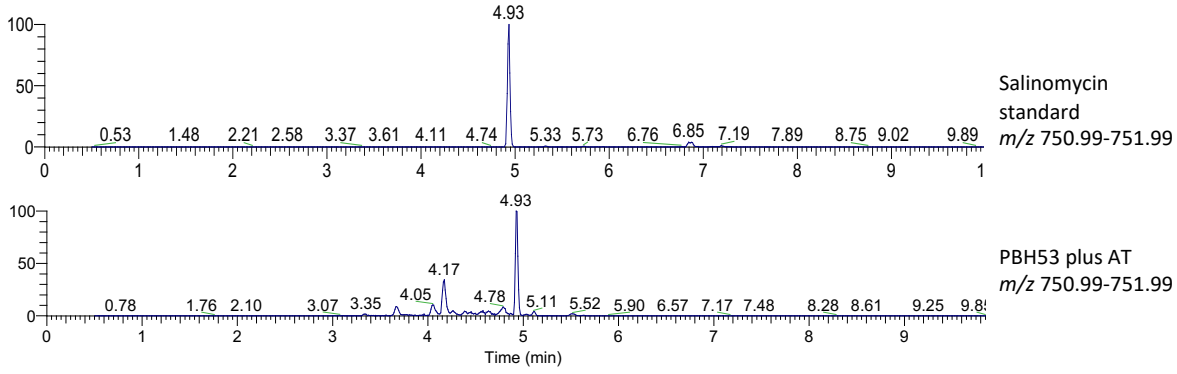

Figure 2.12. Extracted ion chromatogram for *m/z* 750.99-751.99 collected by UPLC-HRMS, corresponding to $[M+H]^+$ 751 as the protonated mass of salinomycin. Both the salinomycin standard and the sample show a peak at 4.93 mins with the protonated mass being present.

2.2.3 *Streptomyces armeniacus* plus trimethoprim

The noise threshold for *S. armeniacus* with 30 µM TMP was set at 5.0E5. The data was normalized by median and pareto scaled for analysis. PCA was used to visualize the variance in the dataset. Sample D without TMP was removed from analysis as it appeared to be an outlier. PC1 represent 82.3% of the variance in the samples, while PC2 represents 10% of the variance

(Figure S3.2). The combined variance of the first two PCs is 92.3%. The Plus and Minus treatments separate from each other along PC1, but variance within the Plus TMP treatments separate along PC2 **(Figure 2.13)**. The cluster of points in the left hand side of the loadings plot indicates features corresponding to the Minus TMP treatments, while the variables that appear farther to the right of the plot are influencing the separation of the Plus TMP samples across PC1 (**Figure 2.13, 2.14**)

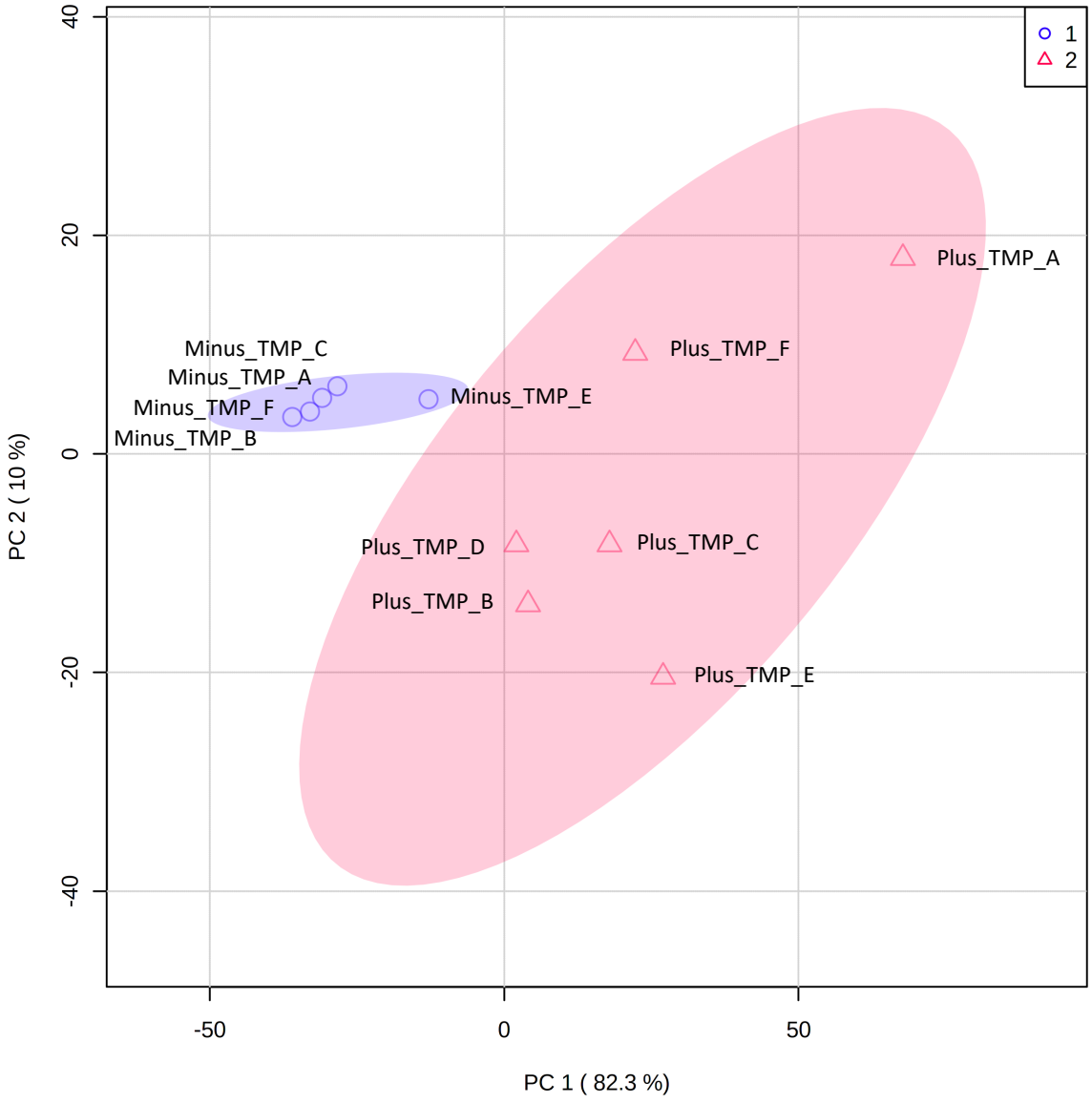

Figure 2.13. PCA scores plot comparing secondary metabolite fingerprints *of S. armeniacus* cultured without TMP (Blue, Class 1; n=5) and with 30 µM TMP (Red, Class 2; n=6). Shaded areas represent the 95% confidence interval for each treatment.

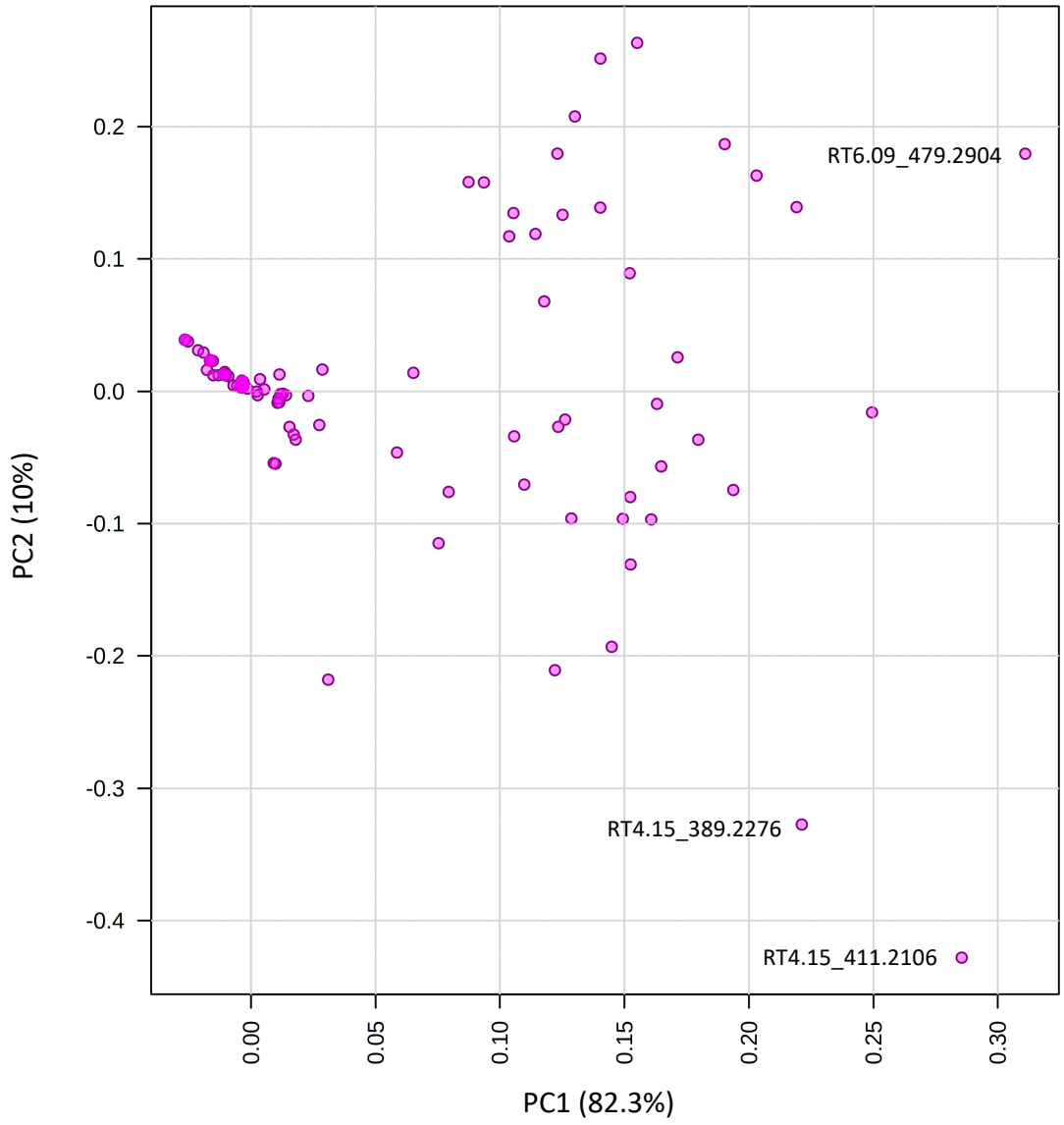

Figure 2.14. Loadings plot of the first two PCs of the PCA for *S. armeniacus* cultured with and without 30 µM TMP. Select features appearing to contribute more to the variance of the PCs are labelled.

Many compounds appeared to be significantly upregulated in the plus TMP treatments according to the volcano plot **(Figure 2.15, Table S3.2)**. Efforts were focused on the mass feature with a mass of 411.2111 based on its apparent high abundance (**Figure 2.16)**. We

determined that this mass represented the $[M+Na]^+$ adduct and had an associated corrected $[M+H]^+$ of 389.2281, as well as masses corresponding to neutral losses of water and -CO (**Figure 2.17)**. The water loss indicates the presence of at least one -OH group, while the -CO loss indicates the presence of a carbonyl. Together these two findings suggest that the molecule likely has a carboxylic acid group. The extract was subject to Kupchan partition and requires further analysis to identify which fraction contains $[M+Na]^+$ 411, and structure elucidation will be carried out by NMR spectroscopy.

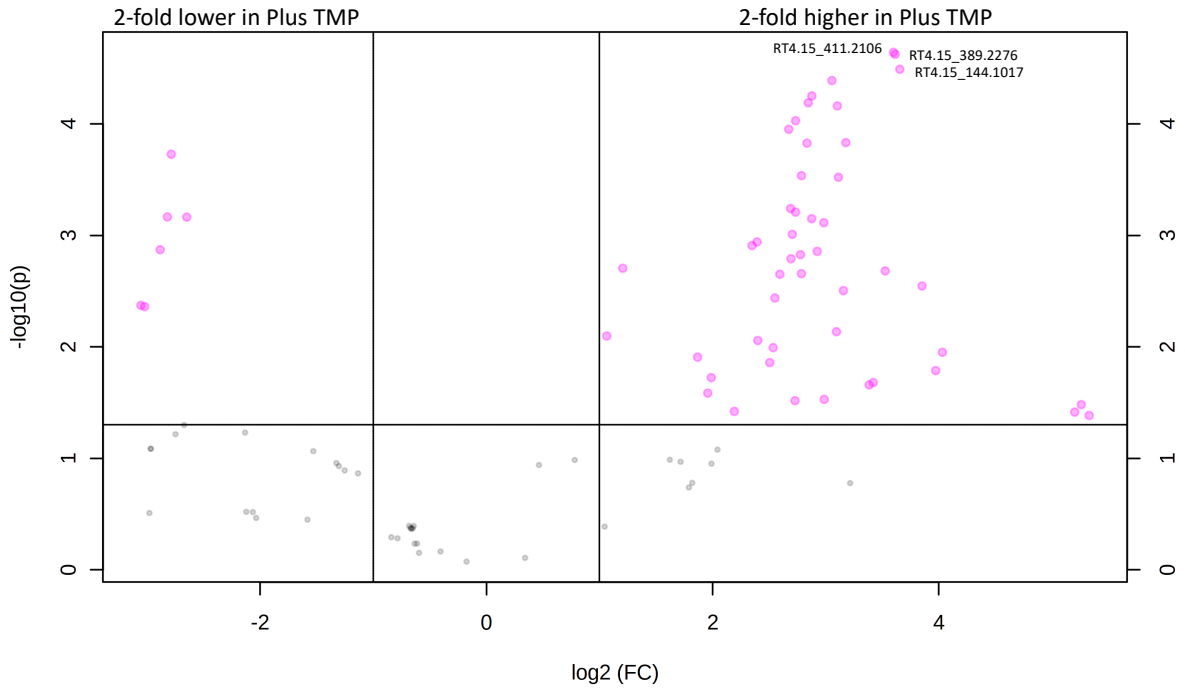

Figure 2.15. Volcano plot of features produced by *S. armeniacus* appearing 2-fold greater in the Plus TMP treatment than the Minus TMP treatment, with p-values generated from a t-test. Cut-offs for points of interest were set at a 2-fold change (vertical lines) and $p<0.05$ (horizontal line). Points of interest are highlighted and labelled (FC = Fold Change).

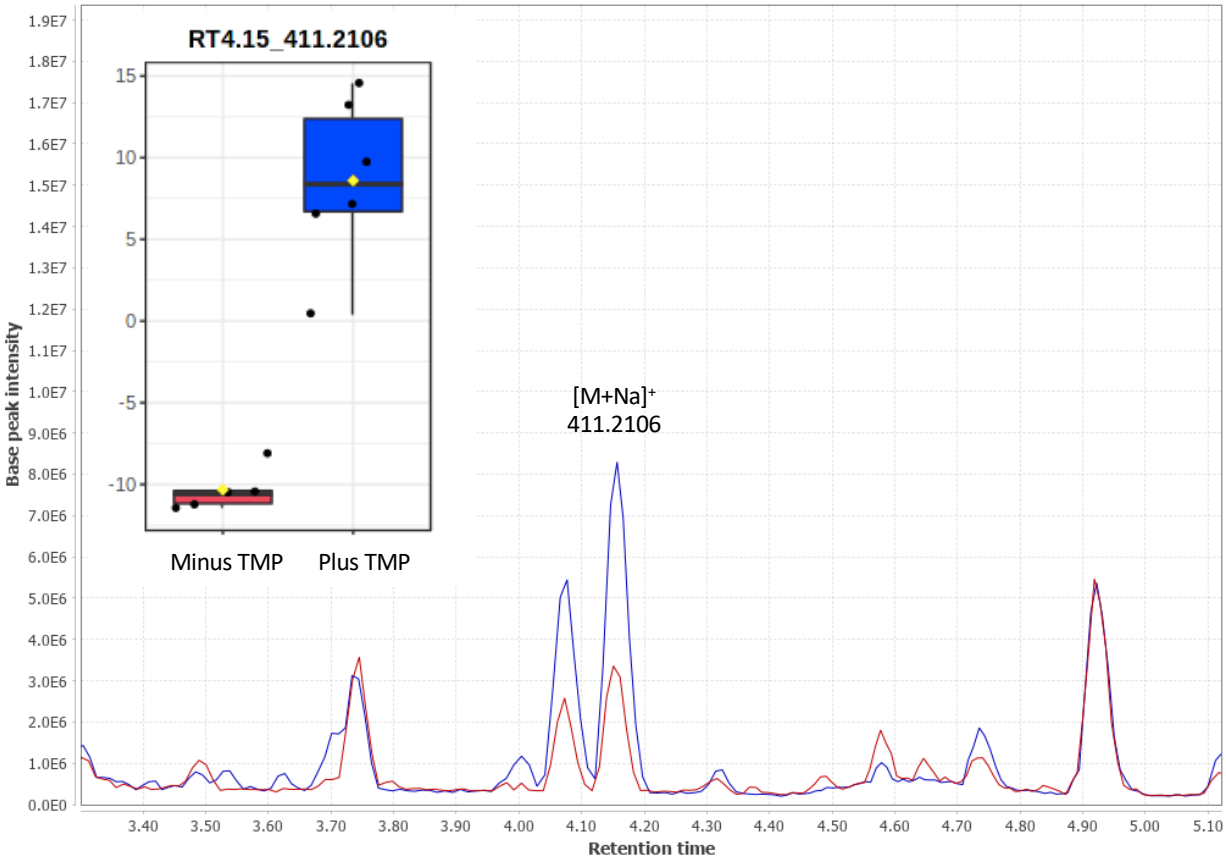

Figure 2.16. TIC overlay of *S. armeniacus* cultured with 30 µM TMP (blue) and without TMP (red). Data were collected between *m/z* 100-2000. Box plot shows the indicated mass feature in all samples.

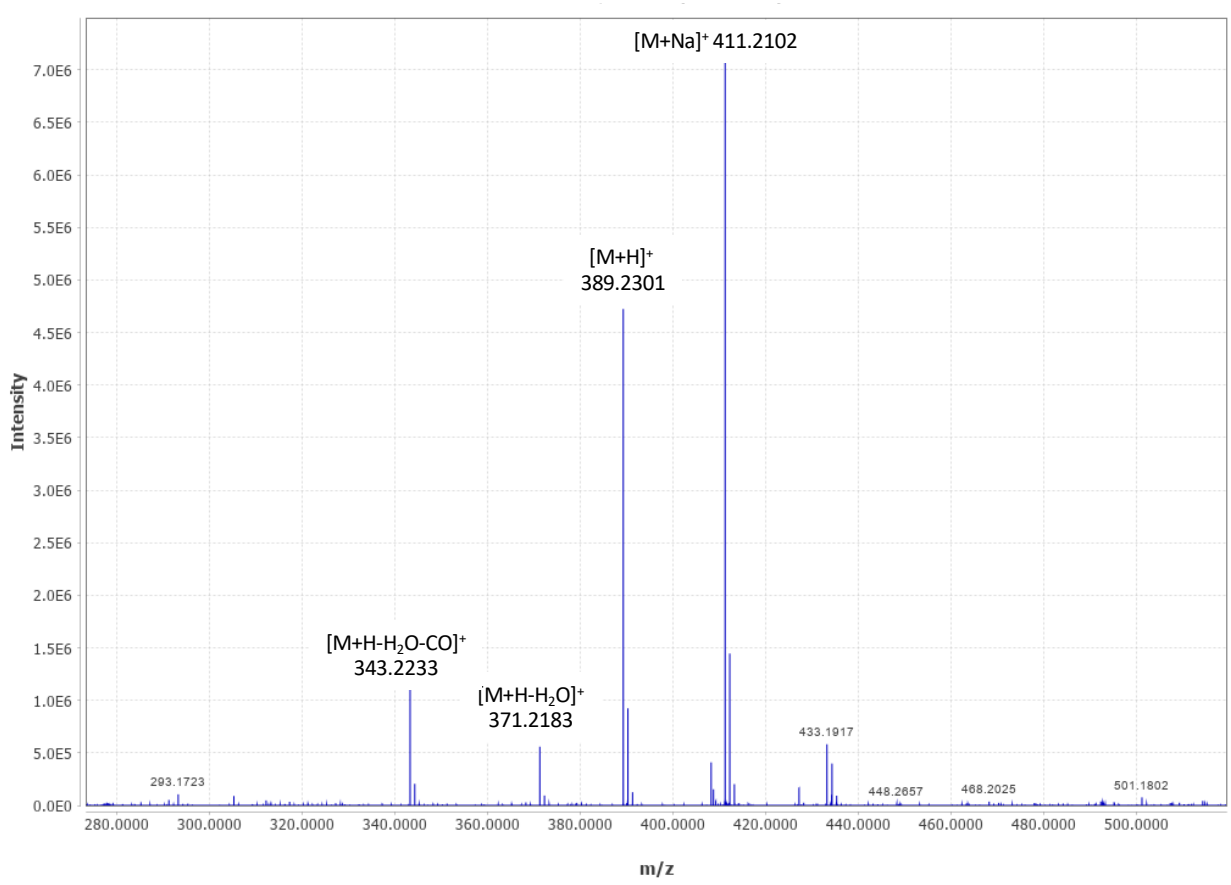

Figure 2.17. Mass spectrum of *S. armeniacus* cultured with 30 µM TMP at a retention time of 4.15 mins. Data were collected between *m/z* 100-2000.

2.2.4 *Streptomyces armeniacus* plus atenolol

The noise threshold for *Streptomyces armeniacus* with 30 µM AT was set at 7E5, and the data were normalized by median and pareto scaled. PCA was used to visualize the variance in the dataset. PC1 represent 49% of the variance in the samples, while PC2 represents 22% of the variance **(Figure S4.2)**. The combined variance of the first two PCs is 71%, Several samples in the scores plot show overlap between the Plus and Minus TMP treatments, but it also appears that the Plus AT samples that represent most of the variance are A, B, and D **(Figure 2.18)**. The

loadings plot the points dispersed across the axes, with a few points appearing at the peripheries of the axes **(Figure 2.19)**.

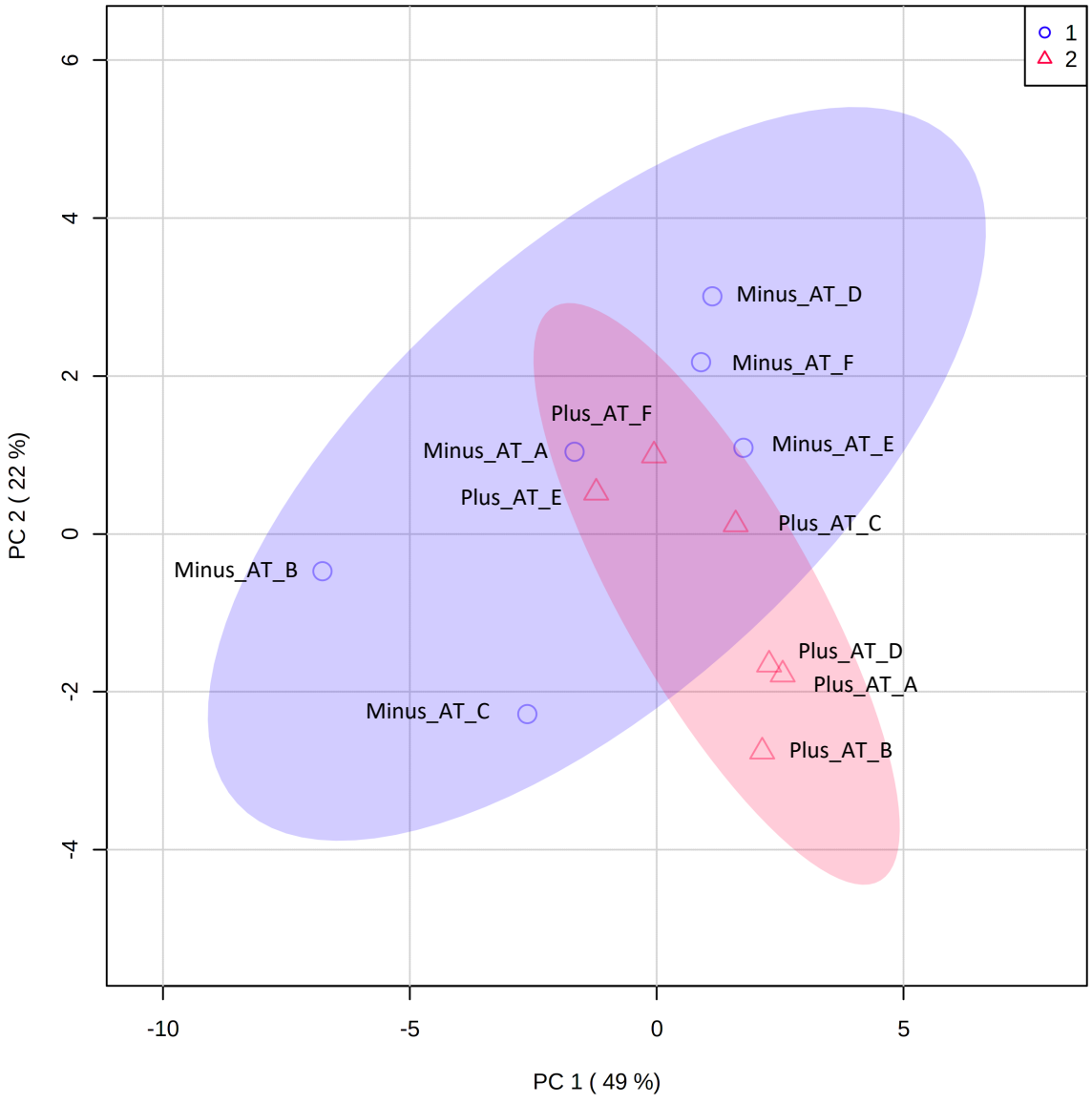

Figure 2.18. PCA scores plot comparing secondary metabolite fingerprints of *S. armeniacus* cultured without AT (Blue, Class 1; n=6) and with 33 µM AT (Red, Class 2; n=6). Shaded areas represent the 95% confidence interval for each treatment.

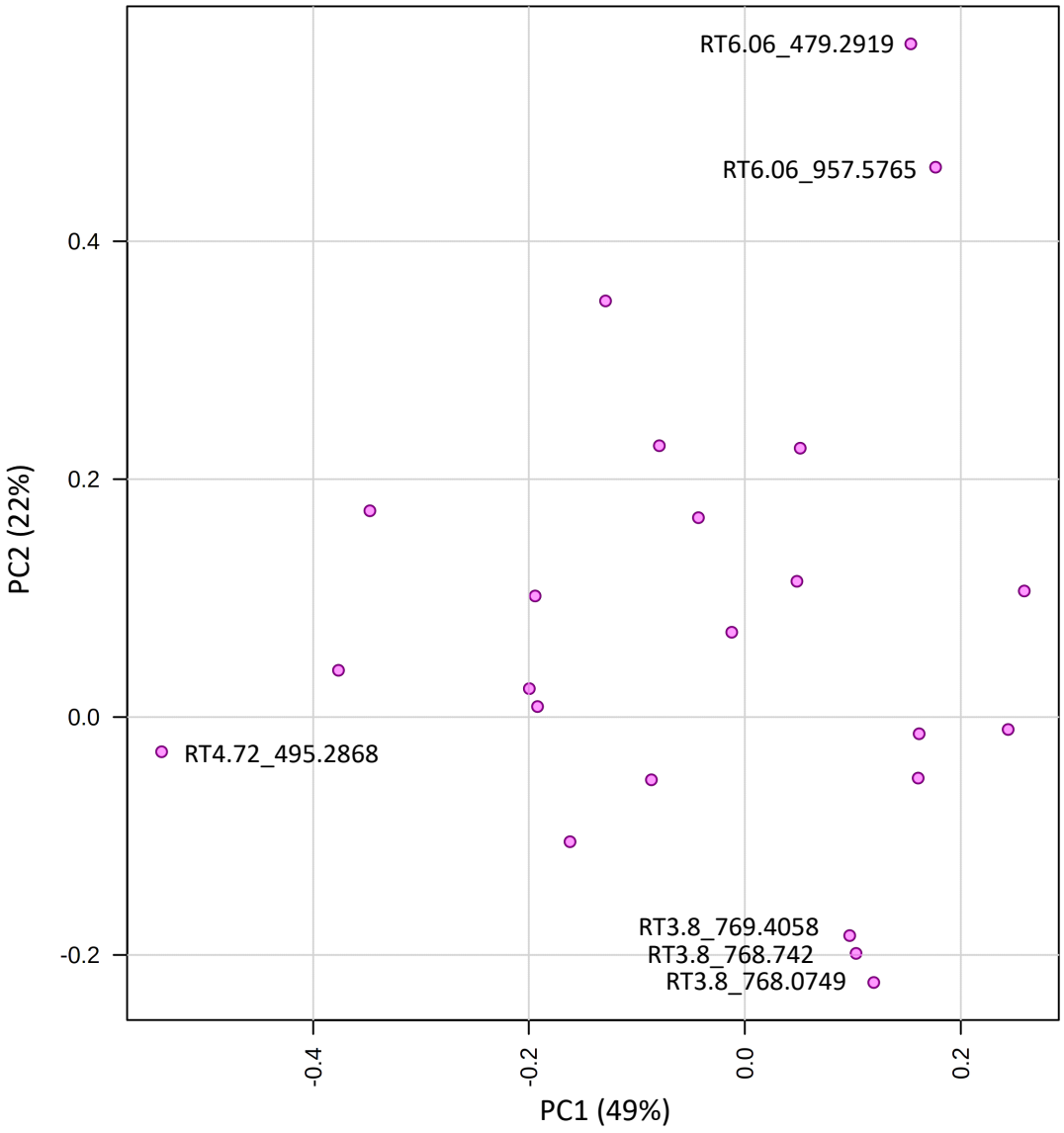

Figure 2.19. Loadings plot of the first two PCs of the PCA for *S. armeniacus* cultured with and without 33 µM AT. Select variables that contribute to the variance seen in the PCs are labelled.

Three mass features appeared on the volcano plot as being upregulated in the Plus AT treatments compared to the Minus AT treatments (**Figure 2.20)** . These three features appeared at retention time 3.8 minutes (**Figure 2.21)**. These peaks seemed to be related as well, as they occurred approximately 0.33 mass units away from each other, suggesting a larger

molecule with three charges resulting in an $[M+3H]^{3+}$ of 768.742. The difference of 0.33 mass units indicates that these three peaks are part of the isotopic peak. Calculations of the $[M+H]^{+}$ adduct revealed a mass of 2303.226, which is outside the range of detection for the mass spectrometer. We also observed the $[M+2H]^{2+}$ ion with a mass of 1152.6085 (**Figure 2.22)**.

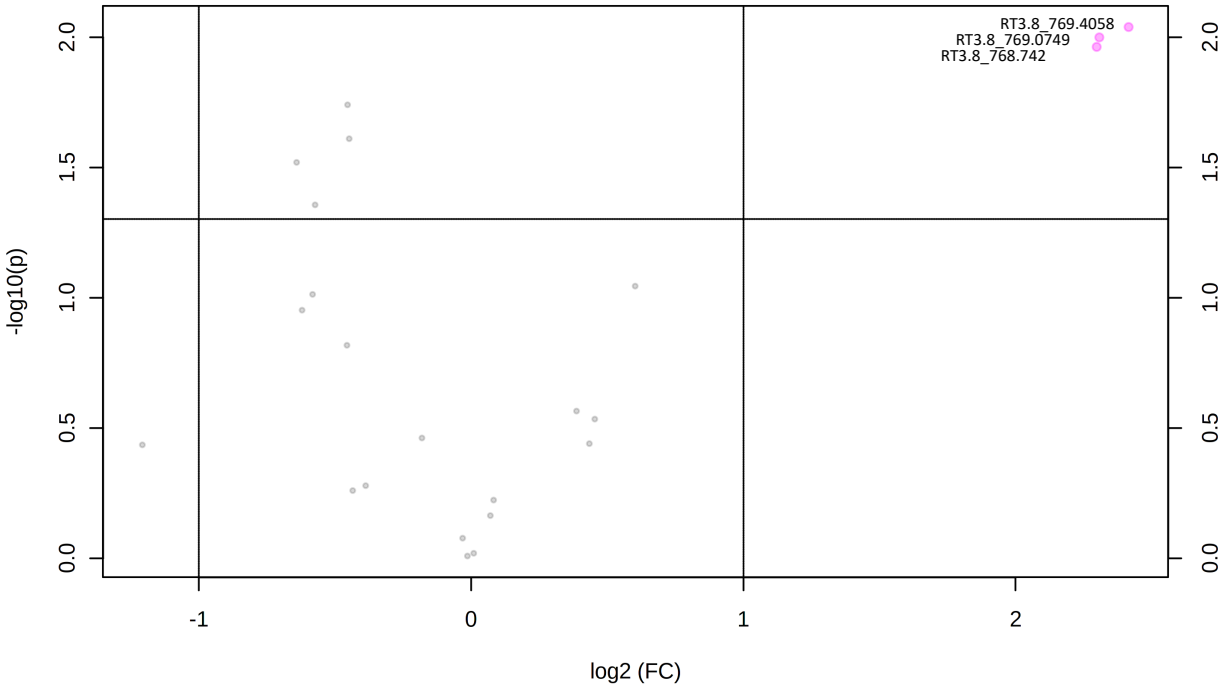

Figure 2.20. Volcano plot of features produced by *S. armeniacus* appearing 2-fold greater in the Plus AT treatment than the Minus AT treatment, with p-values generated from a t-test. Cut-offs for points of interest were set at a 2-fold change (vertical lines) and $p<0.05$ (horizontal line). Points of interest are highlighted and labelled (FC = Fold Change).

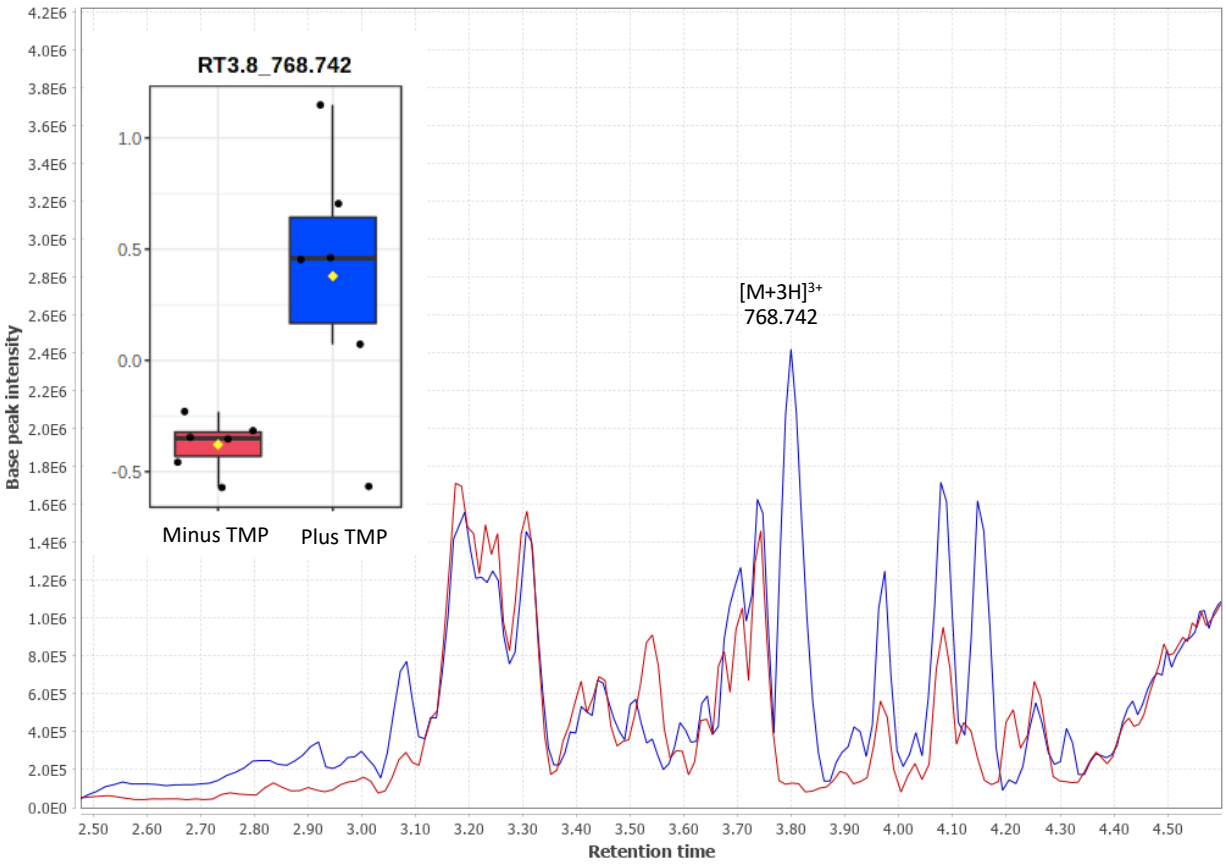

Figure 2.21. TIC overlay of *S. armeniacus* cultured with 30 µM AT (blue) and without AT (red). Data were collected from *m/z* 100-2000. Box plot shows the indicated mass feature in all samples.

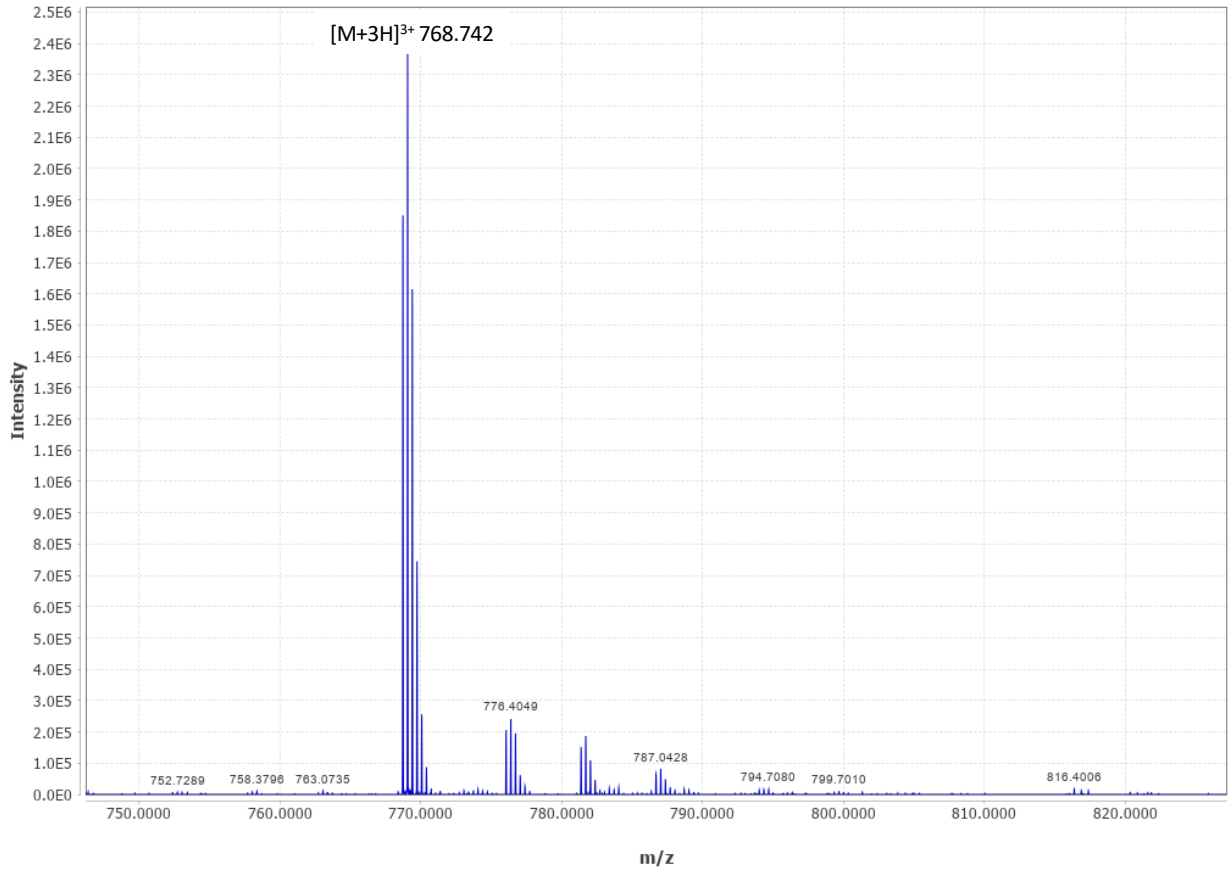

Figure 2.22. Mass spectrum of *S. armeniacus* with 30 µM AT at a retention time of 3.80 mins. Data were collected between *m/z* 100-2000.

A secondary metabolite of this mass indicates that it is likely a peptide around 20 amino acids in length. Further examination of the mass spectrum resulted in the discovery of b ions, which were used to determine 5 amino acids or the sequence. Peaks representing b ions had a similar isotopic peak shape to the $[M+3H]^{+3}$ peak, and mass differences were calculated and compared to the exact masses of amino acids to determine the sequence. It was determined that the amino acid sequence on one end of the molecule is alanine-threonine-glycine-alanine-threonine (**Figure 2.23)**. No other b or y ion signals beyond the first 5 were identified.

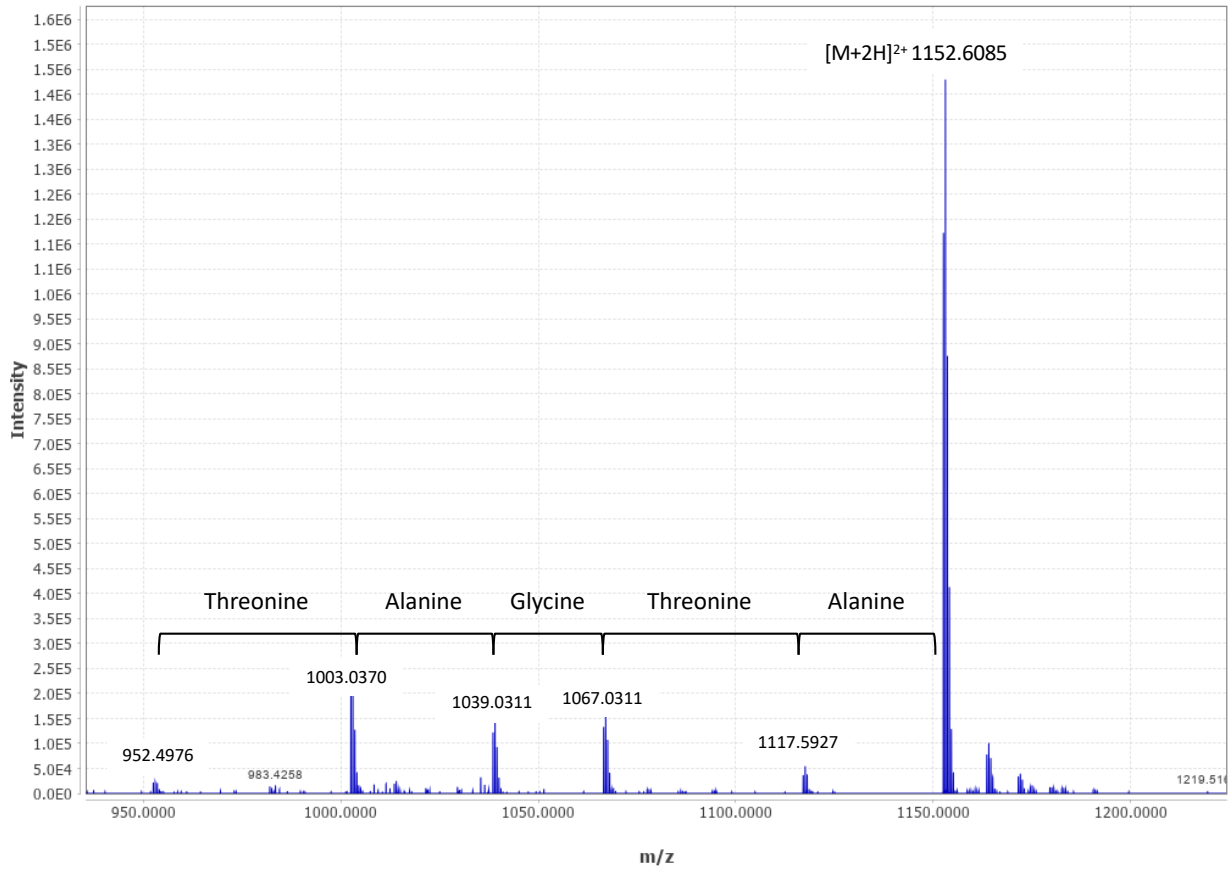

Figure 2.23. Mass spectrum of S. armeniacus cultured with 30 µM AT at a retention time of 3.8 min showing b ions of $\{M+2H]^{2+}$ 1152.6085.

This peptide could originate from one of several types of BGCs in the bacteria. It could be a ribosomally synthesized and posttranslationally modified peptide (RiPP), which is a structurally diverse class of natural products.[8] For this class of secondary metabolite, the peptide is encoded in the DNA and assembled by the ribosome as a linear chain, which is then altered by posttranslational modification machinery.[8] This precursor peptide includes a short leader peptide sequence, the purpose of which has not been conclusively identified but could serve a variety of purposes, including as a recognition site for posttranslational modification

enzymes or to aid in proper folding of the peptide.[9] Because there are many classes of RiPPs, the leader peptide could play different roles depending on the variety of RiPP in question.

The peptide could also be synthesized by nonribosomal peptide synthetases (NRPSs). While these NRP natural products are often smaller than their RiPP counterparts, they can be quite large. For example, chersinamycin is a streptomycetal NRP composed of 17 amino acid residues with a molecular weight of 2574 Da.[10] Peptides originating from NRPS pathways are assembled using modular enzymes with three main types of domains composing a module: adenylation (A) domains, peptidyl carrier protein (PCP) or domains, and condensation (C) domains. Thioester (TE) domains may also exist at the C-terminal of the NRPS to catalyze release of the peptide from the biosynthetic machinery. The A domain activates amino acid monomers in a two-step process. The amino acid is first activated to aminoacyl-adenylate in an ATP-dependent fashion. A domains are substrate-specific, however, they are also able to incorporate non-proteinogenic amino acids.[11] The A domain then transfers the activated amino acid to the PCP domain, where it forms a thioester bond with the 4'-phosphopantetheine of the PCP domain.[12] The PCP delivers the substrate to the C domain located at the beginning of the module, which catalyzes the bond formation between the amino acid monomer and the growing peptide chain. This method of peptide synthesis support very diverse peptides due to the incorporation of non-proteinogenic amino acids.

Further amino acid sequence information could not be inferred from b or y ions from the mass spectrometer, which could be explained if the peptide in question was a lassopeptide RiPP, where some of the amino acid sequence forms a ring while the remaining amino acids form a chain off of the ring.[13] If this peptide belongs to the NRP class of secondary

metabolites, the b and y ions would be more difficult to identify due to the presence of non-proteinogenic amino acids.

The exact origin of $[M+3H]^{3+}$ 768 is difficult to determine due to the lack of genomic evidence that would support either hypothesis. The antiSMASH results of the *S. armeniacus* genome reveals the presence of several RiPP pathways, but none that match the determined amino acid sequence or the mass of the peptide. There are also many NRPS pathways in the genome, but none have enough modules to support the number of amino acids required to make this peptide. It is possible that this genome is not assembled accurately enough to show all of the existing BGCs, given the published *S. armeniacus* ATCC 15676 genome has a fragmented pathway for armeniaspirol.[14] Identification of the full sequence of the peptide by MS^n by further identification of b or y ions is required to classify it as a specific class of natural product.

2.2.5 Trimethoprim and atenolol as elicitors of secondary metabolism

The two elicitors used to change secondary metabolite expression in *S.* sp. PBH53 had different effects. The addition of TMP resulted in the significant upregulation of 6 peaks, all at the same retention time. These peaks correspond to a maximum of 3 compounds, however, the fact that they all elute at the same time could indicate that they are related. AT addition resulted in the significant upregulation of two peaks, both eluting from the column at the same retention time as well. While no obvious relationship between the two molecules has been identified, they could also be related. Both elicitors appear to act as pathway-specific regulators, as opposed to global regulators. The plus and minus AT treatments completely

overlap in these two dimensions, indicating that the treatment did not have a substantial effect on the variance between the two samples. Instead, AT had very specific effects on the metabolome of *S.* sp. PBH53. TMP had a more noticeable effect on the overall variance of the samples ,with more noticeable separation between the plus and minus TMP samples.

The same two elicitors used to change secondary metabolite expression in *S. armeniacus* had very different effects. AT only significantly upregulated the expression of one molecule, $[M+3H]^{3+}$ 768, appearing to have a pathway-specific effect on the metabolome. TMP had more of a global effect on the *S. armeniacus* metabolome, significantly upregulated expression of 10 peaks at 7 different retention times. This effect can be seen in the loadings plot (s. arm TMP), showing that almost every point has a substantial contribution to the variance observed. Where TMP had a pathway-specific effect in *S.* sp. PBH53, it appeared to have pleiotropic effects in *S. armeniacus*.

Given that TMP acts on bacteria by inhibiting DHFR, it stands to reason that it could have widespread regulatory effects throughout the genome. Decreased DHFR function results in decreased DNA, RNA, and protein biosynthesis, which are essential to bacterial life. In contrast, AT has no identified bacterial target and shows no inhibitory effect on bacterial growth.[2] The mechanism by which it modulates gene expression is unknown for this reason. It seems to act primarily as a chemical signal rather than a compound with inhibitory effects. This effect is interesting when viewed in light of research studying pairwise interactions of various *Streptomyces* bacteria. *Streptomyces* bacteria were more likely to secrete molecules that would inhibit their neighbouring bacteria if that strain was closely related or shared BGCs, supporting the hypothesis that secondary metabolites are produced in response to the presence of highly

competitive bacteria and not necessarily to the presence of toxic compounds.[15] Atenolol is a synthetic drug, lacking a pre-existing role in nature. It is possible that it resembles a signalling molecule present in certain *Streptomyces* strains that results in the activation of gene expression. However, further research is required to study the reason why a synthetic drug with no bacterial targets is able to modulate gene expression.

2.2.6 Elicitors for natural product discovery

Natural product discovery is limited by many factors. One such reason is that secondary metabolites aren't constantly expressed in every condition. This problem is what the OSMaC approach addresses by exposing the bacteria to many different culture conditions to differentially activate expression of secondary metabolite pathways.[16] This is also what the use of elicitors attempts to address.[1],[2] By exposing bacteria to potentially stressful conditions by the addition of a foreign compound, the bacteria may change the production of secondary metabolites in response to the chemical signal. The inhibition of some aspect of central metabolism could also be the reason for altered gene expression.[17]-[19]

Another factor that limits novel natural product discovery is the identification of new and relevant molecules in bacterial extracts. Secondary metabolite studies often result in rediscovery of already characterized compounds, as they are often the most readily expressed and easily isolated metabolites. In addition, studying secondary metabolites in a single experimental condition makes it very difficult to recognize what molecules might be of interest. Activity-based identification of compounds of interest results in the same problems, where it is very easy to waste time characterizing an already-known metabolite. Elicitor screens provide at

least two conditions that can be compared to determine differences representing differentially expressed molecules. Elicitor screens analyzed by metabolomics studies prove to be even more powerful, using sensitive instruments and statistical analysis to identify relevant compounds. Dereplication of known metabolites can be performed easily by searching databases for matching masses, leaving unknown masses for further analysis. These reasons make elicitors incredibly useful tools for exploiting the undiscovered potential for secondary metabolite production in bacteria.

This study highlights the benefits of using metabolomics studies for observing the effects of elicitors due to the identification of small peaks from the TLC appearing as significantly upregulated, as seen with $[M+3H]^{3+}$ 768. We also saw that not every replicate of each treatment responded in the same way to the added elicitor, which could be explained by genomic instability and heterogeneity which is characteristic of *Streptomyces*, where antibiotic production occurs in some cells and not others due to a division of labour.[20] This occurrence in elicitor studies makes the study of a single replicate not a reliable source of information, where metabolomics studies uses statistical significance to compare abundances of molecules in every condition to determine if the changes seen are relevant. Metabolomics studies provide relevant information and remove many of the biases and assumptions made by the researcher.

2.2.7 Conclusion

The use of elicitors to induce secondary metabolism is applicable to all cultivatable bacteria, and avoids the need for developing genetic tools in microbes that are difficult to manipulate. The addition of TMP and AT elicitors to *S.* sp. PBH53 and *S. armeniacus* and the use of a metabolomics study enabled the identification of four compounds of interest, three of which may be novel natural products. TMP acted as a global regulator in *S. armeniacus*, but as a pathway-specific regulator in *S.* sp. PBH53, while AT acted as a pathway specific regulator in both bacteria. Further research is required to identify the molecules discovered in this experiment.

2.3 Methods

All solvents were purchased from Fisher Scientific (Ottawa, Ontario, Canada) unless otherwise noted.

2.3.1 Growth of strains

Streptomyces armeniacus ATCC 15676 was grown for one week until isolated colonies formed on Czapek Dox agar with 2 g/L yeast extract. *Streptomyces* sp. PBH53 was grown for one week until isolated colonies formed on ISP2 agar. A single colony was inoculated into 50 mL of the corresponding liquid media in 125 mL Erlenmeyer flasks. Trimethoprim (TMP) and atenolol (AT) (Sigma-Aldrich, Oakville, Ontario, Canada) elicitors stock solutions were prepared to 30 mM in DMSO and added to the cultures at 33 µM for *S. sp.* PBH53[1], and 30 µM for *S. armeniacus*.[2] Six biological replicates were prepared of each of the following treatments:

bacteria with no elicitor, bacteria with elicitor, and media blank with elicitor. The cultures were incubated for 7 days at 27°C with rotatory shaking at 200 rpm.

Table 2.1. Summary of culture conditions and treatments for each *Streptomyces* strain-elicitor interaction.

Bacterial strain	Media type	Elicitor	Elicitor concentration (μM)	Replicates of bacteria, no elicitor	Replicates of bacteria, elicitor	Replicates of media blank, elicitor
S. sp. PBH53	ISP2	TMP	33	6	6	6
S. sp. PBH53	ISP2	AT	33	6	6	6
S. armeniacus	CD+YE	TMP	30	6	6	6
S. armeniacus	CD+YE	AT	30	6	6	6

2.3.2 Metabolite Extraction

Cultures were centrifuged at 4000 rpm for half an hour at 4°C, and the supernatant was decanted into 125 mL Erlenmeyer flasks. Amberlite XAD-16N and XAD-7HP (Sigma-Aldrich, Oakville, Ontario, Canada) resins were prepared by soaking in water, then twice in methanol, then in water again for 30 min each (all at 1% w/v). The beads were added to the culture media at 10 g/L each according to Castro-Falcon *et al.*[21] and the flasks were returned to the rotary shaker for 12 hours, The resin was then vacuum filtered and was placed in clean 125mL Erlenmeyer flask with 40mL methanol for 2 hours. The resin was then filtered and solvent was

removed under reduced pressure and the resulting crude extract was dissolved in 1:1 water:methanol to 0.5 mg/mL.

2.3.3 Metabolic Profiling

High-resolution mass spectrometry (HRMS) data were collected by collaborators at Agriculture and Agri-Food Canada using a LTQ Orbitrap XL Hybrid Ion Trap mass spectrometer (Thermo Scientific, Waltham, MA, USA) coupled to a Dionex Ultimate 3000 ultra-high-performance liquid chromatography (UPLC) system (Thermo Scientific) with a RS diode array detector (Thermo Fisher) and a Corona Veo RS charged aerosol detector (Thermo Fisher). Chromatographic separation was performed using a Phenomenex Kinetex C18 column (2.1 x 50 mm, 1.7 µm, 100 Å; Phenomenex, Torrence, CA, USA) at 30°C with a flow rate of 0.35 mL/min. The mobile phase was composed of water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). The 10 min mobile gradient started with mobile phase A being maintained at 95% for 0.5 min, then increasing to 95% mobile phase B over 4 min. The mobile phase was then maintained at 95% solvent B for 3.5 min, followed by 1.5 min of column equilibration at 95% solvent A. The following parameters were utilized for HRMS analysis (*m/z* 100-2000, resolution 30,000) in ESI+ mode: sheath gas (40), auxiliary gas (5), sweep gas (2), spray voltage (4.0 kV), capillary temperature (320°C), capillary voltage (35 V), tube lens voltage (100 V), maximum injection time (500 ms), and microscans (1).

Data preprocessing was performed in MzMine2.[22] Masses were detected using the exact mass detector, where the noise threshold was set for each dataset according to the noise levels in the chromatograms for methanol and media blanks which were run alongside the

extracts. Next, chromatograms were generated for each mass detected using the Wavelet (ADAP) Chromatogram Builder in MzMine v2.53.[23] The minimum group size in number of scans was set to 5, with an m/z tolerance of 5 ppm. Chromatogram deconvolution was then performed using the local minimum search algorithm, where the search minimum in retention time range was set to 0.1 minutes, and the other values were adjusted to allow for detection of true peaks and exclusion of noise peaks. The isotopic peak grouper was then used to remove masses that are part of an isotopic peak pattern. The m/z tolerance was again set to 5 ppm, with a retention time tolerance of 0.1 minutes and without selecting the monotonic shape option. In addition, the maximum charge to be considered for detecting isotopic peak patterns was set to 1. The Join Aligner was then used to align mass peaks, using the same parameters as in the previous steps. The "Same RT and m/z range gap filler" was used for gap filling the datasets, and the data were then exported into Excel by peak height as a .csv file for further analysis.

Any peaks that were present in the media or methanol blanks above the set noise threshold were removed. In addition, any values that were present above the noise threshold in both the plus drug and minus drug treatments were removed. Then, the methanol and media blank samples were removed and the data was exported to Metaboanalyst v5.0[24]

Data were normalized by median with no data filtering in effect, and scaled according to the best fit for the dataset, using pareto scaling. Data were analyzed using univariate analysis to compare the metabolites produced constitutively to the metabolites with increased production in the presence of an elicitor. Principal Component Analysis was used to visualize the variance between samples within each dataset. Mass features that were enhanced at least two-fold with

the elicitor, that also were significantly different based on a t-test ($p<0.05$) were identified with a volcano plot.

Masses were corrected according to the mass of the reserpine standard.

2.3.4 Production and identification of compounds of interest

The masses of all compounds of interest were queried in the Global Natural Products Social Molecular Networking (GNPS) Public Spectral Libraries[4] and the Natural Products Atlas (NPAtlas)[5] databases.

$[M+H]^+$ 732 was produced by a combination of growing *S.* sp. PBH53 in small cultures and larger cultures. The small cultures corresponded to the original growth conditions, with 50 mL of ISP2 media in a 125 mL flask. Larger cultures were grown by inoculating 1.5L of ISP2 media in a 4L flask. TMP was added at 30 µM, and the cultures were grown on a rotary shaker at room temperature for one week. The cultures were extracted with Amberlite resin (Sigma-Aldrich, Oakville, Ontario, Canada) as described above.

For Liquid Chromatography-Mass Spectrometry (LC-MS) performed to identify the presence of $[M+H]^+$ 732, MS data were collected using a Shimadzu LCMS-2020 coupled to a Shimadzu 20A Prominence system with a SPD-20A Prominence UV/VIS detector at wavelengths 210 and 254 nm (Shimadzu, Columbia, Maryland, USA). Chromatographic separation was performed using a ProntoSIL Eurobond C18 column (125 x 4.0 mm, 5.0 µm; Bischoff Chromatography, Leonburg, Germany) at 30°C with a flow rate of 1.00 mL/min. The mobile phase was composed of water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). The 25 min mobile gradient started with mobile phase A being maintained at 95% for 2 min,

then increasing to 95% mobile phase B over 19 min. The mobile phase was then maintained at 95% solvent B for 1 min, followed by decreasing to 5% solvent B over 0.5 min, then 2.5 min of column equilibration at 95% solvent A. MS was performed in ESI+ mode with the m/z range set to 250-1200.

$[M+H]^+$ 732 was purified by first subjecting the crude extract to silica column chromatography with a gradient from 50% ethyl acetate in hexanes to 100% ethyl acetate, then to 15% methanol in ethyl acetate. The resulting fractions were tested for the presence of $[M+H]^+$ 732 by HPLC-MS. The compound was detected in the fraction with 100% ethyl acetate, and the fraction with 15% methanol in ethyl acetate. Those fractions were combined and further purified by preparative TLC with a mobile phase of 100% ethyl acetate. The silica was scraped from the plate and eluted using methanol. The extract was dried under air and dissolved to 2 mg/mL in methanol.

$[M+H]^+$ 732 was further purified from the extract using an Agilent 1100 Series HPLC with an Agilent 1100 Series Fraction Collector (Mississauga, Ontario, Canada) with a ProntoSIL Eurobond C18 column (125 x 4.0 mm, 5.0 µm; Bischoff Chromatography, Leonburg, Germany) at 30°C with a flow rate of 1.00 mL/min. The mobile phase was composed of water with 0.1% trifluoroacetic acid (TFA) (solvent A) and acetonitrile with 0.1% TFA (solvent B). The 25 min mobile gradient started with mobile phase A being maintained at 95% for 2 min, then increasing to 95% mobile phase B over 19 min. The mobile phase was then maintained at 95% solvent B for 1 min, followed by decreasing to 5% solvent B over 0.5 min, then 2.5 min of column equilibration at 95% solvent A. Fraction collection occurred from 15.35 min to 15.60 min. The resulting fractions were combined and dried under air.

UPLC-HRMS/MS experiments were performed on $[M+H]^+$ 732, where, in addition to the parameters used for UPLC-HRMS analysis above, ions were trapped using an isolation window of 1 m/z, and MS2 scans were acquired using Collision-Induced Dissociation (CID) at 35 eV. The fragmentation pattern was queried using MS-Finder.[3]

NMR analysis was carried out using a Bruker Avance III with cryoprobe operating at 600 MHz for ^{1}H and 150 MHz for ^{13}C data collection.

[M+Na]+ 773 was queried in the NP Atlas[5] database, with its mass matching that of salinomycin. To confirm it was salinomycin, the extracts from the metabolome profiling were compared to a Salinomycin Ready Made Solution (Sigma-Aldrich, Oakville, Ontario, Canada) by UPLC-HRMS. The salinomycin solution was first lyophilized to remove residual DMSO and dissolved to 0.1 mg/mL in methanol. The extracts from *S.* sp. PBH53 were dissolved to 0.5 mg/mL in 1:1 water:methanol. High-resolution mass spectrometry (HRMS) data were collected by collaborators at Agriculture and Agri-Food Canada using a LTQ Orbitrap XL Hybrid Ion Trap mass spectrometer (Thermo Scientific, Waltham, MA, USA) coupled to a Dionex Ultimate 3000 ultra-high-performance liquid chromatography (UPLC) system (Thermo Scientific) with a RS diode array detector (Thermo Fisher) and a Corona Veo RS charged aerosol detector (Thermo Fisher). Chromatographic separation was performed using a Phenomenex Kinetex C18 column (2.1 x 50 mm, 1.7 µm, 100 Å; Phenomenex, Torrence, CA, USA) at 30°C with a flow rate of 0.35 mL/min. The mobile phase was composed of water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). The 10 min mobile gradient started with mobile phase A being maintained at 95% for 0.5 min, then increasing to 95% mobile phase B over 4 min. The mobile

phase was then maintained at 95% solvent B for 3.5 min, followed by 1.5 min of column equilibration at 95% solvent A. The following parameters were utilized for HRMS analysis (*m/z* 100-2000, resolution 30,000) in ESI+ mode: sheath gas (40), auxiliary gas (5), sweep gas (2), spray voltage (4.0 kV), capillary temperature (320°C), capillary voltage (35 V), tube lens voltage (100 V), maximum injection time (500 ms), and microscans (1). Retention times of the mass of interest were compared between the samples and standard to confirm the identity as salinomycin

$[M+Na]^+$ 411 was produced by inoculating twenty 125 mL flasks containing 50 mL Czapek Dox media plus 2 g/L yeast extract and 33 µM TMP with *S. armeniacus* and similarly incubating for 7 days on a rotary shaker at room temperature before extracting for 12 hours with Amberlite resins (Sigma-Aldrich, Oakville, Ontario, Canada) as described above. The beads were eluted in excess methanol for 2 hours, and the beads were filtered off. The solvent was removed under reduced pressure.

The crude extract was subjected to a Kupchan partition. The extract was dissolved in 90% v/v methanol in water. An equal volume of hexanes were added and the upper hexane phase was removed. The aqueous content of the extract was then adjusted to 40% v/v water. The water phase was extracted twice with chloroform at one times the original volume. The chloroform phase was collected, and the remaining water phase was evaporated to remove residual organic solvents. The water phase was then reconstituted to its original volume with water, and was extracted twice with an equal volume of n-butanol. The hexanes, chloroform, and n-butanol partitions were subjected to solvent removal under reduced pressure, and were

dissolved to 0.5 mg/mL in 1:1 water:methanol for UPLC-HRMS analysis to determine which partition contained $[M+Na]^+$ 411.

$[M+3H]^{3+}$ 768 was produced by inoculating three 125 mL flasks containing 50 mL Czapek Dox media plus 2 g/L yeast extract and 33 µM AT with *S. armeniacus* and incubating for 7 days at room temperature on a rotary shaker. The cells were removed by centrifugation at 4000 rpm and 4°C for 30 minutes, and the remaining media was extracted for 12 hours with Amberlite resins (Sigma-Aldrich, Oakville, Ontario, Canada) as described above . The beads were then eluted in 40 mL methanol for two hours, then filtered off. The solvent was removed under reduced pressure.

The crude extract was subjected to a Kupchan partition. The extract was dissolved in 90% v/v methanol in water. An equal volume of hexanes were added and the upper hexane phase was removed. The aqueous content of the extract was then adjusted to 40% v/v water. The water phase was extracted twice with chloroform at one times the original volume. The chloroform phase was collected, and the remaining water phase was evaporated to remove residual organic solvents. The water phase was then reconstituted to its original volume with water, and was extracted twice with an equal volume of n-butanol. The hexanes, chloroform, and n-butanol partitions were subjected to solvent removal under reduced pressure, and were dissolved to 0.5 mg/mL in 1:1 water:methanol for UPLC-HRMS analysis to determine which partition contained $[M+3H]^{3+}$ 768.

2.4 References

[1] Okada, B.K., Wu, Y., Mao, D., Bushin, L.B., and Seyedsayamdost, M.R. (2016). Mapping the Trimethoprim-Induced Secondary Metabolome of *Burkholderia thailandensis*. ACS Chem. Biol. *11*, 2124–2130.

[2] Moon, K., Xu, F., Zhang, C., and Seyedsayamdost, M.R. (2019). Bioactivity-HiTES Unveils Cryptic Antibiotics Encoded in Actinomycete Bacteria. ACS Chem. Biol. *14*, 767–774.

[3] Tsugawa, H., Kind, T., Nakabayashi, R., Yukihira, D., Tanaka, W., Cajka, T., Saito, K., Fiehn, O., and Arita, M. (2016). Hydrogen Rearrangement Rules: Computational MS/MS Fragmentation and Structure Elucidation Using MS-FINDER Software. Anal. Chem. *88*, 7946–7958.

[4] Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T., et al. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. Nat. Biotechnol. *34*, 828–837.

[5] van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., Neto, F.C., Castaño-Espriu, L., Chang, C., Clark, T.N., et al. (2019). The Natural Products Atlas: An Open Access Knowledge Base for Microbial Natural Products Discovery. ACS Cent. Sci. *5*, 1824–1833.

[6] Kim, W.K., Kim, J.H., Yoon, K., Kim, S., Ro, J., Kang, H.S., and Yoon, S. (2012). Salinomycin, a p-glycoprotein inhibitor, sensitizes radiation-treated cancer cells by increasing DNA damage and inducing G2 arrest. Invest. New Drugs *30*, 1311–1318.

[7] Antoszczak, M., Maj, E., Napiórkowska, A., Stefańska, J., Augustynowicz-Kopeć, E., Wietrzyk, J., Janczak, J., Brzezinski, B., and Huczyński, A. (2014). Synthesis, anticancer and antibacterial activity of salinomycin N-benzyl amides. Molecules *19*, 19435–19459.

[8] Link, A.J. (2015). Biosynthesis: Leading the way to RiPPs. Nat Chem Biol *11*, 551–552.

[9] Oman, T.J., and van der Donk, W.A. (2010). Follow the leader: the use of leader peptides to guide natural product biosynthesis. Nat. Chem. Biol. *6*, 9–18.

[10] Morgan, K.T., Zheng, J., and McCafferty, D.G. (2020). Discovery of Six Ramoplanin Family Gene Clusters and the Lipoglycodepsipeptide Chersinamycin. Chembiochem.

[11] Caboche, S., Leclère, V., Pupin, M., Kucherov, G., and Jacques, P. (2010). Diversity of Monomers in Nonribosomal Peptides: towards the Prediction of Origin and Biological Activity. J Bacteriol. *192*, 5143–5150.

[12] Stein, T., Vater, J., Kruft, V., Otto, A., Wittmann-Liebold, B., Franke, P., Panico, M., McDowell, R., and Morris, H.R. (1996). The Multiple Carrier Model of Nonribosomal Peptide Biosynthesis at Modular Multienzymatic Templates. J. Biol. Chem. *271*, 15428–15435.

[13] Arnison, P.G., Bibb, M.J., Bierbaum, G., Bowers, A.A., Bugni, T.S., Bulaj, G., Camarero, J.A., Campopiano, D.J., Challis, G.L., Clardy, J., et al. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. Nat. Prod. Rep. *30*, 108–160.

[14] Fu, C., Xie, F., Hoffmann, J., Wang, Q., Bauer, A., Brönstrup, M., Mahmud, T., and Müller, R. (2019). Armeniaspirol Antibiotic Biosynthesis: Chlorination and Oxidative Dechlorination Steps Affording Spiro[4.4]non-8-ene. Chembiochem *20*, 764–769.

[15] Westhoff, S., Kloosterman, A., Hoesel, S.F.A. van, Wezel, G.P. van, and Rozen, D.E. (2020). Competition sensing alters antibiotic production in *Streptomyces*. BioRxiv 2020.01.24.918557.

[16] Bode, H.B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. ChemBioChem *3*, 619–627.

[17] Li, A., Mao, D., Yoshimura, A., Rosen, P.C., Martin, W.L., Gallant, É., Wühr, M., and Seyedsayamdost, M.R. (2020). Multi-Omic Analyses Provide Links between Low-Dose Antibiotic Treatment and Induction of Secondary Metabolism in *Burkholderia thailandensis*. MBio *11*.

[18] Craney, A., Ozimok, C., Pimentel-Elardo, S.M., Capretta, A., and Nodwell, J.R. (2012). Chemical perturbation of secondary metabolism demonstrates important links to primary metabolism. Chem. Biol. *19*, 1020–1027.

[19] Tanaka, Y., Izawa, M., Hiraga, Y., Misaki, Y., Watanabe, T., and Ochi, K. (2017). Metabolic perturbation to enhance polyketide and nonribosomal peptide antibiotic production using triclosan and ribosome-targeting drugs. Appl. Microbiol. Biotechnol. *101*, 4417–4431.

[20] Zhang, Z., Du, C., de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G.P., Choi, Y.H., Claessen, D., and Rozen, D.E. (2020). Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. Sci. Adv. *6*.

[21] Castro-Falcón, G., Millán-Aguiñaga, N., Roullier, C., Jensen, P.R., and Hughes, C.C. (2018). Nitrosopyridine Probe To Detect Polyketide Natural Products with Conjugated Alkenes: Discovery of Novodaryamide and Nocarditriene. ACS Chem. Biol. *13*, 3097–3106.

[22] Pluskal, T., Castillo, S., Villar-Briones, A., and Orešič, M. (2010). MZmine 2: Modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. BMC Bioinformatics *11*, 395.

[23] Myers, O.D., Sumner, S.J., Li, S., Barnes, S., and Du, X. (2017). One Step Forward for Reducing False Positive and False Negative Compound Identifications from Mass Spectrometry Metabolomics Data: New Algorithms for Constructing Extracted Ion Chromatograms and Detecting Chromatographic Peaks. Anal. Chem. *89*, 8696–8703.

[24] Chong, J., Wishart, D.S., and Xia, J. (2019). Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis. Curr. Protoc. Bioinformatics *68*, e86.

Chapter 3: Genomics-Driven Genome Mining and Assembly in *Streptomyces*

3.1 Introduction

Since the advances in genome sequencing technologies in 2005 with the introduction of 454 (Roche) sequencing and Illumina sequencing, bacterial genome sequencing has become readily accessible. Given that 454 sequencing is no longer available, Illumina short read sequencing has become the main method for sequencing bacterial genomes and metagenomes. With the ability to sequence DNA in a high throughput manner, the need for software tools to help assemble the short DNA reads into genomes grew.

Two main approaches exist for assembling raw DNA reads into a genome. The first is Overlap Layout Consensus (OLC), where overlaps between the reads are found, then a layout is carried out with all of the reads, and the consensus sequence is determined from the resulting graph. Each read is compared to every other read in the dataset for overlaps. The assembly graph is created by assigning reads as nodes and overlaps between reads as edges.[1]

The other main technique for assembling genomes is by De Bruijn Graphs. Instead of viewing the sequence of the entire read and looking for overlaps, De Bruijn Graphs break up the sequence into *k*-mers compare the already-determined consensus sequence of these *k*-mers. De Bruijn Graph assembly is computationally simpler than OLC, making it the first choice of many researchers in genome assembly.[1] One of the most commonly used softwares for *de novo* assembly of bacterial genomes, SPAdes, utilizes De Bruijn Graph construction.[2]

To construct a De Bruijn Graph, the *k*-mers are separated into their beginning and end *k*-1-mers. Each distinct *k*-1-mer is given its own node on the assembly graph. A directed edge is drawn between the nodes if they lie next to each other, and the assembly continues this way. A walk through the graph that only crosses each directed edge once is considered an Eulerian walk **(Figure 3.1).**[3],[4] However, sequencing rarely results in an assembly graph has only one possible walk. Even in the simple example shown, De Bruijn Graph assembly encounters a problem inherent to sequencing and assembly: the occurrence of repetitive sequences (**Figure 3.2)**. Repetitive sequences can have a variable effect on the assembly depending on the algorithm used. In some cases, the number of repetitive sequences becomes underrepresented, leaving out some of the repeats that actually exist in the genome. De Bruijn Graphs are prone to a different type of problem, where there is more than one way to walk through the graph using each edge once.[5] In the example shown, "AB" is a repeated sequence, which causes ambiguity in the sequence following it. In this case, there are four possible walks that cross each edge once **(Figure 3.2)**. The same repetitive sequence concepts occur in genomes and genome assemblies, with magnified effects due to the large size of a genome.[6]

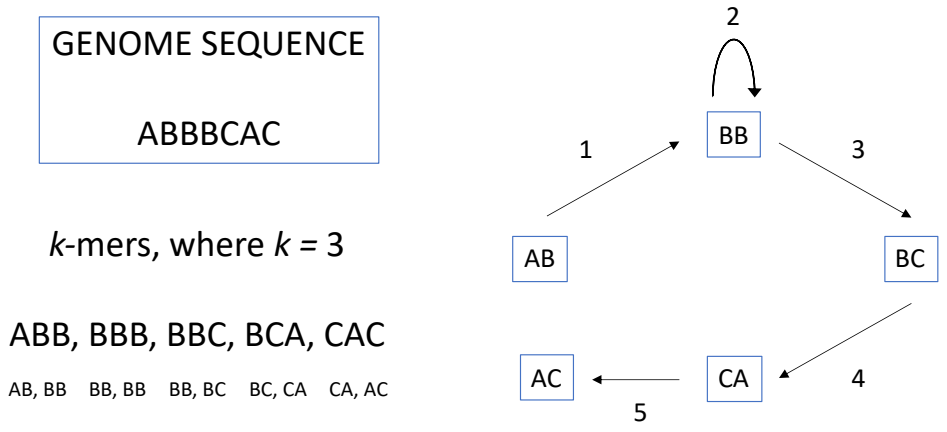

Figure 3.1. A genome sequence of ABBBCAC showing one Eulerian walk through the De Bruijn assembly graph.

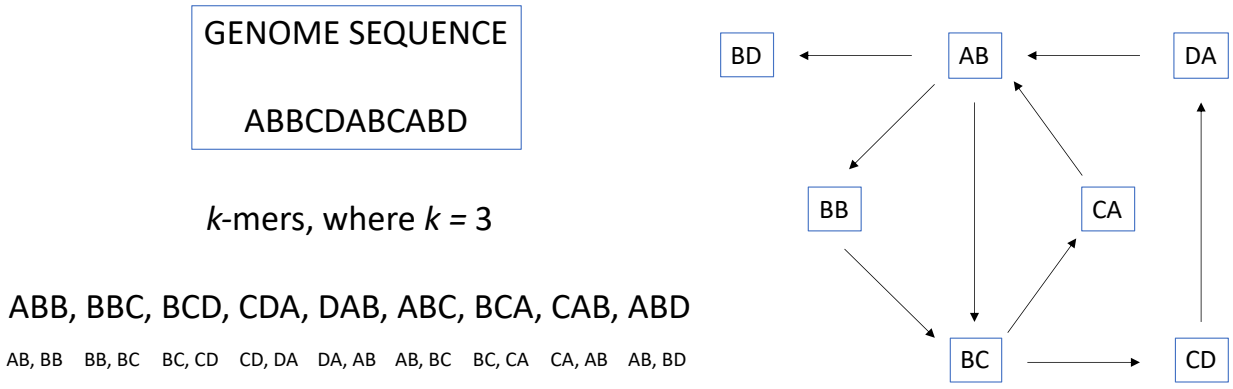

Figure 3.2. A genome sequence of ABBCDABCABD with more than one Eulerian walk through the De Bruijn assembly graph, illustrating the problems that genome assemblies have when encountering repeating sequences.

In bacterial genome sequencing for the purpose of natural product discovery, many biosynthetic gene clusters (BGCs) are identified from a genome sequence, but very few of these BGCs have a verified natural product associated with them. Certain secondary metabolite pathways contain repetitive sequences, specifically, Type I polyketide synthase (PKS) pathways. PKS pathways use acyl-transferase (AT) domains for substrate selection and subsequent

transfer to the acyl carrier protein (ACP).[7] For Type I PKS pathways, there is an AT domain for each module, with some pathways containing up to 22 modules.[8] These AT domains tend to be highly similar in nucleotide sequence, which can result in assembly errors leading to several incorrect assumptions about the genome and the BGCs it encodes **(Table S5.1, S5.2)**.

AntiSMASH[9] analyzes genomic sequences for the presence of BGCs a key step in genomics-guided natural product discovery. Fragmented BGCs from assembly errors could lead to an overestimation of how many clusters and thus natural products are actually encoded in the genome. In addition, fragmented known BGCs likely show low similarity to their correctly assembled known clusters simply because the whole sequence does not occur together. This may lead to the conclusion that a genome with a misassembled known BGC does not possess that cluster when in actuality the full cluster is present in the genome. Thus BGC misassemblies can result in the assumption that genomes possess a larger number of BGCs than they actually do, and that more of these BGS are unique than they actually are. Both would be significant problems for genome-guided natural product discovery.

We propose that genomes sequenced by Illumina could have assembly errors leading to overestimation of the number of gene clusters and underestimation of the number of identified natural products. We also propose that making use of the sequence of biosynthetic gene clusters known to be present in an organism's genome can result in a more accurate genome assembly.

3.2 Results and Discussion

Long Type I PKS pathways from *Streptomyces* species with more than 10 modules were chosen to simulate paired-end Illumina reads using ART.[10] Reads were simulated with a length of 150 bp and average coverage of 30X. The reads were assembled using both SPAdes and Biosynthetic SPAdes **(Table 3.1)**. SPAdes was not able to effectively reassemble the pathways into a single contig, especially as the pathways increased in length. Pathways with 10-12 modules proved easiest to reassemble into one contig, while pathways with 13 or more modules could not be reassembled into a single contig. The reassembled pathways typically had contig edges at AT domains. Biosynthetic SPAdes was also used to see if it had an advantage in assembling BGCs. When viewing the output "contigs" file, Biosynthetic SPAdes was unable to reconstruct complete pathways. However, when viewing the "gene clusters" output file, complete pathways were assembled into a single contig. The "gene clusters" file offers options for how the pathway could be assembled if there is more than one way to assemble it, but for most of the pathways it gave a single option. While Biosynthetic SPAdes appears to be able to accurately assemble BGCs into the "gene clusters" file, it was not always reliable. In some cases, there was no "gene clusters" file generated in the output, as seen with Reveromycin A and Tetronomycin**.** In other cases, the "gene clusters" file was unable to assemble the reads into a single contig, for example, Halstoctacosanolide A and Caniferolide A **(Table 3.1, Table S5.3)**.

To determine if the repetitive nature of highly similar AT domains impacted assembly, reads were also simulated for bacillaene, a *Bacillus* natural product that uses a *trans*-AT domain and thereby avoids the potential problems of multiple AT domain sequences. No assembly errors occurred, with both SPAdes and Biosynthetic SPAdes producing a single complete contig out of the raw reads. Therefore, it appears that the main error in PKS gene cluster assembly occurs due to repetitive AT domain sequences.

Table 3.1. The results of the assembly of simulated Illumina reads of large PKS pathways by SPAdes and Biosynthetic SPAdes.

Cluster Name	**Number of PKS modules**	**Number of assembled contigs (SPAdes)**	**Number of assembled contigs (Biosynthetic SPAdes)**
Oxalomycin B	10	1	1
Sceliphrolactam	10	1	1
Bafilomycin B1	12	1	1
Laidlomycin	13	13	1
Monensin	13	3	1
Nanchangmycin	14	3	2
Reveromycin A	14	9	11*
Sanglifehrin A	14	4	1
Tetronomycin	14	5	4*
Meillingmycin	15	9	2

Nigericin	15	8	2
Oligomycin	17	5	1
Halstoctacosanolide A	19	12	11
Candicidin	21	3	1
Caniferolide A	22	9	12

* indicates no 'gene clusters" file available. Number of contigs retrieved from "contigs" output file.

To determine if the assembly errors seen in the read simulation experiments are present in assembled and deposited genomes sequenced by Illumina, *Streptomyces* genomes from the NCBI database were chosen and analyzed by antiSMASH.[9] Pathways that were assembled incorrectly were identified by looking at the KnownClusterBlast results for each Type I PKS pathway. Misassembled pathways usually had very good matches to known clusters in the tailoring genes, but were lacking full PKS genes. The first KnownClusterBlast result for the fragmented pathways also tended to appear as the match for several pathways in the genome, indicating that the sequence that makes up one pathway in the genome has the potential to be split into multiple pathways in the assembly of raw reads. Out of 35 genomes studied for these assembly errors, 7 genomes had at least one pathway that had evidence of being fragmented, with a total of 10 potential assembly errors identified (**Table 3.2, Table S5.4).** However, these identified pathways do not have associated isolated natural product from the organism, so they may not make the product reported by the KnownClusterBlast.

Table 3.2. Genomes and pathways from the NCBI database found to contain potential assembly errors in PKS gene clusters.

Organism	**KnownClusterBlast Match**	**Contig number**	**Nucleotide start and stop**
S. sparsogenes	Oxalomycin	107	1-15,590
		108	1-7,438
		110	1-23,883
	Nanchangmycin	103	1-11,222
		120	1-6,583
		127	1-29,807
		166	1-13,275
S. diasticus* subsp. *ardesiacus	Butyrolactol A	4	1-43,843
		17	1-41,854
S. rubrogriseus	Coelimycin P1	1	1-35,620
		3	1-8,053
S. violens	Oxalomycin B	2	168,580-178,228
		25	39,159-91,422
S. griseofuscus	Lankamycin	82	1-21,938
		87	1-18,774
	Pentamycin	32	65,088-92,620
S. halstedii	Bafilomycin B1	33	1-29,952
S. yokosukanensis	Sceliphrolactam	2	315,623-447,105
	Lankamycin	45	1-24,944
		49	1-19,347

To validate the findings in published *Streptomyces* genomes, we chose a pathway in *S.* sp. PBH53 that showed evidence of being incorrectly assembled **(Figure 3.3)**. The pathway selected for further study had a KnownClusterBlast match of 37% to Naphthomycin A. The assembled pathway had the necessary tailoring genes, but its PKS genes (*natA-E*) appeared to be truncated and missing from the assembly. We identified the mass of Naphthomycin A from LC-MS analysis, and isolated the compound by HPLC. The identity of the resulting bright yellow compound was confirmed by NMR spectroscopy using Small Molecule Accurate Recognition Technology (SMART) to identify the compound as Naphthomycin A **(Figure S5.1; Table S5.5).**[11] Naphthomycin A was the second match on the result list, with the only matching mass and a cosine score of 0.907. Four more clusters annotated by antiSMASH were identified in *S.* sp. PBH53 that showed evidence of incorrect assembly, corresponding to two natural products: sceliphrolactam and sanglifehrin A.

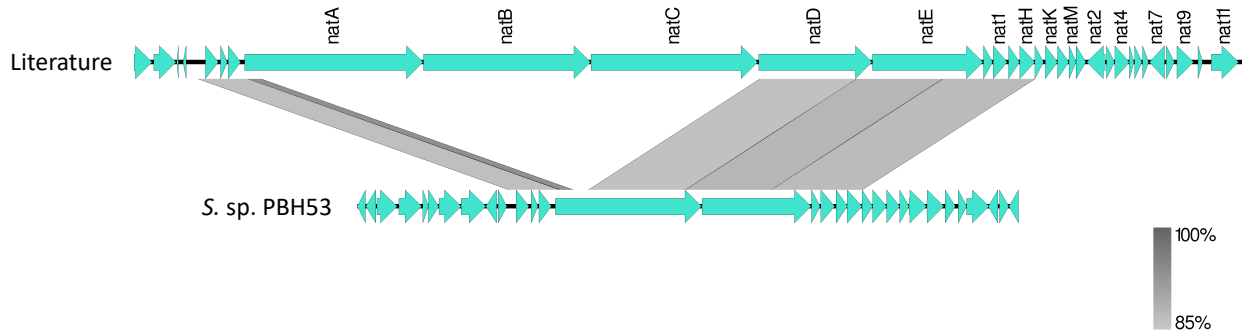

Figure 3.3. Comparison of the Naphthomycin A gene cluster in the published S. sp. PBH53 genome to the reference Naphthomycin A gene cluster. The PKS genes (*natA-E*) show evidence of being truncated, while the tailoring genes on either side assemble correctly. The assembled cluster in *S.* sp. PBH53 shows a KnownClusterBlast match of 37% to the literature Naphthomycin A cluster.

After confirming that *S.* sp. PBH53 produces Naphthomycin A and therefore must have a complete gene cluster, Naphthomycin A paired-end raw reads were simulated from the literature gene cluster (NCBI accession GQ452266.1) in ART. Reads were simulated at 30X coverage with read lengths of 100 bp, 125 bp, and 150 bp (**Table 3.3)**. Reads were also simulated at 60X and 120X coverage with a read length of 150 bp (**Table 3.4)**.

Table 3.3. Number of contigs resulting from assembling the Naphthomycin A cluster using SPAdes at 30X coverage with varying read lengths.

READ LENGTH (BP)	**# OF CONTIGS IN ASSEMBLY**
100	11
125	12
150	8

Table 3.4. Number of contigs resulting from assembling the Naphthomycin A cluster using SPAdes with 150 bp reads and varying depth of coverage.

COVERAGE	# OF CONTIGS IN ASSEMBLY
30X	8
60X	4
120X	4

Further comparison was carried out only for read length of 150 bp and 30X coverage to simulate what could be expected from basic sequencing by Illumina. The simulated Naphthomycin reads were assembled using SPAdes and Biosynthetic SPAdes to see if one worked better for this PKS pathway than the other. SPAdes produced 8 contigs, with only 4 of the 8 contigs recognizable as a BGC by antiSMASH **(Figure 3.4).** Each of the resulting contigs showed that fragmentation in the assembly occurred in the PKS genes. The "contigs" file of Biosynthetic SPAdes revealed 4 contigs in the assembly, while the "gene clusters" file showed two candidate clusters for a single contig. SPAdes was also performed on the Naphthomycin A cluster with the `--untrusted-contigs` command specifying the literature Naphthomycin A cluster sequence to act like a previously assembled contig. The `--untrusted-contigs` command can be used when the reference contig sequence may have errors or be of unknown quality. SPAdes uses the `--untrusted-contigs` command for gap closure and repeat resolution, and also has an option to specify `--trusted-contigs` which have reliable sequence information, and are additionally used for graph construction in the assembly. The `--untrusted-contigs` command was able to accurately reconstruct the Naphthomycin A gene cluster from the simulated reads without being affected by the repetitive AT domains.

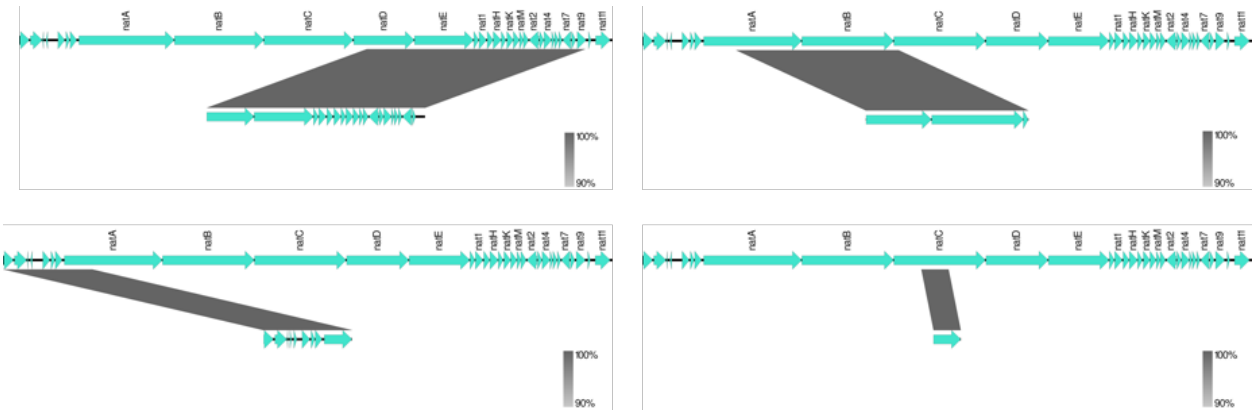

Figure 3.4. Simulated raw reads for the Naphthomycin A BGC assembled using SPAdes and analyzed using AntiSMASH, compared to the complete literature BGC sequence (top sequence of each pairwise alignment).

To test different methods of assembly, the genome of *S.* sp. PBH53 was resequenced by Illumina NextSeq and the quality of assembly was assessed by its ability to reconstruct the Naphthomycin cluster within the genome. First, the assembly was attempted using SPAdes, the same algorithm that was used in the original published genome.[12] SPAdes was unable to correctly assemble the genome, as was expected based on the published genome. SPAdes was then used again, this time including the `--untrusted-contigs` command, again specifying Naphthomycin A as the untrusted contig. The `--untrusted-contigs` command was chosen over the `--trusted-contigs` command due to the possibility of the reference Naphthomycin A cluster sequence differing slightly from the sequence in *S.* sp. PBH53. Despite the successful assemblies seen by using --untrusted-contigs in the single cluster simulations, it was unable to assemble the full genome in a way that resulted in an untruncated Naphthomycin A gene cluster.

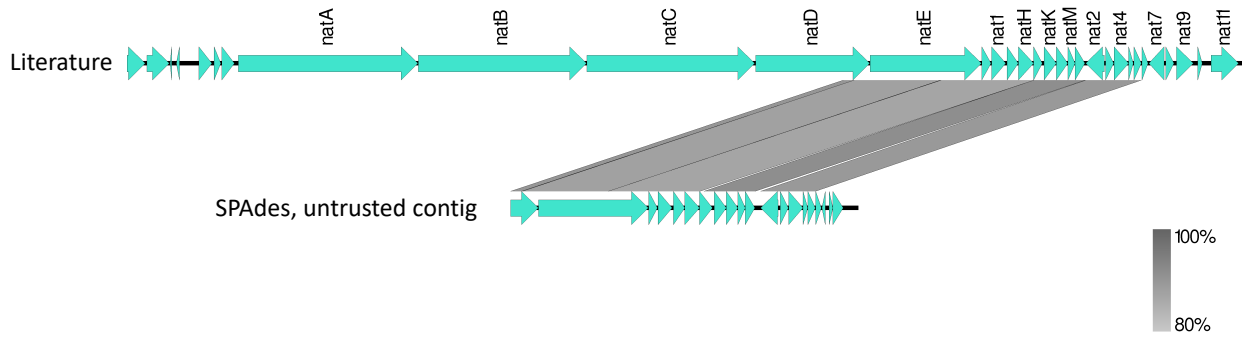

Figure 3.5. The Naphthomycin A cluster resulting from the assembly of the resequenced *S.* sp. PBH53 genome, compared to the literature Naphthomycin A sequence. The assembly was performed using SPAdes, with the literature Naphthomycin A sequence used as an "untrusted contig" in the assembly.

The *S.* sp. PBH53 genome was then assembled using Biosynthetic SPAdes. The resulting "gene clusters" and "contigs" files were run in antiSMASH. The "gene clusters" file contained 43 gene clusters proposed by Biosynthetic SPAdes, with a total of 135 possible ways of assembling those 43 clusters. AntiSMASH recognized 119 candidate gene cluster assemblies belonging to 25 of the 43 clusters, with up to 51 proposed sequences for a single cluster corresponding to sanglifehrin A with 20% similarity. In addition, it was able to correctly assemble the tailoring genes for the Naphthomycin A gene cluster, but often assembled more PKS genes than Naphthomycin A actually has, and had 24 candidates with a top KnownClusterBlast match to Naphthomycin A over 4 gene clusters **(Figure 3.6)**. Out of the 135 candidate clusters, only two did not fall under the category of NRPS, NRPS-like, PKS, or hybrid NRPS/PKS clusters. Given that Biosynthetic SPAdes was specifically designed for assembling NRPS clusters, but also has applications in PKS assembly, it stands to reason that it would have more difficulty recognizing clusters that do not incorporate the easily identified biosynthetic genes of NRPS and PKS pathways.

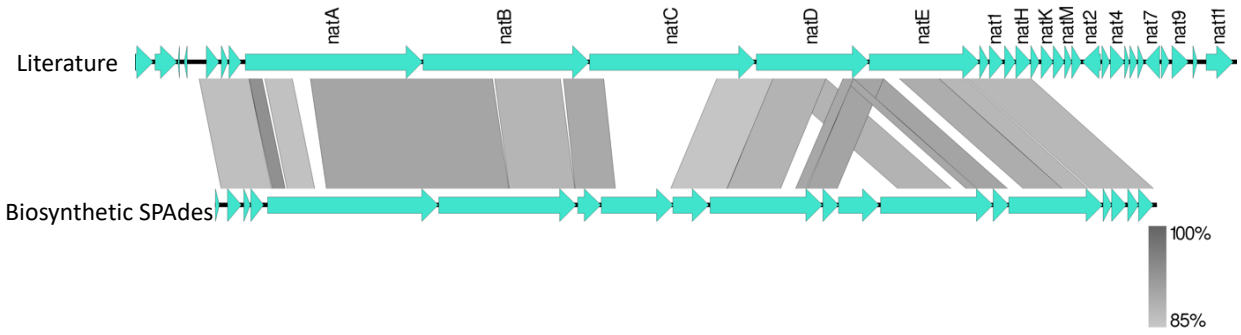

Figure 3.6. An example of one of the proposed Naphthomycin A pathways from the "gene clusters" file of the Biosynthetic SPAdes genome assembly compared to the literature sequence, again showing correctly assembled tailoring genes but problems with correctly assembling the PKS genes (*natA-E*). The assembled cluster shows 37% similarity to the literature sequence.

The “contigs” file resulting from Biosynthetic SPAdes contained the proposed assembled genome with 258 contigs. The assembled genome run in antiSMASH reported 59 putative gene clusters, with more variety than just NRPS and PKS pathways. Two gene clusters had a KnownClusterBlast match to Naphthomycin A, with 25% and 28% matches to the reference gene cluster sequence **(Figure 3.7)**. Only the tailoring genes and the adjacent PKS gene ends were assembled, with significant misassembly occurring in the PKS genes still.

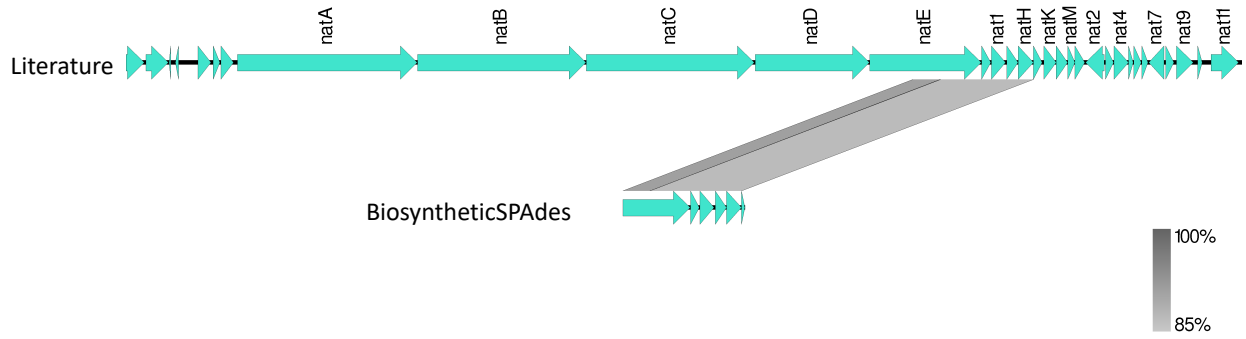

Figure 3.7. One of the two Naphthomycin A clusters in antiSMASH from the "contigs" file of Biosynthetic SPAdes. The tailoring genes assemble, but the assembly cannot cross the entire sequence of the PKS genes (*natA-E*).

Based on these results, Biosynthetic SPAdes doesn't appear to be useful for assembling genomes based on the presence of known BGCs. It is able to identify the main NRPS and PKS pathways present in the genome, but doesn't necessarily assemble them correctly. If Biosynthetic SPAdes was to be used to identify probable gene clusters in the genome without the need for an isolated corresponding molecule, in theory those clusters could be used as references to assemble the genome and resolve ambiguous sequences. However, SPAdes relies on de Bruijn assembly graphs, which will still run into issues when encountering repetitive sequences.

OLC approaches to genome assembly could be used to avoid the issues seen with De Bruijn Graphs. However, it is not a perfect solution. The main OLC assembly program, Newbler (Roche), was designed for assembly of reads from 454 sequencing, and both the sequencing technique and the program have been discontinued. OLC assembly also runs into the same issue that plagues all methods of genome assembly: repetitive sequences make assembly difficult.[13]

To better predict the sequences of BGCs in bacterial genomes, we propose using literature BGC sequences as references to assemble onto, as we attempted with the `--untrusted-contigs` command. One option is to use Bowtie, which is used for aligning RNA-Seq reads onto a reference genome.[14] The reads from Illumina sequencing would be aligned onto BGCs using Bowtie, then combined with the remaining reads and assembled. Another option is to further investigate algorithms for assembling hybrid sequencing data, where long read sequencing data is used to determine the correct order of information in the genome and the short read sequencing data is used to correct the sequence. In this case, the known gene cluster would be treated as if it was long read sequencing data, and the short reads would be scaffolded onto the cluster sequence and assembled around it. These two approaches have yet to be verified as viable options for the assembly problem in PKS clusters.

The approach to avoid many of the problems with assembling short reads is to use long read sequencing. Methods such as Pacific Biosciences (PacBio) Single Molecule, Real-Time (SMRT) sequencing or Oxford Nanopore Technologies are capable of sequencing DNA fragments from 10s to 100s of kb in length. The long reads would be able to span across the repetitive sequences in Type I PKS BGCs, nearly eliminating truncated or fragmented PKS genes. However, long read sequencing technologies have much higher error rates than short read sequencing, with PacBio having an error rate of up to 15% and Oxford Nanopore having an error rate of up to 40%.[14],[16] In contrast, Illumina has been reported to have an error rate of 0.1%[17], sacrificing long read length for a low error rate. We are attempting to mitigate the trade-off between short read and long read sequencing by assembling genomes from short read sequencing more accurately.

3.3 Conclusion

Problems were identified in PKS gene clusters in *Streptomyces* genomes sequenced by Illumina. Simulations showed that PKS clusters often misassemble in the polyketide synthase genes, resulting in truncated genes or fragmented pathways. SPAdes and Biosynthetic SPAdes were compared to evaluate Biosynthetic SPAdes as a possible solution the misassembled pathways, but it did not prove to be reliable. Further research using RNA-Seq alignment algorithms or hybrid assembly algorithms is required to establish a reliable method for assembling PKS gene clusters. Assembly issues like this can lead to incorrect entries in databases that are used to identify the same gene clusters in other genomes, which further compounds the inaccurate estimations of identified BGCs and sequences belonging to the same gene cluster in a genome. Finding a way to mitigate problems in genome assembly allows for correction of reference genomes and BGC sequences for better identification of BGCs.

3.4 Methods

3.4.1 Identification of PKS Sequence Assembly Problems

To study the potential problems associated with assembling reads in areas of repetitive sequences in genomes, long PKS pathways with 10 or more modules from *Streptomyces* genomes were selected and used to simulate Illumina sequencing reads using the program ART version ART-MountRanier-2016-06-05.[10] Reads were simulated at 30X coverage with a read length of 150 bp. These reads were assembled using SPAdes version 3.14[2] and Biosynthetic

SPAdes.[18] These assembled reads were then entered into antiSMASH[9] and evaluated for the ability of the assembly software to correctly reconstruct the gene clusters. Commands used to simulate raw reads and assemble them are available in **Appendix V.**

To evaluate whether the assembly issues seen in isolated BGC simulations are prevalent in sequenced *Streptomyces* genomes, 40 *Streptomyces* genomes that had been sequenced by Illumina sequencing were selected from GenBank and entered into antiSMASH.[9] The resulting gene clusters identified by antiSMASH were inspected for evidence of fragmented pathways. The KnownClusterBlast results were used to identify PKS pathways that had matching regions of tailoring genes but broken or truncated PKS gene sequences.

3.4.2 ART Naphthomycin simulations

To visualize the issues with genome assembly in PKS regions, the sequence of the naphthomycin A gene cluster (MiBIG accession BGC0000106, GenBank accession GQ452266.1) was used to generate simulated Illumina raw reads using ART version ART-MountRanier-2016-06-05.[10] The naphthomycin A cluster was chosen as it appeared to have an assembly error in the PKS genes. Reads were simulated at 30X coverage with read lengths of 100 bp, 125 bp, and 150 bp. Reads were also simulated at 60X and 120X coverage with a read length of 150. These reads were then assembled using SPAdes version 3.14[2] and Biosynthetic SPAdes.[18] These assembled reads were then run through antiSMASH[9] and evaluated for the ability of the assembly software to reconstruct a complete gene cluster.

3.4.3 Naphthomycin isolation

Streptomyces sp. PBH53 was streaked to isolated colonies on ISP2 agar and incubated at room temperature for one week. Single colonies were inoculated into thirty 125 mL flasks containing 50 mL ISP2 broth. The flasks were incubated for one week at room temperature on a rotary shaker. The cells were removed by centrifugation at 4000 rpm and 4°C for 30 minutes. Secondary metabolites were extracted using 10 g/L each of Amberlite XAD-16N and XAD-7HP (Sigma-Aldrich, Oakville, Ontario, Canada) beads for 12 hours. The beads were eluted in 40 mL methanol for 2 hours, and then filtered off. The solvent was removed under reduced pressure, and the resulting crude extract was dissolved to 2 mg/mL in 1:1 water:methanol.

The peak corresponding to naphthomycin A at $[M+H]^+$ 720 was identified by HPLC-MS. MS data were collected using a Shimadzu LCMS-2020 coupled to a Shimadzu 20A Prominence system with a SPD-20A Prominence UV/VIS detector at wavelengths 210 and 254 nm (Shimadzu, Columbia, Maryland, USA). Chromatographic separation was performed using a ProntoSIL Eurobond C18 column (125 x 4.0 mm, 5.0 µm; Bischoff Chromatography, Leonburg, Germany) at 30°C with a flow rate of 1.00 mL/min. The mobile phase was composed of water with 0.1% formic acid (A) and acetonitrile with 0.1% formic acid (B). The 25 min mobile gradient started with mobile phase A being maintained at 95% for 2 min, then increasing to 95% mobile phase B over 19 min. The mobile phase was then maintained at 95% solvent B for 1 min, followed by decreasing to 5% solvent B over 0.5 min, then 2.5 min of column equilibration at 95% solvent A. MS was performed in ESI+ mode with the m/z range set to 250-1200.

$[M+H]^+$ 720 was purified by first subjecting the crude extract to silica column chromatography with a gradient from 50% ethyl acetate in hexanes to 100% ethyl acetate, then

to 15% methanol in ethyl acetate. The resulting fractions were tested for the presence of $[M+H]^+$ 720 by LC-MS. The compound was detected in the fraction with 100% ethyl acetate, and the fraction with 15% methanol in ethyl acetate. Those fractions were combined and further purified by preparative TLC with a mobile phase of 100% ethyl acetate. The silica was scraped from the plate and eluted using methanol. The extract was dried under air and dissolved to 2 mg/mL in methanol.

$[M+H]^+$ 720 was purified from the extract using an Agilent 1100 Series HPLC with an Agilent 1100 Series Fraction Collector (Mississauga, Ontario, Canada) with a ProntoSIL Eurobond C18 column (125 x 4.0 mm, 5.0 µm; Bischoff Chromatography, Leonburg, Germany) at 30°C with a flow rate of 1.00 mL/min. The mobile phase was composed of water with 0.1% trifluoroacetic acid (TFA) (solvent A) and acetonitrile with 0.1% TFA (solvent B). The 25 min mobile gradient started with mobile phase A being maintained at 95% for 2 min, then increasing to 95% mobile phase B over 19 min. The mobile phase was then maintained at 95% solvent B for 1 min, followed by decreasing to 5% solvent B over 0.5 min, then 2.5 min of column equilibration at 95% solvent A. Fraction collection occurred from 15.75 min to 16.05 min. The resulting fractions were combined and dried under air.

NMR analysis was carried out using a Bruker Avance III with cryoprobe operating at 600 MHz for ^{1}H and 150 MHz for ^{13}C data collection. The cross peaks from the HSQC spectrum were queried in the Small Molecule Accurate Recognition Technology (SMART) NMR program.[11]

3.4.4 Whole genome sequencing and assembly

A single colony of *Streptomyces* sp. PBH53 was inoculated into 50 mL of ISP2 broth and cultured for 7 days at room temperature in a shaking incubator. Genomic DNA was isolated using the Monarch Genomic DNA Purification Kit (New England BioLabs, Whitby, Ontario, Canada). The DNA was eluted off of the column with 50 µL of Tris-EDTA (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) heated to 65°C. The resulting DNA was sequenced using the Illumina NextSeq platform by the Microbial Genome Sequencing Center (Pittsburgh, Pennsylvania, USA), with library preparation performed with the Illumina Nextera kit. The resulting paired-end raw reads were trimmed first using the PATRIC Fastq Utilities[19] trimming function, then Trimmomatic[20] with sliding window trimming averaging across 4 bases and an average required quality of 20. The trimmed reads were then assembled using SPAdes version 3.14[2] by setting the untrusted contig to the naphthomycin A BGC sequence. Biosynthetic SPAdes[18] was also used to assemble the *S.* sp. PBH53 genome. The different assembly methods were evaluated by their ability to reconstruct the Naphthomycin A gene cluster.

3.5 References

[1] Li, Z., Chen, Y., Mu, D., Yuan, J., Shi, Y., Zhang, H., Gan, J., Li, N., Hu, X., Liu, B., et al. (2012). Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. Brief Funct. Genomics *11*, 25–37.

[2] Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., et al. (2012). SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. J. Comput. Biol. *19*, 455–477.

[3] Nagarajan, N., and Pop, M. (2013). Sequence assembly demystified. Nat. Rev. Genet. *14*, 157–167.

[4] Langmead, B. (2015). ADS1: De Bruijn graphs and Eulerian walks. Youtube. Accessed 10-29-2020.

[5] Langmead, B. (2015). ADS1: When Eulerian walks go wrong. Youtube. Accessed 10-29-2020.

[6] Rizzi, R., Beretta, S., Patterson, M., Pirola, Y., Previtali, M., Della Vedova, G., and Bonizzoni, P. (2019). Overlap graphs and de Bruijn graphs: data structures for de novo genome assembly in the big data era. Quant. Biol. *7*, 278–292.

[7] Weissman, K.J. (2015). Uncovering the structures of modular polyketide synthases. Nat. Prod. Rep. *32*, 436–453.

[8] Pérez-Victoria, I., Oves-Costales, D., Lacret, R., Martín, J., Sánchez-Hidalgo, M., Díaz, C., Cautain, B., Vicente, F., Genilloud, O., and Reyes, F. (2019). Structure elucidation and biosynthetic gene cluster analysis of caniferolides A-D, new bioactive 36-membered macrolides from the marine-derived Streptomyces caniferus CA-271066. Org. Biomol. Chem. *17*, 2954–2971.

[9] Blin, K., Pascal Andreu, V., de Los Santos, E.L.C., Del Carratore, F., Lee, S.Y., Medema, M.H., and Weber, T. (2019). The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. *47*, D625–D630.

[10] Huang, W., Li, L., Myers, J.R., and Marth, G.T. (2012). ART: a next-generation sequencing read simulator. Bioinformatics *28*, 593–594.

[11] Zhang, C., Idelbayev, Y., Roberts, N., Tao, Y., Nannapaneni, Y., Duggan, B.M., Min, J., Lin, E.C., Gerwick, E.C., Cottrell, G.W., et al. (2017). Small Molecule Accurate Recognition Technology (SMART) to Enhance Natural Products Research. Sci. Rep. *7*, 14243.

[12] Gosse, J.T., Hill, P., Dowd, S.E., and Boddy, C.N. (2015). Draft Genome Sequence of *Streptomyces* sp. Strain PBH53, Isolated from an Urban Environment. Genome Announc *3*.

[13] Langmead, B. (2015) ADS1: Third law of assembly: repeats are bad. Youtube. Accessed 10-29-2020.

[14] Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

[15] Fu, S., Wang, A., and Au, K.F. (2019). A comparative evaluation of hybrid error correction methods for error-prone long reads. Genome Biol. *20*, 26.

[16] Laver, T., Harrison, J., O'Neill, P.A., Moore, K., Farbos, A., Paszkiewicz, K., and Studholme, D.J. (2015). Assessing the performance of the Oxford Nanopore Technologies MinION. Biomol. Detect. Quantif. *3*, 1–8.

[17] Fox, E.J., Reid-Bayliss, K.S., Emond, M.J., and Loeb, L.A. (2014). Accuracy of Next Generation Sequencing Platforms. Next Gener. Seq. Appl. *1*.

[18] Meleshko, D., Mohimani, H., Tracanna, V., Hajirasouliha, I., Medema, M.H., Korobeynikov, A., and Pevzner, P.A. (2019). BiosyntheticSPAdes: reconstructing biosynthetic gene clusters from assembly graphs. Genome Res. *29*, 1352–1362.

[19] Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E.M., et al. (2020). The PATRIC Bioinformatics Resource Center: expanding data and analysis capabilities. Nucleic Acids Res. *48*, D606–D612.

[20] Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120.

Chapter 4: Implications and Future Directions

4.1. Elicitors for Activation of Silent Gene Clusters

Natural product discovery is often limited by transcriptional regulation of biosynthetic gene clusters (BGCs), and the available genomic information indicates enormous potential for secondary metabolite production but not all BGCs are readily expressed. Many clusters require some kind of stimulus to activate their expression.[1] In some cases, this stimulus can take the form of altered culture conditions, as seen with the One Strain-Many Compounds approach.[2] Other approaches activate silent BGCs by altering genetic information to change transcriptional regulation and thus require the host to be transformable.[3]-[5] In this work we use elicitors, which are similar to co-culture in that exposure of the strain to low concentrations of foreign molecules results in changes to secondary metabolism and activation of cryptic gene clusters.[6],[7]

Elicitors are an effective technique for discovery of new natural products. Elicitor screens can be carried out in a high-throughput manner, and can be applied to any cultivatable bacteria, unlike genetic methods which require the host to be transformable. Metabolomics studies increase the ease with which the effects of elicitors can be observed and quantified, due to the use of statistical analysis.

By using trimethoprim (TMP) and atenolol (AT) as elicitors to induce secondary metabolism in *Streptomyces* sp. PBH53 and *Streptomyces armeniacus*, we have identified four upregulated molecules of interest, 3 of which may be novel natural products. To complete this study, the structure of each molecule must be elucidated. Exposing *S.* sp. PBH53 to sublethal

concentrations of TMP resulted in the discovery of $[M+H]^+$ 732, which requires more scaled-up production, isolation, and further NMR experiments to determine its structure. Culturing the same bacteria with AT led to the upregulation of $[M+Na]^+$ 773, which likely corresponds to the known natural product salinomycin, and requires validation against a salinomycin standard by HPLC-MS. The addition of TMP to *S. armeniacus* cultures resulted in global metabolomic changes, with the selection of $[M+Na]^+$ 411 as the compound of interest. Purification of the compound and NMR spectroscopy is required for full characterization of the molecule. In contrast to the global regulation seen as a result of TMP, AT acted as a pathway-specific regulator in *S. armeniacus* leading to the identification of $[M+3H]^{3+}$ 768. This molecule has been identified as a peptide based on the presence of b or y ions corresponding to a five amino acid sequence of alanine-threonine-glycine-alanine-threonine. The peptide has been subject to Kupchan partitioning and the full amino acid sequence will be determined by MS^n experiments.

Beyond structure elucidation of the three unknown molecules, future work will include identification of the biosynthetic gene cluster corresponding to each compound. This will largely rely on structure elucidation to determine what kind of BGC is responsible for the production of each molecule. In addition, activity assays will be used to determine if the metabolites of interest have antibiotic activity and therefore therapeutic relevance. Then, identification of the cellular targets and mechanisms of action of each of the molecules can be determined based on molecular structure, genomic context, and activity assays. Future experiments will also study the mechanism by which TMP and AT activation expression of these cryptic gene clusters in each organism.

The three unknown molecules upregulated in response to the added elicitors may represent novel natural products contributing to closing the gap between an organism's biosynthetic potential and its known biosynthetic output. Elicitor-based methods for natural product discovery prove to be a useful tool for finding and exploiting more of the potential of bacteria to produce secondary metabolites.

4.2 A Secondary Metabolite-Based Approach to Genome Assembly

Advancements in whole genome sequencing techniques have enabled researchers to sequence genomes at low cost with fast turnaround times, especially with Illumina short read sequencing. However, assembling the resulting reads remains a highly challenging problem. Repetitive sequences make assembly difficult, and can result in fragmented sequences or incorrect incorporation of reads into genomic sequences.[8] De Bruijn Graphs, a common algorithm for genome assembly, rely on a single Eulerian walk through the assembly graph, whereas that is often not what is observed in genomic assemblies, especially when repetitive sequences are involved.[9] The resulting genome could be full of errors, with no way to evaluate those errors in *de novo* sequencing. This problem is further compounded in databases such as antiSMASH database[10] where the BGCs used as references may have the same kind of assembly errors, perpetuating the availability of incorrect genomic information.

Type I polyketide synthase (PKS) pathways present a problem in genome assembly in the form of repetitive sequences in the acyl-transferase (AT) domains, which results in truncated PKS genes and fragmented BGCs in the assembly, as seen in sequencing simulations of long Type I PKS pathways but not in pathways that utilize *trans*-AT domains. This assembly

error results in both the overestimation of the number of gene clusters an organism has and an underestimation of the number of pathways with an identified natural product.

This problem exists in published *Streptomyces* genomes available on NCBI, and in the genome of *S.* sp. PBH53 which has five Type I PKS pathways corresponding to three true gene clusters that have matching tailoring genes to their top KnownClusterBlast result in antiSMASH. These BGCs either show a full pathway with truncated and concatenated PKS genes in the case of Naphthomycin A, or show two clusters with partial pathways at each end of the BGC and breaks in the sequence in the PKS genes. In the case of Naphthomycin A, *S.* sp. PBH53 must have a complete pathway as it was isolated from extracts of the culture media.

SPAdes, which relies on De Bruijn Graph assembly, is not able to resolve the repetitive sequences of PKS domains. Biosynthetic SPAdes provides options for how a gene cluster may be assembled, but was not able to correctly assemble the Naphthomycin A cluster.

We propose that more accurate assemblies could be achieved by using verified sequences of BGCs to provide a scaffold for the raw reads to assemble on to. The continuation of this research will involve the use of Bowtie, a program designed for RNA-Seq analysis, to assemble raw reads onto literature sequences for PKS pathways, then assemble the rest of the genome around them. Another approach will be to take advantage of the current practice of sequencing a genome by both long-read and short-read sequencing, and assemble short reads using the literature sequences of PKS pathways as long reads.

We anticipate that these approaches will provide more accurate predictions of BGCs, and that researchers in natural product discovery will be able to use the knowledge of the

problems associated with PKS cluster assembly to expand the practice of predicting products of gene clusters based on genomic sequence.

4.3 References

[1] Zarins-Tutt, J.S., Barberi, T.T., Gao, H., Mearns-Spragg, A., Zhang, L., Newman, D.J., and Goss, R.J.M. (2016). Prospecting for new bacterial metabolites: a glossary of approaches for inducing, activating and upregulating the biosynthesis of bacterial cryptic or silent natural products. Nat. Prod. Rep. *33*, 54–72.

[2] Bode, H.B., Bethe, B., Höfs, R., and Zeeck, A. (2002). Big Effects from Small Changes: Possible Ways to Explore Nature's Chemical Diversity. ChemBioChem *3*, 619–627.

[3] Ji, C.H., Kim, J.P., and Kang, H.S. (2018). Library of Synthetic Streptomyces Regulatory Sequences for Use in Promoter Engineering of Natural Product Biosynthetic Gene Clusters. ACS Synth. Biol. *7*, 1946–1955.

[4] McArthur, M., and Bibb, M.J. (2008). Manipulating and understanding antibiotic production in Streptomyces coelicolor A3(2) with decoy oligonucleotides. Proc. Natl. Acad. Sci. U.S.A. *105*, 1020–1025.

[5] Du, D., Katsuyama, Y., Onaka, H., Fujie, M., Satoh, N., Shin-ya, K., and Ohnishi, Y. (2016). Production of a Novel Amide-Containing Polyene by Activating a Cryptic Biosynthetic Gene Cluster in *Streptomyces* sp. MSC090213JE08. ChemBioChem *17*, 1464–1471.

[6] Traxler, M.F., Watrous, J.D., Alexandrov, T., Dorrestein, P.C., and Kolter, R. (2013). Interspecies Interactions Stimulate Diversification of the *Streptomyces coelicolor* Secreted Metabolome. MBio *4*.

[7] Seyedsayamdost, M.R. (2014). High-throughput platform for the discovery of elicitors of silent bacterial gene clusters. PNAS *111*, 7266–7271.

[8] Langmead, B. (2015). ADS1: Third law of assembly: repeats are bad. Youtube. Accessed 10-29-2020.

[9] Langmead, B. (2015) ADS1: When Eulerian walks go wrong. Youtube. Accessed 10-29-2020.

[10] Blin, K., Pascal Andreu, V., de Los Santos, E.L.C., Del Carratore, F., Lee, S.Y., Medema, M.H., and Weber, T. (2019). The antiSMASH database version 2: a comprehensive resource on secondary metabolite biosynthetic gene clusters. Nucleic Acids Res. *47*, D625–D630.

Appendix I. *Streptomyces* sp. PBH53 and Trimethoprim

Table S1.1. Metabolomics data from *Streptomyces* sp. PBH53 cultured with 33 µM trimethoprim (TMP).

	Minus_TMP_A	Minus_TMP_B	Minus_TMP_C	Minus_TMP_D	Minus_TMP_E	Minus_TMP_F	Plus_TMP_A	Plus_TMP_B	Plus_TMP_C	Plus_TMP_D	Plus_TMP_E	Plus_TMP_F
Class	1	1	1	1	1	1	2	2	2	2	2	2
RT5.22_702.27997	2.08E+07	1.61E+07	2.02E+07	1.32E+07	1.46E+07	1.12E+07	2.11E+07	8813330	1.83E+07	1.38E+07	1.14E+07	1.43E+07
RT3.11_201.12342	2.02E+07	2.09E+07	2.50E+07	1.96E+07	1.81E+07	1.44E+07	1.95E+07	1.34E+07	2.04E+07	1.74E+07	1.53E+07	1.94E+07
RT4.4_829.3353	1.93E+07	1.37E+07	1.30E+07	2.36E+07	1.47E+07	1.78E+07	2.08E+07	1.67E+07	1.72E+07	1.44E+07	2.01E+07	1.45E+07
RT7.59_684.20086	179676.8125	422175.25	17439.55273	954096	7672.045898	1602332.25	201496.3125	11738.24805	365183.4375	1155345.625	1563123.5	6921.479004
RT5.22_742.27281	1.90E+07	1.62E+07	2.00E+07	1.44E+07	1.53E+07	1.26E+07	2.03E+07	9897084	1.80E+07	1.46E+07	1.41E+07	1.48E+07
RT3.43_201.1235	1.65E+07	1.81E+07	2.09E+07	1.83E+07	1.72E+07	1.54E+07	1.66E+07	1.36E+07	1.64E+07	1.68E+07	1.68E+07	1.67E+07
RT3.94_1170.46848	1.22E+07	7957447	1.07E+07	1.29E+07	8041555.5	9341595	9346018	7365403.5	8025825.5	7126571	8917710	7791710
RT7.42_684.20092	171645.25	422175.25	11357.39258	619675	6600.11377	1106432	58566.15234	3490.977539	332159.5938	1155345.625	980107.3125	3597.787354
RT4.4_869.32738	1.20E+07	8553013	8385323	1.45E+07	8479428	1.11E+07	1.34E+07	1.06E+07	1.11E+07	8282898	1.17E+07	9062479
RT3.13_319.23286	1.17E+07	1.22E+07	1.55E+07	1.42E+07	1.24E+07	9815504	1.38E+07	9639808	1.52E+07	1.16E+07	8317369	1.21E+07
RT5.22_684.26946	1.13E+07	8967672	1.14E+07	7519110	8321907.5	6269564.5	1.20E+07	4868262	1.03E+07	7753629.5	6608865.5	7989108.5
RT3.13_583.33937	1.09E+07	1.04E+07	1.17E+07	1.05E+07	9903780	8344417	1.07E+07	8678003	1.05E+07	9795964	9221646	1.06E+07
RT3.07_547.34223	9917755	7923481	9795366	6467754	7359200.5	5208341.5	7947364	4349914.5	8831020	6147193.5	5603153	5553756
RT3.12_243.13293	9900041	1.07E+07	1.35E+07	1.16E+07	1.05E+07	8431434	1.14E+07	8101761	1.25E+07	9722235	7666907.5	1.01E+07
RT4.34_740.30087	9393239	1668736.625	1055994.625	1.40E+07	1.38E+07	1.32E+07	6510378.5	4033013.5	6055866.5	1.09E+07	1.64E+07	5738923
RT5.22_704.2796	8553099	6733585	8526936	5587590	5965224	4599845.5	8851284	3581996	7600969	5770939	4803342.5	5810848
RT4.33_700.30984	8422437	1509711.75	958268	1.41E+07	1.43E+07	1.32E+07	5018855.5	3121474.5	4411498.5	1.08E+07	1.88E+07	5031087.5
RT4.86_684.31501	8367066	2566119.25	3354179.25	4728522.5	2.58E+07	8202447	1368612.125	1970006.25	1659969.125	1.86E+07	2.18E+07	1.33E+07
RT3.23_573.32196	8092133	4653994	5953721	3765971.25	4842282.5	4163849	3989996.25	1888406.625	4550715	3208393.75	3633030	2841704.75
RT5.21_323.19938	7994891	6367175.5	8719412	5175844.5	5698646.5	4398424	8497787	3220432.5	7535017.5	5431848.5	4276566.5	5599184.5
RT3.23_595.30298	7420850	4775420	5610093.5	4013134	4555047	4244946.5	4349246.5	2444270.75	4687962.5	2811297	3607443.5	3235658.25
RT3.58_637.38981	7391755	6425274	8282243.5	5998004.5	7059065.5	5482915	6914477.5	4718838	6975213.5	5999944.5	6737109.5	6537244
RT5.22_744.2719	7385774.5	6537150	8307451	5904206.5	6186813.5	5154666.5	8339573.5	4020518.75	7216379.5	5881410.5	5982214	6032976
RT3.34_201.12348	6497571	5732407.5	6935627.5	5665031.5	6054215.5	5342699	5274884.5	3996159.25	5757087	5316990.5	5296226.5	5018826
RT3.42_401.23761	6271248.5	6897143	7993833	6948715.5	6128336	5197756.5	6522050	5079237.5	6155950	6160051.5	6134728.5	6322462
RT3.07_569.3239	6082654	5058985	5690462.5	4338021	4587863	3735070.25	5448378.5	3952744	6073376.5	4640636	3933006	4321345.5
RT3.59_659.37092	5291350.5	4262400	4812124	4334347	4391505	4040568.5	4492827	3635127.25	4601350	4166600.75	4210554	4200097
RT7.86_684.20077	74090.90625	307253.4375	11599.12012	1342734.125	5864.35498	1501668.625	362742.9063	26311.08203	249628.7813	356780.375	1200984.375	10552.87988
RT4.86_724.30711	5140287.5	1562908	2182940.25	3447949.75	1.16E+07	5401298.5	905354.5625	1260605.875	1069095.125	8516215	1.01E+07	7138796
RT4.4_471.12014	4876453.5	3256312	3242399.25	5893944	3318939.75	4258667	5206455.5	3864518.25	3977292.5	3514668	4974251	3373875.75
RT7.71_684.2008	179676.8125	422175.25	17439.55273	981127.125	7672.045898	1610615.5	336883.375	24547.83594	365183.4375	979612.0625	1563123.5	10294.0752
RT3.26_593.32523	4860332.5	3531627.5	2904418.75	3404038.75	5147657.5	4041641.75	2650616.5	3386093.25	3253699.5	3378315.25	5172329	3301182.75
RT3.58_201.12362	4560937.5	4127835.25	5195929.5	3718490.25	4578159	3298487.5	4583928	2775892.25	4524312	4133431	4313601.5	4084442.5
RT3.36_607.3387	4470861.5	4564499	3086678.75	4984418.5	5968572.5	5063219	3286075.5	4770269.5	3841349	5619391.5	7045752.5	5262503
RT4.4_847.34366	4023679	2831464.75	2757647.25	4946978	3433577	4048851.75	4601598.5	3732891	3610003	3094460.75	4118915.5	3009832.25
RT5.22_686.26818	3906063.5	3136441.75	4135027.25	2582765.25	2927932.25	2181674	4105402.75	1736326.125	3557647.25	2728247.75	2174267	2670896

RT4_722.28953	3903973.75	681020.625	506125.0625	4895989.5	5811433	5061371	1868034.125	1343763	2090239	4306034	6236332	2740889.25
RT7.72_612.18003	9314.318359	1938314.25	13297.18066	2084241.875	2118842.5	14740.52539	2011169.125	8089.919434	44323.50781	2153433.25	6514.444336	2076539
RT3.58_319.23264	3790303.5	3242032.25	3971770	2969553.5	3714559	2849478.5	3356209	2430239	3468195.5	3293962	3563610	3588160.25
RT3.36_585.35764	3789928.75	3429544.75	2423683	3794982	5109370.5	3711171.75	2323934	3375523	2678331.25	4710179	6630451	4001352.25
RT7.64_355.06828	17451.18945	1079142.125	15820.51563	969504.6875	1138833.125	95598.20313	758328.0625	32159.14453	68398.125	1209893.75	102286.7969	744200.875
RT5.22_215.14199	3642165	2997681.5	4159347.5	2453106.25	2684577.75	2137294.5	3969742.25	1525321.5	3718450.75	2514579	2059840.125	2512408.5
RT3.42_601.35307	1.25E+08	1.40E+08	1.83E+08	1.40E+08	1.22E+08	9.98E+07	1.19E+08	8.42E+07	1.24E+08	1.22E+08	1.19E+08	1.23E+08
RT3.34_587.33712	7.74E+07	6.12E+07	8.15E+07	5.84E+07	6.26E+07	5.24E+07	5.26E+07	3.14E+07	5.90E+07	4.94E+07	4.90E+07	4.74E+07
RT3.13_561.35805	3.55E+07	3.63E+07	4.59E+07	3.54E+07	3.10E+07	2.41E+07	3.55E+07	2.30E+07	3.59E+07	2.99E+07	2.52E+07	3.43E+07
RT3.43_623.33518	3.26E+07	3.35E+07	3.90E+07	3.50E+07	3.27E+07	2.95E+07	3.13E+07	2.66E+07	3.26E+07	3.21E+07	3.16E+07	3.30E+07
RT3.34_609.3178	2.55E+07	2.20E+07	2.58E+07	2.12E+07	2.16E+07	2.02E+07	2.05E+07	1.48E+07	2.18E+07	1.91E+07	1.89E+07	1.86E+07
RT4.19_482.25226	735438.8125	340390.3438	2259902.25	1344237.5	1847477.75	1515369	1411483.125	151991.75	627909.4375	1408140.375	1523873.625	1472921
RT4.33_482.25224	1892297.25	702032.4375	2195792.75	3901215.25	5347725	2315203.75	2378686.75	478214.9688	1314599.875	4780146	2430737.75	4071981.25
RT4.86_682.29763	1000253.813	117564.6875	193715.6719	1056747.5	3020111	1703251.25	194367.5781	76338.96094	114143.2891	1815732	3351120.25	878423.5
RT5.11_754.30031	3724325.5	2850379	2520520.75	2018532.5	1342079.625	1592588.125	1.17E+07	1.05E+07	1.08E+07	7843863	7273760	7566345
RT4.97_668.31817	2551721.75	1812571.375	2260309.25	1795915.75	1578332.5	1430964.125	1893585.75	1887666.875	1965480	1673649.75	2519586.75	1660193.75
RT4.23_566.18961	2152593.25	1348087.375	1890234.125	975720.125	804115.4375	795326.75	1579646.375	1149194.625	1178876.5	798948.75	973726.1875	1162978.75
RT4.26_315.08551	2510233.75	2664352.75	3809946.5	2932990	1057160.375	2164809.5	2656425.75	2214049.75	2934273.25	912068.5625	439787.75	1481608.625
RT3.27_571.34287	1735940.125	1319195.625	990674.3125	1264865.25	2326175.5	1596900.125	910862.1875	1211806.75	1145056.25	1699057.75	2740716.75	1223683.125
RT3.07_545.3628	1424267.375	1212110.25	974206.25	1497268.375	2003794	1325916.5	900417.8125	1304704.375	1049258	1550134.375	2092848.75	1324313.625
RT2.95_555.30945	1643363.75	982435.6875	1359361.25	804178.5625	888575.6875	799545.625	1110576.625	553134.375	1139275.625	711384.125	598341.5	675407.875
RT3.01_553.32943	1119067.25	741856.875	611033.5625	800715.25	1275249.875	885221.0625	685804.9375	995443.25	859088.375	1127042.5	1352529.125	833844.375
RT3.34_294.17333	1625821	1418566	1914361.25	1740218.25	2709631.25	1568345	1595255.625	1022973.25	1311612.625	1206534.125	1329980	1194247.125
RT4.05_849.35907	1620178.125	1053579.375	1425212.5	1937435.375	1143671	2445629.75	1853523	1728868.875	1643923.25	1944565.25	1563896.75	2387066.5
RT3.78_319.16384	1580719.5	1311795	1170046.5	1407186.25	1183945	1067453.125	1660200.625	1108372.625	1997695.25	1289636.5	1153578.5	1080481.875
RT4.75_722.29054	1002891.125	91375.63281	77589.13281	920905.0625	3351873.5	1374575.125	92442.4375	55643.17578	103407.0781	1763771	3146634.75	810686.1875
RT3.06_567.34574	1427394.25	1302987.5	937130	1449197.625	1876164.625	1407133.625	1155011.5	2145277	1280442	2075768.5	2460853.75	1856293.5
RT4.25_914.4334	1021948.25	1945949	2066651	1354754.75	452051.2813	747019.5625	956353.0625	953455.25	1190280.75	472265.625	200660.2969	905577.6875
RT4.41_849.35878	1303202.875	994993.875	933586.375	1570405	1111640.75	1387149.5	1584216.75	1196632.75	1248191.875	964652.8125	1451578.875	1033963.938
RT4.25_913.93206	1026337.625	1871320.375	1903872.75	1269941.875	389693.7813	685926.25	921796.875	868427.1875	1078653.25	432158.875	183243.2969	872007.9375
RT4.66_482.25183	1292499.25	354969.25	903986	2583898	3185234.75	1425066.375	1234188.25	263370.1563	705810.9375	2963550.5	1654960.375	2356330.75
RT0.79_120.04383	677068.75	685331.8125	1216752	842926.5	1152773.5	609065.75	868120.1875	556635.3125	578646.25	862254.9375	844522.8125	701480.1875
RT3.72_319.16176	1230754.75	882886.1875	994793.125	1150759.875	471646.375	604540.25	936451	1278188.625	962704.6875	533985.1875	420519.3438	783339.75
RT0.92_120.04377	562153.0625	580877.8125	1021025.375	715888.9375	966247.375	521098.0625	729218.75	471311.7813	483551.0625	758810.1875	736053.625	578114.625
RT3.34_401.23744	1193015.375	979921.125	1288623.625	1041118	1119895.875	908770.9375	980624.3125	610125.3125	910685.375	913841.4375	916861.125	824398.3125
RT5.12_714.30804	2397047.5	1995370.625	1715018.875	1254362.25	797051.875	988210.5	1.00E+07	8926254	9361344	5958670	5476678.5	5751225
RT4.75_682.29762	1001792.25	103803.2813	149846.4375	936426.25	2743522.5	1202861.25	161176.9688	72855.82813	124591.7813	1864239.625	2753746.25	903813.875
RT3.55_645.35546	3552050.75	2307022.25	2706019.5	2134676.75	2389275	2123251.75	2806282.5	1900618.125	3043253	2174405.25	2134944	1797517.625

RT5.11_323.19959		970091.5625	754749.4375	674067.5	474338.5313	343466.9375	393838.5938		4084202.25	3551343.25	3635572.75	2553122.25	2103060.75	2405370.75
RT5.12_215.14196		444716.875	341830.3438	312600.9063	214211.3906	136444.9219	182113.1406		1737743	1560235.125	1607687	1108175.875	942596.25	1016196.188
RT4.36_898.33765		3080701	2259006.75	2169117.75	2561875.75	2600193.25	2394323.5		3114528.75	2658982.25	3219305.25	2199683.75	2531574.25	2049043.625
RT5.12_341.20944		301761.9375	236649.3438	0	151247.125	0	128054.375		1302864.625	1141800.125	1090799.375	779212.3125	644034.4375	750995.625
RT4.4_811.32268		3396807.25	2284059.25	2239678.75	4027501.25	2516780	3075224.25		3656521.25	2699078.25	2998047.75	2409201.25	3612828.25	2448851.5
RT4.07_292.15049		994406.8125	568992.1875	909917.5	1941290.875	1418146.375	1668064.75		1209260.5	2280431	982601.6875	1595198.875	1398699.5	1576846.375
RT3.94_1192.44699		3188750.75	2601755	3258458.25	3356501.5	2345333.75	2857840		2780724.75	2345226.5	2356744.75	2219763.75	2440200	2406678.75
RT3.43_320.15409		3120956.5	3179975.75	3088976.25	3356150.25	3126937.5	3454608		3301666	3145621.75	3252157.75	3109779.75	3055462.25	3065760.25
RT4.59_726.32501		1046883.75	345057.4063	357673.9688	2071163.5	2211467.25	1900299		1079054.875	450327.6875	858611.5	1305752.125	1662325.375	729714.25
RT5.12_696.29538		613374.75	499711.625	448310.75	314018.4688	200258.2344	246179.4219		2572615	2314490	2365028.5	1599542.875	1366225.75	1538259.5
RT3.43_301.17995		3038736	4073987.25	3336640.75	3001805.75	2333702.75	2541262.75		3503030	2591052.75	2364283	2767444.25	2920926.25	3259204.5
RT3.85_686.32908		1164741.625	892705.125	621924.875	1844225	947476.9375	1089243		1745392.375	1249698	1857444.125	997864	1191552.625	897188
RT5.22_341.20933		3020382.75	2480122.75	3215612.75	1949332.125	2131380.25	1643809.875		3217793.25	1256288.875	2892841.25	2117980.75	1688590.25	2137899.5
RT4.85_666.30485		2281141.5	630015.9375	892886.5	1211173.875	7241629	2137736		335072.8438	510296.5313	420427.3125	4939131	6207485.5	3457858
RT3.13_144.10122		2966526.5	3042979.75	3476691.75	2704913.25	2545214	1991111.875		2774916.25	1946895.5	2808968.5	2518667	2410700.5	2924967.75
RT4.59_686.32893		1043297.5	472678.75	552241	1707603.75	1525109.375	1422321.125		1403737.25	696547.625	1182662	1215987.125	1309929	827169.375
RT3.43_283.12766		1421059.25	1557661	1796367.625	1614261.375	1501252.875	1334782.375		1418536.875	1083331	1496552	1488170	1483151.5	1439697.75
RT4.85_590.23535		1202632.125	308678.8438	454994.1563	631987.1875	3800422.25	1129904.25		188095.2344	223365.5781	208275.6563	2598356	3120779	1712863.625
RT3.55_623.37249		2767503.75	1580600.5	2043654.125	1435948.125	1936403.75	1438264.125		2186986.5	1264988.75	2301069.5	1499106.25	1923407.875	1295682.25
RT4.4_323.20016		1184935.625	874006.1875	874695.125	1464546.625	882324.8125	1077006.25		1288067	906156.5625	975898.4375	916623.5625	1278976.25	866059.9375
RT3.33_313.14594		2754327.5	2627336.25	2765924.25	2888908.5	3075405.5	2730457		2526245.5	2285663	2533579.5	2482648	2522239.25	2451528.5
RT3.9_539.3061		546114.25	280365.875	248787.1406	1329804	2709753	1014638.125		566723.0625	601896.25	611199.375	2618495.25	2184756.25	1344111.5
RT4.09_722.28876		605179.75	61868.59766	54802.44922	848399.875	1279007.125	935101.75		148437.8125	97337.35938	165278.8125	825251	1230608.75	438432.9375
RT4_700.30921		2671606.75	354184.5938	257236.8438	3778640.75	5362931.5	3890793		1113174.125	813213.125	1145459.75	3730700	5870160	1878772.375
RT4.4_813.34219		942819.0625	670834.5625	651809.6875	1164554.125	693236.0625	873209.4375		1023788.75	751691.6875	798357	697692.375	1012140.688	767265.125
RT3.34_387.22208		2474317.5	2032564.5	2512659	2147747.25	2326094.5	1990099.875		1976402	1316477.375	1978132.5	1775641.25	1892547.125	1714755.625
RT3.08_305.217		2460107	1834000.75	2484286	1806792.75	1962009.25	1443787.375		2125779.75	1301533.75	2235570	1677746	1604026.5	1542733.625
RT4.86_722.28991		902911.3125	47522.46875	85149.74219	1094934.625	2513067.5	1770110		128944.0625	30112.41016	77186.48438	1394674.375	2700964.75	727734.75
RT3.34_187.10803		2419966.5	2242241.75	2665950	2200351.5	2353915.5	2090412.25		1941707.875	1333972.125	2206762.25	1918712.75	2021913.25	1896721.875
RT3.58_319.16415		2381499.5	1913321.625	2306811.25	1732658	2223147	1736667.625		2006804.625	1421429	2203578.25	1977120.875	2311668.5	2155904.5
RT4.34_682.29772		2294638.25	522063.5313	344216.0625	3590287.25	3836356.75	3557089.25		1462536.625	972974.75	1364566	2816189.75	4851315.5	1342003.125
RT3.94_604.70913		2096224.875	1987149	2328894.25	1995965.125	1633375.75	1828486.375		2054419.25	1877464.5	1766322.875	1649482.625	1638887.875	1915186.5
RT5.22_149.09534		2061683	1715619.625	2373671.5	1418316.75	1582977.25	1204318.75		2260383.5	885553.75	1973812.125	1525985.125	1200765.5	1535883.25
RT4_682.29769		1707210	198781.4688	149210.7656	2316157	3446071.25	2519522.25		662711	384512.7188	705212.25	2353432.5	3565059.5	1156664.25
RT4.58_890.38225		641672.625	415984.2813	130044.4141	35027.20313	98911.16406	114971.5547		142273.7344	3332842.75	188043.2656	905923	60306.57813	87019.5
RT4.79_710.32794		2082611.25	1257189.5	2247753.5	2172065.5	3055895	1720792.875		1694951.375	1301561.875	1475627.625	2184737	2793876	1719642.25
RT4.33_718.32113		1239569.25	264322.375	190164.4219	1769094	1710826.125	1575120.375		769846	489469.1875	702840.5	1291103.875	2283314.5	713119.375

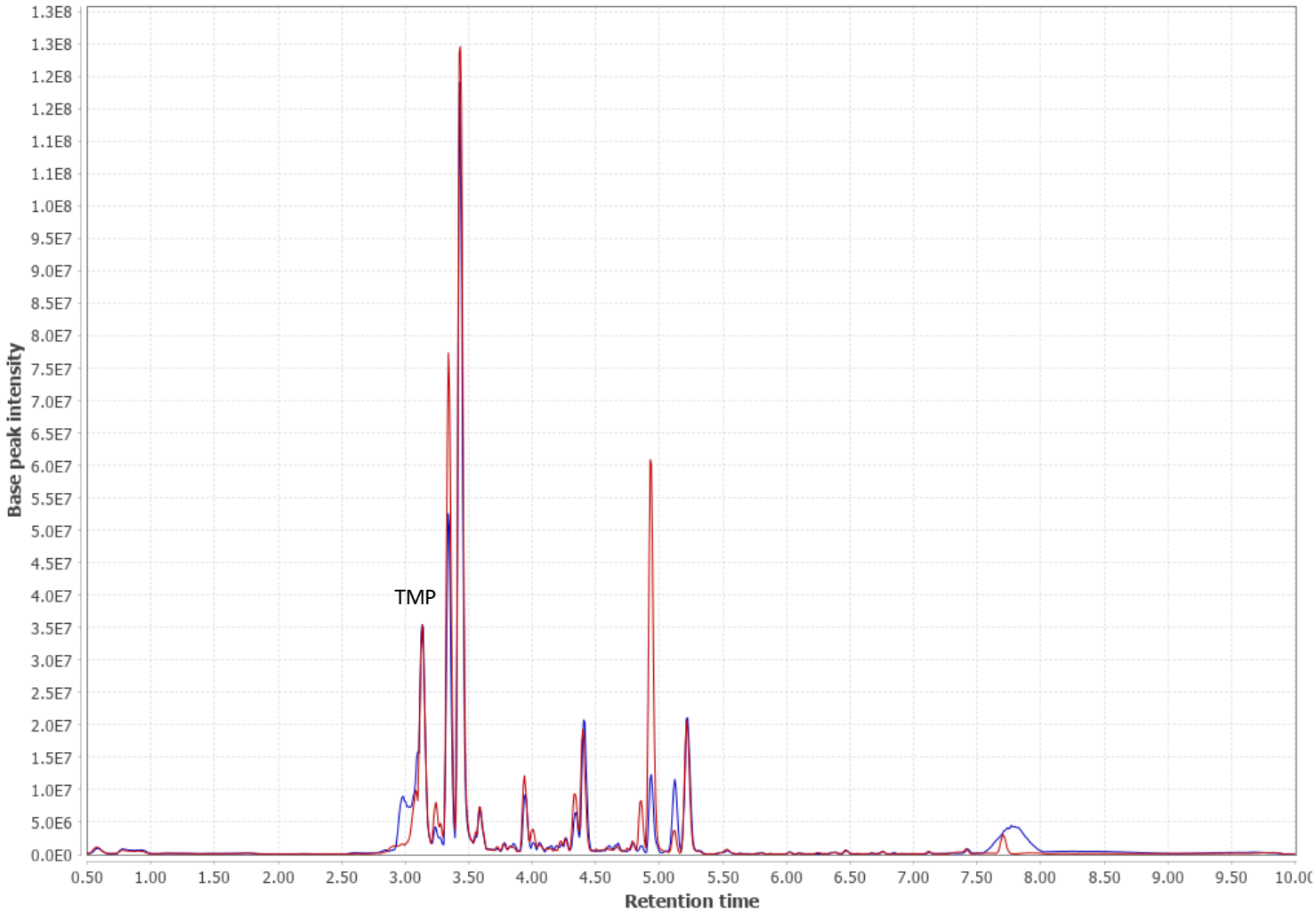

Figure S1.1. Total Ion Chromatogram (TIC) overlay of *S*. sp. PBH53 with 33 µM TMP (blue) and without TMP (red). Data were collected from *m/z* 100-2000.

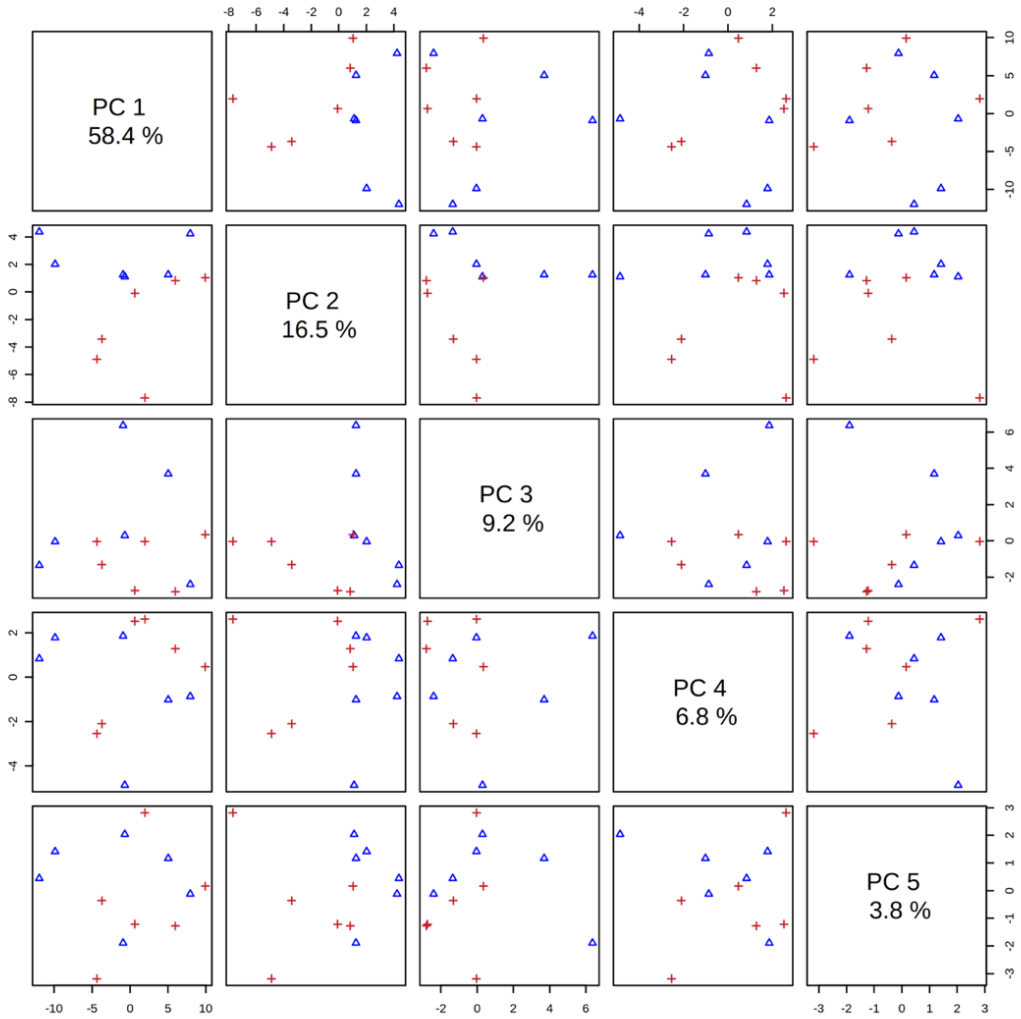

Figure S1.2 Pairwise scores plots for the first 5 principle components of *S.* sp. PBH53 cultured with 33 µM TMP (Red; n=5) and without TMP (Blue; n=6). The first two PCs represent 74.9% of the variance seen in the dataset.

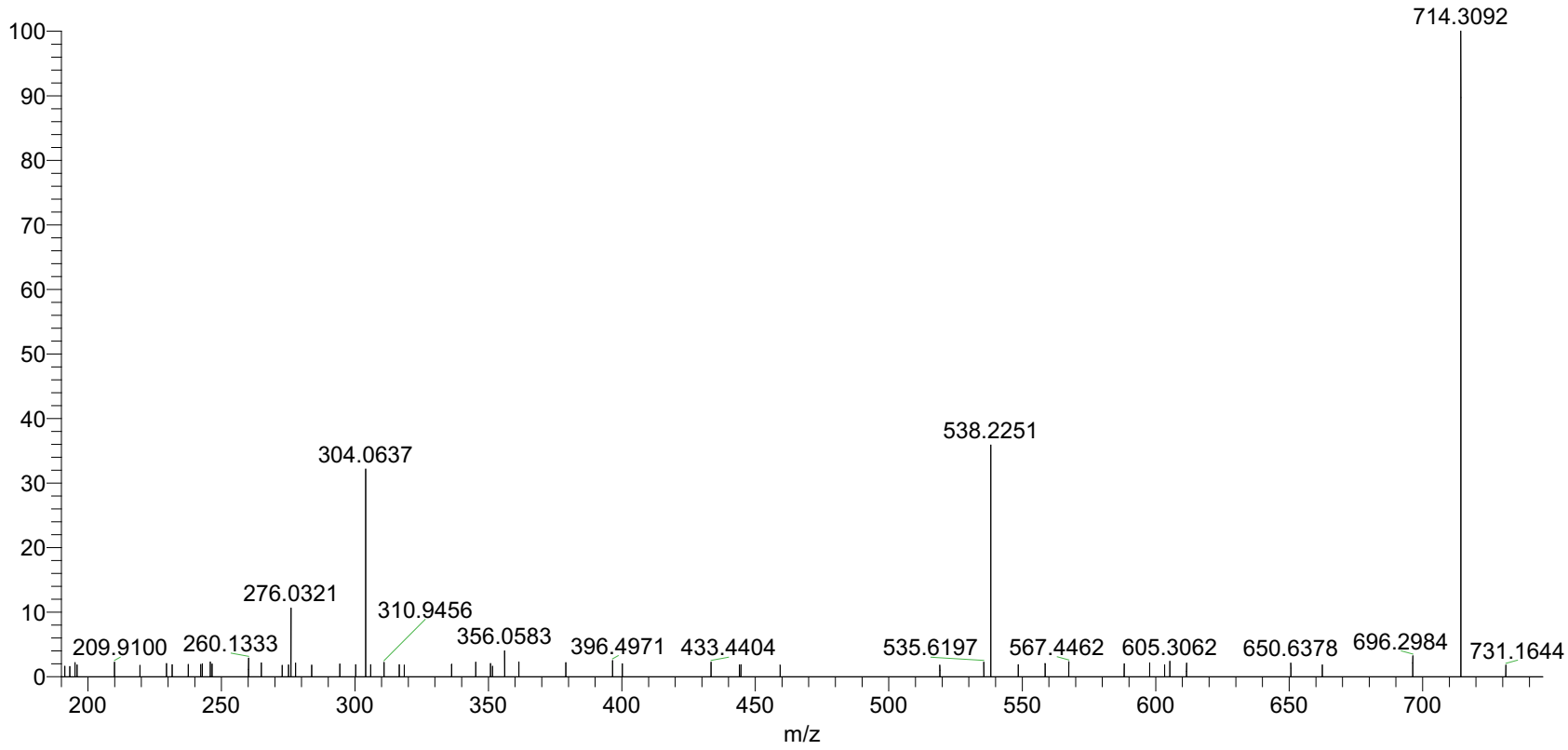

Figure S1.3. Results of tandem MS performed on $[M+H]^+$ 732 at retention time 5.1 mins, with a mass detection range of *m/z* 100-2000.

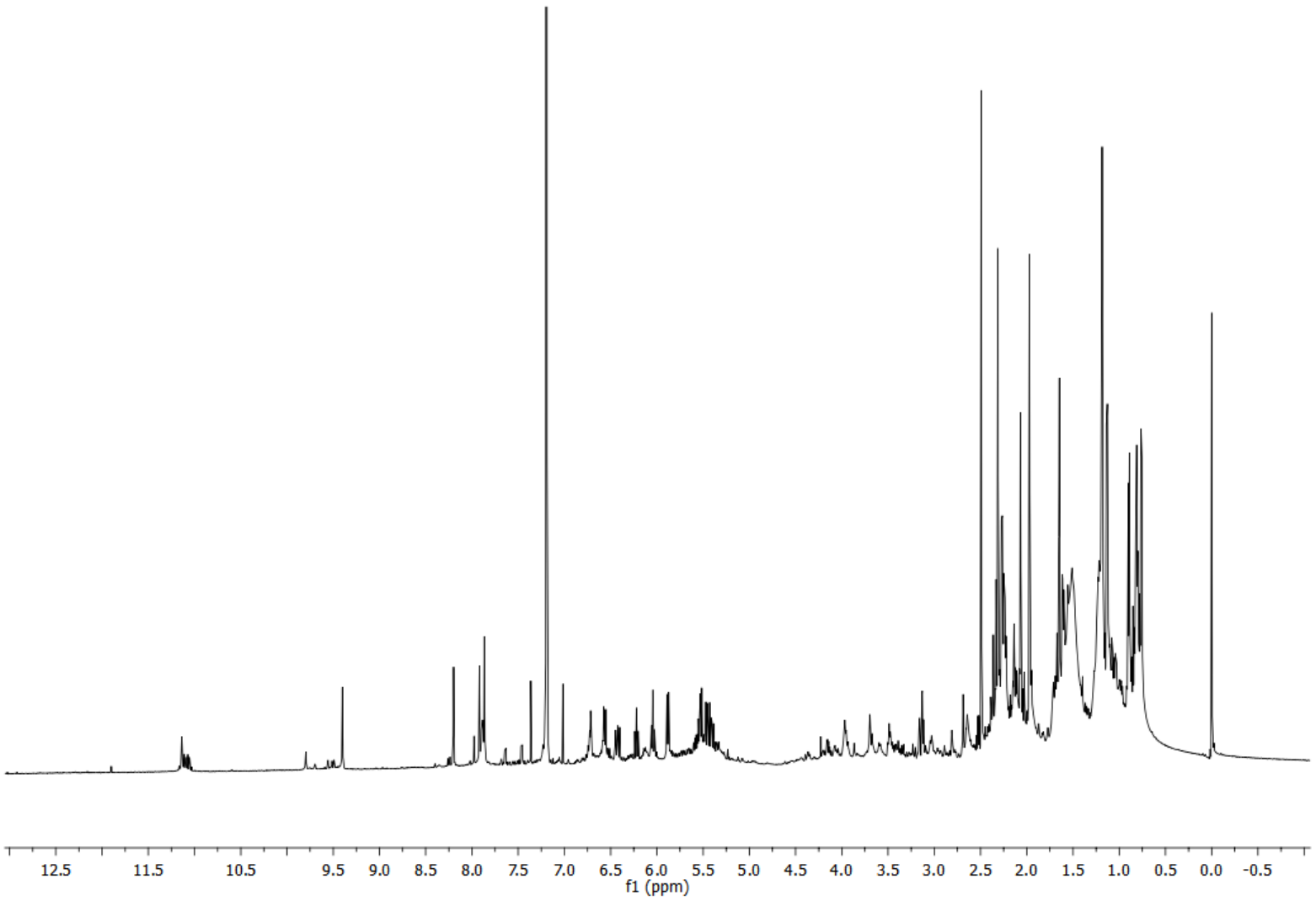

Figure S1.4. ^{1}H spectrum of $[M+H]^{+}$ 732. Not enough compound was purified to obtain further NMR spectra or to elucidate the structure of $[M+H]^{+}$ 732.

Appendix II. *Streptomyces* sp. PBH53 and Atenolol

Table S2.1. Metabolomics data from *S.* sp. PBH53 cultured with 33 µM atenolol (AT).

	MINUS_AT_A	MINUS_AT_B	MINUS_AT_C	MINUS_AT_D	MINUS_AT_E	MINUS_AT_F	PLUS_AT_A	PLUS_AT_B	PLUS_AT_C	PLUS_AT_D	PLUS_AT_F
Class	1	1	1	1	1	1	2	2	2	2	2
RT5.21_380.1251	401004.5	303720.5313	293046	319244.4063	344675.25	287977.7188	367361.2188	356992.7813	243904.9531	311558.4688	287850.75
RT3.27_593.3271	1344142	965185.625	862140.75	763804.3125	752314.1875	1102943.75	1058771.375	1193628.625	540608.8125	699380.1875	1072509
RT5.57_629.3669	307337.2188	147552.5469	201887.6563	136451.7969	153098.3438	129979.2891	172359.0781	167054.125	108177.8281	184118.4219	180294.8281
RT3.34_294.6747	1691815.5	988778.9375	973112.75	883537.9375	947667.3125	1123103.5	1214225.5	1293616.125	740505.25	797152.3125	1034097.688
RT0.52_294.1539	37484.66406	29960.38086	32073.98438	30720.69727	32757.4668	32260.59375	30682.18945	35174.87109	30956.93555	32183.44141	31523.38672
RT4.4_869.3284	806034.1875	1190542.375	815546.125	1860049.875	1534846.75	1137678.875	753959.5	861418.6875	1162988.625	1555434.625	584664.25
RT4.68_529.2767	201791.1875	108356.7578	137123.2031	116346.2344	117154.6875	107284.8828	131599.9375	134292.0313	124648.2891	176345.9219	149220.2188
RT4.33_700.3116	398787.0938	779883.5625	294082.4375	1059941.875	1033124	771888.0625	344201.9688	612524.0625	601955.75	1002967.563	389169.125
RT3.43_320.6559	778652.75	741941.6875	776766.6875	753018.1875	827918.4375	712979.9375	743969.6875	844395.375	692311.25	740147.8125	799328
RT4.85_370.6409	742508.8125	578037	271839.7813	499553.4688	368324.3438	575846.5	707550.875	661682.8125	274280.0313	459854.0313	499958.625
RT3.94_604.7103	689027.8125	820494.25	761179.6875	1218102.125	906209.4375	829777.3125	761131.6875	698050.3125	877654	1117936.25	497649.9688
RT3.43_601.3542	1.97E+07	1.60E+07	1.76E+07	1.70E+07	1.80E+07	1.65E+07	1.81E+07	2.08E+07	1.43E+07	1.94E+07	1.74E+07
RT4.4_443.6513	450506.9375	589572.5	470394.4688	767013.1875	695754.3125	611442.5	435472.1875	459217.9375	616607.75	694887.0625	327393.9375
RT3.06_676.249	296092.7188	323394.5625	452254.3125	378925.4063	368230.0313	391072.125	394034.1875	359028.3438	488164.625	447279.7813	339249.7813
RT3.08_567.3458	485749.8438	549535.0625	486501.5	558651.8125	476853.3125	549361.625	516415	620327.9375	491545.0625	547128.5	642594.8125
RT3.09_547.3439	1961843.625	1162214.75	1291853.75	1230432.875	1205425.375	1323448.125	1336524.75	1622389.625	984741.25	1094192.625	1360010.375
RT3.24_595.3046	2444119	1248845.375	1207702.5	939969.8125	1106829.875	1550692.375	1558504.25	1804639.75	578359.25	797790.25	1619235.5
RT3.58_659.3729	1236893.5	1145959.25	1140783.375	1220173.875	1094167.5	1054216.375	1262083.125	1303846.625	1037408.75	1275713.625	1072232.375
RT4.06_849.3654	510646.5938	618808.0625	565700.0625	923329.5625	882079.5625	783941.25	561279.125	673615.3125	829894.4375	1054323.625	596450
RT4.26_315.0858	609733.5	807228	1078341.125	693180.5625	852523.8125	692967.5625	472673.4375	408833.8438	787831.3125	965850.625	186252.3906
RT4.36_898.3376	575229.25	529328.9375	459873.2813	619939.6875	658220.8125	628024.5625	490531.875	585766.3125	404684	604989.8125	447639.1875
RT3.1_305.2174	1577956	998817.1875	980367.6875	949271.4375	1000080.875	1020342.813	1024469.375	1270272.875	794105.1875	871701.125	1069895.625
RT3.35_387.2229	1423484.125	854480.375	918647.8125	820277.5625	844492.3125	973729.25	1057052.75	1062764.5	672582.6875	726879.5625	947879.875
RT3.58_637.3907	1027267.188	899694.3125	888241.4375	986273.4375	884871.625	822367.1875	1078343.375	1072624.75	846774.4375	1021478.063	781658.3125
RT3.94_605.2107	459314.625	522197.6875	506171.0313	814377.8125	598056	553388.75	485486.625	443751.125	565475.625	796782.1875	346051.875

RT3.57_319.1651		777460.25	670934.5625	665725.5	764808.375	655353.125	619084.3125		818028.4375	802370.5	604596.75	766566.5625	601153.5
RT5.21_684.272		2349556.25	1486315.25	1426835.25	1758138.5	1936194.625	1432109		2132122.75	1899617.25	1073000.125	1497170.75	1381738.125
RT3.37_607.3419		1323507.125	1291346.375	1197782.125	1267400	1203345.75	1303827.375		1316693.25	1345229.875	1126164.25	1251583.75	1389476
RT4.34_740.3048		221688.125	477987.5	145452.7344	713113.8125	706871.125	462099.1875		172309.5938	400595.8125	321094.2188	675848.1875	212558
RT3.34_313.1464		2193789.25	1420108.125	1517263.875	1426539.25	1413765.125	1535862.875		1706342.625	1777849.25	1172067.5	1256042.375	1612747.25
RT4_700.3117		446753.1563	449214.2813	153115.4531	429689.5313	381413.8125	445786.2813		394002.5625	489180.3438	255316.0938	351134.875	296599.4688
RT7.81_903.7443		268893.7188	269236.3438	243918.5	266496.3125	260234.7188	254553.6563		257094.6563	265651.25	270019.7813	262118.9063	264361.5
RT4.4_847.3469		1211877.75	1643966.25	1182424.5	2486120.25	2005242.375	1531137.5		1120197.75	1173697.5	1572167	2016470.375	809990.5625
RT4.92_393.2875		221813.75	209334.3438	254554.6563	277338.0938	179362.375	313384.375		755902.75	741241.25	562643.6875	356628.25	283882.125
RT3.94_1170.4702		1333293	1553406.875	1395998.875	2655616	1754807.625	1529474.375		1347841.375	1350818.25	1683122.5	2428946.25	883280.875
RT5.21_744.2735		1530289.375	1162764.375	1090731.625	1247680.375	1343563.625	1091933.875		1403383.25	1386732.25	848442.3125	1144830.375	1063558
RT5.21_704.2815		1918888.75	1220633.375	1128475.625	1412635.25	1537219.25	1174134.5		1808771.375	1511236.25	834478.5625	1204713.875	1106216.375
RT2.82_292.1172		385138.6563	425454.6875	437578	420767.2188	526729.5625	264014		290401.9063	246709.75	680793.875	550580.6875	168236.9063
RT3.14_319.2333		5313535	4319043.5	5328580	5353834.5	4954326	4167914.25		4341259.5	5609061	4816301.5	5430671.5	4644390.5
RT4.4_443.1502		934076.5625	1228969.625	905061.5625	1498812.375	1320189.375	1118171.75		869576.4375	843184.125	1130150.125	1328060.5	682647.5625
RT4.07_292.1538		330827.2188	235832.9531	246079.25	473244.8125	333079.4688	292614.25		368447.7813	294034.0625	400048.1875	629811.875	239859.625
RT3.34_294.1735		5433035.5	3150403.25	3126830.25	2781915	2978093.75	3576254.75		3949300.75	4125322	2516317.5	2584880.5	3451712.5
RT3.43_320.1542		2668420	2353311.5	2417728	2386726.25	2445348.75	2294317.5		2499870.25	2569575.5	2099177	2456432.75	2503976
RT5.21_323.2007		909560.25	517805.375	560030	643265.5	708844.375	542959.3125		764587.5	737494.1875	394638.3438	533806.25	495420.5625
RT3.14_583.3411		3598357	3073360	3368521.75	3402989	3270661	3081266.5		3129863	3550059.75	3279018.5	3468787.75	3209421.75
RT4.85_724.3087		2556246.75	1991816.375	645965.75	1621723.875	1053607.375	2056504.875		2533907.75	2424595.25	707536.75	1416658.375	1583019.5
RT3.58_319.1985		837295.8125	788072.4375	737030.375	899918.75	795842.8125	661751.1875		941886.625	933310.9375	746302.6875	898318.0625	640620.9375
RT5.21_742.2744		4038985	2854295.5	2820584.25	3249474	3398512.5	2828357.25		3601014	3525801	2185398.5	2881660.5	2741427.25
RT5.21_702.2825		4779343.5	3155436	2901440	3615217.25	4039546.5	3016388.25		4327660	3988424.25	2121625.25	3071577.75	2932973.75
RT3.55_645.3575		819882.9375	497892.9063	503932.625	437334.0313	523641.4375	546868.375		693429.5625	743591.75	393506.125	440630.2188	579728.4375
RT3.43_401.2379		3519939.25	2917652.25	3240270.25	3199268	3380136	3091536		3269571.5	3806524.75	2705019	3525683.75	3204657
RT3.1_274.1752		764620.0625	528855.0625	450406.125	375041.5313	424718.0938	459549.9375		455738.2813	567238.3125	375522.5625	386304.7813	448538.9375

RT3.11_201.1231		3563850.75	2678508.75	3361108.25	3112802	2955596.25	2576003.5		2510084.75	3465401.75	2738469.5	3393542.75	2753378.25
RT3.43_301.6825		2503302.75	2294383	2288585.5	1978548	2510450.5	2227532.75		2471716.75	2576199.25	1893215.625	2440997.25	2515285
RT3.13_243.1336		4107536	3288188.5	3997007.5	4073196.25	3762167.75	3226308.5		3443252.5	4552655.5	3621287.75	4231588	3643624.75
RT4.85_666.3056		752086.625	620902.0625	231588.5469	474378.1875	352826.5625	603397.875		757799.9375	724021.25	262329.5313	423719.5	490682.3125
RT4.92_773.4948		306038.9375	269206.1875	397900.0625	465858.0938	150006.0313	488391.8125		1860266.25	1833116	1173454.25	586133	395649
RT5.21_686.2683		705868.0625	465425.125	434597.8438	558161	576201.4375	461192		693708.1875	625670.5625	334735.2188	474759.9688	461347.0625
RT4.4_829.3362		2188088.75	3059021.5	2166193.25	4609422	3799744.25	2976649.75		2090255.75	2183147	3054329	3893351.5	1619850
RT3.14_281.1834		1938520.75	2043734.25	2229320.75	2517731.5	2254844	2030407.625		2285864.5	2893208.75	2436548.5	2642897.25	2275056.25
RT3.24_287.1667		689571.3125	302415.4688	282094.4063	230441.9531	292981.125	397677.4375		415690.625	436418.4063	151094.3906	179374.4688	350514.0938
RT3.43_623.3357		9192148	7924049	8407926	8648636	8697183	8230646.5		8953006	9970402	7598006.5	9664308	8620934
RT3.34_313.648		610948.625	428741.6563	431771.7813	379257.875	417362.5	458956.7188		505290.4375	496054.625	325709.1563	349762.5	422156.0938
RT5.21_379.6233		801535.3125	629090.3125	657690.25	719069.6875	777752.4375	640416.125		796978.25	733567.1875	525645	634190.5625	645147.8125
RT3.34_587.3387		1.40E+07	8149605.5	8173978.5	7118971	7463610.5	8934163		9779530	1.02E+07	5481629.5	6507329	8693991
RT3.09_569.3259		2442515	1786957.125	1928355.875	1803841.375	1844228.625	2032397.625		1988716.125	2195143.75	1574721.375	1681369.625	2013331.125
RT3.14_281.6849		498953.8125	597008	589603.3125	686438.8125	616752.0625	562768.3125		654252.125	881565.5625	691491.625	709401.625	653262.5625
RT3.43_301.1809		8101165.5	7154398.5	6981122	6045822.5	7903315	6961333		7599648.5	8026469	5959336.5	7777786.5	8003633.5
RT0.89_120.0441		559222.625	457872.2188	536125.9375	462536.625	475969.9688	418620.75		382193.25	427422.2813	468024	473695.4688	359515.375
RT3.34_609.3202		7516471	5666565.5	5754104.5	5221308.5	5373478	5971315		6517871	6407824	4700313.5	5055202.5	5983083.5
RT3.43_283.1285		435297.0313	332996.6875	400391.5	455048.25	384904.2188	397320.8438		438610.9375	493323.6875	345762.625	431281.4063	360326.0313
RT3.44_639.3		372498.3125	349881.25	333975.6563	339031.8438	352144.5	343592.7813		376563.625	382269.625	335438.4063	359469.5313	337195.7188
RT3.42_201.1229		6244897	5073936.5	5965338.5	6241475.5	5905157	5844090		6297174.5	6950047	5015005.5	6305480.5	5572456.5
RT3.24_573.3222		1880956.125	797171.5625	646110.875	533522.5625	697836.5	948587.1875		999389.25	1119064.125	347406.5	443794.125	957840.0625
RT3.14_561.3594		5652696	4581284	5675013	5540526.5	5087953.5	4428198		4570819.5	5859219.5	5045008.5	5689362	4855880
RT3.35_187.1076		919374.375	602561.25	638829.9375	558225.5	586122.25	667371.6875		715845.4375	707159.5625	418047.375	487485.4375	660045.4375
RT4.85_684.3165		4434904.5	3437962.25	1412657.75	2801052.25	1968418.625	3495324.25		4320482	4048130.75	1514961.25	2575545.5	2820969.25
RT7.71_903.7443		254118	269236.3438	232514.4219	260553.375	259804.9375	254553.6563		256307.0625	256170.1875	270019.7813	262118.9063	264361.5

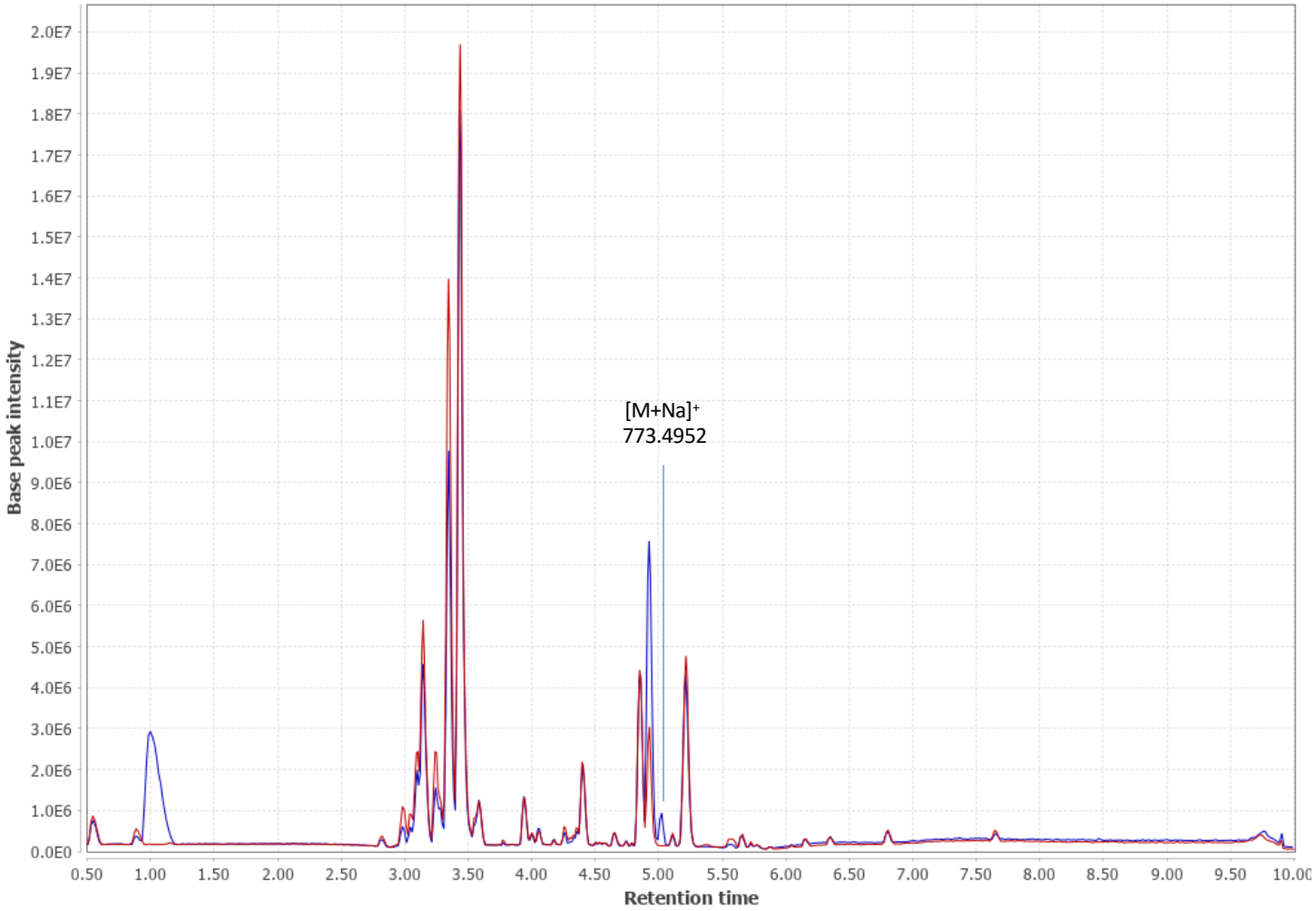

Figure S2.1.TIC overlay of *S*. sp. PBH53 with33 µM AT (blue) and without AT (red). Data were collected from *m/z* 100-2000.

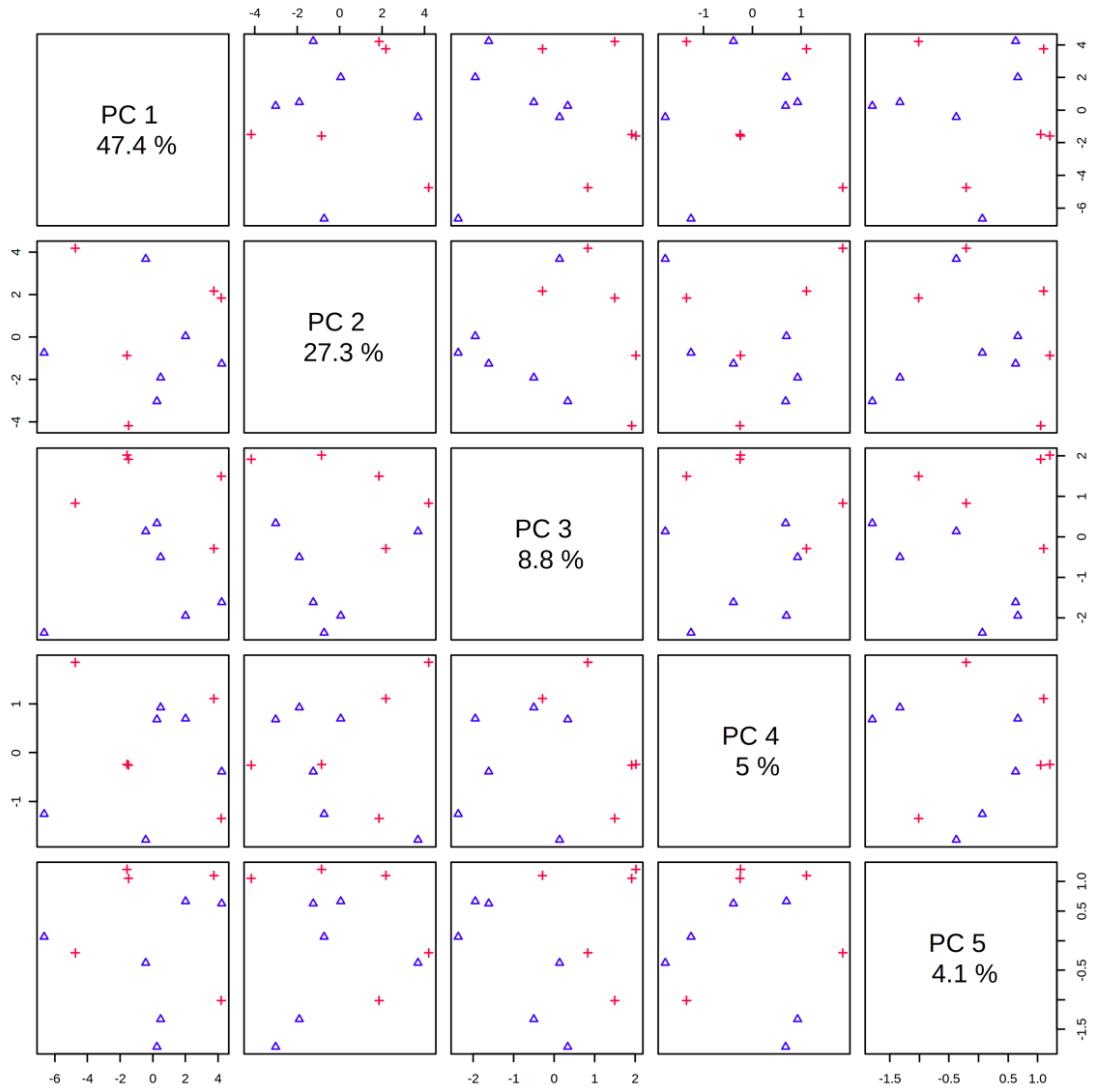

Figure S2.2. Pairwise scores plots for the first 5 principle components of *S.* sp. PBH53 cultured with 33 µM AT (Red; n=5) and without AT (Blue; n=6). The first two PCs represent 74.7% of the variance seen in the dataset.

Appendix III. *Streptomyces armeniacus* and Trimethoprim

Table S3.1. Metabolomics data for *Streptomyces armeniacus* cultured with 30 µM TMP.

	MINUS_TMP_A	MINUS_TMP_B	MINUS_TMP_C	MINUS_TMP_E	MINUS_TMP_F	PLUS_TMP_A	PLUS_TMP_B	PLUS_TMP_C	PLUS_TMP_D	PLUS_TMP_E	PLUS_TMP_F
Class	1	2	1	1	1	2	2	2	2	2	1
RT3.79_769.4086	382180.4688	170549.8438	35557.62109	18872.91406	127482.7734	1006134.438	228970.375	644383.3125	560536.0625	179954.8281	403064.7813
RT3.34_285.1678	619030.875	758342.375	504497.1875	483736	484861.75	917393.125	656494.375	862149	503168.6875	778088.625	563118.125
RT5.38_571.3223	683658.4375	481423.75	678563.8125	599857.75	422015.0313	619736.75	639808.875	706113.4375	376742.5313	839538.1875	693870.5
RT6.06_786.5979	267701.9063	1461278.125	443325.9688	106858.5469	170191.5156	2709.610352	6634.137207	13476.23438	3009.159912	4097.159668	8993.201172
RT6.67_786.5992	1121714.375	2221532.75	653738.125	122617.4922	213749.375	4363.242676	8131.009277	17477.7168	4941.030273	6239.068359	11784.87891
RT6.64_760.5839	925827.875	894948.25	198893.0938	38678.55859	70807.4375	2782.726563	4147.057129	8753.201172	2876.11499	4373.796387	5746.312012
RT6.66_760.5837	925827.875	894948.25	198893.0938	38678.55859	69108.98438	2782.726563	4147.057129	8753.201172	2876.11499	4373.796387	5746.312012
RT9.85_786.5986	751304.9375	977145.125	219139.5938	47559	89618.72656	4739.737793	8296.163086	14869.62891	4411.057617	6172.046875	10477.47754
RT6.58_758.5699	64912.13672	138797.6406	33506.03906	20115.06836	347991.125	5280.114258	7445.223145	10874.6543	5764.736816	6349.043945	9307.334961
RT6.35_786.5984	808966.75	2025105.875	616084.625	144575.6406	239538.4063	4624.027832	9159.791992	18449.0293	5963.83252	6901.084473	11338.03906
RT6.39_760.5838	824295.4375	863667.75	222590.125	43146.66797	81885.78125	3107.406982	4748.056152	9540.415039	3504.687012	4632.255371	6763.647949
RT4.15_411.2106	3356321.25	7332544.5	3457730	2374575.25	1516865.375	4497482.5	8295479.5	5396903.5	6358059.5	6035896	3259913
RT3.75_253.141	3572476.75	3062765.75	4151235.75	4234685	3614194.5	3402576.5	3127381.75	3702816.5	2694278.75	3102412.75	4077904.5
RT4.15_389.2276	2048998	4626463	1960052.25	1430736.375	887197.5	2761549.5	5001946.5	3302706.75	4083780.75	3519436	1855544
RT3.45_130.1227	7945.952148	4975.855957	5406.544922	10165.48047	6637.961914	9502.276367	6542.040527	7477.723633	7752.092773	8613.799805	6740.771484
RT3.16_377.1454	208911.7969	210525.9844	510104.375	257534.6563	272778.0938	336703.4375	516937.5938	256656.4531	473878.8438	463260.1563	415997.2188
RT3.55_764.6593	24680.59375	21457.95898	21916.35938	17433.16406	12621.26953	18122.16992	43086.39453	17563.11328	39706	34758.92969	20841.81641
RT3.55_764.3302	22790.94531	27161.46875	21768.23633	15709.32617	10163.72656	16830.10352	36503.64453	20901.74805	32406.60938	27231.98438	14511.34082
RT3.54_764.9938	13113.98828	17732.54102	20142.35156	16136.50977	6046.671387	15603.14551	29325.58398	15193.68555	41165.46484	21542.18945	13287.30957
RT6.35_758.5691	62404.82422	119268.4219	31130.08398	21078.42773	327015.625	4549.04248	7263.12207	10782.2627	4655.775391	5868.726074	8995.704102
RT3.01_420.2099	432925.8125	536701.625	376983.5938	446521.4063	438579.25	609656.8125	613803.5625	544069.6875	407597.9375	585363.9375	517783.3125
RT6.47_758.5701	62689.42578	138797.6406	30825.10352	19548.21484	347991.125	5280.114258	7445.223145	10874.6543	5323.894531	5868.726074	9307.334961
RT0.77_356.2179	192194.2031	201149.2031	151562.7344	167784.8125	179184.8438	175195.1875	248937.4063	191008.875	179481.8594	223015.6563	214325.8906
RT1.93_116.1069	911.0482788	0	776.600769	1377.681519	1380.800293	704.0160522	1187.055786	0	546.0149536	862.3248291	0
RT6.5_782.5699	134317.6406	9084490	78202.58594	56541.62109	191866.125	15887.17676	23608.38281	33499.58984	18977.4707	20042.88281	30328.91016
RT6.2_716.5247	36914.26172	1654219	26359.39258	22381.92383	60233.83594	10600.06934	13793.36133	19169.98242	11644.36133	11979.21973	12622.4541
RT6.67_758.5698	64912.13672	129457.3203	33506.03906	20115.06836	315437.9375	4741.12793	7140.615234	10511.85352	5764.736816	6349.043945	8337.977539
RT9.84_784.5844	49465.55859	5187878.5	34488.20313	21306.4375	175626.4688	5803.758301	11188.03027	13871.18457	6781.767578	8319.203125	11898.32715
RT6.37_784.5843	84616.0625	1185407.5	40292.66797	27623.60742	483730.6875	5908.343262	9501.647461	13477.44141	6831.301758	6469.80957	11673.30859

RT6.17_758.5695		54580.53516	85874.625	24432.54297	16044.2334	274677.9063		4608.163086	5565.213867	7822.041992	4553.694824	5255.30957	6524.253906
RT5.84_716.5247		30183.94531	1534454.125	23846.99023	18242.72266	50215.58594		7650.759277	10568.64453	13591.22266	9289.886719	9268.827148	10934.47656
RT9.85_758.57		40412.82422	4060265	23397.14844	18196.46094	121679.0781		5637.577148	7055.001953	10225.40723	4914.01709	5910.98877	8426.453125
RT2.78_360.1895		792295.6875	786392.9375	586319.0625	614383.5	745003.5		822111.6875	811221	690267.1875	580767.8125	813295.8125	865993.3125
RT6.63_786.5988		1121714.375	2221532.75	653738.125	122617.4922	213749.375		4363.242676	8131.009277	18671.3125	5090.912598	6239.068359	11784.87891
RT3.49_253.1407		1073258	814782.4375	681918.5625	709851.8125	699344.625		1630585.125	794219.3125	765517.3125	694199.8125	841639.5	657643.125
RT6.16_804.5485		5392.717285	1036580.313	2687.405273	2037.520996	10667.81055		0	1358.672729	1373.354858	876.5323486	918.4016113	1060.631226
RT6.56_784.5845		77466.96875	1095076.75	43743.15625	27028.07227	497820		6471.958496	8802.78125	13920.15137	6545.797363	8295.925781	10677.36035
RT6.19_742.5373		17135.52148	1023806.813	12144.80176	10286.66309	27948.08984		3980.247559	5446.330078	7518.332031	4872.496582	5349.721191	5951.817383
RT6.44_786.5989		964137.875	2210575.25	616084.625	144575.6406	246136.6875		4624.027832	9159.791992	18671.3125	5963.83252	6901.084473	12719.36035
RT4.65_541.2914		1120846.25	953058.375	2731794.5	2114840.25	905717.25		2171697.25	526103.875	904061.1875	415281.1563	451457.4375	1039830.188
RT6.01_742.5374		16278.05859	1021624.063	11764.85156	10446.88867	25615.38867		3639.89502	4903.446777	6701.50293	4306.771973	4302.040527	5739.839355
RT6.18_786.5977		400635.5313	1687297.25	438753.375	108354.75	191038.375		3055.714844	5634.148438	14018.83496	4335.703125	4717.619141	7813.855957
RT6.06_979.5536		1278616	34389.57422	2473835.25	1560155.75	852162.5		1576732.625	421207.75	738894.625	300476.9375	420346.8125	1811664.25
RT6.06_501.2723		1510474.75	406038.9375	1541607.5	1566190.875	1353591.75		1670839.75	1259537.625	1461274.25	1155535.125	1281345.875	1608103.625
RT4.51_511.2821		177601.6875	629165.5	304518.625	464961.7188	262599.0313		1412886.75	290664.25	445215.0625	306817.9063	225090.7813	227616.2188
RT3.18_490.2653		220179.0313	1049217.125	592807.75	1432536.375	448280.8438		965138.75	942402.3125	1460153.75	444460.25	1098458.375	549876.0625
RT6.06_957.5729		1062240.625	29639.08594	2179917	1234091.625	712130.6875		1290979.125	403627.8125	702409.1875	292700.3125	405697.4688	1441927
RT4.07_179.1431		532270.8125	1355298.125	448517.75	447039.2813	306753.125		1350015	1230289	1295329.125	969387.25	1255375.875	427180.6563
RT3.79_768.7406		429221	206383.5625	38089.08984	13656.13086	150512.8906		1212876.375	270014.6563	787678.125	635987.8125	204089.9688	479413.7188
RT4.73_517.2664		350546.0938	552687.625	537959.4375	286191.1563	446342.75		1015067.313	564887.6875	357163.125	667881.3125	462262.2188	402432
RT4.73_495.2854		1131005.25	1808350	2067731.875	895035.4375	1488921		3470479.75	1857633.75	1242878.75	2209868.75	1545919.25	1541202.5
RT6.08_784.5844		73976.11719	1081196.875	37680.30469	19586.25781	376584.75		4828.640137	5159.65625	10447.98633	4319.121094	4469.104492	8763.668945
RT0.73_378.1997		1199010.25	1262320.375	1012201.938	1084982.25	1119579.125		1223185	1390273.375	1216036.875	1154715.25	1332240.125	1271294.625
RT4.75_368.2091		12494.32324	9032.116211	13936.17383	10692.63672	6509.981445		6862.452637	11550.25977	12226.01367	15925.39063	8208.12207	16034.94531
RT3.79_769.072		568786.3125	285019.0938	56571.26953	19426.33984	196347.7031		1493603	361242.4688	994872	874657.75	284510.8125	626121
RT6.38_742.5377		16696.50391	983778.5625	14820.47559	11198.36816	26694.83398		4043.805908	5634.09668	7959.538574	4583.72998	4959.953613	5954.844727
RT4.76_346.224		8492.453125	8322.535156	9517.435547	10661.81348	0		8045.726563	6926.467773	14645.20117	0	7549.551758	9288.740234
RT4.92_393.2874		214040.8594	249631.4219	131802.7656	174758.8125	130444.3281		245807.7031	206296.4063	169837.25	193310.9219	172228.6406	247501.9063
RT5.84_742.5372		15759.08496	932852.125	10175.91602	7618.980469	23923.30469		2671.012695	5339.775879	5695.220703	3455.1521	3704.94043	5126.845703
RT4.15_144.1017		612580.625	1447098.5	670996.0625	428405.875	292359.875		816880.4375	1628800.125	1030539.938	1246463.75	1231362.25	610421.5

RT4.92_751.5128		27554.06055	58139.04688	13448.54492	31164.14453	13831.19336		51670.125	36181.59375	27302.90234	29640.9375	26640.26367	41935.44141
RT3.7_516.2812		657577.625	1188246.875	1131785.25	1415168.125	999700.3125		1042636.313	1728025.875	1731675.75	1309393.375	1585993.125	1194549.75
RT7.64_338.3412		237197.7344	364285.75	264820.875	645532.375	586547.8125		745706.5	587160.875	265478.3125	269963.0313	301061.6875	959968.125
RT8.18_1128.6673		4358577.5	1275267.625	5801155	2721937	1751667.875		2745322.75	1596974.125	1701106.75	1481114.875	1652664.375	2742068.5
RT3.16_510.2892		2443400.5	1880781.125	2085361.125	1584175.625	2117388		2008302.625	2222918	1669344.25	1844682.25	2216658.25	2368061
RT0.53_209.6078		982779.8125	1415307.75	1316717.75	1462133.5	994878.3125		1386694.5	1320165.375	1625838.25	1100082.375	1574114	1149650.625
RT0.53_418.2085		838658.9375	1412526.125	1168637.125	1374331	947875.8125		1525276.375	1167177.125	1616969.5	1079672.5	1455294.875	1037012.313
RT3.74_195.1379		1443359.125	1325312.125	1745806.5	1924363.625	1538250		1472293.375	1316039.625	1573935.875	1232520.625	1304104.5	1661968
RT3.22_253.1409		1593745.5	1348920.75	1663237	1652508	1495893.25		1755584.375	1659041.375	1540373.25	1501609.25	1677901.75	1817410.75
RT4.58_583.2558		1805425.625	1540669.625	666888.0625	305214.0313	798139.9375		1215054	1015677.5	1498272.875	709432.125	1780408.5	1109673.625
RT8.18_1133.6203		3830036.25	1110734	5405998	2460570.5	1446325.25		2503527.25	1400985.375	1463724.875	1288872	1307677.625	2524330.5
RT6.16_758.5689		54580.53516	85874.625	24432.54297	16044.2334	274677.9063		4608.163086	5565.213867	7822.041992	4553.694824	5255.30957	6524.253906
RT4.57_314.1994		9323.083984	8891.912109	11393.20508	9963.493164	7005.870605		11181.92969	8072.982422	7985.865234	10788.93945	11202.18359	9948.50293
RT3.54_577.3137		86385.24219	636775.5	943459.875	1338215.25	364245.25		337445.5938	818443.125	939134	389000.875	536480.5	516289.4063
RT5.36_495.283		705665.625	151938	1845480.25	600385.0625	585767.5625		646268.4375	232395.7813	247249.3906	228214.5156	202224.1094	755583.625
RT5.12_511.2801		778041.375	528771.875	1692859.625	537237.375	603371.625		814208	403655.0938	349754.5938	375072	331728.7813	873763.125
RT4.65_563.2743		350677.5	358662.375	995135.875	762571.8125	368758.9688		778125.625	203744.8125	317674.2188	176614.7031	162132.8125	405106.3438
RT6.47_617.5128		24699.4375	811223.9375	1587.795654	1307.277832	2885.666748		1138.874756	1647.196533	1784.086426	1062.715698	1673.059937	1198.281982
RT4.59_332.204		13186.62598	9877.974609	8725.335938	110883.6563	16338.57422		22347.875	9654.707031	7038.628418	8438.615234	17574.45508	7968.987305
RT6.32_760.5837		725813.5	784298.375	222590.125	43146.66797	78355.19531		2922.447754	4748.056152	8801.234375	3504.687012	4632.255371	6763.647949
RT4.58_354.191		4998.023438	4975.748535	6024.324219	151501.625	20169.91602		10292.24707	6979.729492	8609.046875	5500.149902	12213.82715	9233.811523
RT8.65_453.4776		2512.920654	26080.92773	2018.746094	1271.656494	4172.713867		672.1212158	1078.233032	1054.322754	0	792.5707397	999.5407715
RT3.52_519.8006		52997.60938	1015797.188	852134.1875	1540560.25	677308.9375		33563.89844	603238.5	325129.1875	136668.5469	1003943.438	161714.2969
RT4.56_685.389		26769.73633	20083.63477	25664.66992	24490.17773	27151.17383		29677.80273	26074.05273	26681.24219	25789.29102	25193.25977	26109.60938
RT6.09_479.2904		5451448	1125110.5	7203654	5550438.5	4395881.5		5902585	3613148.5	4533431	3160714	3660124.5	5913811.5
RT4.55_663.4067		47255.39063	43301.46875	50202.24609	41321.59766	38785.50781		37844.62891	47734.89453	57447.99219	41651.17188	51105.35156	45442.13281
RT9.84_782.57		33498.19922	186088.25	23359.39063	14857.6582	57739.50391		4948.780762	7793.810547	8906.645508	4818.740723	5949.907227	9036.427734
RT6.06_324.041		31919.54883	18596.33398	26728.57617	3790.112305	7465.279297		0	83016.89844	61556.77734	35484.44141	17756.51172	24787.86719
RT6.06_322.0443		29483.36523	24304.21094	25224.05469	4709.486816	8407.003906		0	86087.24219	60903.64844	38277.25	19824.73242	30306.58789
RT6.73_740.5226		33103.71875	171533.6094	30798.11523	27130.34961	45424.02344		9264.657227	13136.9043	15374.99023	10763.72266	11785.41699	14291.04199
RT6.66_740.5226		25725.97852	173921.6094	23108.23047	20530.51172	42470.13672		8212.260742	12371.96777	14661.22852	10423.48926	10868.05371	12244.73535
RT6.63_784.5846		77166.92969	926911.3125	43743.15625	27028.07227	454602		6471.958496	8802.78125	13540.47266	6545.797363	8295.925781	10563.43359
RT6.42_784.5843		84616.0625	1185407.5	40292.66797	27623.60742	497820		5908.343262	9501.647461	13920.15137	6831.301758	7363.412598	11673.30859

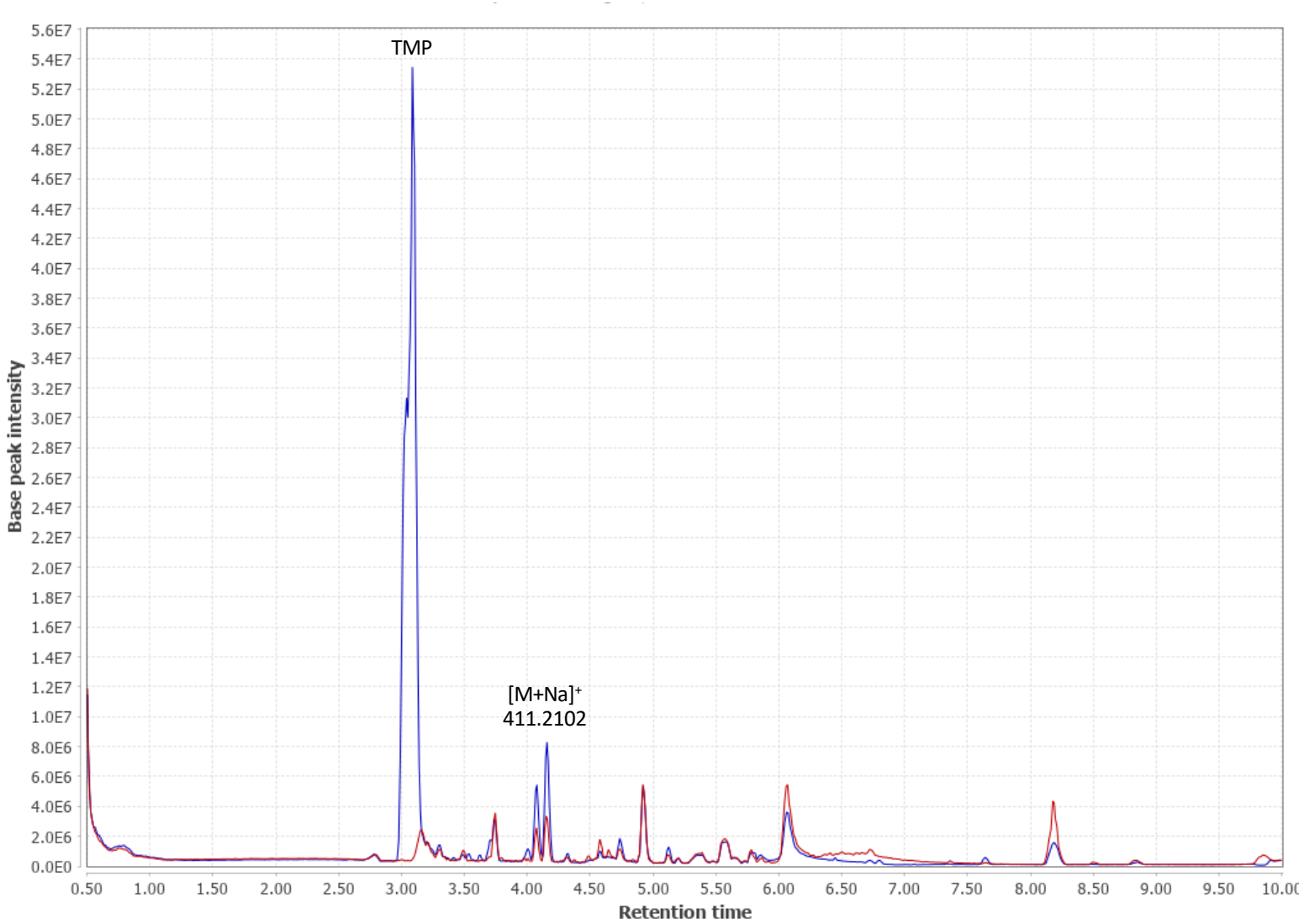

Figure S2.1. TIC overlay of *S. armeniacus* with30 µM TMP (blue) and without TMP (red). Data were collected from *m/z* 100-2000.

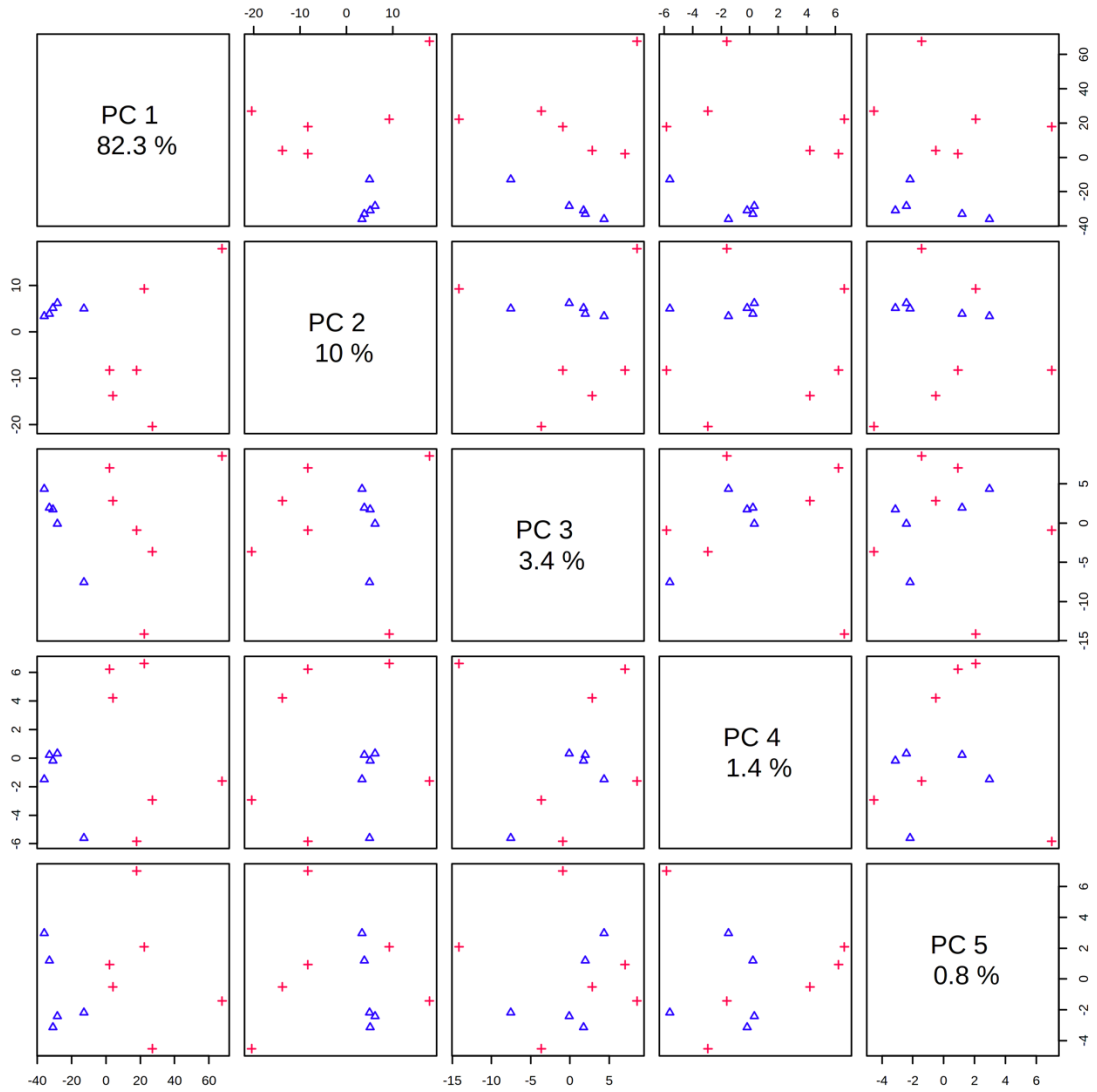

Figure S3.2. Pairwise scores plots for the first 5 principle components for S. armeniacus cultured with 30 µM TMP (Red; n=6) and without TMP (Blue; n=5). The first two PCs represent 92.3% of the variance seen in the dataset.

Table S3.2. Retention times and corrected masses of features upregulated at least two-fold in *S. armeniacus* cultured with 30 µM TMP (n=6) compared to no TMP (n=5) ($p<0.05$).

RT	*m/z*
0.53	209.6078
0.53	418.2085
0.73	378.1997
0.77	356.2179
2.78	360.1895
3.01	420.2099
3.16	377.1454
3.16	510.2892
3.18	490.2653
3.22	253.1409
3.34	285.1678
3.45	130.1227
3.49	253.1407
3.54	764.9938
3.54	577.3137
3.55	764.3302
3.55	764.9938
3.7	516.2812
3.74	195.1379
3.75	253.141
3.79	769.072
3.79	769.4086
3.79	768.7406
4.07	179.1431
4.15	411.2016

4.15	389.2276
4.15	144.1017
4.55	663.4067
4.56	685.389
4.57	314.1994
4.58	583.2558
4.73	495.2854
4.73	517.2664
4.75	368.2091
4.76	346.224
4.92	393.2874
4.92	751.5128
5.12	511.2801
5.38	571.3223
6.06	501.2723
6.06	322.0443
6.06	324.041
6.09	479.2904
6.66	740.5226
6.73	740.5226
7.64	338.3412
8.18	1128.6673
8.18	1133.6203

Appendix IV. *Streptomyces armeniacus* and Atenolol

Table S4.1. Metabolomics data for *S. armeniacus* cultured with 30 µM AT.

	MINUS_AT_A	MINUS_AT_B	MINUS_AT_C	MINUS_AT_D	MINUS_AT_E	MINUS_AT_F	PLUS_AT_A	PLUS_AT_B	PLUS_AT_C	PLUS_AT_D	PLUS_AT_E	PLUS_AT_F
Class	1	1	1	1	1	1	2	2	2	2	2	2
RT3.8_769.0749	0	330479.9063	175148.625	205985.8438	277489.6563	300890.3125	1175511.25	2419729.25	1178168.5	1167183.625	652768.8125	9199.55957
RT3.8_768.742	0	257432.2969	141659.9688	147538.6563	211664.3438	220939.7188	920209.3125	1850432	853707.75	867860.75	477516.875	5081.719727
RT3.8_769.4058	0	190161.2188	118250.9688	119311.9375	172253.2031	185549.4375	775054.375	1614560.75	728531.9375	738750.5	445043.125	5651.533203
RT5.17_525.2972	3027210.75	2565532.5	3030915	1612204.375	1927278.375	2335205.25	2910312.25	3332029.5	2331088	1369397.875	2587101.75	2349061.25
RT6.06_479.2919	1.30E+07	1.01E+07	1.34E+07	1.38E+07	1.44E+07	1.73E+07	1.29E+07	1.36E+07	1.33E+07	1.02E+07	1.34E+07	1.44E+07
RT4.49_511.2828	863955.9375	1337372.25	1760427.5	194493.9063	306282.5625	313180.9063	711199.75	831761.0625	392312.8438	322673.0313	1311765.625	727429.5625
RT5.84_479.2929	1616264.625	110770.8281	2240324	406350.0938	2188019	2193846	2124165	1884539.875	2398357	1701698.125	1781339.75	1926672.875
RT3.54_577.3148	911723.6875	143210.9844	274291.375	899783.625	708372.125	1502495.25	312338.875	362029.4375	524109.3125	303961.5	45203.49219	1779315.75
RT5.36_477.2767	3347118	2112730.75	2281209	1883596.625	1940283.375	2975203	1554840.5	1907687.5	2010897.5	1207704.875	1608359.75	2006224.875
RT4.72_495.2868	7407998	1.28E+07	1.15E+07	2567285.25	3332587	3590869	3710533	2437143	3081273.5	2791664.5	1.05E+07	6112818.5
RT4.63_541.2947	6041329.5	1.14E+07	9696648	5119320	3883769.25	8189872.5	4294014.5	7259841.5	5701455	2825754.5	3665076	5485561
RT5.11_511.2819	7325195.5	6924551	8415617	3900204.25	3768753.5	4849313	2572764.25	3287558	4196594	2463526.75	4228065.5	5536283.5
RT8.74_1133.6215	1719092	279469.8125	263182.125	1097611.125	870955	1399034.25	2255422	1960796.125	1353197.625	1103129.375	479445.6563	585144.1875
RT5.91_477.277	1622395	706379.4375	1341676.25	2589782.25	3096879	3612339.5	3289207.5	4151329.25	3576047.5	1870056.75	1450679.875	2871259.75
RT6.18_786.6017	1546483.375	326270.5625	7870.540039	16405.39844	944296.625	94480.41406	272721.1563	11765.53125	939193.375	6442.369141	4635.001465	4427.067383
RT5.8_477.2772	1363306.25	944568.5	1068821.25	980531.6875	1008640.563	1203736.375	831669.375	817560.375	920575.25	685152.375	1193566.125	1100408.375
RT6.06_957.5765	7463367.5	4902395	7668661.5	8145056	8617578	1.11E+07	7587075	8421159	8163933.5	5372603.5	8021022	9278584
RT5.36_495.2862	7781996	5142855.5	5461750.5	4533123.5	4650287.5	7022110.5	3687533.5	4591995.5	4700730.5	2882547.25	3788435.75	4787537
RT5.11_493.2706	2572948.5	2190608.25	2602325.5	1207852.875	1172590.75	1609673.25	818292	1013429.625	1302481.375	796418.1875	1316077.875	1595996
RT4.64_509.2665	1543619.375	2600841	2248895.5	1133242.5	911342.9375	1722116	1032600.625	1409300.875	1222066.375	831309.3125	1287025.375	1354470.25
RT4.64_491.2564	1352977.875	2647690	2171096	1054096	818158.375	1685326	905515.5	1511470.125	1216525.5	590841.9375	755691	1269053.875
RT6.07_979.559	1334275.625	1234269.5	1716946.875	1723971	1460058	1692758.625	1594736.625	1455992	1465200.625	1557752.5	1574793.125	1589708.125
RT8.75_1128.672	3443207.5	628719.6875	646696.125	2356584.75	2026903.875	3040088.25	4536920	4091487.75	2934209.5	2563996.5	1220860.5	1454582.875

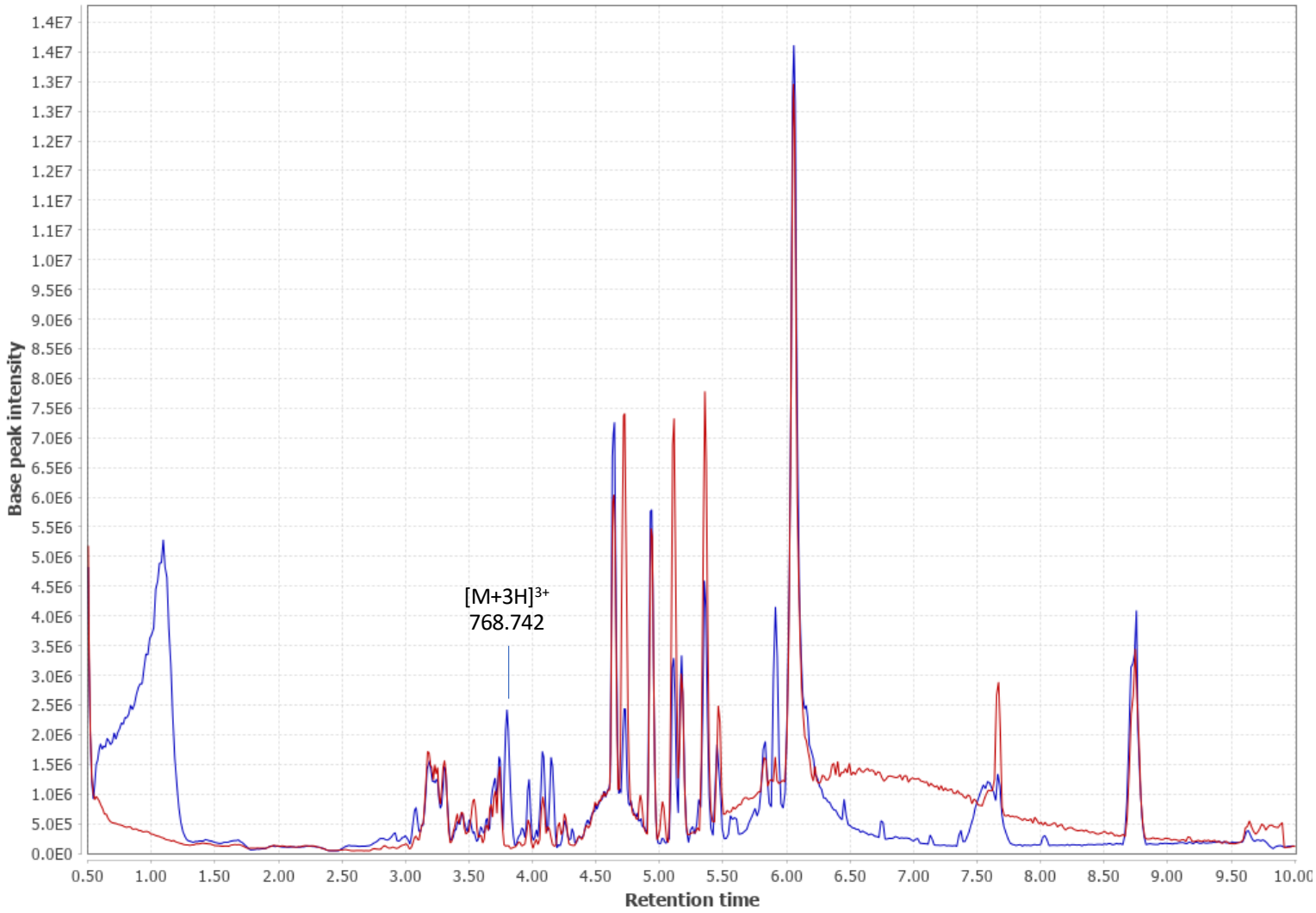

Figure S3.1. TIC overlay of *S. armeniacus* with 30 µM AT (blue) and without AT (red). Data were collected from *m/z* 100-2000.

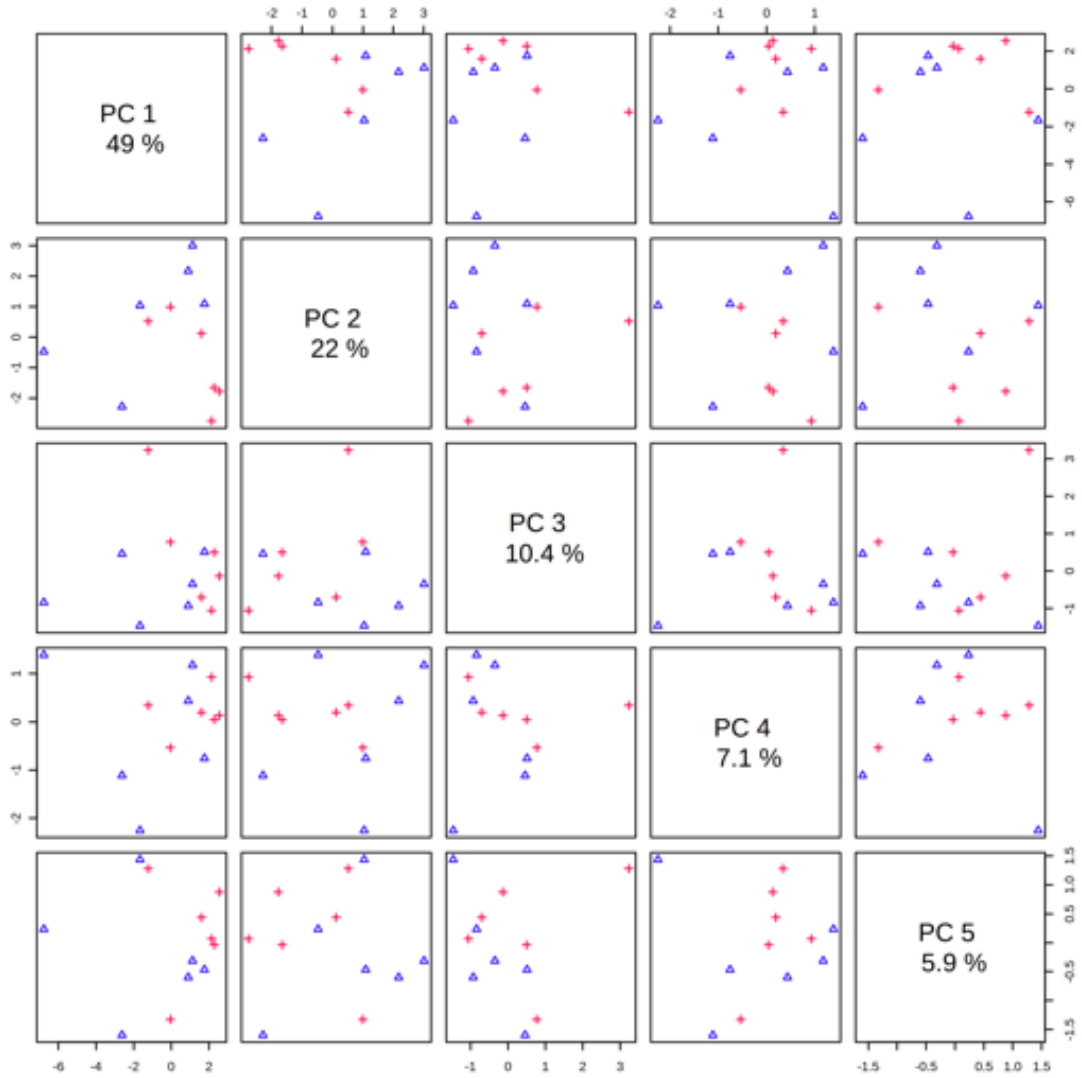

Figure S4.2. Pairwise scores plots for the first 5 principle components for *S. armeniacus* with 30 µM AT (Red; n=6) and without AT (Blue; n=6). The first two PCs represent 71% of the variance seen in the dataset.

Appendix V. Genome Sequencing and Assembly

ART commands used to simulate Illumina reads of biosynthetic gene clusters.

```
art_illumina -ss HS25 -sam -I <filename> -p -l 150 -f 30 -m500 -s20 -o
<outputfilename>
```

Command to assemble simulated raw reads in SPAdes

```
Spades.py -1 <readfile1> -2 <readfile2> -o <outputfoldername>
```

Command to assembled simulated raw reads in Biosynthetic SPAdes

```
Spades.py --bio -1 <readfile1> -2 <readfile2> -o <outputfoldername>
```

Table S5.1. BLAST results of the AT domains from the Naphthomycin A literature sequence using the first AT domain in the cluster (NatA(1) as the query sequence. The monomer selected by the AT domain is given after the gene name and location of the AT domain.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident
Nat A(1) methyl-mal	1615	1615	100%	0.0	100.00%
NatB(2) methyl-mal	1502	1502	100%	0.0	97.21%
NatC(2) methyl-mal	1380	1380	100%	0.0	94.19%
Nat A(3) methyl-mal	1093	1093	95%	0.0	88.30%
NatD(1) methyl-mal	1025	1025	86%	0.0	89.29%
NatE(2) methyl-mal	669	669	99%	0.0	76.73%
Nat B(1) methyl-mal	565	565	99%	5e-164	74.16%
NatE(1) mal	96.0	96.0	39%	2e-22	66.58%
NatD(2) mal	68.9	68.9	39%	3e-14	65.22%
NatC(3) mal	60.8	60.8	39%	5e-12	64.78%
NatC(1) mal	53.6	53.6	8%	7e-10	78.08%

Table S5.2. BLAST results of the AT domains from the Naphthomycin A literature sequence using the first AT domain in the cluster (NatA(2) as the query sequence. The monomer selected by the AT domain is given after the gene name and location of the AT domain.

Description	Max Score	Total Score	Query Cover	E value	Per. Ident
Nat A(2) mal	1626	1626	100%	0.0	100.00%
NatC(3) mal	227	227	49%	6e-62	71.56%
NatE(1) mal	207	207	49%	5e-56	70.72%
NatB(3) mal	191	191	69%	4e-51	66.83%
NatC(1) mal	189	189	49%	1e-50	69.33%
NatD(2) mal	188	188	49%	1e-50	69.21%

Table S5.3. Gene clusters used for raw read simulations.

Cluster Name	Organism	GenBank Accession Number
Monensin	*Streptomyces cinnamonensis*	AF440781.1
Reveromycin A	*Streptomyces* sp. SN-593	AB568601.1
Halstoctacosanolide A	*Streptomyces halstedii*	AB241068.1
Nanchangmycin	*Streptomyces nanchangensis*	AF521085.1
Sceliphrolactam	*Streptomyces* sp. SD85	KX230849.1
Sanglifehrin A	*Streptomyces flaveolus*	FJ809786.1

Oxalomycin B	*Streptomyces albus*	EF552687.1
Bafilomycin B1	*Streptomyces lohii*	GU390405.1
Caniferolide A	*Streptomyces caniferus*	MK303577.1
Candicidin	*Streptomyces* sp. FR-008	AY310323.2
Laidlomycin	*Streptomyces* sp. CS684	JQ793783.1
Nigericin	*Streptomyces violaceusniger*	DQ354110.1
Oligomycin	*Streptomyces avermitilis*	AB070940.1
Meilingmycin	*Streptomyces nanchangensis*	FJ952082.1
Tetronomycin	*Streptomyces* sp. NRRL 11266	AB193609.1
Bacillaene	*Bacillus velezensis* FZB42	AJ634060.2.

Table S5.4. Genomes selected from the NCBI database that were sequenced by Illumina and had evidence of assembly errors in PKS pathways

Organism	GenBank Assembly Accession
S. sparsogenes	GCA_001704635.1
S. diasticus subsp. *ardesiacus*	GCA_002019855.1
S. rubrogriseus	GCA_003112595.1
S. violens	GCA_000717745.1
S. griseofuscus	GCA_000718315.1
S. halstedii	GCA_000720535.1
S. yokosukanensis	GCA_001514035.1

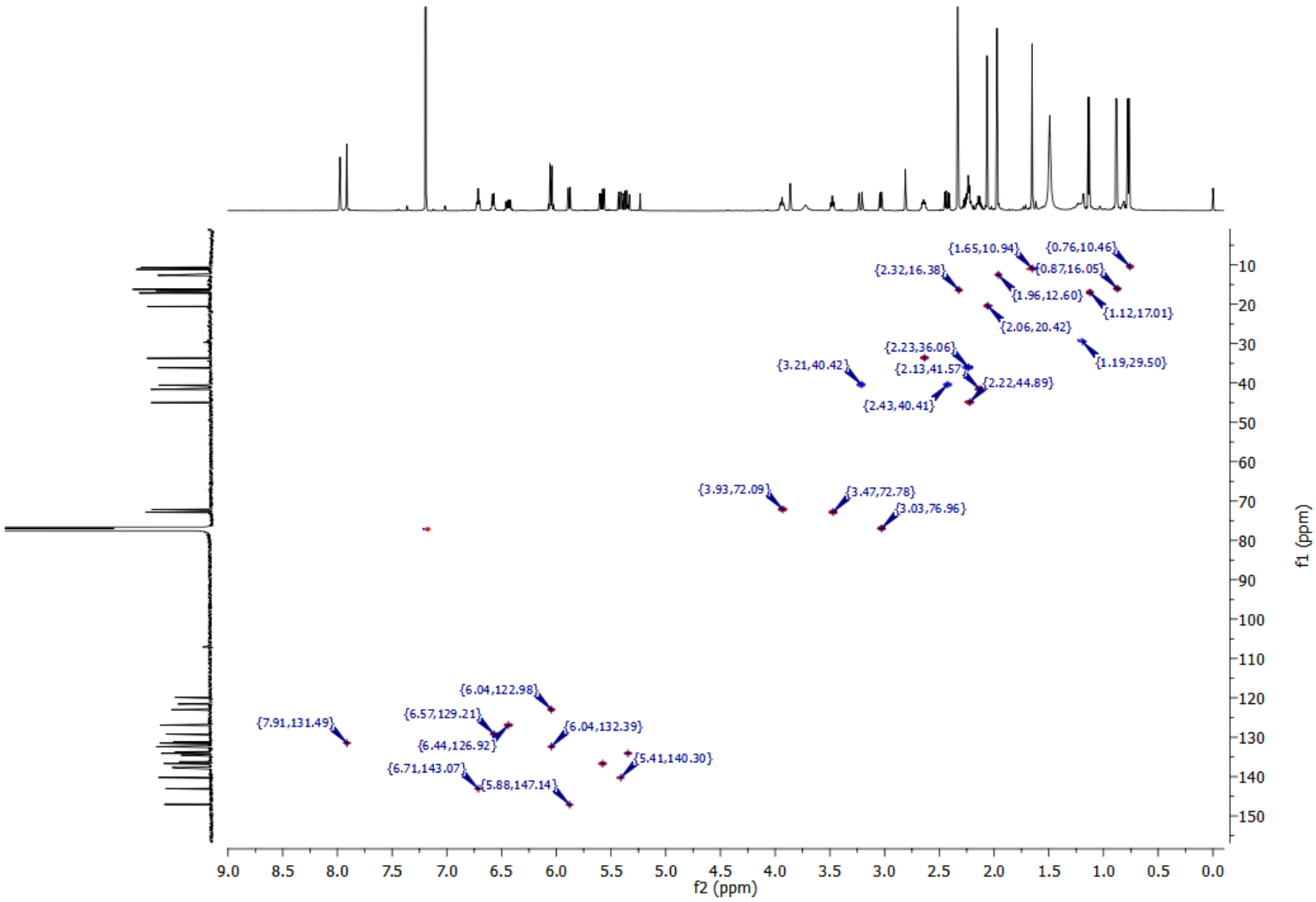

Figure S5.1. HSQC spectrum of Naphthomycin A.

Table S5.5. Cross peaks of the HSQC spectrum used for identification of naphthomycin A.

^{1}H	^{13}C
0.76	10.46
0.87	16.05
1.12	17.01
1.19	29.5
1.65	10.94
1.96	12.6
2.06	20.42
2.13	41.57
2.22	44.89
2.23	36.06
2.32	16.38
2.43	40.41
2.63	33.61
3.03	76.96
3.21	40.42
3.47	72.78
3.93	72.09
5.35	134.1
5.41	140.3
5.58	136.71
5.88	147.14
6.04	122.98
6.04	132.39
6.44	126.92
6.57	129.21
6.71	143.07
7.91	131.49