

BIOPROSPECTION OF METAGENOME TO ACCESS NOVEL BIOLOGICAL PATHWAYS

JESSICA THANDARA GOSSE

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
in partial fulfillment of the requirements
for the Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Ottawa-Carleton Chemistry Institute
Faculty of Science
University of Ottawa

© JESSICA THANDARA GOSSE, OTTAWA, CANADA, 2019

Abstract

This research is focused on finding new secondary metabolites from bacterial sources through genome mining, allowing the study of unknown molecules that could be potential candidates for use as drugs against microbial infections, cancer, and immune diseases. Bioprospection was conducted through targeted environmental sampling, unusual environments exploration and molecular and computational biology.

In the first project, an environmental DNA (eDNA) fosmid library from urban soil samples was built and arrayed for the presence of potential gene clusters. In total, 7% of the library was positive for the presence of secondary metabolite pathways. Sequencing revealed genetic similarity with other organisms that belong to urban areas, showing adaptation of the microbiota to anthropogenic environments.

The second project focuses on the heterologous expression of a lantipeptide from a bacterial isolate from an urban environment. Using a Multigene Expression System (MES), different components of the biosynthetic pathway were cloned and expressed. We were able to successfully express the genes responsible for lantipeptide tailoring but the lantipeptide portion of the pathway was not obtained with the methods used.

The last project examined cave bacterial secondary metabolite production and organism diversity to determine the relationship between the microbiome and to retrieve potential active natural compounds. Through sequence analysis and LC-ESI-HRMS, several metabolites were detected and their potential role for the cave microbiome could be established.

Acknowledgments

The first acknowledgment of this work is not for a person, but for a molecule: caffeine. Without you, in the form of coffee, green tea, black tea, supplements, or a combination of all, the motivation for having the work done would have never been the same. Thank you.

Now for the people: my family. Mom, you are the one who always believed in me and sacrificed so much to help me be who I am. To my grandmother: it is hard to accept that you are gone, but wherever you are I know your light is still guiding me. Your departure gave my work a personal thing to fight for. To my stepmom, aunts, cousins, my in-laws: life is so much fun with you! Blood has some meaning but choosing the ones who will always be by your side no matter what is what defines Family. Thank you.

Andrey, you and I have gone through so much in distant lands, all the pain, tears, doubts...but all the experiences, the growth and things that we learned together made everything worth. Thank you for your love and for being by my side for all these years.

Brazilian friends! You are incredible! All the stories we share, the friendship, the sense of community you help me build in Canada will never be forgotten, no matter where. The Ph.D. and life experiences as an expatriate had a taste of caipirinha and brigadeiro with you here. Thank you.

Now to Boddy lab...Chris, thank you so much for believing in me. In a world where people sometimes don't want to risk, you always do by giving people a chance to do science. Your uplifting, curious and motivating behaviour is inspiring to someone who wants to be a scientist. Thank you for your

supervision. Lab friends from now and from the past: you are awesome! Thank you for being there to teach, listen, learn and share a lab and a part of your life with me.

Special thanks to all the professors, coordinators, students, Department staff, lab staff and lab mates from other groups that I had the pleasure to teach and learn from, in the form of courses, TAing, seminars, paperwork, and informal conversations. Shout out to Ashwag, who was incredibly generous and supportive.

As Science doesn't happen without money, I would like to thank all the financial support from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)- Program Science Without Borders (Brazil), the Boddy Lab, as well as all the scholarships received throughout my degree via University of Ottawa.

Lastly, thank you to all the people involved in my extracurricular activities, from music and dancing to counselling and volunteering, to help me with my work-life balance. You sure help me be less crazy of a person (at least a bit).

And most importantly: DON'T PANIC!

“I’d take the awe of understanding over the awe of ignorance any day.”

— Douglas Adams, *The Salmon of Doubt*

Table of Contents

Abstract.....	ii
Acknowledgments	iii
Table of Contents	vi
List of acronyms and abbreviations	xi
List of figures.....	xiv
List of tables.....	xix
General introduction	1
Secondary metabolites: what, where and why	1
Secondary metabolites as allies for human life (or how the once seen as bad guys became the good guys against the real bad guys)	2
Compound screening approaches and the discovery pitfall.....	3
Make molecules great again: unleash the untapped potential of biomolecules and decrease rediscovery rates.....	6
References.....	8
Introduction	14
Bioprospecting.....	14
Natural products targeted when bioprospecting	14
Actinobacteria, the powerhouse of bioprospection	19
Straight to the point: sampling optimization	20
How to be a Bookworm: library construction and array	22
Project objectives	24
Materials and Methods	25

Sample collection.....	25
eDNA extraction	25
1. Preparation of samples for eDNA extraction	27
2. Extraction at high pH	27
3. Extraction at low pH	28
4. Protein and polysaccharides removal, first DNA concentration	28
5. Protein and humic material removal.....	28
6. eDNA concentration.....	29
Pulsed field electrophoresis	29
PCR evaluation of the presence of natural products pathways.....	31
Library construction	33
7. Preparation of library insert DNA.....	33
8. End-repair of fragments.....	34
9. Size selection of input	34
10. Recovery of DNA fragment.....	35
11. DNA concentration determination	35
12. DNA ligation.....	37
13. Packaging of vectors into phages and library preparation	37
Library array	38
14. Secondary metabolite pathway evaluation.....	38
15. Isolation and identification of inserts.....	39
Permanent library stock preparation.....	40
Results and Discussion	42
Sample collection and eDNA extraction	42
Library construction	44
Optimization of library array system	44
Library evaluation.....	47

Identification of pathways from positive clones.....	50
Conclusion.....	53
Chapter Bibliography	54
Introduction	64
RiPPs, not your average protein	64
You would never break the (thioether) chain – Lantipeptides	65
Strategies for accessing lantipeptides and other RiPPs.....	67
Project objectives	69
Materials and Methods	70
Household isolates library.....	70
16S identification of isolates	76
Whole genome sequencing (WGS) of isolates and secondary metabolite evaluation..	77
Lantipeptide primer design and amplification.....	77
Cloning of lantipeptide pathway.....	79
2. Expression vector cloning	79
3. Construction of the coexpression vector pLanBC.....	80
4. Transformation of vectors in <i>Escherichia coli</i> strains	81
5. Protein expression and evaluation.....	81
6. Cell free expression	83
Results and Discussion	85
Isolate library evaluation	85
Pathway prediction through antiSMASH	86
Amplification and cloning of lantipeptide pathway	88
Expression of lantipeptide pathway	91
Cell free LanA expression.....	95

Conclusion.....	98
Chapter Bibliography	99
Introduction	104
No one knows what is like to be the Batman, to be the sad man... ..	104
Fly away skyline bat, fly... white-nose syndrome treatments.....	105
We are the cavemen – extreme environments as source of compounds	106
Project objectives	108
Materials and Methods	109
Isolate collection.....	109
DNA extraction and Whole Genome Sequencing of isolates.....	110
Genome assembly and annotation	110
Organism identification and secondary metabolites prediction	111
Orphamide pathway primer design and amplification.....	112
Extraction and fractionation of fermented isolates.....	113
Mass spectrometry of positive extract fractions.....	117
Results and Discussion	120
Disk diffusion assays.....	120
Whole Genome sequencing and assembly.....	123
Secondary metabolite profiling and genome mining.....	127
Correlating the Metabolome and genome	130
Conclusions.....	141
Chapter Bibliography	142
Final remarks.....	149

Appendix 1: GNPS Molecular Networking compound identification of n-	
butanolic and aqueous extracts of <i>Streptomyces</i> sp. ICC1.....	151

List of acronyms and abbreviations

A: adenylation domain

ACN: acetonitrile

BODIPY: boron-dipyrromethene

bp: base pair

BSA: bovine serum albumin

C: condensation domain

CFU: colony forming unit

CTAB: cetyltrimethylammonium bromide

Cys: cysteine

Dha: dehydroalanine

Dhb: dehydrobutyrine

DIMBOA: 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one

DMSO: dimethylsulfoxide

DNA: deoxyribonucleic acid

E. coli: *Escherichia coli*

eDNA: environmental DNA

EDTA: Ethylenediaminetetraacetic acid

GEIs: genomic islands

GNPS: Global Natural Product Social Molecular Networking

HGT: horizontal gene transfer

HT: Hickey-Tresner medium

HMW: high molecular weight

IDT: Integrated DNA Technologies

ISP-2: international *Streptomyces* project medium 2

IPTG: isopropyl β -D-1-thiogalactopyranoside

kb: kilobase

kD/kDa: kilodalton

Lan: lanthionine

LB: Lysogeny Broth/Luria Bertani medium

LC-ESI-HRMS: Liquid Chromatography Electrospray Ionization High Resolution Mass Spectrometry

MeLan: methyllanthionine

MES: Multigene Expression System

MH: Mueller Hinton medium

NEB: New England Biolabs

NRPS: nonribosomal peptode synthetases

OD: optical density

PBS:phosphate buffer saline

PCP:peptidyl carrier protein domain

PDB: Phage Dilution Buffer

PEG: polyethylene glycol

PFGE: pulsed field gel electrophoresis

PKS: polyketide synthases

Ppant : 40-phosphopantetheine cofactor

R2A: Reasoner's 2A medium

RiPPs: Ribosomally synthesized and posttranslationally modified peptides

RNA: ribonucleic acid

RNase A: ribonuclease A

RT room temperature

S. sp. PBH53: Streptomyces sp. PBH53

S. sp. ICC1: Streptomyces sp. ICC1

S. sp. ICC4: Streptomyces sp. ICC4

SB: Saboureaud Dextrose medium

SDS-PAGE: sodium dodecyl sulphate-polyacrylamide gel electrophoresis

Ser: serine

SMs: secondary metabolites

SPE: Solid Phase Extraction

Thr: threonine

Tris:Tris(hydroxymethyl)aminomethane

V: volume or volumes

V8: clarified V8 vegetable cocktail broth

WNS: white-nose syndrome

μL: microliter

μg: microgram

List of figures

- Figure 0.1: Cultivation and activity-guided drug discovery flow.** Different stages of a discovery, from sample to lead molecule and steps to achieve a potential drug candidate are shown, along with the compound range needed to complete each stage. Koehn, F. E., & Carter, G. T. *Nat. Rev. Drug Discov.*, 2005, 4, 206-220, by permission.5
- Figure 1.1: Erythromycin, a polyketide.** Synthesis pathway illustrating the enzyme modules involved in the product formation and post-synthase modifications. Domains shown: ACP: acyl-carrier protein; AT: acyltransferase; DH: dehydratase; ER: enoyl reductase; KR: ketoreductase KS: ketosynthase; TE: thioesterase. Weissman, K.J., *Nat. Prod. Rep.*, 2016, 33, 203, by permission. 16
- Figure 1.2: NRPS organization.** Top: Basic module organization, with the three essential domains (A,T and C) highlighted. Bottom: Reactions catalyzed by the three basic NRPS domains. a) Transfer of phosphopantetheine from coenzyme A to a conserved serine in the T-domain b) A domain selects the amino acids and covalently binds it to the T domain c) chain elongation is catalyzed by C domains via peptide bond formation between the monomer (acceptor) and the growing peptide chain (donor). Fischbach, M.A., Walsh, C.T., *Chem Rev*, 2006, 106: 3468–3496, by permission. 18
- Figure 1.3: Bacterial composition of 450 DNA sequences obtained from street sediments and soils.** Different phyla are represented with its respective percentages. Adapted from the work of Hill, P. et al., *Microb. Ecol.*, 2011, 61(2), 286-302, by permission. 21

Figure 1.4: Scheme of the bioprospection strategy used in this work.....	22
Figure 1.5: General scheme of eDNA extraction process.....	25
Figure 1.6: General scheme showing steps of clone library construction and array	33
Figure 1.7: PFGE for eDNA size evaluation. Red box in Low Range PFG Marker indicate bands of size range between 48.5 and 23.1 kb. A: Pretoria Bridge Mussel B: 80 Fulton C: Pretoria Bridge Red Sand, D: Onondaga Lake E: Seep	43
Figure 1.8: Multiplex PCR probing for 3 classes of natural products. Genomic DNA of two bacterial isolates, JTG09 and PBH53 was used as template. Bands corresponding to type I PKS (1.1-1.3kb), NRPS (800bp) and type II PKS (600-700bp) are present in both samples.	45
Figure 1.9: Plate #17 from the original library clone pooling system. Each colony in the grid proceeds from a library member	46
Figure 1.10: Agarose gel showing results for multiplex amplification of DNA from individual clones. First lanes on left: molecular marker (1kb Molecular Marker, NEB). Lanes 2-18, top and bottom: samples. Lane 19: negative control. Lane 20, top and bottom: positive control	48
Figure 1.11: PFGE of digested positive fosmids. Double digestion performed with restriction enzymes HindIII and BamHI. 1 to 14: digested samples. C: fosmid control L: ladder (Low Range PFG Marker).....	51
Figure 2.1: Nisin A biosynthesis. Steps shown: NisA linear peptide formation, dehydration carried by NisB, thioether linkage formation by NisC and activation of peptide by cleavage performed by NisP. Retrieved from Shi, Y. et al., <i>J. Am. Chem. Soc.</i> 2011, 133, 2338–2341, by permission.	66

Figure 2.2: Agarose gel of amplification of NRPS pathways from isolates library. Lane 1: 1kb DNA ladder (NEB). Lanes 2-10: samples. In detail: NRPS amplicon on PBH53 isolate.....	86
Figure 2.3: antiSMASH 3.0.0 prediction of lantipeptide pathway in <i>Streptomyces</i> sp. PBH53.....	88
Figure 2.4: <i>Streptomyces</i> sp. PBH53 lantipeptide cluster. A: LanA, the lantipeptide (189bp). B: LanB, Lantipeptide dehydratase (1.9kb). C: LanC, lanthionine synthetase (1.1kb).....	88
Figure 2.5: PCR amplification of LanA (200 bp), check digestion of LanB (1.9kb) and LanC (1.1kb)	89
Figure 2.6: Check digestion of multiple expression vector pMGX-K containing LanB (~2kb) and LanC (~1.1kb). Two vectors originated from successful transformed colonies are depicted.	91
Figure 2.7: SDS-PAGE of LanB and LanC expression via multiexpression gene system pMGX. Ly: lysate; FT: flow through; In detail: LanB (71kDa) and LanC (40kDa).....	92
Figure 2.8: Attempt of LanA expression; no visible band corresponding to LanA (7.5kDa) appeared. Ly: lysate; FT: flow through. 15kD strong band visible on fractions.	94
Figure 2.9: SDS-PAGE imaging of cell-free expression of lantipeptide. L: molecular marker (not visible); N: negative control; P: BODIPY-Lys labelled LanA expression U: unlabelled LanA expression. Red band is indicative of signal saturation.	96

Figure 3.1: Disk diffusion assays of bacterial extracts. Row 1: <i>E. coli</i> BW25113. Row 2: <i>Bacillus subtilis</i> 168. Row 3: <i>Pseudomonas aeruginosa</i> PA01.	122
Figure 3.2: Inhibition halo in detail. Left: <i>B. subtilis</i> 168, inhibited by ICC1 aqueous fraction. Right: <i>Pseudomonas aeruginosa</i> PA01 inhibited by ICC1 unprocessed broth.	122
Figure 3.3: Genome alignment of genomes of <i>Streptomyces</i> sp. A: alignment of <i>Streptomyces</i> sp. ICC1 against <i>Streptomyces</i> sp. ICC4 and <i>Streptomyces lavendulae</i> strain CCM 3239. B: alignment results of <i>Streptomyces</i> sp. ICC4 against <i>Streptomyces</i> sp. ICC1 and <i>Streptomyces lavendulae</i> strain CCM 3239.	125
Figure 3.4: Genome features of <i>Streptomyces</i> sp. A: <i>Streptomyces lavendulae</i> strain CCM 3239 B: <i>Streptomyces</i> sp. strain ICC1 C: <i>Streptomyces</i> sp. isolate ICC4. Numbers indicate genes per feature group.	126
Figure 3.5: Diketopiperazines identified from ICC1 by GNPS analysis of LC-ESI-HRMS analysis of culture extracts	131
Figure 3.6: Coelichelin, one of the secondary metabolites found in <i>Streptomyces</i> sp. cave isolates. A) biosynthetic gene cluster for coelichelin. B) proposed mechanism for biosynthesis. C) list of genes in the cluster, their annotation, length and similarity based on BLAST library alignment.	134
Figure 3.7. The type III polyketides 2', 5' – dimethoxyflavone and nordentatin. The molecules were identified in metabolomic analysis of culture broth and a potential putative gene cluster encoding these two compounds was identified by antiSMASH analysis of the genomes of both ICC1 and ICC4. A) biosynthetic gene cluster for both type III polyketides. B)	

proposed mechanism for their biosynthesis. C) list of genes in the cluster, annotation, length and similarity based on BLAST library alignment136

Figure 3.8: Proposed diazepinomicin biosynthetic pathway from ICC1 and

ICC4. Diazepinomicin was detected in one of the fractions from ICC1. A) biosynthetic gene cluster diazepinomicin. B) proposed mechanism for diazepinomicin biosynthesis. C) list of genes in the cluster, annotation, length and similarity based on BLAST library alignment.139

List of tables

Table 1.1: Profile of all soil samples used for eDNA extraction. Major sample component is presented.....	26
Table 1.2: primer sequences used for amplification of PKS type I and II and NRPS pathways	31
Table 1.3: Lambda-DNA standard concentrations for standard curve.....	36
Table 1.4: Library array result for positive clones. Positives shown for each step of the probing process, with percentage of positives related to the total library.....	47
Table 1.5: BLAST alignments for fosmid insert sequences. To random library positive hits sent, the closest species match is shown, along with coverage and identity percentages.	50
Table 2.1: List of prospected urban isolates from the laboratory collection	71
Table 2.2: Household isolates library constructed to complement lab isolate collection	75
Table 2.3: primer sequences used for amplification of lantipeptide pathway	78
Table 3.1: Bacterial isolates collected form various environments in British Columbia. Environmental source, cultivation conditions and strain sensitivity test results are presented.....	109
Table 3.2: primer sequences used for amplification of orphamide pathway.	112

Table 3.3: susceptibility of strains tested against cave bacterial extracts and fractions obtained by Kupchan partition	121
Table 3.4: 16S rRNA, whole genome sequencing (WGS), coverage and identity percentage of BLAST alignment of cave bacteria genomic DNA sequences.	123
Table 3.5: summary of antiSMASH 4.1.0 secondary metabolites prediction of <i>Streptomyces</i> sp. strains ICC1 and ICC4.....	128

General introduction

Secondary metabolites: what, where and why

Secondary metabolites (SMs) are small organic molecules present in different living organisms such as bacteria, fungi and plants. About 500,000 secondary metabolites have been reported to date, with 33,500 bioactive microbial metabolites described. Of these 33,500 microbial metabolites, about 12.5% (4,200) are metabolites of unicellular bacteria and cyanobacteria, 41% (13,700) are products of Actinomycete fermentations, and about 47% (15,600) are of fungal origin (Bérdy, J. 2012; Nett M., Ikeda H., Moore B.S. 2009). Although these molecules are usually not key components in biological functions such as growth, development and reproduction, they have an important role in the host's protection, interaction and competition with other organisms, affecting its essential functions.

As an evolutionary perspective, SMs are very dynamic due to selection pressures that result from environmental changes and co-evolution of species that interact via SMs (Reynolds H.T. et al, 2017), with a high impact of horizontal gene transfer (HGT) in bacterial and fungal species adaptation. These peer-acquired adaptive traits form blocks known as genomic islands (GEIs) (Juhas, M. et al, 2009) and might be beneficial for bacteria under certain growth or environmental conditions (Schmidt & Hensel, 2004).

Due to the variety of conditions and very particular environments and evolutionary pressure suffered by organisms, it is no wonder that the variety of molecules identified as secondary metabolites is also diverse. From fungal, plant

and bacterial toxins that can cause serious impact to molecules bringing society-changing benefits, this dichotomy is indicative of the diversity of the secondary metabolites nature (Bills, G. F., Gloer, J. B., 2016).

Secondary metabolites as allies for human life (or how the once seen as bad guys became the good guys against the real bad guys)

In the history of humanity, microbes were usually the hidden forces behind many processes, from food fermentation to diseases, but often disregarded as so by the lack of knowledge of their existence. The beginning of a scientific look at the microbial world that popularized the germ theory of disease and that introduced the hope of prevention and treatment date from late nineteenth century with Louis Pasteur (Smith, K., 2012). Despite his work, the correlation between diseases and microorganisms was only made ten years later, after Robert Koch first proof that microbes could actually cause an infectious disease (Koch, 1876).

The twentieth century brought a new perspective for the treatment of microbial diseases with the discovery of compounds that could be used as effective treatments. Sir Alexander Fleming reported some emblematic discoveries. In 1921, he discovered in “tissues and secretions” an important bacteriolytic substance, lysozyme. In 1928, while working on influenza virus, he observed that mould, *Penicillium notatum*, had developed accidentally on a *Staphylococcus* culture plate and that the mould had created a halo that prevented bacterial growth. Upon investigation, it was found that a so-called

“mould juice” culture prevented the growth of a range of harmful microorganisms. His assistants isolated the active compound, which was named penicillin (Aldrige, S., Parascandola, J., Sturchio, J.L., 1999).

Penicillin opened the doors to what is known as the Golden Age of antibiotics discovery. Within a decade after World War II, a number of important antibiotics were discovered and developed for therapeutic use. They became the foundation for the treatment of infectious diseases, with one-half of the drugs commonly used today discovered between 1950 and 1960 (Davies, J., 2006).

This stellar achievement was heavily based in natural molecules retrieved from bacteria and fungi after streptomycin was isolated in 1944 from *Streptomyces griseus*. Every effort was made to reach all corners of the globe, with extensive sampling from many different places. A sample from Borneo sent in 1952 grew *Streptomyces orientalis*, from which vancomycin was eventually extracted; vancomycin became available for patient use in 1958 (Gould, K., 2016). From being the causes of disease or the inconvenient spoilers of food, microorganisms were now the stars of life-saving science discoveries.

Compound screening approaches and the discovery pitfall

To address the daunting challenge of discovering new microbial compounds from their producing organisms, many new techniques have been developed. With the extensive sampling made during the decades of the Golden Age of antibiotic discovery, the need for a rational and practical approach for sample processing was essential.

With the absence of techniques focusing on molecular and genetic characteristics, the simplest way to assess organism diversity was cultivation. In traditional cultivation, or “brute force” discovery, bacteria were grown in isolation, under specific conditions, inevitably missing out on many of the signals that trigger the production of natural products in the original habitat (Kolter, R., Van Wezel, G.P, 2016).

While highly successful at the beginning – as evidenced by the high antibiotic discovery rates achieved in the 1950s and 1960s – this brute force method faced limitations by the 1970s as the easily discoverable compounds were already described. The approach was also overlooked many organisms that were unable to thrive under laboratory cultivation conditions, leading to a narrow and very time-consuming look on the potential of samples, even with high-throughput methods increasing processing capabilities (MacLeain, J., et al, 2016).

A solution to expand the drug catalogue was the development of natural products derivatives. Many of these derivatives, based on known antibiotics, were approved and have been important contributors to treating clinical infections (Lewis, K., 2012). Designing synthetic drugs failed because many antibiotics do not obey Lipinski's rule of five, making chemical synthesis less predictable (Lewis, K., 2013).

In addition to cultivation, many other challenges limit natural product drug discovery, including activity assays and compound isolation and identification. Increasing amounts of a compound are required for purification and molecular structure elucidation, and even more of the target molecule is needed for medicinal chemistry drug development (Koehn, F. E., Carter, G. T.,

2005). Figure 0.1 illustrates the pipeline of the activity-based approach of discovery.

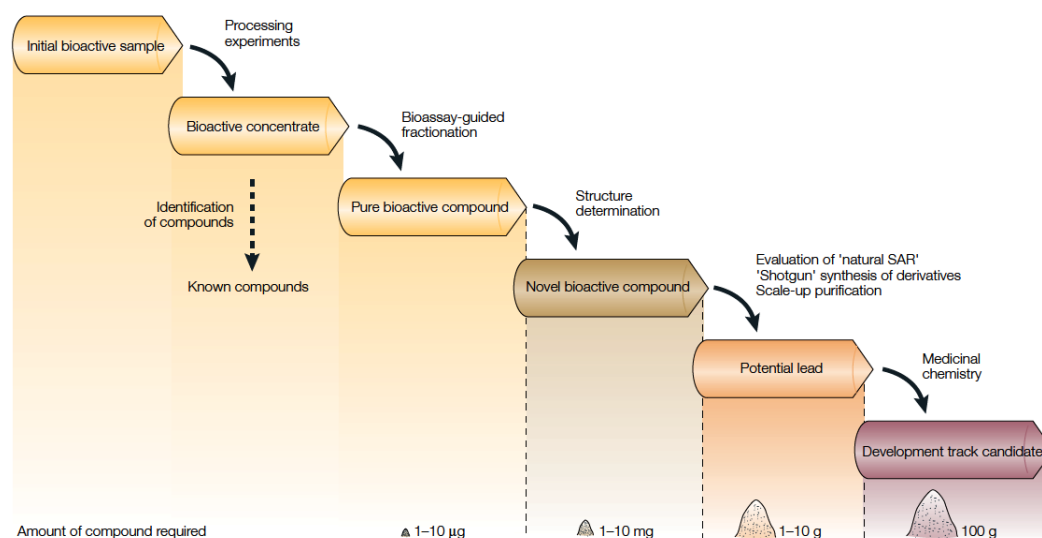


Figure 0.1: Cultivation and activity-guided drug discovery flow. Different stages of a discovery, from sample to lead molecule and steps to achieve a potential drug candidate are shown, along with the compound range needed to complete each stage. Koehn, F. E., & Carter, G. T. *Nat. Rev. Drug Discov.*, 2005, 4, 206-220, by permission.

One of the major pitfalls in natural product drug discovery is rediscovery of known compounds when assessing samples. Re-discovery rates became high using the old paradigm of screening natural product extracts because of the numerous antibiotics already identified in the first 50 years of screening and the time required to isolate and characterize new compounds (Baltz, R.H., 2006). Pharmaceutical companies have long abandoned the search due to costs and labour intensive tasks (Fabbretti, A., Gualerzi, C. O., Brandi, L. 2011; Brown, E. D., & Wright, G. D., 2016).

The second and most serious pitfall is bacterial resistance. There is a general scientific consensus that the use and abuse of antimicrobial drug use in a variety of settings contributes to the burden of a quick development of

antimicrobial resistance (Hoelzer, K. et al, 2017). Antibiotic resistance genes are then vertically passed on to the next generation of microbes; in some cases, they are acquired through horizontal transfer from one microbe to another, when thriving in the same microbial environment. This global proportions problem needs to be urgently addressed as antibiotic-resistant pathogens have caused numerous treatment failures that eventually led to both hospital morbidities and mortalities (Imperial, I.C.V.J., Ibana, J.A., 2016). In 2006, almost 100,000 people in the United States alone died due to hospital-acquired infections, costing about \$8 billion to the US economy. Global projections account a global economic burden of about \$3 trillion/year, with about 444 million people succumbing to infections by 2050 (Aslam, B. et al, 2016). Several claims to rethink and revive antibiotic discovery have been made in the past decades as an effort to prevent the so-called return to the pre-antibiotic era and decrease resistance development rates.

Make molecules great again: unleash the untapped potential of biomolecules and decrease rediscovery rates

The efforts to make antibiotics an effective treatment for infections again and avoid rediscovery of known compounds are increasing over time. Many researchers are still working to expand and revive the antibiotics catalogue despite the lack of interest from the pharmaceutical industry.

One of the research lines focuses on the discovery of new compounds from untouched microbial sources. As studies say that 85-99% of bacteria are unculturable using traditional techniques, with poor phylogenetic diversity in

the culturable ones (Lok, C., 2017; Pham, V.H.T, Kim, J., 2012), we are missing out on the molecular universe of microbes. The use of molecular biology techniques is essential to allow access to this untapped potential.

With the recent development of computational genomic approaches to natural product discovery in prokaryotic genome sequences, tens of thousands of biosynthetic gene clusters (BGCs) have been identified (Tracanna, V. et al, 2017). BCG mining with the use of sequenced genomes has become more and more the usual practice for biological molecules involved in secondary metabolites production, with computer-based tools such as antiSMASH, BAGEL and PRISM being of great help to analyze large data volumes.

The use of mining techniques allows us to rethink sampling. Instead of organism sampling and cultivation from a plethora of environments, mining environmental DNA opens the possibility to study diversity and distribution of compounds directly in their environment. Heterologous expression and synthetic biology allow subsequent capturing of sequenced pathways and enable tailored compound production (Ziemert, N., Alanjary, M., & Weber, T., 2016).

Directed sampling and mining of DNA sequences allow a focused discovery, targeting specific classes of compounds or organisms, thus allowing the selection of potential drug candidates and discarding previously described compounds, leading to a decrease in rediscovery rates. Focused discovery allows researchers to study 'silent' clusters, cryptic compounds that have escaped detection for the reason that they are not produced under standard laboratory conditions (Monciardini, P. et al, 2014), broadening the chances of novel molecule discovery.

In conclusion, there is still a vast potential to be explored in terms of undiscovered metabolites in microorganisms, and the last decades brought many technological advances in terms of molecular biology to allow elucidation of untapped compounds to alleviate the drug crisis we are currently experiencing.

References

Aldrige, S., Parascandola, J., Sturchio, J.L., (1999). *The discovery and development of penicillin*[Brochure]. London:Stairways Communications.

Aslam, B., Wang, W., Arshad, M. I., Khurshid, M., Muzammil, S., Rasool, M. H., ... Baloch, Z. (2018). Antibiotic resistance: a rundown of a global crisis. *Infection and drug resistance*, 11, 1645–1658

Baltz, R. H. (2006). Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *Journal of Industrial Microbiology and Biotechnology*, 33(7), 507–513.

Bérdy J. (2012). Thoughts and facts about antibiotics: where we are now and where we are heading. *The Journal Of Antibiotics*, 65, 385–395.

Bills, Gerald F. and Gloer, James B. (2017). Biologically Active Secondary Metabolites from the Fungi, p 1087-1119. In Heitman, Joseph and Howlett, Barbara J. and Crous, Pedro W. and Stukenbrock, Eva H. and James, Timothy Y. and Gow, Neil A. R.(ed), *The Fungal Kingdom*. Retrieved from doi:10.1128/microbiolspec.FUNK-0009-2016.

Brown, E. D., & Wright, G. D. (2016). Antibacterial drug discovery in the resistance era. *Nature*, 529, 336.

Davies, Julian (2006). Where have all the antibiotics gone? *The Canadian journal of infectious diseases & medical microbiology = Journal Canadien des maladies infectieuses et de la microbiologie medicale*, 17(5), 287-90.

Fabbretti, A., Gualerzi, C. O., & Brandi, L. (2011). How to cope with the quest for new antibiotics. *FEBS Letters*, 585(11), 1673–1681.

Gould, K. (2016). Antibiotics : from prehistory to the present day. *Journal of Antimicrobial Chemotherapy*, 71: 572–575.

Hoelzer, K., Wong, N., Thomas, J., Talkington, K., Jungman, E., & Coukell, A. (2017). Antimicrobial drug use in food-producing animals and associated human health risks: what, and how strong, is the evidence?. *BMC veterinary research*, 13(1), 211.

Imperial, I. C., & Ibana, J. A. (2016). Addressing the Antibiotic Resistance Problem with Probiotics: Reducing the Risk of Its Double-Edged Sword Effect. *Frontiers in microbiology*, 7, 1983.

Juhas, M., van der Meer, J. R., Gaillard, M., Harding, R. M., Hood, D. W., & Crook, D. W. (2009). Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiology Reviews*, 33(2), 376–393.

Koch, R. (1876). Die aetiologie der milzbrand-krankheit, begrundet auf die entwicklungsgeschichte des *Bacillus antracis*. *Beitr. Biol. Pflanz.* 2, 277–310.

Koehn, F. E., & Carter, G. T. (2005). The evolving role of natural products in drug discovery. *Nature Reviews Drug Discovery*, 4(3), 206-220.

Kolter, R., & van Wezel, G. P. (2016). Goodbye to brute force in antibiotic discovery? *Nature Microbiology*, 1, 15020.

Lewis, K. (2013). Platforms for antibiotic discovery. *Nature Reviews Drug Discovery*, 12, 371.

Loc, C. (2015). Mining the microbial dark matter. *Nature*, 522(7556), 270-3.

McLain, J. E., Cytryn, E., Durso, L. M., & Young, S. (2016). Culture-based Methods for Detection of Antibiotic Resistance in Agroecosystems: Advantages, Challenges, and Gaps in Knowledge. *Journal of Environmental Quality*, 45, 432–440.

Monciardini, P., Iorio, M., Maffioli, S., Sosio, M., & Donadio, S. (2014). Discovering new bioactive molecules from microbial sources. *Microbial biotechnology*, 7(3), 209-20.

Nett M, Ikeda H, Moore BS. (2009). Genomic basis for natural product biosynthetic diversity in the Actinomycetes. *Natural Product Reports*, 26:1362–1384.

Pham, V. H. T., & Kim, J. (2012). Cultivation of unculturable soil bacteria. *Trends in Biotechnology*, 30(9), 475–484.

Reynolds, H. T., Slot, J. C., Divon, H. H., Lysøe, E., Proctor, R. H., & Brown, D. W. (2017). Differential Retention of Gene Functions in a Secondary Metabolite Cluster. *Molecular Biology and Evolution*, 34(8), 2002–2015.

Schmidt, H., & Hensel, M. (2004). Pathogenicity islands in bacterial pathogenesis. *Clinical microbiology reviews*, 17(1), 14-56.

Smith, K. A. (2012). Louis Pasteur, the father of immunology? *Frontiers in Immunology*, 3 (April), 1–10.

Tracanna, V., Jong, A. De, Medema, M. H., & Kuipers, O. P. (2017). Mining prokaryotes for antimicrobial compounds: from diversity to function. *FEMS Microbiology Reviews*, 41, 417–429.

Ziemert, N., Alanjary, M., & Weber, T. (2016). The evolution of genome mining in microbes - a review. *Nat. Prod. Rep.*, 33(8), 988–1005.

General Objective of the PhD

This research is focused on finding new secondary metabolites from bacterial sources through genome mining, allowing the study of unknown molecules that could be potential candidates for usage as drugs against microbial infections, cancer, and immune diseases.

Bioprospection of these molecules was conducted with the use of targeted environmental sampling, exploration of unusual environments and molecular and computational biology, to overcome the challenges faced with the use of traditional approaches.

Genetic data generated during the research was deposited on public database, to enable growth of the resources available to the scientific community and to serve as a source of data for heterologous expression efforts throughout the research development.

We expect to bring a significant contribution to expand the understanding of secondary metabolites produced by microorganisms under special environments and to help increase the use and development of new strategies of compound discovery and development of new antimicrobials.

Chapter 1: Bioprospection of an environmental DNA library for assessing new natural products

Introduction

Bioprospecting

Bioprospecting is the act of discovery and commercialization of new products based on biological resources, focusing the screening efforts for bioactive compounds (Saslis-Lagoudakis, C. H. et al., 2012). The need of new compounds for use in the pharmaceutical (e.g. antibiotics against multi-resistant pathogens) and the agro- and food industries (e.g. agrochemicals, biocontrol products, food additives) stresses the importance of novel bio-sources (Müller, C. A., Obermeier, M. M., & Berg, G., 2016). Considering the history of microbiological sources of important treatments, many efforts are concentrated in bioprospecting this specific group of organisms.

Natural products targeted when bioprospecting

A variety of natural products can be retrieved from a single organism, but some families of compounds are more relevant depending on the prospecting goals. In our work, we targeted polyketide synthases (PKS) and nonribosomal peptide synthetases (NRPS), as many antibacterial compounds belong to these families.

PKS are modular enzymes that are evolutionary and mechanistically related to fatty acid synthases. Individual active sites operate sequentially on protein-bound reactive intermediates to yield a polyketide product; in most

cases, a given enzyme within the protein assembly line catalyzes one and only one reaction in the overall biosynthetic pathway (Robins, T. et al. 2016).

The operational unit of an assembly line PKS is called a module. The minimal functional domains in a module are a ketosynthase (KS), an acyltransferase (AT), and an acyl-carrier protein (ACP). The KS receives the growing polyketide chain from an upstream module via an ACP thioester-to-KS active site thiol acyl transfer reaction, hereafter referred to as “chain translocation”. It also catalyzes a decarboxylative Claisen condensation between this incoming polyketide chain and an ACP-bound malonyl (or substituted malonyl) extender unit, a reaction that is hereafter referred to as “chain elongation”. The AT transacylates, via a thioester exchange reaction, the extender unit from an appropriate acyl-CoA species onto the pantetheinyl side chain of an ACP. Thus, the ACP serves as a tethering site that is occupied first by the extender unit and then the newly elongated polyketide chain (Robins, T. et al. 2016). Modification of the product can occur through the synthesis, with the help of a variety of enzymes, or after (post-synthase modifications). Figure 1.1 illustrates the process using DEBS pathway.

PKS pathways are subdivided into categories related to the way modules are used. Type I PKS refers to linearly arranged and covalently fused modules, forming large multifunctional enzymes, working sequentially and playing its role a single time. Type II indicates a dissociable complex of discrete and usually monofunctional enzymes, working in an iterative fashion, with post-synthase processing modifications tailoring the final product. Furthermore, a third group of multifunctional enzymes of the chalcone synthase type is denoted as type III PKSs (Hertweck, C., 2009).

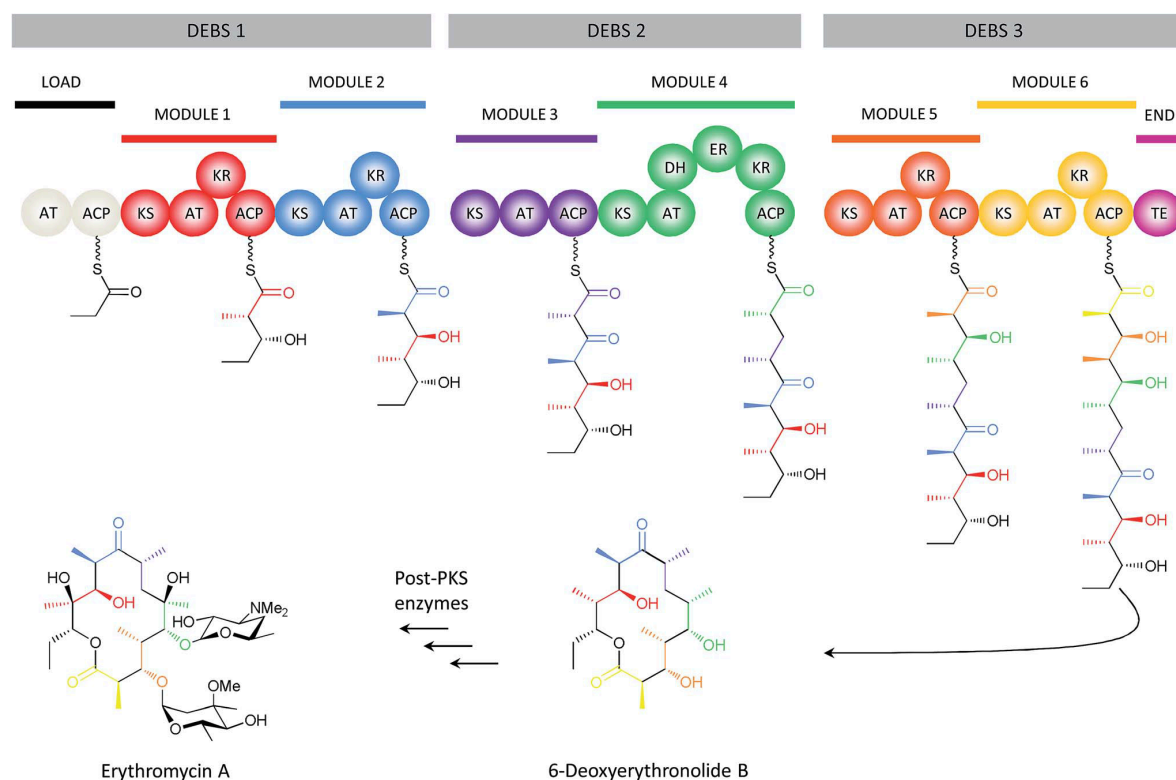


Figure 1.1: Erythromycin, a polyketide. Synthesis pathway illustrating the enzyme modules involved in the product formation and post-synthase modifications. Domains shown: ACP: acyl-carrier protein; AT: acyltransferase; DH: dehydratase; ER: enoyl reductase; KR: ketoreductase KS: ketosynthase; TE: thioesterase. Weissman, K.J., *Nat. Prod. Rep.*, 2016, 33, 203, by permission.

NRPS is another family of compounds targeted in our bioprospection work. Non-ribosomal peptide synthetases are complexed enzymes that assemble secondary metabolites of peptide origin. In contrast to the nucleic acid-dependent ribosomal synthesis that relies on the 20–22 proteinogenic amino acids, NRPSs can process several hundreds of monomers, including D- and β -amino acids, α -keto and α -hydroxy acids as well as heterocycles and fatty acids, among other moieties (Marahiel, M. A., Stachelhaus, T. and Mootz, H. D., 1997; Konz, D. and Marahiel, M. A., 1999; Finking, R. and Marahiel, M. A., 2004).

Three essential domains compose the NRPS module: an adenylation (A) domain responsible for selecting the substrate and activating it as an aminoacyl-

adenylate, peptidyl carrier protein (PCP or T) domain that carries all acyl-intermediates on the terminal –SH group of its phosphopantetheine (PPT) cofactor, and a condensation (C) domain which forms peptide bonds between the acyl-S-PCP intermediates of two adjacent modules (Marahiel, M.A., 2016). Figure 1.2 shows the basic arrangement of NRPS modules and the reactions performed by different enzymes.

During synthesis, the growing peptide chain is handed over from one module to the next and can be edited by enzymes during this process, such as the epimerization (E) domain, which changes residue configuration, until it reaches the final module's PCP-domain. This module contains in most cases a thioesterase (TE) domain that is important for the liberation of the product. Some NRPSs require posttranslational modification and, under certain conditions, editing to be functionally active (Finking, R. and Marahiel, M. A., 2004).

In a real-life context, PKS and NRPS are remarkable in the pharmaceutical industry, as many compounds used as treatments belong to either one of these families. Drugs such as rifamycin (antitumor), lovastatin (a cholesterol-lowering agent), erythromycin (an antibiotic), and FK506 (an immunosuppressant) (Wong, F.T., Khosla, C., 2012) are polyketides, and antibiotics such as penicillin, vancomycin, and cephalosporin (Felnagle, E. A. et al., 2008), and immunosuppressant cyclosporin are nonribosomal peptides. Their versatility, associated with the vast amount of unknown compounds, is still appealing as a target for bioprospecting.

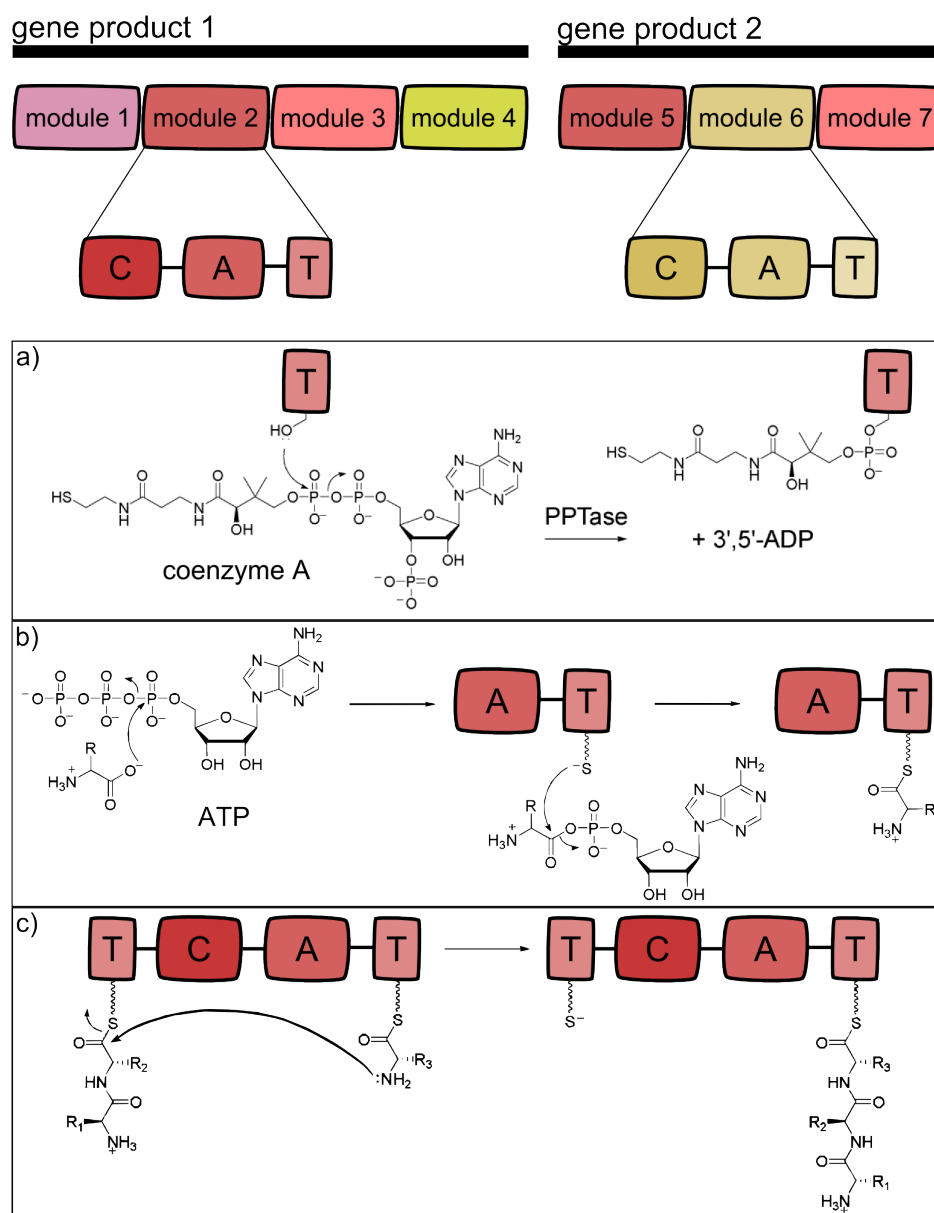


Figure 1.2: NRPS organization. Top: Basic module organization, with the three essential domains (A,T and C) highlighted. Bottom: Reactions catalyzed by the three basic NRPS domains. a) Transfer of phosphopantetheine from coenzyme A to a conserved serine in the T-domain b) A domain selects the amino acids and covalently binds it to the T domain c) chain elongation is catalyzed by C domains via peptide bond formation between the monomer (acceptor) and the growing peptide chain (donor). Fischbach, M.A., Walsh, C.T., *Chem Rev*, 2006, 106: 3468–3496, by permission.

Actinobacteria, the powerhouse of bioprospection

Among bacterial groups targeted as sources of novel bioactive compounds, the phylum Actinobacteria is the main target and the most explored one. Actinobacteria are Gram-positive filamentous bacteria with a high guanine-plus-cytosine (G+C) content in their genomes. They grow by a combination of tip extension and branching of the hyphae. This similarity with fungi is what gave them their old name, actinomycetes, which derives from the Greek words for ray (aktis or aktin) and fungi (mukēs) (Barka et al. 2015).

Actinobacteria sources account for about 45 % of all microbial bioactive secondary metabolites with 7,600 of these compounds (80%) being produced by *Streptomyces* (Bérdy, J., 2005). Despite this astonishing productivity, it has been predicted that only about 10% of the total number of natural products that can be synthesized by these organisms have been discovered (Watve, M.G. et al. 2001).

With the majority of natural products from actinomycetes undiscovered, we can ask ourselves the reasons why this potential is still unexplored. The traditional cultivation approach is one of the reasons, as this phylum tends to present slow growth and peculiar culturing conditions when compared to other organisms (Jiang, Y. et al. 2016).

Another issue was unveiled with the advent of genome sequencing. Even with the use of tailored culturing methods, many pathways encoding secondary metabolites remain silent during cultivation, despite evidence of their presence during DNA analysis. These cryptic biosynthetic pathways may produce many

novel bioactive compounds with the potential to rejuvenate stalled drug discovery pipelines, and activation of these silent biosynthetic pathways are thus of major interest (Ochi, K., Hosaka, T., 2012).

Actinobacteria is still one of the biggest targets for secondary metabolites prospection, with many indescribable molecules. Even with the limitations posed by cultivation methods and the inability of cryptic pathways expression, researchers are developing methods to overcome these difficulties and get the most out of these important organisms.

Straight to the point: sampling optimization

Actinobacteria are widely distributed in both terrestrial and aquatic (including marine) ecosystems, especially in soil, where they play a crucial role in recycling biomaterials by decomposition and humus formation (Ventura M. et al., 2007) The traditional perception of these organisms as autonomous in soil and freshwater is being radically reviewed as it is increasingly evident that members of this phylum are among the most successful colonizers of all environments in the extremobiosphere, often occurring as the dominant population (Bull, T.A., 2011). As Actinobacteria is present everywhere, sampling can focus on sample quality or microbial community composition analysis.

Microbial community analysis seems to be an important determinant of sampling strategies, as studies have shown that land use affects the community structure (Hill, P. et al, 2011). The results point out that the overall actinobacterial community structure appears to be controlled by land use and that street sediments are enriched with Actinobacteria when compared to cultivated and uncultivated soils (Figure 1.3). This is likely to be due to their desiccation resistant nature (Cruz-Martínez, K. et al. 2009).

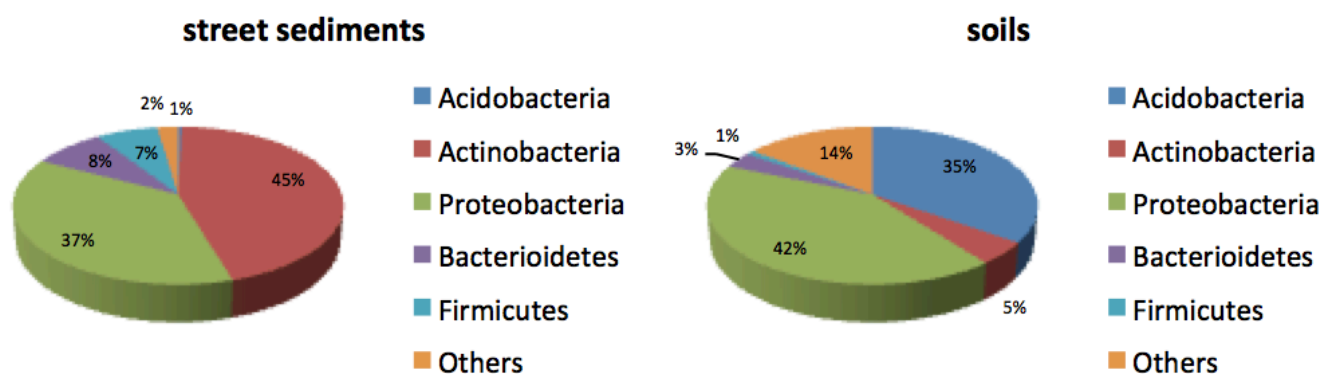


Figure 1.3: Bacterial composition of 450 DNA sequences obtained from street sediments and soils. Different phyla are represented with its respective percentages. Adapted from the work of Hill, P. et al., *Microb. Ecol.*, 2011, 61(2), 286-302, by permission.

Sampling can be tailored to target environments with a higher prevalence of actinobacterial community, causing a positive impact on the number of samples and volume of material collected to maximize the chances of retrieving novel secondary metabolites.

How to be a Bookworm: library construction and array

With optimized sampling methods, there is a need for analysis tools optimization as well. High-throughput metagenomic methods are the preferred ones, with high success rates and a well-established pipeline. Figure 1.4 illustrates the approach used in the present work.

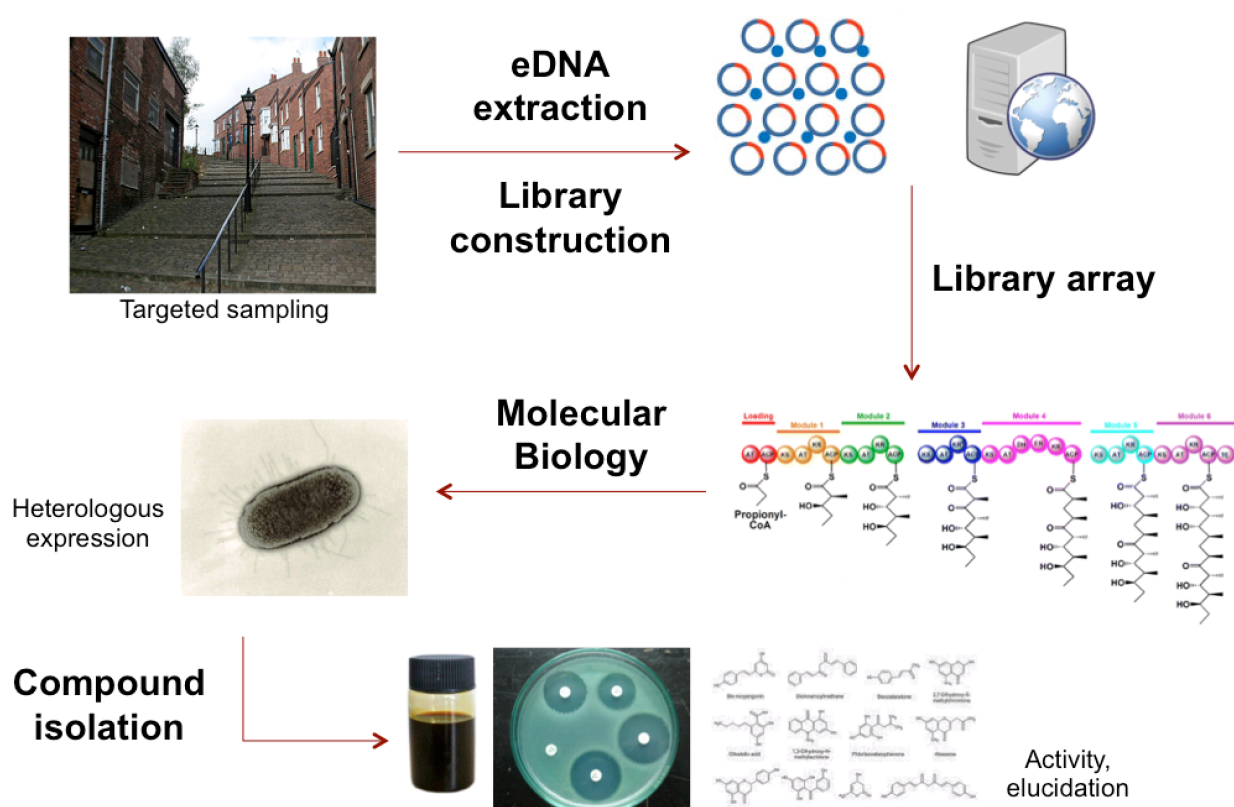


Figure 1.4: Scheme of the bioprospection strategy used in this work

A library-based method of metagenomic analysis is comprised of producing libraries using DNA from the samples, with *E. coli* clones as host cells. These clones undergo homology-based screens using degenerate PCR primers designed to recognize conserved sequences within secondary metabolite biosynthetic genes. An alternative method of amplification is the linker-

amplified shotgun library (LASL) approach, where short linkers are attached to target DNA, and PCR is conducted with primers targeting the linkers (Schmitz, J. E., et al. 2008).

As more metagenomic sequencing data appears in publicly available databases, it should also be possible to use purely bioinformatics-based search strategies to identify new natural product biosynthetic enzymes and gene clusters (Banik, J.J., Brady, S.F., 2010). This novel approach not only saves time and resources but also contributes to the expansion of the sequencing data that is added to public databases.

Many successful cases of library-based molecule discovery were reported in the literature. Some were focusing on extreme environment DNA analysis, such as marine and desert soils, while others reported entire product families retrieved from cosmid libraries made with eDNA from uncultured soils (Yang, J. K. et al., 2018; Singh, M.P. et al. 2007; Brady, S. F., & Clardy, J., 2003; Brady, S. F., Chao, C. J., & Clardy, J., 2002). Commercially available kits make the library construction an easy task and as short as 2 days as per some manufacturers (Epicentre, 2012).

In conclusion, bioprospecting largely benefits from methods that take the focus away from a culture-based approach, bringing metagenomics and high-throughput as solutions for quick and accurate metabolite discovery.

Project objectives

The goal of this project was to access novel gene clusters encoding secondary metabolites with potential for drug development. This will be achieved through the collection and evaluation of environmental DNA from different urban soil samples, followed by the built and array of a fosmid library containing the genetic material.

Fosmids would be analyzed for the presence of potential gene clusters using molecular biology techniques such as amplification and sequencing, and positive clones would be evaluated for the nature of the genes present.

We expect to contribute to the development of an optimized protocol to enable genetic screening of environmental samples, expanding the number of environments studied and the genetic potential of such places.

Materials and Methods

Sample collection

A variety of soils from different localities and having different soil properties were collected for eDNA extraction. The composition of all soils was analyzed (Table 1.1).

eDNA extraction

Environmental DNA was extracted using an enzymatic method proposed by Patrick Hill (unpublished).

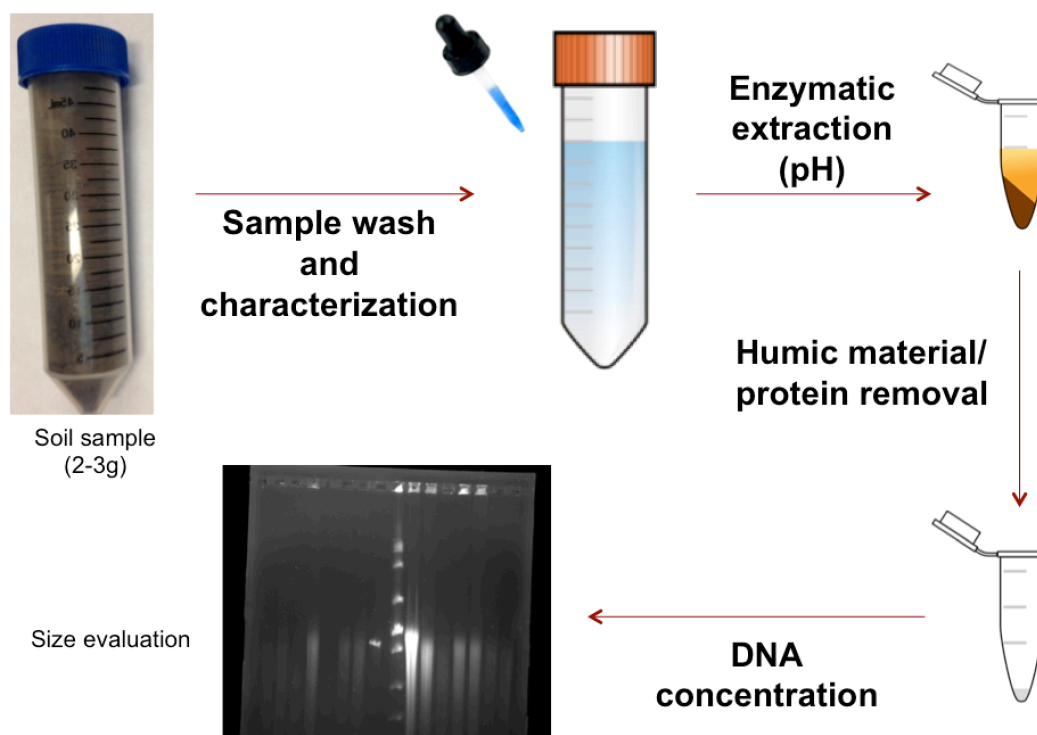


Figure 1.5: General scheme of eDNA extraction process.

Table 1.1: Profile of all soil samples used for eDNA extraction. Major sample component is presented.

Sample name	Location	Landscape	Composition
Holbrock	Arizona, USA	High desert	Sand
4985 Camino del Viento (1)	Arizona, USA	Scrubby desert	Sand
4985 Camino del Viento (2)	Arizona, USA	Mud slide	Mud
Raton Pass	New Mexico, USA	Scrubby open pine forest	Soil/plant debris
Table Mesa	Arizona, USA	Black Canyon	
Cave Creek	Arizona, USA	Desert	Sand
Bethany	Missouri, USA	Cultivated	Soil
Ellisworth	Kansas, USA	Cultivated scrubby grass	Soil
Victor	Iowa, USA	Cultivated	Soil
Termidine nature	San Diego, USA	Marine	Sand
Diamond	San Diego, USA	Marine	Sand
80 Fulton	Ottawa, Canada	Household	Compost
82 Fulton	Ottawa, Canada	Household	Compost
85 Fulton	Ottawa, Canada	Household	Compost
Boston Brick	Boston, USA	Urban	Debris
Byward Market 3	Ottawa, Canada	Urban	Debris
Vancouver Estuary	Vancouver, Canada	Marine	Sandy
Worthing 3	England	Marine	Coarse sand
Wolf	Gatineau, Canada	Park/forest	Soil
Dunlop	Gatineau, Canada	Park/forest	Soil
Vancouver Skate Park	Vancouver, Canada	Marine	Sand
Belcher 1	Hudson Bay, Canada	Marine	Gravel/mud
Belcher 2	Hudson Bay, Canada	Marine	Gravel/mud
Billings Bridge	Ottawa, Canada	Forest	Soil
Church Road	Gatineau, Canada	Forest	Soil
Experimental Farm 1B	Ottawa, Canada	Crop	Grassland
Woodbine East	Ottawa, Canada	Cultivated	Soil
Woodbine West	Ottawa, Canada	Cultivated	Soil
Dune Open Forest	Ottawa, Canada	Forest	Soil

1925 Pine Forest	Ottawa, Canada	Forest	Soil/plant material
Conserved Sand Dune	Ottawa, Canada	Dune	Sand
Fungal Dune	Ottawa, Canada	Dune	Sand
Light Beige Dune	Ottawa, Canada	Dune	Sand
Pretoria Bridge Mussel	Ottawa, Canada	Urban	Mud
Seep	Ottawa, Canada	Urban	Sand
Ottawa Canal Red Sand	Ottawa, Canada	Urban	Sand
Pretoria Bridge Red Sand	Ottawa, Canada	Urban	Sand
Ottawa Canal	Ottawa, Canada	Urban	Mud
Onondaga Lake	New York, US	Shore	Sand
Sommerset Bridge	Ottawa, Canada	Urban	Debris
Experimental Farm 4A	Ottawa, Canada	Crop	Cultivated

1. Preparation of samples for eDNA extraction

2-3 g of soil samples was weighted and washed using 50 mL of Tris-EDTA buffer (50 mM Tris, 50m M EDTA). The addition of buffer was done in two parts to ensure soil disaggregation. Samples were centrifuged for 1 hour (4000 g, 4 °C). The pellet can be taken to DNA extraction or can be frozen at -20 °C for further use.

2. Extraction at high pH

The pellet was suspended in 2.5 mL of solution 1 (0.5 M NaCl, 50 mM Tris, 50mM EDTA, lysozyme 5 mg/mL pH 8.3) and incubated for 1 hour at 37°C. After incubation, 140 µL of an aqueous fresh SDS 20 % solution and 100 µL of a 10 mg/mL proteinase K solution were added. The samples were incubated again for another hour at 37 °C.

3. Extraction at low pH

5 mL of solution 2 (0.5 M NaCl, 10mM EDTA, 0.3 M succinic acid, pH 5.7) was added to the soil solution, along with 700 µL of 20 % aqueous fresh SDS solution. The mixture was incubated at 65 °C for 30 minutes. At this step, solutions can be left at 4° C overnight if needed.

The samples were centrifuged for 20 minutes (15,000 g, 4° C) and the supernatant was transferred to a fresh tube (approximate yield: 8 mL solution).

4. Protein and polysaccharides removal, first DNA concentration

1 mL of 0.5 M NaCl and 2 mL of fresh 10 % CTAB were added to the samples supernatant, and incubated at 65 °C for 30 minutes. To this mixture, 10 mL chloroform and 0.5 volume of freshly prepared PEG 50 % were combined to the sample solution and left overnight at 4 °C for DNA precipitation.

The next day, the DNA was concentrated through centrifugation (1 hour, 14,000 g, 4 °C) and washed with 70 % ethanol. 480 µL of TE buffer (50 mM Tris, 50mM EDTA) were used to resuspend the pellet. The liquid was transferred to a 1.5 mL microfuge vial.

5. Protein and humic material removal

1 volume of phenol:chloroform:isoamylalcohol(25:24:1) was added to the sample. The mixture was centrifuged for 15 minutes (15,000 g, 4° C) and the top layer was kept, avoiding the collection of the white central layer or the bottom solution. The procedure was repeated using 1 volume of chloroform.

0.57M CaCl₂ was added to the sample (10 µL CaCl₂ per 200 µL sample) and incubated for 30 minutes at 65 °C. The liquid was centrifuged at 15,000g for 20 minutes at room temperature.

To the collected supernatant, potassium acetate 3 M (110 µL per 500 µL sample) was combined and left at room temperature for 15 minutes. The insoluble material was removed through centrifugation (15,000 g, 20 minutes, room temperature).

6. eDNA concentration

600 µL of room temperature isopropanol was mixed to the collected supernatant, followed by the precipitation of DNA with the use of a centrifuge (15,000 g, 20 minutes). The pellet was washed with 600 µL of 70 % ethanol, which was removed by centrifugation under the same conditions as the isopropanol removal. The pellets were left to air-dry and were resuspended in 20 µL TE buffer (50 mM Tris, 50 mM EDTA).

Pulsed field electrophoresis

For both eDNA size evaluation and fosmid library construction pulsed field electrophoresis (PFGE) was used. This method allows for separation of high molecular weight (HMW) DNA (Schwartz & Cantor, 1984), proven useful to determine the quality of the DNA used for library and to retrieve natural product pathways.

To prepare the pulsed field gel to be used, 150 mL of 1 % agarose was used, with the separation of 10 mL of the total volume to seal the wells. This agarose was kept warm at 60 °C.

Sample preparation was done by mixing 2-3 µL of a 50 ng/µL solution of each desired sample with 20 µL of ultrapure water and 5 µL of running buffer. Any standards used were prepared the same way.

The pulsed field apparatus (MB-10 Megabase System, Tyler Research Corporation, Alberta, Canada) was prepared by being pre-chilled with 0.5 X TBE buffer (89 mM Tris, 89 mM boric acid, 2 mM EDTA, pH 8.3) to 14 °C.

Loading of samples was done on gel before immersion of it in the apparatus, along with an arbitrary thin slice of marker (Low Range PFG Marker, 1-200 kilobases, New England Biolabs(NEB)) on the middle well, to allow better run and a more realistic size comparison. After the sample loading, the wells were sealed using the warm agarose set aside and the gel was placed inside the running chamber. Running conditions were as followed: 145-150 V, 22-24 h, 14 °C.

Gels were soaked for 1 hour in a solution containing 200 mL 0.5 X TBE buffer and 20 µL of a 10 mg/mL of ethidium bromide solution to allow visualization of bands. For gels prepared for fosmid library construction no soaking was performed. Bands were visualized with the help of a transilluminator apparatus (Alphamager HP, Alpha Innotech).

PCR evaluation of the presence of natural products pathways

As targets for natural product discovery, three classes were picked for analysis: Polyketide Synthases class I and II (PKSI and PKSII) and Non Ribosomal Peptide Synthetases (NRPS). The presence of conserved regions in these enzymes made them targets for identification via polymerase chain reaction (PCR).

The type I PKS and NRPS primers used were designed based on degenerated sequences of alignments of ketosynthase, acyl transferase and adenylation domains (Ayuso-Sacido & Genilloud, 2005). Type II PKS sequences used were based on modular ketoacyl-synthases shared by a number of organisms (Wawrik, Kerkhof, Zylstra, & Kukor, 2005). The alpha subunit was chosen as a target. Primers were provided by Integrated DNA Technologies (USA). Primer sequences are shown in Table 1.2.

Table 1.2: primer sequences used for amplification of PKS type I and II and NRPS pathways

Primer	Sequence
K1S (TI PKS)	5'- TSAAGTCSAACATCGGBCA -3'
M6R (TI PKS)	5'- CGCAGGTTSCSGTACCAGTA -3'
540F (TIIPKS)	5'- GGITGCACSTCIGGIMTSGAC -3'
1100R (TII PKS)	5'- CCGATSGCICCSAGIGAGTG -3'
A3F (NRPS)	5'- GCSTACSYSATSTACACSTCSGG3-3'
A7R (NRPS)	5'- SASGTCVCCSGTSCGGTAS3 -3'

For Type I and NRPS primers, a multiplex reaction was prepared as per previously cited work of Ayuso and Sacido. For a final 50 μ L reaction, reactions were performed containing 50-100 ng DNA template, 5 % DMSO, 25 mM $MgCl_2$, 0.2 mM of dNTP mix, 0.2 μ M of each primer, 1.25 U Taq DNA polymerase (NEB), 5 X reaction buffer without $MgCl_2$ (Promega). Amplification cycles consisted of an initial heating at 95 $^{\circ}C$ for 5 minutes, followed by 35 cycles of 95 $^{\circ}C$ for 30 seconds, 2 minutes at 58 $^{\circ}C$ and 4 minutes at 72 $^{\circ}C$, with a final elongation at 72 $^{\circ}C$ for 10 minutes.

For Type II primers, the reaction consisted in 50 μ L reactions containing 50-100 ng DNA template, 25 mM $MgCl_2$, 0.8 % 10 mg/ml BSA, 0.2 mM of dNTP mix, 0.4 μ M of each primer, 1.25 U Taq DNA polymerase (NEB), 5 X reaction buffer without $MgCl_2$ (Promega). Amplification cycles were as follows: initial heating for 5 minutes at 95 $^{\circ}C$ followed by 40 cycles of 1 min at 95 $^{\circ}C$, 1 min at 64 $^{\circ}C$ and 90 seconds at 72 $^{\circ}C$ with a final 15 min extension step at 72 $^{\circ}C$.

A multiplex reaction containing all three sets of primers combined (Types I and II PKS and NRPS) was used for library analysis. The 50 μ L reaction contained 50 ng DNA template, 8 % DMSO, 25 mM $MgCl_2$, 0.4 mM of dNTP mix, 0.2 μ M of each primer, 0.8 % 10 mg/ml BSA, 1.25 U Taq DNA polymerase (NEB), 5 X reaction buffer without $MgCl_2$ (Promega). Amplification cycles consisted of an initial heating for 5 minutes at 95 $^{\circ}C$, followed by 35 cycles of 1 min at 95 $^{\circ}C$, 90 seconds at 58 $^{\circ}C$ and 90 seconds at 72 $^{\circ}C$, followed by a final 10 min extension step at 72 $^{\circ}C$.

All amplification reactions were analyzed using 1 % (w/v) agarose gel electrophoresis with ethidium bromide staining.

Library construction

For the eDNA library construction, a fosmid kit was used to facilitate the process (CopyControl™ Fosmid Library Production Kit, Epicentre). The procedure described refers to the steps provided on the kit manual and any modifications made were highlighted.

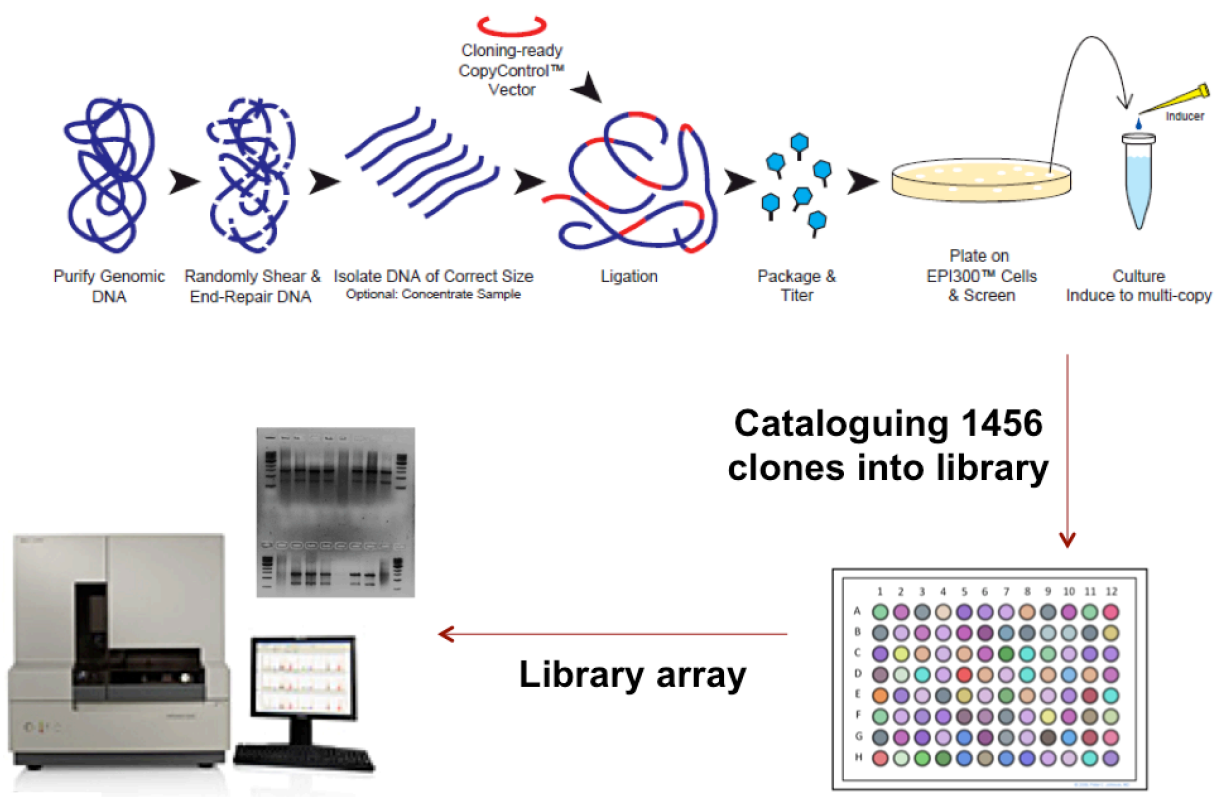


Figure 1.6: General scheme showing steps of clone library construction and array

7. Preparation of library insert DNA

The eDNA from the following samples were used as input: Onondaga Lake, 80 Fulton, Ottawa Canal Red Sand Seep, Pretoria Bridge Red Sand and

Ottawa Canal. There was no distinction between the samples for analysis, considering all input as part of a single library. Isolate PBH53 from a collection of the lab was used as a positive control for the construction, as environmental samples yield can be low when compared to isolate libraries.

No DNA shearing was needed for the samples as all met the size and concentration recommended for library construction (~40 kb, 0.5 µg/µL).

8. End-repair of fragments

DNA fragments were repaired to generate 5' phosphorylated blunt ends for further ligation to the vectors. For a total volume of 80 µL reaction were combined 8 µL 10X End-Repair Buffer, 8 µL 2.5 mM dNTP Mix, 8 µL 10 mM ATP, 20 µg insert DNA and 4 µL End-Repair Enzyme Mix. The kit, with the exception of the DNA input, provided all reagents. Reactions were carried on for 45 minutes at room temperature. 20 µL of 5X Green GoTaq Buffer (Promega) was added to each sample and the mixture was inactivated at 70 °C for 10 minutes.

9. Size selection of input

All samples after inactivation were loaded on a 1 % (w/v) low melting point (LMP) agarose gel. A 42 kb human DNA, provided by the kit, was used as a supplementary control size. Two control bands were prepared for size selection by comparison. All samples were run on a pulsed field gel apparatus for 22 hours at 35 V.

After the run, the size comparison bands were excised and stained with ethidium bromide. The gel was reassembled to help determining the area with desired size of input DNA to be extracted.

10. Recovery of DNA fragment

The samples were excised from the gel, placed in tared tubes, weight, and their masses were calculated. Mass data was used to estimate the amount of buffer required to release DNA from the gel. For each 1 mg of agarose, 1 μ L of 50 X gelase buffer was used. The samples were incubated at 70 °C until the gel was completely molten. Gelase (1 U/100 μ L molten agarose) was added and incubated at 45 °C for 1 hour. A water bath at 70 °C was used to stop enzyme activity.

To recover DNA, the tubes were chilled in an ice bath for 5 minutes and spun in a microcentrifuge (10,000 g, 20 minutes). The supernatant was recovered and precipitated for 10 minutes using 1:10 v/v of 3 M sodium acetate and 2.5 volumes of a solution of ethanol 70 %. The solution was centrifuged at 10,000 g for 20 minutes, repeating the ethanol wash and centrifugation steps two more times. The DNA pellet was left to air-dry for 10 minutes and resuspended in 10 μ L of TE buffer.

11. DNA concentration determination

For further steps, the concentration of DNA was determined with PicoGreen assay (Quant-iT PicoGreen, Invitrogen).

Standard curves using lambda-DNA serial microdilutions were prepared in triplicate according to Table 1.3. 100 μL of each solution was loaded in a microplate, along with the same volume of a working solution of Quant-iT PicoGreen (Quant-iT PicoGreen reagent:1X TE buffer, 1:200).

Table 1.3: Lambda-DNA standard concentrations for standard curve

Point	DNA concentration (ng/200 μL)
1	200
2	100
3	50
4	10
5	2
6	0.5
7	0.3
8	0.05
9	0.02
10	0.005
11	0 (blank)

The mixed solutions were incubated at room temperature for 5 minutes and read using a plate reader (Infinite F200, Tecan). Reading settings: excitation wavelength 485 nm, emission wavelength 535 nm, integration time 20 μs , 25 reads.

12. DNA ligation

The selected DNA was ligated to the fosmid vector provided in the kit (CopyControl Fosmid Library Production Kit with PCC1FOS vector). The 10 μ L reaction was prepared as follows: 1 μ L 10X Fast-Link Ligation Buffer, 1 μ L 10 mM ATP, 1 μ L CopyControl pCC1FOS vector, 6 μ L concentrated insert DNA, 1 μ L Fast-Link DNA Ligase. All reagents were present in the kit. The ligation was left for 4 hours at room temperature and inactivated at 70 °C for 10 minutes.

13. Packaging of vectors into phages and library preparation

To form the bacterial library, the fosmid vectors were packed into lambda bacteriophages to be used for cell infection. The packaged phages were titrated to determine the best dilutions to yield a library.

The day before packaging, an overnight culture of the bacterial cells to be used as library storage (*Escherichia coli* EPI300-T1, kit component) was prepared. 50 mL of LB broth (10 g Bacto-tryptone, 5 g yeast extract, 10 g NaCl, pH 7.5) enriched with 10 mM MgSO_4 and 0.2 % Maltose was inoculated with 0.5 mL of the overnight culture and incubated at 37 °C until A_{600} of 0.8-1.0 was reached.

The ligation was performed in two steps. First, 10 μ L of the ligation reactions were added to 25 μ L of phage solution (MaxPlax Packaging Extract, kit component) and incubated for 2 hours at 30 °C. Secondly, an additional 25 μ L of phage solution was then added and the incubation step was repeated.

At the end of incubation, 940 μ L of Phage Dilution Buffer, PDB (per L: 50 mL 1 M Tris HCl, 2 g MgSO_4 , 2% w/v gelatin) and 25 μ L of chloroform was added to each reaction. The samples were incubated overnight at 4 °C.

Three dilutions of the packaged phages in PDB, 1:10, 1:100 and 1:1000, were prepared to monitor the quality of the packaging. 10 µL of each solution along with an undiluted preparation was incubated for 1 hour at 37 °C with 100 µL of *E. coli* EPI300-T1^R. This mixture was plated on LB agar containing 12.5 µg/mL chloramphenicol and incubated overnight at 37 °C. The number of colonies was monitored to determine if it agreed with the expected value based on the dilution of library preparation. The same cultivation approaches were used to generate the fosmid library, with the clones collected and stored on 1 mL of LB broth enriched with 12.5 µg/ml chloramphenicol and 50 % glycerol.

Library array

14. Secondary metabolite pathway evaluation

For accessing the genetic information contained in the library, an array method was developed to facilitate the process. 1456 individual clones were cultivated in groups of 40 on LB agar + 12.5 µg/mL chloramphenicol plates. These clones were mixed together and used as a template for a colony multiplex PCR as described on “PCR evaluation of the presence of natural products pathways” section.

The positive pools were reassessed again by breaking them into 4 groups of 10 clones each. The colonies were again mixed and used as DNA template for another round of multiplex PCR. The positive results from this assay were separated into individual clones, which were evaluated using the multiplex PCR technique one more time. These individual clones were ready to have their genetic information determined by vector extraction and sequencing.

15. Isolation and identification of inserts

To enable isolation of fosmids, positive individual clones were autoinduced to increase vector copy. 50 mL of LB broth was supplemented with chloramphenicol 12.5 µg/mL and 100 µL/mL Fosmid Autoinduction Solution (provided with Fosmid kit). Clones were cultivated overnight at 37 °C under agitation (200 rpm).

The isolation of fosmids was performed using a no-column method due to low yields usually obtained with column-processed fosmids. Cells were pelleted by centrifugation at 4,000 g for 10 minutes. After decanting supernatant, cell pellet was resuspended in 2.5mL of solution P1 (6.1 g Tris, 3.7 g EDTA.2H₂O, 100 µg/mL RNase, pH 8.0). After resuspension, 2.5 mL of P2 (8 g NaOH in 900 mL H₂O, 100 mL 10% SDS) was added and mixture was incubated at room temperature for 3 minutes or until solution became clear. To this, 3.5 mL of solution P3 (294 g potassium acetate in 500 mL H₂O, pH 5.5 with acetic acid, adjust to 1 L) was introduced in the reaction and tube was inverted 4-6 times to allow protein precipitation. A centrifugation at 4,000 g for 20 minutes was performed to pelletize cell debris.

7 mL of the supernatant was transferred to a clean tube containing 7 mL of isopropanol and cooled on ice for 15 minutes. The reaction was centrifuged at 4,000 g for 30 minutes. To the pelletized DNA, 500 µL of chilled 70 % ethanol were added and the mixture was centrifuged for 16,000 g for 10 minutes. The ethanol was removed and the pellet was air-dried for 20 minutes and

resuspended in 100 µL of elution buffer (10mM Tris, pH 8.5). The DNA concentration was determined by UV detection using a nanodrop.

To determine insert identity, two unique restriction enzyme cutters in the fosmid sequence flanking the insert region were used to digest the vector. For a 50 µL reaction, fosmids (2000 ng) were incubated with 10 units of HindII and BamHI (NEB) and 10X buffer (5 µL 3.1 Buffer, NEB, provided with enzymes) and 41 µL of H₂O at 37 °C for 2 hours and 30 minutes. The enzymes were inactivated for 20 minutes at 80 °C. Digestion products were analyzed through agarose pulsed field gel electrophoresis (1 % agarose, 140 V, 22 hours).

Preliminary characterization of fosmids was achieved through Sanger sequencing. Based on gel analysis, selected samples were sent for sequencing of both of insert ends. Primers used for the process were T7 forward and two custom sequences matching flanking regions of insertion site, provided by the fosmid kit manual: FP (5' GGATGTGCTGCAAGGCGATTAAGTTGG 3') and RP (5' CTCGTATGTTGTGTGGAATTGTGAGC 3').

Permanent library stock preparation

In order to obtain a copy of the library that could be kept as permanent stock, all colonies were transferred to 96 well microplates using a colony picking device (Qpix 420, Molecular Devices). To create the copy, the petri dishes containing the 40 clones pool were used.

The equipment camera was aligned using a dummy plate via Camera Alignment software process. The application also allows determining the depth

of agar to be pin pricked. To sanitize the robotic pin head, solutions of 1 % bleach, 70 % ethanol and sterile distilled water were prepared. The entire equipment was sanitized with UV irradiation for 10 minutes and the pins were washed in solutions and heat dried.

96 microplates were filled with 200 μ L of LB broth supplemented with chloramphenicol 12.5 μ g/mL and placed into the equipment for receiving the samples. The source plates were placed onto the camera plate holder. The General Picking process was used to generate the output, with a customized size threshold for identifying the colonies on the plate. The equipment neglected colonies too close from the edges of the Petri dishes by default, so they were hand picked instead.

The transferred colonies were cultivated overnight at 37 °C. Plates were topped up with 80 μ L of 30 % glycerol per well and stored at -80 °C.

Results and Discussion

Sample collection and eDNA extraction

From all samples processed through eDNA extraction, only a few were considered suitable for library construction: 80 Fulton, Onondaga Lake, Pretoria Bridge Red Sand, Pretoria Bridge mussel, and Seep. The cut-off was made based on size and eDNA concentration evaluation after extraction.

A preliminary analysis of DNA concentration during the extraction process was performed to give a panorama of the extraction process and to calculate DNA loss during humic material and protein removal. Despite promising results obtained for the majority of samples during this initial check, there is a significant loss during purification and concentration of nucleic acids, which is expected during this step, as it is for any other nucleic acids extraction method, including specific kits for soil DNA (Ferrand, J., et al., 2014; Martin-Laurent, F. et al., 2001; Luna, G. M., Dell'Anno, A. and Danovaro, R., 2006; Kennedy, N. A. et al., 2014).

Size evaluation of eDNA was the second and most important parameter to determine the samples for library construction, as high molecular weight DNA correlates to a higher chance of getting intact or mostly intact secondary metabolite pathways. Some organisms have a representative portion of their genomes related to these pathways, such as the genus *Streptomyces*, which can dedicate up to ~ 10 % of their 8-10 Mb genomes for the biosynthesis of bioactive secondary metabolites (Stulberg, E. R et al., 2016; Kallifidas, D., Jiang, G., Ding, Y.,

& Luesch, H., 2018). These pathways can be long: for example, the complete erythromycin synthetic gene cluster is 55kb in length (Zhang, H. et al., 2010).

Selected eDNA samples contain fragments ranging from 35 to 42 kb, the maximum size accepted as an insert fragment by the fosmid used for library construction. Figure 1.7 shows the pulsed-field gel electrophoresis (PFGE) used for quality evaluation of samples.

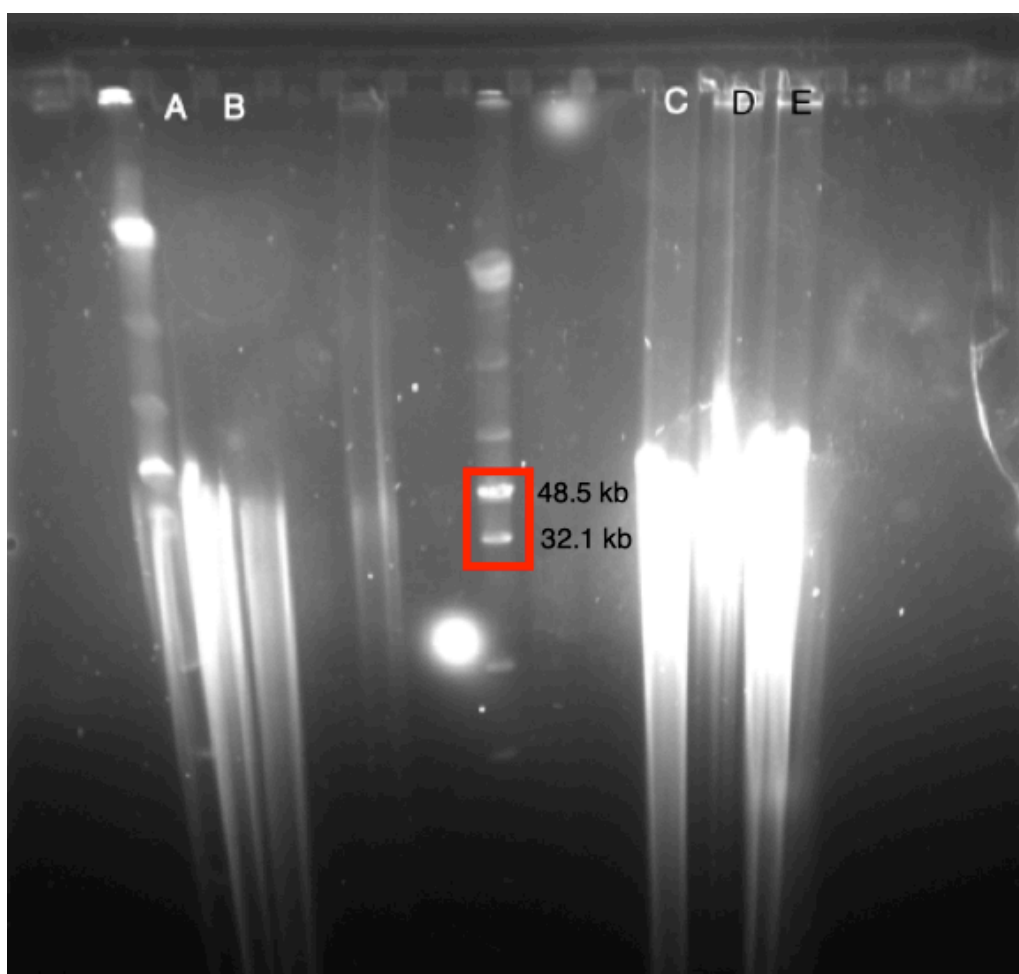


Figure 1.7: PFGE for eDNA size evaluation. Red box in Low Range PFG Marker indicate bands of size range between 48.5 and 23.1 kb. A: Pretoria Bridge Mussel B: 80 Fulton C: Pretoria Bridge Red Sand, D: Onondaga Lake E: Seep

Library construction

The five eDNA selected samples were used for library construction. The CopyControl™ Fosmid Library Production Kit was chosen due to the presence of an optimized vector and *E. coli* strain and for its quick method for clone generation. The versatility of the system allowed the construction of libraries using various organisms as DNA source, such as cyanobacteria, marine sediment and other soil microorganisms (Rabausch, U. et al., 2013; Felczykowska, A. et al., 2014; Hårdeman, F., & Sjöling, S. 2007), showing that the kit would suit our needs.

Our library consisted of 1456 clones obtained from the combination of the five eDNA used as inserts. Each sample contributed to the library differently, with the compost sample (80 Fulton) being the one with the least number of clones. Sand samples (Onondaga Lake, Pretoria Bridge Red Sand, and Seep) were the major contributors. This could be caused by the homogeneous nature of the compost and sampling of the fresh, superficial layer in comparison to a dynamic environment such as Onondaga Lake and the Rideau Canal, both subjected to anthropogenic activities, which could have caused an impact in the bacterial community.

Optimization of library array system

Library evaluation strategies were developed to facilitate the array, as we had as targets 3 major classes of natural products: type I and type II PKS and Non

Ribosomal Peptides. For discovery of positive clones, the library was probed by PCR with the use of three sets of degenerate primers targeting conserved regions.

Even with a library of moderate size, the array could be very slow and time and reagent consuming if individual PCR reactions were carried out for each pair of primers. To solve this problem, a multiplex system was designed.

Two pairs of primers (PKSI and NRPS) were already known to function in a one-pot reaction, as per the work of Ayuso-Sacido and Genilloud (2004). Type II PKS primers however had not yet been evaluated for a multiplex system. To test multiplexing the three pairs of primers, genomic DNA samples from two bacterial isolates from the laboratory that were reported to possess pathways belonging to the three classes of NPs, JTG09 and PBH53, were analyzed. Figure 1.8 shows the result of multiplex PCR test.

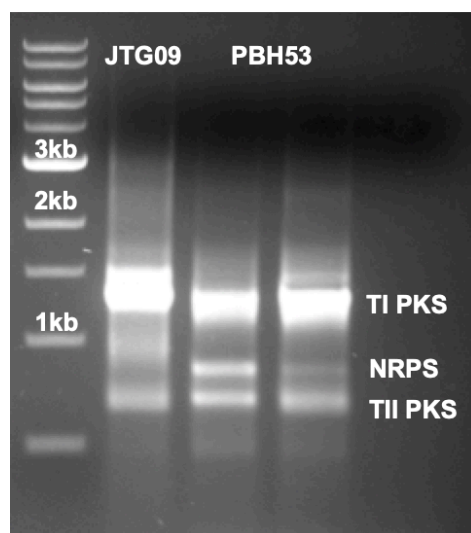


Figure 1.8: Multiplex PCR probing for 3 classes of natural products. Genomic DNA of two bacterial isolates, JTG09 and PBH53 was used as template. Bands corresponding to type I PKS (1.1-1.3kb), NRPS (800bp) and type II PKS (600-700bp) are present in both samples.

Three to four bands were observed in each sample, corresponding to the different targets. Type I PKS amplicons were observed at 1.3 kb, with additional bands at 1.1 kb. NRPS adenylation amplicons were visualized at 800 bp and type II PKS were located below, between 600-700 bp. This data demonstrated that a multiplexed analysis could be used to evaluate the library as a first check for positive clones.

Regarding the clones, the strategy adopted was pooling the samples and refining the analysis for the positive pools by making smaller groups until the individual positive clones were identified. The starting point was the cultivation of all library members in plates that were used as the first pool, containing around 40 colonies. Figure 1.9 depicts one of the plates from the original pooling system.

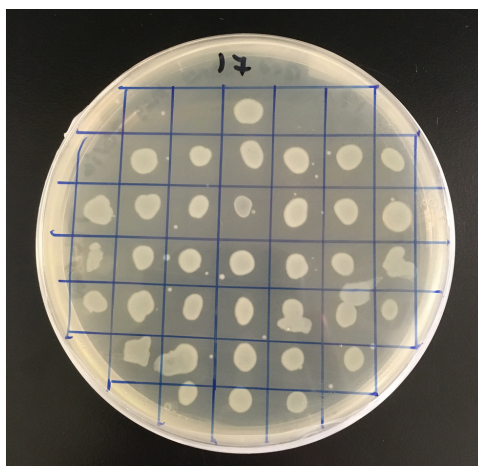


Figure 1.9: Plate #17 from the original library clone pooling system. Each colony in the grid proceeds from a library member

Library evaluation

With the pooling system and the multiplex PCR, all clones were evaluated for the evidence of secondary metabolites pathway. From 1456 clones composing the entire library, 103 individual ones scored positive for a secondary metabolite gene. Table 1.4 condenses the results for all steps of library array. Figure 1.10 illustrates one of the agarose gels for DNA amplification on individual clones.

Table 1.4: Library array result for positive clones. Positives shown for each step of the probing process, with percentage of positives related to the total library.

	Whole library	40 clones pool	10 clones pool	Individual clones
Pools/clones	1456	6	18	180
Total clones	1456	240	180	103
(%)	100%	16.5%	12.4%	7%

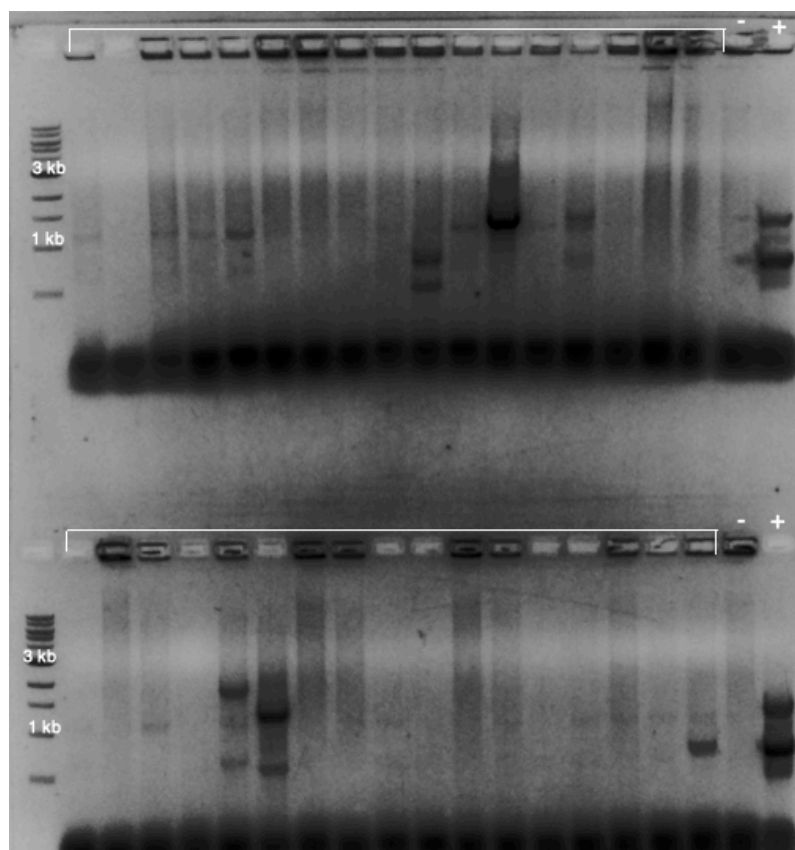


Figure 1.10: Agarose gel showing results for multiplex amplification of DNA from individual clones. First lanes on left: molecular marker (1kb Molecular Marker, NEB). Lanes 2-18, top and bottom: samples. Lane 19: negative control. Lane 20, top and bottom: positive control

Regarding the percentage of positives, our library shows 7% of total clones being positive for the secondary metabolites analyzed. In comparison with other libraries, our numbers are very discrepant. For example, Foerstner et al. (2008), when analyzing farm soils and had a library possessing 183,536 clones, with only 52 of them positive for type I PKS. This corresponds to only 0.0283% of positives. Owen et al. (2013), in a study targeting desert soils from Mexico, constructed a library of 15,000,000 unique clones, with 20,661 positives (0.13%). Cuadrat et al. (2015), when performing a metagenomic analysis of upwelling-affected seawater for type I PKS and NRPS built two libraries from

port/sewage and upwelling sampling. Both libraries were similar in size and number of positives, with 38 out of 409,1110 (0.0092%) and 46 out of 451,722 (0.0101%).

Our high percentage of positives could be related to the sampling on urban environments, which are known to have a higher prevalence of Actinobacteria, increasing the chances of finding genes coding for secondary metabolites. The targeted sampling method seems to be a good method to follow when looking for specific compounds to boost the chances of finding such pathways. The small number of clones in the library could be due to the small amount of starting sample to purify eDNA from (usually 2-2.5 g). Nevertheless, the library size is acceptable as sufficient number of positive clones was recovered.

Analysis of positives identified amplification for 2 NRPS, 88 type I PKS and 59 type II PKS. Many clones were positive for both type I and type II PKS, which could be due to the presence of hybrid or different fragmented pathways. Considering that type II PKS biosynthetic clusters tend to be smaller when compared to type I PKS, as they usually present a minimal PKS architecture of two ketosynthase units ($KS\alpha$ and $KS\beta$) and an acyl carrier protein (ACP), it is not unusual that they could be on the same DNA insert fragment with a partial type I PKS, and get identified by the multiplex technique.

Identification of pathways from positive clones

After the identification of positives, the next step was gene identification. As a way to dereplicate and obtain preliminary information about the pathways, the amplicons from the multiplex PCR were sent to Sanger sequencing. This way we could focus on unique pathways and prevent sequencing of inserts that corresponded to already described and characterized molecules. Unfortunately, the results were inconclusive for this assay, potentially due to the presense of multiple amplicons in each sample.

We decided to proceed to complete sequencing of the insert by fosmid preparation on large scale. Commercial kits and column centrifugation methods were avoided to maximize yields and mostly because they are not suitable for fosmid size. All positive hits were processed, inserts were digested from the fosmids and random samples were sent for sequencing for preliminary results. Figure 1.11 presents a PFGE of samples to recover digested inserts for sequencing. Table 1.5 shows sequencing results after BLAST alignment.

Table 1.5: BLAST alignments for fosmid insert sequences. To random library positive hits sent, the closest species match is shown, along with coverage and identity percentages.

Sample	Species	Coverage	Identity
1A T7	<i>Sulfurifustis variabilis</i> skN76	39%	76%
1G T7	<i>Streptomyces</i> sp. PBH53	100%	99%
2E R	<i>Streptomyces iranensis</i>	90%	84%
3C F	<i>Streptomyces</i> sp. PBH53	96%	99%
3C R	<i>Streptomyces</i> sp. PBH53	94%	99%
3G T7	<i>Streptomyces</i> sp. PBH53	100%	100%

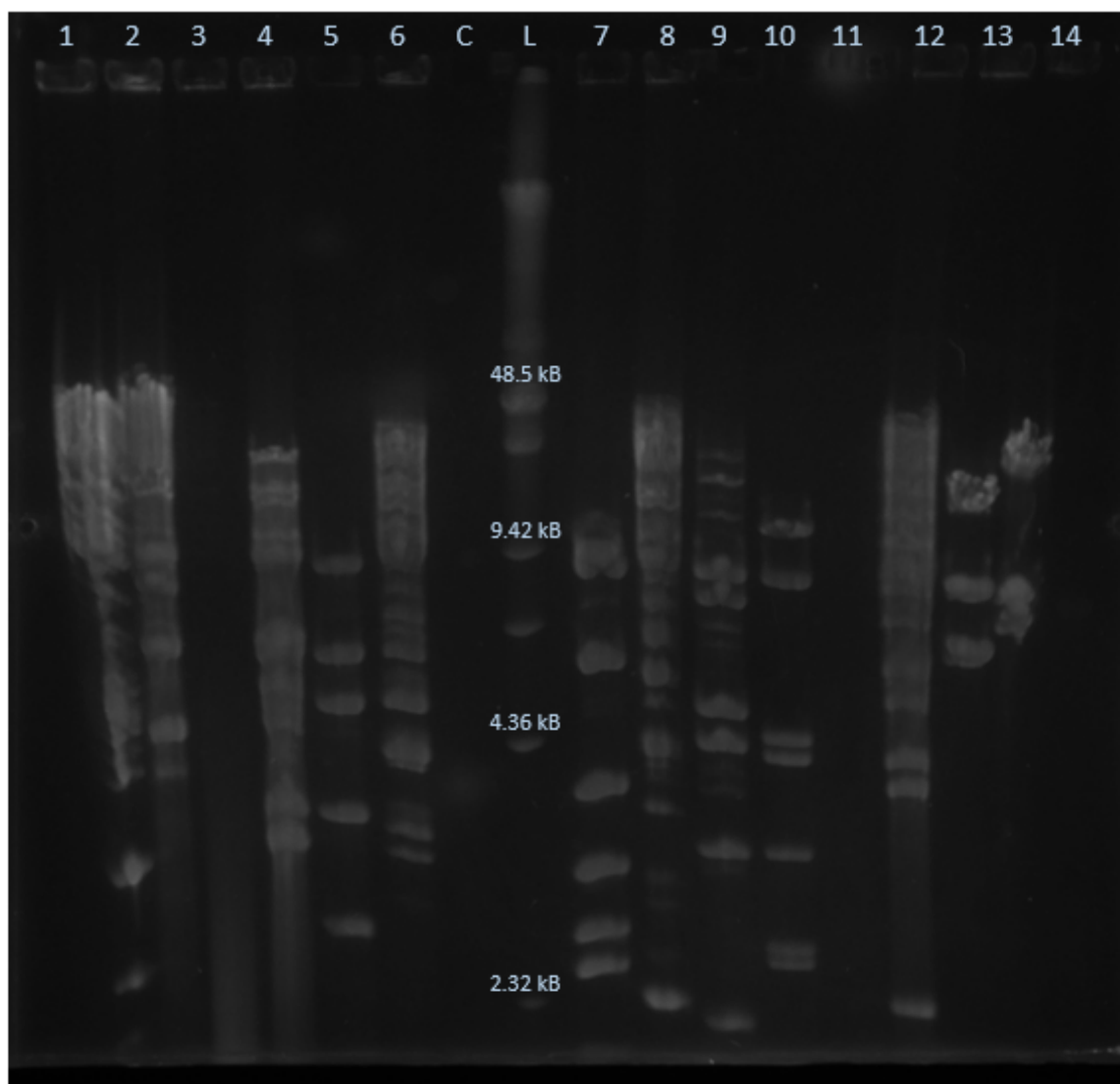


Figure 1.11: PFGE of digested positive fosmids. Double digestion performed with restriction enzymes HindIII and BamHI. 1 to 14: digested samples. C: fosmid control L: ladder (Low Range PFG Marker).

Observing the sequencing results, the majority of the sequences align with one isolate described by our group, *Streptomyces* sp. strain PBH53 (Gosse, J.T. et al., 2015). Considering that a substantial amount of our library is composed of samples from areas not distant from the original isolate collecting point, it is not unrealistic that we are dealing with a local strain.

The other hit obtained from BLAST analysis of the sequencing data, *Sulfurifustis variabilis*, does not possess high similarity. However looking at the strain information, it seems to be related to our fosmid insert as *S. variabilis* was isolated from sediment of a lake in Japan. The strain is able to oxidized thiosulfate, tetrathionate and elemental sulphur (Kojima, H., Shinohara, A., & Fukui, M., 2015). Onondaga Lake, one of the sources of our samples, is described as “the most polluted lake in America”, which could be a bearer of a similar organism. Some of the mechanisms of survival or pathways involved in adaptation to such harsh conditions could be shared by both strains, hence the similarity.

Fosmid sequencing using other methods was considered, but as this preliminary result pointed to an organism which sequence is known and secondary metabolites were predicted by our group, we decided to focus on this information and develop strategies to get a better understanding of some exciting pathways such as lantipeptides (covered on Chapter 2).

Conclusion

The goal of this project was to access novel gene clusters encoding secondary metabolites with potential for drug development through the collection and evaluation of environmental DNA from different urban soil samples, followed by the built and array of a fosmid library containing the genetic material.

A library of 1456 clones from five urban environments was successfully built and arrayed for the presence of type I and II polyketide synthase and nonribosomal peptide synthetase genes, yielding 103 positive individual hits. The majority of positive clones have hits for polyketide synthase genes, with a high prevalence of mixed type I and II PKS genes in the same clone.

Sequencing of positives revealed genetic similarity of fosmid clones with a recently described organism from a geographically close sampling site, which shows the organism distribution and adaptation to a specific urban area. Similarities between a metal-metabolizing lake organism and one of our library hits is a strong indicator of possession of pathways that help these strains adapt in anthropogenic environments.

We expected that this work contributes with the development of an optimized protocol to enable genetic screening of environmental samples, expanding the number of environments studied and the genetic potential of such places and demonstrating the importance of targeted sampling when bioprospecting.

Chapter Bibliography

Ayuso-Sacido, a., & Genilloud, O. (2005). New PCR Primers for the Screening of NRPS and PKS-I Systems in Actinomycetes: Detection and Distribution of These Biosynthetic Gene Sequences in Major Taxonomic Groups. *Microbial Ecology*, 49(1), 10–24.

Banik, J. J., & Brady, S. F. (2010). Recent application of metagenomic approaches toward the discovery of antimicrobials and other bioactive small molecules. *Current opinion in microbiology*, 13(5), 603-9.

Barka, E. A., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., ... van Wezel, G. P. (2016). Taxonomy, Physiology, and Natural Products of *Actinobacteria*. *Microbiology and Molecular Biology Reviews*, 80(1), 1 LP-43.

Bérdy, J. (2005). Bioactive Microbial Metabolites. *The Journal Of Antibiotics*, 58, 1.

Brady, S. F., Chao, C. J., & Clardy, J. (2002). New Natural Product Families from an Environmental DNA (eDNA) Gene Cluster. *Journal of the American Chemical Society*, 124(34), 9968–9969.

Brady, S. F., & Clardy, J. (2003). Synthesis of Long-Chain Fatty Acid Enol Esters Isolated from an Environmental DNA Clone. *Organic Letters*, 5(2), 121–124.

Bull, A.T. (2011). *Actinobacteria of the Extremobiosphere* (Vol. 2).
http://doi.org/10.1007/978-4-431-53898-1_58

Cruz-Martínez, K., Suttle, K. B., Brodie, E. L., Power, M. E., Andersen, G. L., & Banfield, J. F. (2009). Despite strong seasonal responses, soil microbial consortia are more resilient to long-term changes in rainfall than overlying grassland. *The Isme Journal*, 3, 738.

Cuadrat, R. R., Cury, J. C., & Dávila, A. M. (2015). Metagenomic Analysis of Upwelling-Affected Brazilian Coastal Seawater Reveals Sequence Domains of Type I PKS and Modular NRPS. *International journal of molecular sciences*, 16(12), 28285-95.

Epicentre Technologies Corp. (2012). CopyControl™ Fosmid Library Production Kit with pCC1FOSTM Vector [Brochure]. Retrieved from <http://www.epibio.com/docs/default-source/protocols/copycontrol-fosmid-library-production-kit-with-pcc1fos-vector.pdf?sfvrsn=6>

Felczykowska, A., Dydecka, A., Bohdanowicz, M., Gąsior, T., Soboň, M., Kobos, J., ... Węgrzyn, G. (2014). The use of fosmid metagenomic libraries in preliminary screening for various biological activities. *Microbial Cell Factories*, 13(1), 105.

Felnagle, E. A., Jackson, E. E., Chan, Y. A., Podevels, A. M., Berti, A. D., McMahon, M. D., & Thomas, M. G. (2008). Nonribosomal peptide synthetases involved in the

production of medically relevant natural products. *Molecular pharmaceutics*, 5(2), 191-211.

Ferrand, J., Patron, K., Legrand-Frossi, C., Frippiat, J.-P., Merlin, C., Alauzet, C., & Lozniewski, A. (2014). Comparison of seven methods for extraction of bacterial DNA from fecal and cecal samples of mice. *Journal of Microbiological Methods*, 105, 180–185.

Finking, R., & Marahiel, M. A. (2004). Biosynthesis of Nonribosomal Peptides. *Annual Review of Microbiology*, 58(1), 453–488.

Fischbach, M.A., Walsh, C.T. (2006). Assembly-line enzymology for polyketide and nonribosomal Peptide antibiotics: logic, machinery, and mechanisms. *Chemical Reviews* 106: 3468–3496.

Foerstner, K. U., Doerks, T., Creevey, C. J., Doerks, A., & Bork, P. (2008). A computational screen for type I polyketide synthases in metagenomics shotgun data. *PloS one*, 3(10), e3515.

Gosse, J. T., Hill, P., Dowd, S. E., & Boddy, C. N. (2015). Draft Genome Sequence of *Streptomyces* sp. Strain PBH53, Isolated from an Urban Environment. *Genome announcements*, 3(4), e00859-15.

Hårdeman, F., & Sjöling, S. (2007). Metagenomic approach for the isolation of a novel low-temperature-active lipase from uncultured bacteria of marine sediment. *FEMS Microbiology Ecology*, 59(2), 524–534.

Hertweck, C. (2009). The Biosynthetic Logic of Polyketide Diversity. *Angewandte Chemie International Edition*, 48: 4688-4716.

Hill, P., Krištůfek, V., Dijkhuizen, L., Boddy, C., Kroetsch, D., & van Elsas, J. D. (2011). Land use intensity controls actinobacterial community structure. *Microbial ecology*, 61(2), 286-302.

Janda, J. M., & Abbott, S. L. (2007). 16S rRNA gene sequencing for bacterial identification in the diagnostic laboratory: Pluses, perils, and pitfalls. *Journal of Clinical Microbiology*, 45(9), 2761–2764.

Jiang, Y. (2016). Isolation and Cultivation Methods of Actinobacteria. In Q. Li (Ed.) (p. Ch. 2). Rijeka: IntechOpen.

Kallifidas, D., Jiang, G., Ding, Y., & Luesch, H. (2018). Rational engineering of *Streptomyces albus* J1074 for the overexpression of secondary metabolite gene clusters. *Microbial Cell Factories*, 17(1), 25.

Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., & Thomson, J. M. (2014). The Impact of Different DNA Extraction Kits and

Laboratories upon the Assessment of Human Gut Microbiota Composition by 16S rRNA Gene Sequencing. *PLOS ONE*, 9(2), e88982.

Kojima, H., Shinohara, A., & Fukui, M. (2015). *Sulfurifustis variabilis* gen. nov., sp. nov., a sulfur oxidizer isolated from a lake, and proposal of Acidiferrobacteraceae fam. nov. and Acidiferrobacterales ord. nov. *International Journal of Systematic and Evolutionary Microbiology*, 65(10), 3709–3713.

Konz, D., & Marahiel, M. A. (1999). How do peptide synthetases generate structural diversity? *Chemistry & Biology*, 6(2), R39–R48.

Luna, G. M., Dell’Anno, A. and Danovaro, R. (2006) DNA extraction procedure: a critical issue for bacterial diversity assessment in marine sediments. *Environmental Microbiology*, 8: 308-320.

Marahiel, M. A. (2016). A structural model for multimodular NRPS assembly lines. *Natural Product Reports*, 33(2), 136–140.

Marahiel, M. A., Stachelhaus, T., & Mootz, H. D. (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews*, 97(7), 2651–2674.

Marchesi, J. R., Sato, T., Weightman, A. J., Martin, T. A., Fry, J. C., Hiom, S. J., & Wade, W. G. (1998). Design and evaluation of useful bacterium-specific PCR

primers that amplify genes coding for bacterial 16S rRNA. *Applied and Environmental Microbiology*, 64(2), 795–799.

Martin-Laurent, F., Philippot, L., Hallet, S., Chaussod, R., Germon, J. C., Soulas, G., & Catroux, G. (2001). DNA extraction from soils: old bias for new microbial diversity analysis methods. *Applied and environmental microbiology*, 67(5), 2354–9.

Müller, C. A., Obermeier, M. M., & Berg, G. (2016). Bioprospecting plant-associated microbiomes. *Journal of Biotechnology*, 235, 171–180.

Ochi, K., & Hosaka, T. (2013). New strategies for drug discovery: activation of silent or weakly expressed microbial gene clusters. *Applied Microbiology and Biotechnology*, 97(1), 87–98.

Owen, J. G., Reddy, B. V. B., Ternei, M. A., Charlop-Powers, Z., Calle, P. Y., Kim, J. H., & Brady, S. F. (2013). Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proceedings of the National Academy of Sciences*, 110(29), 11797–11802.

Rabausch, U., Juergensen, J., Ilmberger, N., Böhnke, S., Fischer, S., Schubach, B., ... Streit, W. R. (2013). Functional Screening of Metagenome and Genome Libraries for Detection of Novel Flavonoid-Modifying Enzymes. *Applied and Environmental Microbiology*, 79(15), 4551 LP-4563.

Robbins, T., Liu, Y. C., Cane, D. E., & Khosla, C. (2016). Structure and mechanism of assembly line polyketide synthases. *Current opinion in structural biology*, 41, 10-18.

Saslis-Lagoudakis, C. H., Savolainen, V., Williamson, E. M., Forest, F., Wagstaff, S. J., Baral, S. R., ... Hawkins, J. A. (2012). Phylogenies reveal predictive power of traditional medicine in bioprospecting. *Proceedings of the National Academy of Sciences*, 109(39), 15835 LP-15840.

Schmitz, J. E., Daniel, A., Collin, M., Schuch, R., & Fischetti, V. A. (2008). Rapid DNA Library Construction for Functional Genomic and Metagenomic Screening. *Applied and Environmental Microbiology*, 74(5), 1649 LP-1652.

Schwartz, D. C., & Cantor, C. R. (1984). Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell*, 37(1), 67–75.

Singh, M.P.; Janso, J.J.; Brady, S.F. (2007). Cytoskyrins and Cytosporones Produced by *Cytospora* sp. CR200: Taxonomy, Fermentation and Biological Activities. *Marine Drugs*, 5, 71–84.

Stulberg, E. R., Lozano, G. L., Morin, J. B., Park, H., Baraban, E. G., Mlot, C., Heffelfinger, C., Phillips, G. M., Rush, J. S., Phillips, A. J., Broderick, N. A., Thomas, M. G., Stabb, E. V., ... Handelsman, J. (2016). Genomic and Secondary Metabolite

Analyses of *Streptomyces* sp. 2AW Provide Insight into the Evolution of the Cycloheximide Pathway. *Frontiers in microbiology*, 7, 573.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G. F., Chater, K. F., & van Sinderen, D. (2007). Genomics of Actinobacteria: tracing the evolutionary history of an ancient phylum. *Microbiology and molecular biology reviews : MMBR*, 71(3), 495-548.

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

Wawrik, B., Kerkhof, L., Zylstra, G. J., & Kukor, J. J. (2005). Identification of Unique Type II Polyketide Synthase Genes in Soil, 71(5), 2232–2238.

Weissman, K. J. (2016). Genetic engineering of modular PKSs: from combinatorial biosynthesis to synthetic biology. *Natural Product Reports*, 33(2), 203–230.

Wong, F.T, Khosla, C. (2012). Combinatorial biosynthesis of polyketides - a perspective. *Current Opinion in Chemical Biology*, 16:117–123

Yang, J. K., Liang, J. F., Xiao, L. M., Yang, Y., & Chao, Q. F. (2018). Vertical stratification of bacteria and the chemical compounds in crude oil-contaminated soil layers of the semi-deserted Dzungharian Basin. *PloS one*, 13(9), e0203919.

Zhang, H., Wang, Y., Wu, J., Skalina, K., & Pfeifer, B. A. (2010). Complete Biosynthesis of Erythromycin A and Designed Analogs Using *E. coli* as a Heterologous Host. *Chemistry & Biology*, 17(11), 1232–1240.

Chapter 2: Heterologous expression of a lantipeptide from a bacterial isolate

Introduction

RiPPs, not your average protein

The acronym RiPPs stands for ribosomally synthesized and posttranslationally modified peptides. The main difference from RiPPs to nonribosomal peptides is the absence of dedicated machinery, with native ribosomal translation playing the peptide synthesis role.

Most RiPPs are biosynthesized from a larger precursor that can typically be demarcated into an amino-terminal leader sequence (leader peptide) that is removed to liberate the final bioactive product, and a carboxy terminal core peptide that is posttranslationally modified into the active compound (Yu, Y. et al., 2013; Letzel et al., 2014). The final product is a usually small molecule made up of a couple of residues to up to 40-50 amino acids with several modifications (Link, A. J., 2015).

The diversity of RiPPs is shown by the presence of numerous families of compounds. The variety of structural features yielded from modifications allows a range of activities such as antifungal, antibacterial, allelopathic, and antiviral (Essig et al., 2014, Férir et al., 2013, Mohr et al., 2015, Zheng et al., 2015). Due to their versatility, RiPPs are undoubtedly another valuable source of novel metabolites to be explored.

You would never break the (thioether) chain – Lantipeptides

Lanthionine-containing peptides, also known as lantipeptides, are polycyclic peptides originated from a linear precursor peptide (generically termed LanA) and characterized by the presence of thioether cross-linked amino acids lanthionine (Lan) and methyllanthionine (MeLan), which carry an additional methyl group on one of the β - carbons.

Biosynthetically, lanthionines and methyllanthionines originate from Ser and Thr residues that are first dehydrated to generate dehydroalanine (Dha) and dehydrobutyrine (Dhb) residues respectively (Knerr, P.J., van der Donk W.A., 2012). Subsequently, the thiol of a Cys is added across the carbon-carbon double bond of these dehydroamino acids in a Michael-type addition to produce the Lan and MeLan structures, respectively (Arnison, P. G. et al., 2013).

Lantipeptides can be divided into four groups according to the nature of the enzymes responsible for MeLan catalysis. Type I lantibiotics presents two enzymes: LanB, the lanthionine dehydratase which catalyses the dehydration of the amino acids, and LanC, the lanthionine synthetase, which catalyzes the thioether formation. Type II lantibiotics contain a single enzyme which performs dehydration and thioether formation. Type III and IV are lantipeptides which are catalyzed by distinct enzymes such as the RamC-like and LanL enzymes (Marsh, A.J. et al., 2010).

Following the dehydration and thioether linkage formation, the peptide is not yet bioactive. To allow activation, the amino-terminal leader peptide needs to be removed. In lantipeptides, this occurs in different manners depending on the molecule type. Usually in type I lantipeptides the removal of the amino-

terminal leader peptide is carried out by a subtilisin- like serine protease termed LanP (van der Meer, J.R. et al., 1993). Type II leader peptide removal occurs concomitant with the transport of the modified precursor peptide, and is catalyzed by the papain-like Cys protease domain of LanT, and in type III cleavage is performed by prolyl oligopeptidases (van der Donk, W. A., & Nair, S. K., 2014). Figure 2.1 shows an example of lantipeptide synthesis using the pathway for nisin, a type I lantipeptide.

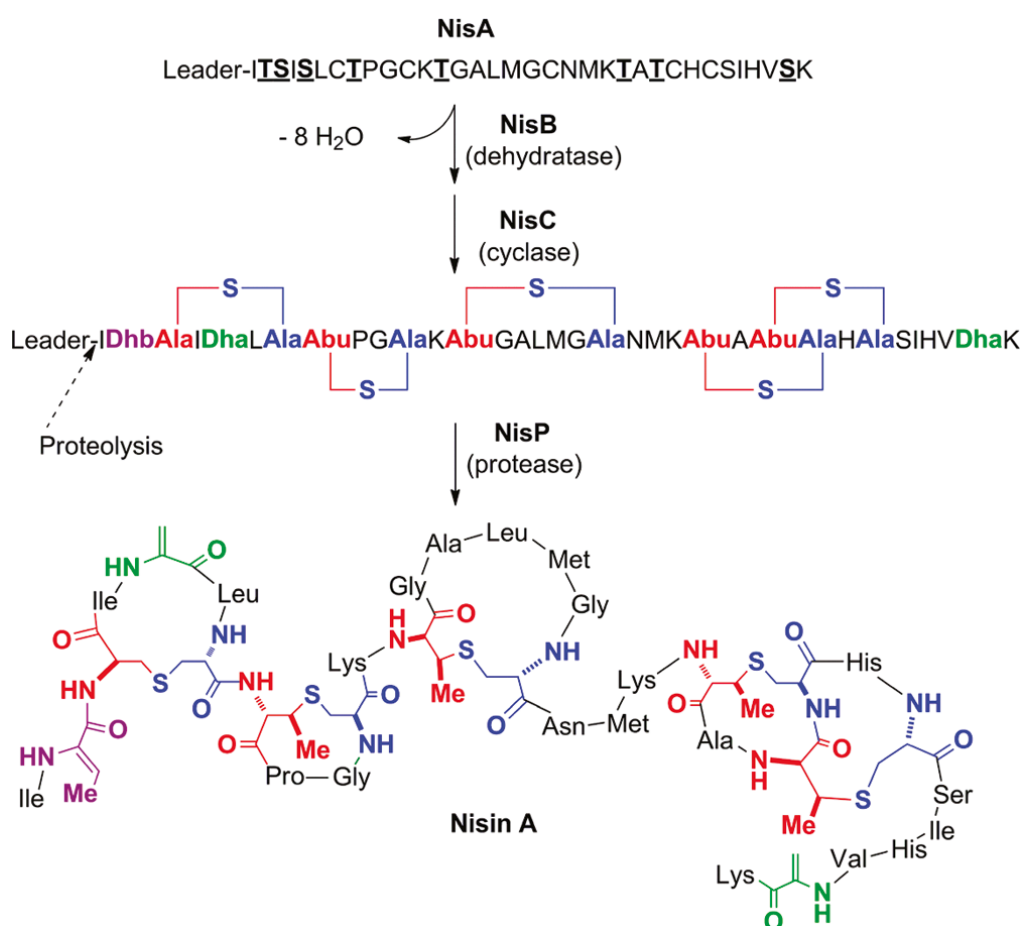


Figure 2.1: Nisin A biosynthesis. Steps shown: NisA linear peptide formation, dehydration carried by NisB, thioether linkage formation by NisC and activation of peptide by cleavage performed by NisP. Retrieved from Shi, Y. et al., *J. Am. Chem. Soc.* 2011, 133, 2338–2341, by permission.

Lantipeptides possess remarkable antimicrobial activity against many clinically relevant gram-positive bacteria, including drug-resistant strains of *Staphylococcus*, *Streptococcus*, *Enterococcus*, and *Clostridium*, as well as against select gram-negative pathogens, such as *Neisseria* and *Helicobacter* (Knerr, P.J., van der Donk, W.A., 2012). This property puts these bioactive molecules as one of our target compounds for bioprospecting.

Strategies for accessing lantipeptides and other RiPPs

As RiPPs have proven to be exciting compounds for a variety of applications, one of the challenges is the development of strategies to identify and synthesize significant amounts of product for testing and further usage. A deeper understanding of their biosynthesis and the use of bioinformatics tools to accurately identify associated genetic components have accelerated the development of new *in vivo* and *in vitro* engineering strategies.

The engineering of bacterial strains to serve as heterologous hosts is currently the protocol for lantipeptide production, with special focus in the use of *E. coli*, which simplifies and generalizes genetic manipulation. In 2005, the first report of lantipeptide biosynthesis in *E. coli* described the production of nukacin through single-vector coexpression (Nagao, J. et al., 2005).

Six years later, in 2011, Shi and colleagues described the preparation of seven different lantipeptides in *Escherichia coli* and demonstrate a methodology to incorporate nonproteinogenic amino acids. The genes for the precursor peptides and the synthetase were cloned as a bicistronic construct, with a N-

terminal hexahistidine tag to allow column purification. Artificial cleavage sites were introduced between the leader and core peptides to *allow in vitro* removal of the leader sequence with the help of introduced proteases (Shi, Y. et al., 2011).

The work performed by Karakas-Sen and Narbad (2012) reports the overexpression of a nisin precursor peptide, prenisin, to be used in vitro investigation of dehydration function of NisB protein. The researchers developed a C-terminally histidine-tagged (6x His) fusion peptide in *Escherichia coli* that enabled purification of His-tagged prenisin using nickel affinity chromatography. The purified prenisin was analyzed by SDS-PAGE, Western blotting and mass spectroscopy, showing the successful heterologous expression using an *E. coli* system (Karakas-Sen, A. and Narbad, A., 2012).

One last example is the work of Hoffman et al. (2012) to identify functional artificial lantipeptides. More than 10^{11} different lanthionine containing peptides were prepared in an *in vitro* translation approach coupled with an enzyme-free protocol for posttranslational modification. The success of production was validated towards Sortase A, a transamidase required for virulence of *Staphylococcus aureus*. The interaction of selected lantipeptides with Sortase A is highly dependent on the presence of lanthionine in the peptide and an active conformation of the protein (Hofmann, F. T., Szostak, J. W., & Seebeck, F. P., 2012).

All the recent successful examples to standardize heterologous expression of lantipeptides using *E. coli* heterologous hosts show that this is a promising methodology and a reasonable starting point for heterologous expression of new lantipeptides. Our efforts should be focused on the challenges of tailoring these compounds.

Project objectives

This project focuses on heterologous expression of a lantipeptide from a bacterial isolate from an urban environment. Using *E. coli* as the expression host and an inducible T7 RNA Polymerase-mediated Multigene Expression System (MES) developed in-house, named pMGX, the different components of the biosynthetic pathway were cloned and evaluated for the production of the propeptide and its active form.

This research aims to expand the knowledge on RiPPs and their presence in different bacterial species and their biological potential, in particular, the ones with antimicrobial potential that could be explored. The development of heterologous expression protocols and the determination of the role of different genes in the molecule biosynthesis is a remarkable effort to expand the studies of this group of poorly known molecules.

Materials and Methods

Household isolates library

In order to complement the urban bacterial isolate collection the lab possesses, a household dust isolate library was generated. Two samples from vacuum cleaner bags belonging to three distinct households plus the contents of an air filter from another location were chosen for analysis.

The vacuum cleaner content was divided into fine, or small particle, and heavy dust, or mixed big particles and hair. 2-3 g of samples was weighted and washed using 50 mL of Tris-EDTA buffer (50 mM Tris, 50 mM EDTA). The supernatant was serial diluted to yield better isolation. 100 μ L of a series of five 1:10 dilutions plus the undiluted sample were plated on ISP-2 (4 g yeast extract, 10 g malt extract, 4 g dextrose, 20 g agar). The medium is specific for *Streptomyces* but allows general bacterial growth. Plates were cultivated for 2 weeks at 30°C with daily observation and collection of new colonies.

New colonies that appeared on plates over the course of the 2-week incubation were isolated by collection on ISP-2 plates and left to grow under the same initial isolation. Permanent stocks of the isolated organisms were prepared with the incubation of selected colonies in 5 mL of LB broth overnight. 500 μ L of the bacterial culture was mixed with the same volume of a 50% v/v glycerol solution and frozen at -80°C. In total 34 colonies were isolated using this method, giving a total of 148 isolates for the collection.

Table 2.1: List of prospected urban isolates from the laboratory collection

Name	Working name	Site collected
PBH1	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH2	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH3	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH4	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH5	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH6	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH7	alley (A)	Alley between Aude Dubliner and Scottish pub, Market, Ottawa
PBH8	Pavement (P)	
PBH9	Pavement (P)	
PBH10	Pavement (P)	
PBH11	Pavement (P)	
PBH12	Pavement (P)	
PBH13	Hcave	
PBH14	Hcave	
PBH15	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH16	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH17	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH18	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH19	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH20	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH21	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH22	Lake	Bowman lake, New York, USA. about 2-3 feet deep
PBH23	Billings Bridge Clay	
PBH24	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH25	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH26	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH27	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH28	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway

PBH29	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH30	Billings Bridge Clay	Billings Bridge estate forest, on rim of valley overlooking transitway
PBH31	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH32	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH33	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH34	Billings Bridge Sand	
PBH35	Billings Bridge Sand	
PBH36	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH37	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH38	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH39	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH40	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH41	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH42	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH43	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH44	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH45	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH46	Billings Bridge Sand	Billings Bridge estate forest, above Sawmill creek on east side
PBH47	Folking alley	
PBH48	Folking alley	
PBH49	Campus	Transitway bus stop Campus
PBH50	Hot Pavement	
PBH51	Hot Pavement	
PBH52	Hot Folking alley	Folking alley, Groningen, The Netherlands
PBH53	Hot Campus	Transitway bus stop Campus
PBH54	Hot Pavement	
PBH55	Hot Pavement	
PBH55	Oxford (Ox)	Oxford street sample, United Kingdom
PBH56	Oxford (Ox)	Oxford street sample, United Kingdom
PBH57	Oxford (Ox)	Oxford street sample, United Kingdom
PBH58	Oxford (Ox)	Oxford street sample, United Kingdom
PBH59	Oxford (Ox)	Oxford street sample, United Kingdom

PBH60	Oxford (Ox)	Oxford street sample, United Kingdom
PBH61	Oxford (Ox)	Oxford street sample, United Kingdom
PBH62	Oxford (Ox)	Oxford street sample, United Kingdom
PBH63	Oxford (Ox)	Oxford street sample, United Kingdom
PBH64	Oxford (Ox)	Oxford street sample, United Kingdom
PBH65	Oxford (Ox)	Oxford street sample, United Kingdom
PBH66	Oxford (Ox)	Oxford street sample, United Kingdom
PBH67	Oxford (Ox)	Oxford street sample, United Kingdom
PBH68	Oxford (Ox)	Oxford street sample, United Kingdom
PBH69	Oxford (Ox)	Oxford street sample, United Kingdom
PBH70	Oxford (Ox)	Oxford street sample, United Kingdom
PBH71	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH72	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH73	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH74	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH75	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH76	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH77	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH78	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH79	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH80	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH81	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH82	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH83	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH84	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH85	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH86	Worthing (Wor)	Durrington Trainstation steps, United Kingdom
PBH87	Hot Billings Bridge Sand	
PBH88	Hot Billings Bridge Sand	
PBH89	Hot Billings Bridge Sand	
PBH90	Hot Billings Bridge Clay	
PBH91	Hot Billings Bridge Sand	

PBH92	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH93	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH94	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH95	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH96	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH97	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH98	Night Air	Plates left open to air at campus bus stop around 10 minutes.
PBH99	Arctic Upland	
PBH100	Arctic Upland	
PBH101	Arctic Upland	
PBH102	Arctic Upland	
PBH103	Arctic Upland	
PBH104	Arctic Upland	
PBH105	Arctic Plateau	
PBH106	Arctic Plateau	
PBH107	Arctic Plateau	
PBH108	Arctic Plateau	
PBH109	Arctic Plateau	
PBH110	Arctic Plateau	
PBH111	Arctic Plateau	
PBH112	Arctic Plateau	
PBH113	Arctic Plateau	
PBH114	Arctic Plateau	

Table 2.2: Household isolates library constructed to complement lab isolate collection

Name	Source	Aspect on ISP2
JTG1	80 Fulton	Yellow, round, small, firm
JTG2	30 Commanche	Green/yellow/orange, fluffy
JTG3	30 Commanche	Yellow, round, small, firm
JTG4	190 Lees ave. 1104	White, round, small/medium, glossy
JTG5	190 Lees ave. 1104	Offwhite, irregular, small
JTG6	190 Lees ave. 1104	Yellow, round, small, firm
JTG7	190 Lees ave. 1104	Pale white, milky, dry (LB)
JTG8	190 Lees ave. 1104	White, small, glossy
JTG9	190 Lees ave. 1104	White, turns to pink as days pass (~5 days),round, glossy
JTG10	80 Fulton	White, starfish shape, hairy but plain motifs
JTG11	190 Lees ave. 1104	White, medium to big, glossy
JTG12	30 Commanche	White, big, glossy
JTG13	30 Commanche	Beige, round colonies with swirls, hairy
JTG14	80 Fulton	Green/white, small, fluffy (velvet-like) (LB)
JTG15	190 Lees ave. 1104	Yellow, round, small, firm
JTG16	30 Commanche	
JTG17	190 Lees ave. 1104	
JTG18	30 Commanche	White, medium to big, glossy
JTG19	Missing	
JTG20	190 Lees ave. 1104	White, non glossy
JTG21	190 Lees ave. 1104	White, looks like velvet
JTG22	190 Lees ave. 1104	White, small, glossy
JTG23	190 Lees ave. 1104	White, small
JTG24	190 Lees ave. 1104	Pink, small/medium,glossy
JTG25	190 Lees ave. 1104	White, parrafin drops aspect
JTG26	190 Lees ave. 1104	White, target-like, glossy
JTG27	190 Lees ave. 1104	White, glossy
JTG28	190 Lees ave. 1104	White, puffy, irregular, non glossy

JTG29	190 Lees ave. 1104	White in the middle of the colony, rounded, glossy
JTG30	190 Lees ave. 1104	White, small
JTG31	190 Lees ave. 1104	White, small
JTG32	190 Lees ave. 1104	White, milky
JTG33	190 Lees ave. 1104	White, velvet-like
JTG34	190 Lees ave. 1104	White, small, glossy

16S identification of isolates

For a preliminary identification of isolates, a PCR targeting the 16S bacterial ribosomal RNA sequence was performed. This housekeeping gene enables the determination of identity due to its widespread presence and to a well-conserved structure (Janda & Abbott, 2007).

Bacterial genetic material was used as a template for PCR after its isolation with the help of a kit (Wizard Genomic DNA Purification Kit, Promega). The 25 µL reactions were prepared containing 100ng DNA template, 0.2mM of dNTP mix, 0.2µM of each primer, 1U Taq DNA polymerase (NEB) with corresponding 10X reaction buffer, 1% 10 mg/ml BSA. The reaction had an initial denaturation step at 95°C and 30 cycles consisting of 95°C for 1 minute, 55°C for 1 minute, and 72°C for 90 seconds followed by a final extension step of 5 minutes at 72°C. Primers used for the amplification were 63f (5'-CAGGCCTAACACATGCAAGTC-3') and 1387r (5'-GGGCGGWTGTACAAGC-3') (Marchesi et al., 1998).

Samples were resolved in 1% w/v agarose gels. Amplicons of 1.3 kilobases were extracted from gel using an isolation kit (E.Z.N.A Gel Extraction

Kit, Omega-Biotek) and sent for sequencing using the services of McGill University and Génome Québec Innovation Centre.

Whole genome sequencing (WGS) of isolates and secondary metabolite evaluation

From the isolate library, two organisms were picked for whole genome sequencing: isolates PBH53 and JTG09. Both strains were cultivated on 5 mL LB broth for 72 hours and had their DNA extracted using a kit (Wizard Genomic DNA Purification Kit, Promega). The yielded material was submitted for sequencing through Mr. DNA - Molecular Research LP (Texas, USA). Sequencing details: Illumina MiSeq, 50x2x300 bp paired-end).

Presence of pathways encoding secondary metabolites array was performed as a multiplex PCR, as described in Chapter 1 Material and Methods.

Lantipeptide primer design and amplification

To amplify PBH53 lantipeptide pathway, three sets of primers were constructed to tackle the three genes composing the pathway. To enable recovery and expression evaluation, the C-terminus of proteins were 6XHis-tagged, with the exception of LanA that was labelled with a N-terminus tag. The primer sequences are expressed in Table 2.3

Table 2.3: primer sequences used for amplification of lantipeptide pathway

Primer	Sequence
LanAR	5'-GAATTCTTACGAGGTGACGCAGGC-3'
LanAF	5'- CATATGCATCATCACCATCACCACGGTGGTGGTATGACCGCCATTCCCGTGGAG -3'
LanAF	5'- CATATGCATCATCACCATCACCACGGTGGTGGTATGACCGCCATTCCCGTGGAG -3'
LanBF	5'- CATATGTGGGCCGACGAGACCTTCCGGCAGGCCG -3'
LanBR	5'- GAATTCTTAGTGGTGATGGTGATGATGACCACCACCTGCGCGCGCCGCA -3'
LanCF	5'- CATATGACCGCCACGACGCTCCCGCCCCG -3'
LanCR	5'- GAATTCTTAGTGGTGATGGTGATGATGACCACCACCGGTGATCAGCAGCACGTC -3'

Amplification conditions for production of fragments suitable for cloning applications follow: for a total volume of 50 μ L reaction were combined 200 ng DNA template, 0.2 mM of dNTP mix, 0.25 μ M of each primer, 8 % DMSO, 0.5 μ L (lanA) or 1 μ L (lanB and lanC) Herculase II DNA Polymerase (Agilent) with corresponding 10 X reaction buffer. Amplification started with a denaturation step at 95 °C and 35 cycles consisting of 95 °C for 1 minute, 65 °C for 30 seconds (lanA)/64 °C for 1 minute/61°C for 1 minute (lanC), and 72 °C for 1 minute followed by a final extension step of 3 minutes at 72 °C.

Fragments amplified were resolved in 1 % w/v agarose gels and extracted from gel using an isolation kit (E.Z.N.A Gel Extraction Kit, Omega-Biotek).

Cloning of lantipeptide pathway

1. Blunt cloning of pathway components

All fragments yielded from the amplification of different genetic components of the lantipeptide pathway were first cloned into blunt plasmids. A kit was used to provide the vector and reagents (Zero Blunt PCR Cloning Kit, Invitrogen Thermo Fisher).

For a 10 μ L reaction were combined 1.5 μ g DNA amplicon input, 1 μ L blunt vector, 5 U of T4 ligase along with 2 μ L 5X blunt buffer. The reaction was carried for 1 hour at room temperature and stored at -20 °C for use in further steps.

2. Expression vector cloning

For the final cloning of genes into expression vectors, the genes were first digested off the blunt vectors using EcoRI and NdeI. The 50 μ L reactions were prepared combining 1.5 μ g of blunt vectors with 10 units of both EcoRI and NdeI enzymes (NEB) and 5 μ L of 3.1 Buffer (NEB, provided with enzymes). Digestion was performed at 37 °C for 1 hour and 30 minutes. Enzyme activity was interrupted with incubation at 65 °C for 20 minutes.

Vectors pMGX-K (lanB and lanC) and pET21 (lanA) were prepared for ligation by being linearized with EcoRI and NdeI. Digestion conditions for the vectors were the same pursued for the inserts. Ligation of fragments happened with the combination of both linearized vectors with their respective fragments, following a ratio of 1 part of vector for 3 parts of fragment. These were mixed

with 1 μ L T4 DNA polymerase (NEB) and 2 μ L of 10 X provided buffer to yield a 10 μ L reaction. The tubes were incubated at 16 °C for 16 hours and heat inactivation of the enzyme occurred at 65 °C for 20 minutes.

3. Construction of the coexpression vector pLanBC

The two tailoring genes of the pathway, lanB and lanC were cloned into the same vector, generating a single coexpression system. For the sake of effectiveness of the process, lanC was the gene of choice to be inserted into the vector already containing the gene lanB.

Vector pMGX-K containing the fragment lanB was linearized in a single digestion process with AvrII. Digestion of lanC was performed with XbaI and AvrII. 50 μ L digestion reactions were prepared combining 1.5 μ g of vectors with 10 units of the respective enzymes (NEB) and 5 μ L of Cutsmart Buffer (NEB, provided with enzymes). Digestion was performed at 37 °C for 1 hour and 30 minutes. Enzyme activity was interrupted with incubation at 65 °C for 20 minutes.

Vectors were dephosphorylated to prevent recircularization, which could lead to false positives. To the digested plasmids, 1.5 μ L of Calf Intestinal Phosphatase (CIP) (NEB) was added and reaction was incubated at 37°C for 30 minutes.

Reactions were monitored for quality by running a 1 % w/v agarose gel. Fragments were extracted using an isolation kit (E.Z.N.A Gel Extraction Kit, Omega-Biotek).

Ligation of fragments happened with the combination of linearized vector

with LanC DNA, fragment following a ratio of 1 part of vector for 3 parts of fragment. These were mixed with 1 μ L T4 DNA polymerase (NEB) and 2 μ L of 10 X provided buffer to yield a 10 μ L reaction. The tubes were incubated at 16 °C for 16 hours and heat inactivation of the enzyme occurred at 65° C for 20 minutes.

4. Transformation of vectors in *Escherichia coli* strains

Vectors constructed throughout the cloning steps were transformed into *E. coli* strains for propagation and protein expression. Strains Top 10 and BL21 chemically competent prepared in the laboratory were used.

100 μ L of the respective competent cells was transferred to a 1.5 mL tube and topped with 50 ng of the desired plasmid. The mixture was incubated on ice for 30 minutes. Cells were heat shocked using a water bath set to 42 °C for 45 seconds. The tubes were put back on ice for 2 minutes. To the cells, 300 μ L of LB broth was added, proceeded by an incubation of 1 hour at 37 °C and 200 rpm.

100 μ L of the incubated transformed cells was spread on LB agar plates containing the antibiotic required for selection based on the vector resistance marker. Plates were incubated for 16 hours at 37 °C and colonies were harvested.

5. Protein expression and evaluation

E. coli BL21 transformants were used to express proteins for the lantipeptide system. Overnight cultures of harvested colonies were prepared and 4 mL of the broth was added to 400 mL of LB broth, followed by an incubation at 37 °C and

200rpm. Aliquots of the broth were taken to determine the optical density (OD) of the culture, which was aimed to reach values between 0.4 to 0.6 at 600 nm.

After the fermented broth reached the desired OD, 400 μ L of 1 M IPTG was added to induce protein expression, and the broth was incubated at 30 °C for 5 to 6 hours or 16° C overnight.

To extract the expressed protein, the culture was spun at 4000 rpm at 4 °C for 40 minutes. The pellets were resuspended using lysis buffer (100 mM sodium phosphate, 300 mM NaCl, 10% v/v glycerol, 1 mg/mL lysozyme, 1 μ g/mL pepstatin A, 1-2 μ g/mL leupeptin, pH 8.0, 0.2 μ l/mL culture). Cells were lysed using a sonicator, in two cycles of 30 seconds each. The solution was spun again to pelletize cell debris at 9000 rpm, 4 °C for 1 hour.

The cell lysate was collected and nickel resin (HisPur Ni-NTA Resin, Thermo Fisher) and a solution of 250 mM imidazole were added to a proportion of 5:5:4 lysate: nickel resin: imidazole. This was followed by a 2-hour incubation period at 4 °C with agitation.

The protein mixture was separated using a column to hold the liquid and to allow fractionation of the protein. The resin was packed with the help of gravity and the flow-through was collected using air pumped into the system to push the fraction down the column. Resin was washed using a set of imidazole solutions in 1 M Tris/5 M NaCl of different concentrations: 0 mM, 20 mM, 100 mM and 250 mM. For the first two solutions, volume used corresponded to 10 X the sample volume and for the remaining concentrations the volume was 5 X. Each fraction was collected in a separate tube and kept on ice for further analysis.

Evaluation of protein expression was performed using SDS-PAGE gel electrophoresis of fractions. 20 μ L of each fraction and the flow-through were

diluted on the same volume of 5 % 2-mercaptoethanol Laemmli buffer and denatured at 85 °C for 5 minutes. Room temperature samples were loaded on gel and run at 220 V for 35 minutes. Gel was activated using UV (302 nm) and bands were visualized after 30 seconds of exposure.

6. Cell free expression

Cell free protein expression is a technique used to bypass cell-dependant expression systems, which could lead to absence of expression, degradation of protein by native proteases present in cells or death of expression host due to toxicity of the expressed protein. Without the use of the entire cell machinery, focusing on the use of translation apparatus only, the expression and stability of the desired protein can be ensured.

LanA, the protein of interest to be cloned using the cell free system, was amplified through PCR, as described above, and ligated into a blunt vector and later into pET21. All the procedures for ligation and digestion followed were the same as the ones used for the traditional cloning approach presented on sections 1-4 of this chapter.

The pET21-LanA vectors were purified to eliminate residues of RNase and salts from the plasmid extraction kits. This step is required to improve quality of translation of the cell free system. One volume of phenol:chloroform:isoamyl alcohol (25:24:1) was added to the vector solution, followed by vigorous shake for 20 seconds and centrifugation at 16,000 g for 5 minutes. The upper part containing the DNA was removed and transferred to a tube. To the solution, 2.5 volumes of 100 % ethanol was added and the solution was left overnight at -20°C

to precipitate DNA. The solution was centrifuged at 16,000 g for 10 minutes to pelletize the genetic material. The pellet was washed with chilled 70% ethanol and centrifuged for 3 minutes. It was left to air dry for 20 minutes and later resuspended in 20 μ L of TE buffer.

The cell free expression was performed with the use of a kit (PURExpress In Vitro Protein Synthesis kit, NEB). A 25 μ L reaction was set up, containing 10 μ L of Solution A, 7.5 μ L of solution B, 0.5 μ L of labeled lysine (FluoroTect GreenLys In Vitro Translation Labeling System, Promega) and 250 ng of DNA template. The reaction was left for 4 hours at 37 °C and interrupted by placing the tubes on ice.

Samples were resolved through SDS-PAGE gel electrophoresis. 10 μ L of each sample was diluted on the same volume of 5 % 2-mercaptoethanol Laemmli buffer and denatured at 85 °C for 5 minutes. Cooled samples were loaded on gel and run at 220 V for 35 minutes. Bands were visualized with the use of 488 nm Ar-ion laser equipment (AlphaInnotech ChemiImager 5500, AlphaInnotech, USA). Coomassie Blue R-250 (ThermoFisher Scientific, USA) staining procedures were carried when necessary as per manufacturer protocol.

Results and Discussion

Isolate library evaluation

Our lab collection of isolates was probed for the presence of secondary metabolite pathways belonging to type I and II PKS and NRPS families. The same set of primers used to perform multiplex eDNA library array was applied here.

A number of samples tested positive for biosynthetic genes, however two strains caught our immediate attention due to presence of all pathways types probed: PBH53 and JTG09. Figure 2.2 shows the agarose gel resolution of a group of samples, including PBH53 amplicon for one of the NP families probed. These isolates were selected for whole genome sequencing.

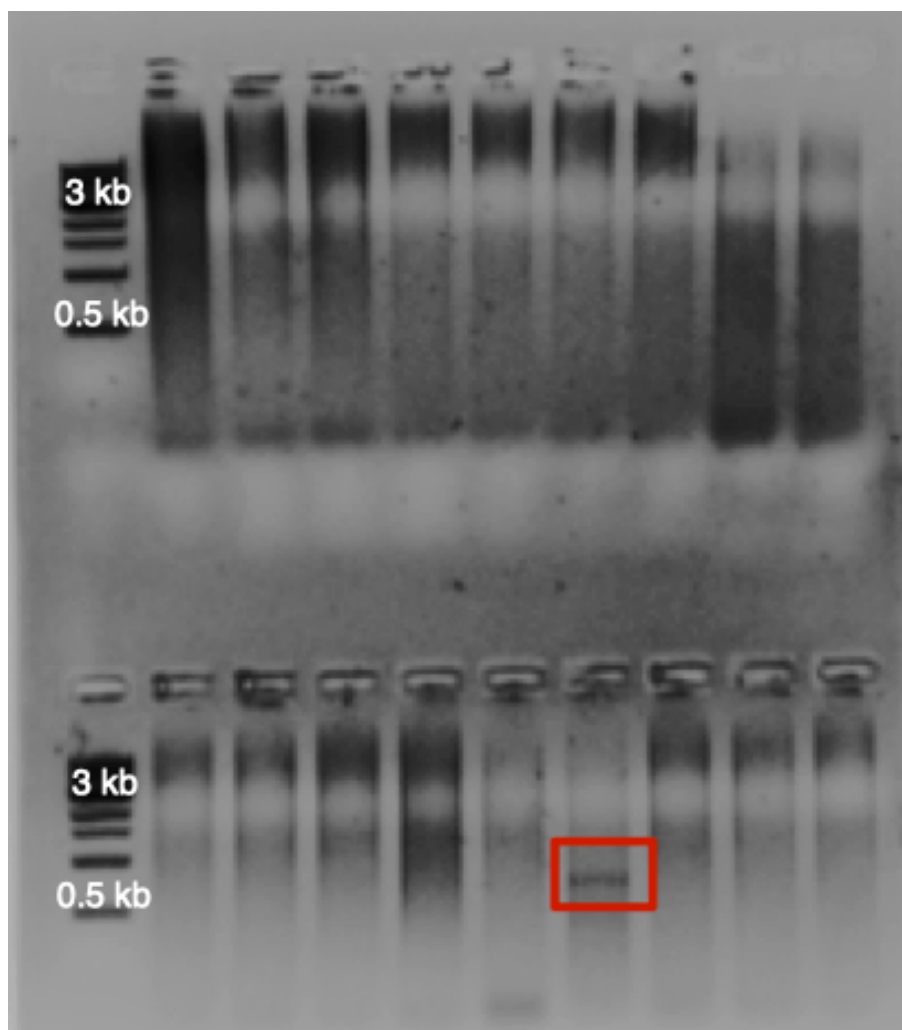


Figure 2.2: Agarose gel of amplification of NRPS pathways from isolates library. Lane 1: 1kb DNA ladder (NEB). Lanes 2-10: samples. In detail: NRPS amplicon on PBH53 isolate.

Pathway prediction through antiSMASH

The WGS of PBH53 generated a draft genome deposit on NCBI database, (accession number CP011799.1) which was analyzed for the presence of secondary metabolites with the help of the bioinformatics tool antiSMASH v. 3.0.0 (Weber T. *et al.*, 2015). In total, 47 secondary metabolite biosynthetic gene

clusters were identified, with the majority predicted to be involved in polyketide or nonribosomal peptide biosynthesis.

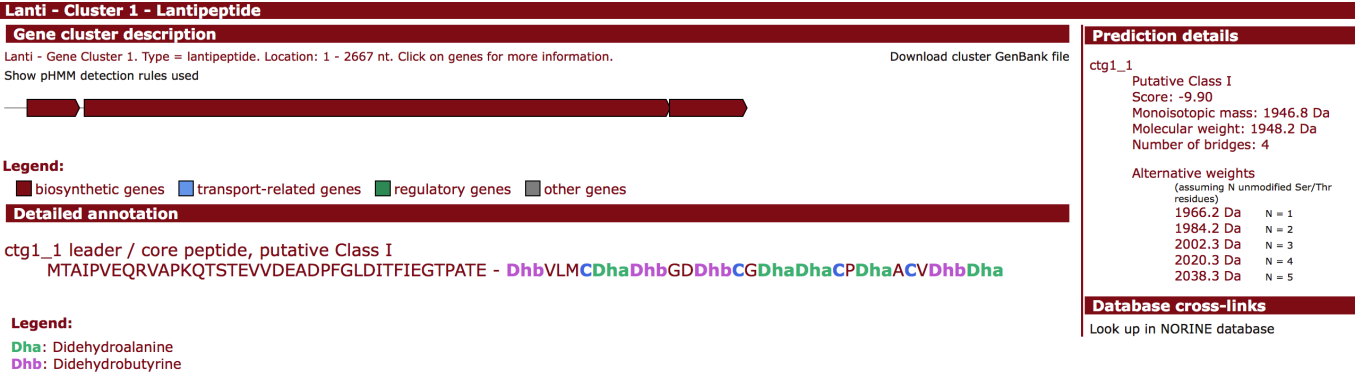
A large number of the predicted PKS and NRPS appeared to be gene cluster fragments, likely due to the short contig size. Some highlights of the prediction are the 9 type I PKS clusters, 2 type II PKS clusters, 2 type III PKS, as well as 8 NRPS and 10 mixed PKS-NRPS or PKS_I-PKS_{II} clusters. The remaining ones were predicted to encode biosynthetic pathways for 2 terpenes, 3 siderophores, a lantipeptide, 3 bacteriocins, ectoin, and 2 melanins. Four clusters were identified but had no predicted compound family. Two predicted pathways showed high homology to known biosynthetic gene clusters (Gosse, J.T. *et al.*, 2015).

Due to the partial nature of many clusters and prediction uncertainties of others, these gene clusters were not chosen as targets for further study. On the other hand, the lantipeptide pathway showed itself as a very exciting cluster to be expressed due to its small size and lower genetic complexity, as few genes are involved in lantipeptide production. The lack of further information on the pathway and on lantipeptides in general was another point that set the cluster as our object of study.

AntiSMASH was able to predict some features of the pathway, such as leader and core peptide positions, modified residues and number of crosslinks (Figure 2.3 and 2.4). It suggested the molecule belongs to Class I lantipeptides, known to comprise antibacterial compounds.

Unfortunately no protease responsible for the separation of the leader from the core sequence was found on the cluster or on the total sequence, but according to the residues involved it was predicted as being an Endoproteinase

Glu-C with the use of ExPASy PeptideCutter (Gasteiger E. et al., 2003). This ended up being beneficial for the heterologous expression of the lantipeptide as the activation of the core peptide could be controlled with the addition of commercial protease.



proved to be very difficult to amplify, probably caused by the rich GC regions from where the primers were generated, which increases T_m and make the process more challenging. These difficulties were solved with the use of DMSO in the reaction mixture, which helps the dissociation of hydrogen bonds due to its association to the major and minor grooves of template DNA destabilizing the double helix structure, (Hardjasa A. *et. al.*, 2010) allowing lower melting temperatures and primer annealing.

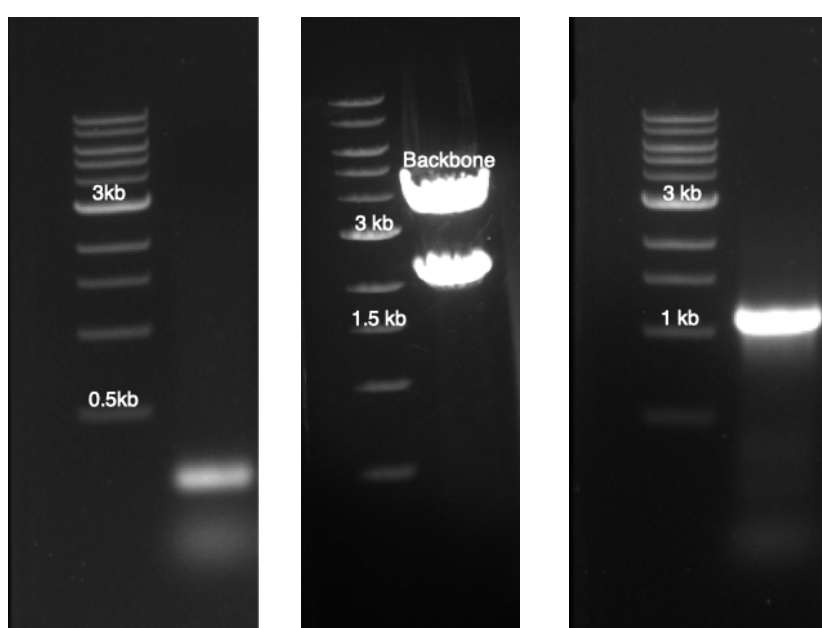


Figure 2.5: PCR amplification of LanA (200 bp), check digestion of LanB (1.9kb) and LanC (1.1kb)

All three genes were successfully extracted from gel for blunt cloning, but LanA due to its small size was more challenging, with low yields for gel extraction. The problems with yield were corrected with blunt cloning, as the process works well even with low amounts of insert and vector extraction had a good recovery, with yields ranging from 300-500 ng vector/ μ L. These high levels

allowed for a better digestion reaction and consequently more LanA fragments obtained for the ligation on expression vector.

For the construction of the multiexpression gene system, the preference was given to ligating genes from an initial pMGx-K-LanB, where the other genes would be added on both extremities of the first gene, simulating the original pathway and respecting the gene order.

The first added gene was LanC, followed by LanB. Ligation was confirmed through digestion of the vector (Figure 2.6). The attempts of ligating LanA on the construct were unsuccessful so the use of a double transformation system was considered. Previous works from Karakas-Sen and Narbad (2011) and Shi *et. al.* (2011) indicated the use of two vectors for nisin transformation in *E. coli*.

LanA was first cloned into pMGX-A and double transformed on *E. coli* BL21 with pMGX-K-LanBC, with the use of two resistance markers, ampicillin and kanamycin, to allow detection of successful transformants. There were no results for colony selection.

The second attempt of cloning used pET21 as the backbone for LanA cloning and produced successful transformants. The double transformation system was formed then by pMGX-K-LanBC plus pET21-LanA.

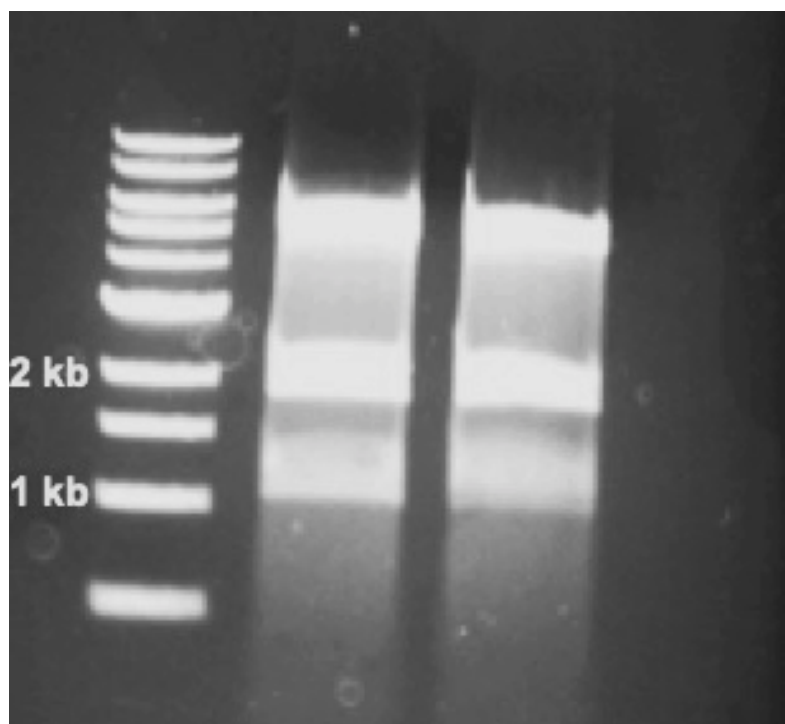


Figure 2.6: Check digestion of multiple expression vector pMGX-K containing LanB (~2kb) and LanC (~1.1kb). Two vectors originated from successful transformed colonies are depicted.

Expression of lantipeptide pathway

Check expression of pathway happened to ensure that all genes were able to produce the respective proteins before performing double transformation. Protein expression of LanBC construct happened using *E coli* BL21 as strain and followed the procedures described on the Materials and Methods session. The SDS-PAGE gel showed the His-tagged proteins corresponding to LanB (71kDa) and LanC (40kDa) (Figure 2.7).

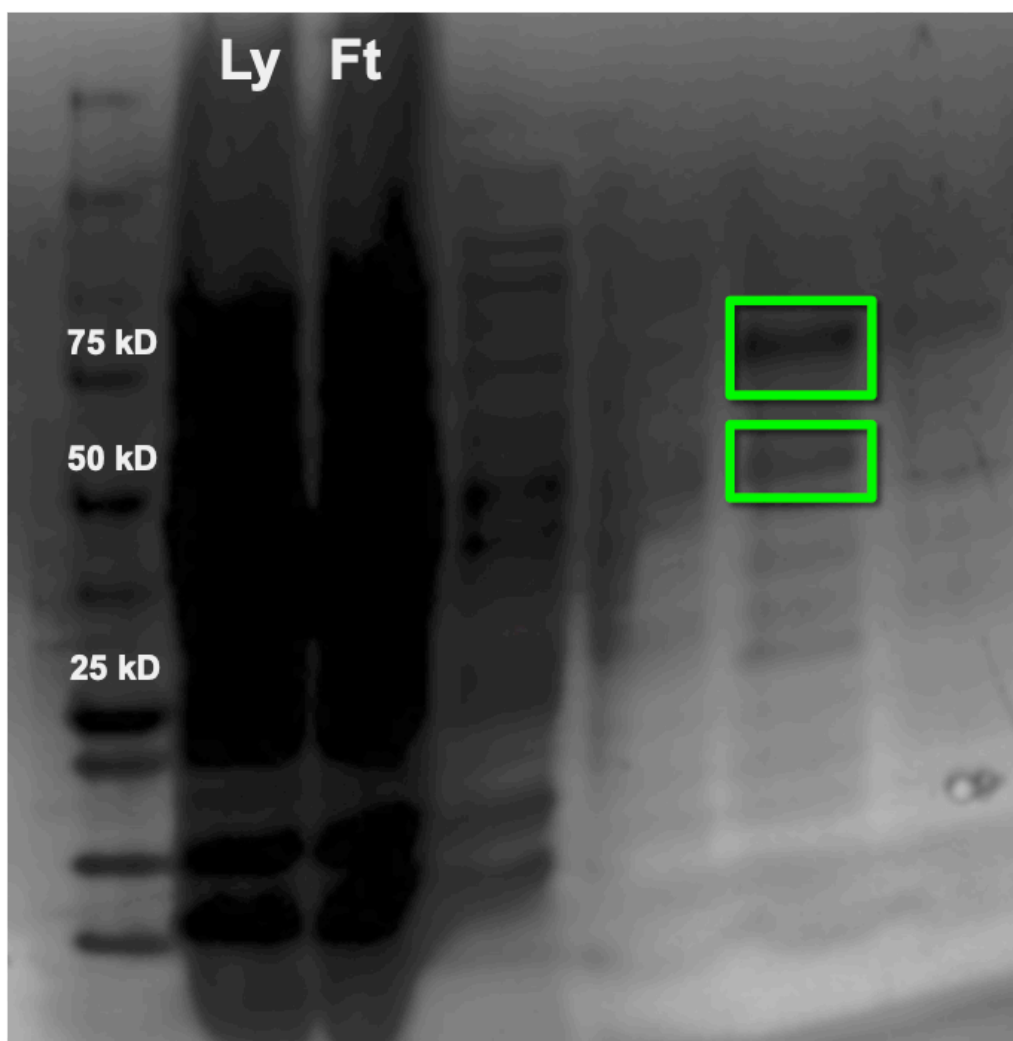


Figure 2.7: SDS-PAGE of LanB and LanC expression via multiexpression gene system pMGX.

Ly: lysate; FT: flow through; In detail: LanB (71kDa) and LanC (40kDa).

Expression of LanA showed to be more difficult than the other proteins, as even after multiple attempts using the same protocol no protein production was observed. The presence of a very prominent 15kDa band on the SDS-PAGE analysis led me to believe that the protein was dimerized, as the expected size for the protein is around 7.5 kDa (Figure 2.8). To check this possibility, a set of tests was conducted to verify the location of the protein as well as the identity of the band.

One of the first tests was a temperature profiling to ensure the expression of the protein. Due to its diminutive structure and no activation, there was a possibility of degradation by proteases if the expression was carried on for long periods of time. To test this possibility, an expression profile of increased temperature for less time (30°C for 4 hours) was tested in opposition to the traditional method (16°C, overnight). The expression carried at 16°C was followed using a time-course approach, with fractions taken at specific time points for protein expression analysis. There was no LanA detected in any of the two methods tested.

Expression of LanA was tested using B-PER Bacterial Protein Extraction Reagent (ThermoFisher Scientific). The use of such reagent, a detergent, would help with the recovery of LanA if the protein was found on the insoluble protein fraction or in inclusion bodies instead of being in solution. This would explain the absence of LanA in fractions collected via regular protein expression methods.

E coli BL21 cell pellets which expression was conducted at 30°C and 16°C were treated with B-PER and soluble and insoluble fractions were analyzed for LanA expression via SDS-PAGE. No visible signs of LanA expression were found in any of the aliquots. As traditional protein expression proved unsuccessful to obtain the lantipeptide, we decided to attempt cell free expression.

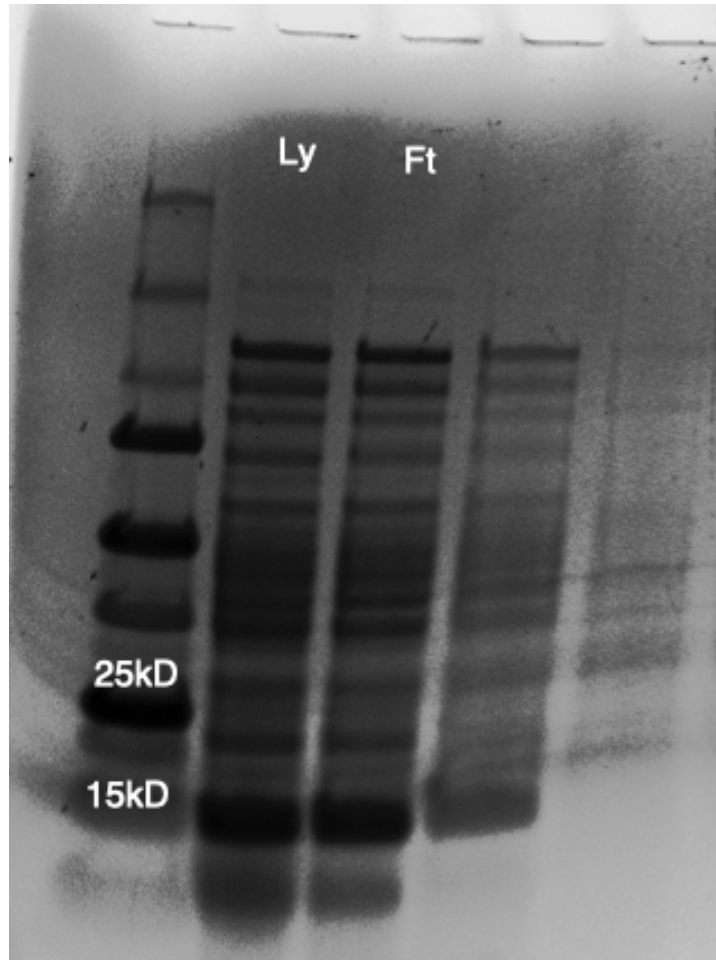


Figure 2.8: Attempt of LanA expression; no visible band corresponding to LanA (7.5kDa) appeared. Ly: lysate; FT: flow through. 15kD strong band visible on fractions.

Cell free LanA expression

One solution for challenging protein expression is to use cell-free protein expression. This method is recommended for proteins for which regular expression represents a bottleneck, such as membrane or posttranslational modified proteins, toxic proteins, and potential medicines (Klammt, C. et al., 2006; Katzen, F., Chang, G., & Kudlicki, W., 2005; .Ogonah, O. W., Polizzi, K. M., & Bracewell, D. G., 2017). By skipping avoiding the bacterial cell and its complex proteome, risk of degradation or toxicity of the produced protein is minimized.

The use of a commercial *E. coli* kit based on reconstituted purified components (Shimizu, Y. et al., 2001) allowed the use of bacterial machinery that is able to express amplified DNA. We found this especially useful as the amplification of LanA was simple and successful, but cloning of the gene was more difficult to achieve.

Association of a fluorolabelling BODIPY-Lys system to cell free expression was an extra step to improve product visualization. Typically [³⁵S]methionine is the label of choice, but other labels such as [³⁵S]cysteine, [¹⁴C]leucine and [³H]leucine could be used as well. The downside of these labels is a more time-consuming, labour-intense protocol and the use of radioactivity (Kobs, B. et al., n.d.). Even with one single lysine residue on the lantipeptide, it could provide a better resolution than the cell free system alone, since the peptide is fairly light.

Cell-free expression was not effective in expressing the lantipeptide. Moreover, imaging of SDS-PAGE was affected by kit components, such as histidine-tagged ribosomal proteins below 16kDa and unspecific labelling of other proteins, even present in the negative control (Figure 2.9). Alternative

staining methods such as Coomassie Blue were not able to help detection of the peptide.

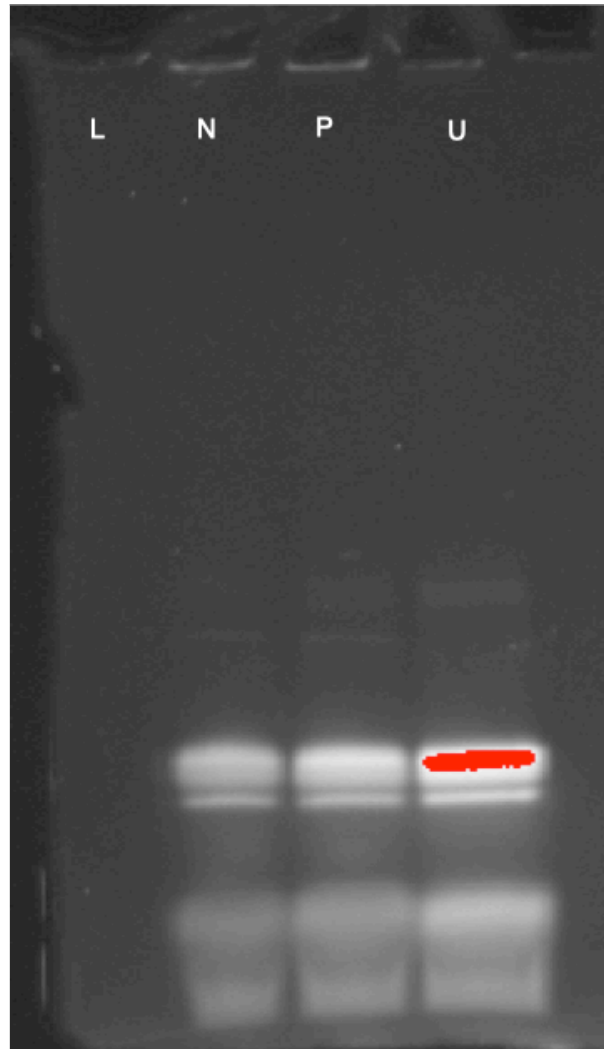


Figure 2.9: SDS-PAGE imaging of cell-free expression of lantipeptide. L: molecular marker (not visible); N: negative control; P: BODIPY-Lys labelled LanA expression U: unlabelled LanA expression. Red band is indicative of signal saturation.

Unfortunately the expression of the lantipeptide portion of the pathway was not possible to be achieved, despite many efforts. Cell-free seems to be the best alternative to obtain protein expression, but further optimization of

commercial kit or use of laboratory-prepared bacterial lysate cell-free systems, such as the one described by Krinsky N. et al. (2016) could be other methods that could provide better results.

Conclusion

This project focuses on heterologous expression of a lantipeptide from a bacterial isolate from urban environment. Using *E. coli* as the expression host and an inducible T7 RNA Polymerase-mediated Multigene Expression System (MES) developed in-lab, named pMGX, the different components of the biosynthetic pathway were cloned and evaluated for the production of the propeptide and its active form.

We were able to successfully clone and express the genes responsible for lantipeptide tailoring, named LanB and LanC using the MES system, but despite multiple attempts, including bacterial dependant and cell-free expression methods, the lantipeptide portion of the pathway was not obtained. This impacted the prospective assays on identification and characterization of peptide activity.

Although some of the work was not successful at first, we expect to keep expanding the knowledge on RiPPs and their presence in different bacterial species and their biological potential. Other methods of expression, including laboratory-produced cell-free expression systems are recognized as alternatives for part of the future work to expand the studies of this group of poorly known molecules.

Chapter Bibliography

Arnison, P. G., Bibb, M. J., Bierbaum, G., Bowers, A. A., Bugni, T. S., Bulaj, G., Camarero, J. A., Campopiano, D. J...Willey, J. M., and van der Donk, W. A. (2013). Ribosomally synthesized and post-translationally modified peptide natural products: overview and recommendations for a universal nomenclature. *Natural Product Reports*, 30, 108–160.

Essig, A., Hofmann, D., Münch, D., Gayathri, S., Künzler, M., Kallio, P.T., Sahl, H.G., Wider, G., Schneider, T., Aebi, M. (2014). Copsin, a novel peptide-based fungal antibiotic interfering with the peptidoglycan synthesis. *Journal of Biological Chemistry*, 289(50):34953-64.

Férir, G., Petrova, M.I., Andrei, G., Huskens, D., Hoorelbeke, B., Snoeck, R., Vanderleyden, J., Balzarini, J., Bartoschek, S., Brönstrup, M., Süßmuth, R.D., Schols, D. (2013). The lantibiotic peptide labyrinthopeptin A1 demonstrates broad anti-HIV and anti-HSV activity with potential for microbicidal applications. *PLoS One*, 8(5):e64010.

Gasteiger E., Gattiker A., Hoogland C., Ivanyi I., Appel R.D., Bairoch A. (2003). ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Research*, 31:3784-3788

Hofmann, F. T., Szostak, J. W., & Seebeck, F. P. (2012). In Vitro Selection of Functional Lantipeptides. *Journal of the American Chemical Society*, 134(19), 8038–8041.

Karakas-Sen, A., Narbad, A. (2012). Heterologous expression and purification of NisA, the precursor peptide of lantibiotic nisin from *Lactococcus lactis*. *Acta Biologica Hungarica* 63(2):301–310.

Katzen, F., Chang, G., & Kudlicki, W. (2005). The past, present and future of cell-free protein synthesis. *Trends in Biotechnology*, 23(3), 150–156.

Kobs, B. G., Hurst, R., Betz, N., Ph, D., & Godat, B. (n.d.). Labeling System, 23–27. Retrieved from: <https://promega.media/-/media/files/resources/promega-notes/77/fluorotecttm-greenlys-in-vitro-translation-labeling-system.pdf?la=en>

Klammt, C. , Schwarz, D. , Löhr, F. , Schneider, B. , Dötsch, V. and Bernhard, F. (2006), Cell-free expression as an emerging technique for the large scale production of integral membrane protein. *The FEBS Journal*, 273: 4141-4153.

Knerr, P.J., van der Donk, W.A. (2012). Discovery, biosynthesis, and engineering of lantipeptides. *Annual Review of Biochemistry*, 81:479–505.

Krinsky N, Kaduri M, Shainsky-Roitman J, Goldfeder M, Ivanir E, Benhar I, et al. (2016) A Simple and Rapid Method for Preparing a Cell-Free Bacterial Lysate for Protein Synthesis. *PLoS ONE*, 11(10): e0165137

Letzel, A., Pidot, S.J. and Hertweck, C. (2014). Genome mining for ribosomally synthesized and post-translationally modified peptides (RiPPs) in anaerobic bacteria. *BMC Genomics*, 15:983

Link, A.J. (2015). Biosynthesis: Leading the way to RiPPs. *Nature chemical biology*, 11(8), 551-2.

Marsh, A. J., O'Sullivan, O., Ross, R. P., Cotter, P. D., & Hill, C. (2010). In silico analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics*, 11(1), 679.

Mohr, K. I., Volz, C. , Jansen, R. , Wray, V. , Hoffmann, J. , Bernecker, S. , Wink, J. , Gerth, K. , Stadler, M. and Müller, R. (2015), Pinensins: The First Antifungal Lantibiotics. *Angewandte Chemie International Edition*, 54: 11254-11258.

Nagao, J., Harada, Y., Shioya, K., Aso, Y., Zendo, T., Nakayama, J., & Sonomoto, K. (2005). Lanthionine introduction into nukacin ISK-1 prepeptide by co-expression with modification enzyme NukM in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, 336(2), 507–513.

Shi, Y., Yang, X., Garg, N., & van der Donk, W. A. (2011). Production of Lantipeptides in *Escherichia coli*. *Journal of the American Chemical Society*, 133(8), 2338–2341.

Shimizu, Y., Inoue, A., Tomari, Y., Suzuki, T., Yokogawa, T., Nishikawa, K., & Ueda, T. (2001). Cell-free translation reconstituted with purified components. *Nature Biotechnology*, 19, 751.

van der Donk, W. A., & Nair, S. K. (2014). Structure and mechanism of lanthipeptide biosynthetic enzymes. *Current opinion in structural biology*, 29, 58-66.

van der Meer, J. R., Polman, J., Beerthuyzen, M. M., Siezen, R. J., Kuipers, O. P., & De Vos, W. M. (1993). Characterization of the *Lactococcus lactis* nisin A operon genes *nisP*, encoding a subtilisin-like serine protease involved in precursor processing, and *nisR*, encoding a regulatory protein involved in nisin biosynthesis. *Journal of bacteriology*, 175(9), 2578-88.

Yu, Y., Zhang, Q., & van der Donk, W. A. (2013). Insights into the evolution of lanthipeptide biosynthesis. *Protein science: a publication of the Protein Society*, 22(11), 1478-89.

Zheng, Q., Wang, Q., Wang, S., Wu, J., Gao, Q., & Liu, W. (2015). Thiopeptide Antibiotics Exhibit a Dual Mode of Action against Intracellular Pathogens by Affecting Both Host and Microbe. *Chemistry & Biology*, 22(8), 1002–1007.

Chapter 3: Assessing cave bacterial natural products: antimicrobials for bats and humans

Introduction

No one knows what is like to be the Batman, to be the sad man...

White-nose syndrome (WNS) is a fungal disease caused by the pathogen *Pseudogymnoascus destructans* (formerly known as *Geomyces destructans*). It initiates a cascade of physiologic disturbances in the hibernating bat host, associated with an unprecedented high bat mortality event. Since the winter of 2006-2007, bat population declines exceeding 75% have been observed at surveyed hibernacula.

Since emerging in the northeastern United States, WNS has been confirmed by gross and histologic examinations of bats at 33 sites (Blehert, D.S. et al., 2009). WNS has spread rapidly and now occurs throughout the northeastern and mid-Atlantic regions and in Ontario and Québec provinces in Canada and currently affects at least seven species of hibernating bats (Frick, F.W. et al., 2010).

The infection causes physiological changes, with frequent arousal from hibernation and loss of fat tissue and water. Infected bats developed respiratory acidosis and as infection progresses. This is initially well tolerated by repair and buffering mechanisms, until they are completely depleted and lead to death of the bat (Verant, M.L. et al., 2014).

Bats that have had contact with *P. destructans* and not developed the disease were tested positive for the presence of the fungus through PCR (Lorch, J. M. et al., 2013), meaning that the disease can be spread to other individuals that could be sensitive to the pathogen. As studies have shown that that only some

immune system pathways remain active, with leucocytes sequestered during hibernation, responses to pathogens in peripheral tissues decrease, leaving animals susceptible (Field, K.A. et al., 2018).

Fly away skyline bat, fly... white-nose syndrome treatments

Recent research found that in the environment of a hibernaculum, *P. destructans* could remain up to 1 year following first detection of infected bats within the site. This accumulation of the pathogen in the hibernaculum environment may reflect deposition of fungal material shed by bats or growth of the fungus (Verant, M.L. et al., 2018). As a result, treatments currently used to tackle the infection focus on the static part of the equation, the cave environment.

Antifungals such as econazole, sulconazole, chloroacetoxyquinoline, pyrrithione zinc, ciclopirox olamine, and chloroxine, fungicides (phenylmercuric acetate) and biocides (benzalkonium chloride) caused prominent growth inhibition of *P. destructans in vitro* (Chaturvedi, S. et al., 2011) and were pointed as good alternatives due to commercial availability.

However, any proposal for the decontamination requires a careful balance considering the likely deleterious effects on the environment. It is worth highlighting the rich life diversity in caves, including many species of fungi and bacteria (Urzi, C. et al., 2010). To date, the only effective chemical treatment is with benzalkonium chloride, which has been used for many years in the famous Lascaux cave (Montignac, France) to control fungal infestation. Although

effective, this treatment has resulted in the replacement of native bacterial communities with more resistant bacteria in less than five years of use (Bastian, F. et al, 2009).

Other proposed treatments are the use of UV light, that was shown to cause growth disruption in *P. destructans*, and natural compounds such as cold-pressed terpeneless valencia orange oil and bacterial volatile organic compounds (Palmer, J.M. et al., 2018; Boire, N. et al., 2016; Cornelison, C. T. et al., 2014). These alternatives are still in experimental stages and there is a risk of disturbance of the cave habitat as external agents. One of the most efficient solutions seems to be the use of agents that are already part of the environment to minimize damage to such fragile ecosystem.

We are the cavemen – extreme environments as source of compounds

Caves are a deep subsurface habitat that possess unique characteristics including limited amounts of light, low amounts of organic nutrients (depends on the caves), and higher humidity and concentration of minerals (Cheeptham, N., 2013). These features create microenvironments, often resulting from an array of microbial processes present within them (Barton and Northup, 2007; Montano and Henderson, 2013).

Nonetheless, microbes have not only been able to survive, but to thrive in these extreme conditions (Barton, 2006; Barton and Northup, 2007; Barton, 2015; Ghosh, S. et al., 2016) because of their adaptation in their physiology and in the exploitation of different and specialized metabolic pathways (Cheeptham,

N. et al., 2013). This unique potential can be examined to yield novel compounds with multiple applications

Microbial diversity studies from cave habitats showed the presence of Actinomycetes, which possess an immense pool of secondary metabolites (Maciejewska et al., 2016; Maciejewska et al., 2017). Studies of two *Streptomyces* isolates, designated as ICC1 and ICC4 from the Iron Curtain Cave in Chilliwack, Canada, have shown that both strains exhibit antagonistic activities against multidrug resistant bacterial strains of *Escherichia coli* and non-resistant strain of *Staphylococcus aureus* (Ghosh et al., 2017).

The potential shown by this WNS-free Canadian cave compelled our study to determine what molecules could be involved in antimicrobial activities presented by some of its organisms.

Project objectives

The objective of this project is to examine cave bacterial secondary metabolite production and organism diversity to determine the relationship between the microbial flora and to retrieve potential natural compounds active against white-nose syndrome (WNS).

In order to achieve the project goals, custom cultivation techniques and whole genome sequencing were used to provide identification of potential biosynthetic clusters. Bacterial extracts were analyzed using LC-ESI-HRMS to yield spectra suitable for computational analysis using a natural product spectral network.

We expect with this project to expand the understanding of bacteria living in extreme environments in multiple approaches and to contribute to the development of mechanisms of fighting against a deadly disease that is decimating bats in North America.

Materials and Methods

Isolate collection

Ten isolates from different areas and caves from British Columbia were used in this study (Table 3.1). Collection and isolation of strains was performed by Dr. Naowarat Cheeptham's laboratory group (Thompson Rivers University, Kamloops, British Columbia). Detailed collection conditions and geographic points can be retrieved from the work of Soumya Ghosh (Ghosh, S. et al., 2017).

Table 3.1: Bacterial isolates collected from various environments in British Columbia. Environmental source, cultivation conditions and strain sensitivity test results are presented.

Isolate nomenclature	Source	Sensitive test microorganisms	Growth Media	Incubation temperature	RPM	Incubation (days)
ICC1	Iron Curtain Cave	<i>*E. coli</i> 15-102, <i>*E. coli</i> 15-124	R2A	8°C	200	10
ICC4	Iron Curtain Cave	<i>Staphylococcus aureus</i> , <i>Pseudomonas aeruginosa</i> , <i>E. coli</i> , <i>*E. coli</i> 15-318, <i>*E. coli</i> 15-102, <i>*E. coli</i> 15-124, <i>Staphylococcus aureus</i> , <i>E. coli</i> 15-318	V8	8°C	200	30
#25	White Rabbit Cave	<i>*Methicillin-resistant Staphylococcus aureus</i> (MRSA), <i>Staphylococcus aureus</i>	R2A	15°C	200	28
16102307-1	Bat's wings	<i>Pseudogymnoascus destructans</i>	SD	15°C	200	15
16102307-1.2	Bat's wings	<i>P. destructans</i>	SD	15°C	200	15
13	Timber/Bush lake	<i>P. destructans</i>	SD	15°C	200	7
90	Timber/Bush lake	<i>P. destructans</i>	SD	15°C	200	7
3	Timber/Bush lake	<i>P. destructans</i>	SD	15°C	200	7
55	Timber/Bush lake	<i>P. destructans</i>	SD	15°C	200	7
84	Timber/Bush lake	<i>P. destructans</i>	SD	15°C	200	7

*Multidrug resistant strains

DNA extraction and Whole Genome Sequencing of isolates

For DNA extraction all strains collected were cultivated. For ICC1, 50 mL of R2A broth (0.5 g proteose peptone, 0.5 g casamino acids, 0.5 g yeast extract, 0.5 g dextrose, 0.5 g soluble starch, 0.3 g potassium phosphate dibasic, 0.05 g magnesium sulphate heptahydrate, 0.3 g sodium pyruvate) was used for culturing and for the other isolates Sabraud Dextrose broth (SDB) (40 g dextrose, 10 g peptone) was the medium of choice. All strains were cultivated at 15 °C and 200 rpm for 7 days.

Cells were pelletized after incubation period and resuspended in 480 µL of 0.5 M Tris. DNA was extracted with the help of a kit (Wizard Genome DNA Extraction, Promega) following manufacturer's instructions.

Whole genome sequencing was performed for six selected cave isolates: ICC1, ICC4, Bat's wings 1, Bat's wings 1.2, isolate #90 and isolate #84. The extracted DNA was submitted for sequencing through Integrated Microbiome Resource – IMR (Centre for Comparative Genomics and Evolutionary Bioinformatics, Dalhousie University, Canada). Sequencing details: Illumina MiSeq, 50x2x300 bp paired-end).

Genome assembly and annotation

Sequencing files came as raw reads and were assembled and annotated. To proceed with both actions, the combined platform PATRIC v3.5.2 /RAST v2.0 (<https://www.patricbrc.org>) (Wattam, A. R, et al., 2017) was used. Raw files

were uploaded under the tab “Services – Assembly” and the paired files were uploaded for each respective genome. Assembly strategy used was “Auto”, which yields both Velvet, Spades and a consensus files.

With the assembly files in hand, annotation was pursued through RAST. The FASTA files were uploaded on the system without any identification of species or NCBI number associated. For the annotation strategy, “Classic RAST” was picked as the standard scheme. Gaps were selected to be automatically backfilled and errors were automatically corrected. The library was sequenced using a 600-cycle v3 reagent kit (Illumina), with an average sequencing coverage of 3.8X. The raw files were assembled using PATRIC Genome Assembly Service and annotated using RAST (Rapid Annotation using Subsystem Technology) online tools. Genomes are available on NCBI database (Accession numbers CP030286 and CP030287).

Organism identification and secondary metabolites prediction

The sequencing files yielded from WGS were used for species identification using NCBI BLAST database and for secondary metabolite prediction through antiSMASH 3.0 (<http://antismash.secondarymetabolites.org>) (Weber, T. et al., 2015). Predictions were made using standard settings from both databases.

Orphamide pathway primer design and amplification

To amplify pathways determined to be present upon analysis using antiSMASH 3.0, three sets of primers were constructed to amplify the three genes composing the orphamide pathway. To enable recovery and evaluation of expression, the C-terminus of proteins were 6XHis-tagged. Primer sequences are presented in Table 3.2.

Table 3.2: primer sequences used for amplification of orphamide pathway.

Primer	Sequence
LpAF	5'-AAAAACATATGCATCATCACCATCACCACATGCAGTATTCCTCCAGCC-3'
LpAR	5'-GCTAGCGAATTCCTAAATCCGTCGACGGCG-3'
LpBF	5'-CGGGGGCCTCCTCATATGATGCAGTTCAGCGACTTACT-3'
LpBR	5'-AAAAAGCTTACCTTAGTGGTGATGGTGATGATGACCACCCAGGACGATCTCCATTG-3'
LpCF	5'-TTTTTTAAGCTTGTGAACATCCTCGAGCTG-3'
LpCR	5'-TTTTTTCATATGTTAGTGGTGATGGTGATGATGACCACCACGACACCACCTCCCC-3'
LpBCR	5'-CAGCTCGAGGATGTTACAGGAC-3'

Amplification conditions for production of fragments suitable for cloning applications follow: for a total volume of 50 μ L reaction were combined 200ng DNA template, 0.2 mM of dNTP mix, 0.25 μ M of each primer, 8 % DMSO, 1 μ L Herculase II DNA Polymerase (Agilent) with corresponding 10 X reaction buffer. Amplification started with a denaturation step at 95 °C and 35 cycles consisting of 95 °C for 1 minute, 65 °C for 30 seconds (LpA)/64 °C for 1 minute (LpB)/61 °C for 1 minute (LpC), and 72 °C for 1 minute followed by a final extension step of 3 minutes at 72 °C.

Fragments amplified were resolved in 1 % w/v agarose gels and extracted from gel using an isolation kit (E.Z.N.A Gel Extraction Kit, Omega-Biotek).

Extraction and fractionation of fermented isolates

The sequenced isolates were cultivated to allow separation and identification of secondary metabolites for further activity assays. For cultivation, 400 mL of SDB and modified V8 broth (20 % v/v low sodium V8 vegetable cocktail supernatant, 1 % Pharmamedia, 0.5 % calcium carbonate, pH 6, tap H₂O) were used as media for secondary metabolites production. Fermentation occurred at 15 °C for 15 to 30 days depending on the organism. Weekly aliquots of 10 mL of the broth were collected to evaluate growth and metabolite production.

1. Ethyl acetate extraction

A simple extraction was carried on using ethyl acetate as a solvent. 200 mL of the fermented broths were pelletized at 4000 rpm for 15 minutes and the collected supernatant had its pH corrected to 2.5 using 3 M HCl. The liquid was then washed with ethyl acetate three times, using respectively 0.5 V of solvent in the first wash and 0.3 V in the subsequent ones. All fractions were reunited and the solvent evaporated. The remaining solvent was left to air dry.

2. Resin extraction

From the remaining fermented broth, 2mL of each were processed using Amberlite XAD-16 resin to remove polar metabolites. Soaking the beads in 1% w/v water for 30 minutes and soaking two times in 1 % v/w methanol for 30 minutes each time activated the resin, that was stored in water to allow swelling.

The active resin was added to the filtrated fermented broth at 10 % w/v and shaken overnight at 200 rpm. Beads were filtered the next day and soaked in excess methanol for 5 hours at 200 rpm. The solvent was evaporated and the samples were resuspended in a volume of methanol suitable for analysis.

3. Kupchan partition

From the fermented isolates, 50 mL were taken for partition and the remaining broth was filtered and frozen at -20 °C for future use. The fractions were spun at 4000 rpm for 15 minutes and filtered using a 0.22 mm syringe filter. The filtrate was freeze-dried and resuspended in 10 mL methanol for solvent fractioning using a modified Kupchan partition method.

To methanol fractions, original volume was determined and water was added to adjust the aqueous content to 10 % v/v. The same volume of n-hexane was added and the solution was vigorously shaken. After phase separation, the upper phase was collected and the n-hexane was removed through rotary evaporation. The remaining phase had the aqueous content to 40 % water v/v. To this, 1 X original volume of chloroform was added and after phase separation the organic lower layer was extracted and chloroform evaporated.

The remaining phase was stripped from traces of methanol and chloroform via evaporation and reconstituted to its original volume with the help of H₂O. The same volume of n-butanol was added and upon phase separation the upper fraction was collected and the solvent evaporated with the help of water. The remaining aqueous phase was concentrated to remove traces of n-butanol and freeze-dried.

Bacterial Kirby-Bauer activity assays

All fermented filtered broths and fractions were tested for activity against a set of microorganisms. To pursue the testing, the chosen method was a modified version of the Kirby-Bauer disk diffusion assay.

Filter paper disks were prepared using a hole punch and Grade 3 Whatman filter papers. The disks were autoclaved and were soaked in 20 µL of the desired testing solutions. Pencil was used to mark the identity of the compounds. The disks were left to air dry completely before use. Antibiotic controls were prepared according to CLSI guideline standards (Clinical and Laboratory Standards Institute, 2012) and the disk preparation occurred following the same protocol.

Bacteria to be tested were prepared through CLSI standards. Three strains were selected for testing: *E. coli* BW25113, *Pseudomonas aeruginosa* PA01 and *Bacillus subtilis* 168. Colonies were selected from overnight culture plates and diluted in Mueller-Hinton broth to reach OD 600 between 0.07-0.1. The bacterial solution was spread on petri dishes containing Mueller-Hinton agar with the use of a cotton swab. The plates were left to dry for some instants, and

disks containing testing solutions and the suitable antibiotic were placed onto the agar. Plates were incubated for 4 to 6 days at 15 °C and examined for the presence or absence of inhibition halos.

Bacterial microdilution assays

Fractions obtained from SPE were tested using bacterial microdilution assay.

With the help of a multipipettor, 100 µL of Mueller-Hinton medium was dispensed into all wells of a 96-well microtitre plate. Samples to be tested were dissolved in the adequate solvent then diluted in Mueller-Hinton medium to 2x the top concentration desired in the test. 100 µL of 2x test solutions and antibiotic control were added into the wells in column 1 of microplate. Using the multipipettor set at 100 µL, mix the test solution into the wells by sucking up and down 6-8 times.

Withdraw 100 µL from column 1 to be added to column 2. This would make column 2 a two-fold dilution of column 1. Proceed with mixing up and down 6-8 times. Column 3 received 100 µL of solution of column 2. The serial dilution was repeated down to column 10 to create a dilution of test compounds and control antibiotic. The 100 µL aspirated from column 10 was discarded rather than put in column 11.

Three strains were selected for testing: *E. coli* BW25113, *Pseudomonas aeruginosa* PA01 and *Bacillus subtilis* 168. Colonies were selected from overnight culture plates and diluted in Mueller-Hinton broth to reach A_{600} between 0.07-0.1. Pour bacteria of the right A_{600} into a sterile petri dish. The appropriate

inoculum size for standard MIC is 10^4 to 10^5 CFU/mL. 5 μ L of bacterial solutions was dispensed into wells in columns 11 to 1 in that order to achieve the standard inoculum. Column 12 served as broth sterility control.

Plates were incubated at 37 °C or other desired temperature, depending on the assay performed. When satisfactory growth was obtained MIC was determined as the lowest concentration of compound that inhibited bacterial growth.

Mass spectrometry of positive extract fractions

Fractions that presented activity in disk diffusion assays were analyzed through Mass Spectrometry. 2 mL of the methanolic fractions were dried out. About 5 mg of each fraction was purified using Solid Phase Extraction (SPE) cartridges (HyperSep C18, Fisher) and a positive pressure manifold.

Dry fractions were resuspended in 1 mL of miliQ water and applied to the columns, which were equilibrated with 1 wash of HPLC-grade acetonitrile (ACN) and 2 washes of H₂O. Columns were filled with 1 volume of water and the liquid was collected and labeled sub fraction A.

To the original fractions, 100 μ L of a solution of 10 % ACN in water was used to wash any sample remaining. The liquid was injected in the separation column. The column was washed with 1 volume of the 10 % ACN and the flow was collected and mixed with sub fraction A.

The next wash consisted of 100 μ L of a 50 % ACN solution to remove remaining sample residues from the original flask, followed by the injection of

the liquid in the separation column. Column was washed with one full volume of same solution and fraction collected was labeled sub fraction B.

The fourth wash was consisted of 100 % ACN, with addition of remaining sample by injecting 100 μ L of an ACN wash from the original sample vial. The column was washed with 1 volume of ACN, generating sub fraction C.

The last fraction generated was made of a solution of methanol: dichloromethane 1:1 (v/v). After wash of original sample vial and injection of sample, the columns were washed with 1 volume of the solvent solution to yield sub fraction D.

Sub fractions C and D were dried using a vacuum evaporator, and sub fractions A and B were first evaporated to get rid of ACN and later freeze dried to eliminate water fraction.

For Mass Spectrometry sub fractions A and B were resuspended in ACN: H₂O 1:1 (v/v) and sub fractions C and D were resuspended in 100 % ACN to a concentration of 500 μ g/mL. Original fractions were also analyzed, being resuspended in ACN:H₂O 1:1 (v/v) if coming from an aqueous partition of ACN: methanol 1:1 (v/v) if coming from organic solvent fractions. The concentrations were adjusted to be the same as the sub fractions.

Molecular composition of the extracts was determined via LC-ESI-HRMS on a LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA). The data collection protocol used was proposed by Mingxun Wang *et. al.* (2016), which generates data suited for Global Natural Products Social Molecular Networking (GNPS) analysis. Full resolution scan MS spectra (m/z 50 – 1500) were acquired in TOF, followed by six scans on low-resolution ion trap. The top six most intense ions in a particular

scan were fragmented using collision-induced dissociation. RAW files were converted to .mzXML using RAWConverter (He *et. al*,2015) to be suited for GNPS spectral analysis.

Spectral Networking Analysis

MS analysis used the Capture Compound Mass Spectrometry Technology (CCMS) ProteoSAFe Workflow Input Form (available at <https://gnps.ucsd.edu/ProteoSAFe/static/gnps-splash2.jsp>). No Data Preset was chosen. Speclibs was the spectral library used as a comparison parameter. Multiple input MS/MS spectra for the raw fractions and subfractions in .mzXML formats were uploaded the same. Precursor Mass tolerance and Fragment Ion Mass tolerance were set as 2.0 and 0.5 Da, respectively. Advanced parameters were set as standards. Results were analyzed based on m/z errors and mass difference. Complete results are presented in Appendix 1.

Results and Discussion

Disk diffusion assays

The assay was performed to test different ways of extraction; however the procedures were the same for all extraction methodologies used, with exception of temperatures tested throughout the analysis.

The first disk diffusion assay tested the activity of ethyl acetate extracts of the cultured strains, as well as XAD-16 resin fractions. 20 µg of extracts were not able to cause inhibition of growth in any of the strains tested. This is not compatible with previous results from plug assays performed by our collaborators. The main cause of this lack of activity was probably the temperature the assays were conducted: 37 °C. As the bacterial extracts were produced in much lower temperatures (8-15 °C) the increase of it during the testing could have caused the loss of activity. Another cause for the inactivity is thought to be the production method of extracts that could have prioritized molecules with no activity on the organisms tested.

To overcome the unsatisfactory results and give a better perspective of analysis for the fermented broths, Kupchan partition was used to generate more fractions with different molecular profiles. Having a wide polarity range covered by different solvents used in the partition would increase the chance of finding the compounds responsible for activity.

The tests with fractions from Kupchan partition and the whole filtered broth presented activity against all strains tested (Table 3.3, Figures 3.1-3.2). Many different fractions had detectable activity, making me wonder if a single,

widespread in fractions or multiple compounds were responsible for bacterial growth inhibition.

The most susceptible strain tested was *Pseudomonas aeruginosa* PA01, with eight fractions active, plus two unprocessed broths, followed by *Bacillus subtilis* 168, with two partitions active and one unprocessed broth. *E. coli* BW25113 was susceptible against only one fraction.

Table 3.3: susceptibility of strains tested against cave bacterial extracts and fractions obtained by Kupchan partition

Strain	Susceptibility
<i>Pseudomonas aeruginosa</i> PA01	ICC1 unprocessed broth
	#90 hexanes fraction
	#90 chloroform fraction
	#90 n-butanol fraction
	BW1 unprocessed broth
	BW1 hexanes fraction
	BW1 chloroform fraction
	BW1.2 hexanes fraction
	BW1.2 chloroform fraction
	BW1.2 n-butanol fraction
<i>Bacillus subtilis</i> 168	#90 unprocessed broth
	#90 hexanes fraction
	ICC1 aqueous fraction
<i>E. coli</i> BW25113	ICC1 n-butanol fraction

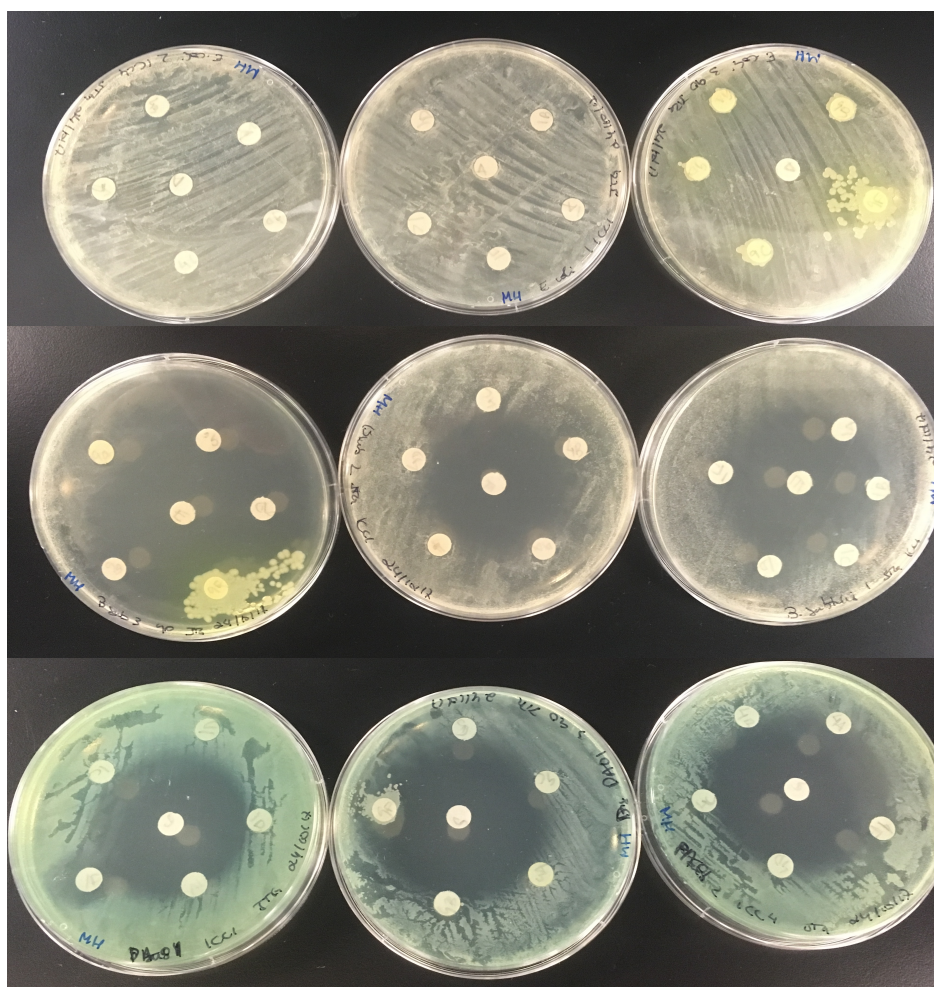


Figure 3.1: Disk diffusion assays of bacterial extracts. Row 1: *E. coli* BW25113. Row 2: *Bacillus subtilis* 168. Row 3: *Pseudomonas aeruginosa* PA01.

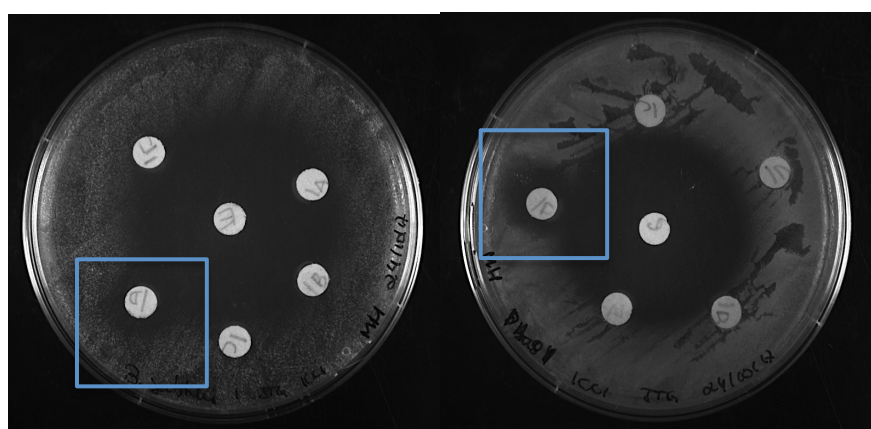


Figure 3.2: Inhibition halo in detail. Left: *B. subtilis* 168, inhibited by ICC1 aqueous fraction. Right: *Pseudomonas aeruginosa* PA01 inhibited by ICC1 unprocessed broth.

Whole Genome sequencing and assembly

Preliminary 16S sequencing of cave strains indicated the genus each strain is related to, which is supported by morphology and growth characteristics of the bacteria. The Iron Curtain Cave strains were identified as belonging to the genus *Streptomyces*, and all 5 Timber strains were associated to *Pseudomonas*. The bacteria collected from the bat's wings were not associated to any strain present in the database.

Whole genome sequencing and de novo genome assembly by PATRIC (Wattam, A.R. et al., 2017) generated draft genomes, which were aligned with BLAST database. Sequence coverage was relatively low, ranging from 53 to 82%, with an identity from 90 to 99 %. The alignment results are shown on Table 3.4.

Table 3.4: 16S rRNA, whole genome sequencing (WGS), coverage and identity percentage of BLAST alignment of cave bacteria genomic DNA sequences.

<i>16S sequencing</i>	<i>WGS</i>	Coverage	Identity
		(%)	(%)
<i>Iron Curtain Cave</i>	<i>Streptomyces</i> sp. <i>Mg1</i>	53	91
<i>2 Actinobacteria</i>	<i>Streptomyces</i> sp. <i>3211</i>	55	94
<i>Bat's wings</i>	<i>Pseudomonas fragi</i> <i>P21</i>	71	99
<i>2 unidentified strains</i>	<i>Pseudomonas fluorescens</i> <i>Pt14</i>	80	96
<i>Timber</i>	<i>Pseudomonas simiae</i>	82	90
<i>5 Pseudomonas</i> sp.	<i>Pseudomonas yamanorium</i> <i>LMG 27247</i>	82	97

The 9,034,309 bp for ICC1 and 9,010,404 bp for ICC4, with a G+C content of 72 % for both strains. The ICC1 genome was assembled into 726 contigs and the ICC4 genome into 734 contigs with N50s of 22,631 and 23,307 respectively. RAST annotation predicted 8235 coding sequences for ICC1 and 8243 coding sequences for ICC4.

Functional profiling and analysis of ICC1 and ICC4

Genomic comparison files were generated online at WebACT (Abbott, J. C. et al., 2005) and functional analysis was performed using Artemis Comparison Tool (ACT) (Carver, T.J. et al., 2005). Strains were aligned using BLAST to compare genomic profiles between each other and with other genomes on the database. The closest genome match to both ICC1 and ICC4 was *Streptomyces lavendulae* strain CCM 3239, with alignments of 54 % of the sequence for both ICC1 and ICC4 and identities of 90 % within both of the alignment (Figure 3.3). When compared to each other, the genomes of ICC1 and ICC4 showed 50 % coverage with 99 % identity, indicating that the two strains are highly related.

Analysis of the features the cave *Streptomyces* genomes with those of *S. lavendulae* using RAST annotation revealed very similar predicted protein content and function between the organisms (Figure 3.4). *Streptomyces* sp. ICC1 and *Streptomyces* sp. ICC4 have a total of 450 subsystems while *S. lavendulae* has 435, indicating all three strains likely possess similar specific biological processes and structural complexes (Overbeek, A. et al., 2005).

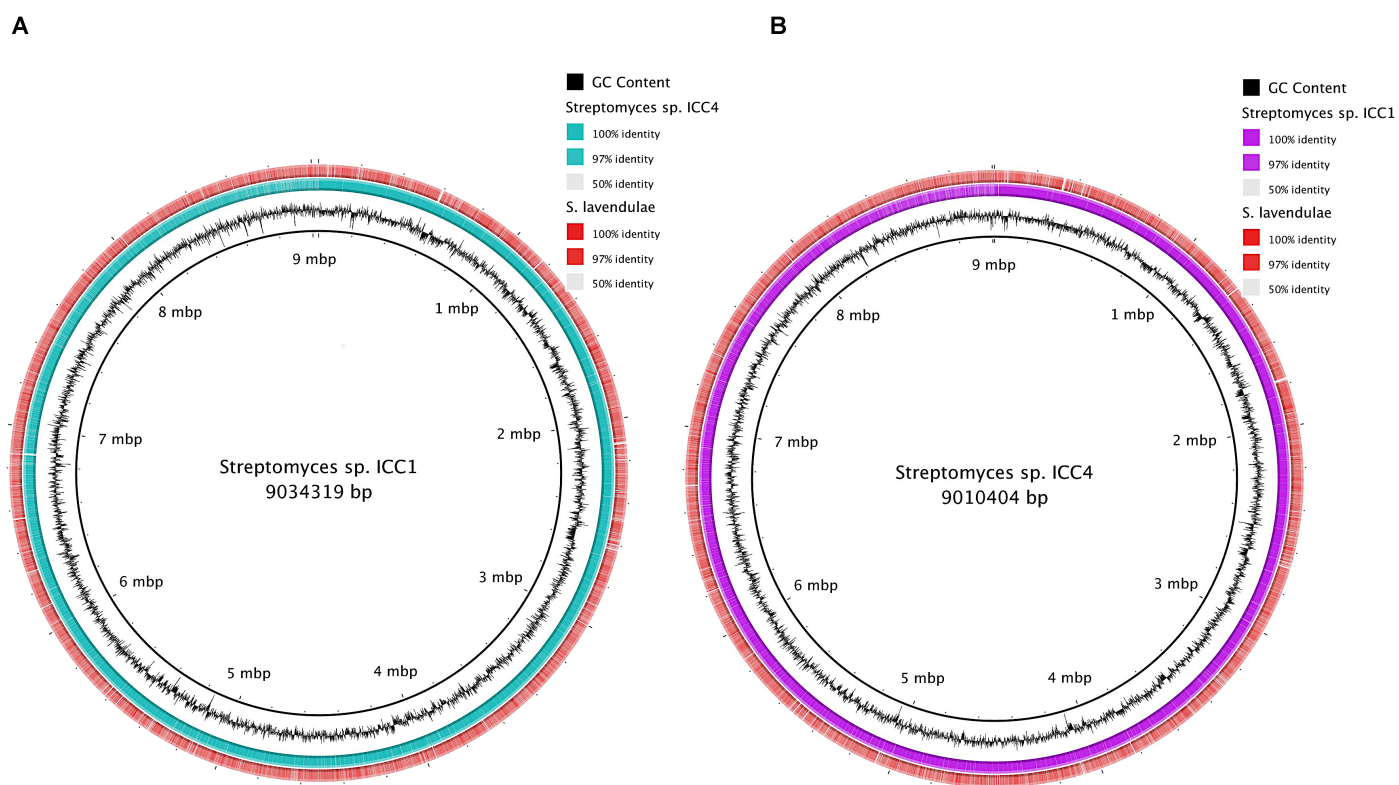


Figure 3.3: Genome alignment of genomes of *Streptomyces* sp. A: alignment of *Streptomyces* sp. ICC1 against *Streptomyces* sp. ICC4 and *Streptomyces lavendulae* strain CCM 3239. B: alignment results of *Streptomyces* sp. ICC4 against *Streptomyces* sp. ICC1 and *Streptomyces lavendulae* strain CCM 3239.

Examination of the subsystem feature counts shows that all three strains have similar numbers of genes in each subsystem, with *S. lavendulae* showing more features involved in secondary metabolite production (22) than the cave strains (12). These data are strongly supportive of all three strains being highly related.

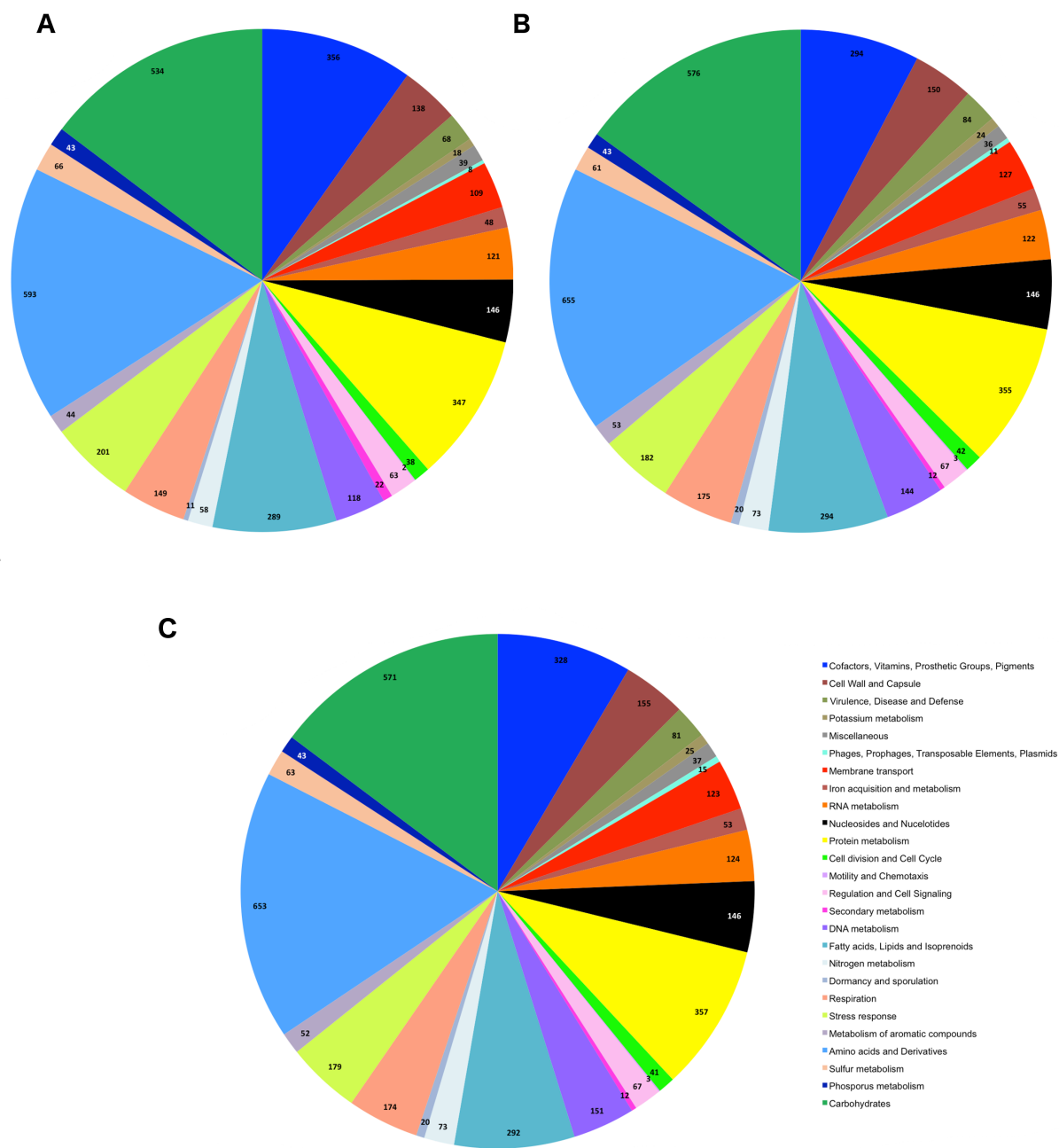


Figure 3.4: Genome features of *Streptomyces* sp. A: *Streptomyces lavendulae* strain CCM 3239 B: *Streptomyces* sp. strain ICC1 C: *Streptomyces* sp. isolate ICC4. Numbers indicate genes per feature group.

Secondary metabolite profiling and genome mining

The genomes of ICC1 and ICC4 were analyzed for the presence of secondary metabolite biosynthetic gene clusters using antiSMASH 4.1.0 and BAGEL4 (van Heel, A. J. et al., 2018). AntiSMASH 4.1.0 uses hidden Markov models and rules based-detection to identify a broad array of biosynthetic gene clusters, including those encoding polyketide, non-ribosomal peptides, terpenes, aminoglycosides and ribosomally synthesized and post-translationally modified peptides (RiPPs) from bacterial genomes. Similarly BAGEL4 uses HMM to detect core RiPP encoding genes however it is independent from the genome ORF calls, enabling it to better detect the small precursor peptides found in RiPP encoding gene clusters.

AntiSMASH predicted 37 putative secondary metabolite clusters for ICC1 and 35 for ICC4 (Table 3.5). The majority of these results were terpenes, and the overall similarity with known natural products was low (under 20 %). Both strains have in common a non-ribosomal peptide biosynthetic pathway with 100% similarity to coelichelin, a ferric iron-chelating peptide (Ziemert, N., Alanjary, M., & Weber, T., 2016) and two lantipeptides, the morphogenic peptide SapB (Marsh, A.J. et al., 2010) and an unidentified molecule (due to the lack of a predicted peptide core). Strain ICC4 has a unique thiopeptide prediction, with very low similarity with rifamycin (5 %).

Table 3.5: summary of antiSMASH 4.1.0 secondary metabolites prediction of *Streptomyces* sp. strains ICC1 and ICC4

Metabolite Class	ICC1 Clusters	ICC4 Clusters	Most similar pathway (%)
PKS I	1	2	Meilingmycin (3%)
PKS II	2	2	Simocyclinone (66%)
PKS III	1	1	Alkylresorcinol (100%)
NRPS	5	3	Coelichelin (100%)
Siderophore	2	-	Salinomycin (4%)
Lantipeptide	2	2	SapB (100%)
Butyrolactone	4	3	Calcimycin/Salinomycin (10%)
Thiopeptide	-	1	Rifamycin (5%)
Melanin	2	2	Melanin (100%)
Bacteriocin	-	2	-
Terpene	10	10	2-methylisoborneol (100%)
Mixed	3	2	Kinamycin (8%)
Other	4	4	BE-14106 (14%)
TOTAL	37	35	

To further characterize the unidentified lantipeptide, both genomes were analyzed with BAGEL (Hart, T. and Moffat, J., 2016), a bioinformatic tool that specializes in detection and annotation of RiPPs such as lantipeptides. BAGEL predicted the same lantipeptides identified by antiSMASH and did not provide further insight into the unidentified lantipeptides. BAGEL predicted an additional RiPP gene cluster in ICC1 and ICC4, which was assigned as the glyocin biosynthetic gene cluster based on sequence similarity. Interestingly the genes for glyocin production were observed within the melanin clusters in both cave

strains and seem to be involved in production and glycosylation of bacteriocins (Aydillo C. et. al., 2014).

A detailed manual analysis of the genomes for biosynthetic gene cluster prediction revealed two independent biosynthetic gene clusters containing the pyoverdinin synthesis genes and an NRPS that was not identified by antiSMASH prediction from 6371985 bp to 6374537 bp in the *Streptomyces* sp. ICC1 genome.

The NRPS alignment detected a *Streptomyces* multispecies synthase with 99% coverage and 81% similarity. A PKS-NRPS mixed cluster was also identified from 6080168 – 6149842 bp, however this is likely a false positive as the BLAST alignment and annotation shows the protein coding sequences as belonging to an uncharacterized protein. The assignment was further confirmed by UniProt and Pfam analysis.

Interestingly, while the overall genomes of ICC1, ICC4 and *S. lavendulae* are highly similar, ICC1 and ICC4 have approximately 40 % more biosynthetic gene clusters than *S. lavendulae* (37 clusters for ICC1, 35 for ICC4, and 26 for *S. lavendulae*). These observations are consistent with little to no adaptation of the hypogean strains versus their terrestrial counterpart and are consistent with a limited restriction on dispersal in structuring microbial communities in the Iron Curtain cave.

Highly related strains that differ in their secondary metabolite potential can exhibit functional adaptation. For example the marine Actinomycetes *Salinispora tropica* and *Salinispora arenicola* share 87.2 % identity among their 3606 orthologs. Their genomes differ primarily in 21 genomic islands, many of which are enriched with large clusters of genes devoted to the biosynthesis of

secondary metabolites (Penn, K. et al, 2009). In total *S. tropica* was found to have 19 biosynthetic gene clusters and *S. arenicola* 30. Biogeographical characterization showed *S.arenicola* to have a cosmopolitan distribution whereas *S. tropica* was shown to be highly restricted in its distribution (Jensen and Mafnas, 2006). The correlation between enhanced biogeographical distribution and enhanced secondary metabolite potential in highly related Actinomycetes suggest that this functional trait may be a driver of ecological diversification in closely related species and may account for the fitness of ICC1 and ICC4 versus *S. lavendulae* in the Iron Curtain cave.

In terms of activity against *P. destructans*, preliminary assays showed no activity against the pathogen. Considering that cultivation methods are restrictive and many pathways can remain silent under these conditions, genome mining would be a better tool to retrieve any missed cluster, but no predicted compound could be promptly related to possess an antifungal role. Still, the strains remain a valuable source to yield interesting secondary metabolites.

Correlating the Metabolome and genome

To characterize the metabolic profile of these two strains, they were cultivated in V8 media for 30 days at 15 °C and the cultures were processed by Kupchan partition to generate four fractions per strain. LC-ESI-HRMS analysis on a LTQ Orbitrap XL Hybrid Ion Trap-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Waltham, Massachusetts, USA) enable the identification of a large number of metabolites from the fractions that showed antimicrobial activity.

In total 67 compounds (Appendix 1) were identified through GNPS Molecular Networking (Wang, M. et al., 2017), the majority of spectra matching with submitted spectra from Bronze datasets in GNPS.

Surprisingly for *Streptomyces* strains, relatively few polyketide and non-ribosomal peptide biosynthetic gene clusters were detected in the genomes. The most commonly associated secondary metabolite pathway were those annotated as terpene gene clusters. The overall metabolite profile shows a large number of lipids and aromatics such as 5-aminovaleric acid, palmitamide and the glucoside DIMBOA. Several cyclic and acyclic dipeptides were identified, with proline, histidine and isoleucine being the most common residues.

Some metabolites could not be readily correlated to biosynthetic gene clusters. For example GNPS spectral analysis clearly identified a number of diketopiperazine products from ICC1. Cyclo (L-Leu- L-Pro), cyclo (L-Pro-L-Val) and cyclo (L-(4-hydroxyPro)-L-Leu) (Figure 3.5) were all identified with low mass error and matched to Bronze datasets in the GNPS database.

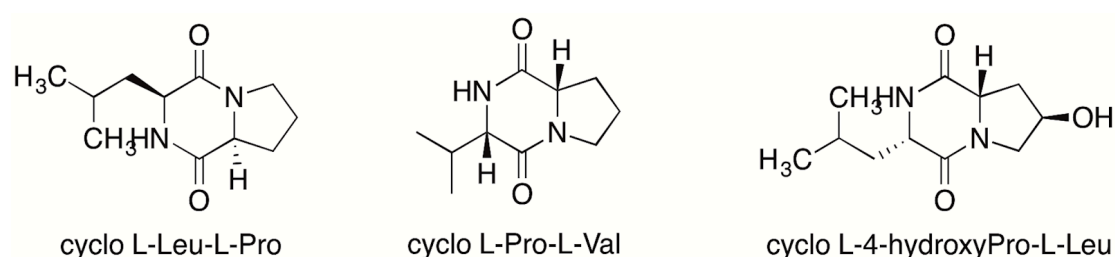


Figure 3.5: Diketopiperazines identified from ICC1 by GNPS analysis of LC-ESI-HRMS analysis of culture extracts

Cyclic peptide biosynthesis can occur via dedicated cyclodipeptide synthases which link and cyclize the two amino acids from two aminoacyl-tRNAs generate the diketopiperazine (Belin et al. 2012), or via nonribosomal peptide synthetases (NRPSs). In our study the genetic origin of the observed diketopiperazines is still unclear, but the lack of a discrete cyclodipeptide synthase and the presence of apparent incomplete or short NRPS encoding gene suggests these may be NRPS-derived metabolites.

Diketopiperazines are commonly seen in bacteria, though were often disregarded as by-products (Da Silva et al.,2017) due to their small size and lack of complexity. With an array of activities, ranging from communication and quorum-sensing to selective antifungal and antibacterial molecules (Belin, P. et al.,2012), they clearly can play an important role in the biology of ICC1 and ICC4. As these diketopiperazines are known to exhibit antifungal activity and likely play a role in the chemical ecology of this cave isolates, future identification of their biosynthetic gene clusters is an important task.

In some cases a biosynthetic gene cluster could be unambiguous identified and assigned to a known compound but that metabolite could not be identified from the culture extracts. For example, antiSMASH identified a NRPS trimodular gene cluster with 100% similarity to the tetrapeptide siderophore coelichelin from the genomes of ICC1 and ICC4 (Figure 3.6). None of the fractions analyzed by LC-ESI-HRMS showed evidence for the presence of coelichelin.

The coelichelin gene cluster was first sequenced and characterized from *Streptomyces coelicolor* (Lautru et al., 2006), a soil-dwelling bacteria responsible for breaking down various organic matter components under variable environmental conditions. As many siderophores are produced only under iron-

limiting circumstances (Bagg, A. and Neilands, J.B., 1987) and *Streptomyces* sp. strains ICC1 and ICC4 were cultivated under iron rich conditions, it is possible that the gene clusters remained silent and unexpressed under our experimental conditions.

Gene clusters encoding siderophore biosynthesis are commonly found in bacterial genomes. Iron is an essential nutrient for all bacteria. However due to its low abundance in some environments, siderophore, secondary metabolites with high affinity for iron, are often produced by the bacteria to scavenge iron from the environment. While the Iron Curtain cave is rich in iron deposits, ferric iron has low solubility (10^{-17} M at pH 7) in aqueous environments (Andrews et al., 2003). Thus coelichelin, the siderophore whose biosynthetic gene cluster was identified in both ICC1 and ICC4, may play a role in resolublizing iron for uptake. Intriguingly, coelichelin may also play a role in Zinc homeostasis as the biosynthesis of the highly homologous compound coelibactin in *S. coelicolor* is regulated by Zn(II) levels (Hesketh et al., 2009).

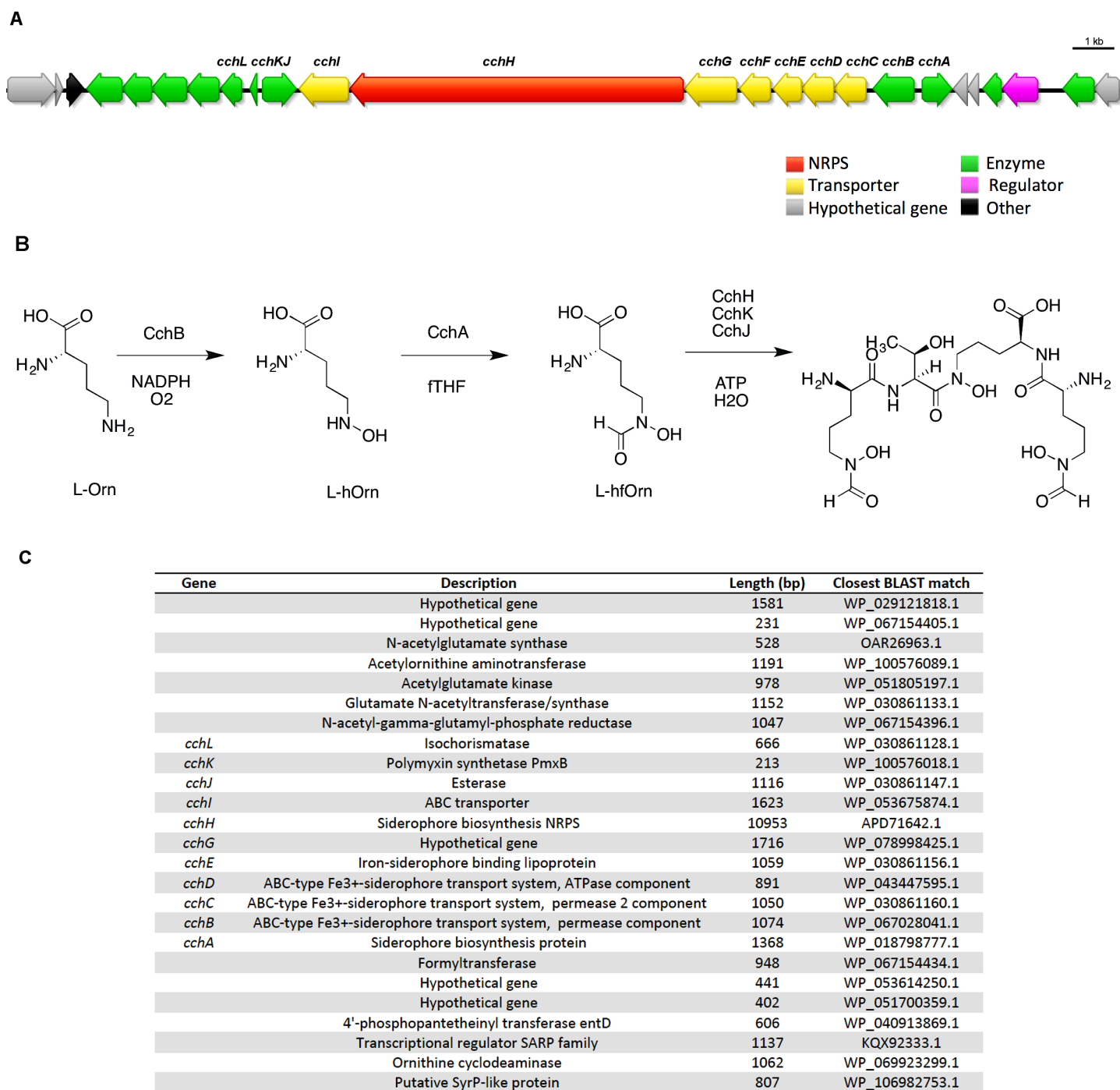


Figure 3.6: Coelichelin, one of the secondary metabolites found in *Streptomyces* sp. cave isolates. A) biosynthetic gene cluster for coelichelin. B) proposed mechanism for biosynthesis. C) list of genes in the cluster, their annotation, length and similarity based on BLAST library alignment.

A number of metabolites identified by LC-ESI-HRMS followed by GNPS analysis could be correlated with biosynthetic gene clusters identified by antiSMASH. In some cases the gene clusters from ICC1 and ICC4 showed high levels of identity to known biosynthetic gene clusters, in others gene cluster identification was done based on the expected biosynthesis of the compound and the presence of the requisite genes required to install that functionality in the gene cluster. Two examples are described below.

2', 5' – dimethoxyflavone and nordentatin

2', 5' – dimethoxyflavone and nordentatin, both type III polyketides were identified by GNPS from the LC-ESI-HRMS data of the aqueous and n-butanolic fractions of *Streptomyces* sp. ICC1, respectively. A biosynthetic gene cluster likely encoding these type III polyketides was identified from the antiSMASH predictions for both strains. A manual genome annotation analysis identified this as the only type III PKS gene cluster in the genome, consistent with the antiSMASH results.

There is very little information about both molecules in the literature, with nordentatin being described as possessing antibacterial activities against Gram-positive and Gram-negative strains (Wu, T-S. and Furukawa, H., 1982). Based on the identified gene cluster and the structure of the metabolites, we provide a putative biosynthesis of both compounds from this single gene cluster (Figure 3.7).

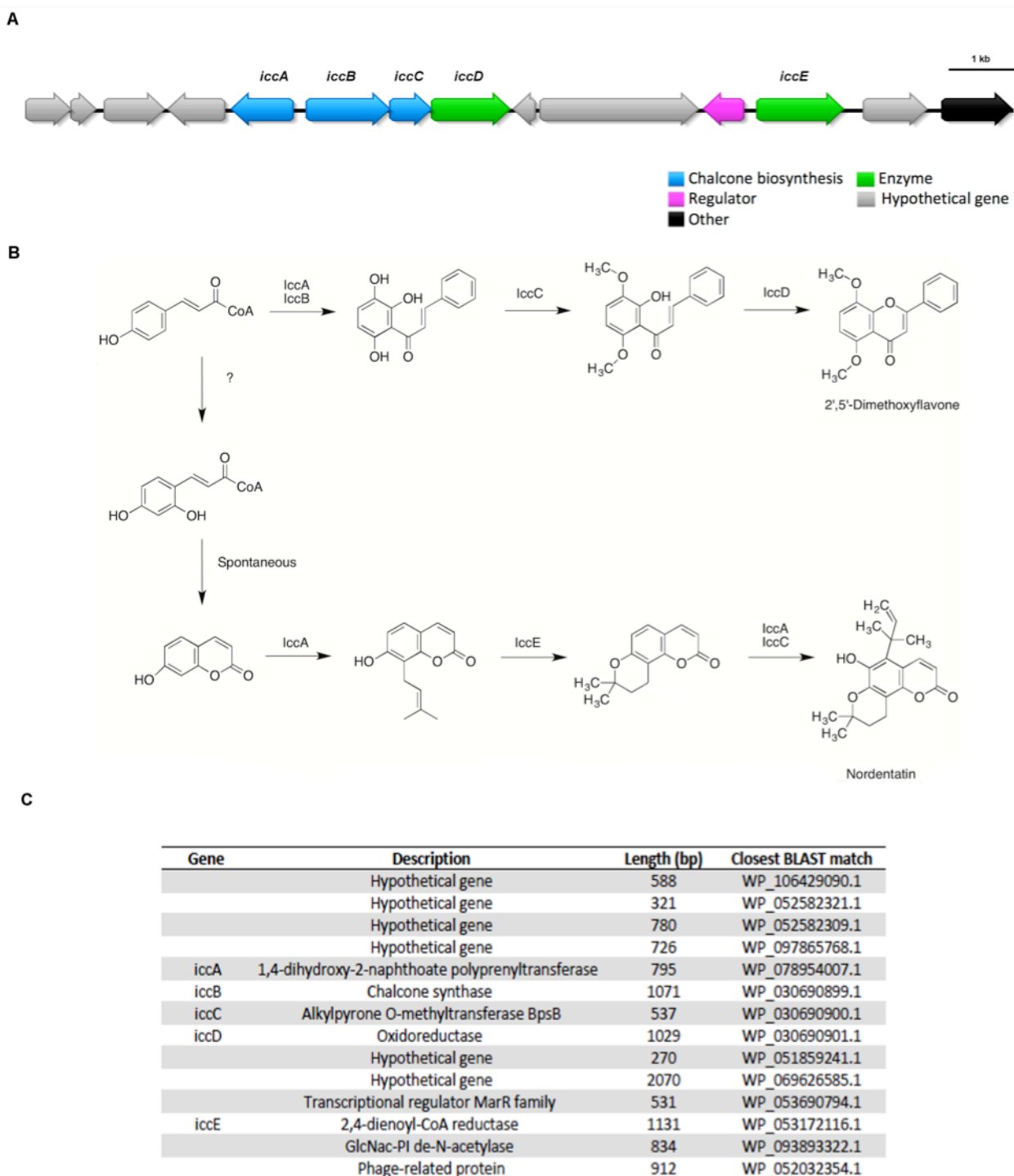


Figure 3.7. The type III polyketides 2', 5' - dimethoxyflavone and nordentatin. The molecules were identified in metabolomic analysis of culture broth and a potential putative gene cluster encoding these two compounds was identified by antiSMASH analysis of the genomes of both ICC1 and ICC4. A) biosynthetic gene cluster for both type III polyketides. B) proposed mechanism for their biosynthesis. C) list of genes in the cluster, annotation, length and similarity based on BLAST library alignment

Diazepinomicin

One of the terpenes gene clusters predicted by antiSMASH possesses a seven-gene pathway that is consistent with diazepinomicin biosynthesis. Diazepinomicin contains a very rare core structure and its biosynthetic gene cluster, to date, has only been identified from the genome of the sponge-associated actinomycete *Micromonospora* sp. DPJ12 genome (McAlpine, J.B. et al., 2008). Diazepinomicin possesses modest antibacterial activity against Gram positive strains (Charan *et. al.*, 2004).

LC-ESI-HRMS analysis of the n-butanolic fraction of *Streptomyces* sp. ICC1 identified the presence of the diazepinomicin core structure. A putative biosynthesis mechanism based on the genes identified in the genome and the proposed biosynthesis by McAlpine J.B. et al. (2008) is shown on Figure 3.8. Key aspects of this assignment are DkpF, DkpD, and DkpE, which we propose are responsible for formation of 3-hydroxyanthranilic acid one of two key building blocks for diazepinomicin biosynthesis. This is fully consistent with the McAlpine identified biosynthetic pathway. Similarly DkpG, DkpC and one of the oxidases encoded in the pathway are proposed to be responsible for conversion of 3-amino-5-hydroxybenzoic acid into 2-amino-6-hydroxy [1,4]benzoquinone, the other a key building. This is also fully consistent with the *Micromonospora* biosynthetic mechanism.

While the ICC1 and ICC4 gene clusters encode the key functionality for diazepinomicin biosynthesis, cluster prediction was non-obvious for diazepinomicin as it differed substantially from the characterized

Micromonospora gene cluster. This suggests that they may have evolved separately.

This gene cluster discrepancy highlights the importance of manual checking annotation of genome mining as well as the potential for highly divergent gene clusters encoding the identical molecule particularly in the case where one of the organisms is a facultative or obligate symbiont.

The similarities in molecule production and genome shared by these two strains could be an indicative of a cooperative mode of living in extreme habitats instead of a competitive one. In challenging and nutrient-limited environments, organisms can assume functions and roles according to bacterial community structure to overcome the unfavourable metabolic panorama. Association between organisms in microbial communities is shown to be essential to the evolutionary dynamics of the environment and optimization of bioprocesses (Cavaliere et al., 2017).

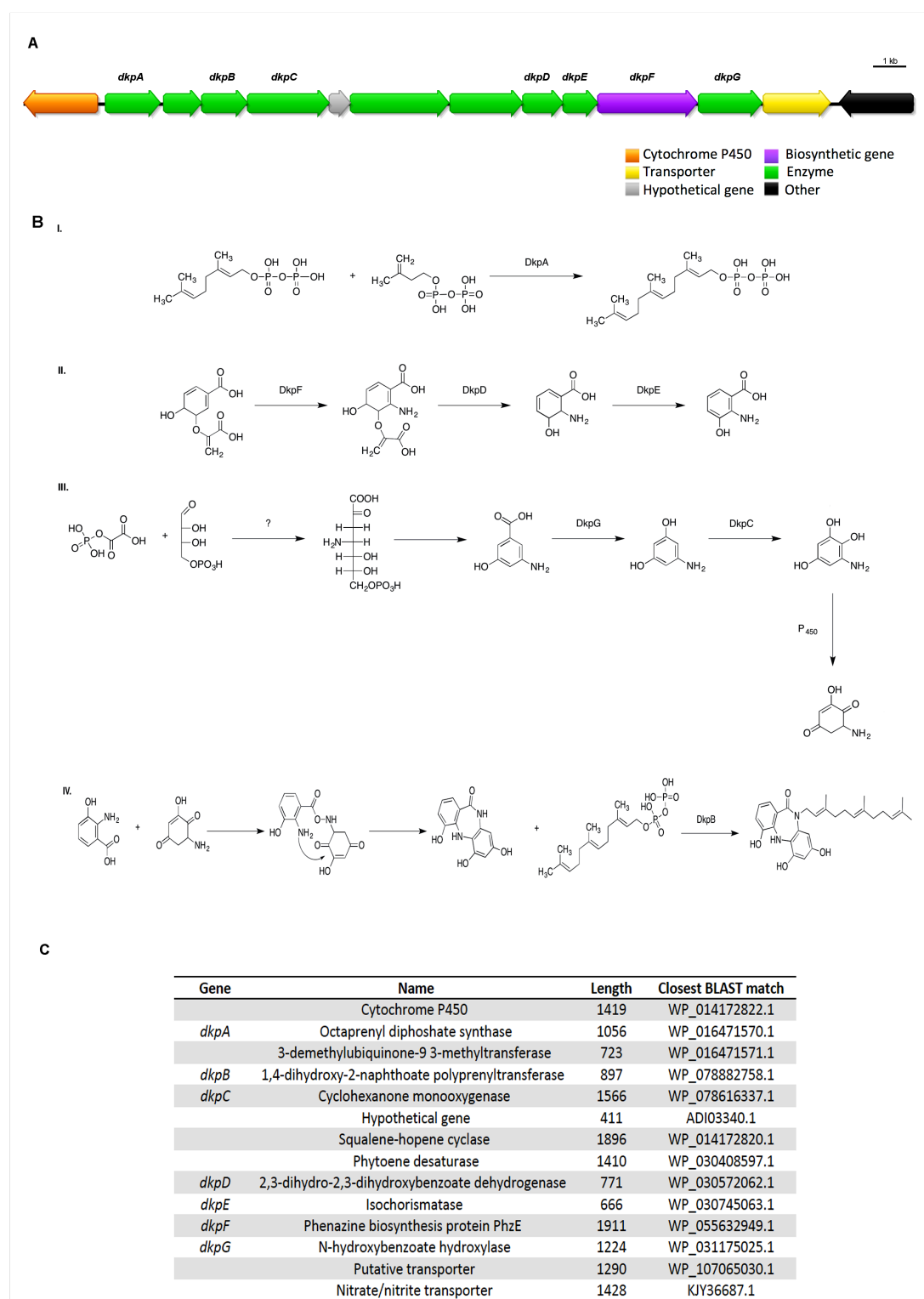


Figure 3.8: Proposed diazepinomicin biosynthetic pathway from ICC1 and ICC4. Diazepinomicin was detected in one of the fractions from ICC1. A) biosynthetic gene cluster diazepinomicin. B) proposed mechanism for diazepinomicin biosynthesis. C) list of genes in the cluster, annotation, length and similarity based on BLAST library alignment.

This synchrony was seen in bacteria living under the Earth's surface in mines, which were shown to work with each other to make conditions favourable in a cross-feeding arrangement (Ravindran, S. 2017). It is also reported that bacteria can cooperate based on "public good" molecules, which benefit an entire population and go far from the competition model usually studied. These associations are based on a large variety of shared molecules: siderophores, enzymes, biosurfactants, components of biofilm matrix, quorum sensing molecules, bacteriocins, etc. (Cavaliere et al, 2017). Considering that the main compounds seen on genomic mining and molecular profiling of our cave isolates, collaboration seems to be the preferred mode of interaction between these organisms in an environment with few resources and slow changing.

Lastly it is unexpected that both cultivated strains from the Iron Curtain cave are so highly similar, though not clonally identical. This could represent a limited microbial diversity of the cave, or limited diversity of cultivatable *Streptomyces*. In either case a more detailed analysis of the bacterial community structure in the Iron Curtain cave is warranted.

Conclusions

Caves are environments characterized by low levels of nutrients and limited physical conditions. Nevertheless these unique habitats harbour a diverse microorganism community that can yield a multitude of unique secondary metabolites. Our work examines and correlates the genomes and metabolomes of two culturable *Streptomyces* isolates from the Iron Curtain cave.

We show a successful genome sequencing, assembly and annotation. With this information we were able to demonstrate that the genomes of these isolates are highly homologous to a known terrestrial *Streptomyces* species, *S. lavendulae*. A key area of difference between the ICC isolates and *S. lavendulae* are the secondary metabolites encoded in the genome.

We examine both through sequence analysis and LC-ESI-HRMS detection a number of these metabolites and discuss their potential role in enabling their fitness in the Iron Curtain cave. Although no direct correlation could be made between these molecules and an antimicrobial compound, we can deduce their role in a cooperation approach to ensure survival rather than competing.

While everything is everywhere, the cave environment does not seem to provide a significant restriction to distribution, enabling the environment to select these strains with enhanced secondary metabolite biosynthesis capability.

Chapter Bibliography

Abbott, J. C.; Aanensen, D. M.; Rutherford, K.; Butcher, S.; B Spratt, G. (2005). WebACT—an online companion for the Artemis Comparison Tool, *Bioinformatics*, 21(18), 3665–3666.

Aydillo, C., Compañón, I., Avenoza, A., Busto, J. H., Corzana, F., Peregrina, J. M., & Zurbano, M. M. (2014). S-Michael additions to chiral dehydroalanines as an entry to glycosylated cysteines and a sulfa-tn antigen mimic. *Journal of the American Chemical Society*, 136(2), 789–800.

Bagg, A., & Neilands, J. B. (1987). Molecular mechanism of regulation of siderophore-mediated iron assimilation. *Microbiological Reviews*, 51(4), 509–518.

Barton, H.A. (2006). Introduction to cave microbiology: A review for the non-specialist. *Journal of Cave and Karst Studies*, 68, 43–54.

Barton, H.A. (2015). "Starving Artists: Bacterial Oligotrophic Heterotrophy in Caves," in *Microbial Life of Cave Systems*, ed. A.S. Engel. (Germany: Walter de Gruyter), 79–95.

Barton, H.A., and Northup, D.E. (2007). Geomicrobiology in cave environments: past, current and future perspectives. *Journal of Cave and Karst Studies*, 69, 163–178.

Bastian, F., Alabouvette, C., Jurado, V., & Saiz-Jimenez, C. (2009). Impact of biocide treatments on the bacterial communities of the Lascaux Cave. *Naturwissenschaften*, 96(7), 863–868.

Blehert, D. S., Hicks, A. C., Behr, M., Meteyer, C. U., Berlowski-zier, B. M., Buckles, E. L., ... Stone, W. B. (2009). Bat White-Nose Syndrome : An emerging fungal pathogen? *Science*, 323(5911), 9–10.

Boire, N., Zhang, S., Khuvis, J., Lee, R., Rivers, J., Crandall, P., Keel, M. K., ... Parrish, N. (2016). Potent Inhibition of *Pseudogymnoascus destructans*, the causative agent of white-nose syndrome in bats, by cold-pressed, terpeneless, valencia orange oil. *PloS one*, 11(2), e0148473.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G., & Parkhill, J. (2005). ACT: The Artemis comparison tool. *Bioinformatics*, 21(16), 3422–3423.

Chaturvedi, S., Rajkumar, S. S., Li, X., Hurteau, G. J., Shtutman, M., & Chaturvedi, V. (2011). Antifungal testing and high-throughput screening of compound library against *Geomyces destructans*, the etiologic agent of geomycosis (WNS) in bats. *PloS one*, 6(3), e17032.

Cheeptham, N. (2013). "Advances and challenges in studying cave microbial diversity," in *Caves microbiome: A novel resource for drug discovery*, ed. N. Cheeptham. Springer Briefs in Microbiology, 1-34.

Cheeptham, N., Sadoway, T., Rule, D., Watson, K., Moote, P., Soliman, L.C., Azad, N., Donkor, K.K., and Horne, D. (2013). Cure from the cave: volcanic cave actinomycetes and their potential in drug discovery. *International Journal of Speleology*, 42, 35-47.

Clinical and Laboratory Standards Institute (2012). Approved standard. NCCLS document M02-A11. Clinical and Laboratory Standards Institute, Villanova, Pa.

Cornelison, C. T., Gabriel, K. T., Barlament, C., & Crow, S. A. (2014). Inhibition of *Pseudogymnoascus destructans* Growth from Conidia and Mycelial Extension by Bacterially Produced Volatile Organic Compounds. *Mycopathologia*, 177(1), 1–10.

Field, K.A., Sewall, B.J., Prokkola, J.M., et al (2018). Effect of torpor on host transcriptomic responses to a fungal pathogen in hibernating bats. *Molecular Ecology*, 27(18), 3727–3743.

Frick, W. F., Pollock, J. F., Hicks, A. C., Langwig, K. E., Reynolds, D. S., Turner, G. G., ... Kunz, T. H. (2010). An Emerging Disease Causes Regional Population Collapse of a Common North American Bat Species. *Science*, 329(5992): 679–683.

Ghosh, S., Kuisiene, N., & Cheeptham, N. (2017). The cave microbiome as a source for drug discovery: Reality or pipe dream? *Biochemical Pharmacology*, 134, 18–34.

Ghosh, S., Paine, E., Wall, R., Kam, G., Lauriente, T., Sa-Ngarmangkang, P.C., Horne, D., and Cheeptham, N. (2017). In Situ Cultured Bacterial Diversity from Iron Curtain Cave, Chilliwack, British Columbia, Canada. *Diversity*, 9, 1-15.

Hart, T., & Moffat, J. (2016). BAGEL: A computational framework for identifying essential genes from pooled library screens. *BMC Bioinformatics*, 17(1), 1–7.

He, L., Diedrich, J., Chu, Y. Y., & Yates, J. R. (2015). Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. *Analytical chemistry*, 87(22), 11361-7.

Kupchan, S.M., Tsou, G., Sigel, C.W. (1973). Datiscacin, a novel cytotoxic cucurbitacin 20-acetate from *Datisca glomerata*. *Journal of Organic Chemistry*, 38(7):1420–1.

Lautru, S., Deeth, R. J., Bailey, L. M., & Challis, G. L. (2005). Discovery of a new peptide natural product by streptomyces coelicolor genome mining. *Nature Chemical Biology*, 1(5), 265–269.

Lorch, J. M., Muller, L. K., Russell, R. E., O'Connor, M., Lindner, D. L., & Blehert, D. S. (2013). Distribution and environmental persistence of the causative agent of

white-nose syndrome, *Geomyces destructans*, in bat hibernacula of the eastern United States. *Applied and environmental microbiology*, 79(4), 1293-301.

Maciejewska, M., Adam, D., Martinet, L., Naome, A., Calusinska, M., Delfosse, P., Carnol, M., Barton, H.A., Hayette, M.P., Smargiasso, N., De Pauw, E., Hanikenne, M., Baurain, D., and Rigali, S. (2016). A Phenotypic and Genotypic Analysis of the Antimicrobial Potential of Cultivable *Streptomyces* Isolated from Cave Moonmilk Deposits. *Frontiers in Microbiology*, 7,1455.

Maciejewska, M., Adam, D., Naome, A., Martinet, L., Tenconi, E., Calusinska, M., Delfosse, P., Hanikenne, M., Baurain, D., Compere, P., Carnol, M., Barton, H.A., and Rigali, S. (2017). Assessment of the Potential Role of *Streptomyces* in Cave Moonmilk Formation. *Frontiers in Microbiology*, 8,1181.

Marsh, A. J., O'Sullivan, O., Ross, R. P., Cotter, P. D., & Hill, C. (2010). *In silico* analysis highlights the frequency and diversity of type 1 lantibiotic gene clusters in genome sequenced bacteria. *BMC Genomics*, 11, 679.

McAlpine, J. B., Banskota, A. H., Charan, R. D., Schlingmann, G., Zazopoulos, E., Pirae, M., Janso, J., Bernan, V. S., Aoudate, M., Farnet, C. M., Feng, X., Zhao, Z., Carter, G. T. (2008) Biosynthesis of diazepinomicin/ECO-4601, a *Micromonospora* secondary metabolite with a novel ring system. *Journal of Natural Products*, 71, 1585-1590.

Overbeek, R., Begley, T., Butler, R. M., Choudhuri, J. V., Chuang, H. Y., Cohoon, M., de Crécy-Lagard, V., Diaz, N., Disz, T., Edwards, R., Fonstein, M., Frank, E. D., Gerdes, S., Glass, E. M., Goesmann, A., ... Stevens, R., Thiele, I., Vassieva, O., Ye, Y., Zagnitko, O., Vonstein, V. (2005) The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research*, 33, 5691-5702.

Palmer, J. M., Drees, K. P., Foster, J. T., & Lindner, D. L. (2018). Extreme sensitivity to ultraviolet light in the fungal pathogen causing white-nose syndrome of bats. *Nature Communications*, 9(1), 35.

Urzi, C., De Leo, F., Bruno, L., & Albertano, P. (2010). Microbial Diversity in Paleolithic Caves: A Study Case on the Phototrophic Biofilms of the Cave of Bats (Zuheros, Spain). *Microbial Ecology*, 60(1), 116–129.

Verant, M. L., Meteyer, C. U., Speakman, J. R., Cryan, P. M., Lorch, J. M., & Blehert, D. S. (2014). White-nose syndrome initiates a cascade of physiologic disturbances in the hibernating bat host. *BMC Physiology*, 14(1), 10.

Verant, M. L., Bohuski, E. A., Richgels, K. L. D., Olival, K. J., Epstein, J. H., & Blehert, D. S. (2018). Determinants of *Pseudogymnoascus destructans* within bat hibernacula: Implications for surveillance and management of white-nose syndrome. *Journal of Applied Ecology*, 55(2), 820–829.

Wang, M., Carver, J. J., Phelan, V. V., Sanchez, L. M., Garg, N., Peng, Y., ... Bandeira, N. (2016). Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nature Biotechnology*, 34(8), 828–837.

Wattam, A. R., Davis, J. J., Assaf, R., Boisvert, S., Brettin, T., Bun, C., ... Stevens, R. L. (2017). Improvements to PATRIC, the all-bacterial bioinformatics database and analysis resource center. *Nucleic Acids Research*, 45(D1), D535–D542.

Weber, T., Blin, K., Duddela, S., Krug, D., Kim, H. U., Bruccoleri, R., ... Medema, M. H. (2015). antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. *Nucleic Acids Research*, 43(Web Server issue), W237–W243.

Wu, T.-S., & Furukawa, H. (1982). Biological and Phytochemical Investigation of *Clausena excavata*. *Journal of Natural Products*, 45(6), 718–720.

Ziemert, N., Alanjary, M., & Weber, T. (2016). The evolution of genome mining in microbes - a review. *Natural Products Report*, 33(8), 988–1005.

Final remarks

With this work my goal was to show bioprospecting as a versatile process with great potential for novel secondary metabolite discovery, taking a different way from the activity-based approach. The use of genetic and computational bases, which are much more consolidated and have fast growing databases and repositories makes the research easier as time passes, and being part of the community that contributes to its growth is of extreme importance. Another advantage of bioprospecting as a tool is that it can be modified and optimized to target different environments, organisms and molecules, as shown in this work with different levels of success. Its association with multiple different techniques gives the plasticity the molecule discovery needs to advance.

One important thing that needs to be highlighted is the audacity of this type of research. Because of its novelty, there is still a lack of information on some of these environments, organisms and molecules that make some of the basic things, such as species composition, very difficult to determine. Again, having the ability to sequence the extracted DNA we can shed some light on the ecology of different microbiomes that could benefit research as a whole. Sometimes even without a direct result, the risk-taking nature of the work developed over my PhD was worth the contribution to the field as a whole to shed light to some unexplored and underresearched places.

Not less important to mention, the addition of results produced during this research to global public access databases, such as the genomes of many of the bacterial strains discovered and analyzed, to the high regarded National

Center for Biotechnology Information (NCBI), and the three publications that primarily came from this work's results are a little bit of the scientific contribution that was possible to generate and that could be the base of many other works to come.

To close this text, I would like to highlight the importance of the development of this fast-growing field not only as a motor for curiosity of the future scientists to come, but for the humanity to find sustainable and promising solutions to some of our global problems. To all the molecular biologists and biological chemists, the world needs each one of us.

Appendix 1: GNPS Molecular Networking compound identification of n-butanol and aqueous extracts of *Streptomyces* sp. ICC1

ICCI	CCID	ICCI name	Compound Name	Compound Source	Excitaxis	Instrument	Ion Source	Library	Qualifying	M/Q	MicroPM	MassDiff	Organism	Precur_MZ	Parent_ID	SharePeak	SpecCharge	SpecMZ
M-H	M-H	74-79-3	Arginine	Commercial	174,112	Orbitrap	LC-ESI	Gold	0.88313	5.7083	0.003078	0.003078	GNPS-EMIL-XCF	391.28	N/A	6	1	175.118
M-H	M-H		Bis(2-ethylhexyl) phthalate	Isolated	381,284	QToF	LC-ESI	Bronze	0.97451	10.1272	0.003078	0.003078	MASSBANK	391.28	N/A	7	1	175.118
M-H	M-H		cyclo-B-(4-hydroxypro)-[4e]	Isolated	227.14	Ion Trap	LC-ESI	Bronze	0.80586	79.706	0.03054	0.03054	GNPS-LIBRARY	197.28	N/A	6	1	227.12
M-H	M-H		Cyclo-B-(4-hydroxypro)-[4e]	Isolated	227.14	Ion Trap	LC-ESI	Bronze	0.80586	79.706	0.03054	0.03054	GNPS-LIBRARY	197.28	N/A	6	1	227.12
M-H	M-H		Flumetol	Isolated	337.25	QToF	LC-ESI	Bronze	0.93867	6.9259	0.001621	0.001621	GNPS-LIBRARY	304.11	N/A	6	1	197.28
M-H	M-H		Flumetol	Isolated	337.25	QToF	LC-ESI	Bronze	0.93867	6.9259	0.001621	0.001621	GNPS-LIBRARY	304.11	N/A	6	1	197.28
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene	Isolated	0	QToF	ESI	Bronze	0.74848	629.51	1.9038	1.9038	GNPS-SILCOELEM-TDA-PART1	304.11	N/A	6	1	302.207
M-H	M-H		Massbank:BM07739-2,5-Dimethylfluorene															

“Let's think the unthinkable, let's do the undoable. Let us prepare to grapple with the ineffable itself, and see if we may not eff it after all.”

— Douglas Adams, Dirk Gently's Holistic Detective Agency