

Characterisation of inducible antibiotic production by streptomycetes isolated from hyper-arid environments

Tiago Santos

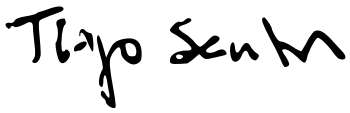
Thesis presented in fulfilment of the requirement for the degree of Doctor of Philosophy (PhD)

2020

Declaration

This thesis is the result of the author's original research. It has been composed by the author and has not been previously submitted for examination which has led to the award of a degree.

The copyright of this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.50. Due acknowledgement must always be made of the use of any material contained in, or derived from, this thesis.

Signed:  Date: 12th May 2020

Acknowledgments

Three years went by very quickly in what I think was a remarkable journey full of ups and downs but also breakthroughs. I am very grateful to my supervisors, Paul and RuAn, for accepting me as a student in their lab and for their continuous encouragement throughout the years. I hope that with this PhD thesis I achieved what you had in mind for this project.

I would like to thank everyone in Paul's and RuAn's lab who helped me immensely, especially Jana, Gaetan, Gillian, Lis, Thaer, David, Jan, Adriana, Bela, Victor and Mohamed. Also, a big thank you to two great people, Talal and Bird, that I look forward to meeting again soon.

I could not forget all people in the 6th floor, past and present, that I had the luck to share the lab with, Natalie, Muattaz, Walid, John, Ainsley, Becca, Liam, Laia, Ally, Gordon, Jonny, Dalia, Alemão and Adam. Also, Fraser and Paul Murray for their help with HSI. My thanks also to all the people that I have met in conferences, workshops and summer school as well as outside university for all good moments they shared with me.

Lastly, I am especially grateful for your friendship Ana, Gaetan and the 'Spanish' duo Alejandro and Parra with whom I created great memories during these years.

Por último, agradecer aos meus pais por me aturarem e ajudarem tanto a chegar até aqui. Agradeço também ao meu irmão Miguel pelo apoio ao longo destes anos.

Para os meus pais João e Armanda, e irmão Miguel.

Table of Contents

ABSTRACT 1

LIST OF FIGURES 2

LIST OF TABLES 6

LIST OF ABBREVIATIONS AND SYMBOLS 8

CHAPTER 1. INTRODUCTION 11

1.1 Antimicrobial resistance as a global threat 11

1.1.1 Impact of AMR in Human health and wealth 11

1.1.2 Origin and antimicrobial resistance mechanisms 14

1.2 *Actinobacteria*: natural source of important metabolites 17

1.3 The unique case of *Streptomyces* 20

1.3.1 *Streptomyces* life cycle 21

1.3.2 Morphological regulation and specialised metabolism 23

1.4 Natural products and where to find them 25

1.5 Re-purposed and novel natural products 27

1.5.1 Advent of next-generation sequencing 31

1.5.2 Genomics as a tool for wider analysis 35

1.6 Metabolomics and the search for new molecules 37

1.7 Scope of the project 38

CHAPTER 2. MATERIALS AND METHODS 39

2.1 General microbiology 39

2.1.1 Microbial strains, plasmids and oligonucleotides 39

2.1.2 Isolation, maintenance and cultivation conditions 43

2.2 Characterisation of the isolates 47

2.2.1 Cultural and phenotypic characteristics 47

2.2.2 Hyperspectral Imaging analysis 48

2.2.3 Phylogenetic characteristics using 16S rRNA 48

2.2.4 Pulse Field Gel Electrophoresis 49

2.3 Genome extraction and sequencing 51

2.4 Whole genome analysis 53

2.4.1 Multilocus sequence analysis (MLSA), Automated Multi-Locus Species Tree and average nucleotide identity 53
2.4.2 Pan-genome analysis and Gene Ontology classification 53
2.4.3 Prophage prediction and genomic islands identification 54
2.4.4 Complete genome sequence for PH28 and PH5 and comparative analysis 55
2.4.5 Whole genome sequence improvement and annotation 56
2.4.6 Secondary metabolites prediction 56
2.5 Analysis of secondary metabolites production 57
2.5.1 Bioactivity assays against clinical pathogens 57
2.5.2 Screening for interactions between isolates using plug assay 57
2.5.3 Potential induction of the isolates by osmotic stress 58
2.6 Small and Large-scale Metabolomics analysis of the crude extracts for the isolates 59
2.6.1 Metabolomics analysis and bioactivity assays of crude extracts from all isolates 59
2.6.2 Mass spectrometry and data analysis 60
2.6.3 Scale-up fermentation of isolate PH28 61
2.6.3.1 Medium pressure liquid chromatography (MPLC) and thin layer chromatography (TLC) 62

CHAPTER 3. PHYLOGENY, PHYSIOLOGICAL AND VISUAL CHARACTERISATION OF *STREPTOMYCES* ISOLATES FROM THE ATACAMA DESERT 65

3.1 Introduction 65
3.2 Isolate identification by 16S rRNA gene sequencing 69
3.3 Phenotypic differences among the isolates from different locations 72
3.4 Hyperspectral Imaging as a promising tool to resolve differences among related isolates 76
3.5 Physiological assays show hidden diversity between isolates 81
3.6 Inhibition and/or neutral interactions emerged between isolates from both locations 84
3.7 Osmotic stress induces bioactive metabolites against indicator strains 86
3.8 Discussion 91

CHAPTER 4. WHOLE-GENOME ANALYSIS OF THE ATACAMA DESERT ISOLATES FROM SOIL 7 AND SOIL 8 95

4.1 Introduction 95

4.2 Whole-genome analysis and mapping uncovers remarkable inter- and intra-specific diversity 98
4.3 Pangenome analysis reveals intra-specific gene diversity and large accessory genome in isolates from both soils 108
4.4 Prophage diversity as a potentiator of niche driven diversity 115
4.5 Assembly and complete genome comparison of two isolates confirms divergence between locations 118
4.5.1 Assembly and *in silico* analysis of representative isolates PH28 and PH5 genomes to understand potential soil specific characteristics 118
4.5.2 Distinct genome structure and synteny between strains from the Atacama Desert and other *Streptomyces* suggests shared traits in this environment 126
4.5.3 Genomic islands predicted mostly in higher frequency at the end of PH28 and PH5 genomes 137
4.6 Common specialised metabolites found among the isolates but also location-specific BGCs were identified 141
4.7 Discussion 148

CHAPTER 5. BIOACTIVE COMPOUNDS AND METABOLOMICS 157

5.1 Introduction 157
5.2 Production of active metabolites against one or more indicator strains observed in solid media 162
5.3 Small scale fermentation and mass spectrometry analysis of the isolates suggests soil-dependent activity profile 164
5.3.1 Disk diffusion assays from crude extracts display activity against Gram-positive bacteria 164
5.3.2 Metabolomics analysis of the isolates indicates potential for specialised metabolites production 171
5.3.3 Genomics and phylogenetic analysis of biosynthetic gene clusters shows cinerubin as a soil-specific BGC 185
5.4 Identification of the active compound using large scale fermentation of PH28 and mass spectrometry analysis 191
5.5 Discussion 207

CHAPTER 6. SUMMARY AND CONCLUSION 210

CHAPTER 7. REFERENCES 215

Abstract

Antimicrobial resistance is a major threat to the future of human society with 10 million deaths projected in 2050. However, the search for new antimicrobials faces a setback in the rediscovery of the already known antibiotics. One strategy to mitigate this problem is the study of isolates from extreme environments, e.g. the Atacama Desert in Chile, adapted to those conditions and potentially encoding new specialised metabolites. Another method is to employ a multi-approach strategy for the discovery and elucidation of new metabolites combining genomics and metabolomics. This research project employed bioactivity assays, whole-genome analysis and metabolomics to determine the antibiotic potential from bioassays, identification of potential biosynthetic gene clusters and the elucidation of the compounds(s) responsible for the activity. Fourteen strains, closely related to *Streptomyces phaeoluteigriseus* by 16S rRNA, were isolated from two locations 30 km apart with similar altitude from the Atacama Desert. Further to taxonomic similarity the isolates showed also observable similarities, e.g. soluble pigment, despite inhabiting different locations. However, differential growth inhibition of several indicator strains, e.g. *Staphylococcus aureus* and *Enterococcus faecalis* was observed between isolates of both locations. Furthermore, phenotypic analysis and hyper-spectral imaging revealed hidden diversity of the isolates. Whole-genome analysis contradicted the phylogenetic relationship observed using 16S rRNA. Unique inter and intraspecific variations between isolates emerged with two distinct clades forming, revealing the possibility of soil 7 isolates to be a new species. Additionally, highly diverse plasmid profiles were observed instead of phages suggesting that the observed unique variations could result from horizontal gene transfer. Antibacterial activity was observed in solid and liquid cultures, especially in isolates from soil 7. Mass spectrometry analysis helped identify cinerubin B and unknown congeners of actinomycin (MW 1307.552). The latter were produced in large quantity by PH28 and further scale-up fermentation allowed the purification of an anti-*E. faecalis* metabolite. Actinomycin Y_1 was its main constituent however further purification is required due to its low purity. In this study, the potential of the Atacama Desert to promote genetic and metabolomic diversification of closely evolutionary strains by niche specific pressures leading to production of known and, most importantly, bioactive metabolite production was demonstrated.

List of Figures

Figure 1.1 Estimation overview of the antimicrobial resistance by 2050. 12
Figure 1.2 Overview of antibiotic targets in the cell. 16
Figure 1.3 Overview of the *Streptomyces* life cycle (left) and fluorescence time-lapse of *S. venezuelae* life cycle using a microfluidic device (right). 23
Figure 1.4 Synthetic generation of derivates using oxytetracycline as the source for the core structure. 28
Figure 1.5 Overview of the Next generation Illumina (A) and PacBio sequencer (B) methodologies. 33
Figure 1.6 Overview of a pipeline combining metabolomics, genome mining and bioactivity for dereplication and natural product discovery. 36
Figure 3.1 Sampling locations of soil 7 (red circle) and soil 8 (dark blue circle) in the Atacama Desert (Chile) compared to those of previous novel *Streptomyces* species. 68
Figure 3.2 Maximum Likelihood tree for 16S rRNA sequences from Chilean isolates. 71
Figure 3.3 Atacama Desert isolates from soil 7 and soil 8, and the closest neighbour *Streptomyces phaeoluteigriseus* grown in ISP3 and showing the aerial spore mass (top) and reverse colour (bottom). 73
Figure 3.4 Dendrogram showing similarities between *Streptomyces* from both soils using reverse colour RGB values in ISP3 based in *Antony-Babu, Stach & Goodfellow (2010)*. 75
Figure 3.5 Example of two spatial dimension and one spectral dimension projection of two isolates grown in ISP2 using hyperspectral imaging. 77
Figure 3.6 Dendrogram showing similarities between *Streptomyces* from both soils using HSI instrument to capture the reverse colour of each isolate in ISP3. 79
Figure 3.7 Dendrogram showing similarities between *Streptomyces* from both soils using the HSI instrument to capture the aerial spore mass colour of each isolate in ISP3. 80
Figure 3.8 Overview of inhibition between isolates from the two locations. 85
Figure 3.9 Induction obtained against *B. subtilis* with isolates from soil 7. 88
Figure 3.10 Induction obtained against *S. aureus* with isolates from soil 7. 89
Figure 3.11 Developmental changes observed at different concentrations of potassium chloride in *S. phaeoluteigriseus* DSM 41896. 90
Figure 4.1 Maximum Likelihood tree constructed using concatenated sequences of four housekeeping genes, *atpD*, *gyrB*, *recA* and *trpB* from each isolate and *S. phaeoluteigriseus* DSM 41896. **102**
Figure 4.2 MLST tree constructed using autoMLST (default parameters) with concatenated sequences of 84 housekeeping genes against *Streptomyces* reference genomes. 103
Figure 4.3 Heatmap of ANI percentage identity calculation using PYANI between isolates from both soils. 104

Figure 4.4 Pulse-Field Gel Electrophoresis of total DNA isolated from *S. phaeoluteigriseus* and the Atacama isolates from soil 7 showing extrachromosomal replicons. 106
Figure 4.5 Pulse-Field Gel Electrophoresis of total DNA isolated from *S. phaeoluteigriseus* and the Atacama isolates from soil 8 showing extrachromosomal replicons. 107
Figure 4.6 Pangenome analysis summary statistics obtained using PIRATE for all of the isolates (>95% isolates). 109
Figure 4.7 Overview of the 14 Atacama Desert isolates pangenome using the complete genomes inferred by PIRATE. 110
Figure 4.8 Pangenome visualization of the gene presence (low (red) to high sequence identity (blue) to a reference sequence) and absence (white lines) are represented by lines using PIRATE revealing the core and accessory genome in both species. 112
Figure 4.9 Distribution of orthologous genes on COG categories identified in the core genome using GO FEAT. 114
Figure 4.10 Visualisation of the complete genomes of PH28 and the putative plasmid pPH28.1 using CGView. 122
Figure 4.11 Visualisation of the complete genomes of PH5 and the two putative plasmids pPH5.1 and pPH5.2 using CGView. 123
Figure 4.12 Overlapping of orthologous gene clusters for the chromosome and plasmids of PH28 and PH5. 125
Figure 4.13 Location of some genes in several *Streptomyces* complete genomes oriented and aligned to *S. coelicolor dnaA*, normally adjacent to the origin of replication (*oriC*). 128
Figure 4.14 Comparison of the archetypal telomeres of *S. coelicolor* and the isolate PH5. 130
Figure 4.15 Dot plot genome alignment of the the soil representative genome of PH28 (soil 7) and type strains using D-Genies. 131
Figure 4.16 Dot plot genome alignment of the the soil representative genome of PH5 (soil 8) and type strains using D-Genies. 132
Figure 4.17 Regions containing mismatches between the alignment of PH28 and PH5 genomes. 133
Figure 4.18 Circular visualisation GI predicted on the genome of PH28 (left) from soil 7 and *S. leeuwenhoekii* (right) and linear of the respective genomes (bottom) in Islandviewer 4. 139
Figure 4.19 Circular visualisation of GI predicted on the genome of PH5 (left) from soil 7 and *S. leeuwenhoekii* (right) and linear of the respective genomes (bottom) in Islandviewer 4. 140
Figure 4.20 Graphic representation and distribution of the number of biosynthetic gene clusters per genome identified using antiSMASH 5.0. 144
Figure 4.21 Clustering in different types of biosynthetic gene clusters, PKSother, RiPPS, PKSI, Terpenes and PKS-NRP hybrids shared between isolates within the same or different location obtained using BiG-SCAPE. 145

Figure 4.22 Clustering in different types of biosynthetic gene clusters, Others and NRPS, shared between isolates within the same or different location obtained using BiG-SCAPE. 146
Figure 4.23 Differences in the biosynthetic gene cluster for SapB of PH12 and comparison with two *Streptomyces* species isolated from soil 8. 147
Figure 4.24 Differences in the orthologous gene clusters and proteins between isolates from both soils. 153
Figure 5.1 Chemical structures of actinomycins. 159
Figure 5.2 Chemical structures of four anthracyclines. 160
Figure 5.3 Different phases of analysis performed to understand the production of metabolites by the isolates. 167
Figure 5.4 Principal component analysis (PCA) plot for all samples from soil 7 and soil 8 isolates. 174
Figure 5.5 OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *E. faecalis* and *S. aureus*. 175
Figure 5.6 Log transformed OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *E. faecalis* and *S. aureus* excluding outlying inactive extract ISP2_PH6. 176
Figure 5.7 OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *B. subtilis*. 177
Figure 5.8 OPLS-DA SIMCA analysis plots overview for all samples from soil 7 and soil 8 isolates against *B. subtilis* excluding outlying inactive extracts. 178
Figure 5.9 Heat map analysis showing the diversity of the metabolomes. From top to bottom, lines correspond to low to high *m/z* ratios. 179
Figure 5.10 Maximum Likelihood showing 16S rRNA gene sequences relationships between the isolates from both soils (soil 7: black; soil 8: blue) and strains with hits obtained by dereplication and SIMCA analysis. 186
Figure 5.11 PKSother network visualisation of the relationship and similarity of predicted cinerubin B biosynthetic gene clusters from soil 7 isolates. 188
Figure 5.12 Clustering of the genes composing the family FAM_00088 shared by soil 7 isolates with activity against *Enterococcus faecalis* and *Staphylococcus aureus*. 189
Figure 5.13 Base peak plot chromatogram of species from soil 7, solvent and media blanks in the positive ionization mode showing a peak *m/z* 826.328 and corresponding to cinerubin B based on dereplication. 190
Figure 5.14 Summary TLC plate of the 27 fractions obtained after fractionation of PH28 with mobile phase DCM:MeOH (9:1, v/v). 194
Figure 5.15 Bioactivity assay showing the inhibition zones against *E. faecalis* by the fractions extracted from the fermentation of PH28. 195

Figure 5.16 Principal component analysis (PCA) plot for the fractions obtained from PH28 fermentation. 196
Figure 5.17 OPLS-DA SIMCA plots of the bioactivity of PH28 fractions against *E. faecalis*. 197
Figure 5.18 Extracted ion chromatograms and high-resolution mass spectra for *m/z* 1303.54 [M-H] and 1305.56 [M+H] dereplicated as actinomycin Y_1 in fractions 2, 3, and 21. 199
Figure 5.19 Extracted ion chromatograms and high-resolution mass spectra for unknown *m/z* 1306.54 [M-H] and 1308.56 [M+H] in fractions 2, 3, and 21. 200
Figure 5.20 Alignment of the biosynthetic gene clusters of actinomycin D (A) and actinomycin G from *S. iakyrus* (B), *Streptomyces* spp. PH28 (C) and *S. flavovariabilis* (D). 202
Figure 5.21 Alignment of the biosynthetic gene clusters of actinomycin G from *S. iakyrus* (A), *Streptomyces* spp. PH28 (B) and *S. flavovariabilis* (C). 203
Figure 5.22 ^{1}H NMR spectra for selected fractions of PH28 obtained from the scale-up fermentation in DMSO-d_6. 204
Figure 5.23 ^{1}H NMR spectra for selected fraction F3 from PH28 scale-up fermentation in DMSO-d_6. 205
Figure 5.24 Two-dimensional COSY (A) and TOCSY (B) of selected fraction F3 from PH28 scale-up fermentation for illustrating the occurrence of individual amino acids for structure elucidation. 206
Figure 6.1 Overview of the phylogeny of the isolates and the inhibition observed between the two locations. 214

List of Tables

Table 1.1 List of some bioactive compounds in *Streptomyces* species. 19
Table 1.2 Natural products discovered from deep sea and hyper-arid desert *Actinobacteria*. 30
Table 1.3 Comparison of high-throughput sequencing platforms. 34
Table 2.1 Bacterial strains used in this project. 40
Table 2.2 Model strains for bioactivity assays. 41
Table 2.3 Primers used in this project for 16S rRNA amplification and sequencing. 42
Table 2.4 Culture media for growth and maintenance of strains. 44
Table 2.5 Gradient elution of crude extract obtained using Biotage MPLC IsoleraTM One 2.0.4 Spektra flash purification. Dichloromehtane was used as solvent A. 64
Table 3.1 16S rRNA gene sequencing results including the similarity (%) obtained in EZBioCloud database. 70
Table 3.2 Reverse colour of each isolate and type strain grown in different ISP media. 74
Table 3.3 Phenotypic characteristics of isolates from soil 7 and its most closely related species, *Streptomyces phaeoluteigriseus* DSM41896. 82
Table 3.4 Phenotypic characteristics of isolates from soil 8 and its most closely related species, *Streptomyces phaeoluteigriseus* DSM41896. 83
Table 4.1 Sequencing statistics obtained after SPAdes assembly and RAST annotation. 99
Table 4.2 Average of the sequencing statistics obtained after SPAdes assembly and RAST annotation for both soils. 100
Table 4.3 Prophage regions identified in the genomes of the isolates from soil 7 using PHASTER. 116
Table 4.4 Prophage regions identified in the genomes of the isolates from soil 8 using PHASTER. 117
Table 4.5 Complete genome statistics of PH28 and PH5 after complete assembly. 120
Table 4.6 DNA-DNA hybridisation performed using GGDC (DSMZ) of selected isolates PH28 and PH5 with the most closely related strain *S. phaeoluteigriseus* and *S. leeuwenhoekii* isolated from the Atacama Desert. 121
Table 4.7 Best-hit obtained by blast search of genes *tap* and *tpg* identified in the chromosome or plasmid of isolates PH28 and PH5. 129
Table 4.8 Genes predicted to be encoded in the gap localised in the central core region of PH28 genome and absent at the same location in isolate PH5. 134
Table 4.9 Genes predicted to be encoded in the gap localised in the central core region of PH5 genome and absent at the same location in isolate PH28. 136
Table 4.10 Sequencing statistics obtained after scaffolding with PH28 and PH5 as reference genomes for soil 7 and soil 8, respectively, using MeDuSa. 143

Table 4.11 Previous pangenome analysis of *Streptomyces* strains and the total percentage of core genes present in the pangenome. 151
Table 4.12 Pangenome analysis of *Streptomyces* strains in each soil and the percentage of shared genes considered as core genome using PIRATE. 152
Table 5.1 Bioactivity results using six ISP media (ISP2-ISP7) and the Atacama isolates (only shown indicator strains inhibited by the isolates, 5 out of 7). 163
Table 5.2 Liquid cultures obtained after 7 days of growth in production and non-production medium of soil 7 isolates. 166
Table 5.3 Liquid cultures obtained after 7 days of growth in production and non-production medium of soil 8 isolates. 166
Table 5.4 Bioactivity by disk diffusion assay against the indicator strains using 100 μg of each crude extract (only shown inhibited strains) and controls used (*). 168
Table 5.5 Variation between bioactivity by plug assay from solid media and disk diffusion assay from liquid broth against the indicator strains. 170
Table 5.6 Dereplication table containing the outlaying metabolites identified in the isolates crude extract that showed activity against *E. faecalis*, *S. aureus* and *B. subtilis*. 182
Table 5.7 List of biosynthetic gene clusters from MIBiG repository and identified in MS dereplication for BGC similarity search against soil 7 isolates BGC's using BiG-SCAPE. 187
Table 5.8 Fractionation of PH28 crude extract from 10 L scale-up yielding 29 fractions. 193
Table 5.9 Outlying metabolites detected by OPLS-DA in PH28 fractions that showed bioactivity against *E. faecalis*. 198

List of Abbreviations and Symbols

%	Percent
ºC	Degrees Celsius
ATCC	American Type Culture Collection
BAM	Compressed Binary version of Sequence Alignment/Map file
bp	Base pair
CCS	Circular Consensus Sequencing
$CDCl_3$	Deuterated chloroform
COG	Clusters of Orthologous Groups
COSY	Correlation Spectroscopy
D-Genies	Dot plot large GENomes in an Interactive, Efficient and Simple Way
D-Val	D-Valine
Da	Dalton
DCM	Dichloromethane
DDH	DNA-DNA Hybridization
ddNTPs	Dideoxynucleotides triphosphates
dH_2O	Distilled Water
DMSO-*d6*	Deuterated Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleoside triphosphates
dsDNA	Double Stranded DNA
DSM	German Collection of Microorganisms and Cell Cultures
EDTA	Ethylenediaminetetraacetic acid
g	Gram
GC	Guanine-Cytosine
gDNA	Genomic DNA
GGDC	Genome-to-Genome Distance Calculator
GO FEAT	Gene Ontology Functional Enrichment Annotation Tool
GTR	Generalised Time Reversible
h	Hour
H_2SO_4	Sulfuric acid
HGAP	Hierarchical Genome Assembly Process
HPLC	High Performance Liquid Chromatography
HSI	Hyperspectral Imaging
iPrOH	Isopropyl alcohol
ISP	International *Streptomyces* Project
Kbp	Kilobases pair

km	Kilometres
L	Litre
L-Pro	L-Proline
L-Thr	L-Threonine
LC-MS	Liquid Chromatography–Mass Spectrometry
M	Molar
mA	Milliampere
Mbp	Megabases pair
MeDuSa	Multi-Draft based Scaffolder
MeOH	Methanol
MeVal	N-Methylvaline
mg	Milligram
MIBiG	Minimum Information about a Biosynthetic Gene cluster
min	Minutes
mL	Millilitre
mM	Millimolar
mm/a	Millimetres per annum
Mol%	Mole percent
MS/MS	Tandem Mass Spectrometry
MW	Molecular Weight
N-Me	N-Methyl
NaCl	Sodium chloride
NaOH	Sodium hydroxide
NASA	National Aeronautics and Space Administration
ng	Nanogram
nm	Nanometre
NMeGly	N-Methylglycine
NRPS	Non-Ribosomal Peptide Synthase
NRRL	Agricultural Research Service Culture Collection
OD	Optical Density
PCR	Polymerase Chain Reaction
pH	Potential of hydrogen
PIRATE	Pangenome Iterative Refinement and Threshold Evaluation
PKS	Polyketide Synthase
PMSF	Phenylmethylsulphonyl fluoride
ppGpp	Guanosine pentaphosphate
ppm	Parts-Per-Million
PRISM	PRediction Informatics for Secondary Metabolomes

qPCR	Quantitative PCR
RAST	Rapid Annotation using Subsystem Technology
RAxML	Randomized Axelerated Maximum Likelihood
RGB	Red Green Blue
rpm	Revolutions Per Minute
rRNA	Ribosomal Ribonucleic acid
RT	Retention Time
SDS	Sodium Dodecyl Sulphate
sec	Seconds
SMRT	Single Molecule Real-Time
SPAdes	St. Petersburg genome assembler
SPRI	Solid Phase Reversible Immobilisation
sRGB	Standard Red Green Blue
ssDNA	Single stranded DNA
T1PKS	Type I Polyketide Synthase
T2PKS	Type II Polyketide Synthase
T3PKS	Type III Polyketide Synthase
T92	Tamura 3-parameter
TIC	Total Ion Chromatogram
TOCSY	Total Correlation Spectroscopy
UV/Vis	Ultraviolet–Visible spectroscopy
v/v	Volume by Volume
w/v	Weight by Volume
x*g*	Gravitational force
µg	Micrograms
µL	Microliters
µM	Micromolar

Chapter 1. Introduction

1.1 Antimicrobial resistance as a global threat

Antimicrobial resistance (AMR) is a natural occurrence regulated by Darwinian rules of selection. Changes in the environment select individuals with mutations or traits that confer a positive advantage for survival, increasing the fitness of later generations, e.g. selection of antimicrobial resistant individuals in a population. Its occurrence has been discovered in locations far from active human activity sources, such as permafrost and caves (Mindlin & Petrova, 2017; Bhullar *et al.*, 2012; Perry, Waglechner & Wright, 2016; Lupo, Coyne & Berendonk, 2012); this suggests an ancient origin for production and resistance to antibiotics (>40 million years ago) (Wilson, 2014).

1.1.1 Impact of AMR in Human health and wealth

The widespread increase of antimicrobial resistance is one of the biggest threats to public health and the lifestyle of modern society. According to The World Health Organization (WHO), the dissemination of resistant species has encompassed all continents (World Health Organization, 2015) and current evaluations reveal that the number of human deaths will rise from 700,000 to 10 million annually by 2050 due to antimicrobial resistant microorganisms (**Figure 1.1**) (Jasovský *et al.*, 2016). In 2013 around 25,000 lives were lost in Europe costing €1.5 billion annually (Paphitou, 2013; Blair *et al.*, 2015). Notwithstanding this, the fact that antimicrobial resistance is a natural process, the irresponsible use of antibiotics and the inefficient prevention and control of infection outbreaks, have boosted this effect (World Health Organization, 2015; Jasovský *et al.*, 2016).

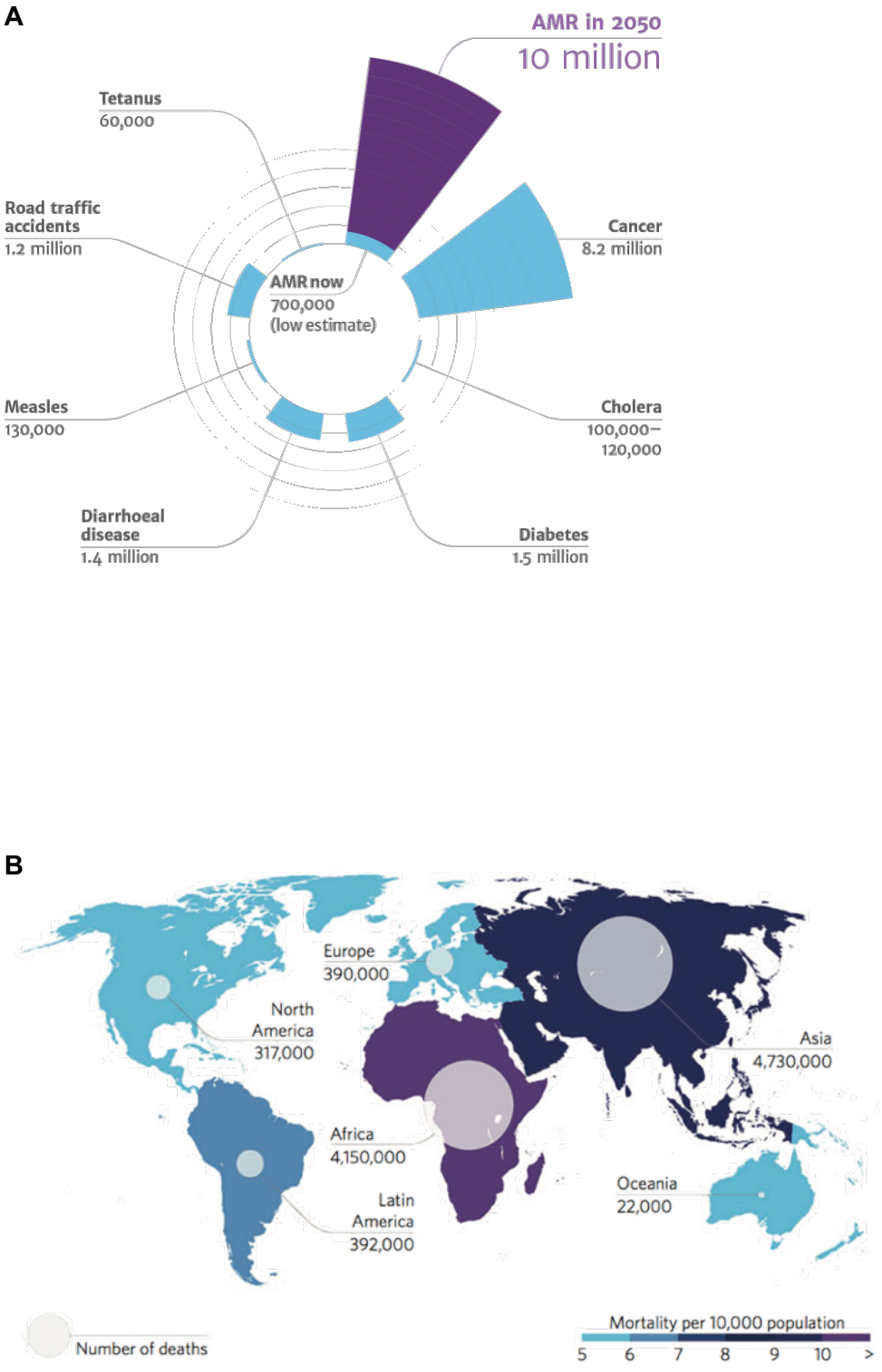


Figure 1.1 Estimation overview of the antimicrobial resistance by 2050.

A. Main causes of human deaths (O'Neill, 2016) and the **B.** number of deaths by continent each year (Sugden, Kelly & Davies, 2016).

In the health sector, antibiotics are used to prevent infections in surgical interventions, transplants and in cancer treatments. The decrease in antibiotic effectiveness and therefore in treatments caused by novel resistance will eventually result in a high number of human deaths (Smith & Coast, 2013; Jasovský *et al.*, 2016). The agricultural sector is one of the major sources promoting resistance in soil by the overuse of antibiotics as prophylactics for the purpose of increasing production due to high demand for food (Jasovský *et al.*, 2016; Laxminarayan *et al.*, 2016). Graham *et al*., studied the implication of manure use and presence of resistance determinants in soils of Denmark, revealing a cause-effect between the two conditions. Furthermore, when the use of antibiotics for non-therapeutic treatments was banned, a decrease in resistance genes for β-lactams were identified. In this way, regulation and control of antibiotic usage is possible and can mitigate the fast growing rate of resistance (Graham *et al.*, 2016).

The 2014 report from the European Antimicrobial Resistance Surveillance Network (EARS-Net) confirms the overall increasing trend of antimicrobial resistance. Moreover, the combined resistance to three or more antimicrobial groups increased in Gram-negative bacteria, namely in *Escherichia coli*, *Klebsiella pneumoniae* and *Pseudomonas aeruginosa* (ECDC: European Centre for Disease Prevention and Control, 2015). For the Gram-positive Methicillin-resistant *Staphylococcus aureus* (MRSA) the level of resistance decreased from 18.6% to 17.4% from 2011 to 2014 (ECDC: European Centre for Disease Prevention and Control, 2015).

Antibiotic resistance also disrupts the economic sector. Health economists consider that the costs previously estimated at approximately US$55 billion per year associated with antibiotic resistance are miscalculated (Smith & Coast, 2013). The 'Review on Antimicrobial Resistance' estimates the cost at US$100 trillion by 2050. However, the value could also be underestimated since it only takes into account the direct use of antibiotics to treat infections (Piddock, 2016). Recently, some initiatives have been taken by the WHO, the European Commission (ND4BB program) and the Infectious Diseases Society of America (10 x '20 Initiative) to tackle this problem. The main goal is to research and discover new antibiotics and encourage this research in the private sector, where investment is declining (Paphitou, 2013; O'Neill, 2016). Between 2003 and 2013 only US$1.8 billion (about 5% of total venture capital) was invested into pharmaceutical Research and Development (R&D) related to

antimicrobials research (O'Neill, 2016).

Antibiotic resistance poses an enormous challenge to overcome. However, with new promises of investment in discovery of novel and effective antibiotics combined with adequate regulation of their use, the return to a "pre-antimicrobial era" is avoidable.

1.1.2 Origin and antimicrobial resistance mechanisms

Naturally, antibiotic-producing bacteria have resistance to antibiotics they produce for self-protection and are considered to be the initial source of resistance traits (Wilson, 2014). Bacteria can have intrinsic resistance to different antibiotics due to structural characteristics, e.g. daptomycin is active against Gram-positive but cannot effectively pass the cytoplasmic membrane of Gram-negative bacteria due to a lower abundance of anionic phospholipids (Blair *et al.*, 2015). Resistance can be developed by mutations or over expression of genes present in the chromosomes or plasmids, for example resistant versions of *gyrA* and *gyrB* can confer resistance to fluoroquinolones and mutations in the genes coding topo-isomerase IV (*par*C and *par*E) confer resistance to linezolid in *Staphylococcus aureus* (Blair *et al.*, 2015; Lupo, Coyne & Berendonk, 2012).

Bacteria can also modify or degrade the antibiotics through different enzymatic reactions, e.g. hydrolysis, group transfer and oxidation (**Figure 1.2**). Macrolides, e.g. erythromycin, interfere with protein synthesis by bonding to the 50S ribosome subunit and blocking the release of the peptide. However, production of hydrolytic enzymes, e.g. esterases, enables the cleavage of the macrocycle ester and inactivation of these antibiotics. Similarly, the β-lactam antibiotics, e.g. penicillins and cephalosporins, can be inactivated by cleavage of their ring by β-lactamases. Furthermore, the action of aminoglycosides that bind to the A-site on the 30S subunit can be blocked by action of acetyltransferases that disrupt important hydroxyl and amine groups or phosphorylation by phosphotransferases producing resistance (Wright, 2005; Lin *et al.*, 2015). Another mechanism of resistance is multidrug efflux systems that are normally encoded in the chromosome and are responsible for general resistance mechanism to heavy metals and quorum sensing signals that are pumped out of the cell. However, systems involved in drug resistance are also located in mobile genetic

elements (e.g. for tetracycline resistance) and can be transferred by horizontal gene transfer. Furthermore, efflux systems are also important in antibiotic producing bacteria, constituting a self-resistance mechanism, e.g. *Streptomyces rimosus*, the producer of oxytetracycline (Blair *et al.*, 2015).

Resistance can also be acquired by bacteria through phages, transposons, plasmids or incorporation of exogenous DNA. D'Costa *et al*. found genes in permafrost (30,000 years old) similar to those found in contemporary pathogenic bacteria, e.g. *vanX*-like and *tetM* coding for vancomycin and tetracycline resistance (D'Costa *et al.*, 2011). This show emergence of antibiotic resistance as a natural phenomenon, predating widespread usage of antibiotics in modern society (D'Costa *et al.*, 2011). Additionally, *floR* and *strB-strA,* found in a plasmid from *Aeromonas bestiarum*, were identified with high sequence similarity in another species, *Vibrio cholerae*, and demonstrates the exchange of antibiotic resistance genes via conjugative plasmids or other mobile genetic mobile elements (Lupo, Coyne & Berendonk, 2012).

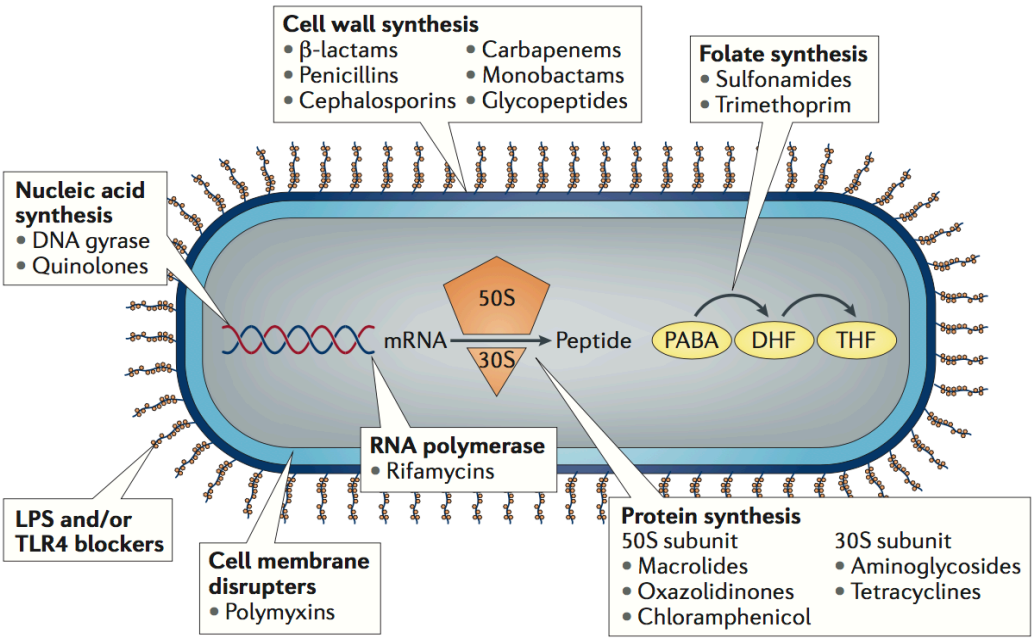


Figure 1.2 Overview of antibiotic targets in the cell.

DHF, dihydrofolic acid; LPS, lipopolysaccharide; PABA, para-aminobenzoic acid; THF, tetrahydrofolic acid; TLR4, Toll-like receptor 4. *From Brown (2015).*

1.2 *Actinobacteria*: natural source of important metabolites

According to Bergey's Manual of Systematic Bacteriology the *Actinobacteria* phylum comprises six classes; the *Actinobacteria*, *Nitriliruptoria*, *Acidimicrobiia*, *Coriobacteriia*, *Thermoleophilia* and *Rubrobacteria*. These are Gram-stain-positive or Gram-stain variable bacteria with high DNA mol% G+C content (50-70 mol%), generally with large genomes (average of 5 Mbp) and phenotypically very different (Lewin *et al.*, 2016; Gao & Gupta, 2012; Goodfellow *et al.*, 2012). *Actinobacteria* have a wide distribution in distinct ecosystems, from aquatic to air, being very common in soil, forming the majority of the population (Goodfellow *et al.*, 2012; Bhat *et al.*, 2013). The combination of large populations and the ability to produce enzymes that degrade biomass demonstrates their major role in the cycling of organic matter (Lewin *et al.*, 2016). Generally, optimal growth of *Actinobacteria* is achieved at a neutral pH, between 6-9, and a temperature of 25-30ºC. However, some are thermophilic with an optimal growth around 55-60ºC and present in extreme environments, such as hyper-arid deserts or deep seas (Bhat *et al.*, 2013; Goodfellow *et al.*, 2012). Many *Actinobacteria* reproduce by sporulation, after the formation of a mycelium through vegetative growth. This is achieved by tip extension combined with branching of the hyphae and several factors can influence it including pH, temperature and humidity. Furthermore, vegetative growth is limited in dry soils due to moisture limitation and during nutrient depletion, these bacteria live as semidormant spores (Barka *et al.*, 2016).

The diversity of niche-specific pressures can promote novel chemical features generating diverse specialised metabolites. These, have great relevance in several areas, i.e. human medicine and biotechnology, ranging from antimicrobial and antiphage activity (streptomycin and chromomycin from *Streptomyces* spp.) to anticancer (salinosporamides from *Salinispora tropica*), boosting the interest and research on this phylum (Goodfellow *et al.*, 2012; Bhat *et al.*, 2013; Barka *et al.*, 2016; Kronheim *et al.*, 2018). The main source of the majority of bioactive metabolites already discovered, around 41%, have been isolated from *Actinobacteria* (Guo *et al.*, 2015). The first commercialised antibiotic from this phylum was isolated in 1944 by Selman Waksman and Albert Schatz. Streptomycin from *Streptomyces griseus* was used to treat tuberculosis (Chater, 2006; Kurtböke, 2012). Following this breakthrough, other natural products with antibiotic activity were discovered and

commercialised. Tetracycline, discovered in 1945 from *Streptomyces aureofaciens*, was used to treat diseases such as typhus and infections caused by *Streptococcus pneumonia*. Cephalosporins and chloramphenicol, discovered in the same decade, from *Streptomyces clavuligerus* and *Streptomyces venezuelae* were used to treat bacterial infections such as typhoid. Interestingly, another antibiotic was discovered later on produced by *S. clavuligerus*, clavulanic acid, that function as beta-lactamase inhibitor, that can be used in combination with β-lactam antibiotics circumventing antimicrobial resistance. In the 50s, the trend continued and comprised the most prolific years with several more metabolites discovered. Among these, erythromycin from *Saccharopolyspora erythraea* (1952), vancomycin from *Amycolatopsis orientalis* (1956), novobiocin from *Streptomyces niveus* (1956), rifamycin from *Amycolatopsis mediterranei* (1957) and kanamycin from *Streptomyces kanamyceticus* (1957) are the most well-known compounds. These compounds are used against a wide range of bacterial infections spanning from veterinary to human health (de Lima Procópio *et al.*, 2012; Chaudhary, Dhakal & Sohng, 2013).

The most important producers of natural products with antibiotic properties belong to the genus *Streptomyces*. Interestingly, in 2012 a database was developed, StreptomeDB, to gather all the published data regarding the genus *Streptomyces* and their natural products (**Table 1.1**) (Klementz *et al.*, 2016).

Table 1.1 List of some bioactive compounds in *Streptomyces* species.

Adapted from Chaudhary, Dhakal & Sohng (2013).

Drug	Organism	Application
Aclacinomycin A	*S. galilaeus*	Antitumor
Bleomycin	*S. verticillus*	Antitumor
Ascomycin	*S. hygroscopicus* var. *ascomyceticus*	Antifungal
Pimaricin	*S. natalensis*	Antifungal
Avermectin	*S. avermitilis*	Antiparasitic
Meilingmycin	*S. nanchangensis*	Antiparasitic
Chloramphenicol	*S. venezuelae*	Antibacterial
Clavulanic acid	*S. clavuligerus*	Antibacterial
Daptomycin	*S. roseosporus*	Antibacterial
Kanamycin	*S. kanamyceticus*	Antibacterial
Neomycin	*S. faradae*	Antibacterial
Novobiocin	*S. niveus*	Antibacterial
Streptomycin	*S. griseus*	Antibacterial
Thiostrepton	*S. azureus*	Antibacterial
Tetracycline	*S. aureofacien*	Antibacterial
Chromomycin	*S. griseus*	Antiviral

1.3 The unique case of *Streptomyces*

The Gram-positive genus *Streptomyces* was effectively identified in 1943 by Waksman and colleagues (Goodfellow *et al.*, 2012; Kämpfer *et al.*, 2014). They produce a wide variety of pigments responsible for the colour of the vegetative and aerial mycelia (Goodfellow *et al.*, 2012). The prevalent diamino acid of the cells walls is the LL-diaminopimelic acid. Moreover, mycolic acids are not present (Goodfellow *et al.*, 2012; Kämpfer *et al.*, 2014). Generally, the members of this genus have large genomes, from 6 to 12 Mbp, with a high DNA mol% G+C content (66-73%). They can use different organic compounds as the only source of carbon for energy and growth. The use of semidormant spores increases the endurance of these organisms to rapid changes in the habitat (Goodfellow *et al.*, 2012).

The different specialised metabolites produced by *Streptomyces* species are a continuous source of interest and research. As an overview, the StreptomeDB platform lists 4,040 compounds and 905 different activities observed (Klementz *et al.*, 2016). Among these, several natural products with antiparasitic activity were discovered, such as avermectin (*Streptomyces avermectilis*) and antifungal activity, for example ascomycin (*Streptomyces hygroscopicus* var. *ascomyceticus*). Antitumor compounds currently in use include daptomycin (*Streptomyces roseosporus*), neomycin (*Streptomyces faradae*), thiostrepton (*Streptomyces azureus*), bleomycin (*Streptomyces verticillus*) and aclacinomycin A (*Streptomyces galilaeus*) (**Table 1.1**) (Chaudhary, Dhakal & Sohng, 2013; Klementz *et al.*, 2016). Some anticancer drugs exhibit also antiphage activity, namely daunorubicin and doxorubicin (*Streptomyces peutecius*) (Kronheim *et al.*, 2018). Secondary metabolites can be produced at the end of vegetative growth to protect nutrient resources when stress or changing conditions occur, allowing a shift to aerial growth and beginning of sporulation (Liu *et al.*, 2013; Barka *et al.*, 2016). Despite this, physiological differentiation can still occur in the absence of morphological differentiation, for example during stationary phase of many liquid-grown *streptomycetes*.

1.3.1 *Streptomyces* life cycle

The *Streptomyces* complex life cycle begins under favourable conditions with the germination of the spore and formation of the germ tubes (**Figure 1.3, panel A-B**). These grow by tip extension (**Figure 1.3, panel B**) and branching forming a vast vegetative mycelium (**Figure 1.3C**). The protein DivIVA is part of a polarisome-like complex responsible for the branching and tip growth (**Figure 1.3, panel A-C**) (Barka *et al.*, 2016; Schlimpert, Flärdh & Buttner, 2016).

The formation of the aerial hyphae is thought to be triggered by depletion or decrease of nutrients (**Figure 1.3, panel D-E**) (Bush *et al.*, 2015). Some parts of the vegetative mycelium start a programmed cell death providing a nutrient source to form the aerial hyphae. This is a crucial phase and the bacteria start the production of metabolites to protect the source of nutrients (Kämpfer *et al.*, 2014; Schlimpert, Flärdh & Buttner, 2016). The *bld* (bald) genes are responsible for the erection of the aerial hyphae. Mutants without functional *bld* genes lack the aerial mycelia due to a differentiation block having a "bald/hairless" appearance in some media (Barka *et al.*, 2016; Kämpfer *et al.*, 2014). The next phase is the aerial mycelium growth. In *Streptomyces coelicolor* and *S. griseus* several genes are implicated exclusively in aerial hyphae growth (*ram* genes). Furthermore, aerial growth is facilitated by a sheath, that in the end of the process, will form the spore's coat. It comprises the hydrophobic proteins chaplins (Chp) and rodlins (RldA and RldB). Additionally, the extracellular presence of the surfactant SapB is also required. This protein is encoded by the *ramCSAB* gene cluster under regulation of the *ramR* gene in *S. coelicolor*. In *S. griseus* the gene is regulated by the *bldH* gene and also by BldD and BldA (Barka *et al.*, 2016; Kämpfer *et al.*, 2014).

Sporulation begins with the replication and partitioning of the chromosomes and laying down of septa (**Figure 1.3, panel E**). This synchronous event is organised by the constriction of multiple FtsZ rings, necessary for the formation of walls in vegetative mycelium and septa (**Figure 1.3, panel E-F**, in green colour) (Schlimpert, Flärdh & Buttner, 2016; Kämpfer *et al.*, 2014). In the end, each compartment contains one copy of the chromosome and will eventually form unigenomic spores (**Figure 1.3, panel F**). The maturation of the spores, involving the curling and septation of the aerial hyphae and spore formation, is controlled by the *whi* (white) genes. Mutants

without these genes are unable to stop aerial growth or to form the grey spore pigment and thus unable to complete the life cycle (Kämpfer *et al.*, 2014; Barka *et al.*, 2016).

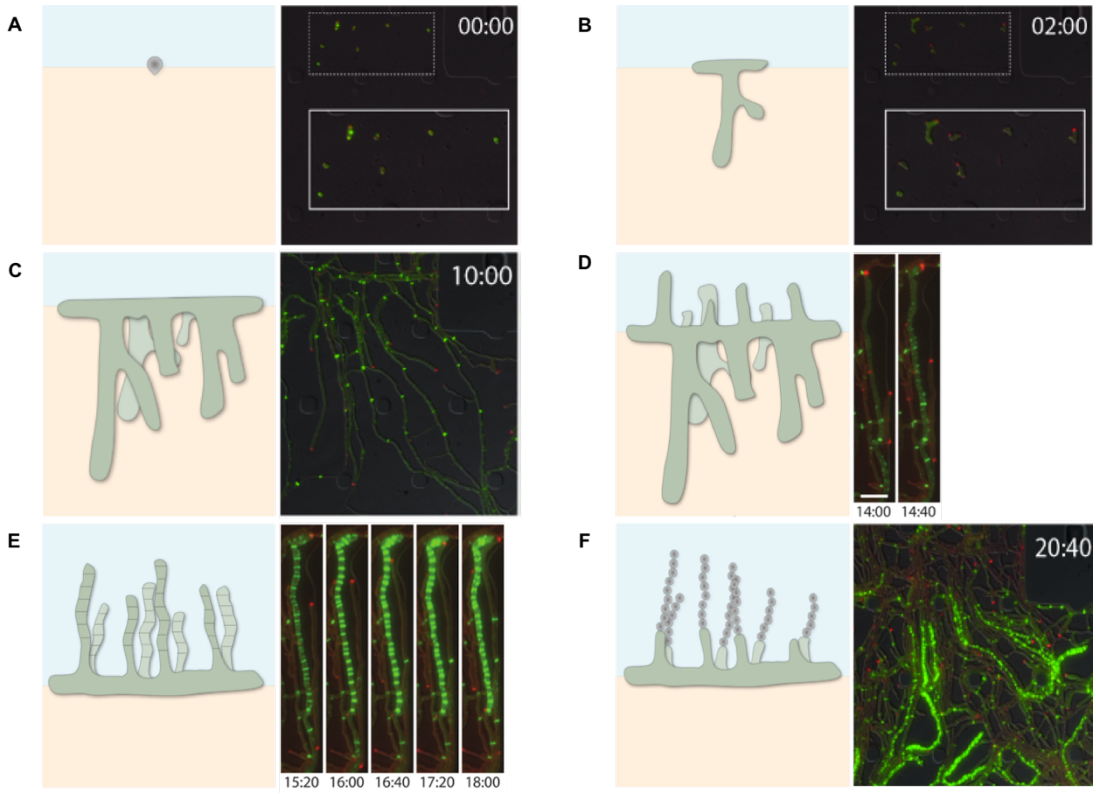


Figure 1.3 Overview of the *Streptomyces* life cycle (left) and fluorescence time-lapse of *S. venezuelae* life cycle using a microfluidic device (right).

A. Free spore growth; **B.** Spore germination tip extension (protein DivIVA, red colour); **C.** Branching of vegetative hyphae and mycelium growth; **D.** Formation of the aerial hyphae; **E.** Segregation of the chromosomes and formation of the sporulation septs; **E.** Spores maturation containing one chromosome copy each. *Adapted from Bush et al. (2015) and Schlimpert, Flärdh and Buttner (2016).*

1.3.2 Morphological regulation and specialised metabolism

Until now, most research has been performed in the model organism *S. coelicolor* to uncover relevant genes. Thus, several *bld* genes have been identified. The *bldA* gene encodes a tRNA molecule that translates the rare leucine codon UUA, unnecessary for growth and not read by any other tRNA molecule but required for a number of cluster-situated regulators of specialised metabolites by regulating the transcription of *bldH* that encodes a regulator of development and antibiotic production (Kim *et al.*, 2005; Guyet *et al.*, 2014; Chater & Chandra, 2008). *bldB* encodes a small peptide and exhibits a pleiotropic effect with mutations on this gene blocking aerial hyphae and synthesis of antibiotics (Tschowri *et al.*, 2014). *bldC* encodes a protein important for the production of actinorhodin (Hunt *et al.*, 2005). The master regulator of development, BldD is a transcription factor controlled by the signalling molecule cyclic-di-GMP (CDG), forming a functional complex that allows the bind to its targets (Tschowri *et al.*, 2014). Consequently, it can repress development related genes during vegetative growth. It can also play an important role in regulating the expression of *whiA* and *whiB* (Tschowri *et al.*, 2014). The gene *bldG* may play a role in aerial hyphae growth and production of pigmented antibiotics and *bldN* is required for the aerial mycelium formation (Bibb, Molle & Buttner, 2000). BldJ is a peptide that signals the start of the *bld* cascade but it requires the presence of the *bldK* gene cluster to import it (Lambert *et al.*, 2014).

Furthermore, the study of the early conversion steps from aerial hyphae to mature spores allowed the identification of other proteins. WhiG controls the development and growth of aerial mycelium and regulates *whiJ* and *whiH* (Mendez & Chater, 1987). Moreover, *whiJ* encodes a transcription factor with a similar function that BldD (Flärdh *et al.*, 2000). It also acts as an expression regulator of *whiH* (Flärdh *et al.*, 2000). The *whiE* gene cluster is responsible for production of the grey colour of the mature spores' coat. WhiA and WhiB halts the aerial hyphae growth promoting the septation and sporulation phases (Flärdh *et al.*, 2000). Septation encompasses the curling, partitioning and replication of the chromosomes in aerial hyphae. WhiH activates the protein FtsZ thus promoting the sporulation-specific cell division (Persson, Chater & Flärdh, 2013). The *ssgA* and *ssgB* genes are transcribed independently of the *whi* genes and are required for correct sporulation-specific cell division. SsgA activates the sporulation stage and SsgB is involved in septa and spore wall localization and

formation (Jiang & Kendrick, 2000; van Wezel *et al.*, 2000). Other important proteins are also involved, ParAB are required in the partitioning of DNA during septation (Jakimowicz *et al.*, 2005, 2007) and protein SmeA influences chromosomal segregation and condensation during sporulation (Ausmees *et al.*, 2007).

Streptomyces have regulatory mechanisms that are shared between morphological and physiological development. As *Streptomyces* undergo morphological differentiation several regulatory proteins that control the onset and production of antibiotics are activated responding to environmental and physiological signals (Bibb, 2005). These can be pleiotropic regulators that do not directly act on specific clusters but affect multiple gene clusters and transmit changes in the environment or in the cellular physiology to cluster-situated regulators (CSRs), that control the transcription of genes within a biosynthetic gene cluster (Niu *et al.*, 2016). Several molecules that form part of the primary metabolism control antibiotic production and morphological differentiation, e.g. *N*-acetylglucosamine that is released during autolysis of cell walls of *Streptomyces* or fungi, the phosphorylated guanosine nucleotide ppGpp that controls growth rate of gene expression and γ-butyrolactone (A-factor) required for both morphological and onset of secondary metabolism (Niu *et al.*, 2016; Bibb, 2005). AfsA is the responsible for A-factor biosynthesis that then binds to cytoplasmic protein ArpA and allows *adpA* transcription and activation of *strR*, the CSR for streptomycin production, and for expression of other genes required for sporulation (Niu *et al.*, 2016; Bibb, 2005).

1.4 Natural products and where to find them

Since ancient times, natural products from diverse origins have been used to improve human health. The first records of the use of natural products are oils from *Cupressus sempervirens* and *Commiphora* species in Mesopotamia. Furthermore, the use of natural products, mainly from medicinal plants, was widespread in most ancient civilizations, spanning ancient Egypt, Greece or China to middle-ages Europe (Dias, Urban & Roessner, 2012). In bacteria, most antimicrobial compounds originate from secondary metabolism and are normally not essential for the organism's development and survival. However, they possess a major role for adaptation to environmental changes or as defence mechanisms against others microorganisms (Katz & Baltz, 2016; Dias, Urban & Roessner, 2012). Although, an idea has emerged that these compounds have other roles other than inter-species chemical warfare mainly due to low quantities found in nature (Davies & Ryan, 2011). Thus, the perception is that these molecules could also function as a way of signalling and communication in microbial communities. Furthermore, studies have proved that sub-inhibitory concentrations of antibiotics change the transcription of the target strain and therefore act as signals (Davies & Ryan, 2011).

Nonetheless, through environmental selective pressures, including competition between strains, novel structural modifications could occur producing a remarkable diversity in bioactive compounds classes (Dias, Urban & Roessner, 2012). Penicillin, discovered by Fleming in the 20s, was the first microbial natural antibiotic produced in large scale and instigated the so-called golden age of antibiotic research, spanning between the 40s and the 70s (Scheffler *et al.*, 2013; Genilloud, 2014). During this period, phenotypic screening was the main approach to discover these bioactive compounds without prior knowledge of their mechanism of action. Essentially, samples from soil containing cultivable microorganisms were replicated, grown in fermentation flasks and the produced culture supernatant screened against susceptible test organisms. Alternatively, plugs of the microorganisms were also used. The potential activity was then observed by the presence or absence of an inhibition zone and the compounds were isolated and structurally characterised (Katz & Baltz, 2016).

Recently, technological innovations and/or exploration of untapped or extreme environments have been a source for metabolites. The latter can include a wide range of habitats such as deep seas, sediments, deserts, mountains, caves or even the human body that possess challenging environmental conditions (Wohlleben *et al.*, 2016). One such example is the salinosporamides produced by *Salinispora* species isolated from Bahamas sediments or closthioamide from *Clostridium cellulolyticum* isolated from compost (Williams *et al.*, 2005; Lincke *et al.*, 2010). However, only 0.1% of prokaryotes are thought to be cultivatable in laboratory conditions (Alain & Querellou, 2009). Culture conditions were changed, such as media components and supplements, and devices created to try to tackle this problem with some success recently using these alternative approaches through the discovery of a new antibiotic, teixobactin, using an innovative device called the isolation chip (iChip) (Nichols *et al.*, 2010; Ling *et al.*, 2015). It comprises hundreds of diffusion chambers in one chip, achieving high-throughput isolation, while assuring that only one cell grows in each chamber. Furthermore, this platform allows the growth of novel bacteria *in situ*, e.g. the producer of teixobactin *Eleftheria terrae*, not isolated by the traditional way (Nichols *et al.*, 2010; Ling *et al.*, 2015). Among extreme environments the Atacama Desert in Chile has gained recently attention. It is a hyper-arid desert combining very low amounts of water, due to lack of precipitation for years, and a very high solar radiation. These strict conditions force microorganisms to adapt and evolve tolerance to dehydration and intense radiation, among other specialisations (Bull & Asenjo, 2013; Neilson *et al.*, 2012; Maki *et al.*, 2019; Warren-Rhodes *et al.*, 2007). Several strains have been isolated and their bioactivity potential studied. Among the compounds identified, some showed antibacterial (chaxamycin D) and anti-inflammatory (abenquines) activity (Bull & Asenjo, 2013). Interestingly, all of these compounds identified were from actinobacterial species, mainly from the genus *Streptomyces* and one of these was a novel species, *Streptomyces leeuwenhoekii*.

1.5 Re-purposed and novel natural products

One consideration when searching for novel natural products is finding the same bioactive compounds already known and described. This aspect led to the decline of pharmaceutical companies' investment and research on this area. However, one way to tackle this problem is to improve the effectiveness of known antibiotics by synergism of two molecules or by synthetic biology (Brown, 2015). The former was applied when microorganisms developed mechanisms of resistance against a specific antibiotic but when this system is directly targeted by another molecule it allows for the antibiotic to take effect. One commercially available example is Augmentin (GlaxoSmithKline) that contains two drugs, amoxicillin and clavulanic acid. Amoxicillin, a derivative of ampicillin, showed antibacterial activity against Gram-negative however these bacteria subsequently developed resistance through the production of β-lactamases that degrade the antibiotic. Clavulanic acid showed low potency as an antibacterial agent but was a potent inhibitor of β-lactamases, thus combining both allowed the inhibition of β-lactamases and restoring amoxicillin antibacterial activity (White *et al.*, 2004; Brown, 2015).

Another strategy focuses on the molecule itself and the tailoring of natural product structure by changing side chains without modifying the main chain/backbone (**Figure 1.4**) (Walsh & Wencewicz, 2014). One example, oxytetracycline discovered during the so-called "golden age" from *S. rimosus*, had several iterations of tailoring creating doxycycline and then tigecycline aiming to tackle resistance to the previous generation of drugs. Other classes of antibiotics were modified synthetically, such as cephalosporins (from cefalotin to 5th generation ceftaroline) and macrolides (from erythromycin A to 4th generation solithromycin) (Walsh & Wencewicz, 2014). However, this approach can only achieve so much and the tailoring of existing natural products will eventually reach a limit (Walsh & Wencewicz, 2014). As a result, this approach can only supplement, but not replace the discovery of novel natural products directly from nature.

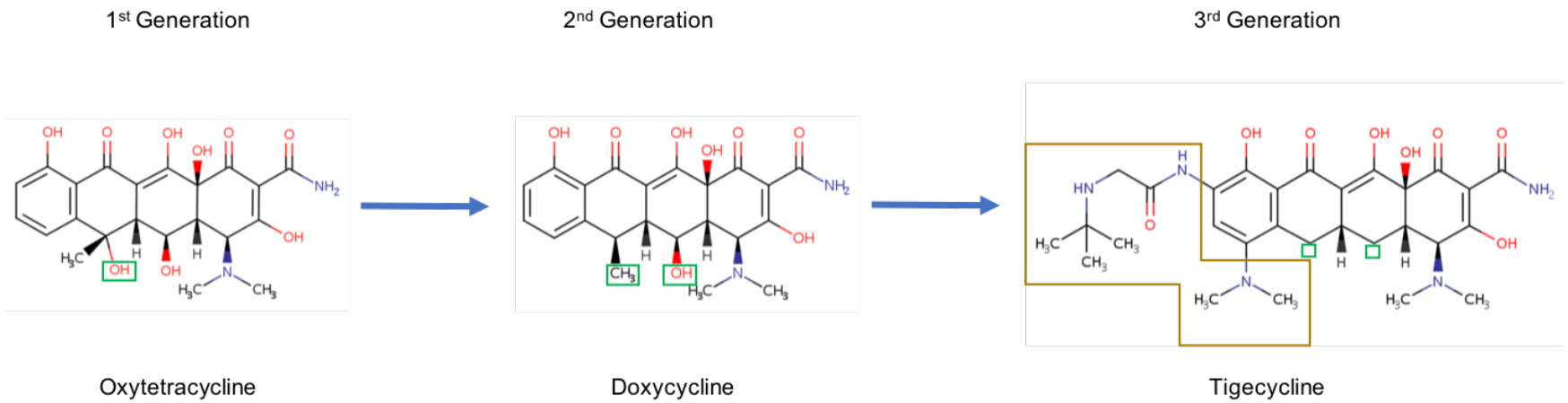


Figure 1.4 Synthetic generation of derivates using oxytetracycline as the source for the core structure.

Green boxes show loss of hydroxy and methyl groups and golden boxes the modifications in the structure to create tigecycline.

Searching new isolation locations (extremobiosphere) and elicitation methods can promote a faster pipeline for natural product discovery. The extremobiosphere encompasses environmental limits to life with strong selective pressures by reaching physico-chemical boundaries, such as high temperatures, salinity and depth, presence of toxic elements or intense radiation. It promotes a paradigm shift based on the idea that studying extreme environments could increase the prospect of finding new taxa and new chemistry (Bull & Goodfellow, 2019). For that, environments such as deep sea, polar, volcanic and desert environments at different levels of depth and extreme conditions were bioprospected for actinobacterial organisms (Bull & Goodfellow, 2019).

In all the locations biological and chemical diversity was found and new species of *Rhodococcus*, *Corynebacterium* and *Mycobacterium* described at around 3,000-9,000 meters below the sea level (Izo Bonin and Japan Trenches) and new species of *Pseudonocardia* and *Saccharothrix* at around 3,000 meters above sea level in the hyper-arid Atacama Desert (Bull & Goodfellow, 2019). These new taxa were able to produce new-in-a-class or first-in-a-class chemicals with a different range of bioactivities, from antibacterial to anticancer (**Table 1.2**) (Bull & Goodfellow, 2019). The full potential of each new taxon could still be untapped but by the one strain many compounds (OSMAC) approach using co-culturing allowed the synthesis of nine specialised metabolites (Bull & Goodfellow, 2019; Romano *et al.*, 2018). However, the discovery of these new metabolites still relies on the conventional strategy of bioassays, metabolomics and more recently whole genome sequencing and genome mining.

Table 1.2 Natural products discovered from deep sea and hyper-arid desert *Actinobacteria*.

Adapted from Bull and Goodfellow (2019).

Class	Natural product	Bioactivity	Location	Altitude (meters)	Reference
Diene glycosides	Lentzeosides	Anti-HIV	Cerro Chajnantor (Chile)	+ 5,046	Wichner *et al.*, 2017
Peptides	Chaxapeptin	Antibacterial and anticancer	Salar de Atacama (Chile)	+ 2,219	Elsayed *et al.*, 2015
Aminoheptosyl glycosides	Spicamycins A-E	Anticancer	Yungay (Chile)	+ 1,016	Hayakawa *et al.*, 1983
Beta-diketones	Asenjonamides A-C	Antibacterial	Yungay (Chile)	+ 1,016	Abdelkader *et al.*, 2018
Spirotetronate polyketides	Levantilide C	Anticancer	Golfo Corcovado (Chile)	- 200	Fei *et al.*, 2013
Benzoxazoles	Benzoxacystol	Anticancer	Canary Basin (Atlantic Ocean)	- 3,814	Nachtigall *et al.*, 2011
Phenazines	Dermacozines	Anticancer and antitrypanosome	Mariana Trench (Pacific Ocean)	- 10,897	Abdel-Mageed *et al.*, 2010

1.5.1 Advent of next-generation sequencing

The development of high-throughput next-generation sequencing (NGS) platforms lowered the cost and time necessary to have the whole information of a genome (whole genome sequencing) providing an overview of all the features of that organism by genome mining, e.g. biosynthetic gene cluster identification and phylogenetic relationships (Katz & Baltz, 2016; Baltz, 2016).

The main steps involve template preparation from the target DNA, sequencing and imaging using distinct approaches, e.g. Illumina or PacBio, and subsequent computational processing of the data obtained as reads (Metzker, 2010). The second-generation Illumina platforms follows this strategy and a characteristic feature is the bridge PCR. After fragmentation, each DNA fragment (500 bp) is primed with a sequence that has a complementary adapter in the flow cell. Then, each bound fragment is amplified by solid phase bridge amplification and bending of the fragments. Subsequently, differently labelled dNTPs are added to the flow cell and in each cycle occurs incorporation in to the DNA strand and detection by live imaging using four distinct laser channels (**Figure 1.5, panel A**). The fluorophores are then cleaved and washed, and the cycle begins again with nucleotide incorporation and imaging (Goodwin, McPherson & McCombie, 2016). However, third generation PacBio has a distinct approach and sequences long DNA molecules (up to 60 kbp). Ligation of hairpin adaptors at both ends of a target double-stranded DNA molecule form a closed single-stranded molecule, called SMRTbell, that is then loaded in to the SMRT cell containing individual zero-mode waveguide. These are photonic nano-structures consisting of a circular hole (waveguide) smaller than the wavelength of light decreasing the background noise and allowing efficient analysis of a single molecule at high fluorophore concentrations (Levene *et al.*, 2003; Vilgis & Deigner, 2018). Each waveguide as one single polymerase attached that binds to the hairpin adaptor and initiates DNA replication. For each dNTP incorporated, light pulses are imaged, and the sequence determined (**Figure 1.5, panel B**).

Each platform has different specifications with Illumina requiring more time to run and using short reads but presenting low error rate and PacBio as fastest with long reads but with higher error rate (**Table 1.3**) (Slatko, Gardner & Ausubel, 2018; Rhoads & Au, 2015). It is possible to use and combine assemblies obtained by two different

technologies, e.g. Illumina and PacBio, to increase the accuracy of the sequencing and in some instances assemble a complete bacterial genome sequence (Slatko, Gardner & Ausubel, 2018).

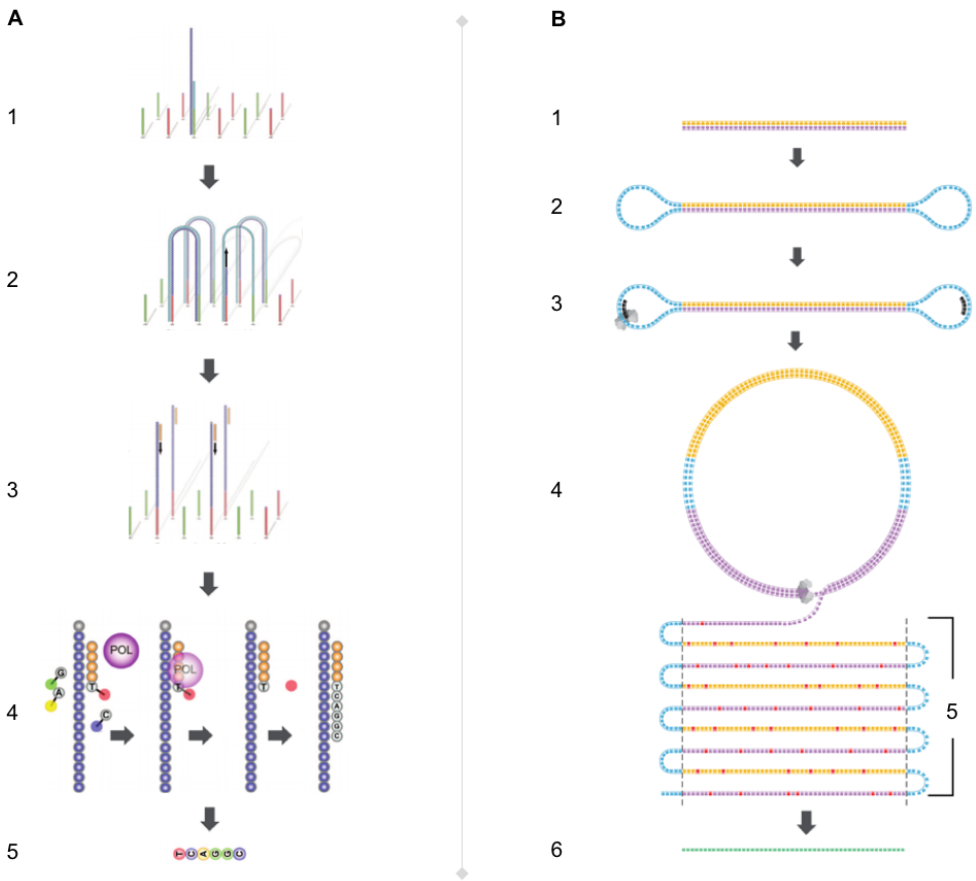


Figure 1.5 Overview of the Next generation Illumina (A) and PacBio sequencer (B) methodologies.

A. Illumina sequencing by synthesis. Short DNA fragments are ligated to adapters and the flow cell **(1)**. Bridge PCR amplifies the fragment obtaining thousands of copies **(2)**. Sequencing starts **(3)** with the addition of labelled dNTPs in the ssDNA strand by a polymerase **(4)** and the signal captured by a camera generating a consensus sequence read **(5)**. *Adapted from Voelkerding, Dames and Durtschi (2009)*. **B.** PacBio long molecule sequencing. The double-stranded DNA **(1)** is ligated to specific hairpin adaptors **(2)** creating the SMRTbell and loaded in the SMRT cell containing a polymerase **(3)**. The incorporation of labelled nucleotides by the polymerase allows live imaging and the molecules sequence several times **(4 and 5)** and generation of a consensus sequence read **(6)**. *Adapted from Wenger et al. (2019).*

Table 1.3 Comparison of high-throughput sequencing platforms.

Adapted from Rhoads and Au (2015).

Technologies	Run time	Read length (bp)	Error rate per read (%)	Cost per 10^6 bases (USD)	Reads per run
Sanger (1st generation)	up to 3 hours	Up to 1 kbp	0.001	500	96
Roche 454 (2nd gen)	up to 23 hours	Up to 700	1	8.57	1×10^6
Ion Torrent (2nd gen)	up to 4 hours	Up to 400	1	0.1	8.2×10^7
Illumina (2nd gen)	up to 6 days	Up to 2 × 300	0.1	0.03-0.04	$1.2 – 8 \times 10^9$
PacBio (3rd gen)	up to 4 hours	Up to 60 kbp	13	0.40-0.80	$3.5 – 7.5 \times 10^4$
Oxford Nanopore (3rd gen)	up to 50 hours	Up to 100 kbp	38	6.44-17.90	$1.1 – 4.7 \times 10^4$

1.5.2 Genomics as a tool for wider analysis

The revolution brought by whole genome sequencing, uncovered hidden genome features; especially the discovery of several cryptic or silent secondary metabolites pathways previously unknown (Katz & Baltz, 2016; Baltz, 2016). Notably, it was found that *Actinobacteria*, mainly the genus *Streptomyces*, have large genomes with many cryptic biosynthetic gene clusters. For instance, the *Streptomyces rapamycinicus* genome encodes 48 secondary metabolite biosynthetic gene clusters (around 3 Mbp or 24%) of a 12.7 Mbp genome (Katz & Baltz, 2016). Using this information, it is possible to try to elicit the activation or production from these gene clusters.

There are several ways of achieving this using biological, chemical or molecular elicitation (Abdelmohsen *et al.*, 2015; Yoon & Nodwell, 2014). One example of biological elicitation was the synthesis of the antibacterial rhodostreptomycins after co-cultivation of *Rhodococcus fascians* and *Streptomyces padanus* (Kurosawa *et al.*, 2008). Increased production of actinorhodin by *S. coelicolor* was also achieved using scandium (Sc^{3+}) as a chemical elicitor (Kawai *et al.*, 2007). Recently, next-generation sequencing and bioinformatics improved the analysis of these gene clusters and provided a new way to activate and produce these cryptic metabolites (Abdelmohsen *et al.*, 2015). Thus, a cryptic cluster encoding chattamycins A and B and produced by *Streptomyces chattanoogensis* L10 was activated after overexpressing a pathway-specific activator. Before, only two products from twelve polyketide biosynthetic gene clusters were identified (Abdelmohsen *et al.*, 2015).

Several bioinformatic tools were created and widely available that can analyse whole genomes for diverse features, autoMLST for phylogenetic relationships using housekeeping genes (Alanjary, Steinke & Ziemert, 2019), putative biosynthetic gene clusters for specialised metabolites, e.g. antiSMASH 5.0 (Blin *et al.*, 2019), BAGEL4 (van Heel *et al.*, 2018), PRISM 3 (Skinnider *et al.*, 2015) and NaPDoS (Ziemert *et al.*, 2012) and even prediction of resistance genes (Gupta *et al.*, 2014). All these tools enhanced the analysis of whole genomes and the potential to establish pipelines for natural product discovery coupled with a metabolomics analysis to identify the molecules responsible for the activity observed (**Figure 1.6**).

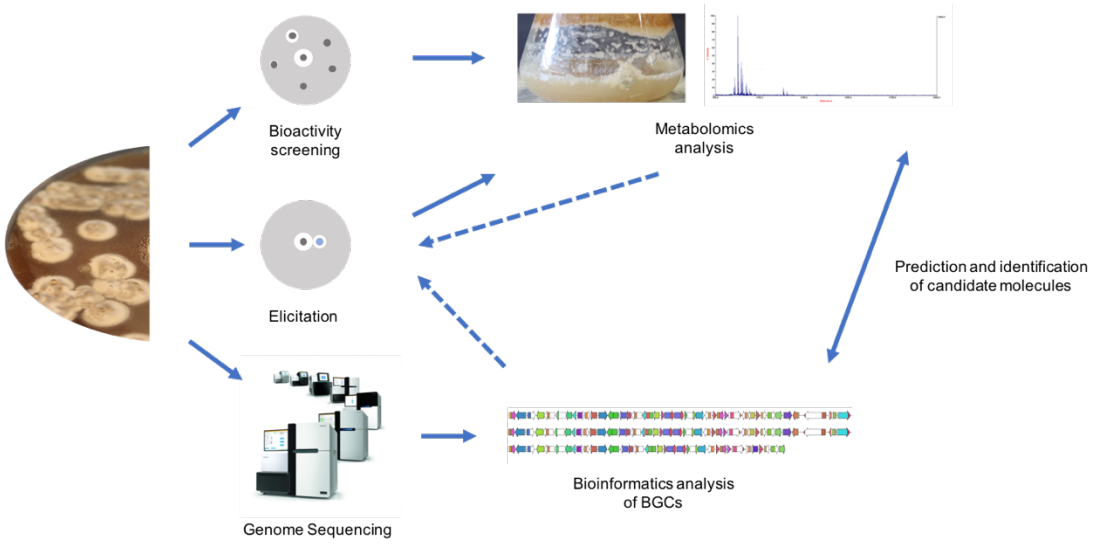


Figure 1.6 Overview of a pipeline combining metabolomics, genome mining and bioactivity for dereplication and natural product discovery.

1.6 Metabolomics and the search for new molecules

Metabolomics aims to understand and characterise metabolites chemically transformed during metabolism and linking both genotype and phenotype using analytical techniques (Patti, Yanes & Siuzdak, 2012; Lopes *et al.*, 2017). The combination of analytical chemistry, Mass Spectrometry (MS) and Nuclear Magnetic Resonance (NMR) spectroscopy provides a wide spectrum of information for data analysis using different tools such as MZmine (Pluskal *et al.*, 2010) and Mnova (Willcott, 2009) for analysis, process and visualisation of mass spectrometry or NMR data, respectively. Additionally, databases such as the Dictionary of Natural Products® (DNP) (Buckingham, 2001), ChemSpider (Pence & Williams, 2010), Global Natural Products Social Molecular Networking (GNPS) (Wang *et al.*, 2016) or in-house databases allow a fast analysis of the MS dataset, effectively lowering the cost and time necessary for dereplication of specialised metabolites (Harvey, Edrada-Ebel & Quinn, 2015; Villas-Bôas *et al.*, 2005; Bouslimani *et al.*, 2014). The processed data is then analysed by multivariate statistical analysis using programs such as SIMCA that allow analysis of differences between samples using unsupervised clustering, e.g. partial component analysis (PCA), or differences between samples based on activity observed using supervised clustering, e.g. orthogonal partial least squares discriminant analysis (OPLS-DA) (Macintyre *et al.*, 2014; Robotti & Marengo, 2016).

In mass spectrometry, gas-phase ionized samples are produced from liquid or solid phases by electrospray ionization (ESI) or matrix-assisted laser desorption (MALDI), respectively. Afterwards, each ion passes through the mass spectrometer, that could be ion trap (coupled with ESI) or time-of-flight (coupled with MALDI), according to their mass-to-charge (*m/z*) ratio and generating an electric current from the incident ions proportional to their abundance (Becker *et al.*, 2012). In ion-trap MS, these *m/z* ratios are obtained by the containing and fragmentation of the ions for a period of time (MS/MS) and with time of flight (TOF) after ionization. This technique is widely used in molecular biology and pharmaceutical research because it requires small volumes of samples and providing high sensitivity and identification of metabolites via mass spectra analysis and comparison (Aebersold & Mann, 2003; Glish & Vachet, 2003; Awad, Khamis & El-Aneed, 2015).

1.7 Scope of the project

The main aim of this thesis is to investigate the ability of *Actinobacteria* isolated from an extreme environment to produce previously unknown metabolites. The isolates will be further characterised using several growth conditions and tests to understand differences between them. The potential to produce antimicrobials will be assessed against selected pathogens. Furthermore, metabolomics and genome sequencing will be done to potentially link the bioactivity observed with the *in silico* analysis and metabolomics results.

- Investigation of the phylogenetic relationship and diversity of similar isolates;
- Characterise morphologically and chemically the Atacama isolates;
- Analysis of specialised metabolism potential with bioactivity assays;
- Isolation and elucidation of secondary metabolite structure by Mass Spectrometry and Nuclear Magnetic Resonance;
- Whole genome sequencing using next-generation sequencing platforms, e.g. Illumina and PacBio;
- Bioinformatics analysis to characterise features of the isolates including biosynthetic gene clusters;
- Understanding underlying interactions between isolates from geographically distant soil samples, and also within the same location

Chapter 2. Materials and Methods

2.1 General microbiology

2.1.1 Microbial strains, plasmids and oligonucleotides

Soil samples from two locations were collected and the isolation of *Actinobacteria* was achieved using several isolation media targeting *Actinobacteria*. The soil named as 7 was taken at Piedras Rojas near an Andean plant at 4052.9 meters (23º 54' 15''S, 67º 42' 31'' W) and soil 8 adjacent to a clump of Peruvian Feathergrass (*Jarava ichu*) at Miscanti Laguna at 4192.2m of altitude (23º 44' 44'' S, 67º 57' 21'' W). Seven isolates from each soil sample were used for this project (**Table 2.1**).

To identify putative bioactivity of the isolates multidrug resistant clinical isolates (ESKAPE pathogens) and other indicator strains were used (**Table 2.2**).

All oligonucleotides used in this project are listed in **Table 2.3**.

Table 2.1 Bacterial strains used in this project.

Strain	Source	Reference
Streptomyces phaeoluteigriseus	DSM 41896	Goodfellow *et al.*, 2007
PH2	Soil No. 7	This study
PH3	Soil No. 7	This study
PH4	Soil No. 7	This study
PH5	Soil No. 8	This study
PH6	Soil No. 8	This study
PH12	Soil No. 8	This study
PH26	Soil No. 8	This study
PH27	Soil No. 7	This study
PH28	Soil No. 7	This study
PH31	Soil No. 7	This study
PH45	Soil No. 8	This study
PH70	Soil No. 7	This study
J57	Soil No. 8	This study
J58	Soil No. 8	This study

Table 2.2 Model strains for bioactivity assays.

Indicator strains	Reference
Enterococcus faecalis	ATCC 51299
Staphylococcus aureus	ATCC 43300
Klebsiella pneumoniae	ATCC 700603
Acinetobacter baumannii	ATCC 19606
Pseudomonas aeruginosa	ATCC 27853
Escherichia coli	ATCC 25922
Bacillus subtilis	ATCC 23857

Table 2.3 Primers used in this project for 16S rRNA amplification and sequencing.

Name	Sequence	Tm (ºC)	Amplicon (bp)	Reference
BAC27S	5'-AGAGTTTGGATCMTGG-3'	55	1,400	(Jiang *et al.*, 2006)
UNIV1492	5'-CGGTTACCTTGTTACGACTT-3'			

2.1.2 Isolation, maintenance and cultivation conditions

All media used were sterilised by autoclaving for 15 minutes at 121°C. For antibiotics and heat sensitive substances filter sterilisation was performed (0.22 μM filter).

The International *Streptomyces* Project (ISP1-ISP7) media (Shirling & Gottlieb, 1966) were used for the identification of the isolates. All the media used, and their respective compositions are described below (**Table 2.4**).

Glycerol spore stocks were obtained following a modified version of Kieser (2000), unless otherwise stated. Briefly, the isolates were grown for 14 days at 30ºC. Sterile distilled water (5 mL) was poured on to the plates and the spores removed with a sterile cotton bud. The solution was then filtered to a 50 mL falcon tube using a sterile syringe containing cotton wool. These were centrifuged at 1,462 x*g* for 5 minutes at 4ºC and the pellet dispersed with 1 or 2 mL of glycerol (20%, v/v). The spore suspension was transferred to cryovials and stored at -80ºC.

For non-*Streptomyces* bacteria, a single colony was grown overnight in liquid culture and 500 μL of culture aliquoted to a cryovial with 500 μL of glycerol (50%, v/v) and stored at -80ºC.

Table 2.4 Culture media for growth and maintenance of strains.

Media	Formula
Tryptone-yeast extract broth (ISP1)	5.0g Tryptone 3.0g Yeast extract 1000 mL distilled water
Yeast extract-malt extract agar (ISP2)	3.0g Malt extract 3.0g Yeast extract 10.0g Dextrose 5.0g Peptone 20.0g Agar 1000 mL distilled water
Oatmeal agar (ISP3)	Cook 20.0g Oatmeal for 20 min Filter and restored the volume (1000 mL of distilled water) Add 1 mL trace salt solution* and adjust pH to 7.2 Add 18.0g Agar 1000 mL distilled water
Inorganic salts-starch agar (ISP4)	10.0g Starch 1.0g Dipotassium phosphate 1.0g Magnesium sulphate.7H_2O 1.0g Sodium Chloride 2.0g Ammonium sulphate 2.0g Calcium carbonate Add 1 mL trace salt solution* 20.0g Agar 1000 mL distilled water
Glycerol asparagine agar (ISP5)	1.0g L-Asparagine 1.0g Dipotassium phosphate Add 1 mL trace salt solution* 20.0g Agar 10.0 mL Glycerol 1000 mL distilled water
Peptone-yeast extract iron agar (ISP6)	36.0g Peptone 1.0g Yeast extract 20.0g Agar 1000 mL distilled water

Tyrosine agar (ISP7)	1.0g L-Asparagine 0.50g L-Tyrosine 0.50g Dipotassium phosphate 0.50g Magnesium sulphate.7H_2O 0.50g Sodium chloride 0.01g Ferrous sulphate.7H_2O 1.0g Trace salt solution** 20.0g Agar 1000 mL distilled water
Mannitol Soya flour (Hobbs *et al.*, 1989)	16.0g Mannitol 16.0g Soya flour 16.0g Agar 1000 mL Tap water
Nutrient broth (Oxoid)	1.0g 'Lab-Lemco' powder 2.0g Yeast extract 5.0g Peptone 5.0g Sodium chloride 1000 mL distilled water
Nutrient agar (Oxoid)	1.0g 'Lab-Lemco' powder 2.0g Yeast extract 5.0g Peptone 5.0g Sodium chloride 15.0g Agar 1000 mL distilled water
Soft nutrient agar	1.0g 'Lab-Lemco' powder 2.0g Yeast extract 5.0g Peptone 5.0g Sodium chloride 7.0g Agar 1000 mL distilled water
LB broth (Bertani, 1951)	10g Tryptone 5g Yeast extract 10g Sodium Chloride 1000 mL distilled water
LB agar (Bertani, 1951)	10g Tryptone 5g Yeast extract 10g Sodium Chloride 1000 mL distilled water

2x YT (Kieser, 2000)	16g Tryptone 10g Yeast extract 5g Sodium Chloride 1000 mL distilled water

* 1g Ferrous sulphate.$7H_2O$; 1g Manganous chloride.$4H_2O$; 1g Zinc sulphate.$7H_2O$

** 1.36mg Ferrous sulphate.$7H_2O$; 0.027mg Copper chloride.$2H_2O$; 0.04mg Cobalt chloride.$6H_2O$; 0.025mg Sodium molybdate.$2H_2O$; 0.02mg Zinc chloride; 2.85mg Boric acid; 1.80mg Manganese chloride.$4H_2O$; 1.77mg Sodium tartarate

2.2 Characterisation of the isolates

2.2.1 Cultural and phenotypic characteristics

To characterise and understand differences between isolates from both soils their ability to grow in Mannitol Soya flour (MS) and ISP media 2 to 7 (HiMedia) was tested (Wink, 2012). Moreover, growth in presence of different concentrations of sodium chloride (0, 1, 2, 5, 7 and 9%, w/v) and incubation at different temperatures (4, 12, 22, 30, 37 and 50°C) were tested in ISP3 for 21 days (Wink, 2012). To evaluate pH tolerance ISP2 broth was inoculated with spores (~10^5) and grown for 7 days in pH 4.0-9.0 (at 1-unit intervals) using Xu *et al.* buffer system (Xu *et al.*, 2005).

Plates were photographed and aerial (top colour of the plates) and substrate mycelial colours and diffusible pigments (reverse colour of the plates) were determined using sRGB centroids comparison against the Inter-Society Colour Council-National Bureau of Standard Colour charts (ISCC-NBS) (Centore, 2016). The conversion uses the Munsell system to correspond sRGB coordinates to object colours by assuming these are viewed under Illuminant C, the same as ISC-NBS. A centroid of sRGB triples that correlate to a particular ISCC-NBS category is then calculated using a Euclidean metric on sRGB space (Centore, 2016). Then, the mean sRGB of the different media for each isolate was used to build a dendrogram (performed by Fraser MacFarlane and Dr Paul Murray from the Hyperspectral Imaging Centre at the University of Strathclyde).

Further physiological and biochemical characterisation was carried out using API® stripes API ZYM and API 20E (bioMérieux) following manufacturer's instructions. Briefly, the isolates were grown for 3 days in ISP1 at 30ºC 220 rpm and were centrifuged 1,462 x*g*, 12ºC for 5 minutes and washed once with dH_2O. An incubation tray and lid were prepared for API ZYM and 20E by distributing water into the honeycomb to create a humid chamber and then the test strip was added. The pellets of each isolate were resuspended in 2 mL resuspension medium and emulsified. Drops of this solution were dropped in a new resuspension medium to visual turbidity of McFarland Standard 4 (by comparing with bioMérieux ready prepare standards)

and used to inoculate only the API ZYM stripe (2 mL). The rest of the emulsified suspension was topped up to 5 mL using dH_2O and used to inoculate the API 20E stripe. The API ZYM stripes were incubated 24 hours at 30ºC and the ZYM A and ZYM B added to each well and the results recorded after 30 minutes. For the API 20E the incubation period was extended until 48 hours and the results recorded. Afterwards, specific reagents were added to TDA (TDA reagent), IND (JAMES reagent), VP (VP 1 and VP 2) and GLU (NIT 1 and NIT 2) wells and the results recorded.

2.2.2 Hyperspectral Imaging analysis

Samples for Hyperspectral Imaging were prepared by spreading 10 μL of each isolate in 6-well plates containing ISP2-ISP5 in triplicate. After 14 days of incubation ISP6 and ISP7 were inoculated with the isolates and the plates incubated for 7 days more. The top and reverse of the plates were imaged using two pushroom hyperspectral imaging systems, the first sensitive to light in the visible range (between 400 and 1,000 nm) and the other in the near-infrared (1,000 and 1,700 nm).

The data analysis and the built dendrograms were generated by Ryan Greer, Fraser MacFarlane and Dr Paul Murray at the Hyperspectral Imaging Centre. An in-house thresholding algorithm was used to segment the bacterial colony from the media using every wavelength in the hyperspectral image (unpublished). Then, for every pixel on the segmented colony it was generated an average spectrum to build a dendrogram (MATLAB's dendrogram tool) by calculating Euclidean distances between the average spectra of each isolate.

2.2.3 Phylogenetic characteristics using 16S rRNA

The isolates were grown in ISP1 overnight at 30ºC and 250 rpm. Genomic DNA was extracted using the ISOLATE II Genomic DNA kit according to the manufacturer protocol (Bioline). The concentration of each sample was measured using Nanodrop 2000c Spectrophotometer (Thermo Scientific).

PCR-mediated 16S rRNA amplification was achieved following the standard MyTaq™ DNA polymerase protocol (Bioline) supplemented with dimethyl sulfoxide solution (2% final concentration) and 16S primers BAC27S and UNIV1492R (**Table 2.3**). DNA was amplified using a T100™ Thermal Cycler (Bio-Rad) with an initial denaturation step at 95ºC for ten minutes, 35 cycles of denaturation at 95°C for 45 seconds, 55°C for 30 seconds for primer annealing and extension at 72 °C for 1 minute, followed by a final extension step at 72ºC for ten minutes. The amplicons were analysed by agarose gel electrophoresis (1%), followed by purification using the QIAquick® PCR Purification kit (Qiagen), and sent to sequence at Eurofins Genomics (Germany). The obtained 16S rRNA gene sequences were assembled using CAP3 sequence program (Huang & Madan, 1999) and identified with the EZBioCloud database (Yoon *et al.*, 2017).

Phylogenies were inferred by the GGDC web server (Meier-Kolthoff *et al.*, 2013) using the DSMZ phylogenomics pipeline (Meier-Kolthoff *et al.*, 2014) and *Kitasatospora setae* KM-6054 (accession number: AB022868) selected as outgroup. A multiple sequence alignment was created with MUSCLE (Edgar, 2004). Maximum likelihood (ML) and maximum parsimony (MP) trees were inferred from the alignment with RAxML version 8 (Stamatakis, 2014) and TNT (Goloboff, Farris & Nixon, 2008), respectively using GTR+GAMMA substitution model. For ML, rapid bootstrapping in conjunction with the autoMRE bootstopping criterion (Pattengale *et al.*, 2009) and subsequent search for the best tree was used. For MP, 1,000 bootstrapping replicates were used in conjunction with tree-bisection-and-reconnection branch swapping and ten random sequence addition replicates. The sequences were checked for a compositional bias using the X^2 test as implemented in PAUP* (Swofford, 2002).

2.2.4 Pulse Field Gel Electrophoresis

Pulse Field Gel Electrophoresis (PFGE) was carried out according to Kieser (Kieser, 2000) but replacing Tris-containing buffer with non-Tris buffer (HEPES) to decrease DNA degradation following a previously described method (Evans & Dyson, 1993). Briefly, all isolates were grown at 30ºC, 250 rpm for 2 days in ISP1 and the mycelia

were harvested by centrifugation at 1,462 x*g*, 5 min and washed twice with HES buffer (25 mM HEPES-NaOH, 25 mM EDTA, 0.3M Sucrose pH 8.0). Then, 2 mL of HES buffer was added and the OD_{600} adjusted to 1.8-2.0. A mixture containing 500 µL of mycelia and 400 µL of 1.5% agarose solution (BioRad) was prepared and 200 µL were added to plug moulds and allowed to solidify for 30 min. Plugs were then placed in 1.5 mL Eppendorf tubes and treated with HES buffer containing 1 mg/mL of lysozyme at 37ºC for 2 hours. Subsequently, the plugs were washed three times with NDS (1% N-laurylsarcosine, 0.5M EDTA, 10 mM Glycine, pH 9.5) and incubated overnight at 50ºC with NDS containing 1 mg/mL of Proteinase K and 1 mM calcium chloride ($CaCl_2$). The plugs were then cooled to 4ºC and incubated with HE buffer (10 mM HEPES-NaOH, 1 mM EDTA, pH 8.0) containing 0.1 mM PMSF for 1 hour at 4ºC. Subsequently, a series of three washes (1h each) with HE buffer were performed, and the plugs were kept for subsequent electrophoresis at 4ºC with fresh HE buffer.

Electrophoresis was carried out using HEPES buffer (1X) and 1% agarose solution in BioRad CHEF-DR II Pulse Field Electrophoresis System. The plugs were placed on the edge of the comb teeth and the agarose gel poured onto the gel cast. The electrophoresis gel was run for 34 hours, with a current of 195-200 mA and voltage gradient of 4 V/cm. The initial and final switch time were set to 70 and 130 seconds, respectively. Afterwards, the gel was stained for 30 min with ethidium bromide (1 µg/mL) and then de-stained with water for 1h. The gel pictures were captured with a UV transilluminator.

2.3 Genome extraction and sequencing

For high quality genomic DNA for genome sequencing a modified phenol/chloroform protocol was used (Salvà-Serra *et al.*, 2018). The isolates were grown in ISP1 overnight 250 rpm at 30ºC and after centrifugation of the cultures the pellet was resuspended with 300 μL of TE buffer (50 mM Tris pH 8.0, 20 mM EDTA) containing lysozyme (5 mg/mL) and RNaseA (0.1 mg/mL) and incubated for 30 min at 37ºC. Fifty microliters of SDS (10%, w/v) was added and thoroughly mixed. Subsequently, 85 μL of NaCl (5M) was added followed by 400 μL of phenol/chloroform/isoamyl alcohol (25:24:1, v/v) and the cells vortexed for 30 seconds. A centrifugation was carried for 10 min at 5,304 x*g*. The aqueous phase was transferred to a new tube and 400 μL chloroform/isoamyl alcohol (24:1, v/v) added, followed by 30 sec vortex and 10 min centrifugation at 5,304 x*g*. Once more the aqueous phase was transferred to a new tube containing 500 μL of isopropanol and mixed by inversion. After incubation at room temperature for 5 min the DNA was pelleted by centrifuging 10 min at 8,288 x*g*. One millilitre of ethanol (70%) was added to wash the pellet and the DNA dried on the bench for 30 min to an hour. The DNA was then resuspended in EB buffer (10 mM Tris-HCl, pH 8.5) or dH_2O. The concentration of DNA was determined using Qubit 2.0 Fluorometer (Invitrogen by Life Technologies) and quality analysed by agarose gel electrophoresis (1%).

The samples ranging from 1.19 to 20.4 ng of DNA, depending on the isolate, were sent to MicrobesNG (Birmingham, UK) that carried out sample preparation, sequencing using Illumina platform and genome assembly. Briefly, three beads were washed with extraction buffer containing lysozyme and RNase A, incubated for 25 min at 37ºC. Proteinase K and RNase A were added and incubated for 5 min at 65ºC. Genomic DNA was purified using an equal volume of SPRI beads and resuspended in EB buffer. DNA was quantified in triplicates with the Quantit dsDNA HS assay in an Ependorff AF2200 plate reader. Genomic DNA libraries were prepared using Nextera XT Library Prep Kit (Illumina, San Diego, USA) following the manufacturer's protocol with the following modifications: two nanograms of DNA instead of one were used as input, and PCR elongation time was increased to 1 min. DNA quantification and library preparation were carried out on a Hamilton Microlab STAR automated liquid handling system. Pooled libraries were quantified using the Kapa Biosystems Library

Quantification Kit for Illumina on a Roche light cycler 96 qPCR machine. Libraries were sequenced on the Illumina HiSeq using a 250 bp paired end protocol. Reads were adapter trimmed using Trimmomatic 0.30 (Bolger, Lohse & Usadel, 2014) with a sliding window quality cut-off of Q15 and *de novo* assembly was performed on samples using SPAdes version 3.7 (Bankevich *et al.*, 2012), and contigs were annotated using Prokka (Seemann, 2014).

Additionally, isolate PH28 and PH5 were re-sequenced (1-5 μg) by NU-OMICS (Northumbria University, UK) that carried out sample preparation, sequencing using PacBio platform and genome assembly. Briefly, genomic DNA was size sheared using Covaris g-Tube (~9 Kbp) and size distribution visualised using Agilent Bioanalyzer. ExoVII was used to remove ssDNA and obtain blunt ended dsDNA. Ligation of blunt hairpin SMRTbell adapters to the polished fragmented gDNA and ExoIII treatment to remove failed ligation products. Association of the adapters and gDNA with PacBio ampure magnetic bead and annealing of sequencing primer (v3) to the SMRTbell templates. PacBio Polymerase binding reaction to SMRTbell template carried out using the Sequel Binding kit 2.1 and sequenced used PacBio Sequel instrument. Samples were sequenced using 10 h movie capture and BAM files of CCS reads used for assembly in HGAP4.

2.4 Whole genome analysis

2.4.1 Multilocus sequence analysis (MLSA), Automated Multi-Locus Species Tree and average nucleotide identity

Due to the limitation of 16S rRNA analysis, four housekeeping genes (*atpD*, *gyrB*, *recA* and *trpB*) were selected to construct a robust phylogenetic tree. The aminoacid sequences were extracted from the whole genome, aligned using ClustalX (2.0) (Larkin *et al.*, 2007) and concatenated using Sequence Matrix (1.8) (Vaidya, Lohman & Meier, 2011). Tree inference was achieve using IQ-TREE web server (Trifinopoulos *et al.*, 2016) with edge-linked partition and best substitution model determined automatically (WAG: *atpD*, *gyrB* and *trpB*; JTTDCMut+I: *recA*). Ultrafast bootstrap approximation (Minh, Nguyen & von Haeseler, 2013), single branch tests and 1,000 bootstraps alignments used . Additionally, an Multi-Locus Species Tree (MLST) was constructed in the autoMLST webserver using 80 housekeeping genes and default parameters (Alanjary, Steinke & Ziemert, 2019). For a wide genetic characterisation of the isolates including similarities and/or differences the average nucleotide identity (ANI) was calculated using PYANI (version 0.2.7; https://github.com/widdowquinn/pyani/) (Pritchard *et al.*, 2016)

2.4.2 Pan-genome analysis and Gene Ontology classification

PIRATE clusters annotated genes by iterative CD-HIT clustering to determine representative sequences for each cluster, followed by a BLAST all-vs-all sequence similarity search (Blast expectation e-value: $1e^{-5}$) (Bayliss *et al.*, 2019). Then, an iterative MCL clustering is performed over a wide range of amino-acid/nucleotide identity thresholds and classifies paralogous genes families (clusters produced at the lowest similarity threshold) into putative gene fission/fusion events or gene duplications while also providing cluster homology and allelic variance (clusters within only a single gene family at higher thresholds) (Bayliss *et al.*, 2019). The pan-genome of both locations, together, was constructed after annotation of the genomes with Prokka (Seemann, 2014) and over a range of amino acid identity values of 45-98%

to identify the presence and absence of genes, the gene content and allelic diversity (Bayliss *et al.*, 2019). GO FEAT (Araujo *et al.*, 2018) was used to perform functional annotation of the core genome obtained from PIRATE (Bayliss *et al.*, 2019). Homology sequence search is performed using BLAST (Blast expectation e-value: $1e^{-5}$) against EBI database to search for homologs and the alignment results mapped to UniProt, NCBI Portein, KEGG and Gene Ontology databases. Orthologs clusters, are classified in biological processes, molecular functions and cellular components depending on the associated Gene Ontology (GO) terms and accession numbers, enabling comparison of these between different organisms. Core genes were aligned using MUSCLE (version 3.8) (Edgar, 2004), cleaned of divergent and ambiguously aligned blocks from the sequence alignment using Gblocks (version 0.91.1) and minimum number of sequences for a conserved position of 50%, minimum number of sequences for a flank position of 85%, maximum number of contiguous non-conserved positions of 8 and minimum length of a block of 10 (Castresana, 2002). FastTree (version 2.1.10) (Price, Dehal & Arkin, 2010) was used to construct a maximum likelihood phylogenetic tree using GTR with gamma distribution as substitution model and 1,000 bootstraps inside the NGPhylogeny web server (Lemoine *et al.*, 2019).

2.4.3 Prophage prediction and genomic islands identification

PHASTER (Arndt *et al.*, 2016) was used to identify and compare prophages sequences encoded in the genome of the isolates. It performs BLAST searches against a custom prophage/phage database and phage-like genes are clustered into prophage regions and non-phage genes undergo a second BLAST search against a non-redundant bacterial protein database (Arndt *et al.*, 2016). Completeness (represented by scores) of prophage regions are assigned based on the number of phage genes in those regions (Arndt *et al.*, 2016). Islandviewer4 (Bertelli *et al.*, 2017) was used to predict clusters of genes that were potentially horizontally transferred, so-called genomic islands, by integrating the genomic islands predictions methods Islander (Hudson, Lau & Williams, 2014), IslandPick (Langille, Hsiao & Brinkman, 2008), SIGI-HMM (Waack *et al.*, 2006) and IslandPath-DIMOB (Hsiao *et al.*, 2003).

2.4.4 Complete genome sequence for PH28 and PH5 and comparative analysis

Illumina paired-end reads and PacBio reads obtained from sequencing were used to create a final hybrid assembly by using short and long reads in Unicycler (excluding contigs shorter than 100 bp, using bold bridging mode and polishing of contigs bigger than 1,000 bp with pilon) (Wick *et al.*, 2017) available at Galaxy web platform (Afgan *et al.*, 2018). The assembly quality was assessed for the N50, total length and largest contig using Quast (Galaxy web platform; version 5.0.2) (Gurevich *et al.*, 2013) and Bandage (version 0.8.1) (Wick *et al.*, 2015) and annotated using Prokka (Galaxy web platform; version 1.13) (Seemann, 2014). The resulting GenBank files were loaded in Artemis (Carver *et al.*, 2012) and used to compare to the assembly obtained from the PacBio sequencing, which was considered as containing most of the genome. Annotated proteins at the end of the contigs from the hybrid assembly allowed were checked and compared to the scaffolds obtained using PacBio. Inverted scaffolds were re-ordered using SnapGene software (from GSL Biotech; available at snapgene.com). To close the gaps and obtain a complete genome, the missing nucleotide sequences were added from the hybrid assembly to the PacBio assembly using SnagGene software (from GSL Biotech; available at snapgene.com).

Afterwards, the assembly was aligned to the Illumina reads using Bowtie2 (Galaxy web platform; version 2.3.4.3) and to the PacBio reads using bwa-mem. The obtained BAM files were analysed with pilon (Galaxy web platform; version 1.20.1) (Walker *et al.*, 2014) to detect variants and polish the final genome assembly. To perform digital DDH for accurate delineation of species the Genome-to-Genome Distance Calculator (GGDC; version 2.1) (Meier-Kolthoff *et al.*, 2013; Meier-Kolthoff, Klenk & Göker, 2014) web service was used. Furthermore, Circular Genome Viewer (CGView) Server (Grant & Stothard, 2008) was used to create interactive genomics maps of the genomes of PH28 and PH5.

Additionally, OrthoVenn2 web server (Xu *et al.*, 2019) was used to compare and annotate orthologous clusters across the different replicons (E-value cut-off for all-to-all protein similarity comparisons: $1e^{-5}$ and inflation value for the generation of orthologous clusters: 1.5). Synteny, breaks and inversions between them were obtained from dot plots alignments using minimap2 software in D-Genies (Cabanettes & Klopp, 2018).

2.4.5 Whole genome sequence improvement and annotation

The contigs assembled from Illumina sequencing (**section 2.4**) were scaffolded *via* scaffolder MeDuSa (Bosi *et al.*, 2015) using the complete genome of isolate PH28 and PH5 as the reference genomes for soil 7 and soil 8, respectively. The new assemblies were annotated using Prokka (Galaxy web platform; version 1.13) (Seemann, 2014).

2.4.6 Secondary metabolites prediction

For rapid identification and analysis of specialised metabolite biosynthetic gene clusters (BGCs) the antibiotics and secondary metabolite analysis shell—antiSMASH 5.0 (Blin *et al.*, 2019) was used. This tool uses an annotated genome as input to identify gene clusters using specific gene signatures by protein alignments with the database. Only gene clusters within a certain cut-off distance defined by the detection rules for each type of gene cluster are shown corresponding to the minimal core features necessary for that BGC (Blin *et al.*, 2019).

Subsequently, similarities between isolates BGCs were determined using the Biosynthetic Genes Similarity Clustering and Prospecting Engine (BiG-SCAPE) (Navarro-Muñoz *et al.*, 2020). Briefly, BiG-SCAPE uses GenBank files containing biosynthetic gene clusters information to calculate a pairwise distance matrix between the gene clusters, based on their protein domain content and sequence identity, and to construct similarity networks of the gene clusters. Furthermore, it incorporates CORASON (CORe Analysis of Syntenic Orthologs to prioritize Natural Product) (Navarro-Muñoz *et al.*, 2020) that uses an aminoacid matrix to compare each genomic core of the BGC's and sort them based on evolutionary relationships and across gene cluster families.

2.5 Analysis of secondary metabolites production

2.5.1 Bioactivity assays against clinical pathogens

For the screening assays of bioactivity, glycerol stock solutions were obtained after growing the isolates on ISP3 for 7 days at 30ºC. Subsequently, glycerol (20%, v/v) was poured on each plate and the spores removed from the surface using sterile cotton buds. The aliquots were then used to obtain lawns of the isolates in different ISP media (ISP2-7) after 21 days. However, plates of ISP6 and IPS7 were only prepared after 14 days.

The antibacterial activity was evaluated against the multidrug-resistant clinical isolates known as ESKAPE pathogens: *E. faecalis*, *S. aureus*, *K. pneumoniae*, *A. baumannii*, *P. aeruginosa* and *E. coli* (**Table 2.2**). Additionally, *B. subtilis* was used. The optical densities used were different for each organism (OD_{600} of 0.1 or 0.01) to obtain confluent growth on the plate. Nutrient agar plates were overlaid with soft nutrient agar inoculated with the pathogen and let to dry. Subsequently, plugs with lawns of each isolate and media only plugs (controls) were placed on top of the overlaid plates and incubated overnight at 30ºC or 37ºC, depending on the organism's requirements. The next day, the inhibition zones were measured in centimetres (cm).

2.5.2 Screening for interactions between isolates using plug assay

An assay was performed to study the potential interactions between isolates from both soils. The plug assay was performed as described before (**section 2.5.1**) however the seeding of soft nutrient agar was made with each of the Atacama strains and incubated overnight at 30ºC.

2.5.3 Potential induction of the isolates by osmotic stress

Since the soil samples were collected from locations with high salinity, the induction of specialised metabolites was tested using a range of different salts. Thus, sodium chloride (0 to 450 mM), potassium chloride (0 to 300 mM) and magnesium sulphate (0 to 300 mM) were added to the test media ISP3 and lawns of selected strains grown for 21 days. Bioactivity assays were then performed in triplicate against the indicator strains using the plug assay method (**section 2.5.1**). Briefly, plugs from the isolate lawns were placed on top of seeded plates and incubated overnight at 30ºC or 37ºC, depending on the organism's requirements. The next day, the inhibition zones were measured.

2.6 Small and Large-scale Metabolomics analysis of the crude extracts for the isolates

2.6.1 Metabolomics analysis and bioactivity assays of crude extracts from all isolates

The media used for liquid fermentation were selected depending on the bioactivity observed as production (inhibition zones observed observed in the plates) and non-production media (no inhibition zones observed) (ISP2, ISP3, ISP5 and ISP7 – **Table 2.4**, without agar). Diaion® HP-20 resin was used as an adsorbent of produced metabolites during growth in liquid following a previous protocol (Purves *et al.*, 2016). Briefly, the resin was activated with ethyl acetate and shaken for 10 minutes then filtrated with filter paper and a Buchner funnel. After drying, 2.5g of resin was added to 50 mL media and autoclaved together.

Overnight cultures of the isolates were grown in ISP1, washed twice using deionised water, resuspended and inoculated in the respective media and grown for 7 days (180 rpm at 30ºC) in Erlenmeyer flasks. Subsequently, the cultures were centrifuged at 1,462 x*g* for 10 minutes at 4ºC. The supernatant was discarded and the pellet with resin/cells frozen overnight at - 80ºC. Lyophilisation was then carried out in a freeze dryer (Christ 100000/Alpha 2-4) for two days and the sample kept afterwards at 4ºC.

The metabolites from the resin were extracted with the addition of 20 mL ethyl acetate to the pellet under agitation for 1 hour at 100 rpm. Then, the ethyl acetate was transferred to a scintillation vial and dried under nitrogen (Stuart sample concentrator SBHC0NC/1 and block heater SBH130D/3). Another 20 mL of ethyl acetate was added to the pellet and mixed overnight in an Erlenmeyer flask at 100 rpm. The extract was transferred to the previously used scintillation vial and dried again using nitrogen. Afterwards the weight of crude extract was determined.

For bioactivity assays the crude extracts were solubilised with ethyl acetate and tested using the disk diffusion assay (100 μg per disk in triplicate). Additionally, disks

containing ethyl acetate, imipenem (10 μg), tetracycline (10 μg) and chloramphenicol (5 μg) were used as controls.
For mass spectrometry analysis, crude extracts were dissolved in acetonitrile to a final concentration of 1 mg/mL.

2.6.2 Mass spectrometry and data analysis

Crude extracts were run at a concentration of 1 mg/mL in MeOH:DCM (80:20, v/v) or 100% MeOH using an Exactive mass spectrometer with an electrospray ionization source attached to an Accela 600 HPLC pump with Accela autosampler and UV/Vis detector (Thermo Scientific, Germany). A mass range of 100-2,000 Da in the positive and negative ionization modes using a spray voltage of 4.5 kV and 10 µL of each sample injected. The column used was an ACE 5 C18 column (75 x 3.0mm; ACE-121-7503 Hichrom Limited, UK). A flow rate of 0.3 mL/min with a binary gradient used. The two solvents consisted of solvent A (distilled water containing formic acid 0.1%) and B (acetonitrile containing formic acid 0.1%) and was carried out for 45 minutes, at the start with 90% A and 10% B followed by a 30 min gradient to 0% A and 100% B. The UV absorption wavelength was set at 254 nm, the sample tray and column maintained at 4°C and 20°C, respectively. Solvent and media blanks were included in the analysis and run first. The data was acquired using Xcalibur (version 2.2; Thermo Scientific, Germany).

The obtained raw files were separated using MassConvert from ProteoWizard in negative and positive ionization mode data files. These were then processed separately using a modified version of MZmine (version 2.10) (Pluskal *et al.*, 2010) following a metabolomics data processing protocol (Macintyre *et al.*, 2014). Then, an in-house Excel macro was used to remove the culture media features (pairs of *m/z*) and dereplication on both set of data files performed as described before (Macintyre *et al.*, 2014). The macro contains a library created by employing an algorithm to use the molecular formula data set from Antibase® (February 2013), Marinlit® (September 2013), and DNP (December 2016) from which the monoisotopic exact masses were recalculated.

To perform statistical analysis and dendrogram generation SIMCA (version 15.0.2) (Umetrics, Umeå, Sweden) was used. For multivariate analysis an unsupervised statistical analysis with principal component analysis (PCA) was used to identify outlying *m/z* in the different bacterial extracts metabolomes, comprising *m/z* located further away from the main group of samples *m/z*, and the culture medium extract. Furthermore, a supervised statistical analysis with orthogonal partial least squares (OPLS) was used for regression analysis and grouping of extracts according to their bioactivity as active or inactive depending on the extract's bioactivity. Both methods were done applying Pareto scaling and the models validated by the multiple correlation (R^2) and cross-validation (Q^2) coefficients as well as by permutation tests for the supervised method. In this analysis, outlying *m/z* in the active side, *m/z* that were located further away from the main clusters, were selected for further analysis.

Additionally, heatmaps were created in R (version 3.03) program (R Foundation for Statistical Computing, Vienna, Austria) using the gplots package to understand the richness of the extract's metabolomes. The standard score (Z-score) of the *m/z* raw data was determined by extracting the mean and then dividing by the standard deviation of the mean-subtracted data.

2.6.3 Scale-up fermentation of isolate PH28

Isolate PH28 was selected based on the dereplication analysis showing great diversity and promising metabolomic profile that diverged from the other strains. For the scale-up fermentation this isolate was grown as described previously. Briefly, an overnight culture grown in ISP1 was washed twice using deionised water, resuspended and inoculated in 1 L Erlenmeyer flasks containing 300 mL of ISP2 and 15g of activated HP-20 resin and grown for 7 days (180 rpm at 30ºC). Afterwards, the culture was filtrated using a Buchner funnel to obtain the resin and a liquid-liquid partitioning was performed by washing three times the cells with HPLC grade ethyl acetate in a 1 L separating funnel (Fisher Scientific, Hemel Hempstead, UK). The organic solvent was evaporated in a rotary evaporator (BÜCHI, Switzerland) under vacuum, reconstituted in a minimum volume of ethyl acetate and transferred to a tared storage vial. The resin was frozen at -80ºC and freeze dried overnight (Christ Alpha 2-4 Martin Christ Gefriertrocknunsanlagen GmbH, Germany). Then the dried resin

was washed with ethyl acetate for a 1 hour and an overnight incubation followed by transfer to storage vials. At this point both liquid and resin extracts were dried over nitrogen.

2.6.3.1 Medium pressure liquid chromatography (MPLC) and thin layer chromatography (TLC)

For medium pressure liquid chromatography, the Isolera One 2.0.4 Spektra flash purification system (Biotage, Sweden) was used. It had a photodiode array (PDA) detector covering an UV wavelength range of 200-400 nm and four solvent channels allowing four solvents to be used in a binary gradient during a single run.

The extracts were dissolved in methanol and loaded in a 1g samplet, dried overnight and placed in the column SNAP Ultra 10g. The appropriate size of the column and the flow rate, which ranges from 4 to 200 mL were set according to the amount and complexity of the number of compounds to be separated, the pressure was automatically adjusted according to the size of column used. Fractionation of crude extracts from liquid were achieved using MeOH:DCM (80:20, v/v) and from resin with iPrOH:DCM (80:20, v/v) (**Table 2.5**). At the end of the runs, columns were then flushed with 100 mL of MeOH:Acetone (1:1, v/v) and collected as wash. Fractions were then pooled based on the UV read-outs for the respective peaks from the Chromatography System and their similarities analysed by normal phase TLC plate (Merck, Germany).

To the TLC chamber was added the solvents used before, MeOH:DCM and iPrOH:DCM, and the samples spotted 1 cm above the bottom edge of normal phase silica gel 60 F_{254} aluminium-plates. Chromatograms were developed until the eluting solvent reached 5 cm from the top of the plate. The plates were dried and observed under the UV lamp (short UV at 254 nm and long UV at 365 nm). TLC plates were also sprayed with Anisaldehyde/H_2SO_4 reagent (0.5 mL of anisaldehyde, 85 mL of methanol, 10 mL glacial acetic acid and 5 mL of concentrated H_2SO_4) and heated at 170°C-200°C using a heat gun to allow visualisation of organic compounds without UV absorbance such as steroids, terpenes or phenolic acids. Fractions considered similar were joined and dried.

A summary TLC plate was prepared, and fractions submitted to LC-MS analysis as described before **(section 2.6.2)**. Additionally, selected fractions were submitted for NMR at the Department of Chemistry of the University of Strathclyde and their structure elucidated by Dr RuAngelie Edrada-Ebel. Bioactivity was found by disk diffusion assay against Gram-positive *E. faecalis* where activity was found previously.

Table 2.5 Gradient elution of crude extract obtained using Biotage MPLC IsoleraTM One 2.0.4 Spektra flash purification. Dichloromethane was used as solvent A.

Solvent	Length (min)	Mix Solvent B (%)
Methanol	5	1
	2.5	1
	25	1-9
	1	9
	34	9-20
	10	20-30
	15	30
	Total time: 92.5	
Isopropanol	5	1
	10	1
	20	1-2
	30	2
	30	2-5
	10	5
	10	5
	50	5-20
	10	20-30
	10	30
	Total time: 185	

Chapter 3. Phylogeny, physiological and visual characterisation of *Streptomyces* isolates from the Atacama Desert

3.1 Introduction

Production of specialised metabolites, which are not essential for the organism's life-cycle, by microorganisms hold important roles in bacterial communication and as growth inhibitors for neighbouring cells (Romero *et al.*, 2011). The latter feature of these producers has important relevance in the Health industry and thus screening for antibiotic properties against clinically relevant pathogens. ESKAPE organisms for instance are able to resist a variety of antimicrobial agents, and that is one of the main focuses of studying bacterial specialised metabolites. Recently, the World Health Organization published the first list of the most critical antibiotic resistant pathogens, with emphasis on *A. baumannii*, *P. aeruginosa* and *Enterobacteriaceae* (such as *K. pneumoniae and E. coli*) (World Health Organization, 2017) that require immediate attention and discovery of new antibiotics.

Thus, environmental interactions between species in the same location play an important role in the production of specialised metabolites (Fajardo & Martínez, 2008; Yim, Wang & Davies, 2007; Traxler *et al.*, 2013). Recently, bistability was found in *Streptomyces* isolated from the same location suggesting that the production and counteraction of antibiotics can lead to coexistence and survival of the common (Wright & Vetsigian, 2016). Furthermore, this phenomenon can also be produced due to intraspecies cooperation when bacteria produce important metabolites for the population (Wright & Vetsigian, 2016). Nonetheless, heterogeneous distribution of nutrients and other environmental pressures also play an important role modulating these interactions.

Lately, it was observed that strains produced more antibiotic in response to cues from a competitor and were capable of inhibiting their growth and as such a form of

competition (Abrudan *et al.*, 2015). Additionally, previous works have shown that salinity can influence the development of bacteria and thus production of antibiotics. For instance, *Streptomyces* isolated from two saline environments were stressed with salts, showing inhibitory effects on growth rate (Killham & Firestone, 1984). In this genus there is a connection between morphological and physiological differentiation and, as such, effects on growth can influence the production of specialised metabolites. This was observed for the model organism *S. coelicolor* that produced more actinorhodin when sucrose was added to the media (Elibol & Mavituna, 1998). Another study uncovered the dual effect of osmotic stress in this bacterium by inducing the production of undecylprodigiosin and inhibiting actinorhodin (Sevcikova & Kormanec, 2004).

One of the oldest hyper arid deserts on Earth is the Atacama Desert in Chile and it is considered to be a landscape similar to those on Mars with several studies conducted there by NASA to develop methods to search for life on that planet (Bull *et al.*, 2016). The conditions present there create an extreme environment that push the limits for sustainable life comprising arid and hyper-arid regions to salt flats and the Chilean Altiplano at 3,500 meters holding some vegetation (Bull *et al.*, 2016). Water availability in the Atacama Desert varies depending on the proximity to rivers (e.g. Río Loa), marshes with groundwater from springs, *salares* or lagoons or through soil moisture by occasional rain or fog. Precipitation increases with altitude and at 5,000 meters above sea level (a.s.l.) the mean annual precipitation is 300 mm/a, at 2,300 a.s.l. it is less than 20 mm/a and below that it is hyper-arid with less than 1 mm/a (Houston, 2006; Bull *et al.*, 2016).

Strict conditions like these force the microorganisms inhabiting this environment to adapt and evolve tolerance to a range of conditions such as dehydration, salinity and intense solar radiation among other specialisations (Bull & Asenjo, 2013; Cortés-Albayay *et al.*, 2019c). These ecological adaptions and genetic differentiations are driven by horizontal gene transfer and habitat isolation promoting microbial speciation (Li *et al.*, 2019). Thus, gene-specific or genome-wide traits caused by selective pressures were observed in the study of 30 *Streptomyces albidoflavus* strains isolated from three distinct sources, insects, sea and soil (Li *et al.*, 2019). For each habitat, genes and alleles that increased the survival in these environments were found indicating a critical role of ecology in microbial speciation (Li *et al.*, 2019).

Interestingly, this extreme environment displayed its diversity with a range of new bacteria from Gram-negative *Pseudomonas* (Campos *et al.*, 2010), Gram-positive *Actinobacteria* as *Lechevalieria* (Okoro *et al.*, 2010), *Pseudonocardia* (Trujillo *et al.*, 2017) or *Streptomyces* (Busarakam *et al.*, 2014) to the archaea *Halorubrum* (Lizama *et al.*, 2002) being identified. Remarkably, six new *Streptomyces* species have been discovered in the Salar de Atacama (**Figure 3.1**), *Streptomyces atacamensis* and *Streptomyces desertii* (23°02'S 68°20'W), *Streptomyces bulli*, *Streptomyces leeuwenhoekii* and *Streptomyces asenjonii* (23°17'S 68°10'W) and *Streptomyces aridus* (23°63'S 67°52'W) (Santhanam *et al.*, 2012a, 2012b, 2013; Busarakam *et al.*, 2014; Goodfellow *et al.*, 2017; Idris *et al.*, 2017). Additionally, two other species were discovered in the Salar de Huasco (20º16'S 68º52'W), *Streptomyces huasconensis* and *Streptomyces altiplanensis* (Cortés-Albayay *et al.*, 2019a, 2019b).

These new species also showed potential to encode the ability to produce novel bioactive compounds such as new lasso peptides, atamycins and chaxapeptin, derived from *S. leeuwenhoekii* strain C38 and C58 with anticancer activity (Elsayed *et al.*, 2015; Nachtigall *et al.*, 2011a). Additionally, two novel compounds from *S. leeuwenhoekii* were identified, chaxamycins and chaxalactins active against MRSA and anti-Gram-positive, respectively (Busarakam *et al.*, 2014). The abenquines from *Streptomyces* sp. DB634 showed weak antimicrobial activity but potential application on inflammatory diseases (Schulz *et al.*, 2011). Considering this potential for novel taxa and chemistry soil samples from the extreme hyper-arid region of the Atacama Desert were collected, (i) Piedras Rojas at 4052.9 meters (soil 7; 23º 54' 15'' S, 67º 42' 31'' W) and (ii) Miscanti Laguna at 4192.2 meters of altitude (soil 8; 23º 44' 44'' S, 67º 57' 21'' W).

The main aims of this chapter were to establish the phylogenetic status of the strains isolated from the Atacama Desert that exhibited phenotypic similarities between both locations by employing (1) phylogenetic studies (16S rRNA), (2) physiological characterisation including pH, temperature and substrate utilisation as well as Pulse-field electrophoresis and hyperspectral imaging. Additionally, (3) their interaction profile among soil 7 and soil 8 and (4) potential effect of osmotic stress in inducing the production of specialised metabolites were also determined.

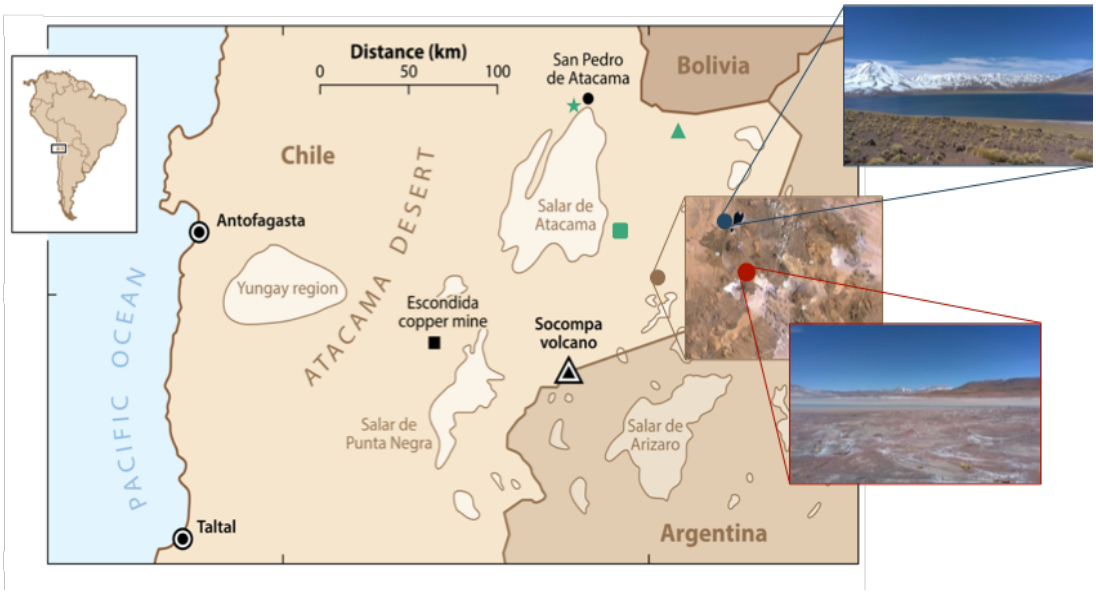


Figure 3.1 Sampling locations of soil 7 (red circle) and soil 8 (dark blue circle) in the Atacama Desert (Chile) compared to those of previous novel *Streptomyces* species. *S. atacamensis* and *S. desertii* (star), *S. aridus* (triangle) and *S. bulli*, *S. leeuwenhoekii* and *S. asenjonii* (square). *Adapted from Bull et al. (2016).*

3.2 Isolate identification by 16S rRNA gene sequencing

Soil samples from two locations (~30 km apart) from the Atacama Desert were collected and prospected for *Actinobacteria*, particularly *Streptomyces*, to search for producers of natural products. From these soil samples several isolates were identified as *Streptomyces phaeoluteigriseus* based on 16S rRNA sequencing and seven from each soil were selected for further analysis (**Table 3.1**). Overall, the isolates from both soils were identified with more than 99% 16S rRNA sequence identity to *Streptomyces phaeoluteigriseus* DSM 41896 (accession number: AJ391815) from an unknown soil source. Initially described as *Streptomyces violaceusniger* (Labeda & Lyons, 1991), subsequent phenotypic and 16S rRNA analysis showed that it was in fact a new species (Goodfellow *et al.*, 2007).

Phylogenetic trees allow some understanding of the relationship between different strains, for instance using housekeeping genes, and as such a maximum likelihood and parsimony phylogenetic trees were constructed using the 16S rRNA gene of the isolates and the type strain, *S. phaeoluteigriseus,* acquired from the German Collection of Microorganisms and Cell Cultures (DSMZ, www.dsmz.de/) (**Figure 3.2**). The phylogenetic tree analysis confirms the previous similarity results where the isolates are closely related forming a well delineated subclade supported by all phylogenetic tree algorithms and by a 97% bootstrap value. Two heterogeneous clades containing soil 7 and soil 8 isolates showed that there are no significant differences between isolates due to soil-specific selection. Furthermore, these isolates were clustered in a distinct clade than the novel Salar de Atacama species (green shapes) indicating that are not particularly closely related. This result shows that strains from both soils were closely related with each other despite spatially segregated using 16S rRNA. However, further studies were carried out to confirm this result since 16S rRNA can be unreliable for *Streptomyces* classification (Antony-Babu *et al.*, 2017).

Table 3.1 16S rRNA gene sequencing results including the similarity (%) obtained in EZBioCloud database.

	Isolate	Length (bp)	Similarity (%)	Top-hit taxon
Soil 7	PH2	1402	99.79	*S. phaeoluteigriseus* DSM 41896
	PH3	1402	99.64	*S. phaeoluteigriseus* DSM 41896
	PH4	1401	99.71	*S. phaeoluteigriseus* DSM 41896
	PH27	1403	99.71	*S. phaeoluteigriseus* DSM 41896
	PH28	1399	99.50	*S. phaeoluteigriseus* DSM 41896
	PH31	1400	99.57	*S. phaeoluteigriseus* DSM 41896
	PH70	1398	99.71	*S. phaeoluteigriseus* DSM 41896
Soil 8	PH5	1376	99.20	*S. phaeoluteigriseus* DSM 41896
	PH6	1402	99.79	*S. phaeoluteigriseus* DSM 41896
	PH12	1400	99.50	*S. phaeoluteigriseus* DSM 41896
	PH26	1403	99.64	*S. phaeoluteigriseus* DSM 41896
	PH45	1402	99.64	*S. phaeoluteigriseus* DSM 41896
	J57	1353	99.33	*S. phaeoluteigriseus* DSM 41896
	J58	1375	99.56	*S. phaeoluteigriseus* DSM 41896

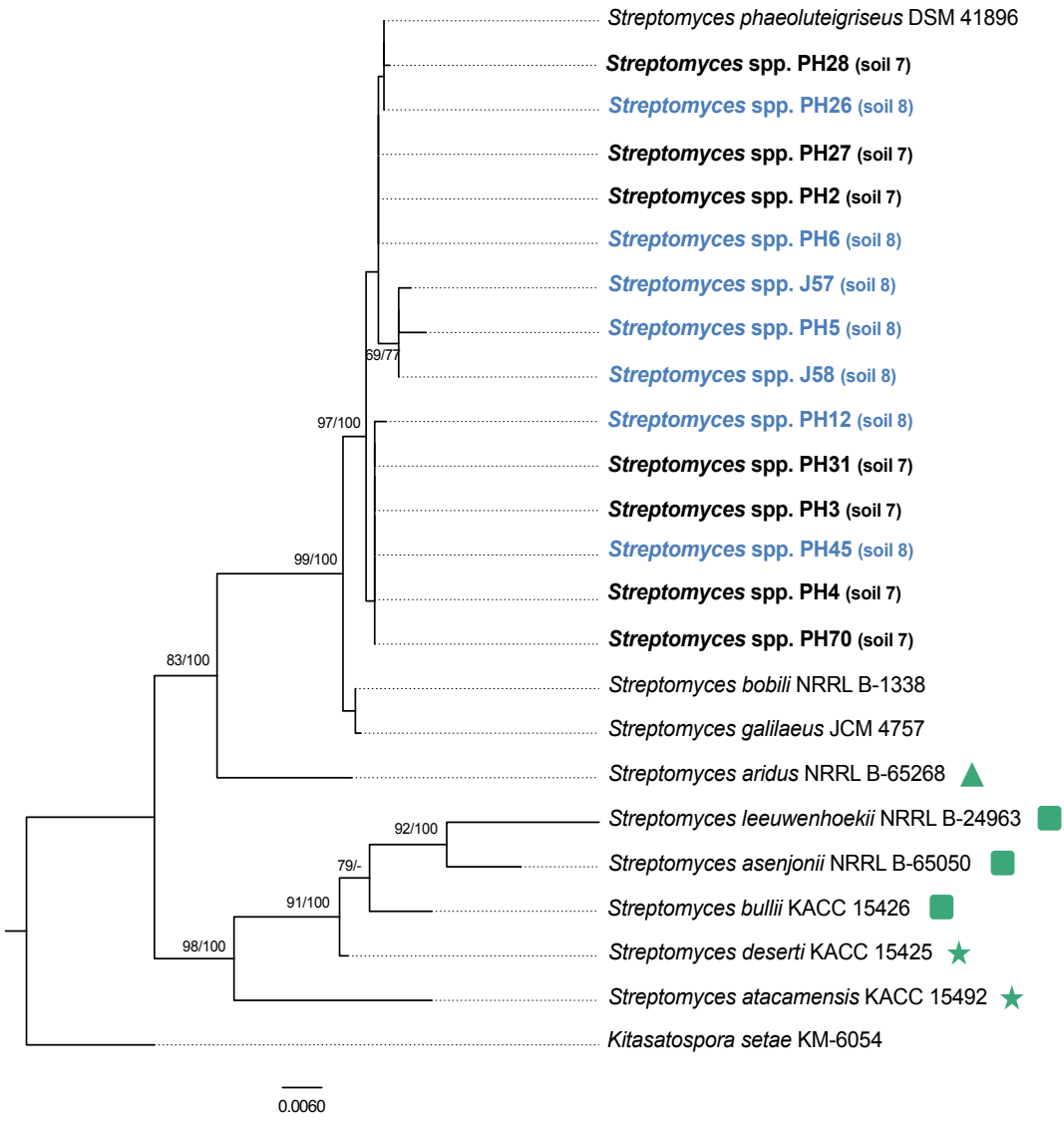


Figure 3.2 Maximum Likelihood tree for 16S rRNA sequences from Chilean isolates. Clades were inferred under a GTR+GAMMA model and using *Kitasatospora setae* KM-6054 as an outgroup. The branches are scaled in terms of the expected number of substitutions per site. The numbers above the branches are support values when larger than 60% from ML (left) and MP (right) bootstrapping. Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**). The known Salar de Atacama species are represented by a star (*S. atacamensis* and *S. desertii*), triangle (*S. aridus*) and square (*S. bulli*, *S. leeuwenhoekii* and *S. asenjonii*).

3.3 Phenotypic differences among the isolates from different locations

To understand the biological differences between isolates, several phenotypic analyses were carried out as, notwithstanding the similar 16S rRNA sequences, the isolates appeared phenotypically different (**Figure 3.3**). For this, the soluble pigment and the reverse colour produced by each isolate was determined by growth in different media (ISP2-ISP7), pictures were taken and the mean RGB identified for comparison (**Table 3.2**). In general, isolates from soil 7 had a darker reverse colour (dark reddish to brownish orange) compared to soil 8 with lighter colours (strong yellow brown to dark greyish yellow). The aerial spore mass was found to be light grey to yellowish grey in all the isolates (**Figure 3.3**). Interestingly, variations in colour within locations were observed, e.g. PH2-PH27 present an overall darker colour than the remaining strains of soil 7 (PH28-PH70) (**Figure 3.3** – top row) and the same was observed for PH5-PH26 with darker colour compared to other soil 8 isolates (PH45-J58) (**Figure 3.3** – bottom row). Furthermore, all the isolates exhibited good growth, especially in ISP3, and produced melanin in ISP6 (**Table 3.2-3.4**).

To quantify these colour dissimilarities a dendrogram was constructed using the average RGB values of each plates and isolates grown in ISP3 in triplicate (performed by Fraser MacFarlane and Dr Paul Murray from the Hyperspectral Imaging Centre at the University of Strathclyde). This technique was employed before to compare and dereplicate hundreds of isolated *Streptomyces* species (Antony-Babu, Stach & Goodfellow, 2010). Two clades emerged (**Figure 3.4**) with PH45 clustering alone in one of those corresponding to the distinct colour observed before (**Table 3.2**). On the other hand, two clades also arose for the other group, forming two distinct subgroups. Isolates PH2, PH3, PH4 and PH27 clustered closely together in one clade while the remaining isolates grouped in the other clade. Soil 8 isolates seemed to share a common colour with PH28, PH31 and PH70 that matched the previous observations (**Figure 3.4** and **Table 3.2**) but were still too distinct to be segregated.

These results contradict the 16S rRNA analysis, where the isolates shared >99% identity, demonstrating the potential phenotypic diversity between isolates from the same and different location.

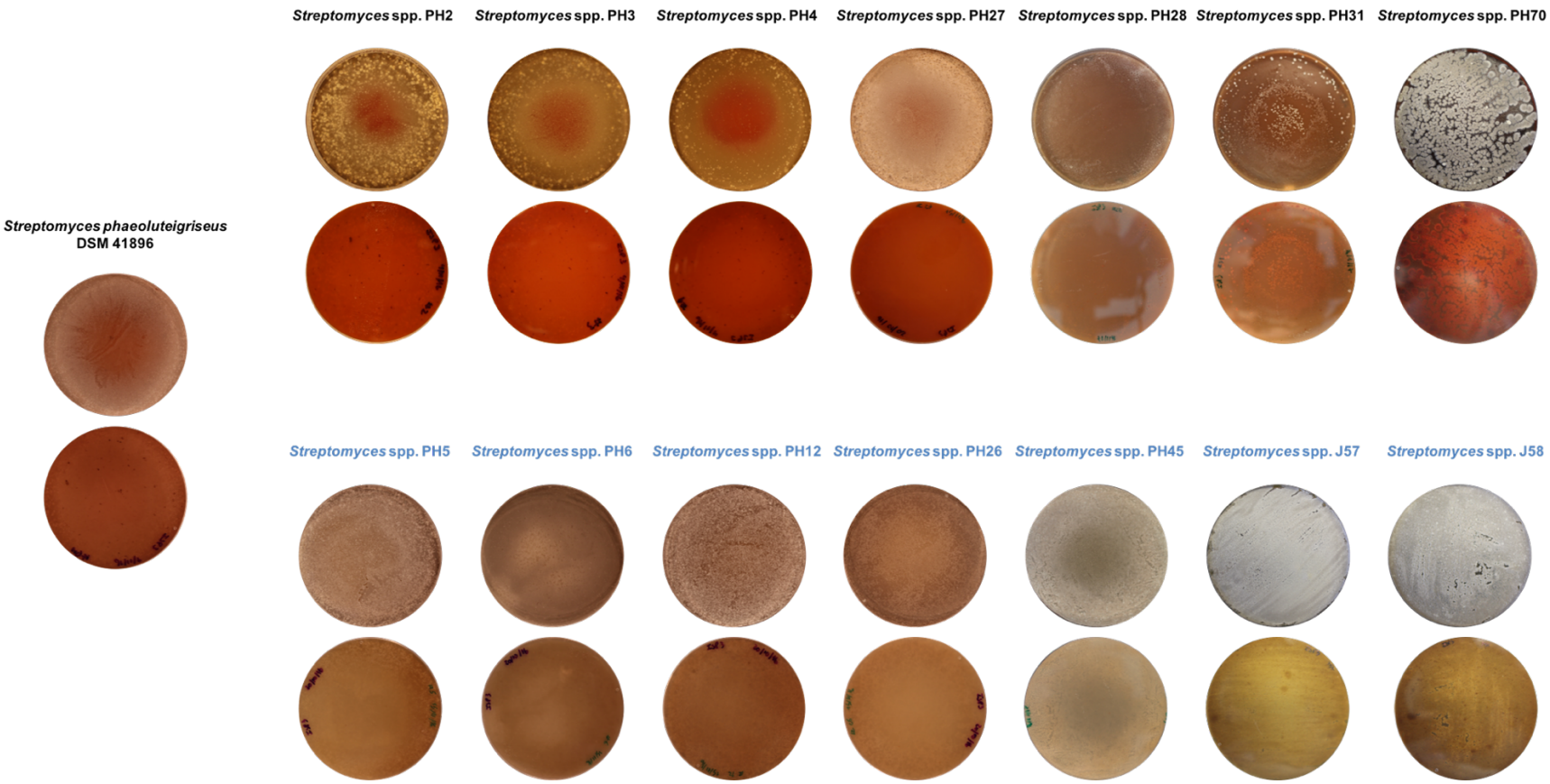


Figure 3.3 Atacama Desert isolates from soil 7 and soil 8, and the closest neighbour *Streptomyces phaeoluteigriseus* grown in ISP3 and showing the aerial spore mass (top) and reverse colour (bottom).

Table 3.2 Reverse colour of each isolate and type strain grown in different ISP media.

Mean RGB colours were obtained for each isolate and soil using Get Average Colour of Image (https://matkl.github.io/average-color/). Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**).

Isolates	ISP2	ISP3	ISP4	ISP5	ISP6	ISP7	Mean	Mean soil
S. phaeoluteigriseus								
PH2								
PH3								
PH4								
PH27								
PH28								
PH31								
PH70								
PH5								
PH6								
PH12								
PH26								
PH45								
J57								
J58								

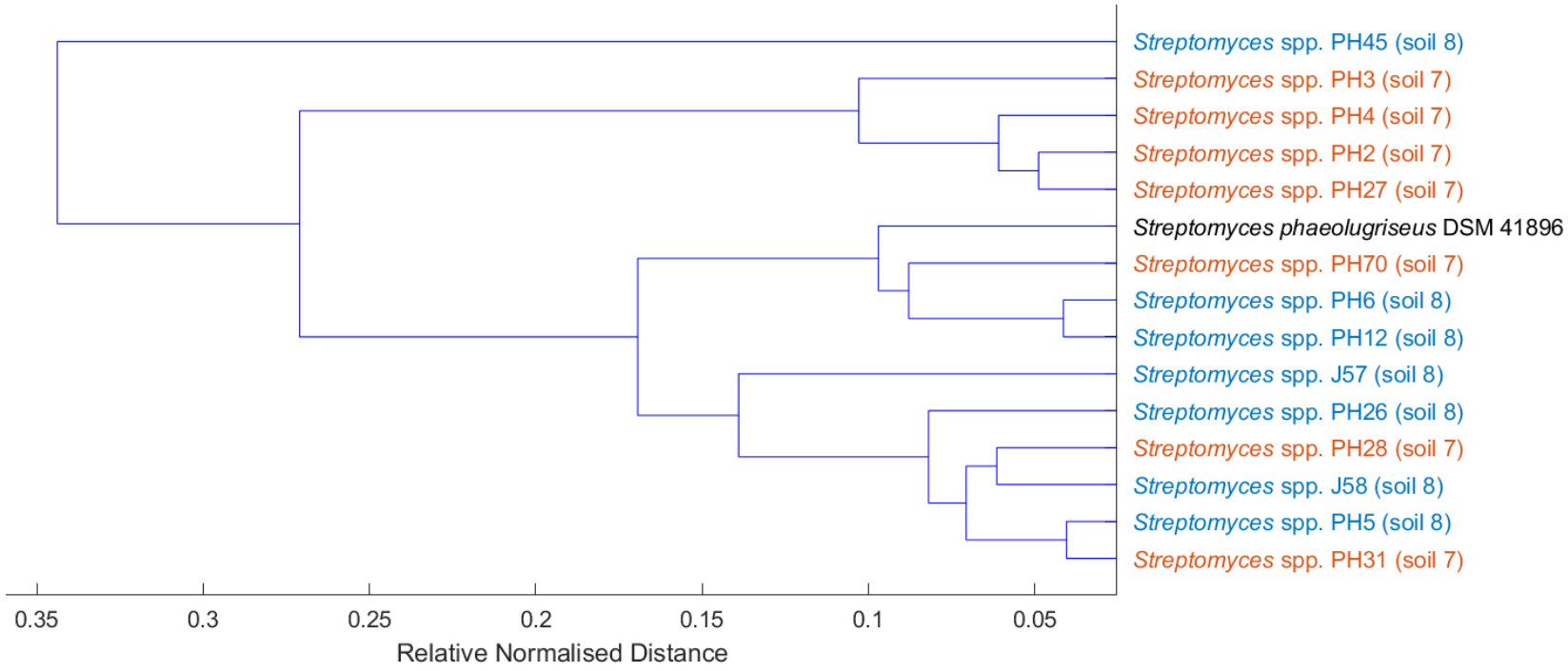


Figure 3.4 Dendrogram showing similarities between *Streptomyces* from both soils using reverse colour RGB values in ISP3 based in *Antony-Babu, Stach & Goodfellow (2010)*.

Similarities between isolates from the two soils (soil 7 and soil 8) emerged with three main heterogeneous groups.

3.4 Hyperspectral Imaging as a promising tool to resolve differences among related isolates

Hyperspectral images are three-dimensional combining digital imaging and spectroscopy and generating a spectrum for each pixel in the image containing multiple layers at different wavelengths of light. Currently it is used in distinct fields, e.g. analytical chemistry (food and pharmaceutical industry) and microbiology (e.g. detect viable cells in food surfaces and Gram-positive and Gram-negative bacteria distinction) among others (Edelman, Van Leeuwen & Aalders, 2012; Tiwari, Arora & Singh, 2011; Lim & Murukeshan, 2016; Liu, Pu & Sun, 2017). One methodology for taxonomic characterisation of *Actinobacteria* is based on the colony features (e.g. aerial and substrate mycelium colours) and formation of melanoid pigment (soluble pigment colour) using different media (Wink, 2012). Thus, the use of HSI could provide high quality differentiation resolution between closely related isolates. The image acquisition, data analysis and dendrogram construction was performed at the Hyperspectral Imaging Centre at the University of Strathclyde with the data analysis and dendrogram generated by Fraser MacFarlane.

To understand and confirm the phylogenetic relationship between the closely related isolates obtained using RGB colours, hyperspectral imaging was performed. This approach can distinguish visually similar samples by generating and comparing spectra for each pixel and thus more accurate than RGB values (Edelman, Van Leeuwen & Aalders, 2012; Tiwari, Arora & Singh, 2011; Lim & Murukeshan, 2016; Liu, Pu & Sun, 2017). HSI analysis was performed by Ryan Greer, Fraser MacFarlane and Dr Paul Murray from the Hyperspectral Imaging Centre at the University of Strathclyde. For this, the isolates were grown in the ISP media (ISP2-ISP7) and imaged using two systems, visible light and near-infrared, with acquisition of two spatial dimensions and one spectral dimension (**Figure 3.5**).

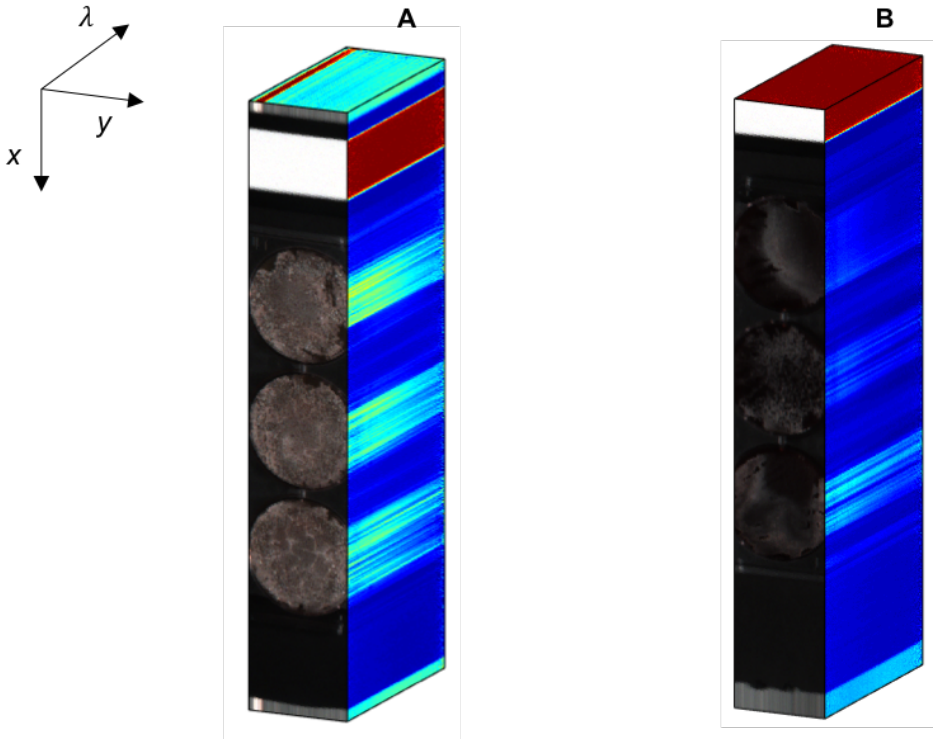


Figure 3.5 Example of two spatial dimension and one spectral dimension projection of two isolates grown in ISP2 using hyperspectral imaging.

A. Isolate PH5 and **B.** Isolate PH28. Colours correspond to the intensity of the image, dark represented in blue and brightest in red. In the hyperspectral image, the *x* and *y* axis represent the spatial dimension while the λ axis represents the wavelength.

Dendrogram were obtained for the reverse and aerial spore mass colours of the isolates grown in ISP3 using visible light. For the reverse colour (**Figure 3.6**) two distinct clades were obtained with a clear separation between soil 7 and soil 8 isolates in line with the comparison of the reverse colours (**Table 3.2**). However, the obtained dendrogram for aerial spore mass colour (**Figure 3.7**) showed isolates clustering independently of the soil location, similar to the previous RGB dendrogram (**Figure 3.4**) and 16S rRNA tree constructed (**Figure 3.2**), demonstrating phenotypical similarities between isolates.

Overall, the isolates have similar aerial spore mass that does not allow a proper distinction between them by HSI. However, it seems to distinguish between isolates more accurately using their reverse colours indicating potential location-specific soluble pigment productio and potentially evolutionary separation.

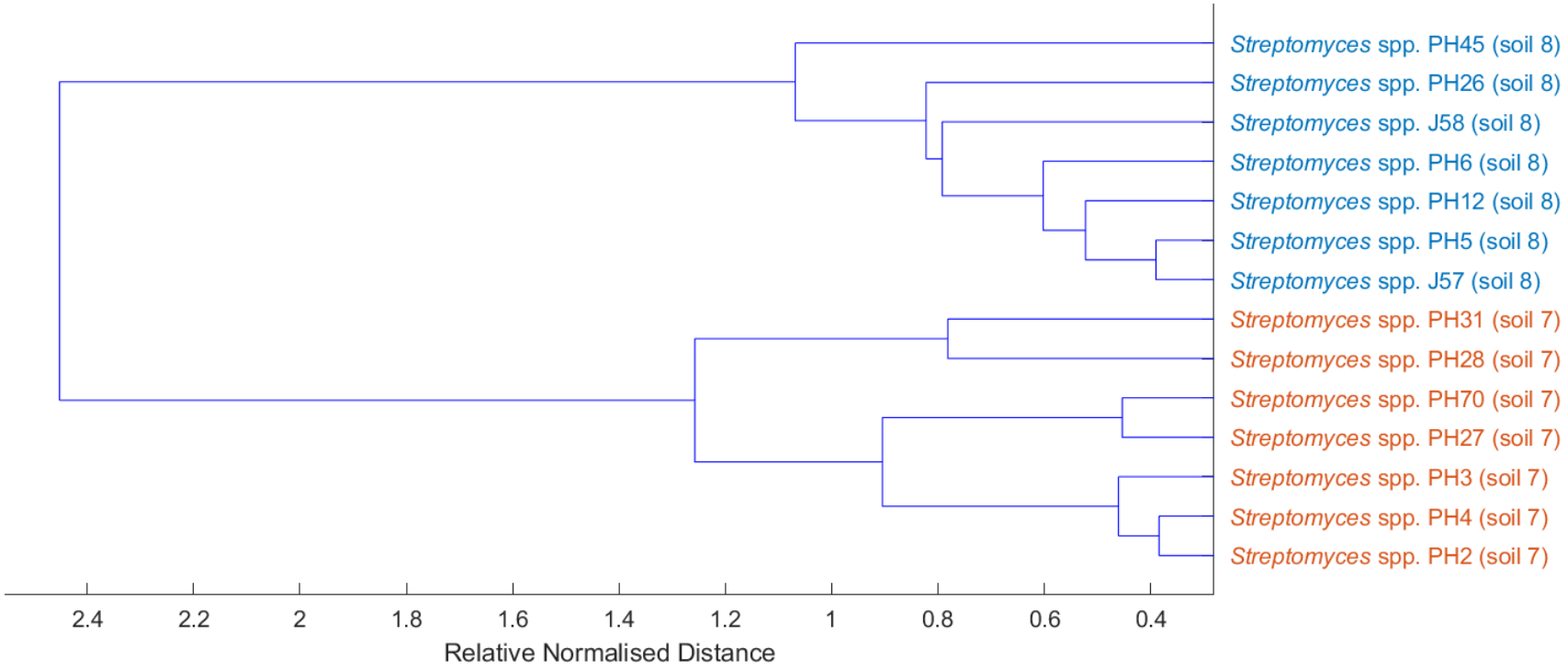


Figure 3.6 Dendrogram showing similarities between *Streptomyces* from both soils using HSI instrument to capture the reverse colour of each isolate in ISP3.

The results indicate the presence of blatant differences between isolates from the two locations creating two distinct groups exclusively comprising of isolates from soil 7 or soil 8. This contrasts with the current methodology results obtained using RGB values for distinction of strains (**Figure 3.4**).

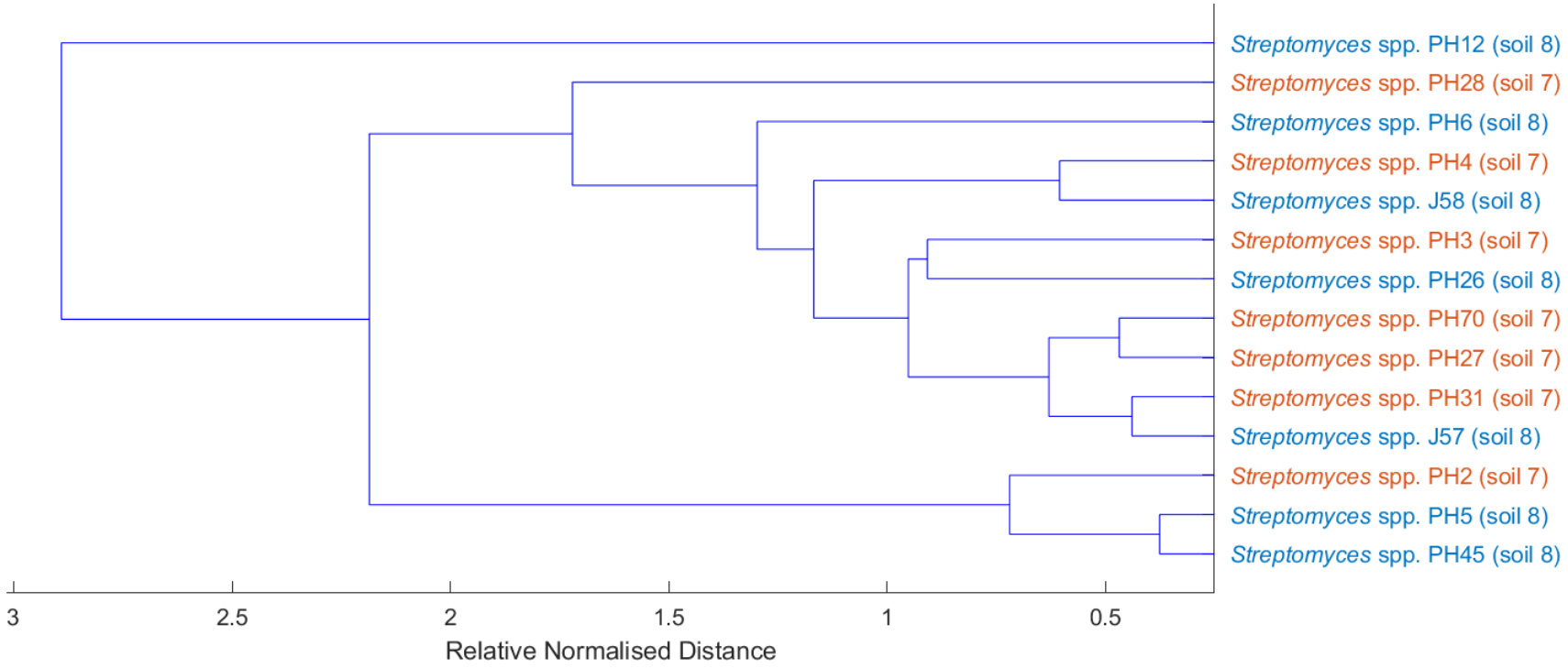


Figure 3.7 Dendrogram showing similarities between *Streptomyces* from both soils using the HSI instrument to capture the aerial spore mass colour of each isolate in ISP3.

The results were similar to the RGB dendogram (**Figure 3.4**) indicate the similarity of the aerial spore mass between isolates from soil 7 or soil 8 and its unreliability for effective distinction of strains when compared to reverse colour.

3.5 Physiological assays show hidden diversity between isolates

To gain insight into the physiological differences between isolates several features were analysed. Growth at different concentrations of sodium chloride was tested and an optimal growth obtained between 1-5% w/v of sodium chloride (NaCl) (**Table 3.3 and Table 3.4**).

Growth was observed at a temperature range between 12-37ºC and isolates PH31, PH5, PH6, PH26, PH45 and J58, as well as the *S. phaeoluteigriseus* DSM 41896, showed some growth at 4ºC (**Table 3.3 and Table 3.4**). Moreover, differences were observed between isolates from the same location and also between locations in regard to pH. Soil 7 isolates were able to grow in a pH range of 6.0-9.0 except PH3 and PH4 that could not grow at pH 9.0 (**Table 3.3 and Table 3.4**). However, soil 8 isolates showed more variability, with isolates PH6, PH12, J57 and J58 having an optimal growth between pH 5.0-8.0, isolate PH26 pH 5-9 and isolates PH5 and PH45 between pH 6.0-8.0, similar to soil 7 isolates (**Table 3.3 and Table 3.4**).

Biochemical tests conducted using API ZYM and 20E strips showed also differences between isolates. For instance, isolate PH28 does not utilise L–arginine, glucose or melibiose but all of the other isolates do, yet it is also possible to observe that many common enzymes are produced among the isolates, e.g. esterase (C4), esterase lipase (C8) and Naphthol-AS-BI-phosphohydrolase. Interestingly, *S. phaeoluteigriseus* DSM 41896 uses mannitol as substrate while the Atacama isolates do not.

These results suggest that there are overall differences among isolates and distinctions can be understood from one another and the type strain.

Table 3.3 Phenotypic characteristics of isolates from soil 7 and its most closely related species, *Streptomyces phaeoluteigriseus* DSM41896.

Characteristic	Type	PH2	PH3	PH4	PH27	PH28	PH31	PH70
Source	Soil *	Soil	Soil	Soil	Soil	Soil	Soil	Soil
Growth on media	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7
Aerial mycelia (ISP3)	Light grey	Light grey	Yellowish grey	Yellowish grey	Light grey	Light grey	Yellowish grey	Light grey
Reverse (ISP3)	Dark red	Dark reddish orange	Deep reddish orange	Dark reddish orange	Dark reddish orange	Strong brown	Brownish orange	Brownish orange
Temperature range (ºC)	12-37 †	12-37	12-37	12-37	12-37	12-37	12-37 †	12-37
Melanin production	+	+	+	+	+	+	+	+
NaCl range (%)	1-5	1-5	1-5	1-5	1-5	1-5	1-5	1-5
pH range	6.0-9.0	6.0-9.0	6.0-8.0	6.0-8.0	6.0-9.0	6.0-9.0	6.0-9.0	6.0-9.0
API 20E and ZYM								
L-arginine	+	+	+	+	+	-	+	+
D-mannitol	+	-	-	-	-	-	-	-
Citrate	+	+	+	+	+	+	+	+
Gelatinase	+	+	+	-	+	+	+	+
Lipase (C14)	+	-	-	+	-	-	+	+
α-Glucosidase	+	-	-	+	-	+	+	-
β-Glucosidase	+	-	-	+	-	-	-	-

* origin unknown; † very small growth at 4ºC

Table 3.4 Phenotypic characteristics of isolates from soil 8 and its most closely related species, *Streptomyces phaeoluteigriseus* DSM41896.

Characteristic	Type	PH5	PH6	PH12	PH26	PH45	J57	J58
Source	Soil *	Soil	Soil	Soil	Soil	Soil	Soil	Soil
Growth on media	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7	ISP2-7
Aerial mycelia (ISP3)	Light grey	Yellowish grey	Greyish yellowish brown	Light grey	Light grey	Yellowish grey	Light grey	Light grey
Reverse (ISP3)	Dark red	Strong yellowish brown	Moderate yellowish brown	Strong brown	Strong yellowish brown	Light grayish olive	Dark greenish yellow	Dark grayish yellow
Temperature range (ºC)	12-37 †	12-37 †	12-37 †	12-37	12-37 †	12-37 †	12-37	12-37 †
Melanin production	+	+	+	+	+	+	+	+
NaCl range (%)	1-5	1-5	1-5	1-5	1-5	1-5	1-5	1-5
pH range	6.0-9.0	6.0-8.0	5.0-8.0	5.0-8.0	5.0-9.0 ‡	6.0-8.0	5.0-8.0 ‡	5.0-8.0 ‡
API 20E and ZYM								
L-arginine	+	+	+	+	+	+	+	+
D-mannitol	+	-	-	-	-	-	-	-
Citrate	+	+	+	+	+	+	+	+
Gelatinase	+	+	+	+	-	+	+	+
Lipase (C14)	+	+	+	+	+	+	+	+
α-Glucosidase	+	-	+	-	-	-	-	+
β-Glucosidase	+	-	+	-	-	-	+	+

***** origin unknown; † very small growth at 4ºC; ‡ very small growth at pH 5.0

3.6 Inhibition and/or neutral interactions emerged between isolates from both locations

An interaction assay was designed to uncover possible inter and intraspecific interactions between isolates. This was achieved by placing plugs of isolates grown in ISP3 on top of seeded plates containing one other isolate.

Overall, isolates from soil 7 showed the ability to inhibit each other in varying degrees, for instance isolate PH31 inhibits all isolates except PH4 but PH2 only exhibits this behaviour against PH70 (**Figure 3.8**). These isolates also showed the ability to inhibit the isolates from the other location, except PH28 did not show any activity against PH5 and J57. Furthermore, PH31 showed greater action against isolates from soil 8 (>2 cm) in comparison with isolates from the other soil (~1.5 cm). Additionally, some auto-inhibition only for isolates from soil 7 (PH2, PH3, PH28, PH31 and PH70) was observed, although less than 1 cm of diameter, likely due to the lack of immunity at that time.

Remarkably, isolate PH4 was the most resistant not affected by any of the isolates except PH5. On the other hand, PH70 was inhibited by all isolates and inhibiting only isolates from the other location. Similarly, isolates from soil 8 only inhibit isolates from soil 7 but not soil 8. Furthermore, isolate PH5 had the ability to inhibit its competitors from the other location (**Figure 3.8**).

These results suggest that despite intraspecific diversity similar metabolites can be produced during growth and immunity genes expressed early in the development in soil 8. However, in soil 7 differences among isolates emerged showing distinct inhibitions and auto-inhibition.

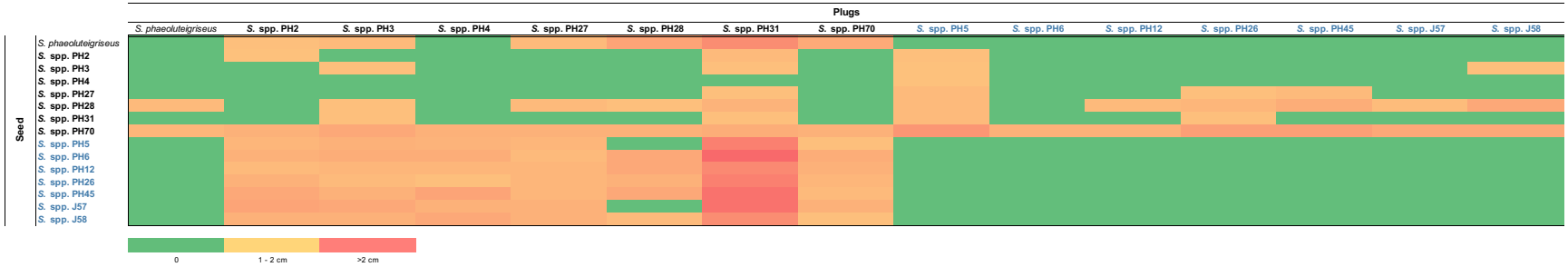


Figure 3.8 Overview of inhibition between isolates from the two locations.

Isolates from soil 7 represented in black and soil 8 in blue. The diameter of the inhibition zones (cm) is represented by a gradient between no (green), intermedium (orange) and high inhibition (red). Columns represent plugs from lawns against isolates seeded on plate (rows).

3.7 Osmotic stress induces bioactive metabolites against indicator strains

Isolates from soil 7 were submitted to osmotic stress to observe how production of secondary metabolites would be affected by either promoting induction or activation of cryptic biosynthetic gene clusters. Since the Atacama Desert contains high salt concentrations sodium chloride (NaCl), potassium chloride (KCl) and magnesium sulphate ($MgSO_4$) were added at six different concentrations to mimic these conditions. Isolates were chosen due to their interspecific inhibition profile, with isolates PH2, PH3, PH4, PH27 and PH31 showing the bigger inhibition zones, intraspecific inhibition zones by variable inhibition of closer related strains and their phylogeny (**Figure 3.8**).

Overall, bioactivity was observed against two of the seven indicator strains and using higher concentrations bigger inhibition zones were seen against *B. subtilis* (~1.5 cm inhibition zone) than *S. aureus* (~2.0 cm inhibition zone) (**Figure 3.9** and **Figure 3.10**). However, PH4 showed bigger inhibition zones against both indicator strains at intermediary concentrations for NaCl (375 mM), KCl and $MgSO_4$ (250 mM). Interestingly, only *B. subtilis* (**Figure 3.9**) and *S. aureus* (**Figure 3.10**) were inhibited suggesting a potential active metabolite against Gram-positive bacteria.

Isolates PH2 and PH27 showed a high induction against *B. subtilis* with KCl, increasing the inhibition diameter about 1 cm (**Figure 3.9, panel B**). This trend was similarly observed against *S. aureus* however with a slightly lower effect (**Figure 3.10, panel B**). Magnesium sulphate (**Figure 3.10, panel C**) showed higher induction than NaCl or KCl contrasting with the results for *B. subtilis* (**Figure 3.9, panel C**).

Production of a possible salt-induced metabolite in *Streptomyces phaeoluteigriseus* DSM 41896 using NaCl, KCl and $MgSO_4$ (>200 mM) active against *B. subtilis* (**Figure 3.9**) and *S. aureus* (**Figure 3.10**; except $MgSO_4$) was detected. Furthermore, using KCl developmental changes were observed with increasing concentrations from normal development to lack of aerial mycelia and sporulation (**Figure 3.11**).

This data suggests the ability to induce the production of specialised metabolites among the selected isolates using high concentrations of salts.

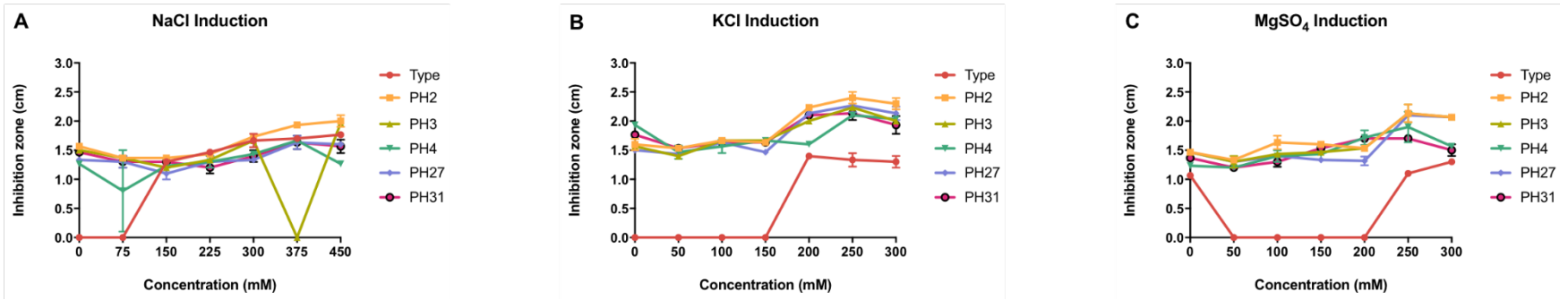


Figure 3.9 Induction obtained against *B. subtilis* with isolates from soil 7.

Increasing concentrations of **A.** sodium chloride (NaCl), **B.** potassium chloride (KCl) and **C.** magnesium sulphate ($MgSO_4$) were tested showing different results.

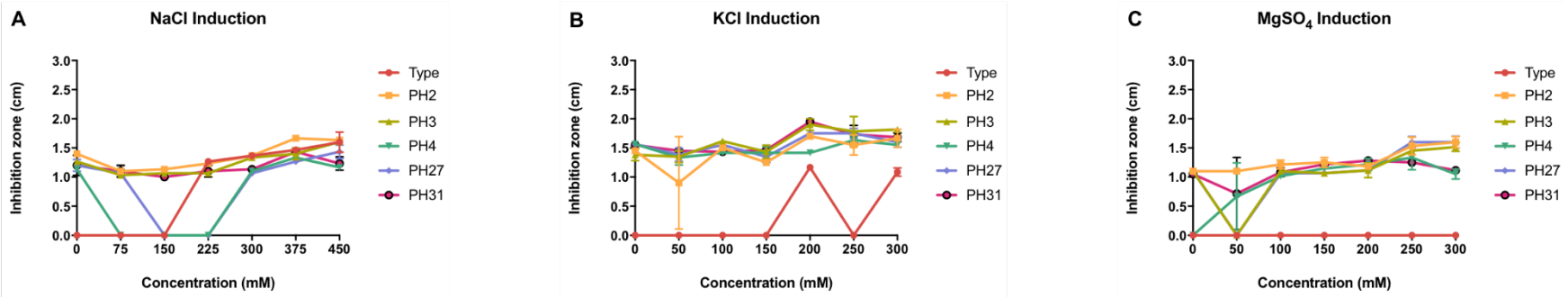


Figure 3.10 Induction obtained against *S. aureus* with isolates from soil 7.

Increasing concentrations of **A.** sodium chloride (NaCl), **B.** potassium chloride (KCl) and **C.** magnesium sulphate ($MgSO_4$) were tested showing wide different results.

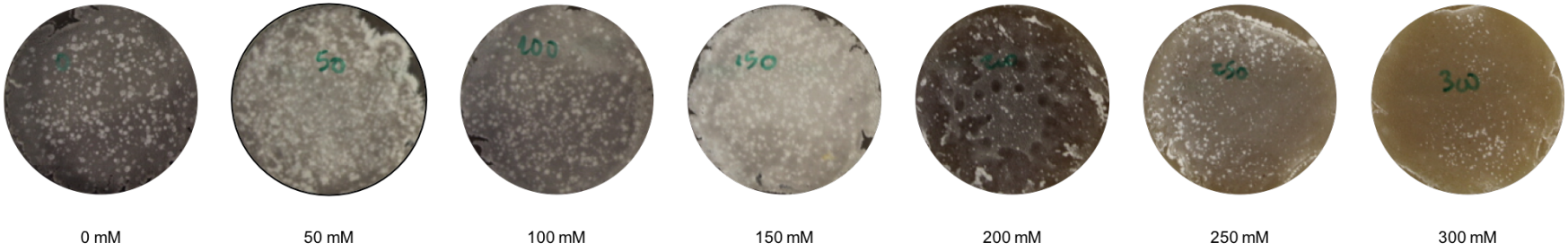


Figure 3.11 Developmental changes observed at different concentrations of potassium chloride in *S. phaeoluteigriseus* DSM 41896.
Effects of the increase in the concentration of KCl, changing from normal development of aerial mycelia (0 mM) to the disruption of erecting aerial mycelia and posterior sporulation (300 mM).

3.8 Discussion

To characterise new environmental isolates and to resolve their evolutionary relationship several features need to be studied. Fourteen abundant and similar *Streptomyces* isolates were recovered from two locations and found to constitute ~16% and ~33% of the total identified *Actinobacteria* species from soil 7 and soil 8, respectively (Dias Mirandela and Schniete, unpublished work). The Atacama Desert was being surveyed at different sites and a wide diversity of bacteria found, although predominantly *Streptomyces* spp. dominate the environment, e.g. 86% of isolated bacteria from the Salar de Huasco belonged to this genus (Cortés-Albayay *et al.*, 2019c).

The first approach in this study was to sequence 16S rRNA and identify the most closely related bacteria in the databases. The collected 14 strains were similar to the same type strain, S. *phaeoluteigriseus* DSM 41896, and as their most closely related neighbour (**Table 3.1**, >99.20% similarity). The construction of phylogenetic trees further demonstrated the close relationship between strains from these two locations approximately 30 km apart, constituting a novel observation in this environment (**Figure 3.2**). Interestingly, closely related strains were already isolated from Salar de Atacama but only from the same soil sample and in a few number, e.g. highly similar strains of *S. asenjonii* (Goodfellow *et al.*, 2017; Cortés-Albayay *et al.*, 2019c). However, several studies showed that 16S rRNA does not provide a sufficient resolution between *Streptomyces* species and further genotypic and phenotypic studies were needed to differentiate strains (Antony-Babu, Stach & Goodfellow, 2008; Rong *et al.*, 2010; Seipke, 2015). Recently, Antony-Babu *et al.* (2017) showed that bacterial strains with identical 16S rRNA sequences (>99%) were diverse enough, metabolic and genomically, to be considered as different species.

Accordingly, growth on different media and relationship analysis by hyperspectral imaging showed differences in the pigment production observed between isolates from the two soils, despite similar aerial spore mass, supported this possibility. Soil 7 isolates showed a dark reddish to brownish orange colour and soil 8 light grey to light brown colour. The HSI colour dendrogram further expands previous work that was

able to dereplicate quickly large numbers of species by correlating colour groups with geographical distribution (Antony-Babu, Stach & Goodfellow, 2010). Additionally, it was shown that isolates in this study clustered in groups dependent on their isolation location however some common colour features were observed, e.g. isolates PH28, PH31 and PH70 and soil 8 isolates PH5, PH6, PH12, PH26.

Physiological tests suggest and confirm the morphological distinctions, in contrast to the 16S rRNA results, providing a general description of the metabolism of the isolates. Several features were identified that differentiate between isolates in each or both locations and the most closely related strain. The optimal pH range was found to be dependent of the location and the tolerance slightly lower in comparison with other Atacama species (average pH 4.0-11.0) although growth and sporulation were higher in ISP3 medium it was the same and optimal temperature range similar to the *S. desertii*, *S. bullii* and *S. aridus* (average 10-40ºC) (Idris *et al.*, 2017; Santhanam *et al.*, 2012b, 2013). Furthermore, distinctive results were observed for utilisation of D-mannitol (only type strain), production of lipase (C14) (specific to type and soil 8 isolates) and β-glucosidase (type, soil 7 PH4 and soil 8 PH6, J57 and J58). Nonetheless, similarities were also found between isolates such as utilisation of citrate, acid phosphatase, esterase (C4), esterase lipase (C8) and Naphthol-AS-BI-phosphohydrolase. Interestingly, some features are shared with the known Atacama type strains such as *S. leeuwenhoekii* (all four), *S. asenjonii* (acid phosphatase and esterase lipase (C8)) and *S. aridus* (acid phosphatase and Naphthol-AS-BI-phosphohydrolase) (Idris *et al.*, 2017; Busarakam *et al.*, 2014; Goodfellow *et al.*, 2017). Some other features are partially shared such as the production of alkaline phosphatases (soil 8 except PH5 and only by PH2, PH4, PH28 and PH31 in soil 7) and valine arylamidase (all except PH3 and PH27) and others not at all present on the collected isolates, e.g. N-acetyl-β-glucosaminidase or cysteine arylamidase.

While isolates from soil 8 do not inhibit themselves they were able to inhibit isolates from soil 7 suggesting a possible difference in metabolites production than the other isolates (**Figure 3.8**). Furthermore, variability was also displayed within soil 7 organisms creating two distinct groups showing different inhibition patterns (**Figure 3.8**). However, this experiment possesses limitations that can bias the results, since we compared the isolates in different stages of their growth cycle. Metabolites produced at a later stage of the growth cycle were tested against

spores/initial vegetative mycelia at the beginning of the life cycle, where immunity genes could yet not be actively expressed. Similarly, Vetsigian, Jajoo & Kishony isolated *Streptomyces* strains from the same soil grain and grew them independently in media to produce conditioned media that was used to grow each of the other strains. They observed that closely related strains by 16S rRNA would somehow inhibit each other less but it was not a strong correlation (Vetsigian, Jajoo & Kishony, 2011). These strains showed different interaction profiles over a short evolutionary time in the production rather than resistance (Vetsigian, Jajoo & Kishony, 2011). Furthermore, Abrudan *et al.*, actively co-cultured *Streptomyces* strains to study the interactions between *Streptomyces* strains from the same location (Abrudan *et al.*, 2015). They concluded that the observed interactions did not depend on phylogenetic relationship but on the identity of the neighbours and available resources (Abrudan *et al.*, 2015). In this study, it was demonstrated that phylogenetically close strains from the same location can either inhibit each other due to distinct production profiles (soil 7) or remain neutral (soil 8) but are able to compete with more phylogenetic distant strains (**Figure 3.8**).

In this research work, it was shown that using a range of different salts present in the Atacama Desert had an inducing effect on the activity against the indicator strains (**Figure 3.9** and **Figure 3.10**). In *Streptomyces* morphological development and production of antibiotics are interconnected with common pathways of gene regulation. Osmotic stress induces stress response genes such as *prsH* and *osaB*, that are required for normal aerial hyphae formation and regulation of antibiotic production in *S. coelicolor* (Viollier *et al.*, 2003; Fuchino *et al.*, 2017; Martínez *et al.*, 2009). Furthermore, *osaB* was disrupted and the resulting *S. coelicolor* mutant strain failed to erect aerial hyphae in media supplemented with salt but overproduced actinorhodin and undecylprodigiosin (Bishop *et al.*, 2004). A similar result was observed with *S. phaeoluteigriseus* that produced a metabolite when exposed to salts but failed to erect aerial mycelia (**Figure 3.11**). To avoid the harmful effect of high salinity in the environment production and accumulation of compatible solutes such as ectoine and 5-hydroxyectoine allow the re-establishing of normal cellular function (Bursy *et al.*, 2008).

In conclusion, the analysis of colour groups, biochemical tests and production of specialised metabolites suggests a hidden diversity present in both locations and

between isolates not observed by 16S rRNA only. However, some of these observations and diversity could be comprehended by whole-genome analysis.

Chapter 4. Whole-genome analysis of the Atacama Desert isolates from soil 7 and soil 8

4.1 Introduction

The first attempt to sequence DNA was achieved when Gilbert and Maxam sequenced 24 base pairs of the lac operator in 1973 (Pareek, Smoczynski & Tretyn, 2011; Gilbert & Maxam, 1973). Just a few years later, Sanger sequencing was developed by Frederick Sanger who used a short oligonucleotide to bind to a single strand DNA and the addition of fluorescent ddNTPs by DNA polymerase in the newly synthesised strand stopped the elongation step and produced molecules with different length (Pareek, Smoczynski & Tretyn, 2011; Sanger & Coulson, 1975). Afterwards, these were separated and identified by polyacrylamide gel or capillary tube gel systems (Pareek, Smoczynski & Tretyn, 2011; Sanger & Coulson, 1975). Almost a decade after, Kary Mullis invented the polymerase chain reaction (PCR) where using short oligonucleotide primers it was possible to rapidly obtain millions of copies of target DNA (Loman *et al.*, 2012; Mullis *et al.*, 1986). These breakthroughs were vital in the development of high-throughput next-generation sequencers able to rapidly and cost-effectively sequence whole genomes (Loman *et al.*, 2012).

The data then available to researchers led to the creation of informatic tools and platforms to compare and attribute biological relevance. The Galaxy web platform (Afgan *et al.*, 2018) holds many different tools to analyse metagenomic, transcriptomic, proteomic and metabolomic data. Furthermore, the National Center for Biotechnology Information (NCBI) was founded in 1988 collaborating with other services, such as the European Molecular Biology Laboratory (EMBL), playing an important role in the provision of tools and information for evolutionary biology, genetics of diseases, gene expression and drug discovery. The increase in data and its availability expanded one's knowledge in the biological world uncovering genetic variants related to disease (Yano *et al.*, 2016), fast identification of virulence factors in bacterial outbreaks (Gilchrist *et al.*, 2015) or even in drug discovery (Brown & Wright, 2016; Tracanna *et al.*, 2017). For the latter, thailandamide A from

Burkholderia thailandensis (Nguyen *et al.*, 2008) and aspoquinolones A–D from *Aspergillus nidulans* (Scherlach & Hertweck, 2006) were found by genome mining and the use of such tools and programs for prediction of specialised metabolites. Furthermore, a better grasp and quick identification of encoded genes and genome structure contributed to an improved understanding of microbial diversity (Mukherjee *et al.*, 2017; Varghese *et al.*, 2015).

Whole-genome sequencing can provide genome structure information such as genome size, number of genes and GC content but also direct species delineation. To achieve this, a multilocus sequence typing (MLST), digital DNA-DNA hybridisation (Meier-Kolthoff *et al.*, 2013) or genome-wide average nucleotide identity (gANI) (Konstantinidis & Tiedje, 2005) can be employed. However, a broader approach combining all the above is more useful due to potential bias present in the individual tools (Figueras *et al.*, 2014). Furthermore, analysis of genome architecture can be performed to infer gene synteny and compare the location of important genes between two species (Tsai *et al.*, 2011; Choulet *et al.*, 2006). This idea was applied to study the organisation of linear chromosomes from *Streptomyces* species by selecting particular genes involved in DNA replication, e.g. DnaA unwinds the DNA at the internal origin (*oriC*), and Tap/Tpg cap initiate end-patching of the telomeres (Tsai *et al.*, 2011; Yang, Tseng & Chen, 2015). The telomeres typically consist of around 160 nucleotides at the end of the chromosome forming stem-loops structures in 'rabbit ears' or 'clover-leafs. The telomeres of *Streptomyces coelicolor* and *S. avermitilis* were considered to be 'archetypal', comprising of a conserved terminal palindrome (Palindrome I) and five more internal palindromes (Tidjani, Bontemps & Leblond, 2020). Interestingly, *oriC* is localised in the central region of the chromosome and this region is also highly conserved, often encoding essential genes (Bentley & Parkhill, 2004; Medema *et al.*, 2010). However, the terminal regions possess more variability and thought to be more niche or strain-specific (Bentley & Parkhill, 2004; Medema *et al.*, 2010).

Insertion of prophage sequences in the genomes and formation of genomic islands could play an important role in this diversity, increasing the fitness of the microorganisms (Ventura *et al.*, 2007). The study of pangenomes can help to understand inter and intraspecific variability by finding genes shared among species and also essential genes that are found in all of them (>95%) comprising the core

genome (McInerney, McNally & O'Connell, 2017). Furthermore, an accessory genome can be defined as the non-essential genes that are not present in all of them thus conferring variability (McInerney, McNally & O'Connell, 2017). Several studies have used this approach to understand evolutionary and genome plasticity of *Streptomyces* species (Kim *et al.*, 2015; Zhou *et al.*, 2012).

Previous analysis showed a high similarity between the collected isolates by 16S rRNA. However, divergent characteristics were observed using HSI and physiological assays. Thus, the fourteen *Streptomyces* isolates from the Atacama Desert were further studied using whole-genome analysis. To understand the difference between isolates (1) their phylogenetic relationship by MLSA and DNA-DNA hybridization, (2) pan- and core genome and (3) their potential for secondary metabolite production was studied employing *in silico* analysis.

4.2 Whole-genome analysis and mapping uncovers remarkable inter- and intra-specific diversity

Whole-genome sequencing was performed using the Illumina platform to gain an insight into the isolate genomes and potential unique gene features. A genome coverage of 30x was obtained for each isolate and the draft genomes were assembled by MicrobesNG using SPAdes (v3.7) (Bankevich *et al.*, 2012) and subsequently annotated with RAST (Overbeek et al., 2013) (**Table 4.1**). In general *Actinobacteria*, namely *Streptomyces*, have large genomes between 6 to 12 Mbp (Kämpfer et al., 2014) while the collected 14 isolates showed an average genome size of 8.5 Mbp (**Table 4.1**). Differences among locations were observed with isolates from soil 7 showing a slightly smaller genome (8.2 ± 0.2 Mbp) than soil 8 (9.0 ± 0.3 Mbp) (**Table 4.2**). Additionally, the average number of contigs assembled for soil 7 were 1,210 contigs (average annotated genes 7,206 ± 248), slightly less than soil 8 with an average of 1,224 contigs (annotated genes 7,970 ± 372). Furthermore, the quality of the genome assembly was determined by assessing the N50. Approximately ~12 Kbp and ~15 Kbp were obtained for isolates from soil 7 and 8, respectively, meaning that 50% of the length of the assembly was contained in contigs with at least 12 or 15 Kbp.

A linear relationship was obtained when comparing the number of genes and genome sizes of each isolate suggesting a continuity of this relationship as more genomes were added (**Table 4.2**). Generally, this trend is observed for bacteria with coding density of one gene per each kilobase of DNA (Bentley & Parkhill, 2004). Konstantinidis & Tiedje found that bacteria with larger genomes had an increase in abundance of genes involved in regulation and secondary metabolism (Konstantinidis & Tiedje, 2004). This observation could explain the broad metabolic diversity of these genomes and their successful adaption to distinct environments and resources depletion.

Overall, isolates from the two soils have similar GC content but distinguishable genome size.

Table 4.1 Sequencing statistics obtained after SPAdes assembly and RAST annotation. N50: represents the statistic in which 50% of the entire assembly is contained in contigs with equal or longer length (Jakubowska *et al.*, 2015).

Sequencing statistics					
PH2	# of contigs	1,479	PH5	# of contigs	955
	Size (*Mbp*)	7.9		Size (*Mbp*)	9.1
	N50 (*Kbp*)	9.5		N50 (*Kbp*)	19.9
	Protein coding genes	6,842		Protein coding genes	8,175
PH3	# of contigs	1,022	PH6	# of contigs	1,122
	Size (*Mbp*)	8.0		Size (*Mbp*)	9.3
	N50 (*Kbp*)	16.1		N50 (*Kbp*)	17.1
	Protein coding genes	7,125		Protein coding genes	8,399
PH4	# of contigs	823	PH12	# of contigs	1,535
	Size (*Mbp*)	8.0		Size (*Mbp*)	8.5
	N50 (*Kbp*)	19.4		N50 (*Kbp*)	9.8
	Protein coding genes	7,134		Protein coding genes	7,342
PH27	# of contigs	1,331	PH26	# of contigs	1,036
	Size (*Mbp*)	8.2		Size (*Mbp*)	9.1
	N50 (*Kbp*)	11.0		N50 (*Kbp*)	17.9
	Protein coding genes	7,318		Protein coding genes	8,176
PH28	# of contigs	1,512	PH45	# of contigs	1,478
	Size (*Mbp*)	8.6		Size (*Mbp*)	8.7
	N50 (*Kbp*)	10.7		N50 (*Kbp*)	11.1
	Protein coding genes	7,534		Protein coding genes	7,593
PH31	# of contigs	870	J57	# of contigs	1,415
	Size (*Mbp*)	8.4		Size (*Mbp*)	9.1
	N50 (*Kbp*)	19.4		N50 (*Kbp*)	10.0
	Protein coding genes	7,469		Protein coding genes	7,971
PH70	# of contigs	1,433	J58	# of contigs	1,024
	Size (*Mbp*)	8.1		Size (*Mbp*)	9.1
	N50 (*Kbp*)	10.5		N50 (*Kbp*)	18.4
	Protein coding genes	7,017		Protein coding genes	8,132

Table 4.2 Average of the sequencing statistics obtained after SPAdes assembly and RAST annotation for both soils.

	Average sequencing statistics		Standard deviation
Soil 7	Mean genome size (*Mbp*)	8.2	0.2
	Mean Protein coding genes	7,206	248
	Mean GC content (%)	71.6	0.1
Soil 8	Mean genome size (*Mbp*)	9.0	0.3
	Mean Protein coding genes	7,970	372
	Mean GC content (%)	71.3	0.1

To confirm the previous results, the phylogenetic relationship between strains by multilocus sequence typing (MLST) analysis using four housekeeping genes *atpD*, *gyrB*, *recA* and *trpB*, extracted from the whole-genome were also further investigated. The phylogenetic tree was constructed using IQ-TREE, concatenated gene sequences and the best substitution model determined automatically (WAG: *atpD*, *gyrB* and *trpB*; JTTDCMut+I: *recA*) (**section 2.4.1**). A clear distinction was observed between both locations exhibiting soil-specific clustering and intra-specific relationships (**Figure 4.1**). Isolates from soil 7 affiliated closely to each other and branched in a distinct clade than isolates from soil 8 and the closely-related type strain *S. phaeoluteigriseus*, implying fundamental phylogenetic differences.

The analysis was expanded by including more housekeeping genes (additional 80 genes) automatically selected using the Automatic Multi-Locus Species Tree (autoMLST) web server (Alanjary, Steinke & Ziemert, 2019). Similar results were obtained with isolates from soil 7 clustering in a distinct clade (**Figure 4.2**). Furthermore, the collected isolates in this study were also distantly-related to other *Streptomyces* species including *S. leeuwenhoekii* isolated from a similar geographic location (**Figure 4.2**). Moreover, calculation of the genome-wide average nucleotide identity (ANI) using PYANI generated the same result (**Figure 4.3**). The ANI values between isolates from different locations was below the threshold for species delineation (Richter & Rosselló-Móra, 2009) (< 95%; **Figure 4.3** - **purple colour**). However, when comparing isolates from the same location, either soil 7 or soil 8, the value was above (**Figure 4.3** – **red colour**).

This data suggests that isolates from soil 7 are distinct from soil 8 isolates and potentially comprising a novel *Streptomyces* species.

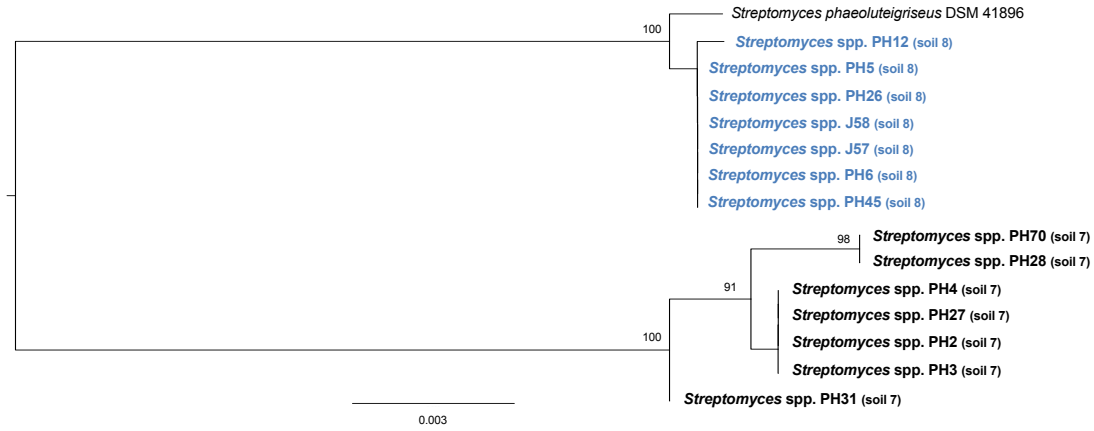


Figure 4.1 Maximum Likelihood tree constructed using concatenated sequences of four housekeeping genes, *atpD*, *gyrB*, *recA* and *trpB* from each isolate and *S. phaeoluteigriseus* DSM 41896.

Tree constructed using IQ-TREE web server (**see section 2.4.1**). Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**).

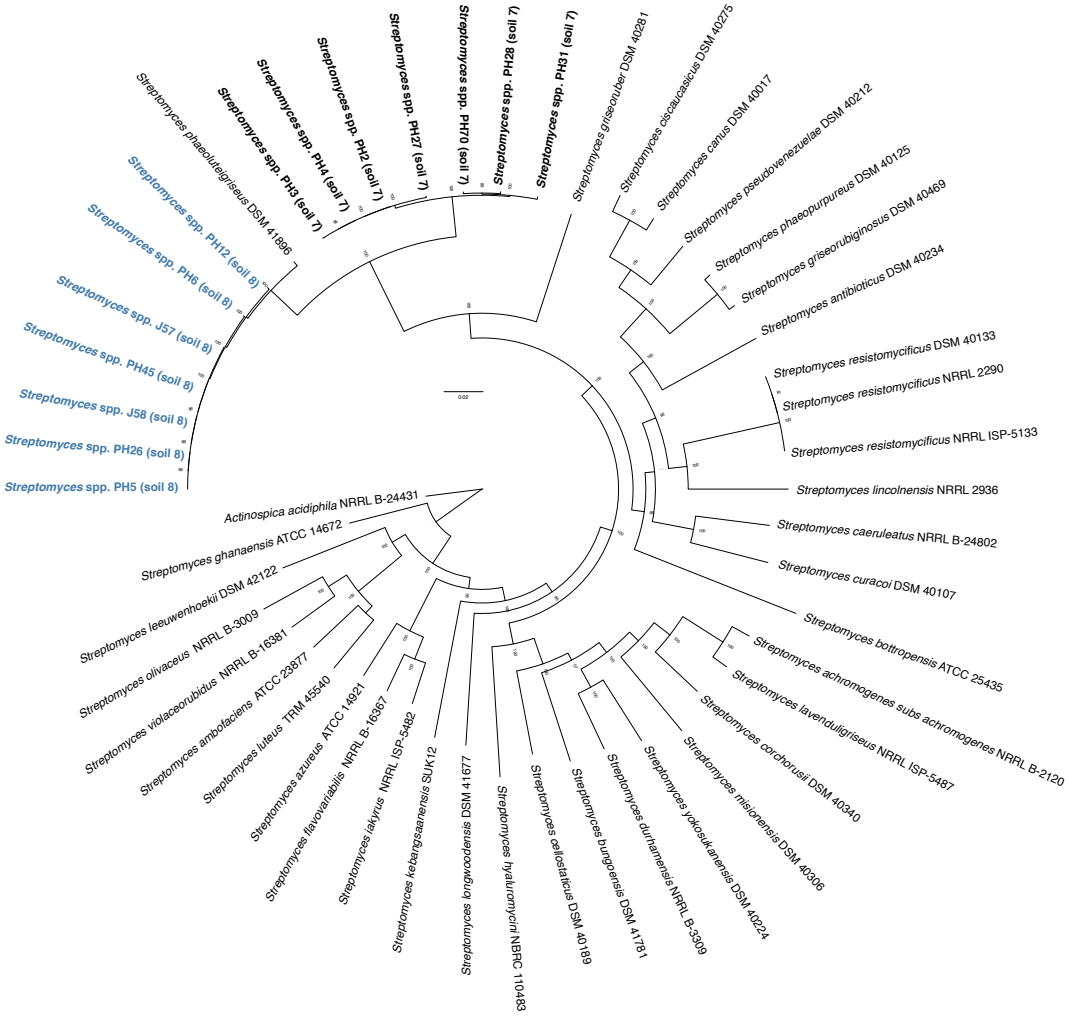


Figure 4.2 MLST tree constructed using autoMLST (default parameters) with concatenated sequences of 84 housekeeping genes against *Streptomyces* reference genomes.

Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**) and only bootstrap values >75% are shown. *Actinospica acidiphila* NRRL B-24431 was automatically selected by the software as outgroup.

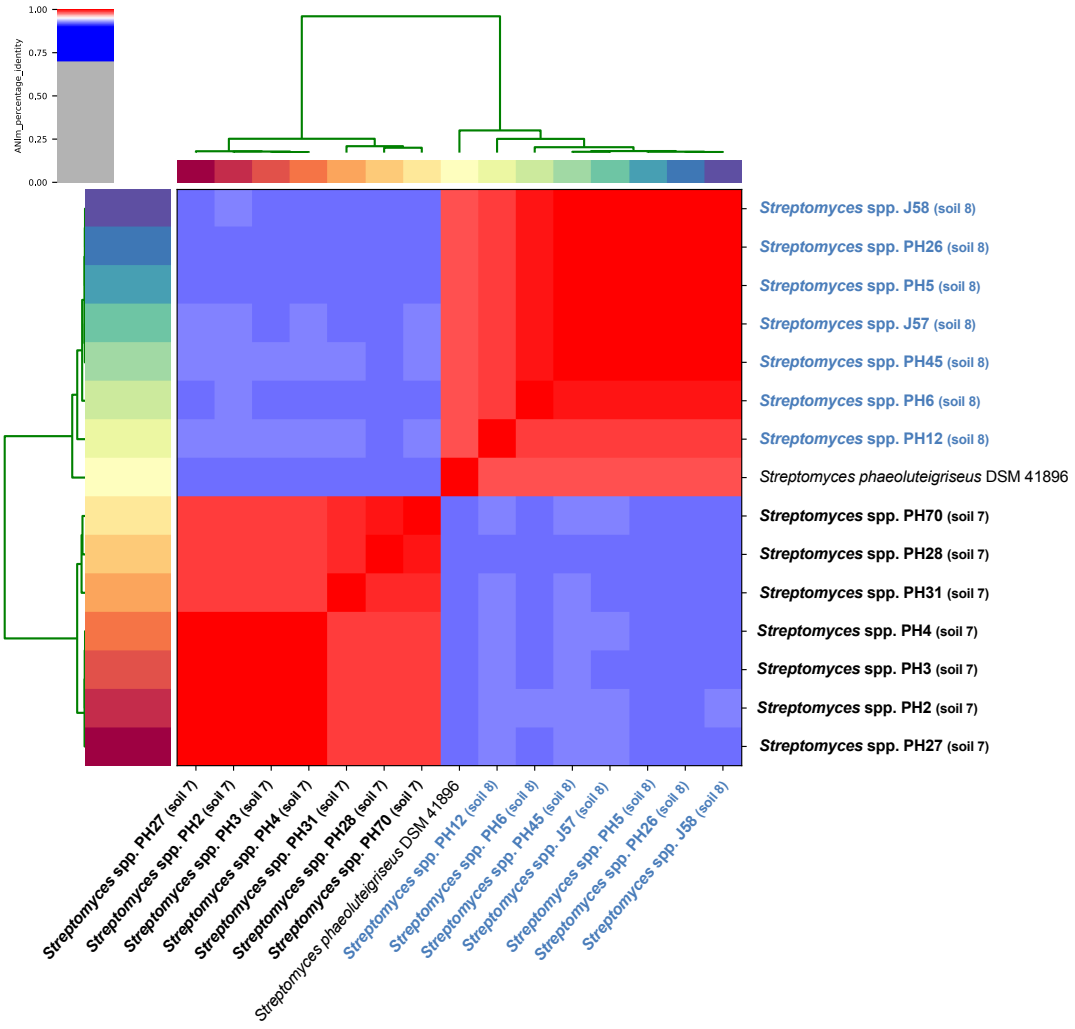


Figure 4.3 Heatmap of ANI percentage identity calculation using PYANI between isolates from both soils.

Cells corresponding to 95% ANI sequence identity (same species) are coloured red and blue cells correspond to isolates that do not belong to the same species. Colour intensity fades as the comparisons approach 95% of sequence identity. The analysis of ANI percentage identity was also used to construct dendrograms representing hierarchical clustering. Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**).

In order to determine if any of the isolates contained giant linear plasmids, PFGE was used to determine the plasmid profile of the isolates and determine if they differed between the two locations. In this way, agarose plugs of the isolates were prepared according to section 2.2.4 (**Chapter 2**) and analysed for the presence or absence of plasmids as well as the determination of their approximate size (**Figure 4.4 and 4.5**). For *S. phaeoluteigriseus* one plasmid of 285 Kbp was detected while for soil 7 isolates varying plasmid sizes were found. These formed two main banding patterns, one from PH2, PH3, PH4, PH27 and PH70 (lower band on the gel) and the other PH28 and PH31 (close to 225 Kbp) (**Figure 4.4**). Interestingly, more diversity in number of plasmids was observed in soil 8 with one, two or three plasmids identified (**Figure 4.5**). PH5 and PH45 have only one plasmid (lower band on the gel) whilst PH26, J57 and J58 also displayed a band of similar size. Furthermore, a second band in PH6 and J58 (225 Kbp) was also present in PH26 and J57. The latter strain also possessed another band between 450-565 Kbp in size.

Collected strains belonging to the same species presented a diversity in their plasmid profile. This suggests that plasmid transfer between closely-related strains from the Atacama Desert is a frequent occurrence and might contribute to their diversification. Similarly, this phenomenon was described in closely-related *Streptomyces* strains where horizontal gene transfer was the main driver of diversification (Tidjani *et al.*, 2019).

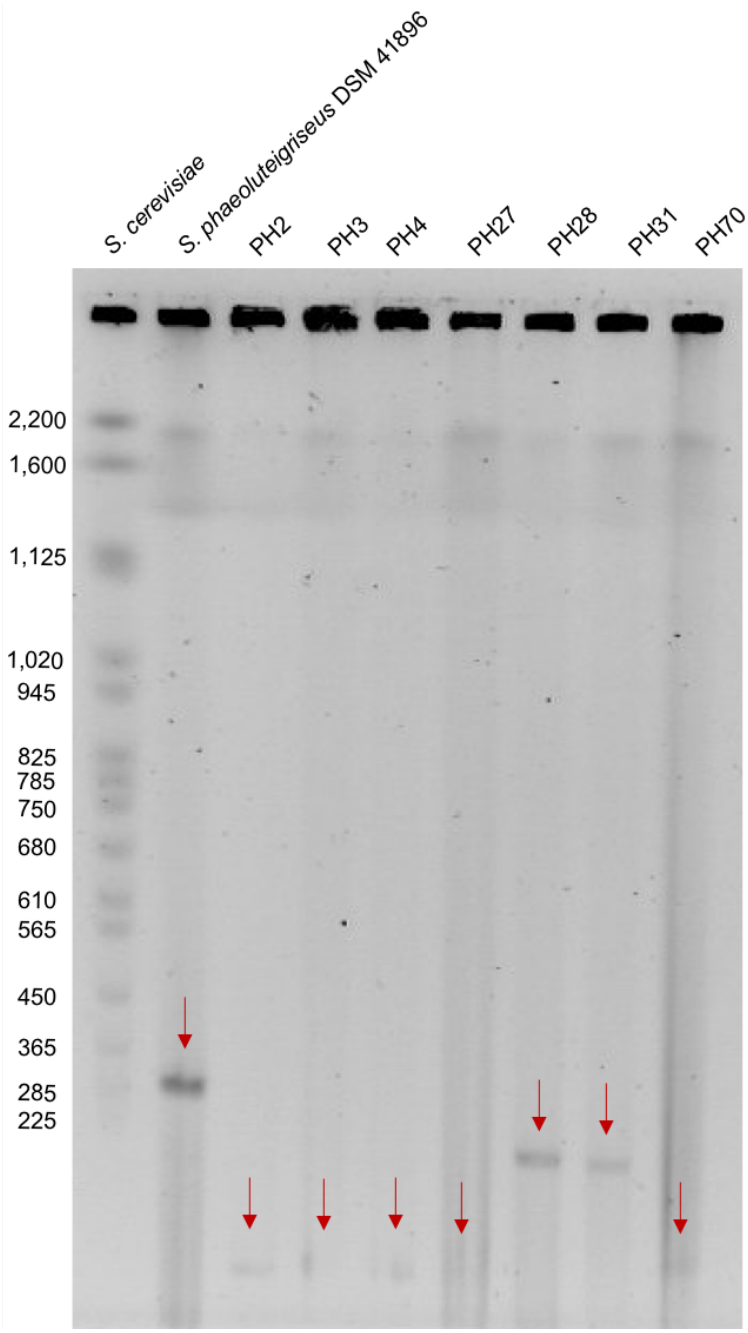


Figure 4.4 Pulse-Field Gel Electrophoresis of total DNA isolated from *S. phaeoluteigriseus* and the Atacama isolates from soil 7 showing extrachromosomal replicons.

Left lane shows *S. cerevisiae* chromosome (size range 0.2-2.2 Mbp) used as DNA size maker (Bio-Rad). Predicted extrachromosomal replicons: ***S. phaeoluteigriseus* DSM 41896:** ~285 Kbp; ***Streptomyces* spp. PH2, PH3, PH4, PH27 and PH70**: <200 Kbp and ***Streptomyces* spp. PH28 and PH31:** ~200 Kbp.

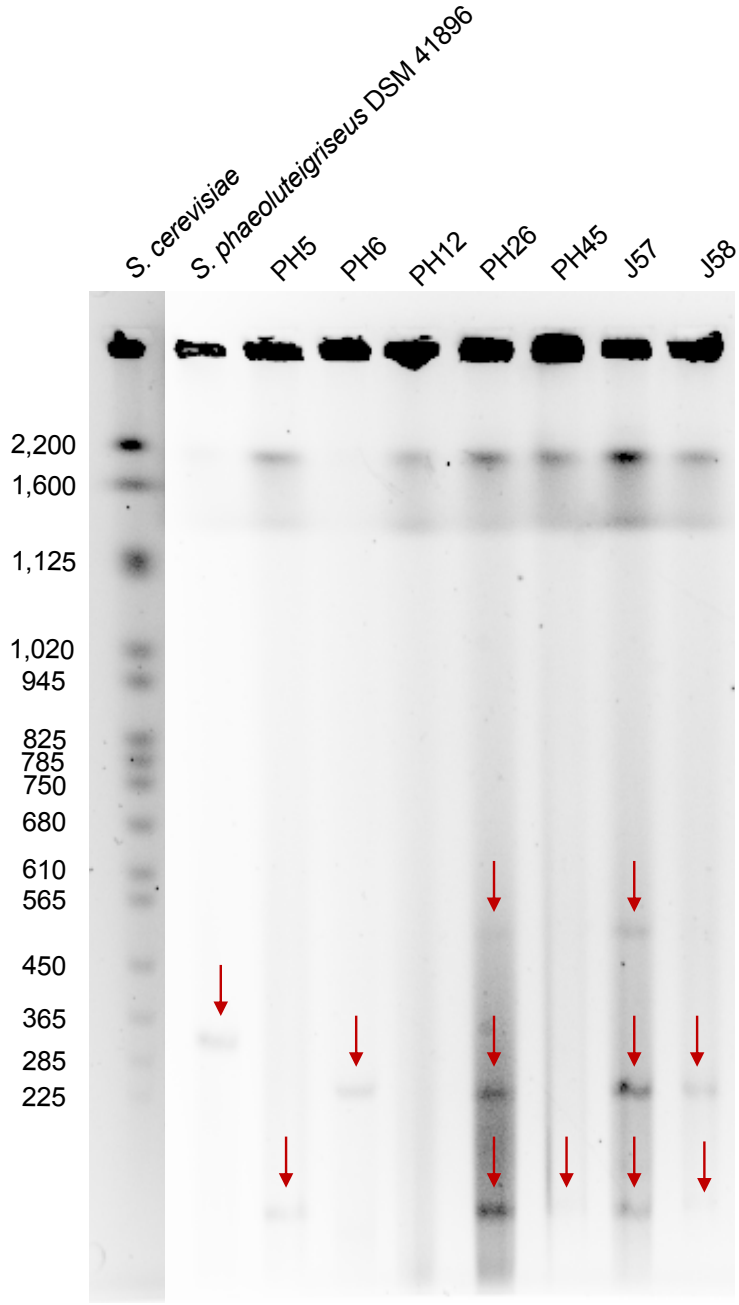


Figure 4.5 Pulse-Field Gel Electrophoresis of total DNA isolated from *S. phaeoluteigriseus* and the Atacama isolates from soil 8 showing extrachromosomal replicons.

Left lane shows *S. cerevisiae* chromosome approximate size (Kbp) used as DNA size maker (Bio-Rad). Predicted extrachromosomal replicons: ***S. phaeoluteigriseus* DSM 41896:** ~285 Kbp; ***Streptomyces* spp. PH5 and PH45**: <225 Kbp; ***Streptomyces* spp. PH6:** ~225 Kbp; ***Streptomyces* spp. PH26 and J57:** <225, ~225 and ~500 Kbp and ***Streptomyces* spp. J58:** <225 and ~225 Kbp.

4.3 Pangenome analysis reveals intra-specific gene diversity and large accessory genome in isolates from both soils

To understand which genes were variable (accessory genome) or conserved (core genome) among the isolates from both locations a comparative pangenome analysis was performed using whole-genome sequences (**section 2.4.2**). The pangenome of the 14 isolates and the strain *S. phaeoluteigriseus* was evaluated using PIRATE, providing an overview of gene content and allelic diversity (Bayliss *et al.*, 2019). This analysis allowed us to measure gene divergence and the categorisation of paralogs in terms of mutation, duplication and/or fission *loci* or insertion and deletion (Bayliss *et al.*, 2019). It was possible to identify 12,975 gene families comprising the pangenome and of those, 3,759 (~31%) were present in all of the isolates and classified as core (>95%), and the remaining 8,307 as accessory (present <95%) (**Figure 4.6**). Furthermore, 909 gene families (7% of pangenome) contained more than one allele, in which 458 gene families were present in >95% of the isolates and 451 gene families in <95% (**Figure 4.7, panel A - red**). Additionally, for fission and fusion (492 gene families; 3.8%) and duplication and loss events (227 gene families; 1.7%) the same pattern was observed. Higher number of these were present in the core genome, 269 fission/fusion and 134 duplication/loss events, than in the accessory genome, 223 and 93 gene families, respectively (**Figure 4.7, panel B**). An increase in the number of clusters was observed at thresholds >90% due to the differences among the two species and the intra-species allelic diversity between them (**Figure 4.7, panel C**). This could be observed in the decrease of core genes estimated above a threshold of 90% (percentage identity limit to assign a cluster) and a steep increase of the accessory genome (**Figure 4.7, panel D**).

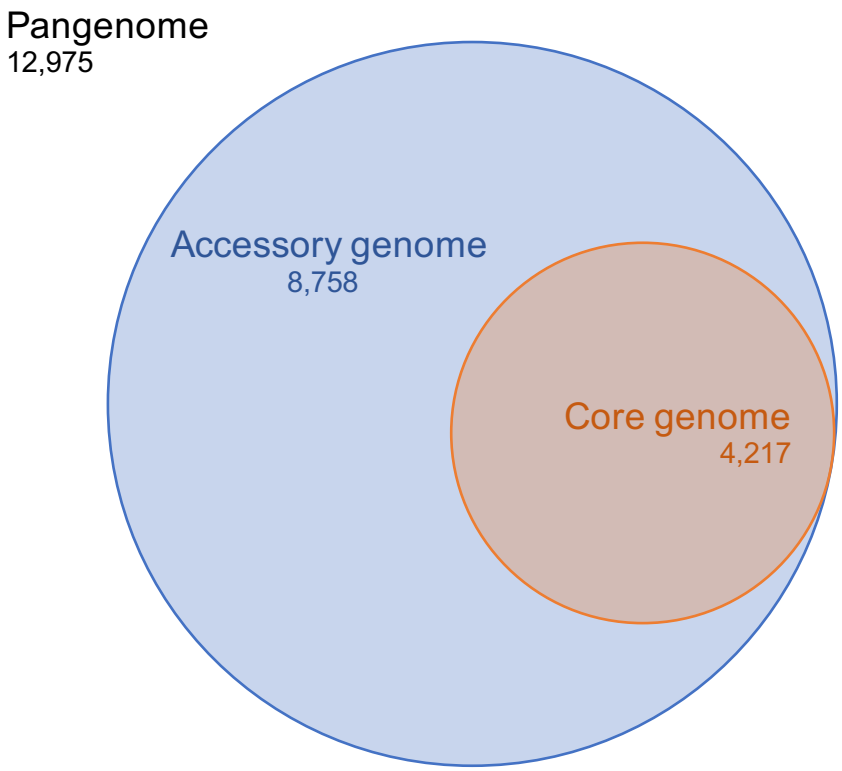


Figure 4.6 Pangenome analysis summary statistics obtained using PIRATE for all of the isolates (>95% isolates).

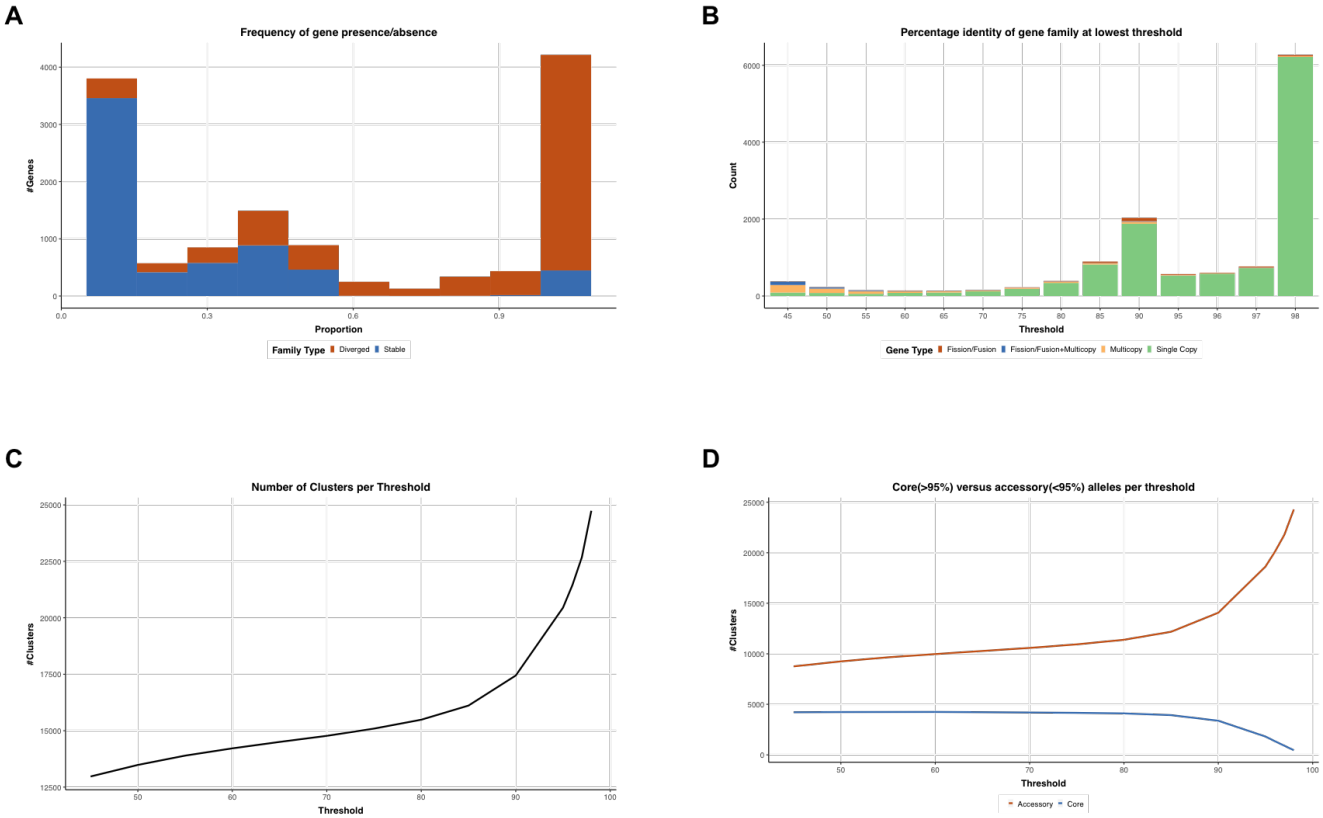


Figure 4.7 Overview of the 14 Atacama Desert isolates pangenome using the complete genomes inferred by PIRATE.

A. Proportion of genomes in which stable gene families were found with a single allele at 98% amino acid identity (**blue**) and diverged with >1 allele (**red**). **B.** Minimum amino acid identity percentage cut-off at which all loci were present per gene family (range between 45-98%). **C.** Selected amino acid percentage thresholds (percentage identity limit to assign gene family cluster) and the corresponding number of clusters. **D.** Core (gene families present in greater than 95% of genomes) and accessory allele estimates for the pangenome at different thresholds.

A dendrogram comprising the pangenome obtained with PIRATE was constructed to observe the presence or absence of genes in each genome and also to compare and identify core regions from each soil (**Figure 4.8**). Once more, phylogeny affiliation showed that isolates within each location affiliated closely with each other but not with isolates from the other location. In this way, isolates from soil 8 clustered closely with the *S. phaeoluteigriseus* type strain forming a clade distinct from soil 7 isolates. However, a core genome was observed containing shared genes between both locations despite some of those showing low sequence identity (**Figure 4.8, red lines**). Furthermore, genes only present within one location were identified for each soil creating two distinct regions that could be part of a species-specific core genome, for soil 7 (**yellow box**)and soil 8 (**green box**). Interestingly, within each soil collection, distinct gene presence in the isolates was discerned suggesting high gene diversity (**black box**). However, mis-annotated genes near contig ends due to fragmented assemblies and over-splitting, where a single gene family is incorrectly split into several smaller clusters, could increase the diversity observed in the accessory genome (Bayliss *et al.*, 2019; Tonkin-Hill *et al.*, 2020).

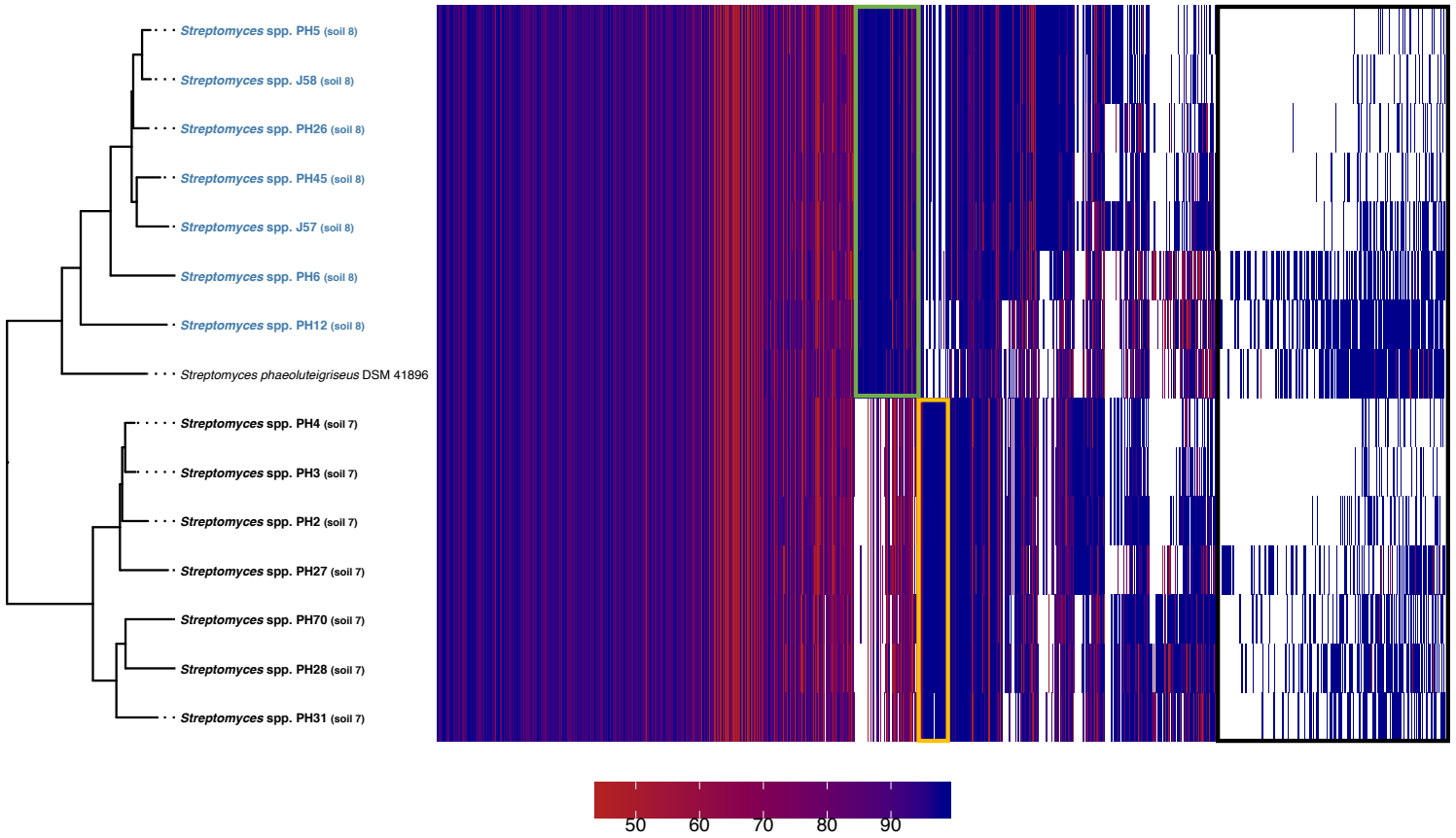


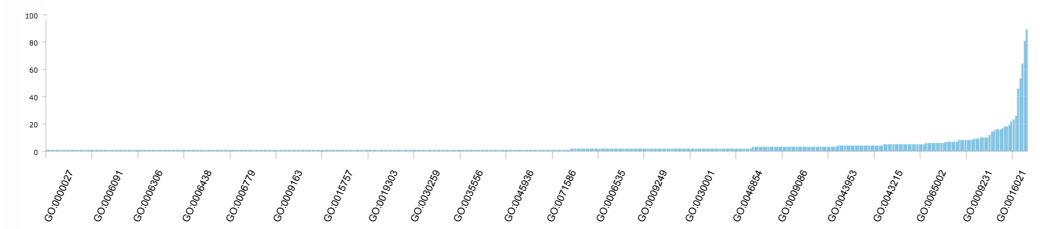
Figure 4.8 Pangenome visualization of the gene presence (low (red) to high sequence identity (blue) to a reference sequence) and absence (white lines) are represented by lines using PIRATE revealing the core and accessory genome in both species.

Species-specific regions are represented in a **green** or **yellow box** comprising genes present in isolates from soil 8 and *S. phaeoluteigriseus* or soil 7 species, respectively. Genes not present in all the isolates are delimited by a **black box**.

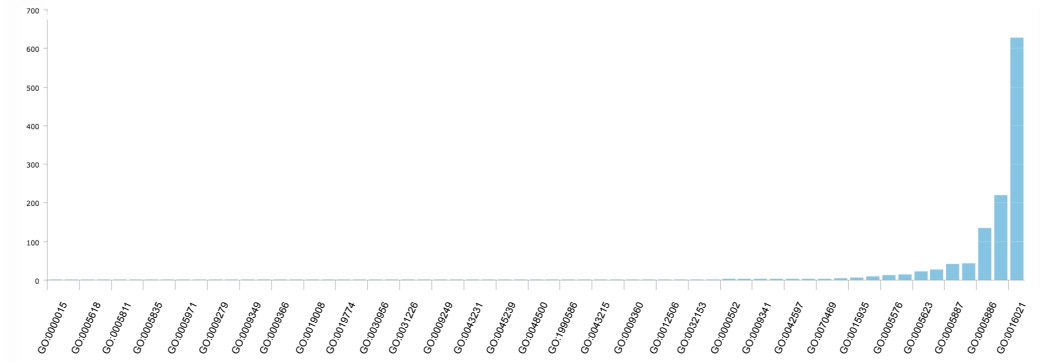
The genes comprising the core genome were analysed by homology search analysis using GO FEAT (Araujo *et al.*, 2018) to obtain functional annotation and genomic enrichment (**section 2.4.2**). Only 74.72% of the core sequences were assigned to three domains of gene ontology (GO): biological process (1,485 hits), cellular component (1,238) and molecular function (3,683). The most abundant biological process GO terms were involved in cell division (GO:0051301) and transmembrane transport (GO:0065002) (**Figure 4.9**). For cellular component the most abundant GO terms were involved in integral component of membrane (GO:0016021) or plasma membrane (GO:0005886 and GO:0005887) (**Figure 4.9**). Most of the molecular function GO terms were mainly involved in binding of polysaccharide (GO:0030247), heme (GO:0020037) and metal ion (GO:0046872) (**Figure 4.9**).

Overall, the analysis of pangenome revealed a large accessory genome associated with higher gene family diversity and could potentially indicate niche driven differences. Additionally, this analysis also shows the presence of intra-specific genetic diversity among isolates.

GO Biological process



GO Cellular component



GO Molecular function

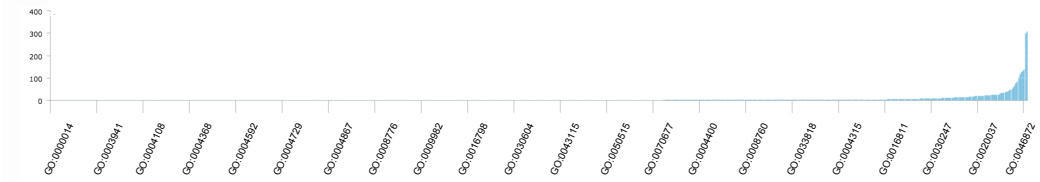


Figure 4.9 Distribution of orthologous genes on COG categories identified in the core genome using GO FEAT.

Bars represent the proportion in each functional category.

4.4 Prophage diversity as a potentiator of niche driven diversity

Temperate phages can integrate extrachromosomal DNA in bacterial genomes by lateral gene transfer creating genetic variances among isolates inhabiting the same location. These belong to the accessory genome and thus to grasp the genomic diversity of the isolates prediction of prophages was performed using PHASTER (Arndt *et al.*, 2016). Overall, 3 intact (score >90), 3 questionable (score 70-90) and 16 incomplete (score <70) prophage regions were identified in the isolates from both locations. Four of these contained more than one region predicted, soil 7 PH28 (1 questionable and 2 incomplete) and PH70 (2 incomplete) and PH6 (3 incomplete), PH12 (1 questionable and 1 incomplete) and J57 (2 incomplete) from soil 8 (**Table 4.3** and **Table 4.4**). Moreover, only incomplete regions were identified for PH5 and PH28 despite having the complete genome. However, two phages, vB-BanS-Tsamsa from *Bacillus anthracis* (Ganz et al., 2014) and Tripp from *Paenibacillus larvae* ATCC 9545 (incomplete) (Abraham et al., 2016), were uniquely identified in the genome of PH28 (**Table 4.3** and **Table 4.4**). Interestingly, the same intact region, matching with phage mu16 from *S. aureofaciens* (Farkašovská et al., 2007) was found only on the isolates PH2, PH3 and PH4 (**Table 4.3**). This phage consists of 38.194 Kbp, 52 predicted genes and 71.2% of GC content. It was found in regions with variable length (38.8 – 43.4 Kbp), number of genes (52 or 61) and GC content (70.61 – 70.69%), similar in the three isolates (~71.6% GC). Furthermore, phage VWB (structurally similar to phage mu16) from *S. venezuelae* ETH14630 (Anne et al., 1984) was identified in both soils on PH31 and PH6 and additionally phage REQ1 from *Rhodococcus equi* (Petrovski, Seviour & Tillett, 2013) identified in PH27 and PH70 from soil 7 and PH12 from soil 8. A potential location specific phage was observed as the phage Minerva from *Mycobacterium* was only present and shared in 4 of seven isolates from soil 8 (**Table 4.4**).

The presence of strikingly diverse prophage regions able to produce genomic rearrangements was observed. However, most of these were potentially inaccurate due to either incomplete or questionable prediction similarities. This could suggest that plasmids rather than phage mediated gene transfer could be the main driver of genetic exchange.

Table 4.3 Prophage regions identified in the genomes of the isolates from soil 7 using PHASTER.

Three classes of completeness considered (intact: score >90), questionable: score 70-90 and incomplete: score <70).

Isolates	Region Position	Region length (Kbp)	Completeness	Number of proteins	Common Phage	GC (%)
PH2	Contig 12	38.8	Intact	52	Strept_mu16 (NC007967/23)	70.61
PH3	Contig 1	43.4	Intact	61	Strept_mu16 (NC007967/24)	70.62
PH4	Contig 1	41.3	Intact	52	Strept_mu16 (NC007967/23)	70.69
PH27	Contig 1	31.7	Questionable	29	Rhodoc_REQ1 (NC016655/4)	70.55
PH28	Chromosome	4.4	Questionable	6	Paenib_Tripp (NC028930/2)	67.50
	Chromosome	6.5	Incomplete	9	Paenib_Tripp (NC028930/2)	67.10
	Chromosome	8.4	Incomplete	7	Bacill_vB_BanS-Tsamsa (NC023007/1)	72.93
PH31	Contig 1	32.3	Incomplete	32	Strept_VWB (NC005345/9)	66.36
PH70	Contig 1	21.4	Incomplete	25	Rhodoc_REQ1 (NC016655/4)	70.94
	Contig 1	12.1	Incomplete	17	Rhodoc_REQ1 (NC016655/3)	70.31

Table 4.4 Prophage regions identified in the genomes of the isolates from soil 8 using PHASTER.

Three classes of completeness considered (intact: score >90), questionable: score 70-90 and incomplete: score <70).

Isolates	Region Position	Region length (Kbp)	Completeness	Number of proteins	Common Phage	GC (%)
PH5	Chromosome	21.5	Incomplete	32	Mycoba_Minerva (NC026584/2)	71.70
PH6	Contig 1	30.5	Incomplete	17	Strept_VWB (NC005345/7)	65.95
	Contig 1	22.4	Incomplete	24	Strept_SF1 (NC028807/8)	70.79
	Contig 1	25.0	Incomplete	26	Strept_YDN12 (NC028974/2)	69.10
PH12	Contig 1	14.7	Incomplete	21	Rhodoc_REQ1 (NC016655/3)	70.27
	Contig 1	24.3	Questionable	30	Strept_SV1 (NC018848/10)	69.51
PH26	Contig 1	20.3	Incomplete	33	Mycoba_Minerva (NC026584/2)	71.42
PH45	Contig 1	20.4	Incomplete	31	Mycoba_Minerva (NC026584/2)	70.73
J57	Contig 1	13.2	Incomplete	16	Strept_R4 (NC019414/3)	69.90
	Contig 1	18.3	Incomplete	26	Mycoba_Minerva (NC026584/2)	71.26
J58	Contig 1	20.4	Incomplete	32	Mycoba_Minerva (NC026584/2)	71.07

4.5 Assembly and complete genome comparison of two isolates confirms divergence between locations

4.5.1 Assembly and *in silico* analysis of representative isolates PH28 and PH5 genomes to understand potential soil specific characteristics

Representatives for the two locations, PH28 and PH5 from soil 7 and soil 8, respectively, were selected for further sequencing using the PacBio platform to obtain long read sequences. The selection was based on the low 16S rRNA identity with the closest related strain *S. phaeoluteigriseus* (**Chapter 3 – Table 3.1**) and the larger genome size obtained after assembly of Illumina reads (**Table 4.1**). Gaps between scaffolds after PacBio assembly were closed manually by comparing gene annotations with two assemblies obtained using Unicycler. The first, a hybrid assembly of Illumina and PacBio reads and the second an assembly of PacBio reads using bold bridging mode. In the end, both genomes were closed with PH28 having one chromosome and one putative plasmid and PH5 one chromosome and two putative plasmids (**Table 4.5**). These results were similar to the observed PFGE (**Figure 4.4** and **Figure 4.5**) except that it was not possible to confirm the presence of the smaller plasmid of PH5 due to its size. All the putative plasmids were categorised as such through identification of genes encoding the partitioning proteins ParAB. Other proteins involved in chromosome segregation and conjugal plasmid transfer traSa (pPH5.2), TraAB (pPH5.1 and pPH28) and TraG/I (pPH5.1 and pPH28) were also identified on the plasmids.

DNA-DNA hybridisation was performed with the GGDC tool (Meier-Kolthoff *et al.*, 2013; Meier-Kolthoff, Klenk & Göker, 2014) to understand species delimitation between the isolates PH28 and PH5, *S. phaeoluteigriseus* and *S. leeuwenhoekii*. The result suggests that PH28 was not related to any of the strains showing DDH values (0.02 – 6.60%) below the threshold for same species delineation (>70%) (**Table 4.6**). However, the threshold obtained for PH5 indicates it is closely related to the species *S. phaeoluteigriseus* (DDH 92.72%) but not to *S. leeuwenhoekii* (DDH 0.02%) or PH28 (DDH 6.36%) (**Table 4.6**).

CGview server (Grant & Stothard, 2008) was used to visualize a genomics map of the genome and plasmids for PH28 and PH5. The majority of biosynthetic gene clusters were near the end of the chromosomes (**Figure 4.10** and **Figure 4.11** – **red lines**) and interestingly, no biosynthetic gene cluster was found in the plasmids, contrary to the previous described plasmids in other *Streptomyces* (Medema *et al.*, 2010; Kinashi, 2011).

Table 4.5 Complete genome statistics of PH28 and PH5 after complete assembly.

Terminal inverted repeats (TIR) on the sequence ends were found for PH5 but not on PH28 whole-genome sequencing.

Final assembly statistics		
PH28	Chromosome (bp)	8,548,382
	GC content (%)	71.6
	Protein coding genes	8,432
	RNAs	86
	TIR (bp)	-
	Plasmid pH28.1 (bp)	239,785
	GC content (%)	68.9
	Protein coding genes	303
	RNAs	0
	TIR (bp)	-
PH5	Chromosome (bp)	9,039,847
	GC content (%)	71.5
	Protein coding genes	8,671
	RNAs	92
	TIR (bp)	8,660
	Plasmid pPH5.1 (bp)	142,705
	GC content (%)	70.4
	Protein coding genes	153
	RNAs	0
	TIR (bp)	458
	Plasmid pPH5.2 (bp)	78,249
	GC content (%)	68.2
	Protein coding genes	85
	RNAs	0
	TIR (bp)	-

Table 4.6 DNA-DNA hybridisation performed using GGDC (DSMZ) of selected isolates PH28 and PH5 with the most closely related strain *S. phaeoluteigriseus* and *S. leeuwenhoekii* isolated from the Atacama Desert.

Considered values obtained from Formula 2 (sum of all identities found in HSPs divided by overall HSP length).

Query	Reference genomes	*S. leeuwenhoekii*	*S. phaeoluteigriseus*	*S.* spp PH5	*S.* spp PH28
PH28	DHH estimate %	26.70 [24.3 - 29.2]	43.90 [41.4 - 46.5]	43.70 [41.2 – 46.3]	-
	Probability that DDH > 70%	0.02	6.60	6.36	
	Probability that DDH > 79%	0.01	1.49	1.44	
	Difference in GC	1.11	0.18	0.19	
PH5	DHH estimate %	26.70 [24.3 – 29.2]	83.20 [80.3 – 85.6]	-	43.70 [41.2 – 46.3]
	Probability that DDH > 70%	0.02	92.72		6.36
	Probability that DDH > 79%	0.01	53.62		1.44
	Difference in GC	1.30	0.01		0.19

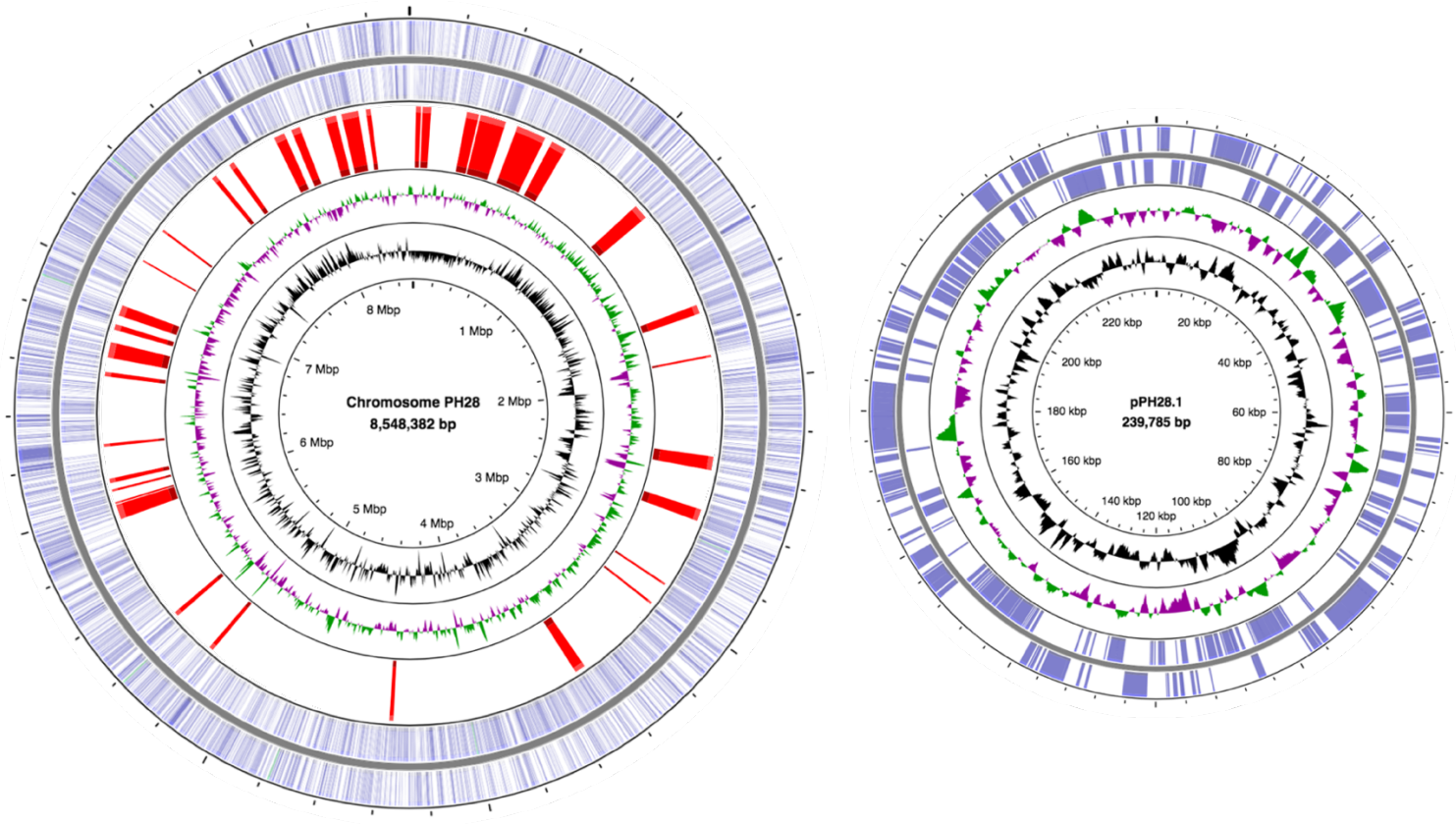


Figure 4.10 Visualisation of the complete genomes of PH28 and the putative plasmid pPH28.1 using CGView.

The linear chromosome is represented as a circle. From outside to inside, the concentric circles represent: nucleotide position; Protein Coding Sequences (PCSs) on the forward strand; PCSs on the reverse strand; the biosynthetic gene clusters (red); the GC-skew indicate below (purple) and above (green) average and the GC-plot (GC %).

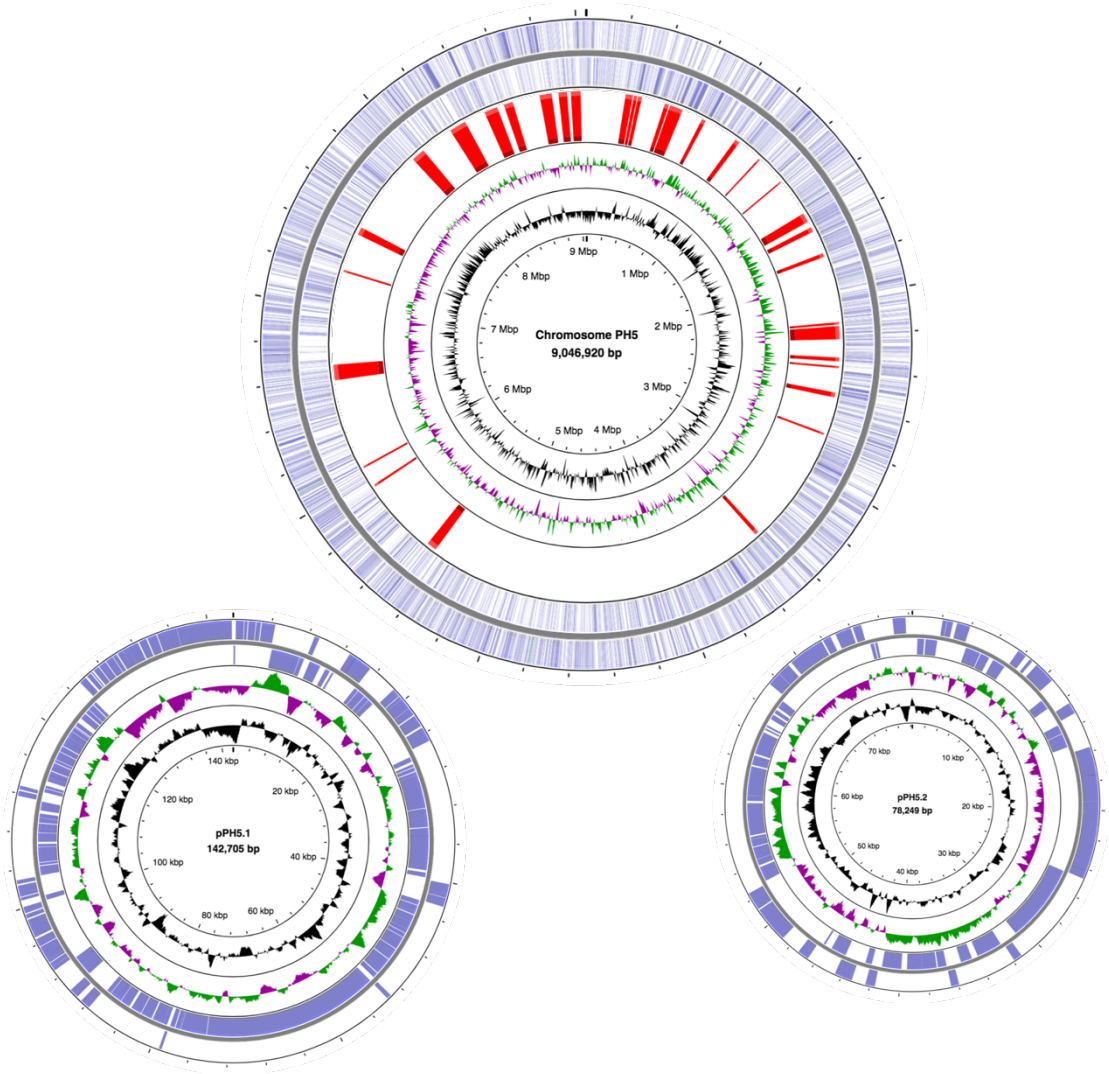


Figure 4.11 Visualisation of the complete genomes of PH5 and the two putative plasmids pPH5.1 and pPH5.2 using CGView.

The linear chromosome is represented as a circle. From outside to inside, the concentric circles represent: nucleotide position; Protein Coding Sequences (PCSs) on the forward strand; PCSs on the reverse strand; the predicted biosynthetic gene clusters (red); the GC-skew indicate below (purple) and above (green) average and the GC-plot (GC %).

In order to investigate if orthologous clusters of essential protein-coding genes were present in plasmids and chromosomes of both PH28 and PH5, a comparison was performed using OrthoVenn2 (Xu *et al.*, 2019). In total, 907 clusters were identified with the majority (843 clusters consisting of 2,939 proteins) present in the chromosome of PH28 and 7 clusters (25 proteins) only present in the plasmid pPH28.1 (**Figure 4.12, panel A**). However, it was not possible to identify the function of these seven clusters as they showed homology with hypothetical proteins (**Figure 4.12, panel A**). Furthermore, only 57 clusters (269 proteins) were shared between plasmid pPH28.1 and chromosome with sporulation and regulation of transcription the most abundant GO (43%) but also processes involved in DNA repair and replication, chromosome segregation and transposition were found (**Figure 4.12, panel C**). In PH5, 883 clusters (3,338 proteins) were assigned only to the chromosome while 19 clusters (51 proteins) were shared between the chromosome and the plasmid pPH5.1 (**Figure 4.12, panel B**). As in pPH28.1, several of these were classified as transcription, DNA replication, chromosome segregation and transposition clusters (**Figure 4.12, panel D** - top). Furthermore, for pPH5.2 and PH5 the majority were related to sensor kinase activity, biotin biosynthetic processes, plasma membrane function and regulation of helicase activity (**Figure 4.12, panel D** - bottom). Interestingly, no homologous clusters were observed between pPH5.2 and pPH5.1 but for the latter, a unique cluster was found comprising three proteins without any homolog in the database (**Figure 4.12, panel B**). Furthermore, no shared gene families were predicted between the plasmid pPH28 and plasmids pPH5.1 and pPH5.2.

Overall, no essential RNAs and biosynthetic gene clusters were found in the plasmids, but the genes identified were involved in important roles of biological process, molecular function and cellular component.

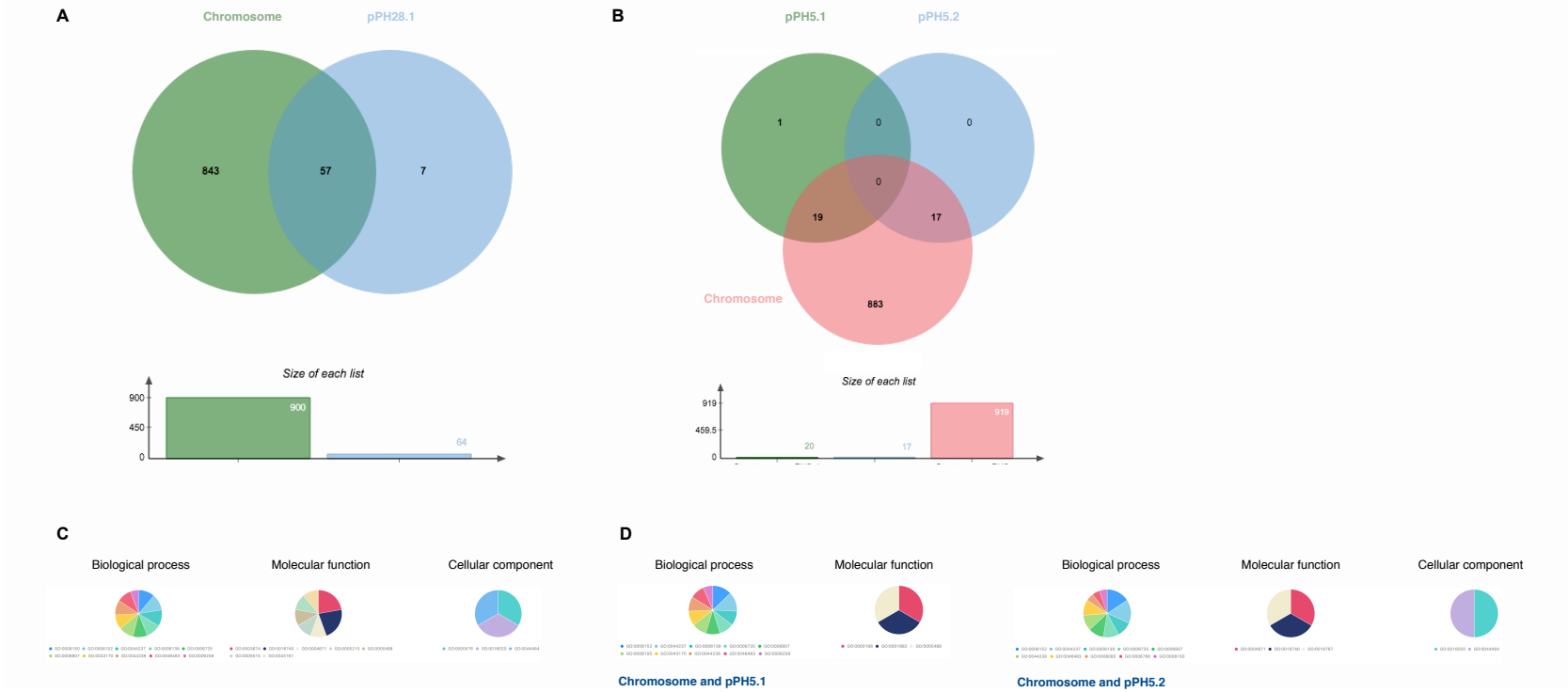


Figure 4.12 Overlapping of orthologous gene clusters for the chromosome and plasmids of PH28 and PH5.

A. Venn diagram showing the total and distribution of shared gene families among PH28 and its plasmid pPH28.1. **B.** Venn diagram showing the total and distribution of shared gene families among PH5 and its two plasmids pPH5.1 and pPH5.2. **C.** Clusters of the overlapping region assigned according to their functional annotation for PH28 and plasmid. **D.** Clusters of the overlapping region assigned according to their functional annotation for PH5 and its plasmids.

4.5.2 Distinct genome structure and synteny between strains from the Atacama Desert and other *Streptomyces* suggests shared traits in this environment

To gain insights if these Atacama Desert isolates showed a similar genomic architecture to other *Streptomyces* species, a comparison was made using conserved genes involved in DNA replication and chromosome segregation, *polA*, *dnaA*, *parA*, *parB*, *recA* and *tpg*. Comparison of these genes and their location in different *Streptomyces* (*S. leeuwenhoekii*, *S. griseus* and *S. coelicolor*) was adapted from Tsai *et al.*, 2011 (**Figure 4.13**). Almost all of the genes selected were identified in the genomes except *tpg* was not found in *S. griseus* corresponding to its atypical telomere sequence (Ohnishi *et al.*, 2008). Moreover, this gene was located on different arms depending on the genomes, e.g. on the right end of *S. coelicolor*, PH28 and PH5. Interestingly, genes *polA* and *recA* from the Atacama Desert species, including *S. leeuwenhoekii*, were translocated when comparing with the other species. It was possible to identify the genes *tap* and *tpg* in the collected isolates and a blast search showed their homology with genes from other *Streptomyces* species (**Table 4.7**). In plasmid pPH5.1 and pPH5.2, only terminal proteins Tap and Tpg, respectively, were identified suggesting that both plasmids could be necessary for replication of the telomeres.

Furthermore, the telomere sequences were extended following work done by Lis Algora using an adaptation of the protocol described by Fan *et al.* (2012). Briefly, genomic DNA was digested with blunt end enzymes *PvuII* and *EcoRV* located at the ends followed by self-ligation of the fragments. Then PCR and sequencing were performed to obtain the telomere sequences. For PH5 chromosome, its terminal regions were extended, however, it was not possible to achieve the same result for PH28.

Isolate PH5 showed conserved and similar telomeres to the archetypal family found in *S. coelicolor* (**Figure 4.14, panel A**). However, palindrome I was incomplete for isolate PH5 due to the missing nucleotides (**Figure 4.14, panel A**). Adding these nucleotides and comparing with the clover-leaf structure of *S. coelicolor* (**Figure 4.14, panel B)** the same archetypal telomere structure was exhibited (**Figure 4.14, panel C**).

Moreover, to infer synteny a dot plot was constructed using the tool D-GENIES (Cabanettes & Klopp, 2018) comparing PH28, PH5 and the type strains. Overall, isolates PH28 and PH5 matched for the most part of their length (matches <50% represents 64.90%) however some gaps/breaks were observed with sequences represented only in one of the isolates (**Figure 4.15 and Figure 4.16**). These breaks were more predominant near the ends, that is thought to be a highly variable region among *Streptomyces* (Bentley & Parkhill, 2004). Furthermore, analysing the breaks in one of the ends of the chromosome (R1: 2.3 - 2.7 Mbp), a lanthipeptide cluster was found in PH28 and a NRPS in PH5 (**Figure 4.17**). Likewise, a cluster for T2PKS-butyrolactone (R2: 6.2 – 6.6 Mbp) was predicted for PH5 but not in PH28 suggesting that these could explain the loss of synteny (**Figure 4.17**). However, a gap (R2) was observed in the central core region (3.8 – 4.7 Mbp) and by aligning them, two main breaks emerged (**Figure 4.17**). This region contained mostly hypothetical proteins however genes involved in transcription regulation, DNA replication and amino acid synthesis were found (**Table 4.8** and **Table 4.9**). Interestingly, a putative integrase was identified as *intQ* facilitating integration of the phage in the genome. Similarly, for the other gap hypothetical proteins composed the majority of the predictions and a putative prophage major tail sheath protein was identified (**Table 4.8** and **Table 4.9**). Nonetheless, no prophage was predicted by PHASTER in this region suggesting uncertainty in the prediction.

When comparing the collected isolates with the type strains, sequence inversion was exhibited in the central region of the genome between 4.1 – 5.2 Mbp in *S. coelicolor* and *S. griseus*. This event could be explained by the inverted orientation of genes in the central region of the genome (**Figure 4.13**). Furthermore, more breaks were observed and even the obtained matches were, for the most part, below 25% for *S. leeuwenhoekii* and *S. coelicolor* or no match for *S. griseus* (**Figure 4.15 and Figure 4.16**). Overall, the genome architecture and synteny between the collected isolates in this study and *S. leeuwenhoekii* was greater than to other more distantly related *Streptomyces* species.

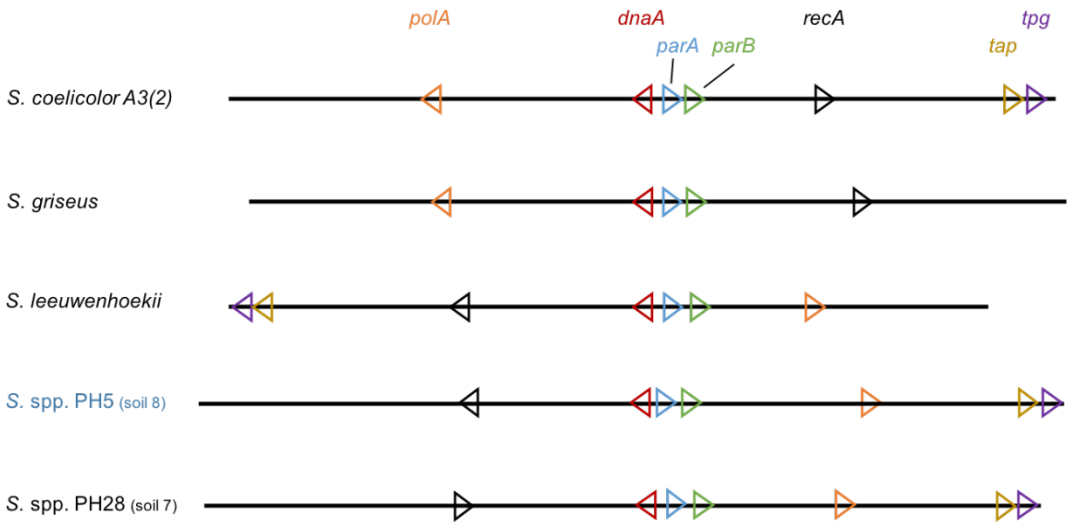


Figure 4.13 Location of some genes in several *Streptomyces* complete genomes oriented and aligned to *S. coelicolor dnaA*, normally adjacent to the origin of replication (*oriC*).

The arrowhead corresponds to the direction of transcription and the colour the different genes.

Table 4.7 Best-hit obtained by blast search of genes *tap* and *tpg* identified in the chromosome or plasmid of isolates PH28 and PH5.

	Gene	Query coverage	E value	Identity (%)	Organism
PH28	*tap*	100	0	86.04	*Streptomyces cellulosae*
	tpg	100	1e-125	95.68	*Streptomyces rishiriensis*
pPH28	*tap*	100	0	88.71	*Streptomyces bikiniensis*
	tpg	100	5e-125	95.68	*Streptomyces parvulus*
PH5	*tap*	100	0	96.80	*Streptomyces albidochromogenes*
	tpg	100	5e-126	96.76	*Streptomyces viridochromogenes*
pPH5.1	*tap*	100	4e-159	81.27	*Streptomyces spp ST5*
	tpg	-	-	-	-
pPH5.2	*tap*	-	-	-	-
	tpg	100	1e-113	89.13	*Streptomyces olindensis*

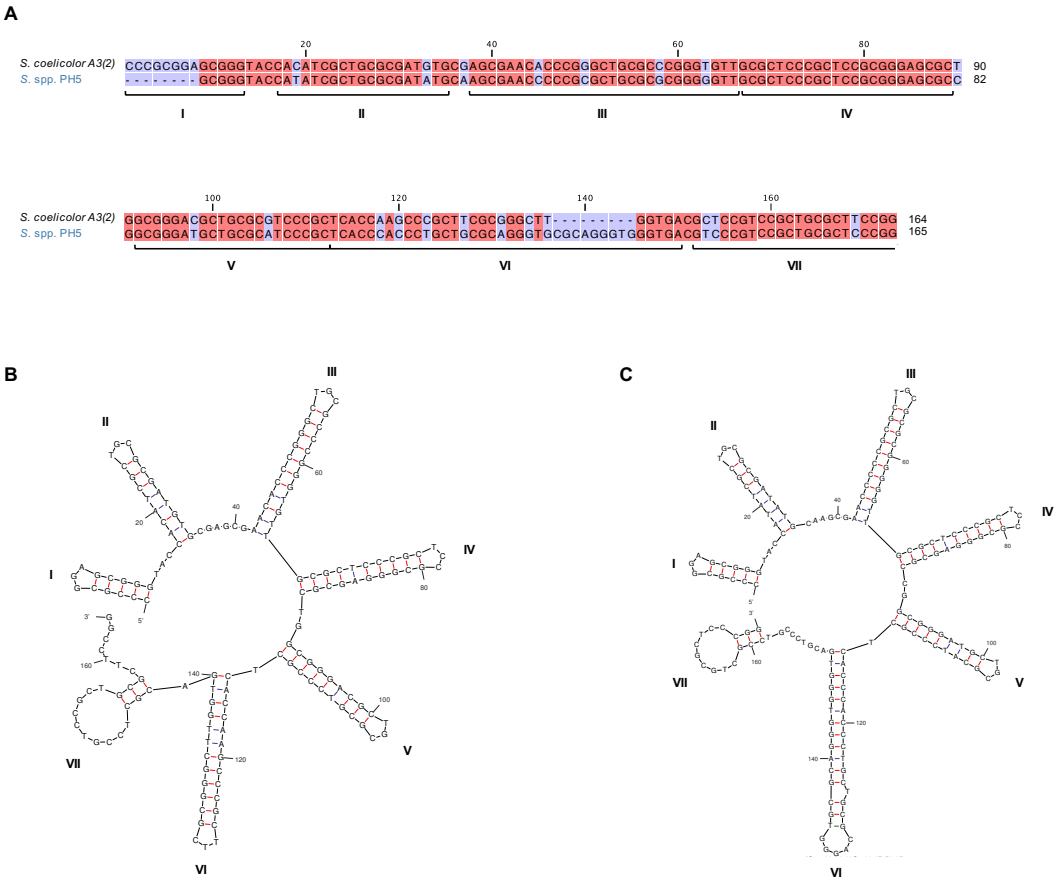


Figure 4.14 Comparison of the archetypal telomeres of *S. coelicolor* and the isolate PH5. **A.** Alignment of the two telomeres sequences showing the initial eight missing nucleotides corresponding to palindrome I. **B.** Clover structure of *S. coelicolor* telomere showing the seven palindromes. **C.** Structure predicted of isolate PH5 showing the archetypal telomere structure.

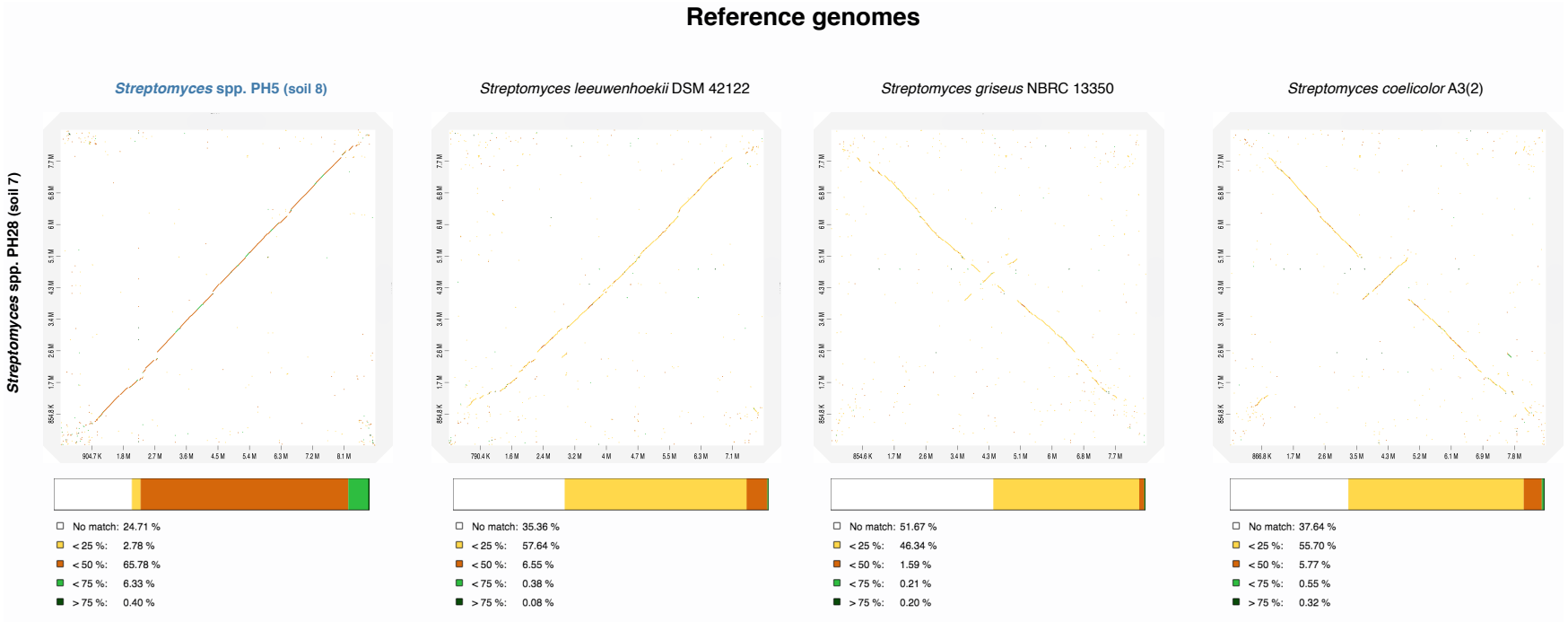


Figure 4.15 Dot plot genome alignment of the the soil representative genome of PH28 (soil 7) and type strains using D-Genies.

The homologous genes and matches for each identity threshold are: dark green lines indicate >75% identity, light green <75%, orange <50% and yellow <25%.

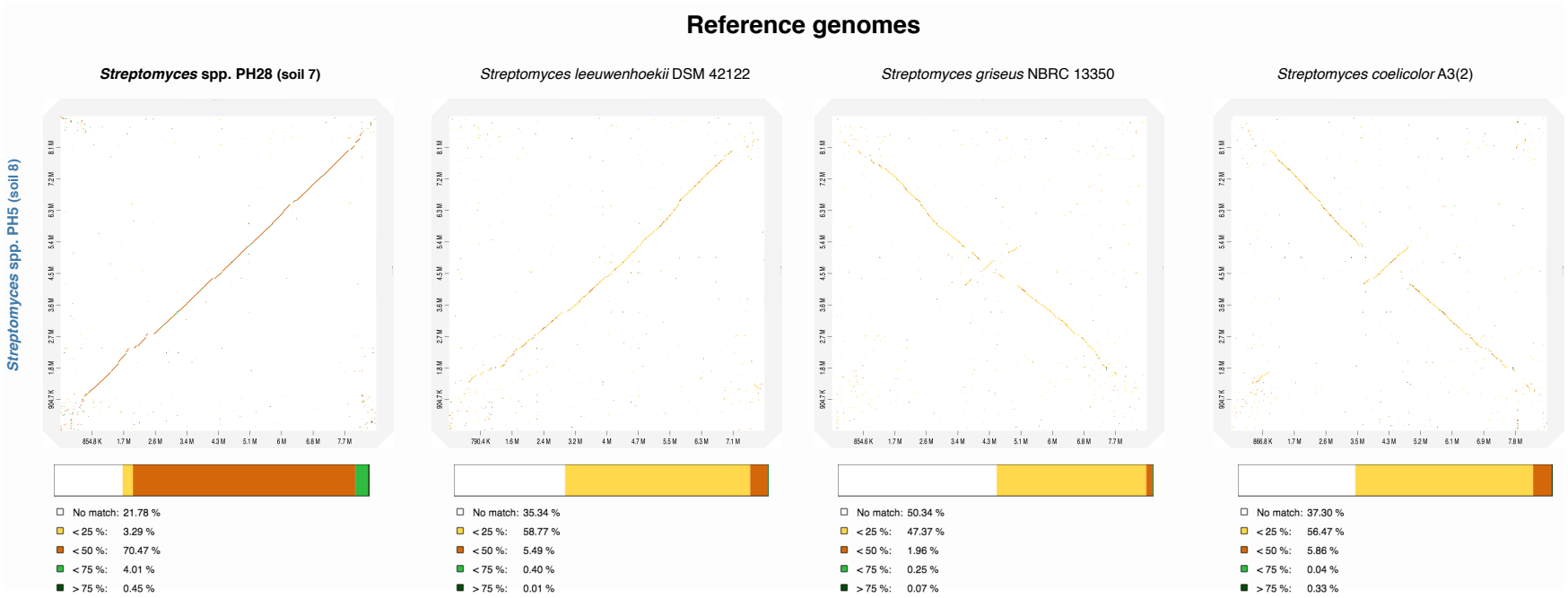


Figure 4.16 Dot plot genome alignment of the the soil representative genome of PH5 (soil 8) and type strains using D-Genies.

The homologous genes and matches for each identity threshold are: dark green lines indicate >75% identity, light green <75%, orange <50% and yellow <25%.

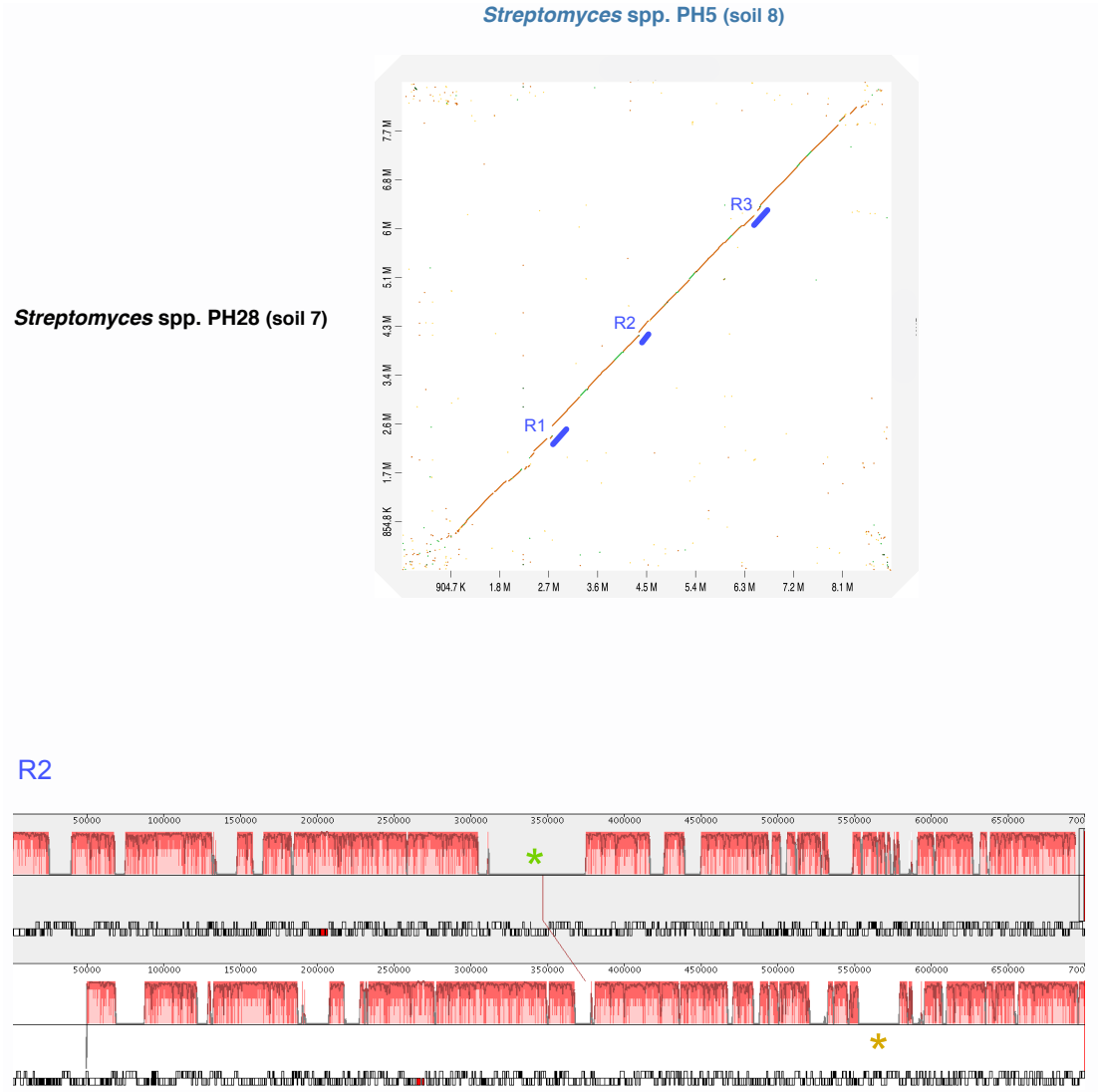


Figure 4.17 Regions containing mismatches between the alignment of PH28 and PH5 genomes.

Alignment of region 2 showing the differences in the core region of the genomes using Mauve (PH28: green asterisk, PH5: orange asterisk).

Table 4.8 Genes predicted to be encoded in the gap localised in the central core region of PH28 genome and absent at the same location in isolate PH5.

Gene	Product	Length (bp)	COG
	hypothetical protein	1725	
	hypothetical protein	3564	
	hypothetical protein	441	
	hypothetical protein	387	
	hypothetical protein	444	
	hypothetical protein	453	
	hypothetical protein	387	
	hypothetical protein	375	
	hypothetical protein	225	
	hypothetical protein	264	
	hypothetical protein	531	
	hypothetical protein	408	
	hypothetical protein	1740	
	hypothetical protein	474	
	hypothetical protein	225	
	hypothetical protein	1038	
whiB1	Transcriptional regulator WhiB1	600	
dnaC	DNA replication protein DnaC	711	COG1484
	hypothetical protein	942	
	hypothetical protein	360	
	hypothetical protein	993	
	putative protein	1176	COG1228
	hypothetical protein	1116	
	hypothetical protein	531	
	F420H(2)-dependent biliverdin reductase	168	
valS	Valine--tRNA ligase	2628	COG0525
thrB	Homoserine kinase	1047	
	hypothetical protein	885	
ipuC	Glutamate--isopropylamine ligase	1428	
pheP	Phenylalanine-specific permease	1536	COG1113
	hypothetical protein	1002	
	hypothetical protein	195	
	hypothetical protein	441	
	hypothetical protein	1239	
gph	Phosphoglycolate phosphatase	810	

	hypothetical protein	1014	
purR	HTH-type transcriptional repressor PurR	1497	
	Glycogen synthase	1362	COG0438
	Putative 2-hydroxyacid dehydrogenase	993	COG1052
moaA	GTP 3',8-cyclase	1122	COG2896
	hypothetical protein	1101	
pqqE	Coenzyme PQQ synthesis protein E	813	
tobZ	nebramycin 5' synthase	1728	
spoT	Bifunctional (p)ppGpp synthase/hydrolase SpoT	606	COG0317
	hypothetical protein	993	
gltX	Glutamate--tRNA ligase	1416	COG0008
	hypothetical protein	780	
	hypothetical protein	555	
	hypothetical protein	1851	
	hypothetical protein	228	
	hypothetical protein	639	
intQ	Putative defective protein IntQ	1641	COG0582

Table 4.9 Genes predicted to be encoded in the gap localised in the central core region of PH5 genome and absent at the same location in isolate PH28.

Gene	Product	Length (bp)	COG
	hypothetical protein	2886	
nreC	Oxygen regulatory protein NreC	702	COG2197
	hypothetical protein	654	
ftsH	ATP-dependent zinc metalloprotease FtsH	1965	COG0465
	Serine protease 1	1320	
	hypothetical protein	1422	
yqjZ	putative protein YqjZ	456	COG2329
	hypothetical protein	660	
	hypothetical protein	717	
gpFI	Putative prophage major tail sheath protein	1593	
	hypothetical protein	444	
	hypothetical protein	474	
	hypothetical protein	153	
	hypothetical protein	2208	
	hypothetical protein	1509	
	hypothetical protein	426	
	putative protein	726	COG0627
vgrG1	Actin cross-linking toxin VgrG1	1905	COG3501
	hypothetical protein	444	
	hypothetical protein	1959	
	hypothetical protein	558	
	hypothetical protein	1308	
	hypothetical protein	141	
	tRNA-Gly(tcc)	72	

4.5.3 Genomic islands predicted mostly in higher frequency at the end of PH28 and PH5 genomes

Genomic islands (GI) comprising large genomic sequences (> 8 Kbp) and transmitted by horizontal gene transfer can encode for antibiotic resistance or virulence genes. Since the presence of prophages was ambiguous, the prediction of genomic islands in the two complete genomes was performed using IslandViewer 4 (Bertelli *et al.*, 2017). Comparing PH28 (**Figure 4.18**) and PH5 (**Figure 4.19**) with the type strain *S. leeuwenhoekii*, selected since it is the only available whole genome *Streptomyces* from the Atacama Desert, the predicted presence of genomic islands was predominantly at the ends of the sequences. Overall, the integrated results from IslandViewer predicted 135 GI and 1,354 genes in the genome of PH28 and 173 GI and 1,852 genes in the genome of PH5, contrasting with the prediction obtained with *S. leeuwenhoekii* with 162 GI and 955 genes.

To understand if these GIs were located near the putative prophage regions (**section 4.4)** a comparison of both results was performed. Both PH28 (GI 1: 78,105 - 120,977 bp) and PH5 (G2: 6,711,323 – 6,726,650 bp and G3: 6,729,901 – 6,746,153 bp) genomic islands were present in some part in other *Streptomyces* species. In the genomic island 1 (GI 1) from PH28, 40 encoded hypothetical proteins and 18 with predicted function were found and were involved in arsenic and mercury resistance as well as potassium transport. Part of this region was present in *S. leeuwenhoekii* (89% identity/24% coverage), *Streptomyces nigra* strain 452 (85%/11%) and *Streptomyces* sp. S10(2016) (84%/24%). Additionally, it was found also in plasmid pFRL5 from *Streptomyces* sp. F8 (85%/11%).

Regarding PH5, the genomic island 2 (GI 2) encoded 28 hypothetical proteins and low sequence similarity was found with *Streptomyces* sp. P3 (80% identity/21% coverage). Additionally, low sequence identity between genomic island 3 (GI 3) and plasmid pFP3 from *Streptomyces* sp. F2 (77% identity/25%); *Streptomyces spongiicola* HNM0071 (78% identity/10% coverage), *Nonomuraea sp.* ATCC 39727 (79% identity/7% coverage) and *Streptomyces sp.* MK-45 (76% identity/13% coverage). Furthermore, it encoded 19 hypothetical proteins and 1 putative *B. subtilis subsp. subtilis 168* prophage skin YqbO (putative tail length tape-measure protein).

Among the GIs predicted close to the prophage regions there were no antimicrobial resistance or virulence factors.

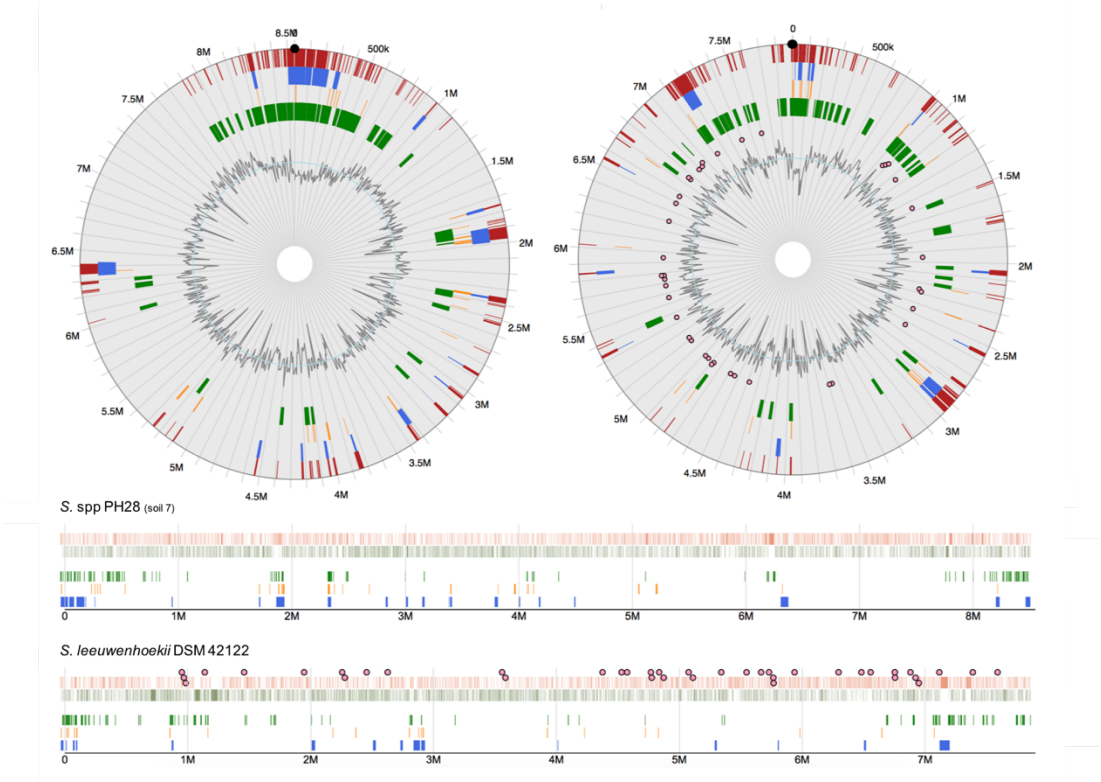


Figure 4.18 Circular visualisation GI predicted on the genome of PH28 (left) from soil 7 and *S. leeuwenhoekii* (right) and linear of the respective genomes (bottom) in Islandviewer 4.

From outside to inside, the blocks represent: predicted by at least one method (red); present in IslandPath-DIMOB (blue) – based on associated GIs features such as tRNAs, transposon elements, integrases and sequence bias (Hsiao *et al.*, 2003); present in SIGI-HMM (orange) (Waack *et al.*, 2006) – based on the codon usage bias with a Hiddden Markov model criterion; present in IslandPick – based on genome comparison (green) (Langille, Hsiao & Brinkman, 2008); GC-skew indicate below and above average (black) and homologs of resistance genes (circular glyphs).

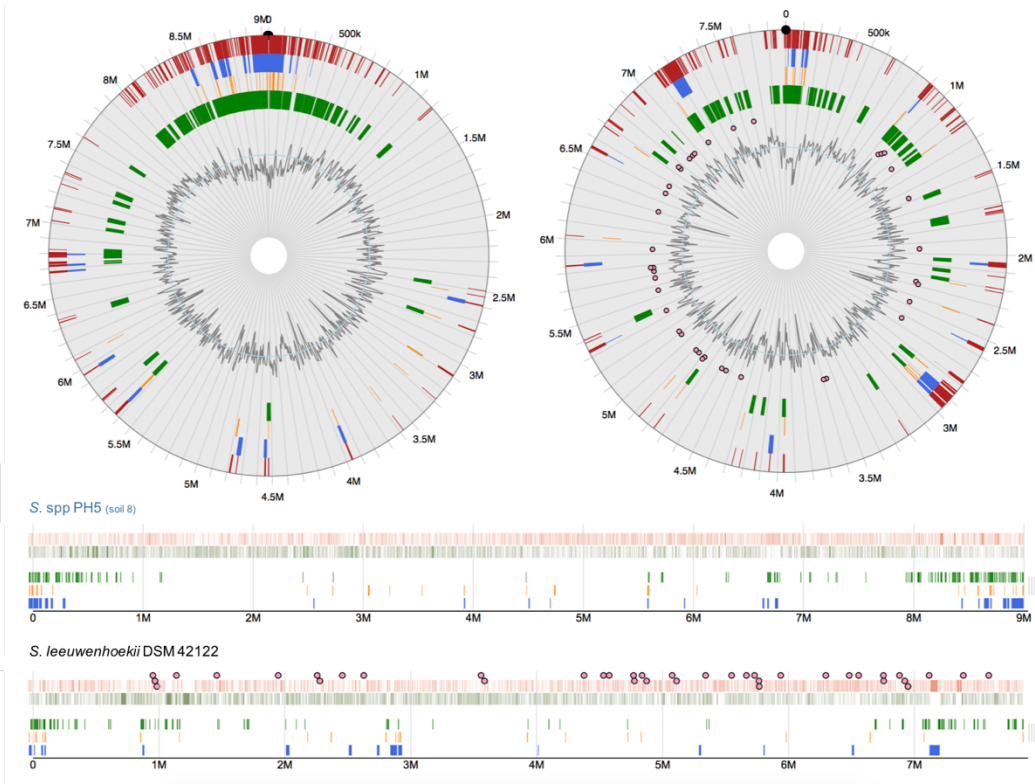


Figure 4.19 Circular visualisation of GI predicted on the genome of PH5 (left) from soil 7 and *S. leeuwenhoekii* (right) and linear of the respective genomes (bottom) in Islandviewer 4.

From outside to inside, blocks represent: predicted by at least one method (red); present in IslandPath-DIMOB (blue) – based on associated GIs features such as tRNAs, transposon elements, integrases and sequence bias (Hsiao *et al.*, 2003); present in SIGI-HMM (orange) (Waack *et al.*, 2006) – based on the codon usage bias with a Hiddden Markov model criterion; present in IslandPick – based on genome comparison (green) (Langille, Hsiao & Brinkman, 2008); GC-skew indicate below and above average (black) and homologs of resistance genes (circular glyphs).

4.6 Common specialised metabolites found among the isolates but also location-specific BGCs were identified

Whole-genome sequencing provides the ability to reveal and predict biosynthetic gene clusters encoded in the genomes. However, a limiting factor is the presence of high number of contigs (e.g. Illumina sequencing) that affects BGCs predictions by truncating genes or lack of annotation. Accordingly, assembly improvement of the remaining isolates was performed using MeDuSa (Bosi *et al.*, 2015) and the complete genomes of PH28 and PH5 as references, for soil 7 and soil 8, respectively. By using this strategy, for each isolate was obtained fewer and larger contigs (N50) (**Table 4.10**).

These were submitted for prediction of secondary metabolites with antiSMASH 5.0 (Blin *et al.*, 2019) and the obtained BGCs classes per genome compared between isolates. Additionally, a phylogenetic tree was constructed using FastTree (Price, Dehal & Arkin, 2010) after aligning the core genes with MUSCLE (Edgar, 2004) and curation with Gblocks (Castresana, 2002) (**Figure 4.20**). On average 30-36 BGCs were predicted to be encoded by the isolates genomes and the pan-biosynthetic genome comprised 470 clusters belonging to 16 types. Several of these types were present in all isolates from both locations including terpenes, non-ribosomal peptides (NRPS), lanthipeptides, bacteriocins, siderophores, melanin, ectoine and indole. Interestingly, some were only predicted in species from soil 8 such as lasso peptides (**Figure 4.20**; except PH45) and nucleosides (**Figure 4.20**). Additionally, using BiG-SCAPE (Navarro-Muñoz *et al.*, 2020), it was possible to identify 76 singletons (single cluster) in 7 distinct groups, Terpenes, NRPS, Ribosomally synthesized and post-translationally modified peptides (RiPPs), Polyketides type I (T1PKS), PKS-other, PKS-NRPS hybrids and “Others” (**Figure 4.21** and **Figure 4.22**).

Seventy-six families were identified in the isolates and several BGCs were present in isolates from the same or different locations (**Figure 4.21** and **Figure 4.22**). A core biosynthetic gene cluster can be defined a BGC shared by several other strains. Independently of location and species core BGC’s were found, composed of albaflavenone and hopene (Terpenes), melanin and ectoine (Others), bacteriocins (RiPPs) and spore pigment (PKSother) among others (**Figure 4.21** and **Figure 4.22**).

However, the majority of the isolates shared more BGCs with isolates inhabiting the same location and even then, showing some diversity within those, e.g. PH12 and PH2 shows differences in the gene cluster SapB (RiPP) (**Figure 4.23**) and cinerubin B (PKSother), respectively (**Figure 4.21**). Remarkably, the gene cluster encoding cinerubin B biosynthesis was only found in genomes from all soil 7 isolates.

Table 4.10 Sequencing statistics obtained after scaffolding with PH28 and PH5 as reference genomes for soil 7 and soil 8, respectively, using MeDuSa.

Scaffolding statistics					
PH2	# of contigs	54	PH6	# of contigs	114
	Length (*Mbp*)	7.9		Length (*Mbp*)	9.3
	N50 (*Mbp*)	7.8		N50 (*Mbp*)	9.1
	# genes	7,028		# genes	8,508
PH3	# of contigs	69	PH12	# of contigs	78
	Length (*Mbp*)	8.1		Length (*Mbp*)	8.5
	N50 (*Mbp*)	8.1		N50 (*Mbp*)	8.3
	# genes	7,244		# genes	7,515
PH4	# of contigs	54	PH26	# of contigs	54
	Length (*Mbp*)	8.0		Length (*Mbp*)	9.2
	N50 (*Mbp*)	8.0		N50 (*Mbp*)	9.1
	# genes	7,195		# genes	8,301
PH27	# of contigs	79	PH45	# of contigs	48
	Length (*Mbp*)	8.3		Length (*Mbp*)	8.7
	N50 (*Mbp*)	8.1		N50 (*Mbp*)	8.7
	# genes	7,463		# genes	7,826
PH31	# of contigs	74	J57	# of contigs	91
	Length (*Mbp*)	8.4		Length (*Mbp*)	9.1
	N50 (*Mbp*)	8.3		N50 (*Mbp*)	8.9
	# genes	7,561		# genes	8,202
PH70	# of contigs	38	J58	# of contigs	66
	Length (*Mbp*)	8.2		Length (*Mbp*)	9.1
	N50 (*Mbp*)	8.1		N50 (*Mbp*)	8.9
	# genes	7,224		# genes	8,259

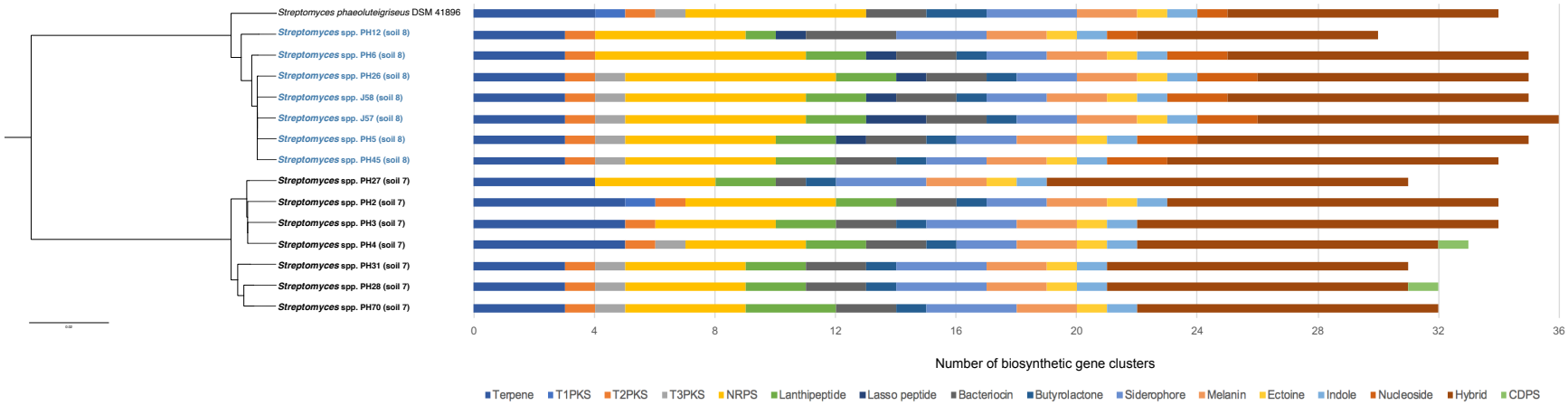


Figure 4.20 Graphic representation and distribution of the number of biosynthetic gene clusters per genome identified using antiSMASH 5.0. Concatenated alignments of 4,217 core genes were aligned using MUSCLE, curated with Gblocks and used to construct a maximum likelihood phylogenetic tree using FastTree (default parameters). T1PKS, type I polyketide synthase; T2PKS, type II polyketide synthase; T3PKS, type III polyketide synthase; NRPS, non-ribosomal peptide synthase; CDPS, tRNA-dependent cyclodipeptide synthases. Colours represents the different biosynthetic gene clusters.



Figure 4.21 Clustering in different types of biosynthetic gene clusters, PKSother, RiPPS, PKSI, Terpenes and PKS-NRP hybrids shared between isolates within the same or different location obtained using BiG-SCAPE.

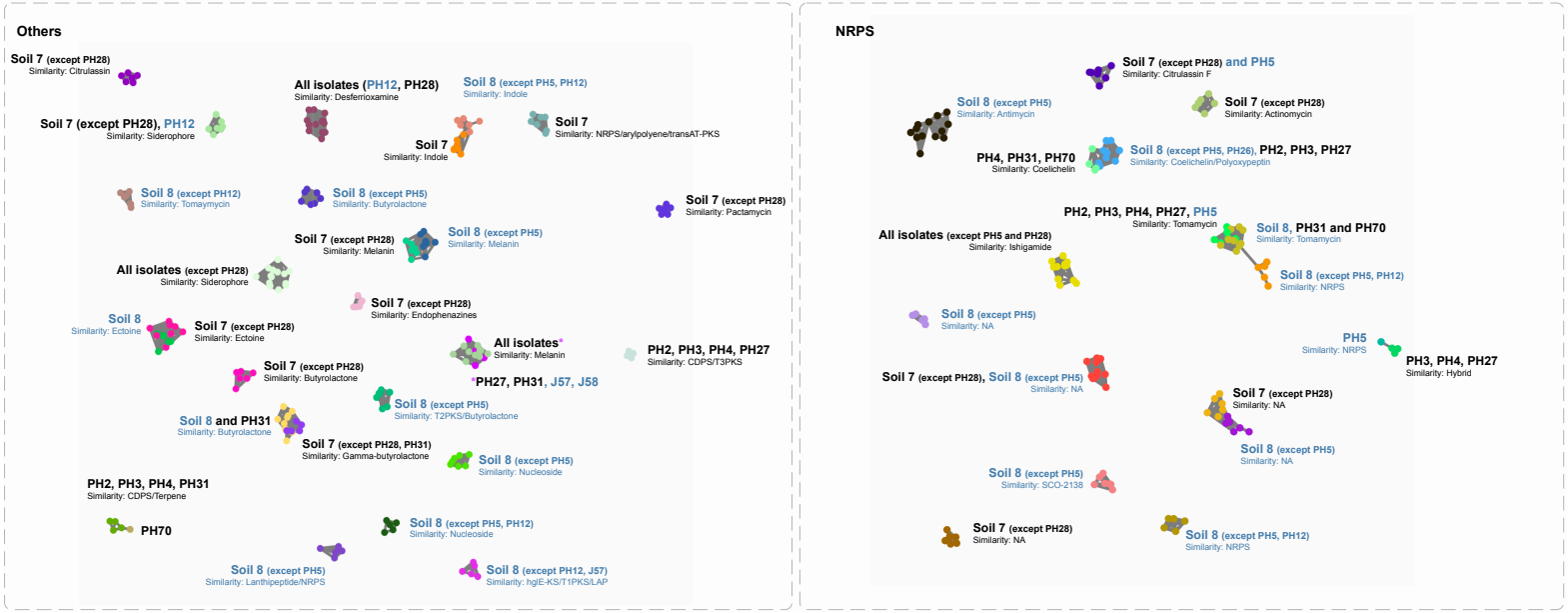


Figure 4.22 Clustering in different types of biosynthetic gene clusters, Others and NRPS, shared between isolates within the same or different location obtained using BiG-SCAPE.

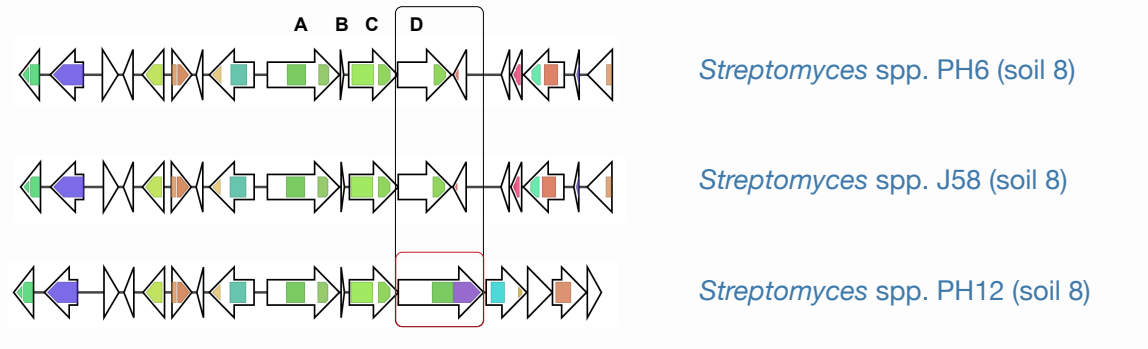


Figure 4.23 Differences in the biosynthetic gene cluster for SapB of PH12 and comparison with two *Streptomyces* species isolated from soil 8.

Letters indicate genes found in the *ramCSAB* gene cluster of *Streptomyces coelicolor*. **A** indicates *ramC* with a serine/threonine protein kinase and lanthionine synthetase domains, **B** indicates *ramS* that will originate the lanthipeptide and **C** and **D** indicate *ramA* and *ramB* encoding ATP-binding cassette transporters. *Streptomyces* spp. PH12 shows the *ramB* gene (**grey box**) replaced by a serine/threonine protein kinase gene (**red box**).

4.7 Discussion

Sequencing the whole genome of an organism allows for a complete overview of the genetic information stored in the DNA. Accordingly, the collected isolates in this study were sequenced to uncover potential novel metabolic pathways, biosynthetic gene clusters to associate with phenotype and to understand genomic differences among them.

Distinct genome sizes and GC content were found among the isolates from both locations (**Table 4.1** and **Table 4.2** and **Figure 4.1**). Genome size and gene numbers differ between bacterial strains owing to several genetic events, e.g. gene duplication/loss, horizontal gene transfer and genome rearrangements, directed by specific niche pressures (Bentley & Parkhill, 2004; Ventura *et al.*, 2007; Nishida, 2012). Similarly, the GC content is thought to be dependent on the same pressures and thus a correlation between these parameters can be established (Bentley & Parkhill, 2004; Nishida, 2012). The collected closely related strains showed differences between them with soil 8 having larger genomes and slightly lower GC content (**Table 4.2**) however this difference is minor (±0.3%). Nevertheless, the genome structure of the isolates is consistent with previous identified *Streptomyces* species in terms of genome size and GC content (Ventura *et al.*, 2007).

Comparison of housekeeping genes and prediction of genome-wide average nucleotide identity was proven useful in previous studies to resolve inter- and intra-species relatedness (Rong *et al.*, 2010; Glaeser & Kämpfer, 2015; Labeda *et al.*, 2017). When comparing these isolates using different methodologies (MLST, ANI and PFGE), a distant relationship was observed between them suggesting that isolates from soil 7 and soil 8 are indeed different species (**Figure 4.2** to **Figure 4.4**). This contradicts the previous 16S rRNA phylogeny obtained previously (**Chapter 3**) and confirming the recent work of Antony-Babu *et al.* (2017) that showed the inaccuracy of 16S rRNA phylogenies to represent genomic variations among related strains.

Furthermore, the widely used PFGE can be applied to characterise and differentiate closely related *Streptomyces* species regarding genome topology and presence of

extrachromosomal DNA (Leblond *et al.*, 1990; Lezhava *et al.*, 1995; Harunari *et al.*, 2018; Gomez-Escribano *et al.*, 2015). The results in this current work showed high diversity and location dependent presence of extrachromosomal replicons between isolates (**Figure 4.5** and **Figure 4.6**). The most diversity was observed in soil 8, with 4 different profiles that shows some degree of correlation with the MLST and ANI constructed where PH5, PH26, PH45, J57 and J58 were more closely related (ANI = 100%) than to PH6 or PH12 (ANI = 99%) (**Figure 4.2**). Additionally, a lower MW band observed in some of the collected isolates could correspond to a plasmid with a similar size (122-130 Kbp) identified in *S. leeuwenhoekii* (Busarakam *et al.*, 2014). However, after aligning plasmid pPH5.1 with pSLE2 only 5 predicted genes were similar: a paralog to *ssbA*, the recombinase *xerC* that converts dimers of the chromosome into monomers allowing their segregation and three hypothetical proteins at the termini. Interestingly, the isolates are not particularly closely related to any of the other Atacama Desert strain described so far (**Figure 4.3)** showing how this environment is a potential source of novel strains and chemistry (Bull & Goodfellow, 2019; Bull & Asenjo, 2013).

Whole genome sequencing and the study of pangenomes contributed to a better understanding of species intra-specific diversity and the role of environmental pressures and natural selection. The pangenome contains all the genes present in a community comprising the core genome and accessory or dispensable genome responsible for the genetic variation (McInerney, McNally & O'Connell, 2017). A study using 5 *Streptomyces* type strains with complete genomes revealed a pangenome and core genome of 17,362 and 3,096 ortholog families (18% of the total), respectively (Zhou *et al.*, 2012). Two more studies using 17 and 31 genomes produced bigger pangenomes (34,592/29,695 ortholog families) but smaller core genomes (2,018/2,048 corresponding to 6-7%) (Kim *et al.*, 2015; Tian *et al.*, 2016). More recently, 122 *Streptomyces* genomes were analysed and 1,048 gene families (3%) were found to compose the core genome of a 39,893 gene families pangenome (McDonald & Currie, 2017). In this way, increasing the numbers of genomes will decrease the predicted core genome due to additional predicted ortholog families not present in the strains (Kim *et al.*, 2015). Accordingly, the collected 14 isolates showed a pangenome of 12,975 gene families and 4,217 constituting the core genome (32% of total gene families) (**Figure 4.7**). A higher prediction of core genes was obtained than in previous studies (**Table 4.11**) and within each Atacama Desert soils

(**Table 4.12**). This could imply a gradual speciation process from a common ancestor of soil 8 derived *Streptomyces phaeoluteigriseus* strains and soil 7 species which could explain their level of relatedness and number of shared genes. Thus, corresponding to the observations from the MLST tree with these two sets of isolates forming a separate clade from the other *Streptomyces* type strains (**Figure 4.3**). However, several genes were specific to only one location or even specific to particular strains inside a location (**Figure 4.24**). Accordingly, indicating unique ecological variation by horizontal gene transfer as observed by the diversity of plasmid profile rather than phage gene transfer, with only complete phage identified (Park & Andam, 2019; Tidjani *et al.*, 2019). Incomplete phage regions were identified in all the isolates and distinctions emerged between locations with some prophages only identified in one soil (**Table 4.3** and **Table 4.4**). Only in isolates PH2, PH3 and PH4 were identified intact phages, more specifically phage mu16 from the Singleton group. Interestingly, only in PH2, this phage (1,112 – 39,920 bp) was observed close to a predicted NRPS cluster (2 - 28,766 bp). In one of the flanking regions, a homologous TcpC protein (position 1,227 bp) from plasmid pCW3 of the Gram-positive *Clostridium perfringens* was identified (Steen *et al.*, 2009). This conjugative plasmid encodes antibiotic resistance or toxin genes and, in this case, tetracycline resistance (Steen *et al.*, 2009). Despite not finding other essential genes belonging to the transfer clostridial plasmid (*tcp*) cluster of pCW3 in this region it could be evidence of some genetic exchange among bacteria in the environment mediated by phages.

Table 4.11 Previous pangenome analysis of *Streptomyces* strains and the total percentage of core genes present in the pangenome.

	Zhou *et al.* 2012	This study 2020	Kim *et al.* 2015	Tian *et al.* 2016	McDonald & Currie 2017
Genomes	5	14	17	31	122
Pangenome	17,362	12,975	34,592	29,695	39,893
Core genome	3,096	4,217	2,018	2,048	1,048
% total	18	32	6	7	3

Table 4.12 Pangenome analysis of *Streptomyces* strains in each soil and the percentage of shared genes considered as core genome using PIRATE.

The analysis performed by Zhou *et al.* was repeated using the same program for comparison purposes.

	Soil 7	Soil 8	Both	Zhou *et al.* 2012
Genomes	7	7	14	5
Pangenome	9,897	10,447	12,975	15,576
Core genome	5,655	6,055	4,217	3,098
% total	57	58	32	20



Figure 4.24 Differences in the orthologous gene clusters and proteins between isolates from both soils.

A. Venn diagram showing the total and distribution of shared gene families among isolates from soil 7 and soil 8 and **B.** proteins identified as shared (2189) but also unique encoded in soil 7 (69) and soil 8 (73).

To obtain the complete genomes of *S. avermitilis* and *S. coelicolor*, intensive hands-on work was performed involving steps of clone growth and isolation, library preparation in vectors, and whole shotgun sequencing (Ikeda *et al.*, 2003; Bentley *et al.*, 2002). Recently, with the advent of next-generation sequencing complete genomes can be obtained using one (Wibberg *et al.*, 2016) or two different sequencing platforms (Rückert *et al.*, 2015), combining short and long reads to improve genome coverage. An example of the later is the recent published genome of *S. leeuwenhoekii* that was sequenced in two different sequencers. Then, it was analysed in-house using computer-intensive software and manually edited to obtain a complete genome (Gomez-Escribano *et al.*, 2015). An approach to manually close the genome was achieved for representatives of each location isolates, PH28 and PH5 (**Table 4.5**), by comparing the ends of the contigs from a PacBio assembly and a polished assembly obtained using Unicycler and pilon.

Subsequently, physical maps for each isolate were obtained with PFGE to confirm the *in silico* assembly following previous studies (Gomez-Escribano *et al.*, 2015; Lezhava *et al.*, 1995; Leblond *et al.*, 1990). One megaplasmid was identified in PH28 and PH5, however, a smaller plasmid (78 kbp) in PH5 was not visible due to the conditions used (**Figure 4.5** and **Figure 4.6**). Furthermore, no RNAs necessary for primary metabolism were found in the megaplasmids, although genes encoding *parAB* (partitioning proteins) and the *tap/tpg* system for telomere replication were found. Additionally, in pPH28.1 and pPH5.2 genes associated with biological process, molecular function and cellular component were found, e.g. involved in DNA repairs or membrane, but in pPH5.1 no relevant genes emerged for cellular components. Some megaplasmids were found to encode specialised metabolites for instance, plasmid SCP1 from *S. coelicolor* carries a BGC to methylenomycin A, pKSL from *Streptomyces lasaliensis* produces lasalocid and pSLA2-L from *Streptomyces rochei* carries four BGCs (Medema *et al.*, 2010; Migita *et al.*, 2009). Interestingly, *S. clavuligerus* pSCL4 carries more biosynthetic gene clusters than the chromosome with 25 and 23 putative BGCs identified, respectively (Medema *et al.*, 2010; Kinashi, 2011). However, for the two megaplasmids identified, pPH28.1 (239 kbp) and pH5.1 (142 kbp), no biosynthetic gene clusters were predicted despite showing similar size to *Streptomyces rochei* pSLA2-L (210 kbp).

DNA-DNA hybridization comparison between the collected isolates and the type strains *S. coelicolor*, *S. leeuwenhoekii* and *S. griseus* showed very low similarity

indicating their distinct nature (**Table 4.6**). However, the genomes match the typical structures of other *Streptomyces* regarding genome size, *S. coelicolor* (8.6 Mbp), *S. griseus* (8.5 Mbp) and *S. leeuwenhoekii* (7.9 Mbp), number of plasmids *S. coelicolor* (2 plasmids) and *S. leeuwenhoekii* (2 plasmids) and location of DNA polymerases and chromosome segregation (**Figure 4.14**) (Gomez-Escribano *et al.*, 2015; Bentley *et al.*, 2002; Lezhava *et al.*, 1995; Tsai *et al.*, 2011).

A conserved central region encoding essential genes is normally found in *Streptomyces* species (Medema *et al.*, 2010; Ikeda *et al.*, 2003). This feature was observed in this study by synteny analysis that showed co-linearity between the collected isolates and *S. leeuwenhoeki* (**Figure 4.16**). However, comparing with *S. griseus* and *S. coelicolor* a large chromosomal inversion formed near the *oriC* region (3.5-5.2 Mbp), thought to be involved in chromosome evolution (Ventura *et al.*, 2007), and breaks appear near the ends (1-2Mbp) with low matches (**Figure 4.15**). This was observed before while comparing the genomes of *S. avermitilis* with *S. coelicolor* (Ikeda *et al.*, 2003) and more recently of 17 *Streptomyces* species (Kim *et al.*, 2015). The analysis of these regions revealed the presence of prophage sequences, non-essential genes and potentially acquired DNA (Ikeda *et al.*, 2003; Bentley & Parkhill, 2004). In fact, likewise to Ikeda *et al.* (2003) and Medema *et al.* (2010), several conserved BGCs among the isolates, such as albaflavenone, ectoine, desferrioxamine and gamma-butyrolactone, were found in the central region of the genomes (between 2.5 - 6.5 Mbp). Furthermore, previous studies demonstrated that most clusters for specialised metabolites were located near the ends and were diverse between species (Ikeda *et al.*, 2003; Medema *et al.*, 2010). Analysing both ends of PH28 chromosome (0-1 Mbp and 7-8 Mbp), the presence of seven and four biosynthetic gene clusters were identified, respectively (**Figure 4.11**). Similarly, in PH5 six biosynthetic gene clusters were encoded in each end (**Figure 4.12**) corresponding in both cases to approximately 34% of the total BGCs predicted confirming previous studies.

After assembly scaffolding the quality was assessed using QUAST to determine scaffold number and N50 (**Table 4.10**). Further complementary tools have also been developed to assess the completeness of assembled genomes. One such example, the Benchmarking Universal Single-Copy Orthologue (BUSCO), accomplishes this by matching sets of genes expected to be present as single-copy orthologues in a

specific taxonomic group (Seppey, Manni & Zdobnov, 2019). It was successfully applied to study insect genomics (Waterhouse *et al.*, 2019), phylogenomics (Shen *et al.*, 2016) and transcriptomics(Cogne *et al.*, 2019).

The genome of each isolate was found to harbour 30-36 BGCs, with the majority of those predicted to produce NRPS-PKS hybrids, NRPS, terpenes and siderophores (**Figure 4.19**). Common BGCs were found such as butyrolactone, ectoine, melanin, indole, siderophore and terpenes among the isolates indicating an important role, e.g. ectoine as response to high salinity. This high number of clusters and common BGC among strains is consistent with previous studies of other *Streptomyces* species (Choudoir, Pepe-Ranney & Buckley, 2018; Seipke, 2015; Park & Andam, 2019) . However, cinerubin B was only predicted in seven isolates all from soil 7 indicating potential biological relevance in that particular location.

In conclusion, by using whole-genome analysis, it was shown that isolates from soil 7 might comprise a new species, with further confirmation by chemotaxonomic studies, and soil 8 a subspecies of *S. phaeoluteigriseus*. Furthermore, high genetic diversity and distinct biosynthetic gene clusters were observed showing unique inter and intraspecific variations. The ability to produce bioactive compounds was analysed using plug assays and metabolomics.

Chapter 5. Bioactive compounds and metabolomics

5.1 Introduction

Secondary metabolites can hold an important role as growth inhibitors of neighbouring cells (Romero *et al.*, 2011). The widely use of antibiotics had an pivotal role in improving Human health but the increase of antimicrobial resistance in clinically relevant pathogens, e.g. *A. baumannii*, *P. aeruginosa* and *Enterobacteriaceae* (such as *K. pneumoniae* and *E. coli*) (World Health Organization, 2017) require immediate attention. Several studies are trying to tackle this problematic by either combining antibiotics ((Chung & Ko, 2019; Ren *et al.*, 2019), using antibiotic scaffolds (Bhamidimarri *et al.*, 2019) or inhibiting bacterial virulence (Feng *et al.*, 2019). Furthermore, several antibiotics that had lower activity against bacteria were repurposed and used in as antitumor complementing chemotherapy, despite undesired side effects such as cardiotoxicity (Krohn, 2009).

One such example are the well-known actinomycins isolated from actinomycetes in 1940 by Waksman and Woodruff (Waksman & Woodruff, 1940). They showed ability to inhibit Gram-positive bacteria, mostly *Staphylococcus* and *Bacillus* species, and more also antitumor activity (Waksman & Woodruff, 1940). However, several studies exposed their toxicity but similar to anthracyclines, they are used nowadays as antitumor agents, e.g. in infantile kidney tumour (Bitzer, Gesheva & Zeeck, 2006; Wang *et al.*, 2013; Philips *et al.*, 1960). All actinomycins share an actinoyl chromophore attached with an amide linkage to two macrocyclic pentapeptide lactones that can possess variations in amino acids. Actinomycin D (**Figure 5.1, panel A**) is the most widely studied among them containing identical amino acids (L-Thr, D-Val, L-Pro, *N*MeGly and *N*MeVal) in each peptidolactone macrocycles (*iso*-actinomycin) and considered as the reference model to study structural modifications in other actinomycins. Variation in the two macrocycles, assigned as the α and β-rings, contributes to distinct structures and cytotoxicity. Actinomycin C_2 (**Figure 5.1, panel B**) for example, holds an additional methyl group in the β-ring while in actinomycins G_2 and G_3 (**Figure 5.1, panel C-D**), threonine was either chlorinated or

hydroxylated, respectively, in their β-ring (Bitzer, Gesheva & Zeeck, 2006; Wang *et al.*, 2013). Furthermore, the latter contain N-Me alanine instead of a N-Me valine in the β-ring (Bitzer *et al.*, 2009; Bitzer, Gesheva & Zeeck, 2006; Wang *et al.*, 2013). It was observed that actinomycin G_2 had higher cytotoxicity and antibacterial activity than G_3. Similar to actinomycin Z_3 and Z_5 (**Figure 5.1, panel E-F**), replacement of the chlorinated group with a hydroxyl moiety substantially decreased the bioactivity (Bitzer, Gesheva & Zeeck, 2006).

Anthracyclines are aromatic polyketides that are widely used in drug combinations to treat patients with lymphoma, breast cancer, carcinomas and sarcomas and are increasing the survival rate of patients by 50% (Mordente *et al.*, 2009; Binaschi *et al.*, 2001). Anthracyclines structurally consist of an tetracyclic molecule with an anthraquinone as their backbone that could be connected to a sugar moiety by a glycosidic linkage trhough a hydroxyl unit resulting to an *O* linkage or the less common C-C linkage, that exhibit its activity by forming a complex with DNA topoisomerase II (Binaschi *et al.*, 2001). Members of the anthracycline group of compounds have been isolated from *Actinobacteria*, mainly *Streptomyces* spp., e.g. daunomycin, doxorubicin, aclacinomycins and cinerubins (**Figure 5.2**) (Lown, 1993; Krohn, 2009; Kronheim *et al.*, 2018). Cinerubins (1-hydroxyaclacinomycin) were first isolated by Keller-Schierlein & Richle (1970) and are structurally similar to aclacinomycins. However, they differ in the number of hydroxyl substituents on the anthraquinone moiety (**Figure 5.2, panel C-D**). Cinerubins were described to exhibit antiparasitic and antitumor activity (Oki *et al.*, 1979; Krohn, 2009; Hayase *et al.*, 2015; Wang, Mata & Hernandez, 1975).

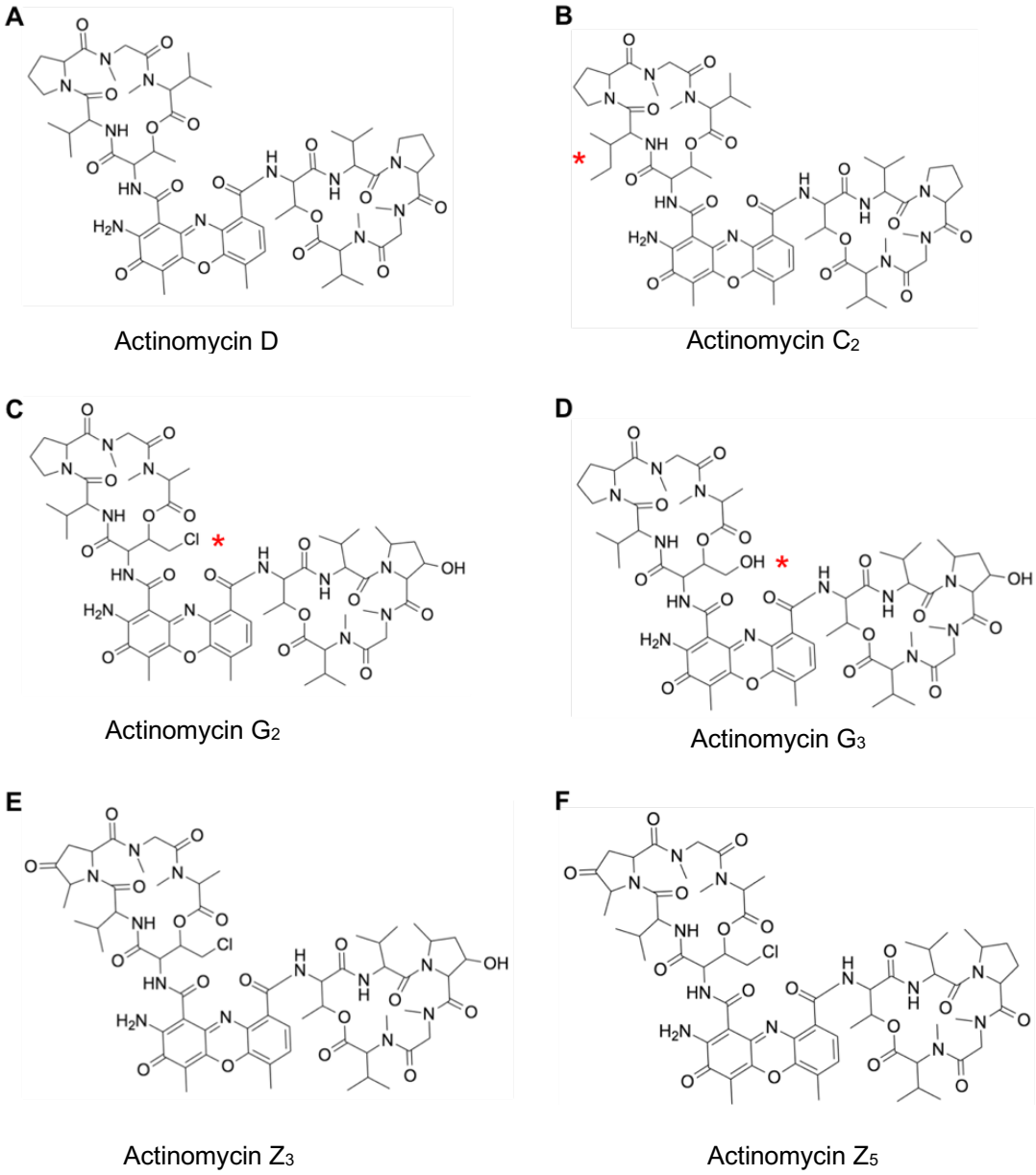


Figure 5.1 Chemical structures of actinomycins.

A. actinomycin D ($C_{62}H_{86}N_{12}O_{16}$), **B.** actinomycin C_2 ($C_{63}H_{88}N_{12}O_{16}$), **C.** actinomycin G_2 ($C_{61}H_{83}ClN_{12}O_{17}$), **D.** actinomycin G_3 ($C_{61}H_{84}ClN_{12}O_{18}$), **E.** actinomycin Z_3 ($C_{62}H_{83}ClN_{12}O_{18}$) and **F.** actinomycin Z_5 ($C_{42}H_{51}NO_{16}$). The structural differences in the rings are indicated with an asterisk. Created using ChemDraw JS (available at https://chemdrawdirect.perkinelmer.cloud/js/sample/index.html).

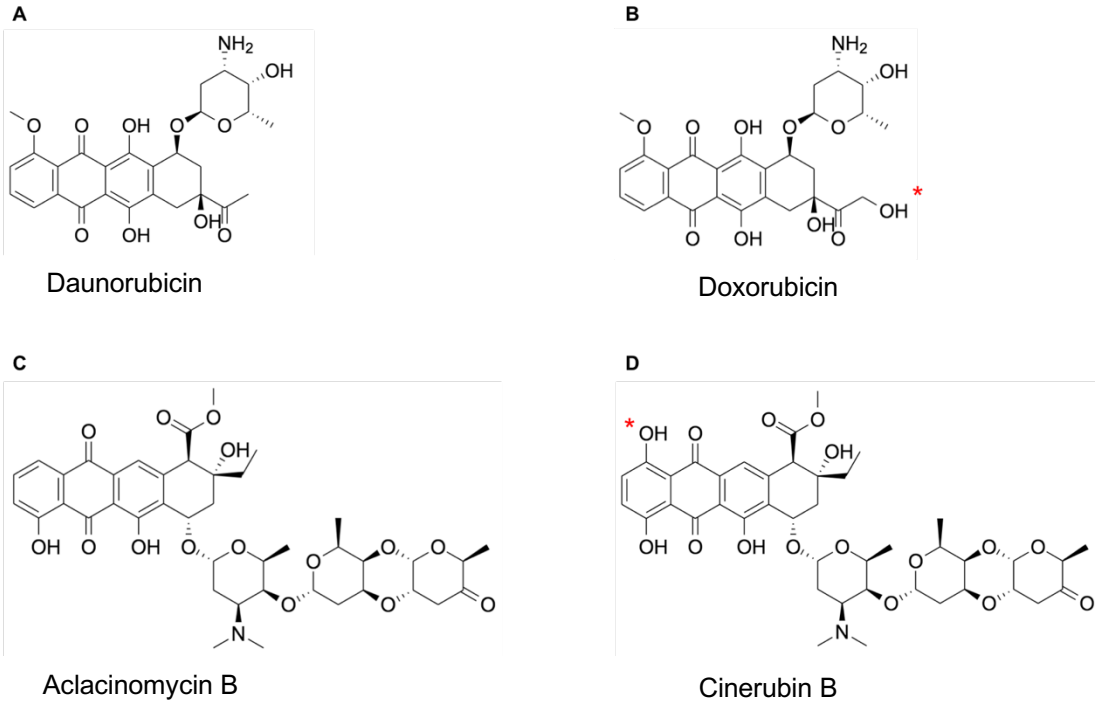


Figure 5.2 Chemical structures of four anthracyclines.

A. daunorubicin ($C_{27}H_{29}NO_{10}$), **B.** doxorubicin ($C_{27}H_{29}NO_{11}$), **C.** aclacinomycin B ($C_{42}H_{51}NO_{15}$) and **D.** cinerubin B ($C_{42}H_{51}NO_{16}$). The extra hydroxyl substituent is indicated with an asterisk. Created using ChemDraw JS (available at https://chemdrawdirect.perkinelmer.cloud/js/sample/index.html).

To discover novel antibiotics, methodologies involving heterologous expression of cryptic biosynthetic gene clusters, derivatisation of known chemical structures or even exploration of microorganisms from extreme environments are the current approaches. Isolation of bacteria from extreme environments in combination with strain characterisation and metabolomics, as a mean to accelerate dereplication and discovery of new compounds, is a successful strategy (Tomova *et al.*, 2015; Arnau *et al.*, 2016; Mahajan & Balachandran, 2017). Metabolomics is a high throughput technique that uses a range of analytical techniques, e.g. LC-MS and NMR, and multivariate data analysis to detect and narrow the potential bioactive metabolites present in biological samples by untargeted spectra acquisition (Wu *et al.*, 2015b).

Comparison of secondary metabolomes of bacterial fermentations by NMR and/or MS provide information related to chemical structure and molecular mass for quick but efficient dereplication and prioritisation of compounds to be targeted for scale-up and isolation work (Wu *et al.*, 2015b). This methodology has proven successful for the discovery of several antibiotics such as the anti-MRSA compound bottromycin D (Hou *et al.*, 2012), the antifungal forazoline A (Wyche *et al.*, 2014) and 7-prenylisatin active against *B. subtilis* (Wu *et al.*, 2015a). The rapid improvement of sequencing technologies and subsequent bioinformatics programs for genome characterisation and mining afford a linkage between chemistry, genomics and microbiology (Soldatou *et al.*, 2019). The active development of BGC prediction tools for genome mining, such as antiSMASH 5.0 (Blin *et al.*, 2019), BAGEL (van Heel *et al.*, 2018) and PRISM (Skinnider *et al.*, 2015), delivers an overview of secondary metabolites encoded by one strain and the similarity between those metabolites can be observed using BiG-SCAPE. One such example is the Natural Product Peptidogenomics (NPP) developed for chemotype-to-genotype associations of peptides using mass shifts obtained from MALDI-TOF spectra and biosynthetic gene clusters from sequenced organisms. (Kersten *et al.*, 2011).

In this study, fourteen *Streptomyces* isolates from the Atacama Desert were investigated for (1) their bioactivity against the indicator strains including ESKAPE pathogens, (2) batch fermentation (small- to large scale-up) and analysis of the metabolome for further characterisation (LC-MS/MS and NMR), and (3) linking the bioactive compounds with the corresponding biosynthetic gene clusters (BGC) predicted with antiSMASH 5.0.

5.2 Production of active metabolites against one or more indicator strains observed in solid media

The strains were tested for their bioactivity potential against indicator strains including ESKAPE pathogens along with *B. subtilis* by plug assays after 14 and 21 days of incubation in the standard ISP medias used for *Streptomyces* characterisation. The bioassay results for the inhibited indicator strains are shown in **Table 5.1**.

All the isolates collected from the different soil samples showed activity against *B. subtilis*, but this was also dependent on the type of media used. For instance, isolate PH3 had the largest diameter of inhibition zone against *B. subtilis* in ISP3 (1.0-1.1 cm) but isolate PH5 showed inhibition in ISP5 (0.5-0.9 cm). Almost all the strains produced a compound active against *S. aureus* however, with globally smaller diameters of inhibition zones (0.3-0.4 cm). Additionally, only species from soil 7 showed activity against *E. faecalis*. Interestingly, isolates PH2, PH3 and PH4 showed activity against *A. baumannii*. According to WHO, finding new antibiotics against *A. baumannii* pathogen is one of the top priorities. *A. baumannii* is considered as opportunistic and nosocomial with wide resistance to currently available antibiotics in the market. Furthermore, isolate PH2 was the most promising, showing also bioactivity against *E. coli* (inhibition zone 0.1-0.2 cm) as well as against four other pathogens.

Overall, it was observed that these isolates belonging to the *Streptomyces* genus were able to produce metabolites with activity against mostly Gram-positive bacteria and posed different potencies, when the isolates were grown in different media.

Table 5.1 Bioactivity results using six ISP media (ISP2-ISP7) and the Atacama isolates (only shown indicator strains inhibited by the isolates, 5 out of 7).

Medium and inhibition zones size represented by numbers and coloured circles, respectively. Size of inhibition zones (cm): **green** (0.1-0.2); **yellow** (0.3-0.4); **orange** (0.5-0.9) and **red** (1.0-1.1). Numbers: medium (ISP2-7) where the inhibition zone was observed. Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**).

Isolates	*E. faecalis*	*S. aureus*	*A. baumannii*	*E. coli*	*B. subtilis*
S. phaeoluteigriseus					7 4 3 2
PH2	3	3 2 7 6 4	3	3	3 2 7 6
PH3	3		3		3 2 7 6
PH4	3	3 2 6 7	3		3 2 6 7
PH27	3	7 4 3 2 6			3 2 6 7
PH28	7	6 7			5 7 4 3 6
PH31	3 7	7 3 6 4			3 4 6 7
PH70	7	2 7			2 6 7 4
PH5		3 6			4 7 3
PH6		3			3 7 6
PH12		7 3 6			7 3
PH26					4 3
PH45		7			5 3
J57		7			7
J58		7			7

5.3 Small scale fermentation and mass spectrometry analysis of the isolates suggests soil-dependent activity profile

5.3.1 Disk diffusion assays from crude extracts display activity against Gram-positive bacteria

To gain an insight on the production of metabolites, dereplication was done on isolates grown in two types of liquid media for 7 days (**Table 5.1**). The production (^{PRO}) and non-production media (^{NPRO}) (**section 2.6.1**) shown in **Table 5.2** and **Table 5.3** were selected for each strain considering if activity was observed or not on that media against any of the indicator strains on the plug assay results (**Table 5.1**). Visible differences in production of pigmentation were observed with the same isolate when cultured in different media. ISP3 showed the most detectable pigmentation (**Table 5.2** and **Table 5.3**).

The crude extracts obtained were then used for bioactivity assays against different indicator strains by disk diffusion method using 100 μg per disk (**Figure 5.3**). Differences in bioactivity were observed when strains were inoculated in solid and liquid media. Isolates PH6 and J57 showed the ability to inhibit *B. subtilis* in ISP3 and ISP7 agar plates (**Table 5.1** – 0.3-0.4 and 0.1-0.2 cm, respectively), but lost their activity when cultured in the correspondent liquid media (**Table 5.4**). Furthermore, the crude extract from isolate J58 did not show any activity in liquid ISP7 ^{PRO} but in ISP2 ^{NPRO} yielded a small inhibition zone (0.1 cm) contrary to the previous result in agar (**Table 5.1**). Isolate PH70 produced distinct results in the different media. In ISP2 agar, it showed activity against *B. subtilis* and *S. aureus* (**Table 5.1**). On the other hand, the extract prepared from ISP2 broth ^{PRO} exhibited bioactivity against *E. faecalis* as well as *S. aureus* while no activity was recorded against *B. subtilis*. (**Table 5.4**). Furthermore, isolate PH28 in ISP2 agar ^{NPRO} did not inhibit any of the indicator strains (**Table 5.1**). However, when using ISP2 broth ^{NPRO} activity was observed against *E. faecalis*, *S. aureus* and *B. subtilis* (**Table 5.4**).

Overall, isolates from soils 7 and 8 showed larger inhibition zones in liquid culture than agar (**Table 5.5**). Interestingly, no activity was observed against *S. aureus* from soil 8 isolates and also no activity was observed for PH2, PH3 or PH4 against

A. baumannii or *E. coli*, which differed from the previous results when using plug assays. All crude extracts from the isolates were then analysed for both non-production and production media using mass spectrometry to understand bioactivity differences and diversity.

Table 5.2 Liquid cultures obtained after 7 days of growth in production and non-production medium of soil 7 isolates.

Production and non-production media were used depending if activity or no activity was observed in the plug assay (**Table 5.1**).

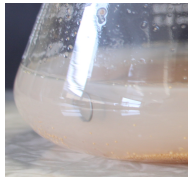
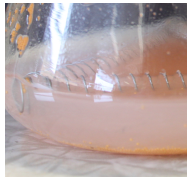
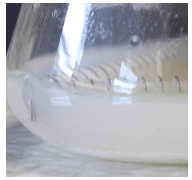
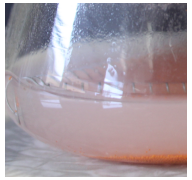
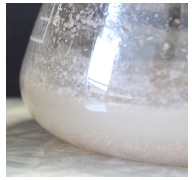
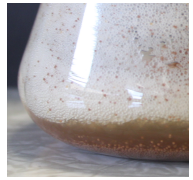
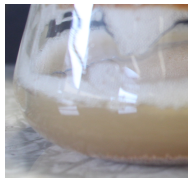
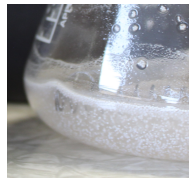
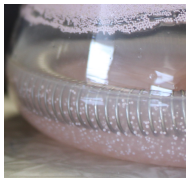
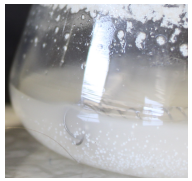
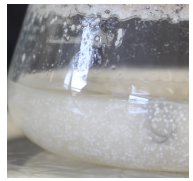
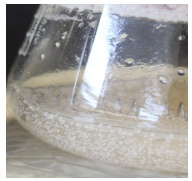
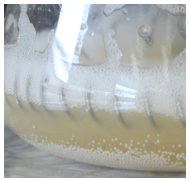
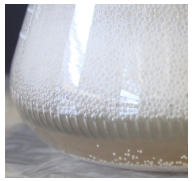
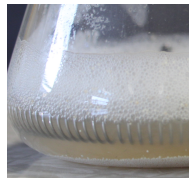
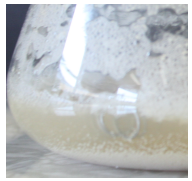
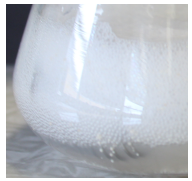
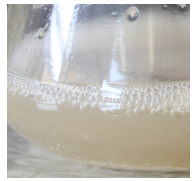
Isolates	PH2	PH3	PH4	PH27	PH28	PH31	PH70
Production medium							
Non-production medium							

Table 5.3 Liquid cultures obtained after 7 days of growth in production and non-production medium of soil 8 isolates.

Production and non-production media were used depending if activity or no activity was observed in the plug assay (**Table 5.1**).

Isolates	PH5	PH6	PH12	PH26	PH45	J57	J58
Production medium							
Non-production medium							

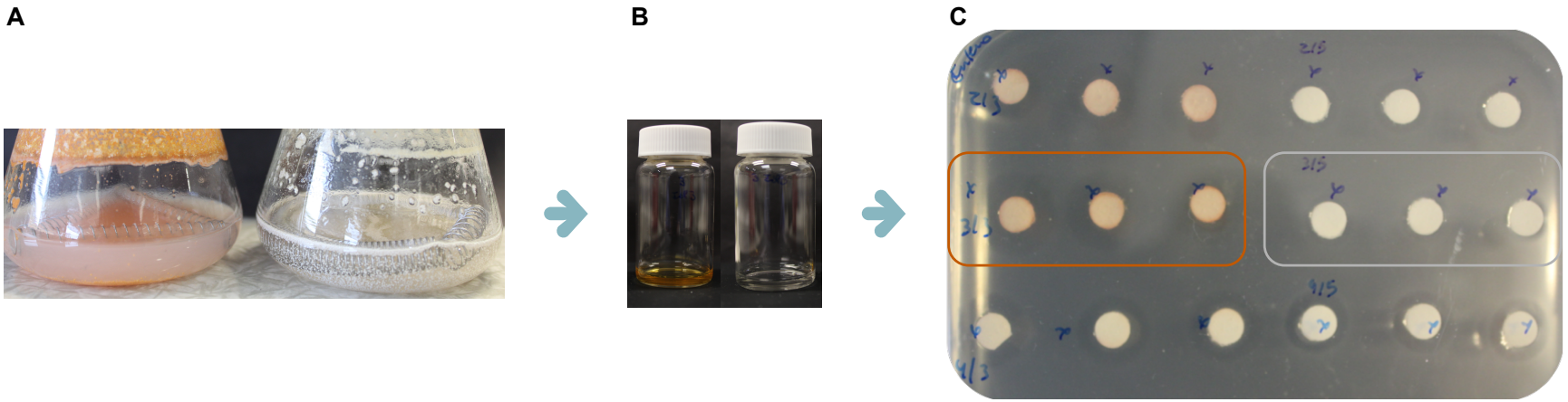


Figure 5.3 Different phases of analysis performed to understand the production of metabolites by the isolates.

A. Liquid culture of isolate PH3 after growth for 7 days in ISP3 (left) and ISP5 (right); **B.** Extracts obtained after the extraction using ethyl acetate, ISP3 (left) and ISP5 (right) and **C.** Bioactivity assay of extracts (100 μg each disc) against *E. faecalis* from ISP3 (**orange box**) and ISP5 (**grey box**).

Table 5.4 Bioactivity by disk diffusion assay against the indicator strains using 100 μg of each crude extract (only shown inhibited strains) and controls used (*).

Size of inhibition zones (cm): **green** (<0.1); yellow (0.1-0.5); **orange** (0.6-1.0); **light red** (1.1-1.5) and **red** (>1.5). **PRO:** medium where previous bioactivity was identified on the plug assay and **NPRO:** medium where no previous bioactivity was identified against any indicator strain in plug assay. The two distinct species are shown with different colours (soil 7: **black**; soil 8: **blue**).

Isolates and Media	*E. faecalis*	*S. aureus*	*B. subtilis*
PH2 ISP3 – PRO			
PH2 ISP5 – NPRO			
PH3 ISP3 – PRO			
PH3 ISP5 – NPRO			
PH4 ISP3 – PRO			
PH4 ISP5 – NPRO			
PH27 ISP3 – PRO			
PH27 ISP5 – NPRO			
PH28 ISP7 – PRO			
PH28 ISP2 – NPRO			
PH31 ISP3 – PRO			
PH31 ISP2 – NPRO			
PH70 ISP2 – PRO			
PH70 ISP5 – NPRO			
PH5 ISP3 – PRO			
PH5 ISP2 – NPRO			
PH6 ISP3 – PRO			
PH6 ISP2 – NPRO			
PH12 ISP3 – PRO			
PH12 ISP2 – NPRO			
PH26 ISP3 – PRO			
PH26 ISP2 – NPRO			
PH45 ISP3 – PRO			
PH45 ISP2 – NPRO			
J57 ISP7 – PRO			

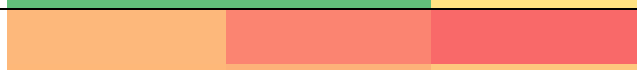
J57 ISP2 – NPRO	
J58 ISP7 – PRO	
J58 ISP2 – NPRO	
* Imipenem (10 μg)	
* Tetracycline (10 μg)	
* Chloramphenicol (5 μg)	

Table 5.5 Variation between bioactivity by plug assay from solid media and disk diffusion assay from liquid broth against the indicator strains.

Larger inhibiton zones diameter observed in liquid broth represented as **green (+)**, same diameter in liquid and solid media represented in **yellow (=)** and larger inhibition zone diameter in solid media represented as **orange (-)**. The variation was classified '+' or '-' when a difference between solid or liquid were ≥ 0.1. Isolates from different soils are shown with different colours (soil 7: **black**; soil 8: **blue**).

Isolates & Media	*E. faecalis*	*S. aureus*	*B. subtilis*
PH2 ISP3 – PRO	-	-	-
PH2 ISP5 – NPRO	+	+	+
PH3 ISP3 – PRO	+	+	-
PH3 ISP5 – NPRO	+	+	+
PH4 ISP3 – PRO	+	-	-
PH4 ISP5 – NPRO	+	+	+
PH27 ISP3 – PRO	+	-	-
PH27 ISP5 – NPRO	+	+	+
PH28 ISP7 – PRO	-	-	-
PH28 ISP2 – NPRO	+	+	+
PH31 ISP3 – PRO	+	-	-
PH31 ISP2 – NPRO	+	+	+
PH70 ISP2 – PRO	+	-	-
PH70 ISP5 – NPRO	+	+	=
PH5 ISP3 – PRO	=	-	-
PH5 ISP2 – NPRO	=	=	+
PH6 ISP3 – PRO	=	-	-
PH6 ISP2 – NPRO	=	=	=
PH12 ISP3 – PRO	=	-	-
PH12 ISP2 – NPRO	=	=	+
PH26 ISP3 – PRO	=	=	-
PH26 ISP2 – NPRO	=	=	+
PH45 ISP3 – PRO	=	-	-
PH45 ISP2 – NPRO	=	=	+
J57 ISP7 – PRO	=	-	-
J57 ISP2 – NPRO	=	=	=
J58 ISP7 – PRO	=	-	-
J58 ISP2 – NPRO	=	=	+

5.3.2 Metabolomics analysis of the isolates indicates potential for specialised metabolites production

The crude extracts obtained for both non-production ^{NPRO} (no activity observed in plug assays) and production ^{PRO} (activity observed in plug assays) broths (**Table 5.2** and **Table 5.3**) were analysed by LC-MS and the resulting spectral data was subjected to multivariate analysis by using PCA and OPLS models to pinpoint the potential metabolites responsible for the bioactivity observed in the disk diffusion assay (**Table 5.4**). PCA plots were generated from four components using Pareto scaling, a R^2X (coefficient of determination) value of 0.79 was achieved but Q^2 of 0.34, which specified that predictability of the model was not so good where Q^2 must be at least > 0.6 (Eriksson *et al.*, 2004). As done under the SIMCA software by Umetrics, the number of components was determined upon the optimum Q^2 as obtained by the summary of fit graph that follows a model complexity curve. The goodness of fit (R^2) follows a linear curve while the goodness of prediction (Q^2) is sigmoidal. Here, in this model, the Q^2 do not increase more than 4 components, the point at which the R^2 and Q^2 curve diverges. The presence of outliers and diversity on chemical profiles of fractions caused the poor predictability of the model.

PCA enabled the visualisation of outliers and the samples clustered according to their chemical similarities and differences in various media (**Figure 5.4**). Outliers were identified from the scores plot as ISP2 isolates PH6 and PH70 (**Figure 5.4, panel A**). Most of the ISP2 samples dominated the top and bottom left quadrant while samples in ISP3, ISP5 and ISP7 clustered in the middle of the right quadrant. The loadings plot (**Figure 5.4, panel B**) represented the molecular ion masses corresponding to the same location and quadrant as the samples found in the scores plot (**Figure 5.4, panel A**). Media ISP3 and ISP7 produce a higher density of higher molecular weight (MW) compounds (**Figure 5.4, panel B** – green (*m/z* 1,436) to red dots (*m/z* 2,000)) than ISP2 (**Figure 5.4, panel B** – blue (*m/z* 120) to green dots (*m/z* 872). The latter medium could comprise more promising bioactive compounds for drug discovery following Lipinski's rule of five, where an active drug should have a molecular mass of less than 500 Da (Lipinski, 2004).

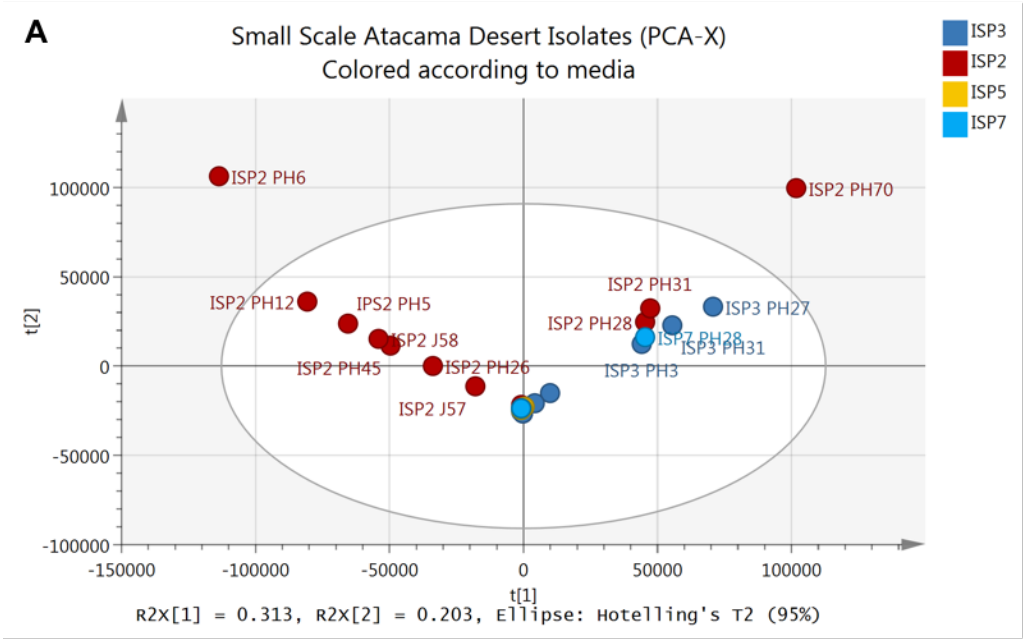
OPLS-DA scores, loadings scatter and S–plots were generated and validated by permutation test by grouping the fractions into active or inactive to identify the

metabolites responsible for the bioactivity. OPLS-DA scores, loadings scatter and S–plots were generated and validated by permutation test by grouping the fractions into active or inactive to identify the metabolites correlated with the bioactivity. OPLS-DA for the bioactivity of all isolates from soils 7 and 8 against *E. faecalis* and *S. aureus* (**Figure 5.5**) afforded an R^2Y of 0.503 while the Q^2 was only 0.45. The poor model was due to the presence of the outlying sample ISP2 PH6. Since ISP2 PH6 was inactive in the bioassay, its removal from the dataset improved the model to indicate a more linear trend in the occurrence of the various metabolites (**Figure 5.6**). The decision to remove the outlier was due to the inactivity of the extract, considering it to be a feature of non-importance to the dataset since the objective of the model is to pinpoint for the bioactive features of extracts. To further produce a normally distributed variance, as the dataset was skewed due to the unequal numbers of observations between active and inactive extracts, the dataset was log transformed affording a good model with R^2Y and Q^2 values of 0.99 and 0.93, respectively, while passing the permutation test with a Y-intercept of less than zero (**Figure 5.6, panel D**). The permutation test assessed that the current OPLS-DA model just fits the training set well. The validation done by the SIMCA algorithm is to compare the goodness of fit (R2 and Q2) of the original model with the goodness of fit of several models based on data where the order of the Y-observations has been randomly permuted, while the X-matrix has been kept intact. A difference between R2 and Q2 must not be >0.3 to classify against overfitting while the permutation test also strongly indicates that the model is valid when Q2Y-intercept is below zero. The active quadrant (**Figure 5.6, panel A** – top and bottom left quadrant) was selected as they represented unique chemistry with potential bioactivity. Similarities in chemical profiles were found with compounds sharing similar higher MW metabolites between 600 and 1500 Da in the active extract. The loadings S-plot (**Figure 5.6, panel B-C**) indicated that the bioactive extracts consisted mostly of metabolites with higher MWs greater than 1300 Da. The “end components” of the S-plot (**Figure 5.6, panel B**) hypothetically specified ($P<0.01$) the target metabolites correlated with the bioactivity against *E. faecalis* and *S. aureus* (**Table 5.6**).

OPLS-DA for the bioactivity of sample isolates from soils 7 and 8 against *B. subtilis* (**Figure 5.7**) gave an R^2Y of 0.92 and Q^2 of 0.354. The predictive ability of the model was poor due again to the presence of inactive outlying samples ISP2 PH6 and ISP2 PH70, the same outliers observed for the OPLS-DA of the bioactivity against *E.*

faecalis and *S. aureus* (**Figure 5.5, panel A**). Since the outlying samples were rendered inactive, while they rendered higher internodal variation of 27.8% than the group separation of only 6.81%, these samples were then excluded from the dataset then log transformed to attain a good model with R^2Y of 0.96 and Q^2 of 0.68 (**Figure 5.8**). As disclosed by the scores plot, the inactive extracts clustered with the media blank (**Figure 5.8, panel A**). The loadings S-plot exhibited a higher density of higher MWs metabolites found in the active samples (**Figure 5.8, panel C**). The "end components" of the S-plot (**Figure 5.8, panel C**) specified the target metabolites potentially correlated with the bioactivity against *B. subtilis*, which consisted of actinomycin congeners (**Table 5.6**) indicated by ion peaks at *m/z* 1303.54, 1304.54, 1305.54, 1305.56, 1306.54, 1307.56, and 1308.56. Permutation test afforded a Y-intercept of < 0 while the difference between R^2Y and Q^2 was 0.227, which was less than 0.3 indicating that the model was not over-fitted (**Figure 5.8, panel D**).

Additionally, a heatmap was created to assess the diversity and abundance of metabolites in the crude extracts using the program R with its gplots tool. Most of the samples produced a range of compounds as indicated by the number and intensity of blue bands (**Figure 5.9**). Some samples (ISP2: PH70 and PH6; **blue box**) had higher diversity for low and high molecular weight metabolites. In contrast, there were samples that had only low molecular weight (ISP2: PH5, PH12, PH26, PH45, J57 and J58; **yellow box**) or just high molecular weight metabolites (ISP7: PH28 and ISP3: PH31; **red box**).



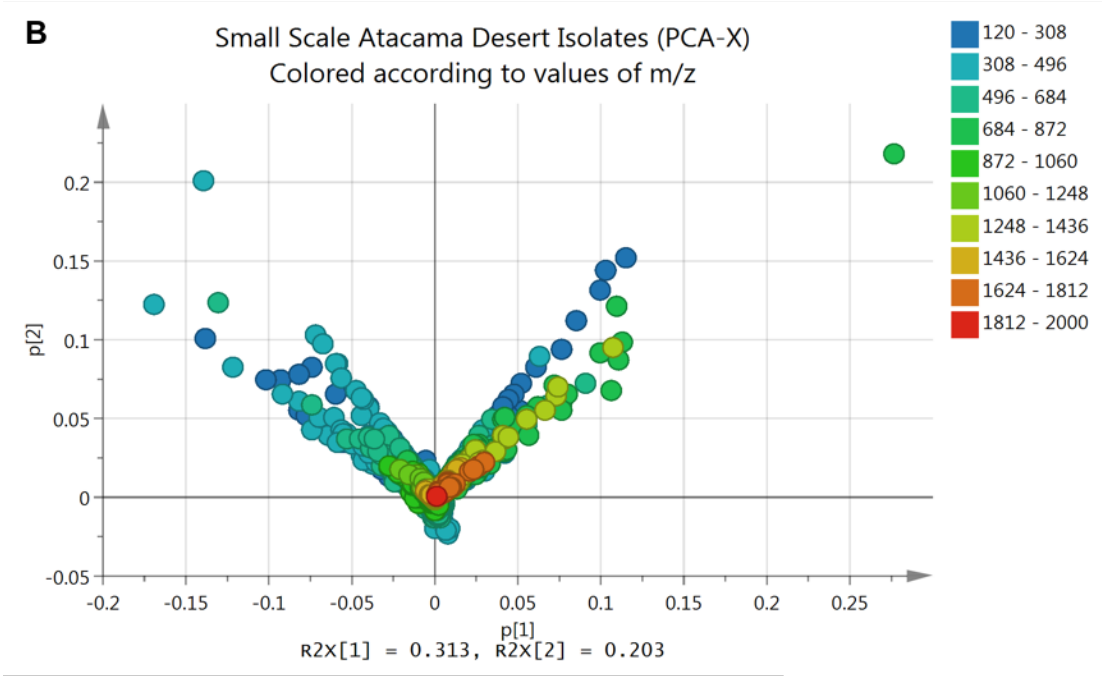


Figure 5.4 Principal component analysis (PCA) plot for all samples from soil 7 and soil 8 isolates.

A. PCA-X score plot of samples coloured by media used. **B.** Loadings plot acquired from the PCA-X model correlated with the score plot.

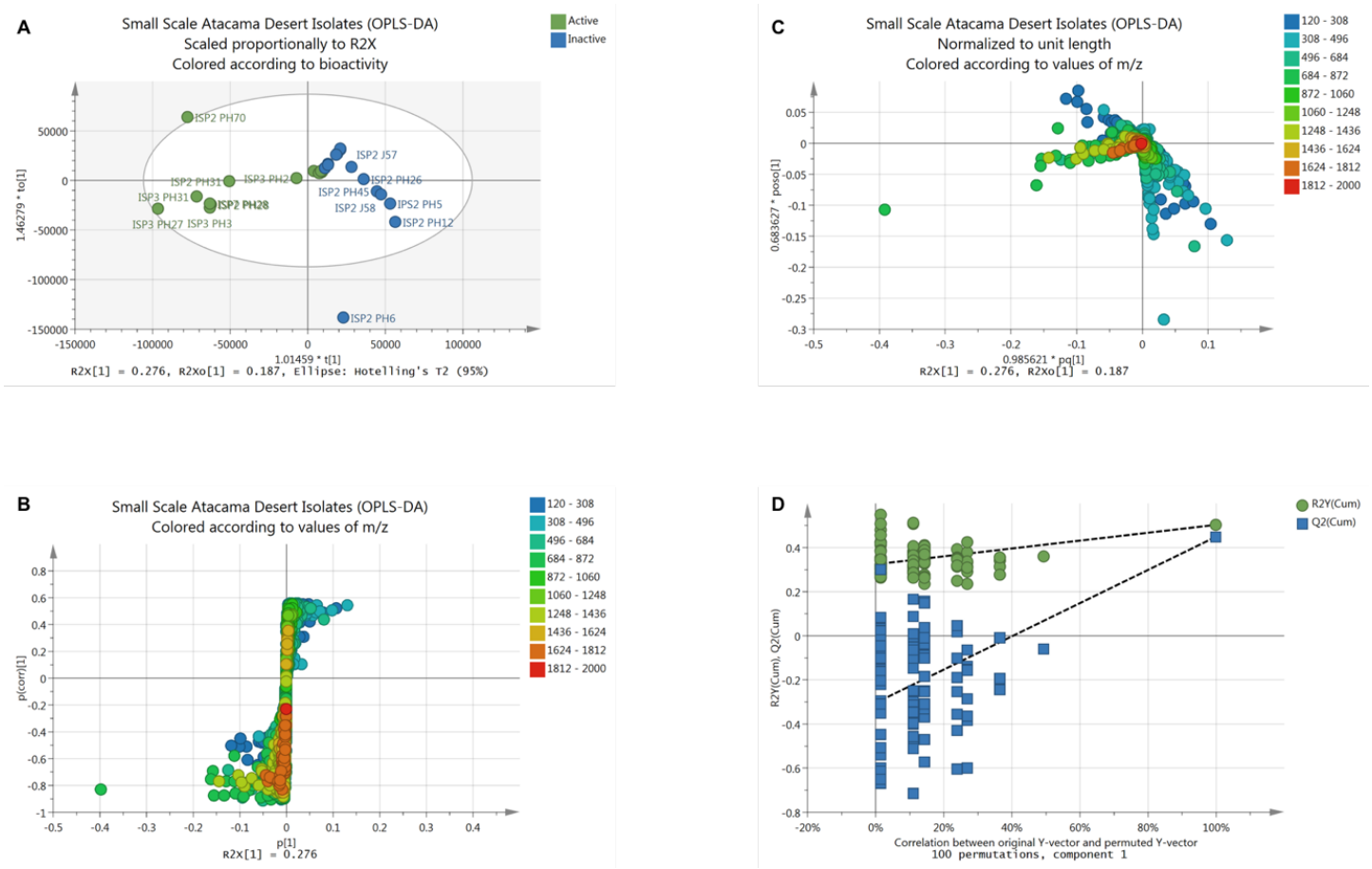


Figure 5.5 OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *E. faecalis* and *S. aureus*.

A. Samples classed as active or inactive on the disk diffusion assay. **B.** Top and bottom left quadrant of the S-plot showing *m/z* corresponding to the active fractions in A. **C.** Loadings plot acquired from the OPLS-DA model. **D.** Permutation test to validate OPLS model where Q^2Y intercept must be less than zero.

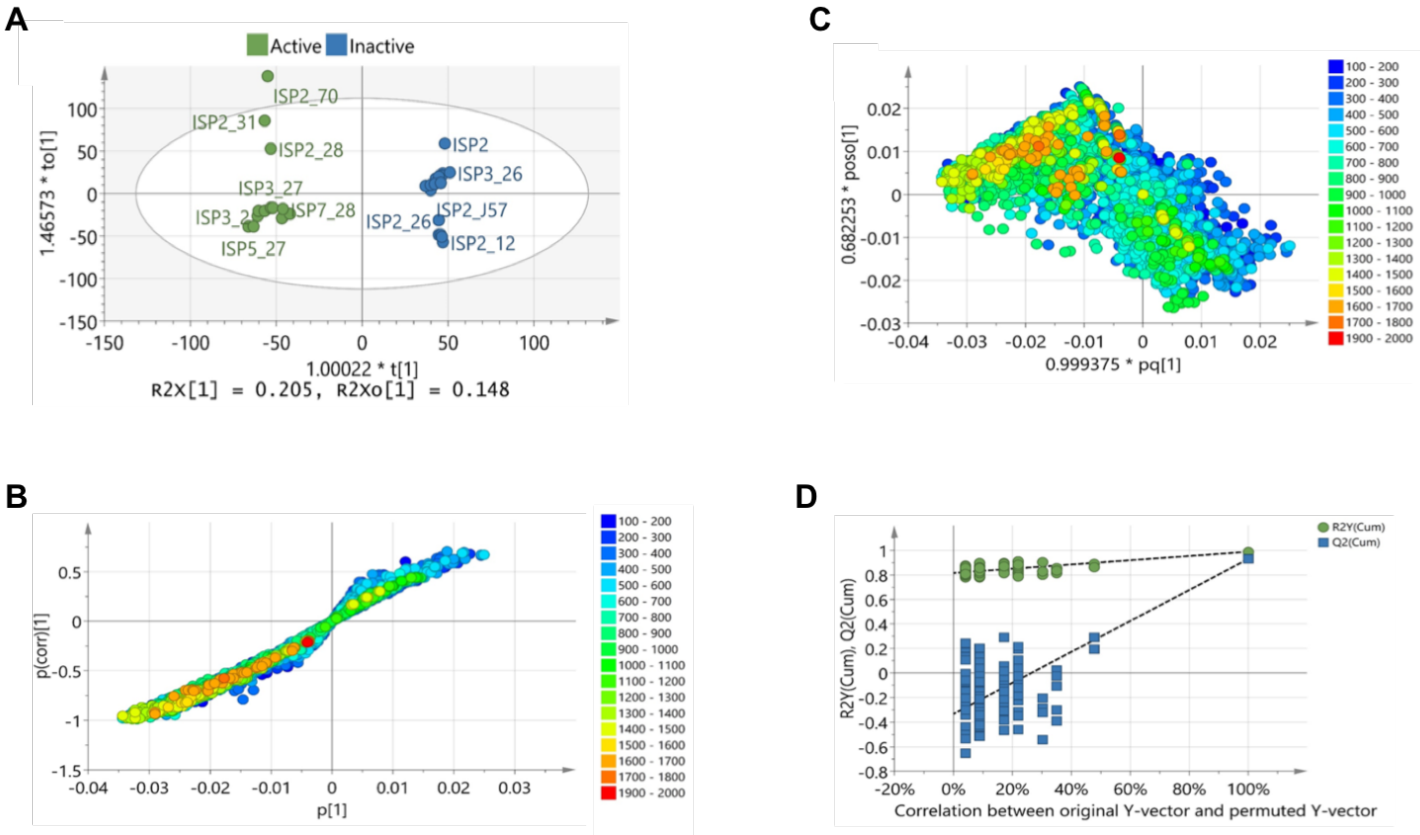


Figure 5.6 Log transformed OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *E. faecalis* and *S. aureus* excluding outlying inactive extract ISP2_PH6.

A. Samples classed as active or inactive on the disk diffusion assay. **B.** Top and bottom left quadrant of the S-plot showing *m/z* corresponding to the active fractions in A. **C.** Loadings plot acquired from the OPLS-DA model. **D.** Permutation test to validate OPLS model where Q^2Y intercept must be less than zero.

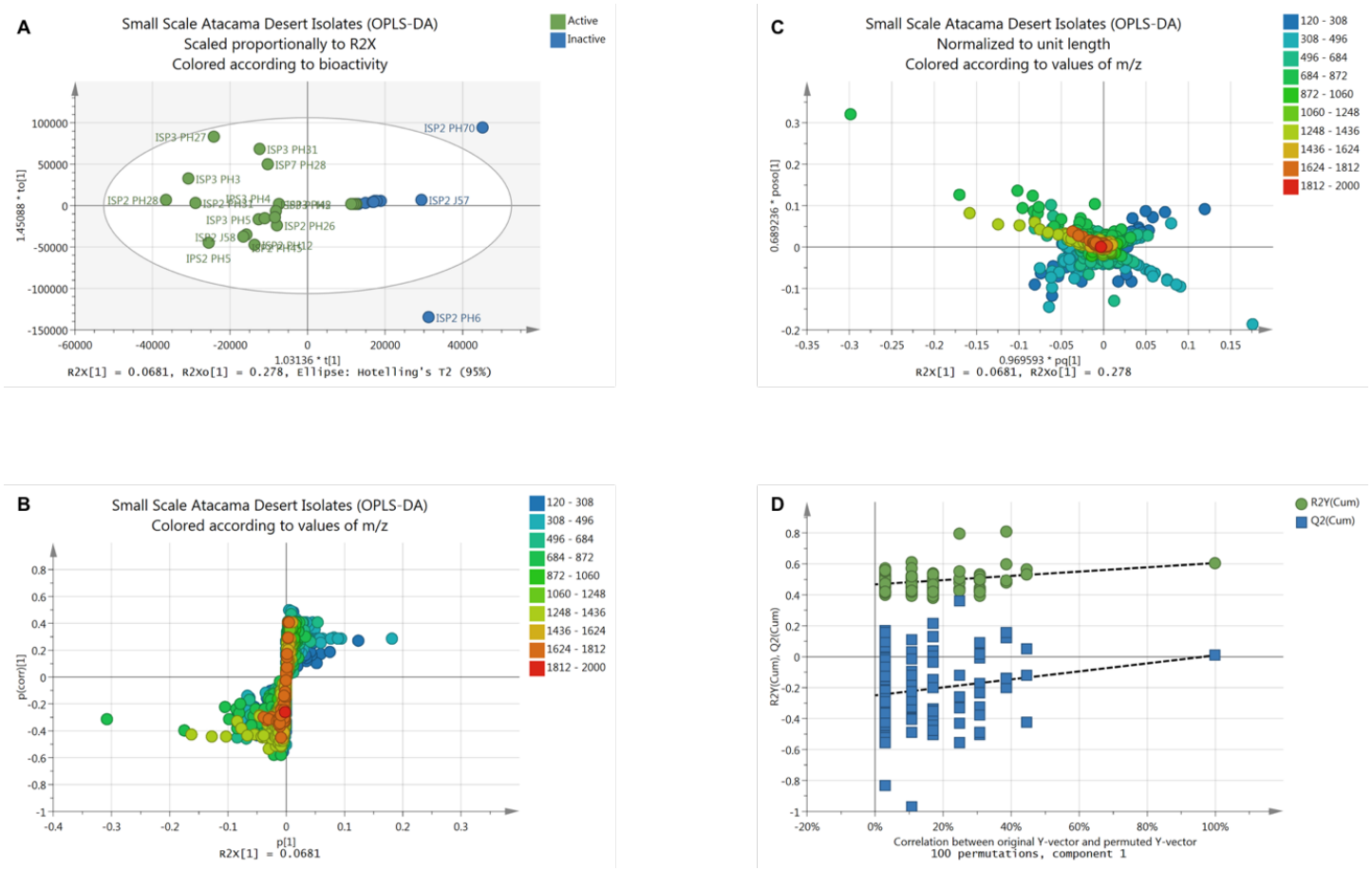


Figure 5.7 OPLS-DA SIMCA plots of bioactivity of samples from soils 7 and 8 isolates against *B. subtilis*.

A. Samples classed as active or inactive on the disk diffusion assay. **B.** Top and bottom left quadrant of the S-plot showing m/z corresponding to the active fractions in A. **C.** Loadings plot acquired from the OPLS-DA model. **D.** Permutation test to validate OPLS model where Q^2Y intercept must be less than zero.

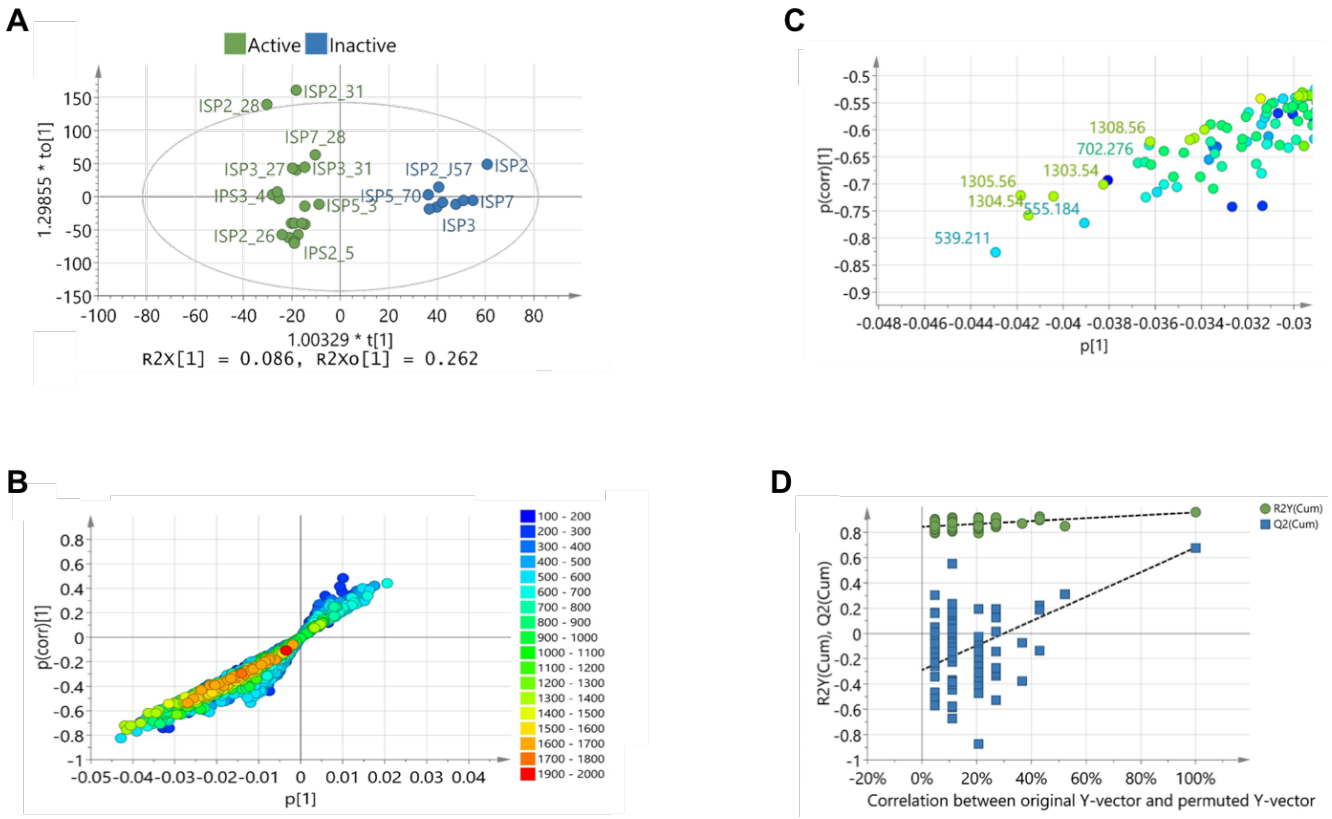


Figure 5.8 OPLS-DA SIMCA analysis plots overview for all samples from soil 7 and soil 8 isolates against *B. subtilis* excluding outlying inactive extracts.

A. Samples coloured by active or inactive considering the disk diffusion assay. **B.** Top and bottom left quadrant of the S-plot showing *m/z* corresponding to the identified active fraction in A. **C.** Expansion of the "end point components" for the active side of the S-plot in B. **D.** Permutation tests for the samples where Q^2Y intercept is less than zero.

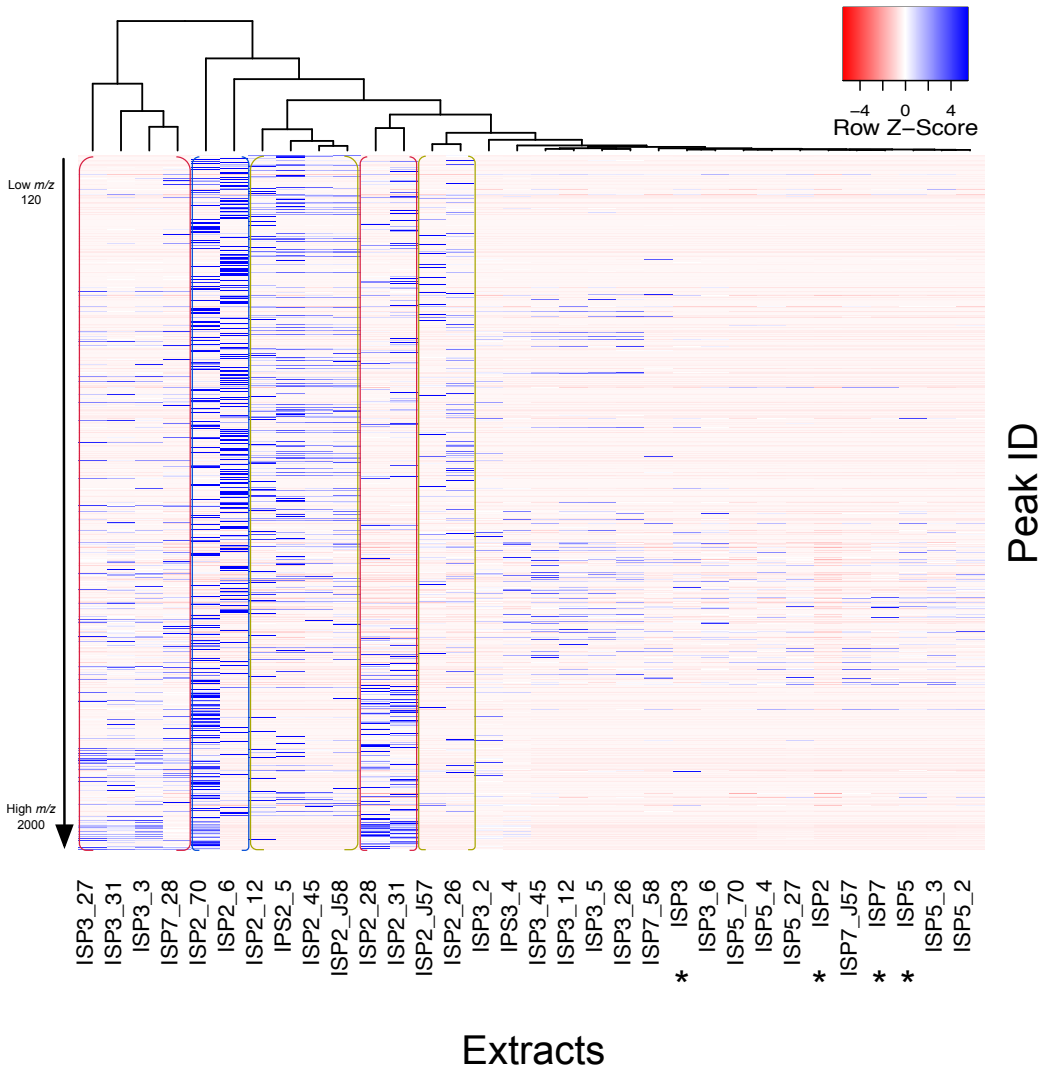


Figure 5.9 Heat map analysis showing the diversity of the metabolomes. From top to bottom, lines correspond to low to high *m/z* ratios.

Blue bands indicate richness of metabolome and the media controls were indicated with an asterisk. Samples in blue box (ISP2: PH70 and PH6) showed higher diversity for low and high molecular weight metabolites. Samples in yellow box had higher diversity only in low molecular weight (ISP2: PH5, PH12, PH26, PH45, J57 and J58) and samples in red box for high molecular weight metabolites (ISP7: PH28 and ISP3: PH31). Standard score (Z-score) indicates the relative abundance of the individual features occurring in the respective extracts where blue is the increase of the abundance or occurrence of a metabolite, while red is the decrease of the abundance of the metabolites.

The compounds identified in the active quadrant of the S-plots, mostly actinomycin congeners (**Figure 5.8, panel B**) as indicated by *m/z* 1200-1500 and thought to be correlated with the bioactivity observed were selected for dereplication and comparison with DNP databases (Buckingham, 2001). The extracts were later fractionated to confirm the occurrence of actinomycins by 1D and 2D NMR. The compound hits active against *E. faecalis*, *S. aureus* and *B. subtilis* are listed in **Table 5.6**. The significant ($P < 0.01$) bioactive target metabolites were listed according to their molecular weights. Most of the active metabolites were identified from the positive ionisation mode (n=22) while the negative mode identified only three metabolites. Furthermore, some of the metabolites were present in all isolates from soil 7, e.g. P_5738 (*m/z* 1308.559194) corresponding to an unknown metabolite and P_5729 (*m/z* 826.3284789) corresponding to aclacinomycin M and cinerubin B, only identified in these isolates (**Table 5.6**). Actinomycin Y_1 was identified in both ionisation modes. Aklavinone and enshumycin were only identified with the negative mode. Interestingly, the predominant masses identified were consistent with known polyketides or non-ribosomal peptides, e.g. griseusin B, cinerubin B and aclacinomycin M and actinomycin Y_1/Y_5, respectively with ion peaks *m/z* ranging between 700-1500 Da (**Table 5.6**). Additionally, most of these identified metabolites have been described for their antitumor activity particularly aklavinone, cinerubin B and aclacinomycin M (Oki *et al.*, 1979). In addition, griseusin B, actinomycin Y_1/Y_5 and enshumycin were earlier reported for their bioactivity against Gram-positive bacteria (Tsuji *et al.*, 1976; Bitzer, Gesheva & Zeeck, 2006; Bycroft & Payne, 2013; Bitzer *et al.*, 2009). Five metabolites comprising of low (222.043 and 744.286) and high (1307.560, 1308.560 and 1426.630) molecular weights did not show any hit from the database and thus, considered as unknowns or as new natural products (**Table 5.6**). Overall, the isolates showed activity primarily against Gram-positive bacteria. As deduced from the metabolomics analysis, ion peaks at *m/z* 1306.550 and 1307.552 afforded by PH28 in ISP2 (with peak areas of 5E+7 to 7E+7) showed a related predicted molecular formula to actinomycin Y_1 and were selected for further target isolation work, characterisation and scale-up fermentation.

In conclusion, a high diversity of metabolomes was exhibited by the collected isolates from both soils of the Atacama Desert that could potentially encode bioactive compounds (**Figure 5.9**). Furthermore, dereplication analysis of the crude extracts' activity showed that species isolate from soil 7 were the most promising isolates with

several known bioactive metabolites identified (**Table 5.6**). Additionally, the unique biosynthetic gene cluster cinerubin B identified only in isolates from soil 7 was detected in variable high peak areas (**Table 5.6**). This analysis confirms the remarkable ability of these isolates to encode a wide range of bioactive metabolites.

Table 5.6 Dereplication table containing the outlaying metabolites identified in the isolates crude extract that showed activity against *E. faecalis*, *S. aureus* and *B. subtilis*.

Listed metabolites only included those with p-values < 0.01 and were considered significant while hits found from the DNP database (Buckingham, 2001) were filtered for *Streptomyces* strains. Duplicated hits at different retention times were not possible to achieve the "true" identification of the compounds due to absence of MS/MS data.

Mzmine ID	*m/z*	RT (min)	MW (Da)	Formula	Name	Source
P_5754	223.051	11.48	222.043	$C_{13}H_6N_2O_2$	no hits	-
N_3532	411.109	16.06	412.116	$C_{22}H_{20}O_8$	aklavinone	*Streptomyces* spp.
P_5755	586.228	12.94	585.221	$C_{30}H_{35}NO_{11}$	griseusin B, pyrromycin	*S. griseus*, *S. capoamus* and *S. violaceus*
P_5752	586.228	15.64	585.221	$C_{30}H_{35}NO_{11}$	griseusin B, pyrromycin	*S. griseus*, *S. capoamus* and *S. violaceus*
P_5775	702.276	12.02	701.269	$C_{35}H_{43}NO_{14}$	antibiotic CG 17B, respinomycin A1,	*S. galilaeus* KE 303 and *S. xanthocidicus*
P_5739	716.291	12.78	715.284	$C_{36}H_{45}NO_{14}$	antibiotic SN 706, musettamycin, respinomycin A1	*Streptomyces* spp.
P_5804	744.286		743.279	$C_{38}H_{41}N_5O_{11}$ $C_{33}H_{41}N_7O_{13}$ $C_{37}H_{45}NO_{15}$ $C_{26}H_{45}N_7O_{18}$	no hits	-
P_5800	810.333	15.15	809.326	$C_{42}H_{51}NO_{15}$	aclacinomycin M, cinerubin B	*S. galilaeus*, *S. cinereoruber, S. eurythermus* and *S. lavendofoliae*

P_5759	810.333	17.28	809.326	$C_{42}H_{51}NO_{15}$	aclacinomycin M, cinerubin B	*S. galilaeus*, *S. cinereoruber, S. eurythermus* and *S. lavendofoliae*
P_5892	812.311	13.75	811.304	$C_{41}H_{49}NO_{16}$	auramycin B, cinerubin B, taurimycin B	*S. galilaeus*, *S. melanogenes* and *S. tauricus*
P_5756	812.313	14.68	811.305	$C_{41}H_{49}NO_{16}$	auramycin B, cinerubin B, taurimycin B	*S. galilaeus*, *S. melanogenes* and *S. tauricus*
P_5733	814.365	15.72	813.358	$C_{42}H_{55}NO_{15}$	aclacinomycin M	*Streptomyces* spp.
N_3520	824.315	14.68	825.322	$C_{38}H_{49}N_8O_{11}S$	enshumycin	*S. enshuensis*
P_5750	826.328	14.73	825.321	$C_{42}H_{51}NO_{16}$	aclacinomycin M, cinerubin B	*Streptomyces* spp.
P_5729	826.328	15.70	825.321	$C_{42}H_{51}NO_{16}$	aclacinomycin M, cinerubin B	*Streptomyces* spp.
P_5736	828.343	14.68	827.336	$C_{42}H_{53}NO_{16}$	aclacinomycin G_2, streptovaricin C	*S. galilaeus*, *S. violaceus* and *S. spectabilis*
P_5741	830.360	13.84	829.352	$C_{42}H_{55}NO_{16}$	aclacinomycin M	*Streptomyces* spp.
P_5742	844.337	15.31	843.329	$C_{42}H_{53}NO_{17}$	aclacinomycin M	*S. galilaeus* KE 303
P_5831	844.339	13.83	843.331	$C_{42}H_{53}NO_{17}$	aclacinomycin M	*S. galilaeus* KE 303
P_5732	1269.580	14.27	1268.573	$C_{61}H_{80}N_{12}O_{18}$	actinomycin Y_5	*S.* sp. Gu-GS12
N_3517	1303.540	18.55	1304.549	$C_{61}H_{81}ClN_{12}O_{18}$	actinomycin Y_1	*S.* sp. Gu-GS12
P_5730	1305.560	18.56	1304.550	$C_{61}H_{81}ClN_{12}O_{18}$	actinomycin Y_1	*S.* sp. Gu-GS12
P_6954	1307.560	18.58	1306.550	$C_{62}H_{83}ClN_{10}O_{19}$	a putative new actinomycin congener	-
				$C_{63}H_{78}N_{12}O_{19}$	a putative new actinomycin congener	
P_5738	1308.560	18.55	1307.552	$C_{61}H_{82}ClN_{11}O_{19}$	a putative new actinomycin congener	-
P_5748	1426.630	18.53	1425.623	$C_{71}H_{87}N_{13}O_{19}$	a putative new actinomycin congener	-

	$C_{70}H_{92}ClN_{11}O_{19}$	a putative new actinomycin congener

5.3.3 Genomics and phylogenetic analysis of biosynthetic gene clusters shows cinerubin as a soil-specific BGC

To complement the metabolomics analysis and to understand if common biosynthetic gene clusters correspond to potential phylogenetic relatedness, the 16S rRNA of the matching *Streptomyces* spp. (**Table 5.6**) were used to construct a maximum likelihood phylogenetic tree (**Figure 5.10**). Subsequently, analysis of selected biosynthetic gene clusters was performed. The Atacama isolates clustered closely together in one clade with the closest neighbour *S. phaeoluteigriseus*. Overall, no relationship could be drawn from the phylogenetic tree in terms of correlation between metabolites produced and phylogenetic distance. However, *S. galilaeus*, the closest species to the Atacama isolates clade, is a known producer of metabolites such as aclacinomycin and cinerubin (**Figure 5.10**; **see section 5.3.2**). Subsequently, a relationship analysis of the biosynthetic gene clusters identified in the isolates with biosynthetic gene clusters of known function deposited in MIBiG (Kautsar *et al.*, 2019) was performed using BiG-SCAPE. Interestingly, the biosynthetic gene clusters for cinerubin B were only identified in species from soil 7 suggesting it as a potential species-specific gene cluster (**Table 5.7**). The constructed relationship network (**Figure 5.11**) showed the clustering of cinerubin B and aclacinomycin M from *Streptomyces* sp. SPB74 and *S. galilaeus*, respectively, along with the Atacama isolates. However, genetic comparison of the BGCs revealed differences between these strains with the emergence of three different types of clusters (**Figure 5.12**). One group consisted of PH4 and PH28 that shared the same gene features as the BGC of cinerubin B deposited the in MIBiG repository. The other group comprising of PH3, PH27, PH31 and PH70 showed a longer biosynthetic gene cluster with more gene features predicted (**Figure 5.12, panel A**). However, PH2 was classified to a different family (average distance of 0.47) and lacked some genes for the cinerubin BGC but shared other gene features from the rest of the isolates (**Figure 5.12, panel B**). Nevertheless, cinerubin B was found in the MS spectrum of crude extracts of soil 7 isolate that were active against *E. faecalis* and *S. aureus* (**Figure 5.13)** with a peak eluting at ~15 min.

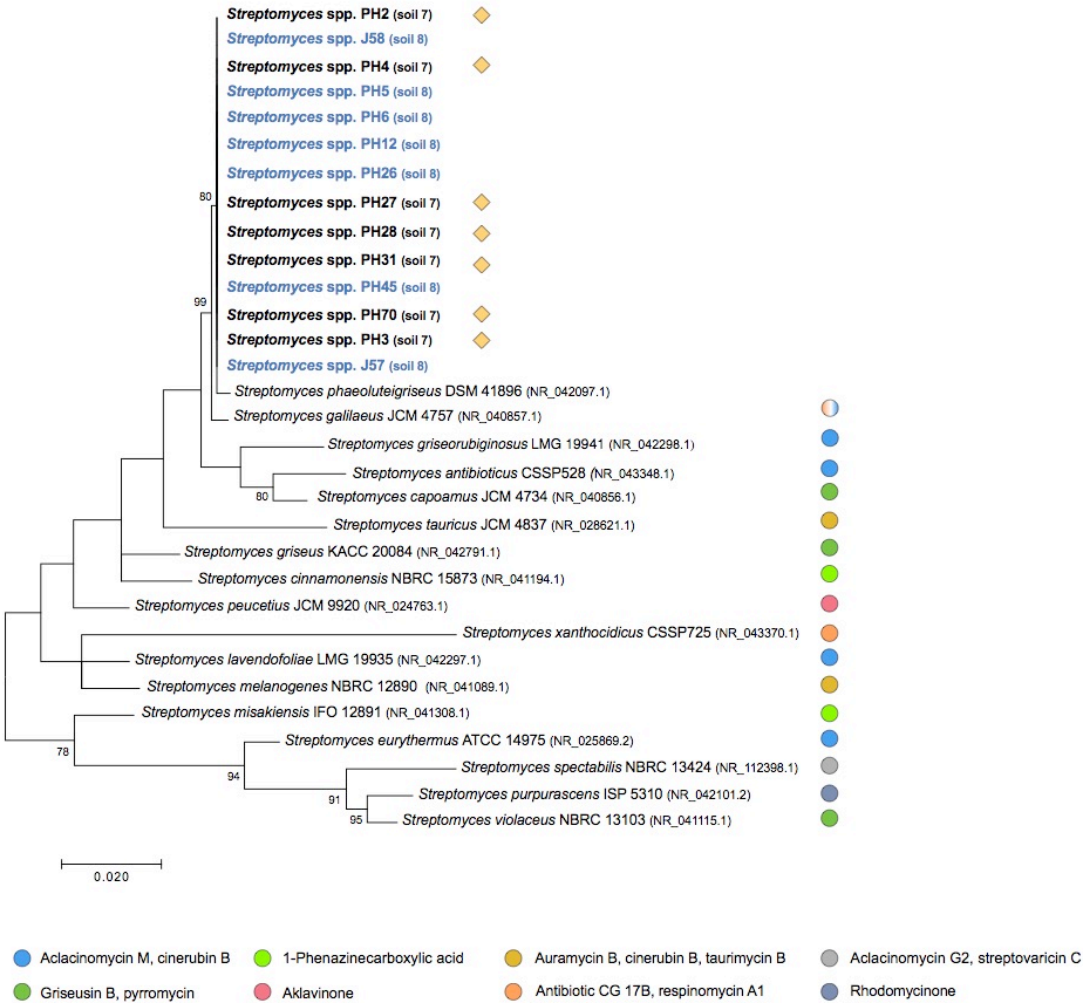


Figure 5.10 Maximum Likelihood showing 16S rRNA gene sequences relationships between the isolates from both soils (soil 7: black; soil 8: blue) and strains with hits obtained by dereplication and SIMCA analysis.

For each *Streptomyces* species the corresponding hit obtained from SIMCA analysis is represented by a coloured circle and for the Atacama *Streptomyces* species a coloured diamond represents the prediction of a cinerubin B biosynthetic gene cluster by antiSMASH. Clustal X (2.0) was used to align the sequences and the phylogenetic tree constructed using model Tamura 3-parameter (T92+G+I) and 1,000 bootstraps (only values >75% shown) in MEGA7.

Table 5.7 List of biosynthetic gene clusters from MIBiG repository and identified in MS dereplication for BGC similarity search against soil 7 isolates BGC's using BiG-SCAPE.

Strain	Specialised metabolite	MIBiG	Class
Streptomyces anulatus	actinomycin	BGC0000296	NRP
Amycolatopsis orientalis HCCB10007	vancomycin	BGC0000455	NRP
Streptomyces galilaeus	aclacinomycin	BGC0000192	PK
Streptomyces sp. SPB74	cinerubin	BGC0000212	PK
Streptomyces griseus	griseusin	BGC0000231	PK
PH2	-	-	
PH3	-	-	
PH4	-	-	
PH27	-	-	
PH28	-	-	
PH31	-	-	
PH70	-	-	

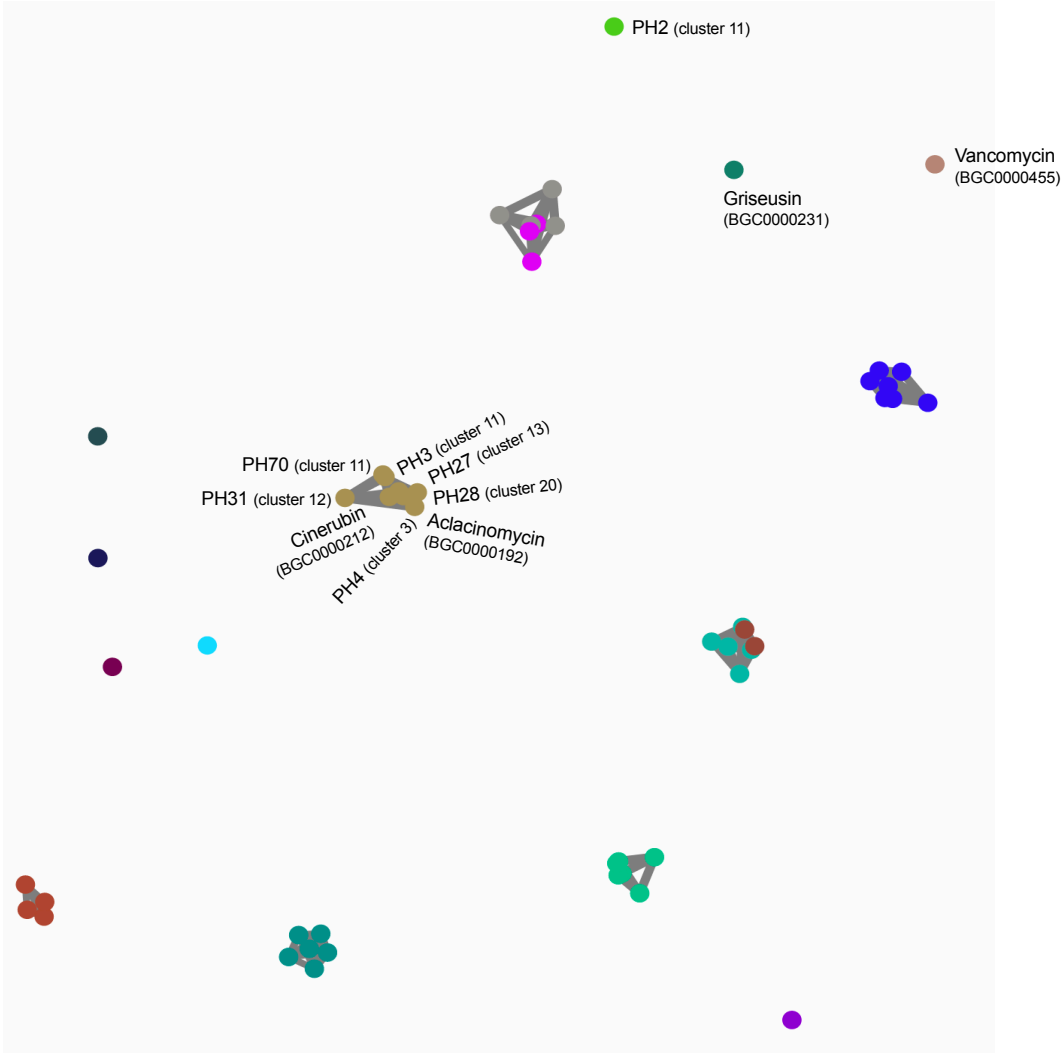


Figure 5.11 PKSother network visualisation of the relationship and similarity of predicted cinerubin B biosynthetic gene clusters from soil 7 isolates.

Predicted PH2 cinerubin B cluster shows no relationship with any gene clusters of the soil 7 isolates biosynthetic gene clusters, that cluster in one group with the reference cinerubin B gene cluster. Actinomycin was identified as a singleton in the NRPS network. BiG-SCAPE was used to construct the network (**section 2.4.6**).

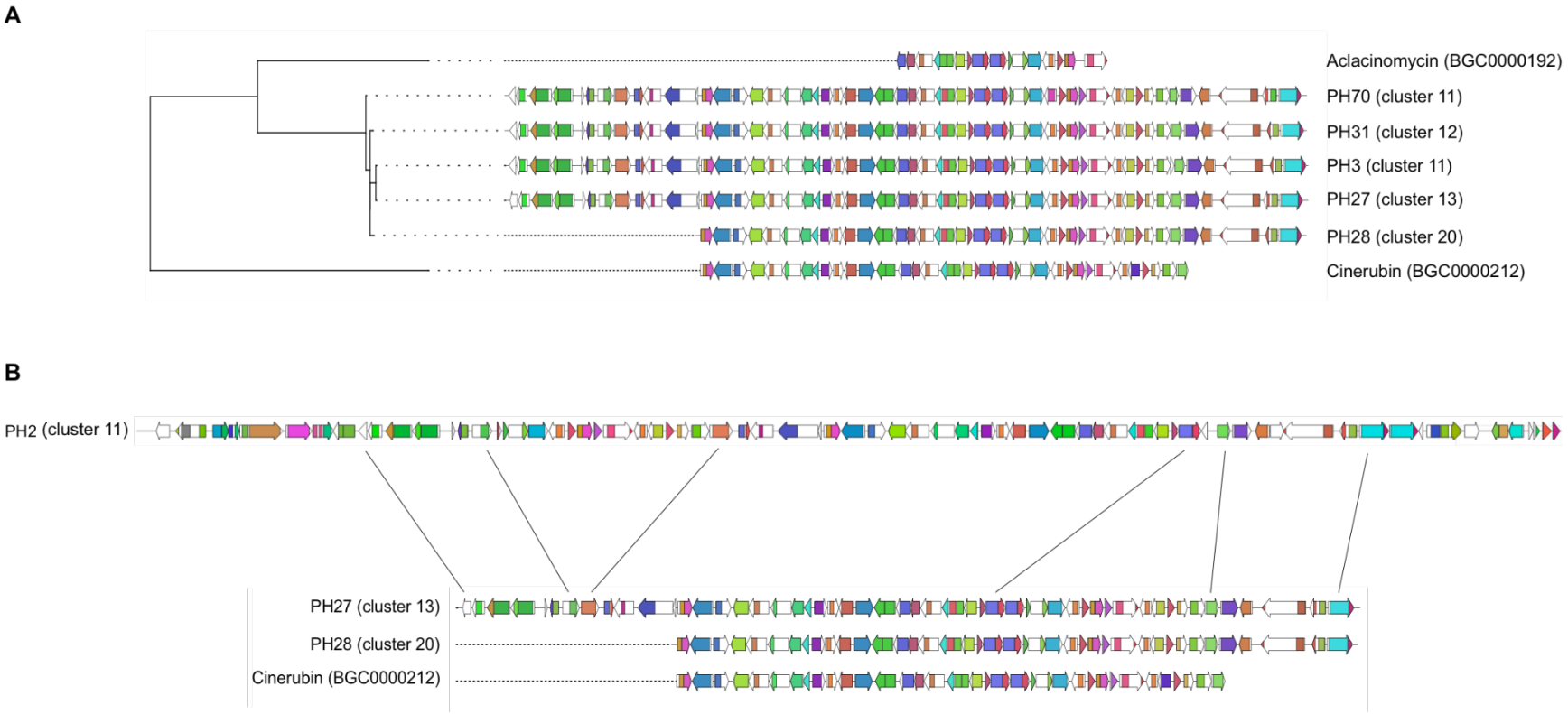
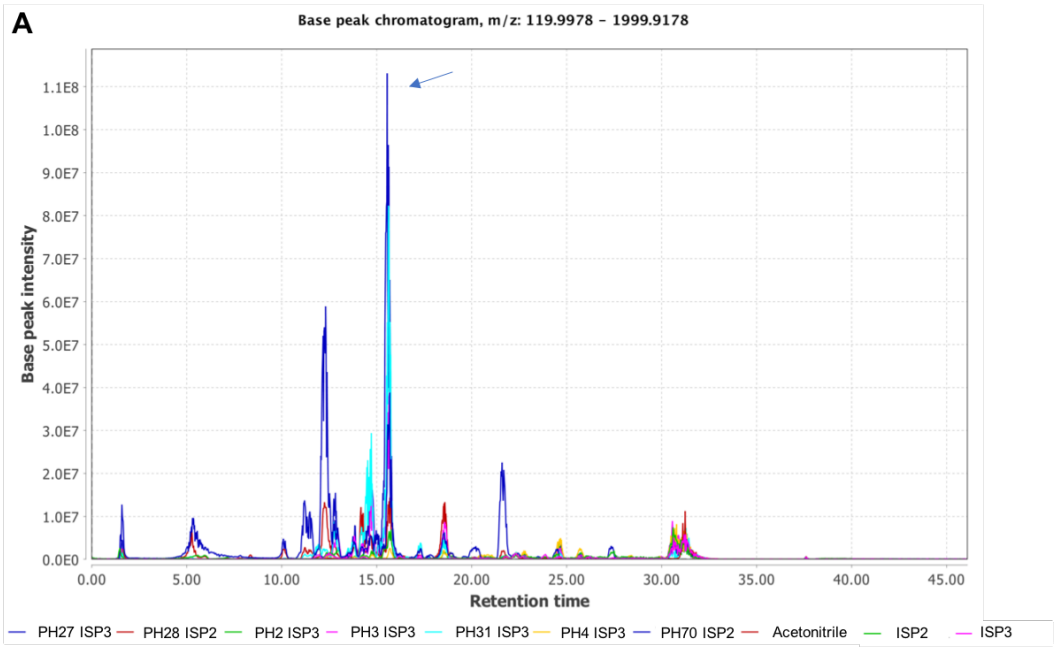


Figure 5.12 Clustering of the genes composing the family FAM_00088 shared by soil 7 isolates with activity against *Enterococcus faecalis* and *Staphylococcus aureus*.

A. Relationship between biosynthetic gene clusters (except PH2) with aclacinomycin M and cinerubin B. **B.** Comparison of PH2 with PH27 and PH28 cinerubin B gene clusters showing gene differences between them.



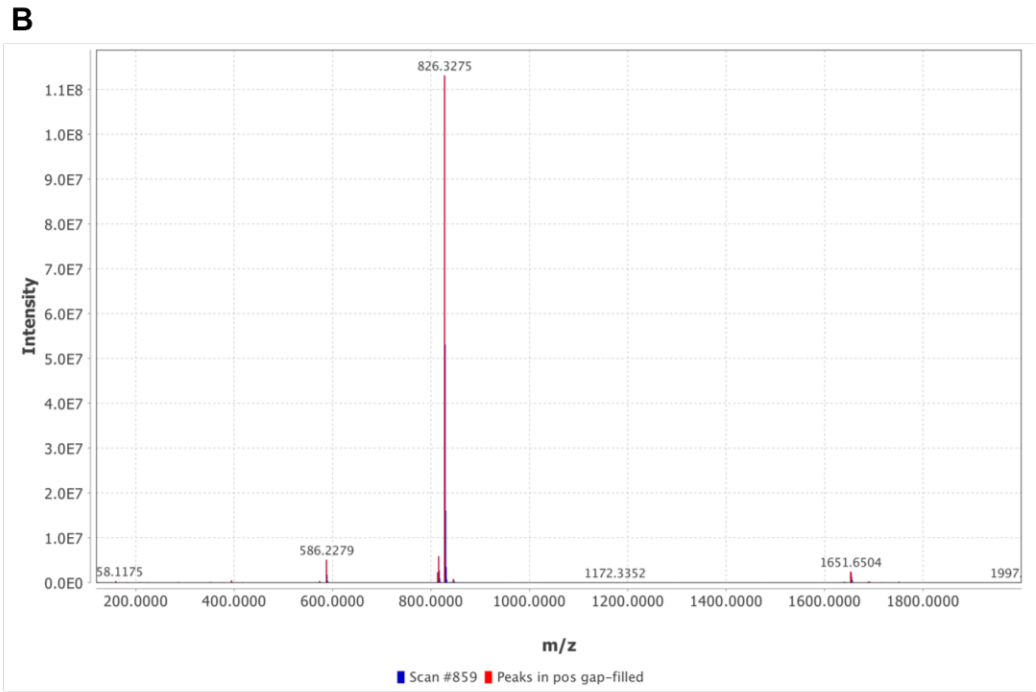


Figure 5.13 Base peak plot chromatogram of species from soil 7, solvent and media blanks in the positive ionization mode showing a peak *m/z* 826.328 and corresponding to cinerubin B based on dereplication.

A. Overlay of the *m/z* from crude extracts of species from soil 7 showing a peak (arrow) at 15 minutes of retention time (intensity between 10^7-10^8) and **B.** corresponding to cinerubin B (*m/z* 826.3275). Created on MzMine 2.20.

5.4 Identification of the active compound using large scale fermentation of PH28 and mass spectrometry analysis

In order to identify the unknown metabolites from the small scale (*m/z* 1307.560 and/or 1308.560), a bacterial culture of PH28 was scaled-up to 10 L (29 x 350 mL) in ISP2. This isolate produced the highest concentrations of both metabolites compared to the remaining isolates in the small-scale fermentation. Additionally, the prediction of biosynthetic gene clusters using its complete genome could potentially elucidate and identify these metabolites. After seven days of incubation, bacterial cultures were extracted using ethyl acetate and HP-20 resin as described before (**Chapter 2 - section 2.6.3**). Further fractionation of the extracts was performed using the Biotage MPLC Isolera^{TM} One 2.0.4 Spektra Flash Purification System with variable gradient elution method (**Chapter 2 - section 2.6.3.1**).

5.4.1 Fractioning and MS analysis of the fractions

Fractionation was accomplished using gradient ratios of isopropanol and methanol affording 130 test tubes which were pooled according to their TLC profiles (**Figure 5.14**) yielding 29 fractions (**Table 5.8**). F1 to F18 and F28 were obtained from the extract prepared by maceration of HP-20 beads through which the culture broth was pass through to absorb the secondary metabolites. F19 to F27 and F29 were from the extract attained by liquid-liquid partitioning of the culture filtrate that passed through the HP-20 beads. The pooled fractions were analysed for bioactivity by disk diffusion assay (**Figure 5.15**). Activity was found against *E. faecalis* for fractions F2 and F3 (1.1 and 1.4 cm, respectively; **red box**) and F20 and F21 (0.8 cm; **orange box**). However, the inhibition zones obtained were smaller than those obtained for the positive control imipenem (10 μg/disc) with diameter of 2.9 cm.

The mass spectral data was processed, dereplicated and analysed by multivariate analysis. PCA plots were generated from twenty-nine components, a R^2X value of 1 and Q^2 of 0.651 were achieved using Pareto scaling with a difference of 0.345 indicating a good fit and prediction for the model (**Figure 5.16**). According to their chemical similarities, fractions clustered into two groups, the first group comprised of F2, F3, F4 and F5 while the second group had the rest of fractions together with the wash and media blank. Fraction F1 was an outlier (**Figure 5.16, panel A**). The

loadings plot represented the molecular ion masses of the metabolites found in the fractions on the corresponding quadrant position on the scores plot (**Figure 5.16, panel B**). Fractions F3 and F4 were located on the bottom middle quadrant and indicated similar chemical profiles but only F3 did show bioactivity against *E. faecalis*. However, F2, situated closer to the centre of the score plot than F3 and F4, was able to inhibit *E. faecalis*. Interestingly, F1 showed a distinctive chemical profile from the other fractions while F16 and F17 were observed to be furthest from the rest of the fractions.

OPLS-DA scores, loadings scatter and S–plots were generated and validated by permutation test by grouping the fractions into active or inactive to identify the metabolites responsible for the bioactivity (**Figure 5.17**). The R^2Y was 0.923 and Q^2 was 0.346, while the Q^2Y intercept was -0.217. These values specified good fitting but poor prediction as the R^2Y was close to 1 and $Q^2 < 0.6$, while the Q^2Y intercept in the permutation test was less than 0 indicating validity of the model. However, the difference between R^2Y and Q^2 was greater than 0.3 that could imply an over fitted dataset. The $R^2X(1)$ was 0.0539, which indicated a variation between groups of 5.39% and $R^2X0(1)$ of 10.6% as variation within groups revealing chemical diversity between fractions. Discriminating metabolites for the active end of the S-plot were dereplicated using the DNP database (Buckingham, 2001) (**Table 5.9**).

Actinomycins Y_1 and Y_5 (negative mode) were putatively identified and six unknown metabolites. The unknown metabolite P_7574 (1307.552 MW) along with actinomycins Y_1 and Y_5 were found in fractions F2, F3, and F21 (**Figures 5.18** and **5.19**) suggesting a role on the bioactivity observed. P_7574 afforded a molecular formula of $C_{61}H_{82}ClN_{11}O_{19}$, which was unusual having only 11 nitrogens. All reported actinomycin congeners have 12 nitrogens with the exception of actinomycin G_5 (**Figure 5.19**). Actinomycin G_5 was isolated from *Streptomyces iakyrus* DSM 41873 and showed antibacterial against *B. subtilis*. The presence of chlorine in the metabolites found in fractions F2, F3, and F21 was established by the isotopic pattern of the high-resolution data. The highest abundance of the actinomycin derivatives were found in F3. However, the putatively new congener was overlapping with the known metabolite actinomycin Y_1 and was found to be at lower concentration than the known derivative.

Table 5.8 Fractionation of PH28 crude extract from 10 L scale-up yielding 29 fractions. F1 to F18 and F28 were obtained from extracted HP-20 beads of the culture broth. F19 to F27 and F29 were from the liquid-liquid partitioned culture broth filtrate.

Fraction	Tested tube numbers	Yield (mg)	Solvent gradient
F1	1-5	3.7	DCM/MeOH
F2	6	22.3	
F3	7-9	5.7	
F4	10-11	7.0	
F5	12-13	6.0	
F6	14	0.2	
F7	15-16	5.6	
F8	17-20	4.1	
F9	21-24	3.4	
F10	1-2	0.1	DCM/iPrOH
F11	3-4	3.5	
F12	5	5.6	
F13	6-11	12.5	
F14	12-19	11.4	
F15	20-24	8.4	
F16	25-39	11.6	
F17	40-49	5.0	
F18	50-63	10.5	
F19	1-2	58.4	DCM/MeOH
F20	3-9	6.0	
F21	10-11	20.9	
F22	12-13	14.2	
F23	14-15	4.0	
F24	16	2.8	
F25	17-19	5.3	
F26	20-39	51.4	
F27	40-41	16.6	
F28	Wash run	6.2	
F29	Wash run	7.0	

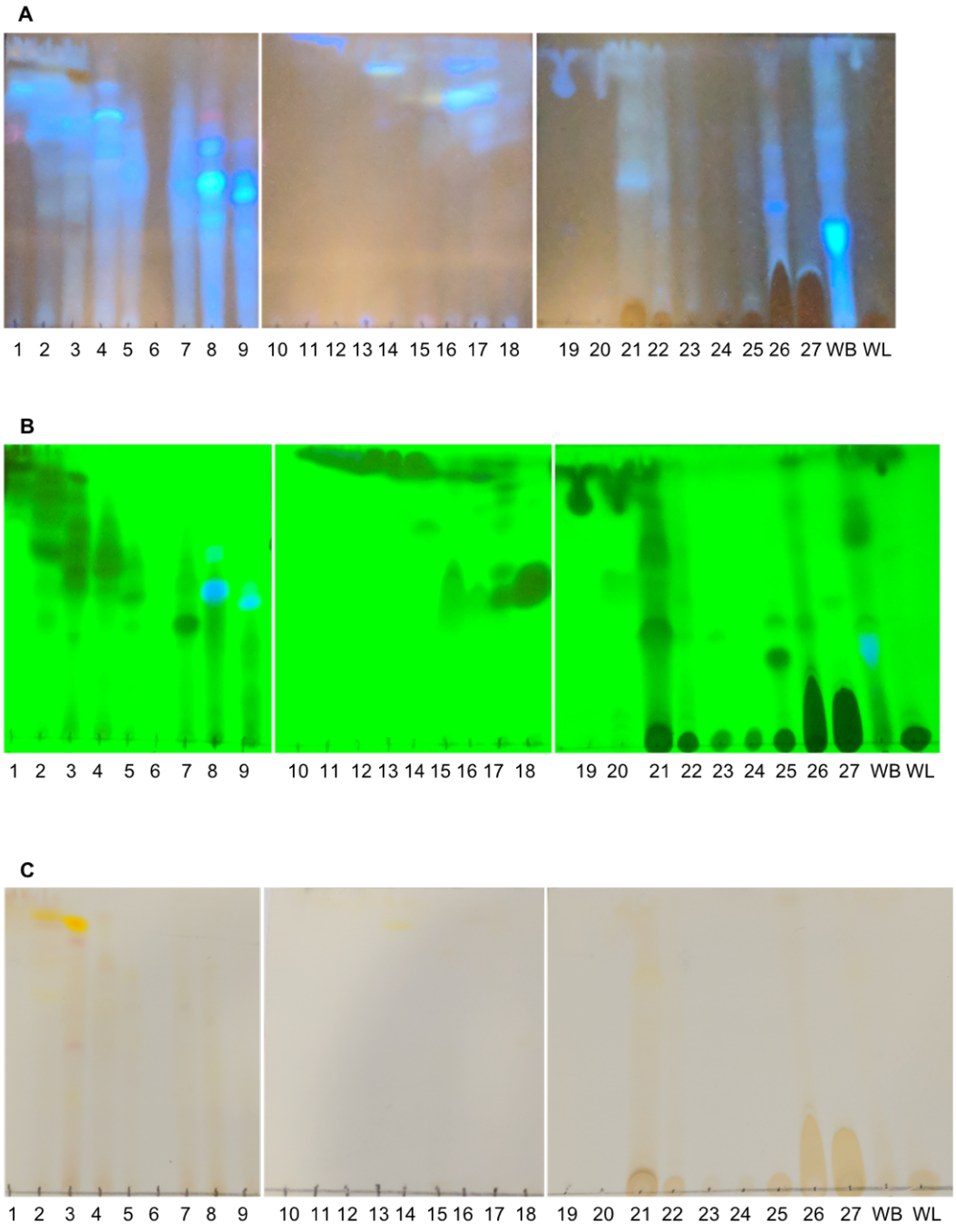


Figure 5.14 Summary TLC plate of the 27 fractions obtained after fractionation of PH28 with mobile phase DCM:MeOH (9:1, v/v).

A. TLC plate observed under long wavelength ; **B.** under short wavelength and **C.** under visible light. Fractions WB and WL contain the wash produced at the end of the runs for beads and liquid extracts, respectively.

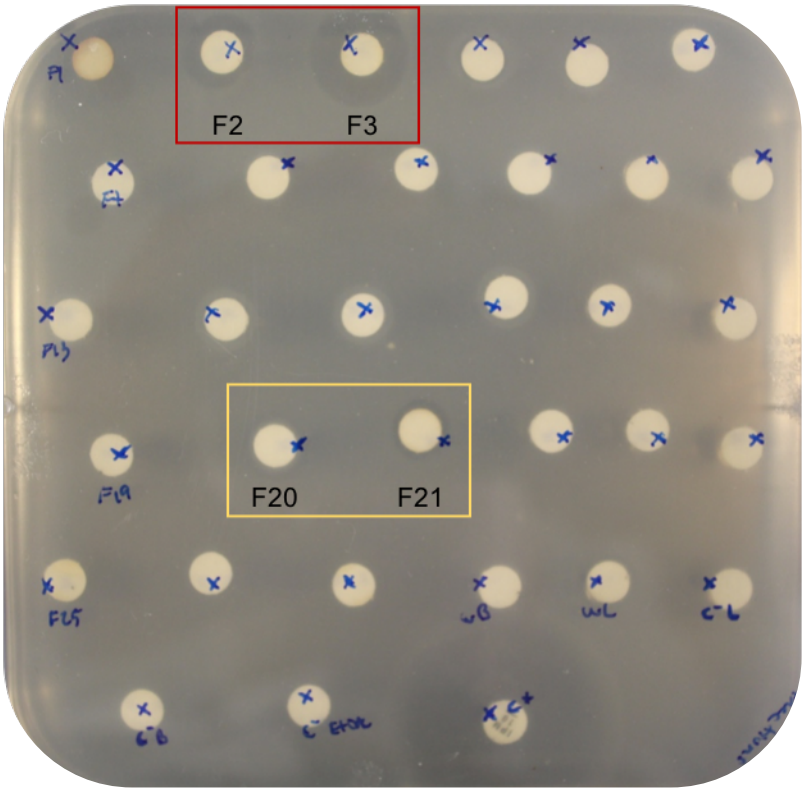


Figure 5.15 Bioactivity assay showing the inhibition zones against *E. faecalis* by the fractions extracted from the fermentation of PH28.

Active fractions are represented by a red box (F2 and F3) and orange box (F20 and F21) according to activity. Imipenem (10 μg/disc) was used as control.

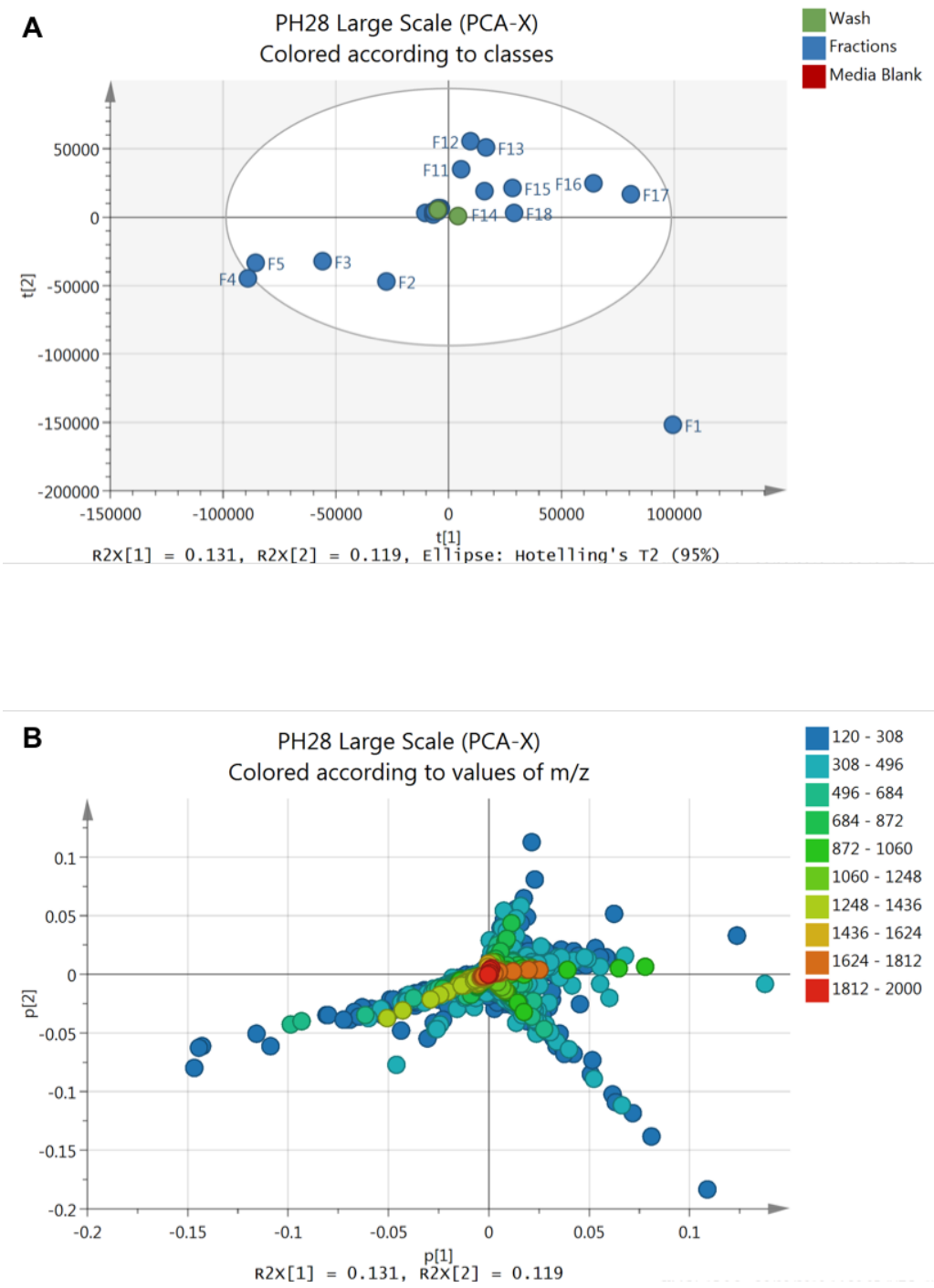


Figure 5.16 Principal component analysis (PCA) plot for the fractions obtained from PH28 fermentation.

A. PCA-X score plot of samples coloured by media used. **B.** Loadings plot acquired from the PCA-X model correlated with the score plot.

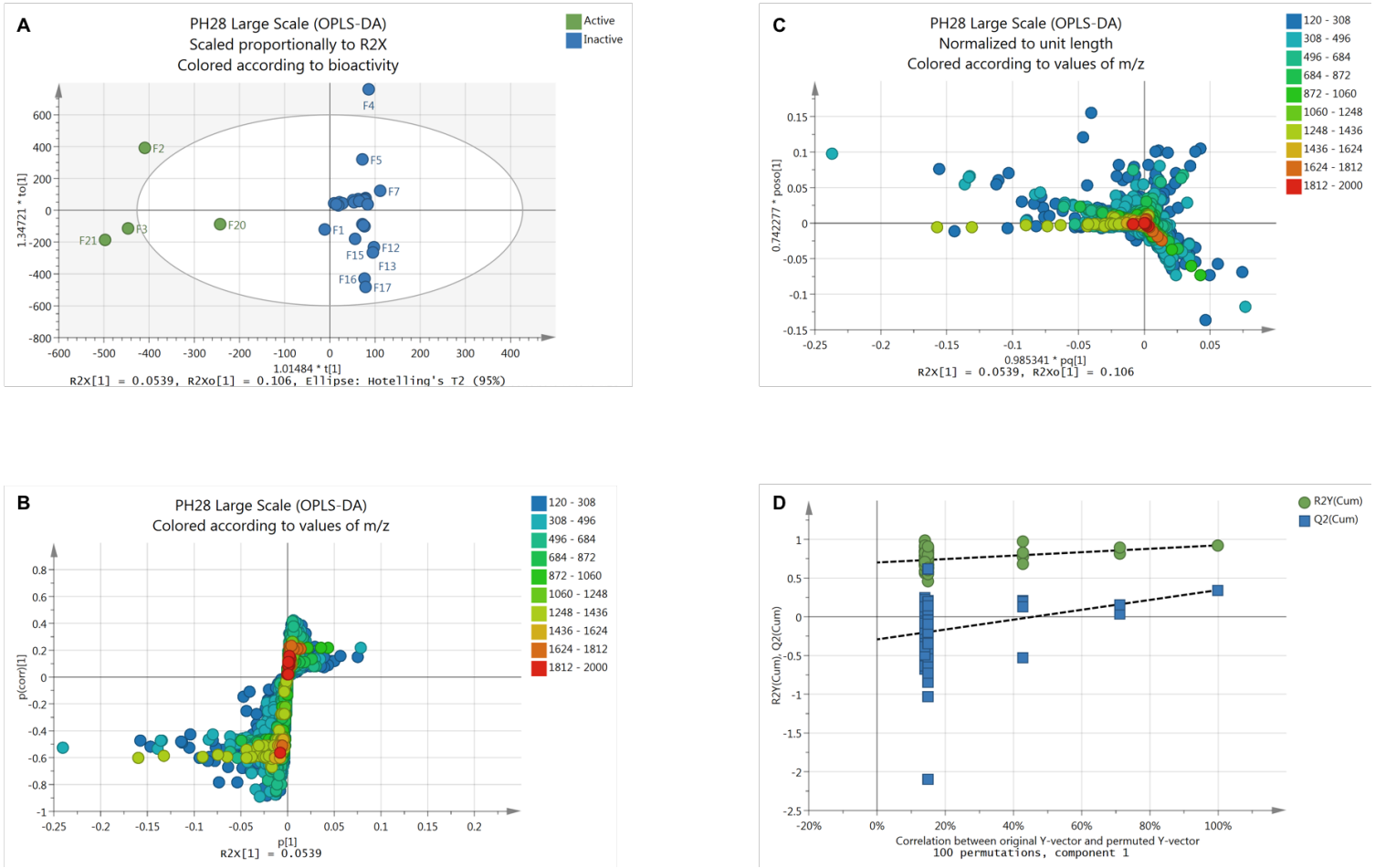


Figure 5.17 OPLS-DA SIMCA plots of the bioactivity of PH28 fractions against *E. faecalis*.

A. Samples classed as active or inactive on the disk diffusion assay. **B.** Bottom left quadrant of the S-plot showing *m/z* corresponding to active fraction in A. **C.** Loadings plot acquired from the OPLS-DA model. **D.** Permutation test to validate OPLS model where Q^2Y intercept must be less than zero.

Table 5.9 Outlying metabolites detected by OPLS-DA in PH28 fractions that showed bioactivity against *E. faecalis*.

Listed metabolites only included those with p-values < 0.05 and were considered significant while hits found from the DNP database (Buckingham, 2001) were filtered for *Streptomyces* strains.

Mzmine ID	*m/z*	RT (min)	MW (Da)	Formula	Name	Source
P_1407	164.071	5.27	163.063	$C_9H_9NO_2$	3,4-dihydro-1(2*H*)-isoquinolinone	*S. griseus* 2002-104
P_7607	164.071	5.45	163.063	$C_9H_9NO_2$	3,4-dihydro-1(2*H*)-isoquinolinone	*S. griseus* 2002-104
P_2054	271.144	5.29	270.137	$C_{16}H_{18}N_2O_2$	3-benzylidene-6-isobutylidene-2,5-piperazinedione	*S. albus* ATCC 3004, *S.* sp. SUC1
P_2037	277.027	5.04	276.020	no prediction		
P_1738	365.189	5.52	364.182	$C_{27}H_{24}O$	no hits	-
P_7579	429.320	26.78	428.313	$C_{24}H_{44}O_6$	2-*O*-octadecylascorbic acid	*S. antibioticus*, *Streptomyces* spp.
P_7567	429.321	27.21	428.313	$C_{24}H_{44}O_6$	2-*O*-octadecylascorbic acid	*S. antibioticus*, *Streptomyces* spp.
P_9874	489.249	6.81	488.242	$C_{25}H_{34}N_4O_4$	cyclo(phenylalanylprolylphenylalanylprolyl)	*S. barakatei*, *S.* sp. TN17
P_1474	489.249	6.83	488.242	$C_{25}H_{34}N_4O_4$	cyclo(phenylalanylprolylphenylalanylprolyl)	*S. barakatei*, *S.* sp. TN17
P_1707	521.206	5.90	520.198	$C_{28}H_{24}N_8O_3$ $C_{32}H_{28}N_2O_5$	no hits	-
N_831	1267.565	18.57	1268.572	$C_{61}H_{80}N_{12}O_{18}$	actinomycin Y5	*S.* sp. Gu-GS12
N_829	1303.541	18.58	1304.548	$C_{61}H_{81}ClN_{12}O_{18}$	actinomycin Y1	*S.* sp. Gu-GS12
P_7564	1305.56	18.59	1304.549	$C_{61}H_{81}ClN_{12}O_{18}$	actinomycin Y1	*S.* sp. Gu-GS12
P_7574	1308.56	18.57	1307.551	$C_{61}H_{82}ClN_{11}O_{19}$	a putative new actinomycin congener	-
N_830	1306.543	18.58	1307.550	$C_{61}H_{82}ClN_{11}O_{19}$	a putative new actinomycin congener	-

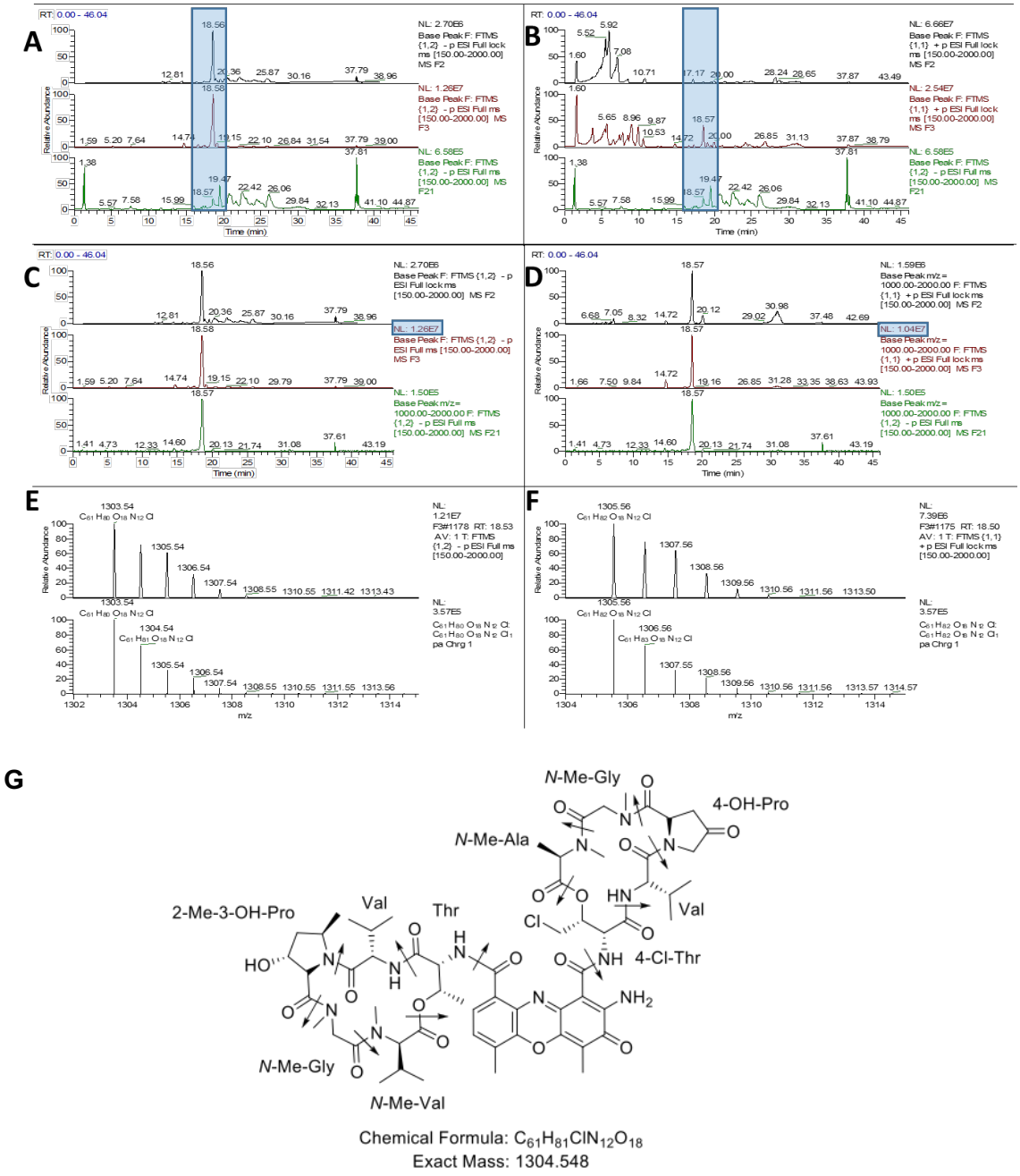


Figure 5.18 Extracted ion chromatograms and high-resolution mass spectra for *m/z* 1303.54 [M-H] and 1305.56 [M+H] dereplicated as actinomycin Y_1 in fractions 2, 3, and 21.

A. TICs in negative mode and **B.** in positive mode **C.** Extracted ion chromatograms for *m/z* 1303.54 [M-H] and **D.** *m/z* 1305.56 [M+H] indicating highest peak intensity (in blue box) found in F3. **E.** Expanded mass spectra showing the isotopic patterns for *m/z* 1303.54 [M-H] and simulated spectrum for $C_{61}H_{80}ClN_{12}O_{18}$ and **F.** *m/z* 1305.56 [M+H] for $C_{61}H_{82}ClN_{12}O_{18}$. **G.** Structure of actinomycin Y_1.

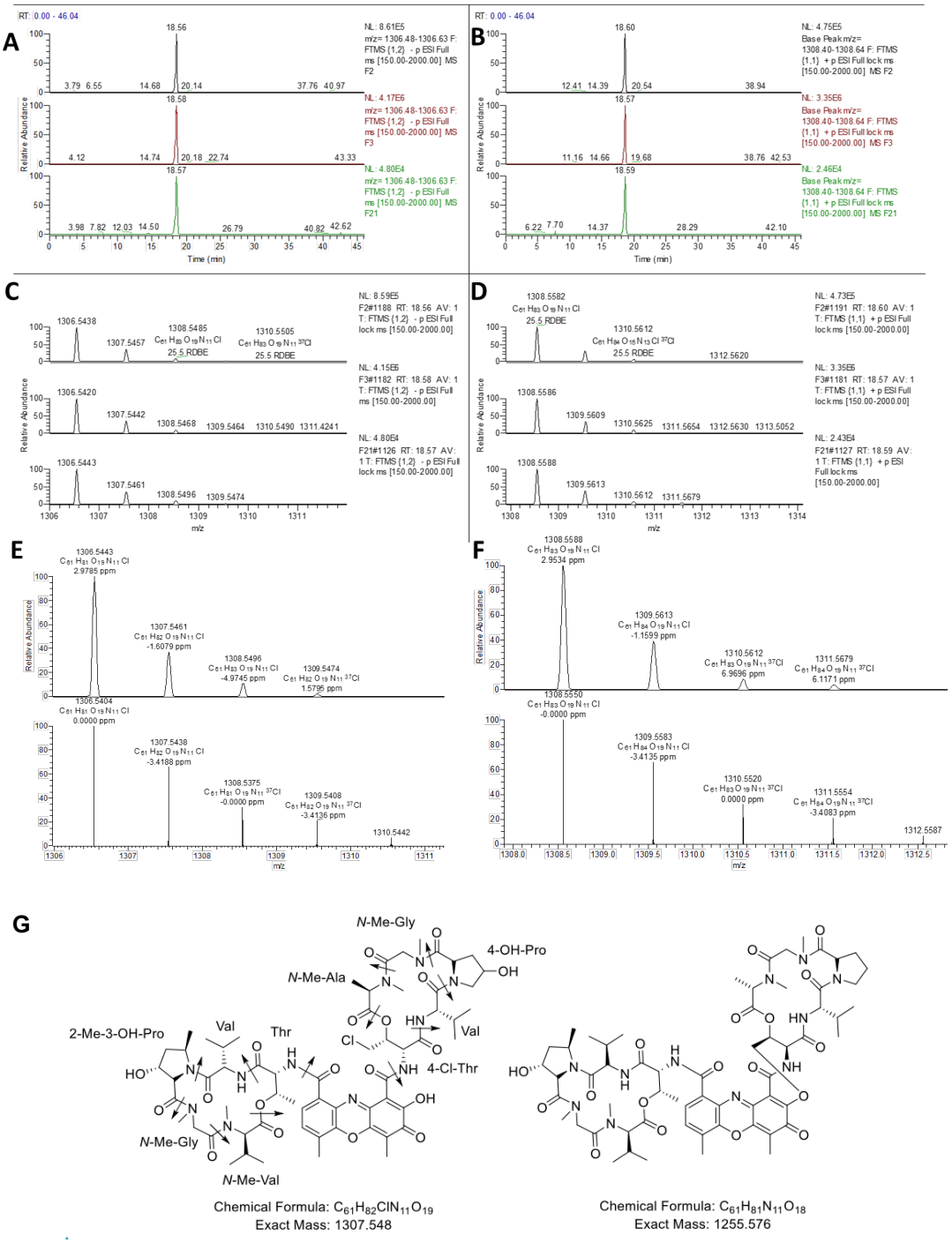


Figure 5.19 Extracted ion chromatograms and high-resolution mass spectra for unknown *m/z* 1306.54 [M-H] and 1308.56 [M+H] in fractions 2, 3, and 21.

A. Extracted ion chromatograms for *m/z* 1306.54 [M-H] and **B.** *m/z* 1308.56 [M+H]. **C.** Expanded mass spectra showing the isotopic patterns for *m/z* 1306.54 [M-H] and **D.** *m/z* 1305.56 [M+H]. **E.** Simulated spectra for $C_{61}H_{80}ClN_{12}O_{18}$ and **F.** $C_{61}H_{82}ClN_{12}O_{18}$. **G.** Putative structure for $C_{61}H_{82}ClN_{12}O_{18}$ and actinomycin G_5. Arrows indicate the respective amino acids as determined by TOCSY-NMR.

From the previous analysis using BiG-SCAPE (**Figure 5.11**) no homologs of the actinomycin D biosynthetic gene clusters were found in the collected isolates. However, the encoded putative actinomycin was distinct from actinomycin D (**Figure 5.20, panel A**) and similar to actinomycin G from *S. iakyrus* (**Figure 5.20, panel B**) and *Streptomyces flavovariabilis* (**Figure 5.20, panel D**). Furthermore, after comparison between the actinomycin encoded in isolate PH28 and the two other clusters (**Figure 5.21**), an overall good alignment and gene conservation was observed. Interestingly, a unique gene (405 bp) in PH28 (**Figure 5.21** - asterisk) was identified that showed homology to transcriptional modulators of the PemK-like protein family impairing cell growth (Pellegrini *et al.*, 2005). PemK-like proteins have been identified in several bacteria including *E. coli*, *B. subtilis* and *Streptomyces* species (Pellegrini *et al.*, 2005; Saleh *et al.*, 2012). A homolog was found in *Streptomyces anulatus* nearby the endophenazine biosynthetic gene cluster but it was not required for its correct biosynthesis (Saleh *et al.*, 2012).

From the mass spectral data of F2, F3, and F21, it was deduced that these three bioactive fractions did contain the actinomycin congeners at varying concentrations (**Figure 5.18** and **Figure 5.19**). Similarly, the NMR spectral data of F2, F3, and F21 were compared to simulated spectra of known metabolites actinomycins Y_1 and Y_2 as well as the proposed structure (**Figure 5.22**). Analysis of the NMR spectral data confirmed that F3 afforded the highest abundance of the major metabolite actinomycin Y_1 albeit F3 affording the lowest yield of 5.7 mg in comparison with F2 and F21 at around 20 mg each. F2 and F21 contained more compounds other than the actinomycins as their major metabolite and hence it was impossible to do further elucidation work. The ^{1}H NMR spectrum of F3 was most related and compatible to those reported in the literature for actinomycin Y_1 (Bitzer *et al.*, 2009) (**Figure 5.23**). Furthermore, F3 was still a mixture of possibly two to three actinomycin derivatives, where the presence of the related G_5 congener could be deduced from the extra overlapping resonances between 7.0 to 7.5 ppm. Two-dimensional COSY (**Figure 5.24, panel A**) and TOCSY (**Figure 5.24, panel B**) were performed to elucidate the individual amino acids present in the structure. Due to the relatively low yield, it was impossible to isolate the new actinomycin derivatives.

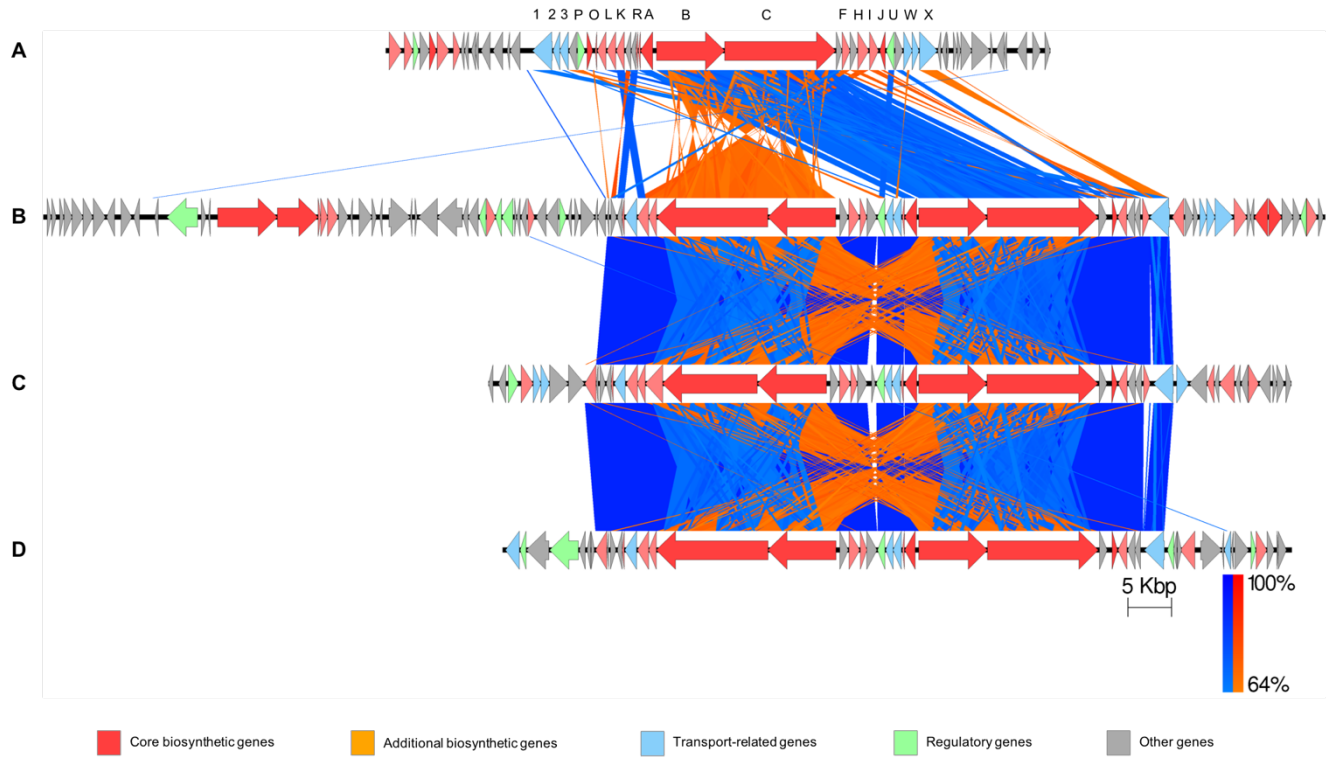


Figure 5.20 Alignment of the biosynthetic gene clusters of actinomycin D (A) and actinomycin G from *S. iakyrus* (B), *Streptomyces* spp. PH28 (C) and *S. flavovariabilis* (D).

Alignment created with Easyfig (version 2.1) (Sullivan, Petty & Beatson, 2011). The different colours were matched based on the antiSMASH 5.0 gene annotation. Letters and numbers indicate genes important for the production of actinomycin D. Numbers 1: *acmrC*; 2: *acmrB*; 3: *acmrA*. Inverted BLAST matches represent with blue gradient depending on identity (lower as aqua and higher as blue) and normal BLAST represented by gradient from orange (lower identity) to red (higher identity).

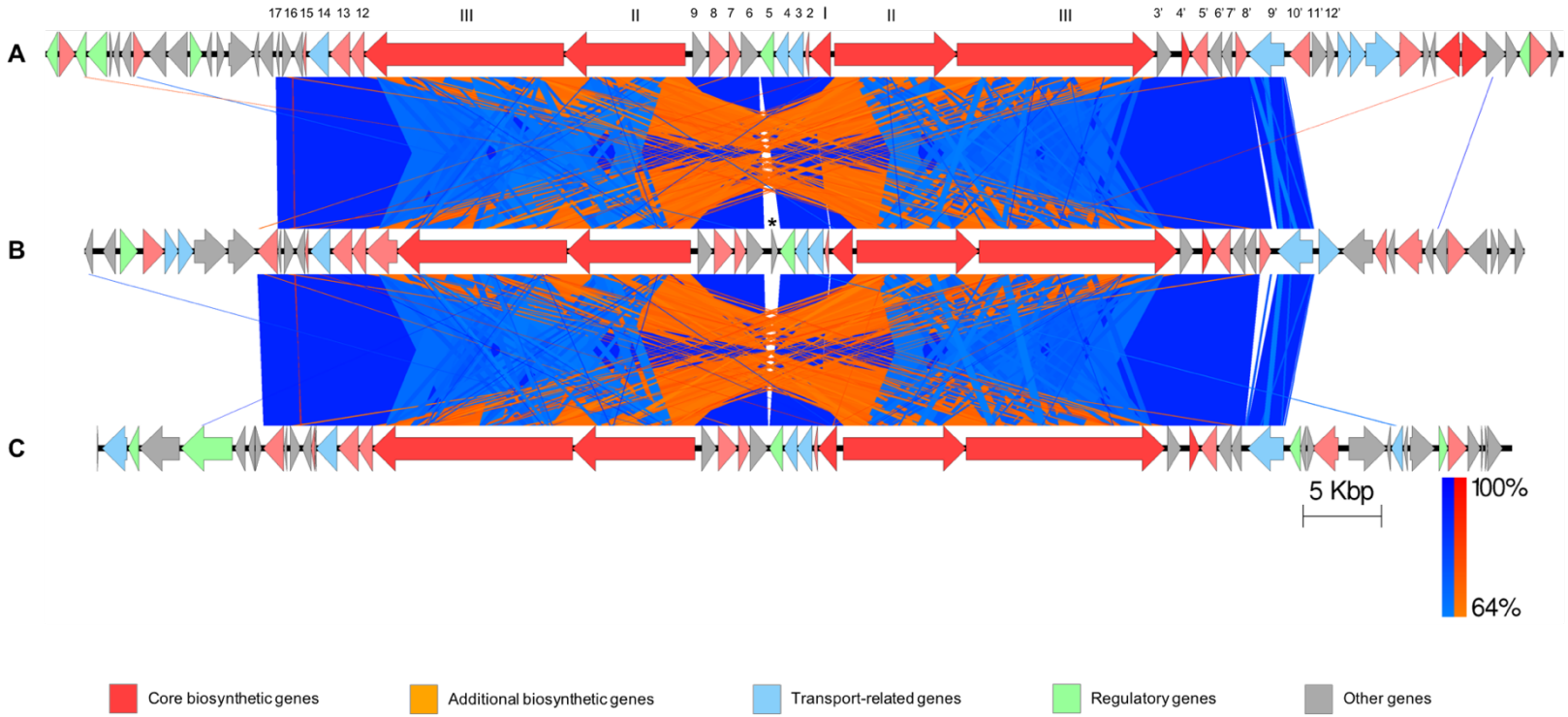


Figure 5.21 Alignment of the biosynthetic gene clusters of actinomycin G from *S. iakyrus* (A), *Streptomyces* spp. PH28 (B) and *S. flavovariabilis* (C).

Alignment created with Easyfig (version 2.1) (Sullivan, Petty & Beatson, 2011). The different colours were matched based on the antiSMASH 5.0 gene annotation. Actinomycin synthetase I (**I**) is the initiation module and two NRPS multidomain (**II** and **III**) responsible for the assembly and release of the peptide. The asterisk represents an additional gene present in *Streptomyces* spp. PH28 encoding a transcriptional modulator. Tailoring and regulatory genes are indicated by numbers according to Wang *et al.* (2013). Inverted BLAST matches represent with blue gradient depending on identity (lower as aqua and higher as blue) and normal BLAST represented by gradient from orange (lower identity) to red (higher identity).

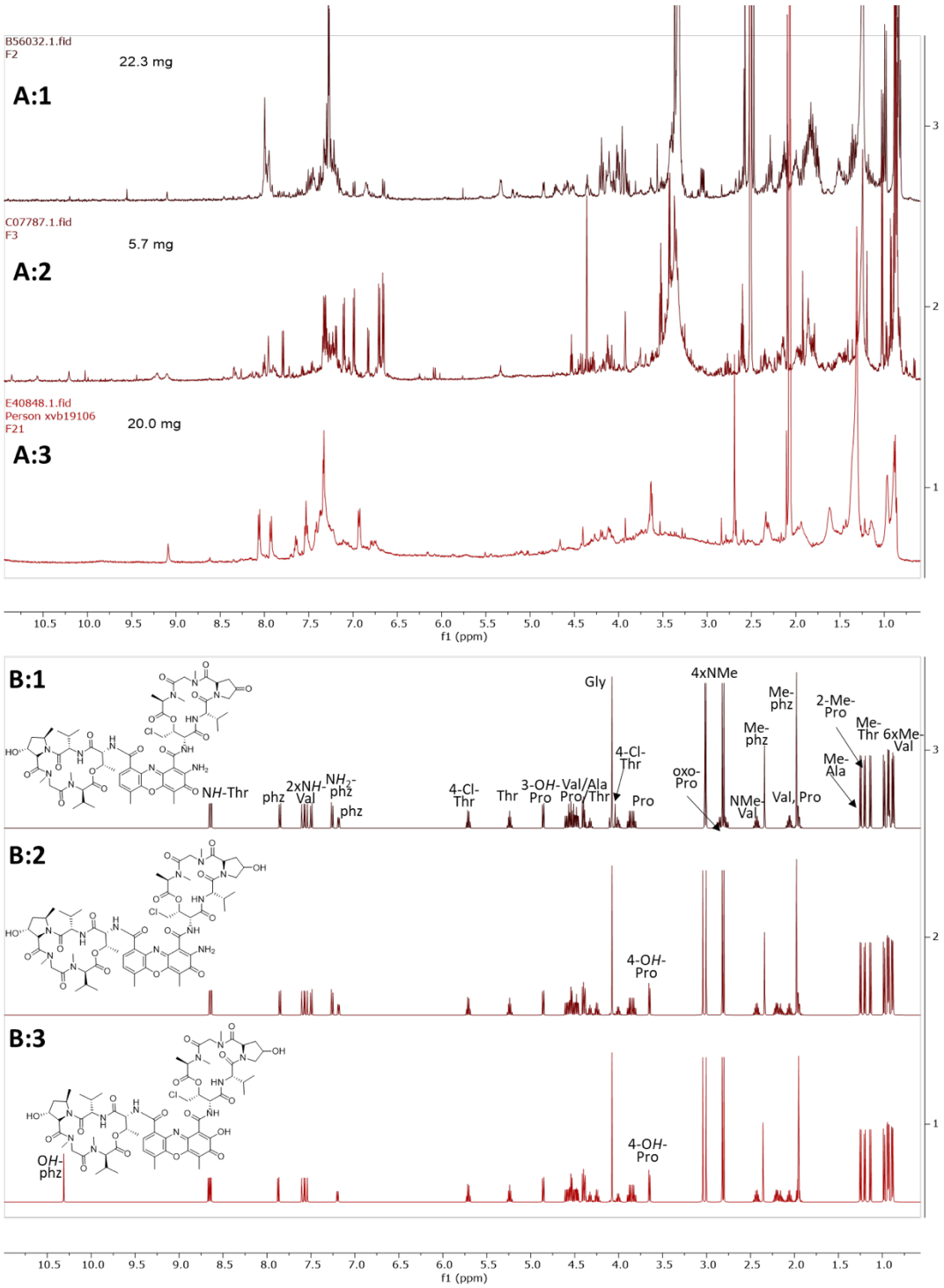


Figure 5.22 ^{1}H NMR spectra for selected fractions of PH28 obtained from the scale-up fermentation in DMSO-*d*$_6$.

A:1-A:3. ^{1}H NMR spectra of fractions 2 (**A:1**), 3 (**A:2**), and 21 (**A:3**) in DMSO-*d*$_6$. **B:1-B:3.** Simulated ^{1}H NMR spectra of actinomycins Y_1 (**B:1**), Y_2 (**B:2**), and for the proposed structure for $C_{61}H_{82}ClN_{12}O_{18}$ (**B:3**).

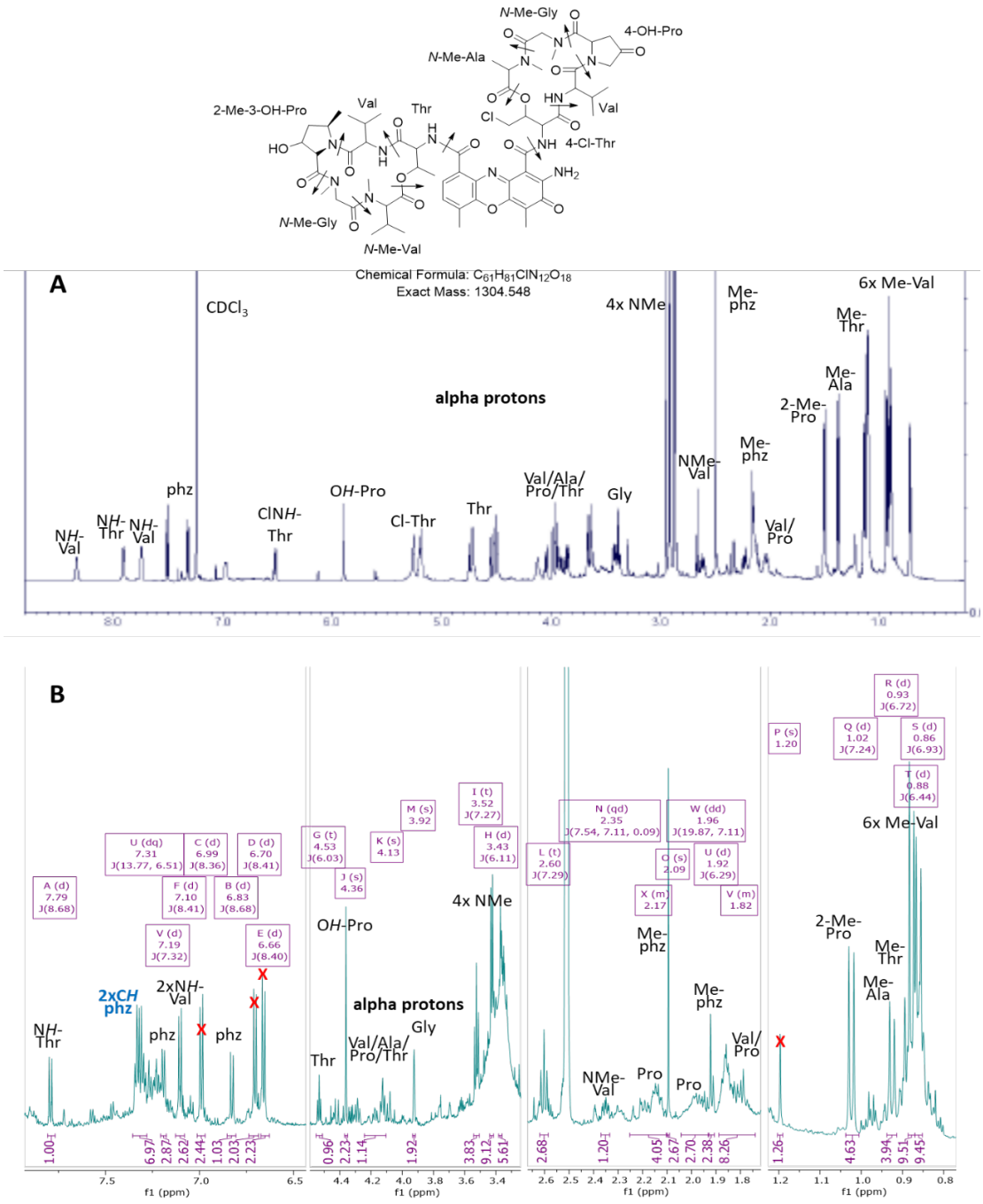


Figure 5.23 ^{1}H NMR spectra for selected fraction F3 from PH28 scale-up fermentation in DMSO-d_6.

A. ^{1}H NMR spectra of actinomycins Y_1 in $CDCl_3$ from Bitzer *et al.* (2009). **B.** ^{1}H NMR spectrum of fraction 3 in DMSO-d_6. The extra pair of phenaxine methine protons are shown in blue.

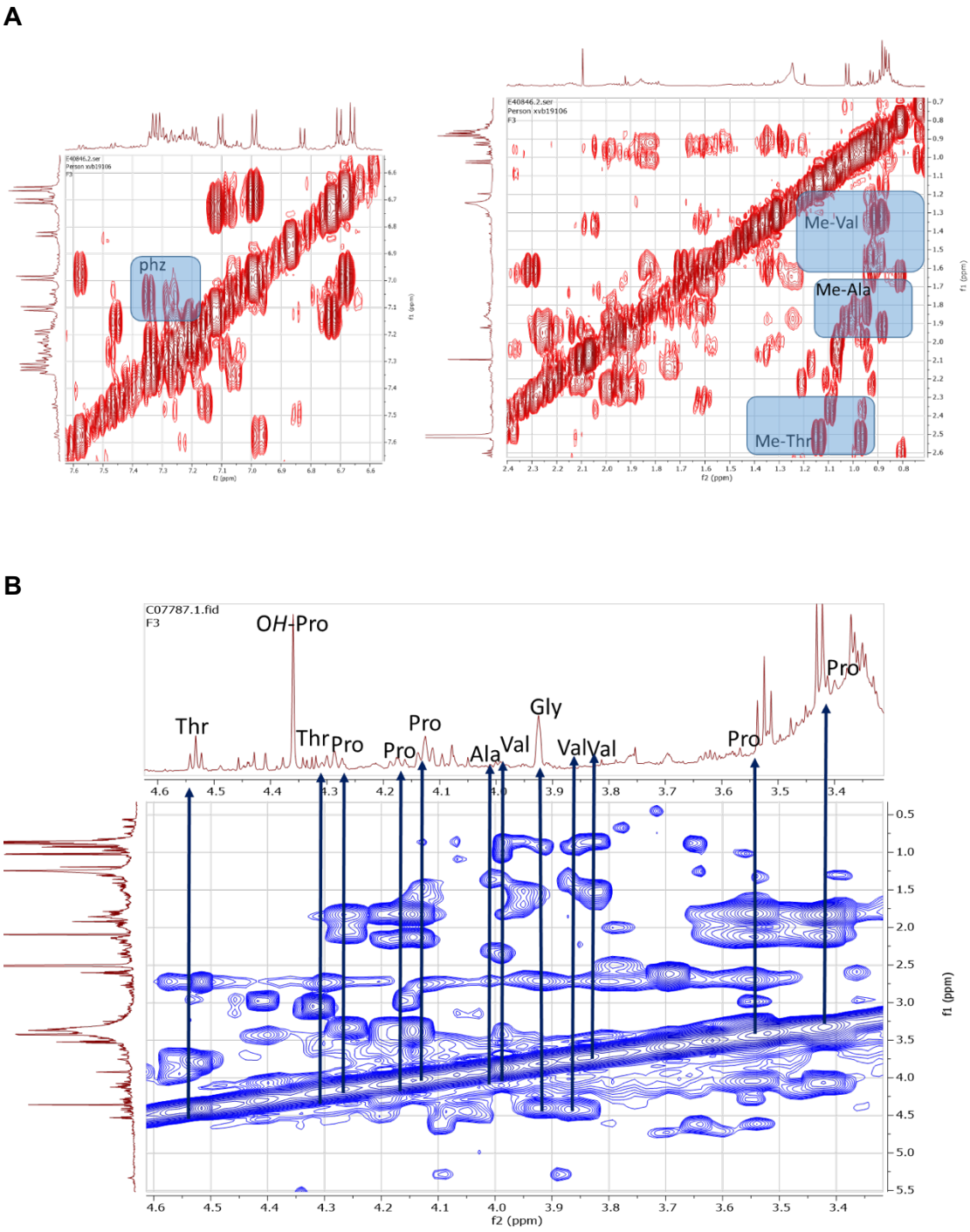


Figure 5.24 Two-dimensional COSY (A) and TOCSY (B) of selected fraction F3 from PH28 scale-up fermentation for illustrating the occurrence of individual amino acids for structure elucidation.

5.5 Discussion

Among the fourteen isolates bioactivity was observed from solid media against one or more indicator strains (**Table 5.1**) and mainly against Gram-positive bacteria (*E. faecalis*, *S. aureus* and *B. subtilis*). Species from soil 7, including PH28, showed a promising activity profile with isolates PH2, PH3, PH4 also able to inhibit Gram-negative bacteria (*A. baumannii* and *E. coli*). Furthermore, previous *Streptomyces* species isolated from the Atacama Desert showed the same potential with chaxamycins and chaxalactins, active against MRSA and other Gram-positive bacteria, and atacamycins, chaxapeptin and abenquines being identified as anticancer and anti-inflammatory, respectively (Busarakam *et al.*, 2014; Schulz *et al.*, 2011; Nachtigall *et al.*, 2011a; Elsayed *et al.*, 2015). However, the same activity profile was not observed in liquid cultures of the isolates, only *S. aureus* and *B. subtilis*, were inhibited by the crude extracts. Overall, crude extracts from PH2 and PH4 grown on liquid ISP5 showed activity against *S. aureus* (**Table 5.4**), contradicting that observed previously in agar (**Table 5.1**). Similarly, the same occurred against *B. subtilis*, where isolates, e.g. PH28 in ISP2 liquid, that had no activity in one of the agar media showed activity in the corresponding liquid medium (**Table 5.4**). More surprisingly, isolates PH6 and J57 in liquid media did not show any activity (**Table 5.4**) against any of the indicator strains despite showing inhibition against two of the 8 indicator strains tested before (**Table 5.1**). Additionally, no activity was observed against *A. baumannii* or *E. coli* from PH2, PH3 or PH4, which was in contrast to the previous results when using plug assays (**Table 5.1**).

The distinct inhibition of *E. faecalis* between soil isolates 7 and 8 was observed in both solid and liquid assays with no zones of inhibition observed for the latter (**Table 5.4**) due to differential antibiotic production. Furthermore, no activity was observed against *S. aureus* from crude extracts of soil 8 isolates contrasting with the agar media assay (**Table 5.1**). Clearly, fermentation of the isolates in liquid culture produced different chemical profiles when compared to those obtained with solid media. Similar results were described in previous studies where different metabolites were identified depending on the media, solid or liquid. One explanation could be the distinct stages of growth in liquid and solid or the composition of the media used (Romano *et al.*, 2018). Other limiting factors between both protocols could be the inconsistent levels of oxygenation, not the correct growth stage for starting the

production and pH during growth in liquid cultures. Additionally, detection of the metabolites can be dependent on the solvents used for extraction, freeze drying steps or the capability of the metabolites to ionise during mass spectrometry analysis. Thus, optimisation of growth conditions or the use of elicitors can enable production in liquid cultures (Manteca & Yagüe, 2018; Romano *et al.*, 2018).

The analysis of the small-scale fermentation of the isolates revealed the presence of several known metabolites, predominantly aclacinomycin M, cinerubin B and actinomycin Y1/Y5, with known anticancer or anti-Gram-positive bioactivity. Furthermore, Oki *et al.* (1979) isolated red and yellow coloured compounds that corresponded to aclacinomycin and cinerubin from *Streptomyces galilaeus*, respectively. The same observation was found in some crude extracts obtained in this work that showed a similar pigmentation and additionally, these metabolites were also extracted previously using ethyl acetate and/or Diaion HP-20 (Krohn, 2009) suggesting a correlation between the pigmentation and the identified metabolites by dereplication. Differences between soil 7 and soil 8 isolates emerged specifically in terms of production of cinerubin B that was confirmed by analysis of the whole-genome showing the absence of the biosynthetic gene cluster in soil 8 isolates. Nakata *et al.* (1992) showed that anthracyclines, such as cinerubin B and R, produced by *Streptomyces eurythermus* strain HM1715MY2 inhibited the growth of several Gram-positive bacteria including *S. aureus* and *B. subtilis*. This suggests that cinerubin B produced by the Atacama isolates could play a role in inhibiting *S. aureus* and/or *E. faecalis*. Recently, *Streptomyces* sp. CMAA 1527, isolated from Antarctica, was identified as producing cinerubin B and suggests a widespread production among this genus (Silva, 2018).

Interestingly, some metabolites present in soil 7 isolates could not be identified from the databases and two that were chemically similar to actinomycin Y_1/Y_5 were selected for isolation and purification work. The crude extract from isolate PH28 despite its low quantity (~5.6 mg) showed the best inhibition zones against the indicator strains. Furthermore, completeness of its genome and the ability to produce cinerubin B made this isolate the best candidate for large-scale fermentation. Only one of the two metabolites was found, MW 1307.552, in fractions F2, F3, and F21 that showed bioactivity (**Figure 5.13**) and thus suggesting that this metabolite could be the responsible for the antimicrobial activity. Nevertheless, the two actinomycins

Y_1 and Y_5, previously observed, were also identified from the fractions F2, F3, and F21, implying that these are not pure. NMR analysis confirmed the presence of actinomycin Y_1 as the main constituent of the active fractions in varying concentrations. Actinomycin Y_1 showed the ability to inhibit Gram-positive bacteria as *S. aureus* and *B. subtilis* but not *E. coli* (Bitzer, Gesheva & Zeeck, 2006; Bitzer *et al.*, 2009). This observation is consistent with the results in this current work and thus responsible for the activity observed in liquid and solid media.

Overall, the Atacama Desert isolates were able to produce a wide range of secondary metabolites in solid and liquid culture with activity against the tested indicator strains. Small-scale fermentation shed light in the metabolomic diversity among the chosen isolates while common metabolites e.g. aclacinomycin, cinerubin B and actinomycin among others were also produced especially in soil 7. Further, the production of an active metabolite associated with actinomycin Y_1 was identified.

Chapter 6. Summary and Conclusion

From two distinct locations in the Atacama Desert but with quite similar conditions, several isolates were identified by 16S rRNA sequencing as *Streptomyces phaeoluteigriseus* DSM 41896. Moreover, these isolates constituted the majority of the *Actinobacteria* isolated in both soils. However, further analysis showed that these were distinct species with soil 7 comprising a new species.

Seven isolates from each location were selected for further studies. A phylogenetic tree was constructed, and all of the isolates grouped, as expected, in one branch demonstrating the close relatedness between them (>99% identity of 16S rRNA sequence). However, phenotypic and biochemical analysis showed significant differences between isolates and hinted the possibility of two different species. Previous work showed the inaccuracy of 16S rRNA analysis (Antony-Babu *et al.*, 2017) and thus MLSA (multi-locus sequence analysis), average nucleotide identity and pulse-field gel electrophoresis were performed and confirmed this result. The presence and absence of plasmids could contribute to rapid genome diversification of co-occurrent isolates as previously proposed, with emerging strain-specific gene content (Vetsigian, Jajoo & Kishony, 2011; Tidjani *et al.*, 2019).

These strains shared a recent common ancestor and possibly distinct gene sequences emerged as niche-specific pressures materialised creating two different populations (Bentley & Parkhill, 2004; Ventura *et al.*, 2007; Nishida, 2012). Events, such as horizontal gene transfer, creating genomic islands, and/or genomic rearrangements triggered by prophages can originate intraspecific variations harbouring functional adaptations to that niche (Ventura *et al.*, 2007; Penn *et al.*, 2009). However, it was hypothesized that plasmids could be the main contributor to genetic exchange and the specific adaptions in the Atacama Desert due to the distinct PFGE profiles observed. Further sequencing and comparative genomic analysis of the plasmids from each species could help delineate their evolutionary relatedness.

Overall, lateral gene transfer in *Streptomyces* is common mostly between closely related strains and the observed diversity is likely to be cause by this phenomenon (McDonald & Currie, 2017). Yet, genes involved in secondary metabolism, e.g. natural

products, are mostly acquired through lateral gene transfer of one gene or the full cluster followed by gene exchange with other clusters (McDonald & Currie, 2017). Most of these were found in the accessory genome localised in *Streptomyces* at the terminal regions of the linear chromosome (McInerney, McNally & O'Connell, 2017). Accordingly, a large proportion of clusters encoding specialised metabolites were found in this hotspot of variability, important to increase fitness of a strain and/or population to a niche (Tidjani *et al.*, 2019). Furthermore, this microevolution with unique genes and alleles could promote strain survival and population stabilisation within a niche (Li *et al.*, 2019).

Several clusters were found in all isolates including terpenes, NRPS, RiPPS, bacteriocins, butyrolactones, siderophores, melanin and ectoines. The latter plays an important role for the survival of the isolates in hypersaline environments demonstrating a broader adaption to the Atacama Desert (Gunde-Cimerman, Plemenitaš & Oren, 2018). Nonetheless, unique clusters found in only one of the location suggests that each species inherited these clusters from its common ancestor. Furthermore, production of unique antibiotics could be linked to different roles, e.g. competitive interaction against co-occurrent and/or foreigner bacteria or act as a public good to stabilise a population (Tidjani *et al.*, 2019). In this study, it was confirmed that these two hypotheses can occur, demonstrating a differential role for antibiotics in the two populations (**Figure 6.1**).

Isolates from soil 7 showed intense inhibition towards isolates from soil 8 but also some ability to inhibit each other. The latter observation could be due to relatedness differences among species from soil 7 which comprised of two distinct clades (**Figure 6.1 – green** and **yellow** box). Previous studies found that closely related isolates were unable to inhibit each other (Vetsigian, Jajoo & Kishony, 2011; Tidjani *et al.*, 2019). In this work, none of isolates from soil 8 was able to inhibit its closely related strains (**Figure 6.1 – red box**), despite the results for species from soil 7. Interestingly, isolate PH31 seemed to highly inhibit isolates from soil 8 (**Figure 6.1 –** inhibition zone sizes > 2 cm) but only showed moderate inhibition against its closely related strains from soil 7. From the assays performed on PH31, it did not show any exceptional features so further analysis of this strain is required to understand why it possesses such a profile. For future studies, tagging and co-culturing the isolates with fluorescent genes, *mcherry* and *egfp*, could improve the understanding on socially mediated

interactions of competition and/or cooperation between them (Vetsigian, Jajoo & Kishony, 2011; Abrudan *et al.*, 2015).

In this study, it was found that cinerubin B and actinomycin gene clusters were only identified in the genome of isolates from soil 7. Actinomycin was produced in the scale-up fermentation and was able to inhibit *E. faecalis* while possibly responsible for the activity observed in previous assays. However, cinerubin B was only found in the metabolome of isolates from soil 7 during the small-scale fermentation but not in the active fractions generated after the scale-up. This could suggest that certain metabolite production in specific conditions or in sub-inhibitory concentrations was acting as signalling molecules within the microbial niche rather than taking an antibacterial role. Surfactin and bacillaene from *B. subtilis* were previously linked as signalling molecules inhibiting the development of aerial hyphae and spores in *S. coelicolor* (Tyc *et al.*, 2017). Furthermore, other signalling compounds such as γ-butyrolactones, furans, and γ-butenolides have also been reported to impact antibiotic production (Arakawa, 2018). For instance, butenolides produced in low concentrations by *Streptomyces albus* J1074 were found to stimulate avermectin production in *S. avermectilis* demonstrating the potential environmental competitive interactions between species (Nguyen *et al.*, 2018). Similarly, another study showed that avermectin B1a induced the production of cryptic secondary metabolites in *S. albus* J1074 functioning also as a signaling mechanism between interspecies (Xu *et al.*, 2017). A successful gene knockout and complementation of cinerubin B could shed light on its ecological role in competition and/or cooperation interactions. Isolates from soil 7 highly inhibited strains from the other soil (**Table 6.1**) and was therefore hypothesised that production of cinerubin B or actinomycin could aid *kin* selection by cooperation of highly similar strains for their survival against other species (Hibbing *et al.*, 2010). Furthermore, inhibition between isolates in soil 8 was not observed, which contradicted previous studies were *Streptomyces* from the same location inhibited each other more than *Streptomyces* from other locations (Tyc *et al.*, 2017).

In the Atacama Desert novel metabolites have been discovered, for example chaxamycin D and abenquines (aminobenzoquinones) (Bull & Asenjo, 2013). Production of specialised metabolites against clinically relevant pathogens similar to those previously reported (Elsayed *et al.*, 2015; Rateb *et al.*, 2011; Abdelkader *et al.*, 2018) were also found in this study. Actinomycin Y_1 was identified in this work as the

main bioactive metabolite inhibiting the two ESKAPE pathogens *E. faecalis* and *S. aureus*. Actinomycins have already been earlier described to inhibit Gram-positive bacteria, mostly *Staphylococcus* and *Bacillus* species, in 1940 (Waksman & Woodruff, 1940). However, in this current work, actinomycins were afforded in low-yielding active fractions that were not pure enough to determine the total structure of the derivatives. A further 30 L scale-up is recommended for future work and optimisation studies will be needed to increase the yield and concentration of the actinomycins.

Elicitation of secondary metabolites following the OSMAC approach, using chemicals or stress, to activate cryptic biosynthetic clusters has been shown to be viable (Yoon & Nodwell, 2014). Enhanced production of jadomycin B by *S. venezuelae* and validamycin A by *S. hygroscopicus* were achieved using higher growth temperatures or adding ethanol (Yoon & Nodwell, 2014). Similarly, ectoine produced by *S. coelicolor* was increased using high concentrations of salt since it acts as a osmoprotectant conferring high osmolality endurance (Yoon & Nodwell, 2014; Bursy *et al.*, 2008). It has been shown in this work that high concentrations of salts, especially Na^+, K^+ and Mg^{2+} led to an increase in inhibition against *B. subtilis* and *S. aureus* (increase of ~1.0 cm in diameter). Interestingly, the non-Atacama Desert *S. phaeoluteigriseus* DSM 41896 started producing an active metabolite under osmotic stress. Changes in the development caused by salts are intimately related to production of antibiotics and could explain this observation (Elibol & Mavituna, 1998; Sevcikova & Kormanec, 2004). Other conditions that can act as elicitors of biosynthetic gene clusters, such as co-cultures or chemicals, could be used in future studies (Abdelmohsen *et al.*, 2015; Rutledge & Challis, 2015; Okada & Seyedsayamdost, 2017).

In conclusion, this work has confirmed that 16S rRNA is not appropriate to resolve phylogenetic relationship between closely related strains and propose the use of HSI as a new promising tool complemented by whole-genome analysis. This study has shown the effect of specific pressures, e.g. plasmid mediated gene transfer, on species diversification due its niche/location resulting in genomic and metabolomic differences between close evolutionary strains. This work also confirmed the potential for *Streptomyces* from the Atacama Desert to produce a wide range of metabolites, and the identification of active congeners of actinomycin.

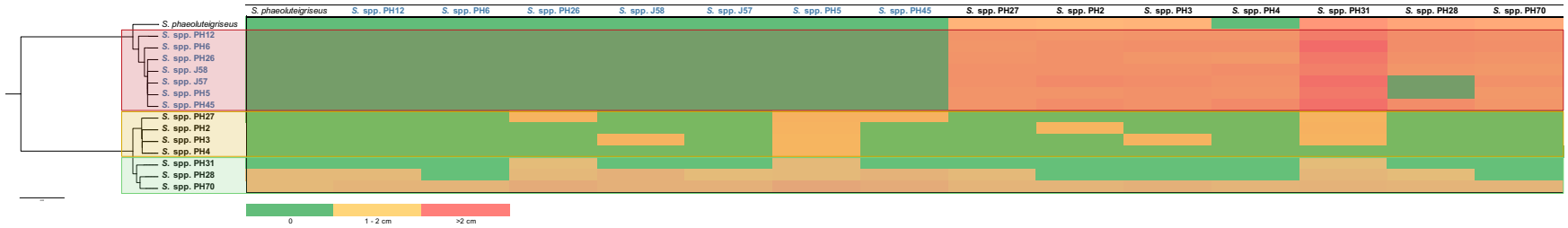


Figure 6.1 Overview of the phylogeny of the isolates and the inhibition observed between the two locations.

Isolates from soil 7 represented **in black** and soil 8 in **blue**. The diameter of the inhibition zones (cm) is represented by a gradient between no (green), intermediate (orange) and high inhibition (red). Columns represent plugs from lawns against isolates seeded on plate (rows). Phylogenetic affiliations obtained using PIRATE and core genome.

Chapter 7. References

Abdel-Mageed, W., Milne, B., Wagner, M., Schumacher, M., Sandor, P., Pathom-aree, W., Goodfellow, M., Bull, A., Horikoshi, K. & Ebel, R. (2010) Dermacozines, a new phenazine family from deep-sea dermacocci isolated from a Mariana Trench sediment. *Organic & biomolecular chemistry*. 8 (10), 2352–2362.

Abdelkader, M., Philippon, T., Asenjo, J., Bull, A., Goodfellow, M., Ebel, R., Jaspars, M. & Rateb, M. (2018) Asenjonamides A–C, antibacterial metabolites isolated from *Streptomyces asenjonii* strain KNN 42.f from an extreme-hyper arid Atacama Desert soil. *The Journal of antibiotics*. 71 (4), 425.

Abdelmohsen, U., Grkovic, T., Balasubramanian, S., Kamel, M., Quinn, R. & Hentschel, U. (2015) Elicitation of secondary metabolism in actinomycetes. *Biotechnology advances*. 33 (6), 798–811.

Abraham, J., Bousquet, A., Bruff, E., Carson, N., Clark, A., Connell, A., Davis, Z., Dums, J., Everington, C. & Groth, A. (2016) *Paenibacillus larvae* phage Tripp genome has 378-base-pair terminal repeats. *Genome Announc.* 4 (1), e01498-15.

Abrudan, M., Smakman, F., Grimbergen, A., Westhoff, S., Miller, E., van Wezel, G. & Rozen, D. (2015) Socially mediated induction and suppression of antibiosis during bacterial coexistence. *Proceedings of the National Academy of Sciences*. 112 (35), 11054–11059.

Aebersold, R. & Mann, M. (2003) Mass spectrometry-based proteomics. *Nature*. 422 (6928), 198.

Afgan, E., Baker, D., Batut, B., Van den Beek, M., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Coraor, N. & Eberhard, C. (2018) The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. *Nucleic acids research*. 46 (W1), W537–W544.

Alain, K. & Querellou, J. (2009) Cultivating the uncultured: limits, advances and future challenges. *Extremophiles*. 13 (4), 583–594.

Alanjary, M., Steinke, K. & Ziemert, N. (2019) AutoMLST: an automated web server

for generating multi-locus species trees highlighting natural product potential. *Nucleic acids research*. 47 (W1), W276–W282.

Anne, J., Wohlleben, W., Burkardt, H., Springer, R. & Pohler, A. (1984) Morphological and molecular characterization of several actinophages isolated from soil which lyse *Streptomyces cattleya* or *S. venezuelae*. *Microbiology*. 130 (10), 2639–2649.

Antony-Babu, S., Stach, J. & Goodfellow, M. (2010) Computer-assisted numerical analysis of colour-group data for dereplication of streptomycetes for bioprospecting and ecological purposes. *Antonie Van Leeuwenhoek*. 97 (3), 231–239.

Antony-Babu, S., Stach, J.E.M. & Goodfellow, M. (2008) Genetic and phenotypic evidence for *Streptomyces griseus* ecovars isolated from a beach and dune sand system. *Antonie Van Leeuwenhoek*. 94 (1), 63–74.

Antony-Babu, S., Stien, D., Eparvier, V., Parrot, D., Tomasi, S. & Suzuki, M. (2017) Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Scientific Reports*. 7 (1), 11089.

Arakawa, K. (2018) Manipulation of metabolic pathways controlled by signaling molecules, inducers of antibiotic production, for genome mining in Streptomyces spp. *Antonie Van Leeuwenhoek*. 111 (5), 743–751.

Araujo, F., Barh, D., Silva, A., Guimarães, L. & Ramos, R. (2018) GO FEAT: a rapid web-based functional annotation tool for genomic and transcriptomic data. *Scientific reports*. 8 (1), 1794.

Arnau, G., Danilovich, M., Sánchez, L., Acosta, F. & Delgado, O. (2016) Novel sources of antimicrobials from pristine and poorly explored environments. The patagonia microbiota case. In: *Biology and Biotechnology of Patagonian Microorganisms*. Springer. pp. 127–146.

Arndt, D., Grant, J., Marcu, A., Sajed, T., Pon, A., Liang, Y. & Wishart, D. (2016) PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*. 44 (W1), W16–W21.

Ausmees, N., Wahlstedt, H., Bagchi, S., Elliot, M., Buttner, M. & Flärdh, K. (2007) SmeA, a small membrane protein with multiple functions in *Streptomyces* sporulation including targeting of a SpoIIIE/FtsK-like protein to cell division septa. *Molecular Microbiology*. 65 (6), 1458–1473.

Awad, H., Khamis, M. & El-Aneed, A. (2015) Mass spectrometry, review of the basics: ionization. *Applied Spectroscopy Reviews*. 50 (2), 158–175.

Baltz, R. (2016) Gifted microbes for genome mining and natural product discovery. *Journal of Industrial Microbiology & Biotechnology*. 44 (4–5), 573–588.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A., Dvorkin, M., Kulikov, A., Lesin, V., Nikolenko, S., Pham, S. & Prjibelski, A. (2012) SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology*. 19 (5), 455–477.

Barka, E., Vatsa, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Klenk, H.-P., Clément, C., Ouhdouch, Y. & van Wezel, G. (2016) Taxonomy, physiology, and natural products of Actinobacteria. *Microbiology and Molecular Biology Reviews*. 80 (1), 1–43.

Bayliss, S., Thorpe, H., Coyle, N., Sheppard, S. & Feil, E. (2019) PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. *GigaScience*. 8 (10).

Becker, S., Kortz, L., Helmschrodt, C., Thiery, J. & Ceglarek, U. (2012) LC–MS-based metabolomics in the clinical laboratory. *Journal of Chromatography B*. 883, 68–75.

Bentley, S., Chater, K., Cerdeño-Tárraga, A., Challis, G., Thomson, N., James, K., Harris, D., Quail, M., Kieser, H. & Harper, D. (2002) Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3 (2). *Nature*. 417 (6885), 141.

Bentley, S. & Parkhill, J. (2004) Comparative genomic structure of prokaryotes. *Annu. Rev. Genet.* 38, 771–791.

Bertani, G. (1951) Studies on lysogenesis I.: The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of bacteriology*. 62 (3), 293.

Bertelli, C., Laird, M., Williams, K., Group, S.F.U.R.C., Lau, B., Hoad, G., Winsor, G. & Brinkman, F. (2017) IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic acids research*. 45 (W1), W30–W35.

Bhamidimarri, S., Zahn, M., Prajapati, J., Schleberger, C., Söderholm, S., Hoover, J., West, J., Kleinekathöfer, U., Bumann, D., Winterhalter, M. & van den Berg, B. (2019) A Multidisciplinary Approach toward Identification of Antibiotic Scaffolds for *Acinetobacter baumannii*. *Structure*. 27 (2), 268-280.e6.

Bhat, S., Nazir, R., Malik, T. & Shah, F. (2013) Secondary metabolites of actinomycetes as potential source of antibiotics. *Stem Cell*. 4 (2), 41–46.

Bhullar, K., Waglechner, N., Pawlowski, A., Koteva, K., Banks, E., Johnston, M., Barton, H. & Wright, G. (2012) Antibiotic resistance is prevalent in an isolated cave microbiome. *PloS one*. 7 (4), e34953.

Bibb, M. (2005) Regulation of secondary metabolism in streptomycetes. *Current opinion in microbiology*. 8 (2), 208–215.

Bibb, M.J., Molle, V. & Buttner, M.J. (2000) ςBldN, an Extracytoplasmic Function RNA Polymerase Sigma Factor Required for Aerial Mycelium Formation in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology*. 182 (16), 4606 LP – 4616.

Binaschi, M., Bigioni, M., Cipollone, A., Rossi, C., Goso, C., Maggi, C.A., Capranico, G. & Animati, F. (2001) Anthracyclines: selected new developments. *Current Medicinal Chemistry-Anti-Cancer Agents*. 1 (2), 113–130.

Bishop, A., Fielding, S., Dyson, P. & Herron, P. (2004) Systematic insertional mutagenesis of a streptomycete genome: a link between osmoadaptation and antibiotic production. *Genome research*. 14 (5), 893–900.

Bitzer, J., Gesheva, V. & Zeeck, A. (2006) Actinomycins with altered threonine units in the β-peptidolactone. *Journal of natural products*. 69 (8), 1153–1157.

Bitzer, J., Streibel, M., Langer, H.-J. & Grond, S. (2009) First Y-type actinomycins from *Streptomyces* with divergent structure-activity relationships for antibacterial and cytotoxic properties. *Organic & biomolecular chemistry*. 7 (3), 444–450.

Blair, J., Webber, M., Baylay, A., Ogbolu, D. & Piddock, L. (2015) Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*. 13 (1), 42.

Blin, K., Shaw, S., Steinke, K., Villebro, R., Ziemert, N., Lee, S.Y., Medema, M. & Weber, T. (2019) antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic acids research*. 47 (W1), W81–W87.

Bolger, A., Lohse, M. & Usadel, B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 30 (15), 2114–2120.

Bosi, E., Donati, B., Galardini, M., Brunetti, S., Sagot, M.-F., Lió, P., Crescenzi, P., Fani, R. & Fondi, M. (2015) MeDuSa: a multi-draft based scaffolder. *Bioinformatics*. 31 (15), 2443–2451.

Bouslimani, A., Sanchez, L., Garg, N. & Dorrestein, P. (2014) Mass spectrometry of natural products: current, emerging and future technologies. *Natural product reports*. 31 (6), 718–729.

Brown, D. (2015) Antibiotic resistance breakers: can repurposed drugs fill the antibiotic discovery void? *Nature Reviews Drug Discovery*. 14 (12), 821.

Brown, E. & Wright, G. (2016) Antibacterial drug discovery in the resistance era. *Nature*. 529 (7586), 336.

Buckingham, J. (2001) *Dictionary of natural products on CD-ROM*.

Bull, A. & Asenjo, J. (2013) Microbiology of hyper-arid environments: recent insights from the Atacama Desert, Chile. *Antonie Van Leeuwenhoek*. 103 (6), 1173–1179.

Bull, A., Asenjo, J., Goodfellow, M. & Gómez-Silva, B. (2016) The Atacama Desert: technical resources and the growing importance of novel microbial diversity. *Annual review of microbiology*. 70, 215–234.

Bull, A. & Goodfellow, M. (2019) Dark, rare and inspirational microbial matter in the extremobiosphere: 16 000 m of bioprospecting campaigns. *Microbiology*. 165 (12), 1252–1264.

Bursy, J., Kuhlmann, A., Pittelkow, M., Hartmann, H., Jebbar, M., Pierik, A. & Bremer, E. (2008) Synthesis and uptake of the compatible solutes ectoine and 5-hydroxyectoine by *Streptomyces coelicolor* A3 (2) in response to salt and heat stresses. *Applied and environmental microbiology*. 74 (23), 7286–7296.

Busarakam, K., Bull, A., Girard, G., Labeda, D., van Wezel, G. & Goodfellow, M. (2014) *Streptomyces leeuwenhoekii* sp. nov., the producer of chaxalactins and chaxamycins, forms a distinct branch in *Streptomyces* gene trees. *Antonie Van Leeuwenhoek*. 105 (5), 849–861.

Bush, M., Tschowri, N., Schlimpert, S., Flardh, K. & Buttner, M. (2015) c-di-GMP signalling and the regulation of developmental transitions in streptomycetes. *Nat Rev Micro*. 13 (12), 749–760.

Bycroft, B.W. & Payne, D.J. (2013) *Dictionary of Antibiotics and Related Substances*. CRC Press.

Cabanettes, F. & Klopp, C. (2018) D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ*. 6, e4958.

Campos, V., Valenzuela, C., Yarza, P., Kämpfer, P., Vidal, R., Zaror, C., Mondaca, M.-A., Lopez-Lopez, A. & Rosselló-Móra, R. (2010) *Pseudomonas arsenicoxydans* sp nov., an arsenite-oxidizing strain isolated from the Atacama desert. *Systematic and Applied Microbiology*. 33 (4), 193–197.

Carver, T., Harris, S., Berriman, M., Parkhill, J. & McQuillan, J. (2012) Artemis: an integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics*. 28 (4), 464–469.

Castresana, J. (2002) GBLOCKS: selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Version 0.91 b. Copyrighted by J. Castresana, EMBL*.

Centore, P. (2016) sRGB centroids for the ISCC-NBS colour system. *Munsell Colour Sci. Painters*.

Chater, K. (2006) *Streptomyces* inside-out: a new perspective on the bacteria that provide us with antibiotics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*. 361 (1469), 761–768.

Chater, K. & Chandra, G. (2008) The use of the rare UUA codon to define “Expression Space” for genes involved in secondary metabolism, development and environmental adaptation in Streptomyces. *The Journal of Microbiology*. 46 (1), 1–11.

Chaudhary, A., Dhakal, D. & Sohng, J. (2013) An insight into the “-omics” based engineering of streptomycetes for secondary metabolite overproduction. *BioMed research international*. 2013.

Choudoir, M., Pepe-Ranney, C. & Buckley, D. (2018) Diversification of secondary metabolite biosynthetic gene clusters coincides with lineage divergence in Streptomyces. *Antibiotics*. 7 (1), 12.

Choulet, F., Aigle, B., Gallois, A., Mangenot, S., Gerbaud, C., Truong, C., Francou, F.-X., Fourrier, C., Guérineau, M. & Decaris, B. (2006) Evolution of the terminal regions of the *Streptomyces* linear chromosome. *Molecular biology and evolution*. 23 (12), 2361–2369.

Chung, E. & Ko, K. (2019) Eradication of persister cells of *Acinetobacter baumannii* through combination of colistin and amikacin antibiotics. *Journal of Antimicrobial Chemotherapy*. 74 (5), 1277–1283.

Cogne, Y., Degli-Esposti, D., Pible, O., Gouveia, D., François, A., Bouchez, O., Eché,

C., Ford, A., Geffard, O. & Armengaud, J. (2019) De novo transcriptomes of 14 gammarid individuals for proteogenomic analysis of seven taxonomic groups. *Scientific Data*. 6 (1), 1–7.

Cortés-Albayay, C., Dorador, C., Schumann, P., Andrews, B., Asenjo, J. & Nouioui, I. (2019a) *Streptomyces huasconensis* sp. nov., an haloalkalitolerant actinobacterium isolated from a high altitude saline wetland at the Chilean Altiplano. *International journal of systematic and evolutionary microbiology*. 69 (8), 2315–2322.

Cortés-Albayay, C., Dorador, C., Schumann, P., Schniete, J., Herron, P., Andrews, B., Asenjo, J. & Nouioui, I. (2019b) *Streptomyces altiplanensis* sp. nov., an alkalitolerant species isolated from Chilean Altiplano soil, and emended description of *Streptomyces chryseus* (Krasil'nikov et al. 1965) Pridham 1970. *International journal of systematic and evolutionary microbiology*. 69 (8), 2498–2505.

Cortés-Albayay, C., Silber, J., Imhoff, J., Asenjo, J., Andrews, B., Nouioui, I. & Dorador, C. (2019c) The Polyextreme Ecosystem, Salar de Huasco at the Chilean Altiplano of the Atacama Desert Houses Diverse *Streptomyces* spp. with Promising Pharmaceutical Potentials. *Diversity*.11 (5).

D'Costa, V., King, C., Kalan, L., Morar, M., Sung, W., Schwarz, C., Froese, D., Zazula, G., Calmels, F. & Debruyne, R. (2011) Antibiotic resistance is ancient. *Nature*. 477 (7365), 457.

Davies, J. & Ryan, K. (2011) Introducing the parvome: bioactive compounds in the microbial world. *ACS chemical biology*. 7 (2), 252–259.

Dias, D., Urban, S. & Roessner, U. (2012) A historical overview of natural products in drug discovery. *Metabolites*. 2 (2), 303–336.

ECDC: European Centre for Disease Prevention and Control (2015) *Antimicrobial resistance surveillance in Europe 2014. Annual Report of the European Antimicrobial Resistance Surveillance Network*.

Edelman, G., Van Leeuwen, T. & Aalders, M. (2012) Hyperspectral imaging for the age estimation of blood stains at the crime scene. *Forensic science international*. 223 (1–3), 72–77.

Edgar, R. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*. 32 (5), 1792–1797.

Elibol, M. & Mavituna, F. (1998) Effect of sucrose on actinorhodin production by *Streptomyces coelicolor* A3 (2). *Process Biochemistry*. 33 (3), 307–311.

Elsayed, S., Trusch, F., Deng, H., Raab, A., Prokes, I., Busarakam, K., Asenjo, J., Andrews, B., van West, P. & Bull, A. (2015) Chaxapeptin, a lasso peptide from extremotolerant *Streptomyces leeuwenhoekii* strain C58 from the hyperarid Atacama Desert. *The Journal of organic chemistry*. 80 (20), 10252–10260.

Eriksson, L., Antti, H., Gottfries, J., Holmes, E., Johansson, E., Lindgren, F., Long, I., Lundstedt, T., Trygg, J. & Wold, S. (2004) Using chemometrics for navigating in the large data sets of genomics, proteomics, and metabonomics (gpm). *Analytical and bioanalytical chemistry*. 380 (3), 419–429.

Evans, M. & Dyson, P. (1993) Pulsed-field gel electrophoresis of *Streptomyces lividans* DNA. *Trends in Genetics*. 9 (3), 72.

Fajardo, A. & Martínez, J. (2008) Antibiotics as signals that trigger specific bacterial responses. *Current opinion in microbiology*. 11 (2), 161–167.

Fan, Y., Dai, Y., Cheng, Q., Zhang, G., Zhang, D., Fang, P., Wu, H., Bai, L., Deng, Z. & Qin, Z. (2012) A self-ligation method for PCR-sequencing the telomeres of *Streptomyces* and *Mycobacterium* linear replicons. *Journal of microbiological methods*. 90 (2), 105–107.

Farkašovská, J., Klucar, L., Vlček, Č., Kokavec, J. & Godány, A. (2007) Complete genome sequence and analysis of the *Streptomyces aureofaciens* phage μ1/6. *Folia microbiologica*. 52 (4), 347–358.

Fei, P., Chuan-xi, W., Yang, X., Hong-lei, J., Lu-jie, C., Uribe, P., Bull, A., Goodfellow, M., Hong, J. & Yun-yang, L. (2013) A new 20-membered macrolide produced by a marine-derived *Micromonospora* strain. *Natural product research*. 27 (15), 1366–1371.

Feng, C., Huang, Y., He, W., Cheng, X., Liu, H., Huang, Y., Ma, B., Zhang, W., Liao, C. & Wu, W. (2019) Tanshinones: First-in-Class Inhibitors of the Biogenesis of the Type 3 Secretion System Needle of *Pseudomonas aeruginosa* for Antibiotic Therapy. *ACS central science*. 5 (7), 1278–1288.

Figueras, M.J., Beaz-Hidalgo, R., Hossain, M. & Liles, M. (2014) Taxonomic affiliation of new genomes should be verified using average nucleotide identity and multilocus phylogenetic analysis. *Genome Announc.* 2 (6), e00927-14.

Flärdh, K., Leibovitz, E., Buttner, M. & Chater, K. (2000) Generation of a non-

sporulating strain of *Streptomyces coelicolor* A3(2) by the manipulation of a developmentally controlled ftsZ promoter. *Molecular Microbiology*. 38 (4), 737–749.

Fuchino, K., Flärdh, K., Dyson, P. & Ausmees, N. (2017) Cell-Biological Studies of Osmotic Shock Response in *Streptomyces* spp. *Journal of bacteriology*. 199 (1), e00465-16.

Ganz, H., Law, C., Schmuki, M., Eichenseher, F., Calendar, R., Loessner, M., Getz, W., Korlach, J., Beyer, W. & Klumpp, J. (2014) Novel giant siphovirus from *Bacillus anthracis* features unusual genome characteristics. *PLoS One*. 9 (1), e85972.

Gao, B. & Gupta, R. (2012) Phylogenetic framework and molecular signatures for the main clades of the phylum Actinobacteria. *Microbiology and Molecular Biology Reviews*. 76 (1), 66–112.

Genilloud, O. (2014) The re-emerging role of microbial natural products in antibiotic discovery. *Antonie van Leeuwenhoek*. 106 (1), 173–188.

Gilbert, W. & Maxam, A. (1973) The nucleotide sequence of the lac operator. *Proceedings of the National Academy of Sciences*. 70 (12), 3581–3584.

Gilchrist, C., Turner, S., Riley, M., Petri, W. & Hewlett, E. (2015) Whole-genome sequencing in outbreak analysis. *Clinical microbiology reviews*. 28 (3), 541–563.

Glaeser, S. & Kämpfer, P. (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy. *Systematic and applied microbiology*. 38 (4), 237–245.

Glish, G. & Vachet, R. (2003) The basics of mass spectrometry in the twenty-first century. *Nature Reviews Drug Discovery*. 2 (2), 140–150.

Goloboff, P., Farris, J. & Nixon, K. (2008) TNT, a free program for phylogenetic analysis. *Cladistics*. 24 (5), 774–786.

Gomez-Escribano, J.P., Castro, J.F., Razmilic, V., Chandra, G., Andrews, B., Asenjo, J. & Bibb, M. (2015) The *Streptomyces leeuwenhoekii* genome: de novo sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC genomics*. 16 (1), 485.

Goodfellow, M., Busarakam, K., Idris, H., Labeda, D., Nouioui, I., Brown, R., Kim, B.-Y., del Carmen Montero-Calasanz, M., Andrews, B. & Bull, A. (2017) *Streptomyces asenjonii* sp. nov., isolated from hyper-arid Atacama Desert soils

and emended description of *Streptomyces viridosporus* Pridham et al. 1958. *Antonie van Leeuwenhoek*. 110 (9), 1133–1148.

Goodfellow, M., Kämpfer, P., Busse, H.-J., Trujillo, M., Suzuki, K., Ludwig, W. & Whitman, W. (2012) *Bergey's Manual® of Systematic Bacteriology: Volume Five The Actinobacteria, Part A*. Springer New York.

Goodfellow, M., Kumar, Y., Labeda, D. & Sembiring, L. (2007) The *Streptomyces violaceusniger* clade: a home for streptomycetes with rugose ornamented spores. *Antonie van Leeuwenhoek*. 92 (2), 173–199.

Goodwin, S., McPherson, J. & McCombie, W. (2016) Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics*. 17, 333.

Graham, D., Knapp, C., Christensen, B., McCluskey, S. & Dolfing, J. (2016) Appearance of β-lactam Resistance Genes in Agricultural Soils and Clinical Isolates over the 20th Century. *Scientific reports*. 6.

Grant, J. & Stothard, P. (2008) The CGView Server: a comparative genomics tool for circular genomes. *Nucleic acids research*. 36 (suppl_2), W181–W184.

Gunde-Cimerman, N., Plemenitaš, A. & Oren, A. (2018) Strategies of adaptation of microorganisms of the three domains of life to high salt concentrations. *FEMS microbiology reviews*. 42 (3), 353–375.

Guo, X., Liu, N., Li, X., Ding, Y., Shang, F., Gao, Y., Ruan, J. & Huang, Y. (2015) Red soils harbour diverse culturable actinomycetes that are promising sources of novel secondary metabolites. *Applied and environmental microbiology*. AEM-03859.

Gupta, S., Padmanabhan, B., Diene, S., Lopez-Rojas, R., Kempf, M., Landraud, L. & Rolain, J.-M. (2014) ARG-ANNOT, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. *Antimicrobial agents and chemotherapy*. 58 (1), 212–220.

Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 29 (8), 1072–1075.

Guyet, A., Benaroudj, N., Proux, C., Gominet, M., Coppée, J.-Y. & Mazodier, P. (2014) Identified members of the *Streptomyces lividans* AdpA regulon involved in differentiation and secondary metabolism. *BMC Microbiology*. 14 (1), 81.

Harunari, E., Komaki, H., Ichikawa, N., Hosoyama, A., Kimura, A., Hamada, M. &

Igarashi, Y. (2018) Draft genome sequence of *Streptomyces hyaluromycini* MB-PO13 T, a hyaluromycin producer. *Standards in genomic sciences*. 13 (1), 2.

Harvey, A., Edrada-Ebel, R. & Quinn, R. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nature reviews drug discovery*. 14 (2), 111.

Hayakawa, Y., Nakagawa, M., Kawai, H., Tanabe, K., Nakayama, H., Shimazu, A., Seto, H. & Otake, N. (1983) Studies on the differentiation inducers of myeloid leukemic cells. *The Journal of antibiotics*. 36 (7), 934–937.

Hayase, H., Watanabe, N., Lim, C., Nogawa, T., Komatsuya, K., Kita, K. & Osada, H. (2015) Inhibition of malaria parasite growth by quinomycin A and its derivatives through DNA-intercalating activity. *Bioscience, biotechnology, and biochemistry*. 79 (4), 633–635.

van Heel, A., de Jong, A., Song, C., Viel, J., Kok, J. & Kuipers, O. (2018) BAGEL4: a user-friendly web server to thoroughly mine RiPPs and bacteriocins. *Nucleic acids research*. 46 (W1), W278–W281.

Hibbing, M., Fuqua, C., Parsek, M. & Peterson, S. (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*. 8 (1), 15–25.

Hobbs, G., Frazer, C., Gardner, D., Cullum, J. & Oliver, S. (1989) Dispersed growth of *Streptomyces* in liquid culture. *Applied microbiology and biotechnology*. 31 (3), 272–277.

Hou, Y., Tianero, M.D., Kwan, J., Wyche, T., Michel, C., Ellis, G., Vazquez-Rivera, E., Braun, D., Rose, W., Schmidt, E. & Bugni, T. (2012) Structure and Biosynthesis of the Antibiotic Bottromycin D. *Organic Letters*. 14 (19), 5050–5053.

Houston, J. (2006) Evaporation in the Atacama Desert: An empirical study of spatio-temporal variations and their causes. *Journal of Hydrology*. 330 (3), 402–412.

Hsiao, W., Wan, I., Jones, S. & Brinkman, F. (2003) IslandPath: aiding detection of genomic islands in prokaryotes. *Bioinformatics*. 19 (3), 418–420.

Huang, X. & Madan, A. (1999) CAP3: A DNA sequence assembly program. *Genome research*. 9 (9), 868–877.

Hudson, C., Lau, B. & Williams, K. (2014) Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Research*. 43 (D1),

D48–D53.

Hunt, A., Servín-González, L., Kelemen, G. & Buttner, M. (2005) The bldC Developmental Locus of *Streptomyces coelicolor* Encodes a Member of a Family of Small DNA-Binding Proteins Related to the DNA-Binding Domains of the MerR Family. *Journal of Bacteriology*. 187 (2), 716 LP – 728.

Idris, H., Labeda, D., Nouioui, I., Castro, J., del Carmen Montero-Calasanz, M., Bull, A., Asenjo, J. & Goodfellow, M. (2017) *Streptomyces aridus* sp. nov., isolated from a high altitude Atacama Desert soil and emended description of Streptomyces noboritoensis Isono et al. 1957. *Antonie van Leeuwenhoek*. 110 (5), 705–717.

Ikeda, H., Ishikawa, J., Hanamoto, A., Shinose, M., Kikuchi, H., Shiba, T., Sakaki, Y., Hattori, M. & Ōmura, S. (2003) Complete genome sequence and comparative analysis of the industrial microorganism *Streptomyces avermitilis*. *Nature biotechnology*. 21 (5), 526.

Jakimowicz, D., Gust, B., Zakrzewska-Czerwinska, J. & Chater, K. (2005) Developmental-Stage-Specific Assembly of ParB Complexes in *Streptomyces coelicolor* Hyphae. *Journal of Bacteriology*. 187 (10), 3572 LP – 3580.

Jakimowicz, D., Żydek, P., Kois, A., Zakrzewska-Czerwińska, J. & Chater, K. (2007) Alignment of multiple chromosomes along helical ParA scaffolding in sporulating Streptomyces hyphae. *Molecular Microbiology*. 65 (3), 625–641.

Jakubowska, A., Nalcacioglu, R., Millán-Leiva, A., Sanz-Carbonell, A., Muratoglu, H., Herrero, S. & Demirbag, Z. (2015) In search of pathogens: transcriptome-based identification of viral sequences from the pine processionary moth (Thaumetopoea pityocampa). *Viruses*. 7 (2), 456–479.

Jasovský, D., Littmann, J., Zorzet, A. & Cars, O. (2016) Antimicrobial resistance—a threat to the world's sustainable development. *Upsala Journal of Medical Sciences*. 121 (3), 159–164.

Jiang, H., Dong, H., Zhang, G., Yu, B., Chapman, L. & Fields, M. (2006) Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. *Applied and environmental microbiology*. 72 (6), 3832–3845.

Jiang, H. & Kendrick, K. (2000) Characterization of ssfR and ssgA, Two Genes Involved in Sporulation of *Streptomyces griseus*. *Journal of Bacteriology*. 182

(19), 5521 LP – 5529.

Kämpfer, P., Glaeser, S., Parkes, L., van Keulen, G. & Dyson, P. (2014) The Family Streptomycetaceae. In: Eugene Rosenberg, Edward F DeLong, Stephen Lory, Erko Stackebrandt, et al. (eds.). *The Prokaryotes: Actinobacteria*. Berlin, Heidelberg, Springer. pp. 889–1010.

Katz, L. & Baltz, R. (2016) Natural product discovery: past, present, and future. *Journal of industrial microbiology & biotechnology*. 43 (2–3), 155–176.

Kautsar, S., Blin, K., Shaw, S., Navarro-Muñoz, J., Terlouw, B., van der Hooft, J., van Santen, J., Tracanna, V., Suarez Duran, H., Pascal Andreu, V., Selem-Mojica, N., Alanjary, M., Robinson, S., Lund, G., Epstein, S., Sisto, A., Charkoudian, L., Collemare, J., Linington, R., Weber, T. & Medema, M. (2019) MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Research*. 48 (D1), D454–D458.

Kawai, K., Wang, G., Okamoto, S. & Ochi, K. (2007) The rare earth, scandium, causes antibiotic overproduction in *Streptomyces* spp. *FEMS Microbiology Letters*. 274 (2), 311–315.

Keller-Schierlein, W. & Richle, W. (1970) Metabolic products of microorganisms. LXXXVI. Structure of cinerubine A. *Antimicrobial agents and chemotherapy*. 10, 68.

Kersten, R., Yang, Y.-L., Xu, Y., Cimermancic, P., Nam, S.-J., Fenical, W., Fischbach, M., Moore, B. & Dorrestein, P. (2011) A mass spectrometry–guided genome mining approach for natural product peptidogenomics. *Nature chemical biology*. 7 (11), 794.

Kieser, T. (2000) *Practical streptomyces genetics*. John Innes Foundation.

Killham, K. & Firestone, M. (1984) Salt stress control of intracellular solutes in streptomycetes indigenous to saline soils. *Appl. Environ. Microbiol.* 47 (2), 301–306.

Kim, D.-W., Chater, K., Lee, K.-J. & Hesketh, A. (2005) Changes in the Extracellular Proteome Caused by the Absence of the bldA Gene Product, a Developmentally Significant tRNA, Reveal a New Target for the Pleiotropic Regulator AdpA in *Streptomyces coelicolor*. *Journal of Bacteriology*. 187 (9), 2957 LP – 2966.

Kim, J.-N., Kim, Y., Jeong, Y., Roe, J.-H., Kim, B.-G. & Cho, B.-K. (2015) Comparative genomics reveals the core and accessory genomes of *Streptomyces* species. *J*

Microbiol Biotechnol. 25 (10), 1599–1605.

Kinashi, H. (2011) Giant linear plasmids in *Streptomyces*: a treasure trove of antibiotic biosynthetic clusters. *The Journal of antibiotics*. 64 (1), 19.

Klementz, D., Döring, K., Lucas, X., Telukunta, K., Erxleben, A., Deubel, D., Erber, A., Santillana, I., Thomas, O., Bechthold, A. & Günther, S. (2016) StreptomeDB 2.0—an extended resource of natural products produced by streptomycetes. *Nucleic Acids Research*. 44 (D1), D509–D514.

Konstantinidis, K. & Tiedje, J. (2005) Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*. 102 (7), 2567–2572.

Konstantinidis, K. & Tiedje, J. (2004) Trends between gene content and genome size in prokaryotic species with larger genomes. *Proceedings of the National Academy of Sciences*. 101 (9), 3160–3165.

Krohn, K. (2009) *Anthracycline Chemistry and Biology I: Biological Occurence and Biosynthesis, Synthesis and Chemistry*. Springer.

Kronheim, S., Daniel-Ivad, M., Duan, Z., Hwang, S., Wong, A., Mantel, I., Nodwell, J. & Maxwell, K. (2018) A chemical defence against phage infection. *Nature*. 564 (7735), 283–286.

Kurosawa, K., Ghiviriga, I., Sambandan, T., Lessard, P., Barbara, J., Rha, C. & Sinskey, A. (2008) Rhodostreptomycins, antibiotics biosynthesized following horizontal gene transfer from *Streptomyces padanus* to *Rhodococcus fascians*. *Journal of the American Chemical Society*. 130 (4), 1126–1127.

Kurtböke, I. (2012) From Actinomycin onwards: Actinomycete success stories. *Microbiology Australia*. 33 (3), 108–110.

Labeda, D., Dunlap, C., Rong, X., Huang, Y., Doroghazi, J., Ju, K.-S. & Metcalf, W. (2017) Phylogenetic relationships in the family Streptomycetaceae using multi-locus sequence analysis. *Antonie Van Leeuwenhoek*. 110 (4), 563–583.

Labeda, D. & Lyons, A. (1991) The *Streptomyces violaceusniger* cluster is heterogeneous in DNA relatedness among strains: emendation of the descriptions of *S. violaceusniger* and *Streptomyces hygroscopicus*. *International Journal of Systematic and Evolutionary Microbiology*. 41 (3), 398–401.

Lambert, S., Traxler, M., Craig, M., Maciejewska, M., Ongena, M., van Wezel, G.,

Kolter, R. & Rigali, S. (2014) Altered desferrioxamine-mediated iron utilization is a common trait of bald mutants of *Streptomyces coelicolor*. *Metallomics*. 6 (8), 1390–1399.

Langille, M., Hsiao, W.W. & Brinkman, F. (2008) Evaluation of genomic island predictors using a comparative genomics approach. *BMC Bioinformatics*. 9 (1), 329.

Larkin, M., Blackshields, G., Brown, N., Chenna, R., McGettigan, P., McWilliam, H., Valentin, F., Wallace, I., Wilm, A., Lopez, R., Thompson, J., Gibson, T. & Higgins, D. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics*. 23 (21), 2947–2948.

Laxminarayan, R., Matsoso, P., Pant, S., Brower, C., Røttingen, J.-A., Klugman, K. & Davies, S. (2016) Access to effective antimicrobials: a worldwide challenge. *The Lancet*. 387 (10014), 168–175.

Leblond, P., Francou, F., Simonet, J. & Decaris, B. (1990) Pulsed-field gel electrophoresis analysis of the genome of *Streptomyces ambofaciens* strains. *FEMS microbiology letters*. 72 (1–2), 79–88.

Lemoine, F., Correia, D., Lefort, V., Doppelt-Azeroual, O., Mareuil, F., Cohen-Boulakia, S. & Gascuel, O. (2019) NGPhylogeny. fr: new generation phylogenetic services for non-specialists. *Nucleic acids research*.

Levene, M., Korlach, J., Turner, S., Foquet, M., Craighead, H. & Webb, W. (2003) Zero-mode waveguides for single-molecule analysis at high concentrations. *science*. 299 (5607), 682–686.

Lewin, G., Carlos, C., Chevrette, M., Horn, H., McDonald, B., Stankey, R., Fox, B. & Currie, C. (2016) Evolution and Ecology of Actinobacteria and Their Bioenergy Applications. *Annual Review of Microbiology*. 70 (1).

Lezhava, A., Mizukami, T., Kajitani, T., Kameoka, D., Redenbach, M., Shinkawa, H., Nimi, O. & Kinashi, H. (1995) Physical map of the linear chromosome of *Streptomyces griseus*. *Journal of bacteriology*. 177 (22), 6492–6498.

Li, Y., Pinto-Tomás, A., Rong, X., Cheng, K., Liu, M. & Huang, Y. (2019) Population genomics insights into adaptive evolution and ecological differentiation in streptomycetes. *Appl. Environ. Microbiol.* 85 (7), e02555-18.

Lim, H.-T. & Murukeshan, V. (2016) Spatial-scanning hyperspectral imaging probe for bio-imaging applications. *Review of Scientific Instruments*. 87 (3), 33707.

de Lima Procópio, R., da Silva, I., Martins, M., de Azevedo, J. & de Araújo, J. (2012) Antibiotics produced by Streptomyces. *The Brazilian Journal of infectious diseases*. 16 (5), 466–471.

Lin, J., Nishino, K., Roberts, M., Tolmasky, M., Aminov, R. & Zhang, L. (2015) Mechanisms of antibiotic resistance. *Frontiers in microbiology*. 6, 34.

Lincke, T., Behnken, S., Ishida, K., Roth, M. & Hertweck, C. (2010) Closthioamide: an unprecedented polythioamide antibiotic from the strictly anaerobic bacterium *Clostridium cellulolyticum*. *Angewandte Chemie International Edition*. 49 (11), 2011–2013.

Ling, L., Schneider, T., Peoples, A., Spoering, A., Engels, I., Conlon, B., Mueller, A., Schäberle, T., Hughes, D. & Epstein, S. (2015) A new antibiotic kills pathogens without detectable resistance. *Nature*. 517 (7535), 455–459.

Lipinski, C. (2004) Lead-and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Technologies*. 1 (4), 337–341.

Liu, G., Chater, K., Chandra, G., Niu, G. & Tan, H. (2013) Molecular regulation of antibiotic biosynthesis in *Streptomyces*. *Microbiology and Molecular Biology Reviews*. 77 (1), 112–143.

Liu, Y., Pu, H. & Sun, D.-W. (2017) Hyperspectral imaging technique for evaluating food quality and safety during various processes: A review of recent applications. *Trends in food science & technology*. 69, 25–35.

Lizama, C., Monteoliva-Sánchez, M., Suárez-García, A., Roselló-Mora, R., Aguilera, M., Campos, V. & Ramos-Cormenzana, A. (2002) *Halorubrum tebenquichense* sp. nov., a novel halophilic archaeon isolated from the Atacama Saltern, Chile. *International journal of systematic and evolutionary microbiology*. 52 (1), 149–155.

Loman, N., Constantinidou, C., Chan, J., Halachev, M., Sergeant, M., Penn, C., Robinson, E. & Pallen, M. (2012) High-throughput bacterial genome sequencing: an embarrassment of choice, a world of opportunity. *Nature Reviews Microbiology*. 10 (9), 599.

Lopes, A.S., Santa Cruz, E.C., Sussulini, A. & Klassen, A. (2017) Metabolomic strategies involving mass spectrometry combined with liquid and gas chromatography. In: *Metabolomics: From Fundamentals to Clinical Applications*. Springer. pp. 77–98.

Lown, J.W. (1993) Anthracycline and anthraquinone anticancer agents: current status and recent developments. *Pharmacology & therapeutics*. 60 (2), 185–214.

Lupo, A., Coyne, S. & Berendonk, T. (2012) Origin and evolution of antibiotic resistance: the common mechanisms of emergence and spread in water bodies. *Frontiers in microbiology*. 3, 18.

Macintyre, L., Zhang, T., Viegelmann, C., Martinez, I., Cheng, C., Dowdells, C., Abdelmohsen, U., Gernert, C., Hentschel, U. & Edrada-Ebel, R. (2014) Metabolomic tools for secondary metabolite discovery from marine microbial symbionts. *Marine drugs*. 12 (6), 3416–3448.

Mahajan, G. & Balachandran, L. (2017) Sources of antibiotics: Hot springs. *Biochemical pharmacology*. 134, 35–41.

Maki, T., Lee, K., Kawai, K., Onishi, K., Hong, C., Kurosaki, Y., Shinoda, M., Kai, K., Iwasaka, Y. & Archer, S. (2019) Aeolian dispersal of bacteria associated with desert dust and anthropogenic particles over continental and oceanic surfaces. *Journal of Geophysical Research: Atmospheres*. 124 (10), 5579–5588.

Manteca, Á. & Yagüe, P. (2018) *Streptomyces* Differentiation in Liquid Cultures as a Trigger of Secondary Metabolism. *Antibiotics*. 7 (2), 41.

Martínez, L.F., Bishop, A., Parkes, L., Del Sol, R., Salerno, P., Sevcikova, B., Mazurakova, V., Kormanec, J. & Dyson, P. (2009) Osmoregulation in *Streptomyces coelicolor*: modulation of SigB activity by OsaC. *Molecular microbiology*. 71 (5), 1250–1262.

McDonald, B. & Currie, C. (2017) Lateral Gene Transfer Dynamics in the Ancient Bacterial Genus *Streptomyces*. *mBio*. 8 (3), e00644-17.

McInerney, J., McNally, A. & O'Connell, M. (2017) Why prokaryotes have pangenomes. *Nature microbiology*. 2 (4), 17040.

Medema, M., Trefzer, A., Kovalchuk, A., van den Berg, M., Müller, U., Heijne, W., Wu, L., Alam, M., Ronning, C. & Nierman, W. (2010) The sequence of a 1.8-Mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biology and Evolution*. 2, 212–224.

Meier-Kolthoff, J., Auch, A., Klenk, H.-P. & Göker, M. (2013) Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics*. 14 (1), 60.

Meier-Kolthoff, J., Hahnke, R., Petersen, J., Scheuner, C., Michael, V., Fiebig, A., Rohde, C., Rohde, M., Fartmann, B. & Goodwin, L. (2014) Complete genome sequence of DSM 30083 T, the type strain (U5/41 T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Standards in genomic sciences*. 9 (1), 2.

Meier-Kolthoff, J., Klenk, H.-P. & Göker, M. (2014) Taxonomic use of DNA G+ C content and DNA–DNA hybridization in the genomic age. *International journal of systematic and evolutionary microbiology*. 64 (2), 352–356.

Mendez, C. & Chater, K. (1987) Cloning of whiG, a gene critical for sporulation of *Streptomyces coelicolor* A3 (2). *Journal of bacteriology*. 169 (12), 5715–5720.

Metzker, M. (2010) Sequencing technologies - the next generation. *Nat Rev Genet*. 11 (1), 31–46.

Migita, A., Watanabe, M., Hirose, Y., Watanabe, K., Tokiwano, T., Kinashi, H. & Oikawa, H. (2009) Identification of a gene cluster of polyether antibiotic lasalocid from *Streptomyces lasaliensis*. *Bioscience, biotechnology, and biochemistry*. 73 (1), 169–176.

Mindlin, S. & Petrova, M. (2017) On the Origin and Distribution of Antibiotic Resistance: Permafrost Bacteria Studies. *Molecular Genetics, Microbiology and Virology*. 32 (4), 169–179.

Minh, B.Q., Nguyen, M. & von Haeseler, A. (2013) Ultrafast Approximation for Phylogenetic Bootstrap. *Molecular Biology and Evolution*. 30 (5), 1188–1195.

Mordente, A., Meucci, E., Silvestrini, A., Martorana, G.E. & Giardina, B. (2009) New developments in anthracycline-induced cardiotoxicity. *Current medicinal chemistry*. 16 (13), 1656–1672.

Mukherjee, S., Seshadri, R., Varghese, N., Eloe-Fadrosh, E., Meier-Kolthoff, J., Göker, M., Coates, R.C., Hadjithomas, M., Pavlopoulos, G. & Paez-Espino, D. (2017) 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life. *Nature biotechnology*. 35 (7), 676.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G. & Erlich, H. (1986) Specific enzymatic amplification of DNA *in vitro*: the polymerase chain reaction. In: *Cold Spring Harbor symposia on quantitative biology*. 1986 Cold Spring Harbor Laboratory Press. pp. 263–273.

Nachtigall, J., Kulik, A., Helaly, S., Bull, A., Goodfellow, M., Asenjo, J., Maier, A.,

Wiese, J., Imhoff, J. & Süssmuth, R. (2011a) Atacamycins A–C, 22-membered antitumor macrolactones produced by *Streptomyces* sp. C38. *The Journal of antibiotics*. 64 (12), 775.

Nachtigall, J., Schneider, K., Bruntner, C., Bull, A., Goodfellow, M., Zinecker, H., Imhoff, J., Nicholson, G., Irran, E. & Süssmuth, R. (2011b) Benzoxacystol, a benzoxazine-type enzyme inhibitor from the deep-sea strain *Streptomyces* sp. NTK 935. *The Journal of antibiotics*. 64 (6), 453.

Nakata, M., Saito, M., Inouye, Y., Nakamura, S., Hayakawa, Y. & Seto, H. (1992) A new anthracycline antibiotic, cinerubin R taxonomy, structural elucidation and biological activity. *The Journal of antibiotics*. 45 (10), 1599–1608.

Navarro-Muñoz, J., Selem-Mojica, N., Mullowney, M., Kautsar, S., Tryon, J., Parkinson, E., De Los Santos, E., Yeong, M., Cruz-Morales, P., Abubucker, S., Roeters, A., Lokhorst, W., Fernandez-Guerra, A., Cappelini, L., Goering, A., Thomson, R., Metcalf, W., Kelleher, N., Barona-Gomez, F. & Medema, M. (2020) A computational framework to explore large-scale biosynthetic diversity. *Nature Chemical Biology*. 16 (1), 60–68.

Neilson, J., Quade, J., Ortiz, M., Nelson, W., Legatzki, A., Tian, F., LaComb, M., Betancourt, J., Wing, R. & Soderlund, C. (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles*. 16 (3), 553–566.

Nguyen, T., Ishida, K., Jenke-Kodama, H., Dittmann, E., Gurgui, C., Hochmuth, T., Taudien, S., Platzer, M., Hertweck, C. & Piel, J. (2008) Exploiting the mosaic structure of trans-acyltransferase polyketide synthases for natural product discovery and pathway dissection. *Nature biotechnology*. 26 (2), 225.

Nguyen, T., Kitani, S., Shimma, S. & Nihira, T. (2018) Butenolides from *Streptomyces albus* J1074 act as external signals to stimulate avermectin production in *Streptomyces avermitilis*. *Appl. Environ. Microbiol.* 84 (9), e02791-17.

Nichols, D., Cahoon, N., Trakhtenberg, E., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K. & Epstein, S. (2010) Use of ichip for high-throughput in situ cultivation of “uncultivable” microbial species. *Applied and environmental microbiology*. 76 (8), 2445–2450.

Nishida, H. (2012) Evolution of genome base composition and genome size in bacteria. *Frontiers in microbiology*. 3, 420.

Niu, G., Chater, K., Tian, Y., Zhang, J. & Tan, H. (2016) Specialised metabolites regulating antibiotic biosynthesis in *Streptomyces* spp. *FEMS microbiology reviews*. 40 (4), 554–573.

O'Neill, J. (2016) Tackling drug-resistant infections globally: final report and recommendations. *London: Wellcome Trust & HM Government*.

Ohnishi, Y., Ishikawa, J., Hara, H., Suzuki, H., Ikenoya, M., Ikeda, H., Yamashita, A., Hattori, M. & Horinouchi, S. (2008) Genome sequence of the streptomycin-producing microorganism *Streptomyces griseus* IFO 13350. *Journal of bacteriology*. 190 (11), 4050–4060.

Okada, B. & Seyedsayamdost, M. (2017) Antibiotic dialogues: induction of silent biosynthetic gene clusters by exogenous small molecules. *FEMS microbiology reviews*. 41 (1), 19–33.

Oki, T., Kitamura, I., Matsuzawa, Y., Shibamoto, N., Ogasawara, T., Yoshimoto, A., Inui, T., Naganawa, H., Takeuchi, T. & Umezawa, H. (1979) Antitumor anthracycline antibiotics, aclacinomycin A and analogues. *The Journal of antibiotics*. 32 (8), 801–819.

Okoro, C., Bull, A., Mutreja, A., Rong, X., Huang, Y. & Goodfellow, M. (2010) *Lechevalieria atacamensis* sp. nov., *Lechevalieria deserti* sp. nov. and *Lechevalieria roselyniae* sp. nov., isolated from hyperarid soils. *International journal of systematic and evolutionary microbiology*. 60 (2), 296–300.

Overbeek, R., Olson, R., Pusch, G., Olsen, G., Davis, J., Disz, T., Edwards, R., Gerdes, S., Parrello, B. & Shukla, M. (2013) The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic acids research*. 42 (D1), D206–D214.

Paphitou, N. (2013) Antimicrobial resistance: action to combat the rising microbial challenges. *International journal of antimicrobial agents*. 42, S25–S28.

Pareek, C.S., Smoczynski, R. & Tretyn, A. (2011) Sequencing technologies and genome sequencing. *Journal of applied genetics*. 52 (4), 413–435.

Park, C. & Andam, C. (2019) Within-species genomic variation and variable patterns of recombination in the tetracycline producer *Streptomyces rimosus*. *Frontiers in Microbiology*. 10, 552.

Pattengale, N., Alipour, M., Bininda-Emonds, O., Moret, B.M. & Stamatakis, A. (2009) How many bootstrap replicates are necessary? In: *Annual International*

Conference on Research in Computational Molecular Biology. 2009 Springer. pp. 184–200.

Patti, G., Yanes, O. & Siuzdak, G. (2012) Innovation: Metabolomics: the apogee of the omics trilogy. *Nature reviews Molecular cell biology*. 13 (4), 263.

Pellegrini, O., Mathy, N., Gogos, A., Shapiro, L. & Condon, C. (2005) The *Bacillus subtilis* ydcDE operon encodes an endoribonuclease of the MazF/PemK family and its inhibitor. *Molecular microbiology*. 56 (5), 1139–1148.

Pence, H. & Williams, A. (2010) *ChemSpider: an online chemical information resource*.

Penn, K., Jenkins, C., Nett, M., Udwary, D., Gontang, E., McGlinchey, R., Foster, B., Lapidus, A., Podell, S. & Allen, E. (2009) Genomic islands link secondary metabolism to functional adaptation in marine Actinobacteria. *The ISME journal*. 3 (10), 1193.

Perry, J., Waglechner, N. & Wright, G. (2016) The prehistory of antibiotic resistance. *Cold Spring Harbor perspectives in medicine*. 6 (6), a025197.

Persson, J., Chater, K. & Flärdh, K. (2013) Molecular and cytological analysis of the expression of *Streptomyces* sporulation regulatory gene whiH. *FEMS Microbiology Letters*. 341 (2), 96–105.

Petrovski, S., Seviour, R. & Tillett, D. (2013) Genome sequence and characterization of a *Rhodococcus equi* phage REQ1. *Virus genes*. 46 (3), 588–590.

Philips, F., Schwartz, H., Sternberg, S. & Tan, C. (1960) The toxicity of actinomycin D. *Annals of the New York Academy of Sciences*. 89 (2), 348–360.

Piddock, L. (2016) Reflecting on the final report of the O'Neill Review on Antimicrobial Resistance. *The Lancet Infectious Diseases*. 16 (7), 767–768.

Pluskal, T., Castillo, S., Villar-Briones, A. & Orešič, M. (2010) MZmine 2: modular framework for processing, visualizing, and analyzing mass spectrometry-based molecular profile data. *BMC bioinformatics*. 11 (1), 395.

Price, M., Dehal, P. & Arkin, A. (2010) FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one*. 5 (3), e9490.

Pritchard, L., Glover, R., Humphris, S., Elphinstone, J. & Toth, I. (2016) Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. *Analytical Methods*. 8 (1), 12–24.

Purves, K., Macintyre, L., Brennan, D., Hreggviðsson, G., Kuttner, E., Ásgeirsdóttir, M., Young, L., Green, D., Edrada-Ebel, R. & Duncan, K. (2016) Using molecular networking for microbial secondary metabolite bioprospecting. *Metabolites*. 6 (1), 2.

Rateb, M., Houssen, W., Arnold, M., Abdelrahman, M., Deng, H., Harrison, W., Okoro, C., Asenjo, J., Andrews, B. & Ferguson, G. (2011) Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. *Journal of natural products*. 74 (6), 1491–1499.

Ren, H., Liu, Y., Zhou, J., Long, Y., Liu, C., Xia, B., Shi, J., Fan, Z., Liang, Y., Chen, S., Xu, J., Wang, P., Zhang, Y., Zhu, G., Liu, H., Jin, Y., Bai, F., Cheng, Z., Jin, S. & Wu, W. (2019) Combination of Azithromycin and Gentamicin for Efficient Treatment of *Pseudomonas aeruginosa* Infections. *The Journal of Infectious Diseases*. 220 (10), 1667–1678.

Rhoads, A. & Au, K.F. (2015) PacBio sequencing and its applications. *Genomics, proteomics & bioinformatics*. 13 (5), 278–289.

Richter, M. & Rosselló-Móra, R. (2009) Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*. 106 (45), 19126–19131.

Robotti, E. & Marengo, E. (2016) Chemometric multivariate tools for candidate biomarker identification: LDA, PLS-DA, SIMCA, Ranking-PCA. In: *2-D PAGE Map Analysis*. Springer. pp. 237–267.

Romano, S., Jackson, S.A., Patry, S. & Dobson, A.D.W. (2018) Extending the “one strain many compounds” (OSMAC) principle to marine microorganisms. *Marine Drugs*.16 (7).

Romero, D., Traxler, M., López, D. & Kolter, R. (2011) Antibiotics as signal molecules. *Chemical reviews*. 111 (9), 5492–5505.

Rong, X., Liu, N., Ruan, J. & Huang, Y. (2010) Multilocus sequence analysis of *Streptomyces griseus* isolates delineating intraspecific diversity in terms of both taxonomy and biosynthetic potential. *Antonie Van Leeuwenhoek*. 98 (2), 237–248.

Rückert, C., Albersmeier, A., Busche, T., Jaenicke, S., Winkler, A., Friðjónsson, Ó.H., Hreggviðsson, G., Lambert, C., Badcock, D. & Bernaerts, K. (2015) Complete genome sequence of *Streptomyces lividans* TK24. *Journal of biotechnology*.

199, 21–22.

Rutledge, P. & Challis, G. (2015) Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature reviews microbiology*. 13 (8), 509.

Saleh, O., Flinspach, K., Westrich, L., Kulik, A., Gust, B., Fiedler, H.-P. & Heide, L. (2012) Mutational analysis of a phenazine biosynthetic gene cluster in *Streptomyces anulatus* 9663. *Beilstein journal of organic chemistry*. 8 (1), 501–513.

Salvà-Serra, F., Svensson-Stadler, L., Busquets, A., Jaén-Luchoro, D., Karlsson, R., Moore, E.R.B. & Gomila, M. (2018) A protocol for extraction and purification of high-quality and quantity bacterial DNA applicable for genome sequencing: a modified version of the Marmur procedure. *Protoc Exchange*. 10.

Sanger, F. & Coulson, A. (1975) A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *Journal of molecular biology*. 94 (3), 441–448.

Santhanam, R., Okoro, C., Rong, X., Huang, Y., Bull, A., Weon, H.-Y., Andrews, B., Asenjo, J. & Goodfellow, M. (2012a) *Streptomyces atacamensis* sp. nov., isolated from an extreme hyper-arid soil of the Atacama Desert, Chile. *International journal of systematic and evolutionary microbiology*. 62 (11), 2680–2684.

Santhanam, R., Okoro, C., Rong, X., Huang, Y., Bull, A.T., Andrews, B., Asenjo, J., Weon, H.-Y. & Goodfellow, M. (2012b) *Streptomyces deserti* sp. nov., isolated from hyper-arid Atacama Desert soil. *Antonie Van Leeuwenhoek*. 101 (3), 575–581.

Santhanam, R., Rong, X., Huang, Y., Andrews, B., Asenjo, J. & Goodfellow, M. (2013) *Streptomyces bullii* sp. nov., isolated from a hyper-arid Atacama Desert soil. *Antonie van Leeuwenhoek*. 103 (2), 367–373.

Scheffler, R., Colmer, S., Tynan, H., Demain, A. & Gullo, V. (2013) Antimicrobials, drug discovery, and genome mining. *Applied microbiology and biotechnology*. 97 (3), 969–978.

Scherlach, K. & Hertweck, C. (2006) Discovery of aspoquinolones A–D, prenylated quinoline-2-one alkaloids from *Aspergillus nidulans*, motivated by genome mining. *Organic & biomolecular chemistry*. 4 (18), 3517–3520.

Schlimpert, S., Flärdh, K. & Buttner, M. (2016) Fluorescence time-lapse imaging of

the complete *S. venezuelae* life cycle using a microfluidic device. *Journal of visualized experiments: JoVE*. (108).

Schulz, D., Beese, P., Ohlendorf, B., Erhard, A., Zinecker, H., Dorador, C. & Imhoff, J. (2011) Abenquines A–D: Aminoquinone derivatives produced by *Streptomyces* sp. strain DB634. *The Journal of antibiotics*. 64 (12), 763.

Seemann, T. (2014) Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 30 (14), 2068–2069.

Seipke, R. (2015) Strain-level diversity of secondary metabolism in Streptomyces albus. *PLoS One*. 10 (1), e0116457.

Seppey, M., Manni, M. & Zdobnov, E. (2019) BUSCO: assessing genome assembly and annotation completeness. In: *Gene Prediction*. Springer. pp. 227–245.

Sevcikova, B. & Kormanec, J. (2004) Differential production of two antibiotics of *Streptomyces coelicolor* A3 (2), actinorhodin and undecylprodigiosin, upon salt stress conditions. *Archives of microbiology*. 181 (5), 384–389.

Shen, X.-X., Zhou, X., Kominek, J., Kurtzman, C., Hittinger, C. & Rokas, A. (2016) Reconstructing the backbone of the *Saccharomycotina* yeast phylogeny using genome-scale data. *G3: Genes, Genomes, Genetics*. 6 (12), 3927–3939.

Shirling, E. & Gottlieb, D. (1966) Methods for characterization of *Streptomyces* species. *International journal of systematic bacteriology*. 16 (3), 313–340.

Silva, L. (2018) *Actinobactérias da Antártica produtoras de compostos anticâncer*. Universidade de São Paulo.

Skinnider, M., Dejong, C., Rees, P., Johnston, C., Li, H., Webster, A., Wyatt, M. & Magarvey, N. (2015) Genomes to natural products prediction informatics for secondary metabolomes (PRISM). *Nucleic acids research*. 43 (20), 9645–9662.

Slatko, B., Gardner, A. & Ausubel, F. (2018) Overview of Next-Generation Sequencing Technologies. *Current protocols in molecular biology*. 122 (1), e59.

Smith, R. & Coast, J. (2013) The true cost of antimicrobial resistance. *BMJ (Clinical research ed.)*. 346, f1493.

Soldatou, S., Eldjarn, G.H., Huerta-Uribe, A., Rogers, S. & Duncan, K. (2019) Linking biosynthetic and chemical space to accelerate microbial secondary metabolite discovery. *FEMS microbiology letters*. 366 (13), fnz142.

Stamatakis, A. (2014) RAxML version 8: a tool for phylogenetic analysis and post-

analysis of large phylogenies. *Bioinformatics*. 30 (9), 1312–1313.

Steen, J., Bannam, T., Teng, W., Devenish, R. & Rood, J. (2009) The putative coupling protein TcpA interacts with other pCW3-encoded proteins to form an essential part of the conjugation complex. *Journal of bacteriology*. 191 (9), 2926–2933.

Sugden, R., Kelly, R. & Davies, S. (2016) Combatting antimicrobial resistance globally. *Nat Microbiol*. 1, 16187.

Sullivan, M., Petty, N. & Beatson, S. (2011) Easyfig: a genome comparison visualizer. *Bioinformatics*. 27 (7), 1009–1010.

Swofford, D.L. (2002) PAUP*: Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4.0 b10. *Sinauer Associates, Sunderland*. 4, b10.

Tian, X., Zhang, Z., Yang, T., Chen, M., Li, J., Chen, F., Yang, J., Li, W., Zhang, B. & Zhang, Z. (2016) Comparative genomics analysis of *Streptomyces* species reveals their adaptation to the marine environment and their diversity at the genomic level. *Frontiers in microbiology*. 7, 998.

Tidjani, A.-R., Bontemps, C. & Leblond, P. (2020) Telomeric and sub-telomeric regions undergo rapid turnover within a *Streptomyces* population. *Scientific Reports*. 10 (1), 7720.

Tidjani, A.-R., Lorenzi, J.-N., Toussaint, M., van Dijk, E., Naquin, D., Lespinet, O., Bontemps, C. & Leblond, P. (2019) Massive Gene Flux Drives Genome Diversity between Sympatric Streptomyces Conspecifics. *mBio*. 10 (5), e01533-19.

Tiwari, K., Arora, M. & Singh, D. (2011) An assessment of independent component analysis for detection of military targets from hyperspectral images. *International Journal of Applied Earth Observation and Geoinformation*. 13 (5), 730–740.

Tomova, I., Stoilova-Disheva, M., Lazarkevich, I. & Vasileva-Tonkova, E. (2015) Antimicrobial activity and resistance to heavy metals and antibiotics of heterotrophic bacteria isolated from sediment and soil samples collected from two Antarctic islands. *Frontiers in Life Science*. 8 (4), 348–357.

Tonkin-Hill, G., MacAlasdair, N., Ruis, C., Weimann, A., Horesh, G., Lees, J., Gladstone, R., Lo, S., Beaudoin, C. & Floto, R. (2020) Producing Polished Prokaryotic Pangenomes with the Panaroo Pipeline. *BioRxiv*.

Tracanna, V., de Jong, A., Medema, M. & Kuipers, O. (2017) Mining prokaryotes for

antimicrobial compounds: from diversity to function. *Fems microbiology reviews*. 41 (3), 417–429.

Traxler, M., Watrous, J., Alexandrov, T., Dorrestein, P. & Kolter, R. (2013) Interspecies interactions stimulate diversification of the *Streptomyces coelicolor* secreted metabolome. *MBio*. 4 (4), e00459-13.

Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A. & Minh, B.Q. (2016) W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research*. 44 (W1), W232–W235.

Trujillo, M.E., Idris, H., Riesco, R., Nouioui, I., Igual, J.M., Bull, A.T. & Goodfellow, M. (2017) *Pseudonocardia nigra* sp. nov., isolated from Atacama Desert rock. *International journal of systematic and evolutionary microbiology*. 67 (8), 2980–2985.

Tsai, H.-H., Shu, H.-W., Yang, C.-C. & Chen, C. (2011) Translesion-synthesis DNA polymerases participate in replication of the telomeres in *Streptomyces*. *Nucleic acids research*. 40 (3), 1118–1130.

Tschowri, N., Schumacher, M., Schlimpert, S., Chinnam, N., Findlay, K., Brennan, R. & Buttner, M. (2014) Tetrameric c-di-GMP Mediates Effective Transcription Factor Dimerization to Control *Streptomyces* Development. *Cell*. 158 (5), 1136–1147.

Tsuji, N., Kobayashi, M., Wakisaka, Y., Kawamura, Y., Mayama, M. & Matsumoto, K. (1976) New antibiotics, griseusins A and B. *The Journal of antibiotics*. 29 (1), 7–9.

Tyc, O., Song, C., Dickschat, J., Vos, M. & Garbeva, P. (2017) The ecological role of volatile and soluble secondary metabolites produced by soil bacteria. *Trends in microbiology*. 25 (4), 280–292.

Vaidya, G., Lohman, D. & Meier, R. (2011) SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information. *Cladistics*. 27 (2), 171–180.

Varghese, N., Mukherjee, S., Ivanova, N., Konstantinidis, K., Mavrommatis, K., Kyrpides, N. & Pati, A. (2015) Microbial species delineation using whole genome sequences. *Nucleic acids research*. 43 (14), 6761–6771.

Ventura, M., Canchaya, C., Tauch, A., Chandra, G., Fitzgerald, G., Chater, K. & van Sinderen, D. (2007) Genomics of Actinobacteria: tracing the evolutionary history

of an ancient phylum. *Microbiol. Mol. Biol. Rev.* 71 (3), 495–548.

Vetsigian, K., Jajoo, R. & Kishony, R. (2011) Structure and evolution of *Streptomyces* interaction networks in soil and in silico. *PLoS biology*. 9 (10), e1001184.

Vilgis, S. & Deigner, H.-P. (2018) *Precision Medicine: Tools and Quantitative Approaches - Sequencing in Precision Medicine*. In: Hans-Peter Deigner & Matthias B T - Precision Medicine Kohl (eds.). Academic Press. pp. 79–101.

Villas-Bôas, S., Mas, S., Åkesson, M., Smedsgaard, J. & Nielsen, J. (2005) Mass spectrometry in metabolome analysis. *Mass spectrometry reviews*. 24 (5), 613–646.

Viollier, P., Kelemen, G., Dale, G., Nguyen, K., Buttner, M. & Thompson, C. (2003) Specialized osmotic stress response systems involve multiple SigB-like sigma factors in *Streptomyces coelicolor*. *Molecular microbiology*. 47 (3), 699–714.

Voelkerding, K., Dames, S. & Durtschi, J. (2009) Next-generation sequencing: from basic research to diagnostics. *Clinical chemistry*. 55 (4), 641–658.

Waack, S., Keller, O., Asper, R., Brodag, T., Damm, C., Fricke, W., Surovcik, K., Meinicke, P. & Merkl, R. (2006) Score-based prediction of genomic islands in prokaryotic genomes using hidden Markov models. *BMC Bioinformatics*. 7 (1), 142.

Waksman, S. & Woodruff, H. (1940) Bacteriostatic and Bactericidal Substances Produced by a Soil Actinomyces. *Proceedings of the society for Experimental Biology and Medicine*. 45 (2), 609–614.

Walker, B., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C., Zeng, Q., Wortman, J. & Young, S. (2014) Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one*. 9 (11), e112963.

Walsh, C. & Wencewicz, T. (2014) Prospects for new antibiotics: a molecule-centered perspective. *The Journal of antibiotics*. 67 (1), 7.

Wang, C., Mata, J.M. & Hernandez, S. (1975) *Cinerubin A and B As Antiparasitic Agents*.p.5.

Wang, M., Carver, J., Phelan, V., Sanchez, L., Garg, N., Peng, Y., Nguyen, D., Watrous, J., Kapono, C. & Luzzatto-Knaan, T. (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social

Molecular Networking. *Nature biotechnology*. 34 (8), 828.

Wang, X., Tabudravu, J., Rateb, M., Annand, K., Qin, Z., Jaspars, M., Deng, Z., Yu, Y. & Deng, H. (2013) Identification and characterization of the actinomycin G gene cluster in *Streptomyces iakyrus*. *Molecular BioSystems*. 9 (6), 1286–1289.

Warren-Rhodes, K., Dungan, J., Piatek, J., Stubbs, K., Gómez-Silva, B., Chen, Y. & McKay, C. (2007) Ecology and spatial pattern of cyanobacterial community island patches in the Atacama Desert, Chile. *Journal of Geophysical Research: Biogeosciences*. 112 (G4).

Waterhouse, R., Seppey, M., Simão, F. & Zdobnov, E. (2019) Using BUSCO to assess insect genomic resources. In: *Insect Genomics*. Springer. pp. 59–74.

Wenger, A., Peluso, P., Rowell, W., Chang, P.-C., Hall, R., Concepcion, G., Ebler, J., Fungtammasan, A., Kolesnikov, A. & Olson, N. (2019) Highly-accurate long-read sequencing improves variant detection and assembly of a human genome. *bioRxiv*. 519025.

van Wezel, G., van der Meulen, J., Kawamoto, S., Luiten, R., Koerten, H. & Kraal, B. (2000) ssgA Is Essential for Sporulation of *Streptomyces coelicolor* A3(2) and Affects Hyphal Development by Stimulating Septum Formation. *Journal of Bacteriology*. 182 (20), 5653 LP – 5662.

White, A., Kaye, C., Poupard, J., Pypstra, R., Woodnutt, G. & Wynne, B. (2004) Augmentin®(amoxicillin/clavulanate) in the treatment of community-acquired respiratory tract infection: a review of the continuing development of an innovative antimicrobial agent. *Journal of Antimicrobial Chemotherapy*. 53 (Suppl. S1), i3–i20.

Wibberg, D., Al-Dilaimi, A., Busche, T., Wedderhoff, I., Schrempf, H., Kalinowski, J. & de Orué Lucana, D. (2016) Complete genome sequence of *Streptomyces reticuli*, an efficient degrader of crystalline cellulose. *Journal of biotechnology*. 222, 13–14.

Wichner, D., Idris, H., Houssen, W., McEwan, A., Bull, A., Asenjo, J., Goodfellow, M., Jaspars, M., Ebel, R. & Rateb, M. (2017) Isolation and anti-HIV-1 integrase activity of lentzeosides A–F from extremotolerant *lentzea* sp. H45, a strain isolated from a high-altitude Atacama Desert soil. *The Journal of antibiotics*. 70 (4), 448.

Wick, R., Judd, L., Gorrie, C. & Holt, K. (2017) Unicycler: resolving bacterial genome

assemblies from short and long sequencing reads. *PLoS computational biology*. 13 (6), e1005595.

Wick, R., Schultz, M., Zobel, J. & Holt, K. (2015) Bandage: interactive visualization of de novo genome assemblies. *Bioinformatics*. 31 (20), 3350–3352.

Willcott, M. (2009) MestRe Nova. *Journal of the American Chemical Society*. 131 (36), 13180.

Williams, P., Buchanan, G., Feling, R., Kauffman, C., Jensen, P. & Fenical, W. (2005) New cytotoxic Salinosporamides from the marine actinomycete *Salinispora tropica*. *The Journal of organic chemistry*. 70 (16), 6196–6203.

Wilson, D. (2014) Ribosome-targeting antibiotics and mechanisms of bacterial resistance. *Nature Reviews Microbiology*. 12 (1), 35.

Wink, J. (2012) *Compendium of Actinobacteria: methods for the taxonomic description of Actinobacteria*. 2012.

Wohlleben, W., Mast, Y., Stegmann, E. & Ziemert, N. (2016) Antibiotic drug discovery. *Microbial biotechnology*. 9 (5), 541–548.

World Health Organization (2017) *Global priority list of antibiotic-resistant bacteria to guide research, discovery, and development of new antibiotics.*

World Health Organization (2015) *Worldwide country situation analysis: response to antimicrobial resistance*.

Wright, E. & Vetsigian, K. (2016) Inhibitory interactions promote frequent bistability among competing bacteria. *Nature communications*. 7, 11274.

Wright, G. (2005) Bacterial resistance to antibiotics: enzymatic degradation and modification. *Advanced drug delivery reviews*. 57 (10), 1451–1470.

Wu, C., Du, C., Gubbens, J., Choi, Y. & van Wezel, G. (2015a) Metabolomics-driven discovery of a prenylated isatin antibiotic produced by *Streptomyces* species MBT28. *Journal of natural products*. 78 (10), 2355–2363.

Wu, C., Kim, H., van Wezel, G. & Choi, Y. (2015b) Metabolomics in the natural products field–a gateway to novel antibiotics. *Drug Discovery Today: Technologies*. 13, 11–17.

Wyche, T., Piotrowski, J., Hou, Y., Braun, D., Deshpande, R., McIlwain, S., Ong, I., Myers, C., Guzei, I., Westler, W., Andes, D. & Bugni, T. (2014) Forazoline A: Marine-Derived Polyketide with Antifungal In Vivo Efficacy. *Angewandte Chemie*

International Edition. 53 (43), 11583–11586.

Xu, F., Nazari, B., Moon, K., Bushin, L. & Seyedsayamdost, M. (2017) Discovery of a cryptic antifungal compound from *Streptomyces albus* J1074 using high-throughput elicitor screens. *Journal of the American Chemical Society*. 139 (27), 9203–9212.

Xu, L., Dong, Z., Fang, L., Luo, Y., Wei, Z., Guo, H., Zhang, G., Gu, Y., Coleman-Derr, D. & Xia, Q. (2019) OrthoVenn2: a web server for whole-genome comparison and annotation of orthologous clusters across multiple species. *Nucleic acids research*. 47 (W1), W52–W58.

Xu, P., Li, W.-J., Tang, S.-K., Zhang, Y.-Q., Chen, G.-Z., Chen, H.-H., Xu, L.-H. & Jiang, C.-L. (2005) *Naxibacter alkalitolerans* gen. nov., sp. nov., a novel member of the family 'Oxalobacteraceae' isolated from China. *International journal of systematic and evolutionary microbiology*. 55 (3), 1149–1153.

Yang, C.-C., Tseng, S.-M. & Chen, C. (2015) Telomere-associated proteins add deoxynucleotides to terminal proteins during replication of the telomeres of linear chromosomes and plasmids in *Streptomyces*. *Nucleic acids research*. 43 (13), 6373–6383.

Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H. & Hirano, K. (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nature genetics*. 48 (8), 927.

Yim, G., Wang, H. & Davies, J. (2007) Antibiotics as signalling molecules. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 362 (1483), 1195–1200.

Yoon, S.-H., Ha, S.-M., Kwon, S., Lim, J., Kim, Y., Seo, H. & Chun, J. (2017) Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *International journal of systematic and evolutionary microbiology*. 67 (5), 1613.

Yoon, V. & Nodwell, J. (2014) Activating secondary metabolism with stress and chemicals. *Journal of industrial microbiology & biotechnology*. 41 (2), 415.

Zhou, Z., Gu, J., Li, Y.-Q. & Wang, Y. (2012) Genome plasticity and systems evolution in *Streptomyces*. *BMC Bioinformatics*. 13 (10), S8.

Ziemert, N., Podell, S., Penn, K., Badger, J., Allen, E. & Jensen, P. (2012) The natural

product domain seeker NaPDoS: a phylogeny based bioinformatic tool to classify secondary metabolite gene diversity. *PloS one*. 7 (3), e34064.