Developing a sampling strategy for Antibiotic Bioprospecting from bacteria in Terrestrial Ecosystems.

Patrick Hill

A thesis submitted in partial fulfillment of the requirements for the Doctorate in Philosophy degree in Biology

Department of Biology, Ottawa-Carleton Institute of Biology
Faculty of Science
University of Ottawa

© Patrick Hill, Ottawa, Canada, 2023

Abstract

This thesis investigates the problem of how to sample terrestrial environments for bioprospecting for pharmaceutically useful bacterial secondary metabolites. There is already a considerable literature on the ecology of bacteria in terrestrial ecosystems as well as on attempts to bioprospect from particular terrestrial environments. The first chapter reviews both of these kinds of studies and outlines unresolved questions that need to be answered to efficiently sample for bioprospecting.

Most studies of bacterial communities in soil use molecular methods and start by extracting DNA by bead beating. The extracted DNA is then PCR amplified using 16S bacterial primers. Bead beating is usually done with the Mobio kit, however, there is no one standard 16S primer pair that is used in most studies. The second chapter tests how comparible results from these studies are. DNA was extracted from soil and compost using the Mobio kit and a contrasting method and amplified with four different primer pairs. Results showed that primers have more effect on results than extraction method but that the Mobio kit underextracts several genera by orders of magnitude.

Most antibiotics from bacteria that are in use come from the Actinobacterial phylum. The third chapter uses several primer pairs for Sanger sequencing and bacterial community fingerprinting to compare the Actinobacterial community structure of a range of environments under different Anthropogenic impact. It found that street sediments were enriched in Actinomycetes compared to soil.

In the fourth chapter, environmental DNA from soil and street sediments was amplified with primers for type I polyketide synthase (PKSI) genes. Amplicons were sequenced with Sanger sequencing. Several PKSI amplicons were common to soil or street sediment from 1000s of km apart. Soil PKSI sequences were more often non Actinomycetal and more novel than street PKSI sequences.

Acknowledgments

My mother originally encouraged me to pursue my interests in Science and was indirectly the reason for enrolling as a doctoral student at the University of Ottawa. She died in 2017, but would have enjoyed reading this work. As someone capable enough to begin her career as a secretary and end it running an Economic research consultancy, she would probably have understood most of it.

Thanks as well to Maggie Scott, a fellow resident on my mothers floor who was a friend to both of us.

The seeds of this research began, when I was working on Remote Sensing and Carbon sequestration in soil in Colombia. I became interested in a third project, developing a DNA extraction method from soil working with a biochemist, Peter Wenzl. The soils that we were working on were organic matter rich clayey vertisols, so what was intended as a side project that would take one or two months of part time work ended up taking over a year. In spite of this I had the unwavering support of Dr Edgar Amézquita Collazos. Outside work, a close friend known to all as “Henry el mono” also provided emotional support.

In the Czech Republic I had a post helping setting up soil DNA extraction and the beginnings of soil molecular biology techniques. I would like to thank Dr. H. Šantrůčková and Ota Rauch with assistance. Dr Lubbert Dijkhuizen, from the Netherlands, helped further develop this work.

Dr Václav Krištůfek made my stay in the Czech Republic possible. He helped me understand the Czech Republic and has a very broad knowledge of both soil biology and agronomy that is rare in today’s world of academic specialization. I would also like to apologize to him for calling him Vaklav rather than Václav for the first months of my stay there. Typically for Václav, he was too polite to correct me.

Family circumstances meant that I had to return to Ottawa, the city where I grew up. I wrote to Dr John Arnason asking for advice about where I could do a Ph.D. on the distribution of bacterial secondary metabolites in the environment. He said that my ideas were interesting but that I would be better off talking to a recent arrival in the Chemistry Department, Dr Christopher Boddy.

The Boddy laboratory has been the best place that I have worked in my life. A list of everyone in the laboratory who made my stay pleasant, undergraduates, graduate students and post docs would be too long to be convenient and I would run the risk of missing names by mistake. I can only say thank you to all of you.

Chris Boddy has been an excellent supervisor and shown many examples of both his professional and ethical qualifications as a scientist. Most of all he has genuine scientific curiosity that goes beyond his career needs and meant that working with him gave me a broad view of science.

Finally, I would like to thank John Phillip Melville for help in editing this document.

Table of contents

Abstract ii

Acknowledgements iii

Table of contents v

Acronyms and abbreviations x

List of Figures xii

List of Tables xvii

Introduction Introduction 1

References 10

Chapter 1 Literature Review 13

Outline 13

Abstract 15

1 Introduction 16

2 Unresolved Issues in Microbial Ecology that Limit Bioprospecting for New Polyketides 23

2.1. Data Comparability 23

2.2. Soil Bacterial Habitats Do Not Correspond to Eukaryotic Habitats 25

2.3. Actinobacterial Soil Community Structure May Not Follow Bacterial Soil Community Structure 28

2.4. Non-Actinobacterial Bacterial Polyketide Producers 31

2.5. The Question of Biogeography 33

3. Bioprospecting for New Polyketide Discovery 39

3.1. Eukaryotic Associated Bacteria 40

3.2. Insect-Associated Bacteria 40

3.3. Social Insects as Sources for Bacterial Polyketide Discovery 41

3.4. Non-Social Insects as Sources for Polyketides 51

3.5. Desert Soils 55

3.6. Disease Suppressive Sediments 63

3.7. Other Unique Environments 67

3.7.1. Caves 67

3.7.2. Extremophiles 71

3.8. Potential New Environments for Bioprospecting 73

3.8.1. Cities 74

3.8.2. Airborne Bacteria 75

4. Conclusions 76

4.1. Taking Samples 76

4.2. Characterising Samples 77

4.3. Non Actinomycetal Polyketide Producers 79

Acknowledgements 80

References 80

Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

Outline 91

Abstract 93

Importance 94

Introduction 94

Results and Discussion 96

Results and Discussion Minor effects of extraction method on Alpha and Beta diversity compared to primers and sample. 98

Results and Discussion STAMP analysis shows differential extraction of Enterobacteriales and Pseudomonades 102

Results and Discussion Extraction efficiency 108

Conclusion 111

Methods. 113

Methods Sample sites. 113

Methods Sample collection and storage 113

Methods DNA Extraction 113

Methods Measuring DNA concentration 116

Methods Sequencing Overview 117

Methods Pyrosequencing 117

Methods Illumina Sequencing 118

Methods Sequence Processing. 119

Methods Alpha and Beta diversity analysis with Qiime 120

Methods 16S library comparison with Library compare/STAMP 120

Accession Numbers 121

Acknowledgements. 121

References 122

Chapter 3. Investigation of how the Actinobacterial community structure changes across land use.

Outline 129

Abstract 131

Introduction 131

Methods 134

Methods Sampling for Characterisation with Actinobacterial-Specific Primers 134

Methods Samples for Characterisation with Eubacterial Primers 135

Methods DNA Extraction from Soils and Sediments 137

Methods 16S rDNA Primers 138

Methods PCR Amplification-F-Act/R-Bact 138

Methods PCR Amplification—F-Act/R-Act 139

Methods PCR Amplification—Bacterial Primers 139

Methods Cloning and Sequencing from F-Act/R-Bact Primer Amplicons 140

Methods Cloning and Sequencing from F-Act/R-Act and Bacterial Primers Amplicons 140

Methods Sequence Data analysis and Accession Numbers 141

Methods ARDRA F-Act/R-Bact 142

Methods T-RFLP F-Act/R-Act 142

Methods Analysis of ARDRA and T-RFLP Patterns 142

Results 143

Results Sequences Generated Using the Single 16S rDNA Actinobacterial-Specific Primer System (F-Act/R-Bact) 143

Results Sequences Generated Using the 16S rDNA Double Actinobacterial-Specific Primers (F-Act/R-Act) 145

Results ARDRA Patterns Generated Using the 16S rDNA Single Actinobacterial-Specific Primer System (FAct/R-Bact) 147

Results T-RLFP Patterns Generated Using the 16S rDNA Double Actinobacterial (F-Act/R-Act) Primers 149

Results Reverse T-RFLP 149

Results Forward T-RFLP 150

Results Canonical Correspondence Analysis of Fingerprints 151

Results Sequences Generated with the Eubacterial Primers of Marchesi et al.154

Discussion 156

Acknowledgements 160

References 160

Chapter 4 Comparing the Type I polyketide synthase pathways of soil and street sediments.

Outline 164

Abstract 166

Introduction 167

Methods 168

Methods Sampling and DNA extraction 168

Methods PCR amplification 170

Methods Cloning and Sequencing 170

Methods Sequence Analysis 170

Results 172

Results Sequence length and GC ratios 172

Results AT and KS tree topology 172

Results Soil and street sediments are distinct PKSI habitats 176

Results Earthworm-specific clades 178

Results Sequence comparisons 179

Discussion 180

Discussion Many common environmental PKSIs are habitat-specific but widely distributed 182

Acknowledgments 183

References 184

Chapter 5 Conclusions 187

References 193

Appendix 1 Chapter 2 Supplementary Data 195

Appendix 2 Chapter 3 Supplementary Data 205

Appendix 3 Chapter 4 Supplementary Data 213

Appendix 4 Supplemental Tables Chapter 2 219

Appendix 5 Supplemental Tables Chapter 4 259

List of acronyms and abbreviations (explanations for those that are less common or unique to this thesis)

ACP Acyl carrier protein

ARDRA Amplified Ribosomal DNA Restriction Analysis (Bacterial fingerprinting method based on restriction analysis).

ASV Amplicon Sequence Variant (An alternative to OTUs used in microbial ecology, that uses the exact sequence rather than percentage similarity)

AT domain Acyltransferase domain.

bp Base pair.

CTAB Cetyltrimethylammonium bromide

DH Dehydratase

DGGE Denaturing Gradient Gel Electrophoresis (Bacterial community fingerprinting method based on GC content)

DMSO Dimethylsulfoxide

EL External Laboratory (Four External Laboratories were used for sequencing amplicons in Chapter 2).

ER Enoylreductase

F-Act/R-Act Forward Actinobacterial Specific and Reverse Actinobacterial Specific 16S PCR primers. (Used for Sanger sequencing and T-RFLP fingerprinting in Chapter 3)

F-Act/R-Bact Forward Actinobacterial Specific and Reverse Bacterial Specific 16S PCR primers. (Used for Sanger sequencing and ARDRA fingerprinting in Chapter 3)

FDR False discovery rate

KR Ketoreductase

KS Ketosynthase

Mbt Mycobactin (A mycobacterial siderophore, Chapter 4)

MEGA Molecular Evolutionary Genetics Analysis (Treeing software used in Chapter 3)

MW Molecular Weight

MUSCLE MUltiple Sequence Comparison by Log-Expectation (Sequence alignment software used in Chapter 3)

NRP Non-ribosomal peptide synthetase

OTU Operational Taxonomic Unit.

PCoA Principal Coordinate Analysis (Multidimensional Scaling method used to display data in Chapters 2 and 3)

PEG Polyethylene glycol

PLFA Phospholipid fatty acids (PFLA analysis is used to characterize microbial communities)

PKS Polyketide synthase.

SOSED Soil and sediment group (Non Actinomycetal, Actinobacterial clade identified in chapter 3).

STAMP Statistical analysis of taxonomic and functional profiles (Used to identify bacterial clades that were disproportionately from either soft of bead beating extracted DNA in chapter 2)

TE Thioesterase

T-RFLP Terminal Restriction Fragment Length Polymorphism (Bacterial community fingerprinting method based on restriction analysis)

UPGMA Unweighted pair Group Method with Arithmetic mean.

List of Figures

Chapter 1 Literature Review

Figure 1.1. Microbial ecology approaches used to examine both cultivatable and uncultivatable fraction of the microbial community structure. Page 19

Figure 1.2. The overlap between Sanger sequencing results of three 16S clone libraries at 99% identity from a beach sand using two DNA extraction methods and two primer pairs. After Hong, S.; Bunge, J.; Leslin, C.; Jeon, S.; Epstein, S.S. Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J. 2009, 3, 1365–1373. Page 24

Figure 1.3. Bacterial 16S 454 pyrosequencing results from 88 non-agricultural soils from across North and South America. Modified from Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. Appl. Environ. Microbiol. 2009, 75, 5111–5120 Page 27

Figure 1.4. Forward T-RFLP community fingerprints of the Actinobacteria from Uncultivated soil (blue) After Hill, P.; Krištůfek, ; Dijkhuizen, L.; Boddy, C.; Kroetsch, D.; Van Elsas, J.D. Land use intensity controls actinobacterial community structure. Microb. Ecol. 2011, 61, 286–302. Page 31

Figure 1.5. Overlap of sequences NRP, PKSI and PKSII sequences amplified from three soils from the desert sites in Arizona), Utah and California. After Reddy, B.V.B.; Kallifidas, D.; Kim, J.H.; Charlop-Powers, Z.; Feng, Z.; Brady, S.F. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. Appl. Environ. Microbiol. 2012, 78, 3744–3752. Page 37

Figure 1.6. (A) Leaf cutter ant colonies in Texas. Photo courtesy of Texas A&M AgriLife Extension/Josh Blanek; (B) Leaf cutting ant nest in Costa Rica. Photograph courtesy of Herster Barres, Reforest the Tropics. Page 42

Figure 1.7. Polyketide and non-ribosomal peptide natural products isolated from bacteria associated with social ants. Page 44

Figure 1.8. Microtermolides, produced by termite-associated microbes, and the originally proposed and revised structure of the related NRP vinylamycin. Page 47

Figure 1.9. The known antibiotic geldanamycin and the related analogs discovered from a termite associate Streptomyces Page 48

Figure 1.10. Polyketides discovered from beetle associated bacteria and known related compounds Page 50

Figure 1.11. Beewolf with honey bee prey. Courtesy of Simon Jenkins http://www.simon-jenkins. Photography Page 52

Figure 1.12. Bacterial secondary metabolites associated with non-social wasp species. Page 53

Figure 1.13. Tripartilactam and pederin from non-social beetles and related natural products from non-insect associated microbes. Page 54

Figure 1.14. Type II polyketide molecules coded for by desert soil metagenomic clones. After Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. After Feng, Z.; Kallifidas, D.; Brady, S.F. Proc. Natl. Acad. Sci. USA. 2011, 108, 12629–12634. Page 57

Figure 1.15. Pentangular polyphenols discovered from desert soil derived metagenomic libraries Page 59

Figure 1.16. Natural products from desert isolates. Page 62

Figure 1.17. Natural products from disease suppressive soils. Page 66

Figure 1.18. Cervimycins, isolated from cave derived Actinomycetes, are related to known tetracycline antibiotics 74 and 75. Page 69

Figure 1.19. Type II polyketides isolated from an Actinomycetes isolate from Hardin's cave. Page 69

Figure 1.20. (A) An example of Moonmilk (above) in Goatherds Chasm in Switzerland. Photograph provided by Olivier Gallois of the Groupe Spéléologique Archéologique Mandeure; (B) Yellow Microbial mats from the volcanic cave Gruta de Terra Mole in the Azores. Photo courtesy of Pedro Cardoso. Page 70

Figure 1.21. Polyketides discovered from cultured Actinomycete strains isolated from highly acidic mine drainage and highly basic mine tailings. Page 72

Figure 1.22. Polyketides identified from Actinopolyspora erythraea Page 73

Figure 1.23 Average percent composition at the phylum-level classification of bacterial communities at the soil surface or at 1.5, 4.5, 7.5, or 18 m above the surface. Figure as originally published in Weber, C.F.; Werth, J.T. Is the lower atmosphere a readily accessible reservoir of culturable, antimicrobial compound-producing Actinomycetales? Front. Microbiol.2015, Page 76

Figure 1.24. An example of actinobacterial 16S fingerprinting. Forward and reverse Terminal Restriction Fragment Length Polymorphism (T-RFLP) from a Sanger sequencer for a Russian taiga forest soil (RusforestYenisei). Page 79

Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

Figure 2.1. Primers affected the apparent beta diversity of soils more than extraction method Page 99

Figure 2.2. Varying primers and extraction method does not alter overall community differences between soils and other biomes. Page 100

Figure 2.3. Extraction method affects abundance of different phylogenetic groups from the V4-V5 primer datasets for 11 soils and three compost samples. Page 104

Figure 2.4. Comparisons of extraction method give different results with different primers and PCR conditions. Examples of STAMP Extended error bar plots at the genus level of soil Exp1B comparing extraction methods from four external sequencing laboratories and primer pairs. Page 106

Figure 2.5. The Bacillales order and Stenotrophomonas genus were preferentially extracted by Soft extraction in two soils. STAMP Extended error bar plots at the genus and order levels of soils Exp2B and Exp4A. Page 107

Figure 2.6. Preferential extraction of the Streptophyta order in compost. STAMP Extended error bar plots at the order level from the top layer of garden compost 80F Page 108

Chapter 3. Investigation of how the Actinobacterial community structure changes across land use.

Figure 3.1 Neighbour-joining tree of 16S sequences generated with the F-Act/R-Bact primers of Heueretal. Page 144

Figure 3.2 Neighbour-joining tree of 16S clones generated with the F-Act/R-Act primers of Monciardini et al. Page 146

Figure 3.3 UPGMA tree of amplified ribosomal DNA restriction analysis (ARDRA) patterns generated using the F-Act/R-Bact primers of Heuer et al. [15] and TaqI digestion. Page 148

Figure 3.4 UPGMA tree of terminal restriction fragment polymorphism (T-RFLP) patterns generated using the A3R primer of Monciardini et al. [30] F-Act/R-Act primers and HhaI digestion

Page 150

Figure 3.5 UPGMA tree of terminal restriction fragment polymorphism (T-RFLP) patterns generated using the F243 primer of Monciardini et al's[30] F-Act/R-Act primers and HhaI digestion.

Page 151

Figure 3.6 Redundancy analysis ordination plots of fingerprint patterns for all characterised samples.

Page 153

Figure 3.7 Neighbour-joining tree of Actinobacterial 16S clones generated with the F-Act/R-Act primers of Marchesi et al. Page 155

Figure 3.8 Unifrac Jackknife environmental clustering of the 12 eubacterial clone libraries generated with the primers of Marchesi et al. [28] using the weighted Unifrac algorithm. Page 156

Chapter 4 Comparing the Type I polyketide synthase pathways of soil and street sediments

Fig. 4.1 Rooted maximum likelihood phylogenetic amino acid tree of AT domains. Page 174

Fig. 4.2 Rooted maximum likelihood phylogenetic amino acid tree of KS domains. Page 176

List of Tables

Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

Table 2.1 Sample description and amplicon sequencing plan Page 97

Chapter 3. Investigation of how the Actinobacterial community structure changes across land use.

Table 3.1 Soil Samples and sequenced clones Page 136

Table 3.2 Insect, Earthworm and Urban sediments and sequenced clones Page 137

Table 3.3 Gradient lengths and Monte Carlo Permutation Values Page 152

Table 3.4 Eubacterial clones sequenced. Page 155

Chapter 4 Comparing the Type I polyketide synthase pathways of soil and street sediments

Table 4.1 Soil Samples Page 169

Table 4.2 Street Sediment and Earthworm affected Samples Page 169

Table 4.3 Reeve's signature model for AT domain motifs Page 180

Introduction

Infectious diseases were the main cause of death for most of human history (Smith, Watkins, & Hewlett, 2012). In the mid 19th century this began to change as the role of microbial infection and the importance of sanitation and aseptic medical techniques was understood (Cruckshank, 1944) (Smith et al., 2012). A second decrease in deaths from infectious disease began with the discovery of penicillin, an anti-bacterial compound. While penicillin is produced by the fungal *Penicillium* genus, soon after its discovery bacteria that produced anti-infective compounds were also identified (Chain, 1944). From the 1940s until the 1980s (the golden age of antibiotics) bacteria isolated from the environment were a source of new compounds with antibacterial and antifungal activity. These compounds are known as antibiotics.

Antibiotics are secondary metabolites. Secondary metabolites are specific to particular organisms. They are low weight (MW<3000), very structurally diverse, and it is often unclear why organisms produce them (Bérdy, 2005). While antibiotics act on a range of biological targets, the majority of them in bacteria are the products of three sets of biochemical pathways, the polyketides, non-ribosomal peptides and aminoglycosides.

The bacterial isolates which produced antibiotics during the “Golden Age” were not distributed equally across the bacterial subdivisions. The Actinomycetes, an order within the phylum Actinobacteria, were the best known antibiotic producers and within the Actinomycetes 90 % of known antibiotics came from a single genus: the *Streptomyces* (Watve *et al* (2001). Several other bacterial division also produced antibiotics. Gerth *et al* (2003) describe the main producers of secondary metabolites as: the Actinobacteria (70 %), the fungi (18%), the bacilli (6.5%) and the Myxobacteria (5 %). Bérdy (Bérdy, 2005) also lists the cyanobacteria and *Pseudomonas* as important producers of secondary metabolites.

Strohl (2004) suggests the Gamma proteobacteria and Bacilli as good antibiotic producers. Antibiotic producing bacteria usually have large genomes (> 5 Mb), and appear to be generalists rather than specialists, occupying a wide variety of environments (Davies, 1999).

Bacteria that were resistant to penicillin were found within a few years after it's discovery (Ventola, 2015). Bacterial resistance to antibiotics can be found in nature and has increased as antibiotics were overused. Now antibiotic resistant bacterial pathogens, particularly Gram negative bacteria, are a growing problem killing 23 000 a year in the U.S.A. alone (Dhingra, Rahman, Peile, & Rahman, 2020).

While antibiotics that we have are less and less effective because of resistance, fewer antibiotics are being discovered. Since the 1980s, the rate of discovery of novel antibiotics has slowed. This is only partly a scientific problem. Economic and regulatory reasons make pharmaceutical companies reluctant to invest in antibiotic drug discovery (Projan 2003). Antibiotics are not as profitable as pharmaceuticals that are used to treat long term, chronic conditions. Under present regulatory regimes, antibiotics such as penicillin and gentamicin would probably not make it to market given their side effects. It is considered unethical to test antibiotics in clinical trials against placebos. Instead, comparisons must be made with other antibiotics, this means that many more patients must be enrolled in these trials increasing costs (Wright, 2014). Furthermore, once an antibiotic is discovered there is economic pressure to use it as much as possible before patents expire, which encourages the rapid development of antibiotic resistance (Fernandes, 2006). Angell (Angell, 2005) was less sympathetic to the problems of pharmaceutical companies, doubting their commitment to research.

Even the most public spirited pharmaceutical researchers, working under ideal economic and regulatory conditions, would still face a clear scientific challenge. Antibiotic producing microorganisms were first isolated for medical use in the 1940s. Since the early 1980s, fewer and fewer new antibiotics

have been found through traditional isolation techniques, and the fraction of antibiotics found that are what Watve *et al* (2001) called 'old friends' has increased.

There was dispute as to whether there is any point in further bioprospecting for antibiotics in the environment. Watve *et al* (2001) suggested that this decline in discovery was simply because of less investment starting a vicious circle, the drop in the discovery rate in the early 1980s led to less work screening isolates, which in turn resulted in fewer novel antibiotics being discovered.

Others suggested that this was because that many antibiotics have already been discovered, Strohl (2004) noting that while new antibiotics are continually being discovered, no new bacterially derived chemical scaffolds that have led to novel marketable antibiotics have been found since 1981. Because of this, there has been research into alternatives to antibiotics and in new methods of antibiotic discovery. While developing new antibiotics *de novo* artificially have not proved successful (Bérdy, 2005), several alternatives to antibiotics seem promising.

In the early 20th century bacteriophages were used to fight bacterial infections. In western countries they fell out of use with the introduction of antibiotics however, they continued to be used in Eastern Europe and the Soviet Union. Much of the literature on their use was never published in English and is reviewed by Sulakvelidze et al (Sulakvelidze, Alavidze, & Morris, 2001). Phage therapy has more recently been reviewed by Altamirano and Barr (Altamirano & Barr, 2019). Phages are highly specific and only target a narrow range of bacteria, this means that the bacterial community is less disturbed. However, It also means that pathogens have to be identified before they can be used. In contrast anti-bacterial antibiotics are usually broad spectrum. This means that the pathogen does not have to be identified exactly, but also that the antibiotic can cause further diseases by altering the gut microbiome (Wright, 2014).

Resistance to bacteriophages can develop quickly, this can be countered by using cocktails or sequences of bacteriophages. If bacteriophage binding sites are on antibiotic efflux pumps, developing bacteriophage resistance can reduce antibiotic resistance. Increasingly bacteriophages are used in combination with antibiotics because of synergistic affects (Altamirano & Barr, 2019).

Anti-microbial peptides also known as Host Defense peptides are short (12-50 amino acid) cation peptides that break down bacterial membranes. They are produced by a broad range of organisms, most of known ones come from animals, in particular amphibians (Mahlapuu, Björn, & Ekblom, 2020). In contrast to bacteriophages, host defense peptides have a broader range of activity than antibiotics and are often limited by their toxicity. They typically have less activity than antibiotics and are limited by *in vivo* proteolytic breakdown. There is some evidence that bacteria can develop resistance to them although this seems less of a problem than for antibiotics (Shu, Ting, Beuerman, & Dua, 2020).

Those who have kept looking for new antibiotics have increasingly abandoned the old method of cultivating Actinomycetes from soil and tried one or more of three new approaches.

1-Looking for antibiotic production genes in already cultivated bacteria that are not expressed in standard cultivation conditions.

With the development of cheaper and faster sequencing it became possible to quickly sequence bacterial genomes. Sequencing the genomes of known antibiotic producers meant that the pathways that produced these antibiotics were better understood, it also found that for every antibiotic that the bacteria were known to produce there were often ten to twenty more antibiotic producing pathways in their genomes (Rutledge & Challis, 2015). As well as this, many pathways were found in bacteria that were not known to product antibiotics (Gavriilidou et al., 2022).

Researchers have attempted to express these “cryptic or silent” genes by changing cultivation techniques (Bode, Bethe, Hofs, & Zeeck, 2002) or by expressing the genes heterologously in easily controlled expression systems (Stevens et al., 2013).

2-Changes in methods of extracting antibiotic production genes from soil.

During the “Golden age of Antibiotic discovery” antibiotic producing bacteria were found by isolating Actinomycetes from soil. With time, more and more isolates had to be screened to find new antibiotics. Further scaling up cultivation, so that tens of millions of soil isolates are screened against bacteriaresistant to most common soil antbiotics might find even rarer actinomycetal antibiotic producers (Baltz, 2006).

The overwhelming majority of bacteria in the environment do not grow in the laboratory conditions used to isolate antibiotic producers so far. Cultivation methods can be changed to encourage the growth of a different small group of bacteria that are potentially good antibiotic producers. For example, normally most Actinomycetes isolated from soil are from the *Streptomycete* genus. Novel Actinomycetal genera such as the *Catenulispora* and *Actinospica* can be isolated by changing the pH of cultivation media (Busti et al., 2006).

Alternatively, cultivation techniques can be designed to isolate the unknown majority of soil bacteria. This most successful way of doing this has been the iChip. This involves keeping bacteria in a chamber buried with membrane walls that allow chemical contact with bacteria in soil for weeks to months (Nichols et al., 2010). It was used to isolate a novel NRP proteobacterial antibiotic producer (Ling et al., 2015).

Finally, rather that cultivating soil bacteria, metagenomics DNA from the whole soil microbial community can be extracted from soil and used to make metagenomics libraries. These libraries can

then be screened for antibiotic production. A successful example of this is work by the Brady laboratory on desert soils (Kang & Brady, 2014).

3-Sampling other environments than soils.

In the last 20 years there has been more antibiotic bioprospecting in new environments such as hyper arid desert soils (Castro et al., 2015), in insect communities (Carr, G.; Poulsen, M.; Klassen, J.L.; Hou, Y.; Wyche, T.P.; Bugni, T.S.; Currie, C.R.; Clardy, 2012) (Haeder, Wirth, Herz, & Spiteller, 2009), caves (Cheeptham et al., 2013) and marine sediments (Gerwick & Fenner, 2013).

Deciding how to sample the environment for microbial pharmaceuticals is a question that has become more difficult. Tribe (Tribe, 1998) describes how pharmaceutical company Sandoz bioprospected for anti-fungal antibiotics is the 1960s and 70s. When employees went on holidays or business trips around the world, they would take plastic bags with them and sample soils. Sampling was thus random and geographically widespread.

This kind of sampling would be impractical today. Beginning in the early 1990s, international agreements have increasingly regulated bioprospecting. The most important one was the Nagoya Protocol signed in 2010 which covered national territory. Most recently, in March of 2022 the UN Treaty on the Conservation and Sustainable use of Marine Biodiversity Beyond National Jurisdiction (the High Seas Treaty) held it's fourth session and set up a timetable for a fifth session later this year. This treaty will extend the regulation of bioprospecting out into the open oceans, well away from the territorial limits or continental shelves of maritime countries (Kantai & Bettelli, 2022). Because of restrictions like these, bioprospecting projects reported in the literature are usually for one area, so that agreements only have to be made with a single government.

While modern antibiotic bioprospectors have been forced to adopt new methods to access secondary metabolite pathways and do not have the freedom to sample as widely as they did in much of

the 20th century they do have one advantage that their predecessors did not. The random sampling described by Tribe (1998) made sense as little was known about how actinomycetal and bacterial communities are distributed. Since the early 2000s molecular methods have been used to characterize the prokaryote and fungal communities of soils and other environments across the earth. The results of these methods can show how to sample efficiently for antibiotic discovery.

While there has been a lot of research on biosynthetic pathways within organisms and on how to access these pathways from environmental samples there has been relatively little work done on how to take samples for antibiotic discovery (Hernandez & Murphy, 2021). This topic could include a wide range of secondary metabolite pathways and environments. In this thesis by article I will limit this to the distribution of Actinomycetes and Bacterial Type I polyketide synthase pathway in terrestrial ecosystems.

Chapter 1 Extensively reviewing literature.

Published as: Sampling terrestrial environments for bacterial polyketides (Hill, Heberlig, & Boddy, 2017) Open Access Article.

Antibiotic prospecting projects have often concentrated on a single environment such as ant nests (Barke et al., 2010), disease suppressive soils (van Elsas et al., 2008) or volcanic caves (Riquelme et al., 2015). This work has been done without reference to the extensive literature of studies that asks basic questions about bacterial distribution in the environment (e.g.(Fierer & Jackson, 2006)). In chapter One both kinds of studies are reviewed.

This article was written for Molecules, a Chemistry journal, so it also covers basic topics in Microbial Ecology (e.g. Phospholipid fatty acid analysis, community fingerprinting. It was written to teach anyone without a background in Microbial Ecology all they need to know for terrestrial bioprospecting.

Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

A comparison of hard and soft direct methods for DNA extraction from soil. Hill P, Dextraze MF, Kroetsch D & CN Boddy (2022). Microbiology Spectrum Revisions requested. Open Access Article.

There are many studies of soil bacteria in the literature that can be used to guide sampling for antibiotic discovery. Most of these studies extract DNA from soil using bead beating extraction kits, PCR amplify with 16S primers and sequence the amplicons with next generation sequencing. The results of these studies are influenced by extraction method and PCR primers used. There are several studies in the literature that compare similar bead beating extraction methods with a narrow range of soils using a single bacterial primer pair e.g. (Soliman, Yang, Yamazaki, & Jenke-Kodama, 2017). Chapter two compares two very different extraction methods (bead beating vs enzymatic lysis) across a range of samples that are both very similar and different to get the full range of effects of methods. DNA samples were amplified with four different 16S primer pairs and sent to four different laboratories for either pyrosequencing or Illumina Next Generation Sequencing.

Chapter 3 Investigation of how the Actinobacterial community structure changes across land use.

Published as: Land Use Intensity Controls Actinobacterial Community Structure (Hill et al., 2011) Open Access Article.

The Actinomycetes are the bacteria best known for antibiotic production. They are an order of the Actinobacteria. Most studies of Actinobacterial ecology look at how their population structure changes within a range of treatments within a single environment. This is not very helpful for antibiotic bioprospectors who have to choose between many environments for sampling. These studies also often use a single primer pair for PCR to measure differences between samples.

This study compares a broad range of environments of varying anthropogenic impact. As shown in chapter two, results of amplicon sequencing can vary depending on the primers used. This study uses a range of primer pairs and community fingerprinting to compare Actinobacterial community structure in contrasting environments.

Chapter 4. Comparing the Type I polyketide synthase pathways of soil and street sediment.

Published as: Habitat-specific type in polyketide synthases in soils and street sediments (Hill et al., 2014). Included in thesis under Oxford University Press License 5337270512235

Type I polyketide synthase pathways produce many secondary metabolites which have been used as antibiotics and other medicines. It is unclear if these pathways are found everywhere in the same environments or their distribution is limited by dispersion. Studies that have looked at their distribution so far have typically done so for soil. This study compares PKSI distribution in contrasting soils and street sediments from sites that are geographically distant from each other. It found no evidence for PKSI distributuion being limited by dispersion..

Chapter 5 Conclusions.

The literature review in Chapter 1 summarizes literature up to April of 2017 and lists five key problems in sampling terrestrial environments for antibiotic bioprospecting. In the conclusion these five problems are re-evaluted in the light of more recent literature and the results of chapters 2 to 4.

References

Altamirano, F. L. G., & Barr, J. J. (2019). Phage Therapy in the Postantibiotic Era. *Clinical Microbiology Reviews*, *32*(2), e00066-18.

Angell, M. (2005). *The Truth about the Drug Companies:How they Decieve us and What to do about it.* (2nd ed.). New York: Random House.

Baltz, R. H. (2006). Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *Journal of Industrial Microbiology and Biotechnology*, *33*(7), 507–513. http://doi.org/10.1007/s10295-005-0077-9

Barke, J., Seipke, R. F., Grüschow, S., Heavens, D., Drou, N., Bibb, M. J., … Hutchings, M. I. (2010). A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus. *BMC Biology*, *8*(1), 109. http://doi.org/10.1186/1741-7007-8-109

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of Antibiotics*, *58*(1), 1–26. http://doi.org/10.1038/ja.2005.1

Bode, H. B., Bethe, B., Hofs, R., & Zeeck, A. (2002). Big Effects from Small Changes: Possible Waysto Explore Nature's Chemical Diversity. *ChemBioChem*, *3*(7), 619–623.

Busti, E., Monciardini, P., Cavaletti, L., Bamonte, R., Lazzarini, A., Sosio, M., & Donadio, S. (2006). Antibiotic-producing ability by representatives of a newly discovered lineage of actinomycetes. *Microbiology*, *152*, 675–683. http://doi.org/10.1099/mic.0.28335-0

Carr, G.; Poulsen, M.; Klassen, J.L.; Hou, Y.; Wyche, T.P.; Bugni, T.S.; Currie, C.R.; Clardy, J. (2012). Microtermolides A and B from Termite-associated actinomyctes.pdf. *Org. Lett.*, *14*(11), 2822–2825.

Castro, J. F., Razmilic, V., Gomez-Escribano, J. P., Andrews, B., Asenjo, J. A., & Bibb, M. J. (2015). Identification and heterologous expression of the chaxamycin biosynthesis gene cluster from Streptomyces leeuwenhoekii. *Applied and Environmental Microbiology*, *81*(17), 5820–5831. http://doi.org/10.1128/AEM.01039-15

Chain, E. (1944). OTHER ANTIBACTERIAL SUBSTANCES FROM BACTERIA AND MOULDS. *British Medical Bulleting*, *2*(1), 8–9.

Cheeptham, N., Sadoway, T., Rule, D., Watson, K., Moote, P., Soliman, L. C. Azad, N., Dunkor, K,.Horne, D. (2013). Cure from the cave: Volcanic cave actinomycetes and their potential in drug discovery. *International Journal of Speleology*, *42*(1), 35–47. http://doi.org/10.5038/1827-806X.42.1.5

Cruckshank, R. (1944). HOSPITAL INFECTION : A HISTORICAL REVIEW. *British Medical Bulleting*, *2*, 272–276.

Dhingra, S., Rahman, N. A. A., Peile, E., & Rahman, M. (2020). Microbial Resistance Movements : An Overview of Global Public Health Threats Posed by Antimicrobial Resistance , and How Best to Counter. *Frontiers in Public Health*, *8*(November), 1–22. http://doi.org/10.3389/fpubh.2020.535668

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc*

Natl Acad Sci U S A, *103*(3), 626–31. http://doi.org/10.1073/pnas.0507535103

Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., & Ziemert, N. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol*, *7*(May). http://doi.org/10.1038/s41564-022-01110-2

Gerwick, W. H., & Fenner, A. M. (2013). Drug Discovery from Marine Microbes. *Microbial Ecology*, *65*(4), 800–806. http://doi.org/10.1007/s00248-012-0169-9

Haeder, S., Wirth, R., Herz, H., & Spiteller, D. (2009). Candicidin-producing Streptomyces support leaf-cutting ants to protect their fungus garden against the pathogenic fungus Escovopsis. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(12), 4742–6. http://doi.org/10.1073/pnas.0812082106

Hernandez, A., & Murphy, B. T. (2021). Natural Product Reports The need to innovate sample collection and library generation in microbial drug discovery : a focus on academia. *Natural Product Reports*, *38*, 292–300. http://doi.org/10.1039/d0np00029a

Hill, P., Heberlig, G. W., & Boddy, C. N. (2017). Sampling terrestrial environments for bacterial polyketides. *Molecules*, *22*(5). http://doi.org/10.3390/molecules22050707

Hill, P., Krištůfek, V., Dijkhuizen, L., Boddy, C., Kroetsch, D., & Van Elsas, J. D. (2011). Land use intensity controls actinobacterial community structure. *Microbial Ecology*, *61*(2), 286–302. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20924760

Hill, P., Piel, J., Aris-Brosou, S., Kri??t??fek, V., Boddy, C. N., & Dijkhuizen, L. (2014). Habitat-specific type i polyketide synthases in soils and street sediments. *Journal of Industrial Microbiology and Biotechnology*, *41*(1), 75–85. http://doi.org/10.1007/s10295-013-1362-7

Kang, H. S., & Brady, S. F. (2014). Arixanthomycins A-C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chemical Biology*, *9*(6), 1267–1272. http://doi.org/10.1021/cb500141b

Kantai, T., & Bettelli, P. (2022). Summary of the Fourth Session of the Intergovernmental Conference on an International Legally Binding Instrument under the UN Convention on the Law of the Sea on the Conservation and Sustainable Use of Marine Biodiversity of Areas Beyond National Jurisdic. *Earth Negotiations Bulletin*, *25*(225), 1–19.

Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., … Lewis, K. (2015). A new antibiotic kills pathogens without detectable resistance. *Nature*, *517*(7535), 455–459. http://doi.org/10.1038/nature14098

Mahlapuu, M., Björn, C., & Ekblom, J. (2020). Critical Reviews in Biotechnology Antimicrobial peptides as therapeutic agents : opportunities and challenges. *Critical Reviews in Biotechnology*, *40*(7), 978–992. http://doi.org/10.1080/07388551.2020.1796576

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Lewis, K.,Epstein, S. S. (2010). Use of Ichip for High-Throughput In Situ Cultivation of “ Uncultivable ” Microbial Species ▯. *Applied and Environmental Microbiology*, *76*(8), 2445–2450. http://doi.org/10.1128/AEM.01754-09

Riquelme, C., Marshall Hathaway, J. J., Enes Dapkevicius, M. de L. N., Miller, A. Z., Kooser, A., Northup, D. E., … Cheeptham, N. (2015). Actinobacterial Diversity in Volcanic Caves and Associated Geomicrobiological Interactions. *Frontiers in Microbiology*, *6*(December), 1–16.

http://doi.org/10.3389/fmicb.2015.01342

Rutledge, P. J., & Challis, G. L. (2015). Discovery of microbial natural products by activation of silent biosynthetic gene clusters. *Nature Reviews Microbiology*, *13*(8), 509–523. http://doi.org/10.1038/nrmicro3496

Shu, D., Ting, J., Beuerman, R. W., & Dua, H. S. (2020). Strategies in Translating the Therapeutic Potentials of Host Defense Peptides. *Frontiers in Immunology*, *11*(May), 1–16. http://doi.org/10.3389/fimmu.2020.00983

Smith, P. W., Watkins, K., & Hewlett, A. (2012). Infection control through the ages. *American Journal of Infection Control*, *40*(1), 35–42. http://doi.org/10.1016/j.ajic.2011.02.019

Soliman, T., Yang, S.-Y., Yamazaki, T., & Jenke-Kodama, H. (2017). Profiling soil microbial communities with next-generation sequencing: the influence of DNA kit selection and technician technical expertise. *PeerJ*, *5*, e4178. http://doi.org/10.7717/peerj.4178

Stevens, D. C., Conway, K. R., Pearce, N., Villegas-Peñaranda, L. R., Garza, A. G., & Boddy, C. N. (2013). Alternative sigma factor over-expression enables heterologous expression of a type II polyketide biosynthetic pathway in Escherichia coli. *PLoS One*, *8*(5), e64858.

Sulakvelidze, A., Alavidze, Z., & Morris, J. G. (2001). Bacteriophage Therapy. *Antimicrobial Agents and Chemotherapy*, *45*(3), 649–659.

Tribe, H. T. (1998). The Discovery and Development of Cyclosporin. *Mycologist*, *12*(February), 20–22. http://doi.org/10.1016/S0269-915X(98)80100-6

van Elsas, J. D., Costa, R., Jansson, J., Sjöling, S., Bailey, M., Nalin, R., Vogel, T.M., van Overbeek, L. (2008). The metagenomics of disease-suppressive soils - experiences from the METACONTROL project. *Trends in Biotechnology*, *26*(11), 591–601. http://doi.org/10.1016/j.tibtech.2008.07.004

Ventola, C. L. (2015). The Antibiotic Resistance Crisis Part 1 : Causes and Threats. *Pharmacy and Therapeutics*, *40*(4), 277–283.

Wright, G. D. (2014). Something old , something new : revisiting natural products in antibiotic drug discovery 1. *Canadian Journal of Microbiology*, *60*(January), 147–154.

Chapter 1 Literature Review

Previously published as: Patrick Hill, Graham W. Heberlig, Christopher N. Boddy. 2017. "Sampling Terrestrial Environments for Bacterial Polyketides" *Molecules* 22, no. 5: 707 2017.

https://doi.org/10.3390/molecules22050707

The first section of this literature review introduces the topic of how to sample for antibiotic discovery. The second section of this chapter reviews five essential problems in sampling for antibiotic discovery.

1-Most of the available data about bacterial communities in the environment come from molecular biology studies. The results of these studies depend on the methods that they use. This means that it is uncertain if we can compare results between studies making the scientific literature an unreliable guide for bioprospectors.

2-While there is over a hundred years of study of how macrobial eukaryotic communities are distributed across the earth, bacterial communities may not follow the same patterns. Environments such as tropical rainforest may have a diverse plant and insect populations, but this does not mean that they will also have diverse bacterial populations.

3-Actinomycetes have been the source for most bacterial antibiotics. Actinomycetes distribution may not follow the same patterns as the whole bacterial community. Diverse bacterial communities may not be equally diverse in Actinomycetes.

4-Non-Actinomycetal bacteria may also be source of many novel antibiotics. Actinomycetal distribution may not be a reliable guide to Antibiotic production genes in the terrestrial environment.

5-The biogeography of antibiotic genes distribution is not understood. If it resembles that of eukaryotes, similar environments that are geographically distant will have different antibiotic genes. Other work suggests that similar environments will have similar antibiotics wherever they are and microorganisms spread easily.

Most of the examples cited in section two are from soil, the environment from which most polyketide producing bacteria were isolated in the "Golden Age" of antibiotic discovery from the late 1940s to 1989s.

The third section reviews the ecological reasons for several other terrestrial environments which have been targets for bioprospecting and reviews the polyketides that have been found in them (Insect associated environments, Desert soils, Disease suppressive soils and compost, Caves and Extreme Environments (with extremes of pH, salinity and heat). Two terrestrial environments suggested for bioprospecting are also discussed (Cities, Airborne bacteria).

The final section lays out plans for a program of sampling and antibiotic discovery well beyond the resources for this thesis.

Role of Ph.D. candidate in Chapter 1.

The Ph.D. candidate carried out:

-Finding and reviewing all literature.

-Summarizing (from the literature) or developing all Ecological reasons for bioprospecting.

-Developing recommendations for sampling for bioprospecting.

Sampling Terrestrial Environments for Bacterial Polyketides

Abstract: Bacterial polyketides are highly biologically active molecules that are frequently used as drugs, particularly as antibiotics and anticancer agents, thus the discovery of new polyketides is of major interest. Since the 1980s discovery of polyketides has slowed dramatically due in large part to the repeated rediscovery of known compounds. While recent scientific and technical advances have improved our ability to discover new polyketides, one key area has been under addressed, namely the distribution of polyketide-producing bacteria in the environment. Identifying environments where producing bacteria are abundant and diverse should improve our ability to discover (bioprospect) new polyketides. This review summarizes for the bioprospector the state-of-the-field in terrestrial microbial ecology. It provides insight into the scientific and technical challenges limiting the application of microbial ecology discoveries for bioprospecting and summarizes key developments in the field that will enable more effective bioprospecting. The major recent efforts by researchers to sample new environments for polyketide discovery is also reviewed and key emerging environments such as insect associated bacteria, desert soils, disease suppressive soils, and caves are highlighted. Finally, strategies for taking and characterizing terrestrial samples to help maximize discovery efforts are proposed and the inclusion of non-actinomycetal bacteria in any terrestrial discovery strategy is recommended.

Keywords: polyketides; bioprospecting; microbial ecology

1. Introduction

Polyketide natural products are exquisite molecules, often with extraordinary and diverse structures ranging from macrolides to polyethers and polyphenols. They typically bind with high affinity and selectivity to a biological target (usually a protein or RNA) and can cross biologically relevant membranes, enabling them to modulate the biology of organisms. These features have made polyketide natural products an indispensable resource for the discovery and development of new pharmacological agents including the antibiotics, like erythromycin A and tetracycline, antitumor agents, like doxorubicin; and the anti-fungal drugs, like amphotericin B [1].

While structurally diverse, all polyketides are related by their highly conserved biosynthetic origins. The polyketide backbone is assembled, analogously to fatty acid biosynthesis, through sequential additions of two carbon building blocks (ketide units) derived from malonyl-CoA. This assembly is catalyzed by enzymes called polyketide synthases (PKS). In bacterial these polyketide synthases typically occur with one of two different architectures, modular type I polyketide synthases and the iterative type II polyketide synthases.

The type I PKS pathways contain multiple modules consisting of ketosynthase (KS) domains, which catalyze the addition of ketide units to the growing polyketide chain via decarboxylative Claisen condensation of an acyl carrier protein (ACP)-linked malonyl [2,3]. In addition to the KS and ACP many type I PKS pathways also possess an acyltransferase (AT) domain embedded into each module which loads the appropriate malonyl extender unit onto the ACP. The trans AT type I PKS pathways possess a separate stand-alone AT domain responsible for loading the ACP domains [4]. In addition to these catalytic domains which extend the polyketide chain, type I PKS pathways can also possess reductive domains such as a ketoreductase (KR), dehydratase (DH), and enoylreductase (ER) domains that reduce the growing polyketide intermediate, similar to mammalian fatty acid biosynthesis prior to the next round of polyketide elongation [2,3]. Elongation continues until eventual the completed polyketides is

release from the ACP by a thioesterase (TE) domain [5]. Type I PKS are responsible for the production of a wide variety of bioactive compounds including the antibiotic erythromycin and the antifungal amphotericin B.

Type II polyketide synthases are most frequently found in the order of bacteria actinomycetales and function similarly to bacterial fatty acid synthases. Like type I PKS, these synthases rely on a KS domain, which heterodimerizes with a protein called chain length factor (CLF), and an ACP domain. A malonyl acyltransferase (MAT) domain is used to load malonyl onto the ACP for decarboxylative condensation with the growing KS-bound polyketide chain. Unlike type I PKS, these pathways typically use these four catalytic domains iteratively and possess a single reductive domain [3,6]. Typical products from type II PKS pathways are aromatic compounds that are often highly tailored to generate complex molecules such as the antibiotic tetracycline or the anticancer agent doxorubicin.

Often type I polyketide synthases form hybrid pathways with non-ribosomal peptide synthetases, producing natural products containing both polyketide portions and peptide portions [7–9]. The anticancer agent epothilone B is a prime example of a product from this type of pathway. In this review article we broadly define bacterial polyketides as compounds produced by type I and type II polyketide synthases biosynthetic pathways as well as hybrid polyketide-non-ribosomal peptide pathways.

The discovery of new polyketides has slowed noticeably since the late 1980s as the traditional method of discovery, isolating and cultivating bacteria (usually Actinomycetes) from soil, found fewer and fewer new polyketides. At the same time bacterial resistance to antibiotics has been increasing. To restart polyketide discovery, bioprospectors are using methods that improve or replace bacterial cultivation and sampling alternative environments to soil. Both of these approaches mean that it is important to know how bacterial polyketides are distributed to sample them efficiently.

Soil cultivation methods can be scaled up to screen millions of isolates a year to find rare polyketides [10]. Sequencing of the genomes of isolated bacteria shows that most polyketide genes are

"cryptic", i.e., not expressed under normal cultivation conditions. In some cases these cryptic pathways have been coaxed to express by changing the cultivation conditions [11], adding antibiotics that stimulate polyketide production [12], overexpressing regulatory genes [13], or blocking competing pathways [14]. Alternatively these cryptic pathways can be expressed in heterologous hosts, enabling new compound isolation. While this approach has a number of technical hurdles, including ensuring transcription, codon usage, protein folding, ensuring the presence of precursors, and concerns that the polyketide product itself may kill the host [15], there have been some major discoveries in this area [16].

While most bacteria in soil cannot be readily grown, the cultivatable fraction of the soil bacterial community can be increased by keeping bacteria in contact with the soil in buried "chips" for weeks [17]. To circumvent cultivation, bacterial DNA can be extracted from soil and heterologously expressed in bacteria that can grow in the laboratory [18,19].

These new techniques are changing the way polyketides are discovered (for reviews see [20–25]), but they require much or more time and effort for each sample as the low tech isolation methods that they replace. The problem of how to sample is no longer limited to soil. Bioprospectors can and should consider the many alternatives to soil such as caves, insects and desert soils. Bioprospectors need to compare the microbial polyketide producers within and between environments to find the most polyketide diversity with the fewest samples.

Until recently it was impossible to compare the bacterial community structure between environments. It had long been known that only one of every hundred to thousand bacteria from the environment that could be seen through a microscope could be grown in the laboratory [26]. In 1990 DNA was extracted from soil, denatured, and its reassociation time measured, allowing researchers to estimate that there were the equivalent of 4000 *E. coli* genomes in each gram of the soil [27]. This was one of the first uses of cultivation-independent, nucleic acid-based methods to study soil bacteria.

Modern nucleic acid based methods usually use the polymerase chain reaction (PCR) to measure which bacteria are in the soil and estimate how diverse the bacterial community is (Figure 1.1).

PCR products amplified from soil DNA templates typically contain thousands or more different sequences. Cloning and sequencing with Sanger sequencing (e.g., [28]) was used until the early 2000s to sequence these amplicons. Due to the intrinsically low throughput of Sanger sequencing and thousands to hundreds of thousands of bacterial species in soil, it was difficult to get a full picture of the bacterial community of a single sample, let alone to compare samples. An alternative was to separate the PCR amplicons by gel electrophoresis to get a "community fingerprint" and compare fingerprint patterns to get a rough idea of how bacterial communities differed. Gel fingerprinting can separate amplicons several ways, including gradients of temperature or urea, which separate amplified DNA by GC content (e.g., Denaturing Gradient Gel Electrophoresis, DGGE [29]), or cutting amplicons into smaller fragments by restriction endonucleases and separating by size (e.g., Terminal Restriction Fragment Length Polymorphism, T-RFLP [30]). Community fingerprinting gives less information than Sanger sequencing, but can quickly and cheaply compare many samples (Figure 1.1).

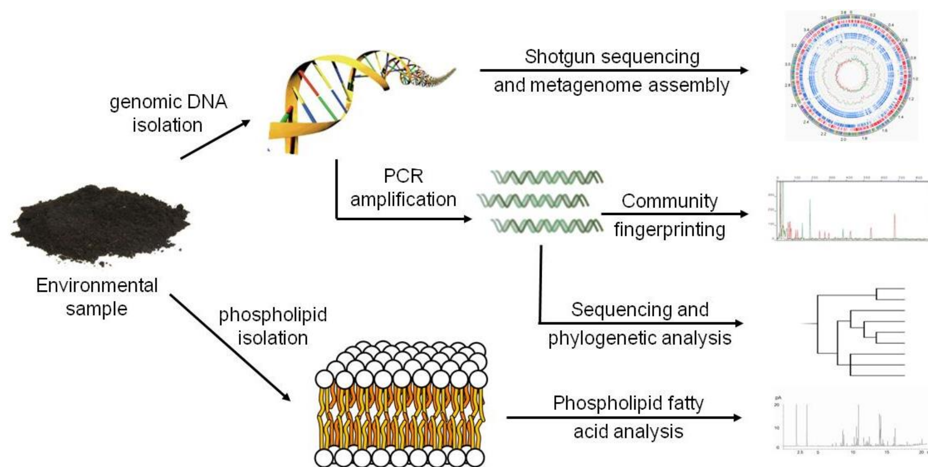


Figure 1.1. Microbial ecology approaches used to examine both cultivatable and uncultivatable fraction of the microbial community structure. This includes molecular methods like shotgun metagenome sequencing, community fingerprinting of PCR products, sequencing of PCR products, and phospholipid fatty acid analysis.

In the 2000s next generation sequencing replaced Sanger sequencing and fingerprinting of PCR products. With next generation sequencing thousands to hundreds of thousands of different amplicons could be sequenced from a single sample and many samples could be multiplexed together to give enormous data sets [31]. This increased the volume of sequencing results and the number of sequenced genomes. Researchers could even sequence unamplified DNA from the soil metagenome (Figure 1.1) and estimate which phylogenetic groups it belonged to [32].

Nucleic acids are not the only biomolecules that can be used to measure microbial community structure. Analysis of phospholipid fatty acids (PLFA) from the cell membranes of soil microbial communities began in the 1980s (Figure 1.1) [33]. Like nucleic acid methods, it provides information on the uncultivated majority of bacteria. While PLFA has largely been replaced by next generation sequencing, there is a large amount of PLFA data in the literature. Unlike nucleic acid sequencing, PLFA cannot identify bacteria at the genus or species level and is often specific to groups that are even broader than phyla (e.g., Gram-positive bacteria). In spite of this there are advantages to PLFA. It is cheaper and faster than nucleic acid methods and gives a better picture of fungal to bacterial ratios. Results from PLFA community analysis showed that pH controlled bacterial community structure well before this was found by nucleic acid-based methods [34]. It has also been suggested that phospholipid fatty acids are quickly degraded in soil, unlike DNA, so PLFA may give a better picture of the live bacterial community than DNA methods that also capturing relic DNA [35].

The most important bacterial group for polyketide production is the Gram-positive phylum Actinobacteria. Within this phylum, the Actinomycetales order, have produced more pharmaceutically useful polyketides than all the rest of the bacteria [36]. This review will use both the terms Actinobacteria and Actinomycete. This is as molecular microbial ecology studies usually use the term Actinobacteria while cultivation studies typically usually use the term Actinomycetales.

The reader might think that deciding where to sample for discovery of new polyketides is a straightforward question of reviewing the microbial ecology literature to find environments where Actinomyces are diverse and a large fraction of the bacterial community. While there are many studies that use molecular biology methods to characterize the microbial communities, using this literature to decide where to sample to find new polyketides is difficult. There are several reasons for this.

Most modern microbial ecology studies use PCR to amplify from nucleic acids extracted from the environment. The results from PCR-based methods depend significantly on the choice of extraction methods, PCR primers, and reaction conditions. Many studies use a unique version of each of these, so results are often not comparable between studies. Recently several large scale projects, such as The Earth Microbiome project [37], have been applying standardized methods on samples from across the planet.

Bacterial habitats do not follow the same pattern as the habitats of eukaryotic life that we see around us. Environments that to the human eye seem very different, can have similar bacterial communities [38]. In addition the actinobacterial community may not be controlled by the same factors as the whole bacterial community. Furthermore, the Actinobacteria, which have so far supplied most of our medicinally useful polyketides, may not produce most of the type I polyketides in the environment [39]. Understanding what controls bacterial community structure and how this affects the distribution of polyketide producing bacteria is still unresolved.

Finally, the major ecological question of bacterial biogeography has still not been settled. Biogeography, the pattern of species distribution across geographical area, affects where multicellular eukaryotes such as plants and animals are found both through the distribution of habitats and through barriers to dispersion. Barriers to the dispersion of plant and antimals such as oceans or deserts can last for millions of years, so that separated populations can have different evolutionary histories. It is not clear if this is true for bacteria as a whole or for polyketide producing bacteria specifically. Questions

such as “Are the same bacterial communities found wherever there are the same conditions?” decide whether bioprospectors should sample particular environments in many places over the earth or many environments that may be in the same area.

We review where to sample for bacterial polyketides in terrestrial environments. The first section of this review discusses problems in using the current scientific literature for bioprospecting as briefly outlined above. Because terrestrial bioprospecting for new polyketides is already underway, the second part of this review discusses several of the terrestrial environments where new polyketides have been found or which have been suggested for bioprospecting.

The final section suggests how to sample to find new polyketides in terrestrial environments. We recommend sampling different environments and measuring samples properties in several stages. Physical/chemical properties such as pH, organic carbon or texture can be measured for a broad range of samples. The microbial community of a subset of these samples can be determined with rapid, low cost community fingerprinting, before next generation sequencing of subsets with contrasting fingerprints. An advantage of using community fingerprinting is that they give results quickly. If certain samples give unusual or promising fingerprints, the environment or area where they are from can be resampled. Bioprospectors should also try to find non-actinomycetal bacterial polyketides as they are more likely to be novel.

2. Unresolved Issues in Microbial Ecology that Limit Bioprospecting for New Polyketides

Molecular biological methods have changed our view of soil bacterial communities. Some Phyla, such as the Acidobacteria, are a much larger fraction of soil bacteria than was previously known [40]. Soil properties such as texture and pH are often more important in determining community structure than vegetation or climate [41]. While we know much more about bacterial communities than we did 20 years ago, there are still gaps in our knowledge that can make it difficult to know where to bioprospect for new bacterial polyketides.

2.1. Data Comparability

Many microbial ecology studies produced data that cannot be compared with other studies. Studies often look at a single environment and look at changes in the microbial community caused by treatments, such as the effect of different tillage on soil bacteria [42], the changes in the microbial community during composting [43], or the bacteria found in different kinds of office dust. These studies extract DNA from the environment, amplify it with primers and analyze the fingerprints or sequences of the amplicons. They often use a unique combination of extraction method, primers, and amplification conditions. Changing any one of these steps can significantly change the results. For example, Delmont et al. [44] compared a range of DNA extraction methods on two grassland soils and found that they gave very different pictures of the microbial community. Hong et al. [45] used two extraction methods and two primer sets to measure the bacterial community of a beach sand sample, only 11 of 1098 operational taxonomic units (OTUs, defined as 16S ribosomal RNA genes that have less than 99% similarity) sequences were found in all three data sets (figure 1.2). In our own work two sets of "actinobacterial specific" and one set of eubacterial 16S primers amplified very different groups of the actinobacteria [46]. Even when the same methods are carried out in different laboratories there can be significant differences in results [47].

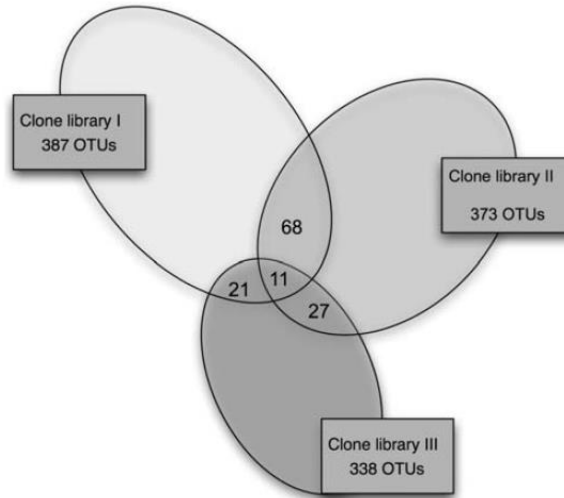


Figure 1.2. The overlap between Sanger sequencing results of three 16S clone libraries at 99% identity from a beach sand using two DNA extraction methods and two primer pairs. Clone library 1 Extraction method 1 PCR primer pair 1, Clone library II Extraction method 2 PCR primer pair 1, Clone library III II Extraction method 1 PCR primer pair 2. After Hong, S.; Bunge, J.; Leslin, C.; Jeon, S.; Epstein, S.S. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* 2009, *3*, 1365–1373.

Bioprospectors need to know where to sample. The most useful results will come from studies that use the same method over a broad range of samples from different environments. As methods of analysis are continually improving, the samples (along with information about where and when they were collected and what they are like) should be stored for reanalysis as new methods come online [48]. Several large collaborative projects have been set up to do exactly this [37,49–51], and this is an area of research that is expanding [52].

These consortia have the choice of using methods that are labour and time intensive but give a lot of information for a few samples or using methods that are faster and cheaper, but less detailed for a lot of samples. The best example of the time and labour intensive approach is the Terragenome project [49]. This consortium began by sequencing the metagenome of a few well characterized soil samples, starting with samples from the oldest field trial in the world at Rothamsted. Eventually this work will provide very detailed information on polyketide content in a few soils.

Shotgun metagenome sequencing of the diverse microbial communities from soil is difficult. An enormous volume of sequence data is needed and it is difficult to assemble genomes or polyketide

biosynthetic gene clusters from these large data sets [53]. Terragenome involves investing a lot of time, work and funds in a few soils before it is known how representative their microbial communities are. It has been suggested that it would be more efficient to survey many soils using simple 16S bacterial community fingerprinting methods, before picking samples for direct metagenomics sequencing [54]. This initial step has been done in soil microbial surveys of Great Britain and France, where hundreds of samples from across landscapes have been fingerprinted [55,56]. These studies provide reproducible but low resolution information on bacterial distribution and diversity in soil.

The most ambitious of these projects, which uses identical molecular methods on many samples, is the Earth Microbiome project. This project samples broadly from environments including soil, marine sediment, and the digestive system of mammals. It uses next generation sequencing to sequence 16S bacterial and 18S fungal ribosomal RNA amplicons, as well as shotgun metagenome sequencing. Recently a host of additional projects looking at the microbiome of humans have used next generation sequencing of 16S amplicons to characterize thousands of samples [57,58]. These studies are beginning to provide a clearer picture of bacterial community structure of many environments.

The laboratory of Brady, a pioneer in polyketide bioprospecting, has developed a database, the Environmental Surveyor of Natural Product Diversity (eSNaPD) that includes information on the distribution of polyketide synthases in the environment. So far it is limited to soils and marine sediments from the Southwestern US and New England [59].

2.2. Soil Bacterial Habitats Do Not Correspond to Eukaryotic Habitats

For over a hundred years, terrestrial ecologists have studied how eukaryotic life on Earth is distributed. Plant and animal life is distributed over the Earth in biomes, such as tundra, rainforest, prairie, etc. These biomes are more diverse in the tropics and diversity decreases with distance from the equator [60]. When biomes such as tropical rainforest are converted to crops or pasture [61], biological

diversity decreases. Given what we know about these biomes, it would seem reasonable to bioprospect for new polyketide natural products in diverse tropical biomes undisturbed by human activity. However, while fungal distribution may follow these biomes [62], bacterial distribution does not.

The plants that grow on soil are probably not the best guide to the bacteria that live in them. Soil properties better predicted bacterial community structure than vegetation in several early studies that looked at soils and land uses [63–65]. In 2006, a study by Fierer compared 88 soils under natural vegetation from the Peruvian Amazon and across the continental United States [66]. Soil DNA was used for 16S community fingerprinting (T-RFLP). Fingerprints from soils close to each other but of different pH differed but resembled those from thousands of miles away with the same pH. Soil pH also determined how diverse the fingerprints were, with the most diversity at neutral pH. This effect of pH on bacteria community structure has been confirmed by studies that compare soils from narrower geographical regions such as the Arctic [38] and Malaysia [41].

The importance of pH to bacterial community structure leads to non-intuitive conclusions for microbial bioprospectors. Biomes such as tundra may have similar soil bacterial communities to biomes with far more and diverse vegetation [38]. Tropical rainforest soil, underneath one of the most biodiverse plant and animal ecosystems on earth, has an undiverse bacterial community, compared to soils from ecosystems that have orders of magnitude fewer species of plants and are thousands of miles further north [66]. Converting Amazonian tropical rainforest to cropland or plantation may increase soil bacterial diversity [67].

While pH may be the major controlling factor for soil bacterial community structure, it is not the only one. When comparing soils over a narrower range of pH, other soil properties are important. A study in The Netherlands compares 25 sites over a broad range of soils and land uses. When three acid pine forest soils were excluded (pH 3.7–4.1), soil phosphorous controlled bacterial community structure

in the remaining soils (pH 5.1–7.6) [68]. On a range of potato soils in Germany of similar pH (5.2–6.2), parent material controlled bacterial community structure [69].

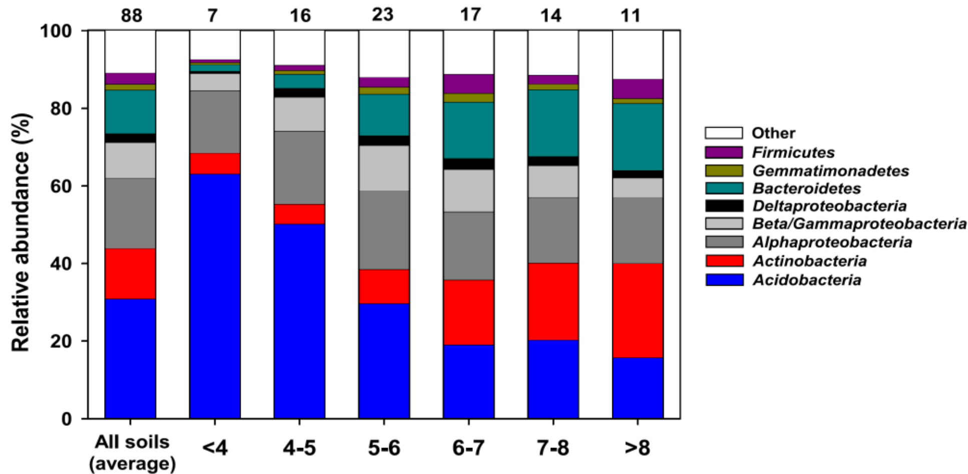


Figure 1.3. Bacterial 16S 454 pyrosequencing results from 88 non-agricultural soils from across North and South America. Modified from Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* 2009, *75*, 5111–5120.

Microbial bioprospectors cannot rely on the vegetation that they see around them to know where to sample soil. The will need to know a range of soil properties that can be readily determined in the laboratory or through soil surveying. Unfortunately many studies of both bacterial and polyketide distribution in soil do not include soil properties.

Eukaryotic and bacterial communities do not just differ in how they are distributed, bacterial communities are far more diverse. Fierer and Lennon discuss why this is, two of the most important reasons being phylogenetic breadth and geographical scale [70]. Bacteria are one of the three domains of life so a fair comparison of the bacterial diversity of soil would be with the diversity of all of the eukaryotes living in the soil (i.e., plants, animals, fungi, protozoa etc.). More important for bioprospectors is the question of scale. Grasses are sampled at in 1 m square plots which can have 10,000 individual grasses, to sample on this scale for bacteria microbiologists would have to sample a single fine sand grain. To sample grasses on the scale that we sample bacteria (tens to hundreds of billions) plots would

have to be 10 km^2. Grass in a 10 km^2 area would have differences in their environment, the same would be even truer of bacteria in soil, a complex environment, where there are changes in conditions at the µm scale.

2.3. Actinobacterial Soil Community Structure May Not Follow Bacterial Soil Community Structure

Actinobacteria cultured from soils are the traditional source of new polyketides. A follow up to the Fierer study, which found that the soil bacterial community structure was controlled by soil pH [26], used next generation sequencing of 16S amplicons obtained from their original samples and found that Actinobacteria were most common at higher pH [71] (Figure 1.3). The actinobacterial fraction of the bacteria increased from 5.35% at pH < 4 to 24.3% at pH > 8 but actinobacterial diversity was greatest at an intermediate pH between 6 and 7. The same relationship between the relative abundance of Actinobacteria and pH was found for the bacterial communities of polar soils [38]. This suggests that neutral to alkaline soils may be the best target for actinobacterial bioprospecting.

However, actinobacterial abundance and distribution may not be controlled by soil pH in the same way that the whole bacterial community is. At the continental scale soils in dry climates are neutral to alkaline and soils in wet climates are acidic. At the landscape level soil parent material can determine soil pH. Desert soils are enriched in Actinobacteria compared to soils from wetter climates [72]. Field plot studies where the pH of a plot of soil is adjusted over a gradient can separate the effect of pH from climate. Four field plot studies have been carried out on cropped [49,73] and pasture soils [74,75] in the United Kingdom. While they have found that the actinobacterial community structure changes with pH [74–76], they have not found that the relative abundance of the actinobacterial fraction depends on pH through sequencing of 16S amplicons [46,47,49] and phospholipid fatty acid abundance [48]. In two of these cases the relative abundance of other major bacterial phylogenetic groups changed [46,47], in another there was no change for any of the bacterial phyla, although there were changes at lower

phylogenetic levels [49]. These results suggest that actinobacterial relative abundance is not controlled by pH but do not say what does control it.

All of the above studies have either compared soils under natural vegetation or looked at a single agricultural land use (cropping/pasture). Converting native vegetation to agricultural use often increases the actinobacterial fraction of the soil bacterial community [65,77], in particular when humid tropical forest is converted to crops [67,78] although this effect can be below statistical significance [79]. This may be because under native forest these soils are acidic and deforestation raises the soil pH slightly. This is the explanation given for the effect of land use (cultivation, pasture, pine forest, and deciduous forest) on soil bacterial community structure in a well cited paper by the Fierer group. The authors examined twelve sites in South Carolina and it was shown that land use changed the bacterial community by changing pH rather than any other inherent effect of the land use [77]. However, the relative abundance of actinobacteria can increase as pH decreases during conversion of forested land to cropping [80]. In our own study of contrasting soils from Colombia, Canada and Europe using community fingerprinting with two sets of actinobacterial 16S primers, we found that fingerprints clustered by land use (cultivated versus uncultivated), with no effect of pH [46] (Figure 1.4).

There are several ways that land use could affect the relative abundance and community structure of soil Actinobacteria. The higher level of Actinobacteria in cultivated and pasture lands versus forest in Alabama [81] or palm oil plantation versus forest in Borneo, Malaysia [82], led researchers to suggest that Actinobacteria are selected by human disturbance. A study in the savanna region of Ghana found that soils with no or little vegetation between corn cropping seasons had a higher relative actinobacterial fraction [83] than soils with more vegetation.

Land use can also affect the bacterial community structure by adding or removing substrate for bacterial growth. Bacteria can be classified as oligotrophs or copiotrophs, similar to K- and r-selected plants and animals. Oligotrophs have an advantage when food is limited and grow slowly. Copiotrophs

are more competitive when food is plentiful. Fierer et al. measured carbon and nitrogen mineralization in 71 soils and compared this with the relative percentage of the Actinobacteria and five other bacterial groups, α-Proteobacteria, β-Proteobacteria, Acidobacteria, Bacteroidetes and Firmicutes, as measured by quantitative real-time PCR. They concluded that while the Acidobacteria were oligotrophs and the Bacteroidetes and β-Proteobacteria were copiotrophs, the α-Proteobacteria, Firmicutes and Actinobacteria were neither [84].

Whether Actinobacteria are oligotrophic or copiotrophic may depend on the nature of the substrate. Additions of carbon rich straw or reduced tillage that leaves straw on the surface of the soil reduces the relative amount of Actinobacteria in the soil bacterial community [42]. Adding nitrogen fertilizer increases the actinobacterial fraction of the soil bacteria. An increase in the Actinobacteria with nitrogen was found for 28 soils from natural vegetation across the continental United States by next generation sequencing of 16S amplicons [85], as well as at a grassland and cultivated corn site through next generation sequencing of 16S amplicons [86,87] and shotgun metagenome sequencing [60]. Both nitrogen and potassium addition increased the actinobacterial fraction of a Dutch grassland soil as measured by shotgun metagenome sequencing [88].

A 100 year old field trial in Alabama compared the effect of adding lime to raise pH of an acid soil versus adding lime and fertilizer. The bacterial community structure was measured using PLFA analysis. While pH had more of an effect on the overall bacterial community structure than nitrogen fertilizer, nitrogen fertilizer had a major effect on the actinobacterial and fungal levels [89]. An increase in the actinobacterial fraction of the bacterial population has been ascribed to the flush of nutrients produced by burning after deforestation in the Amazon [51].

More work is needed comparing the relative importance of pH, moisture, land use and nitrogen on actinobacterial community structure and relative abundance in soil. There may not be a simple relation between these factors. Just as the actinobacterial community structure does not follow that of the

whole bacterial community, different sub groups of the Actinobacteria may be selected for by different factors. For example, Actinobacteria can be the largest fraction of the bacterial communities of deserts [72] and lake waters [90,91]; probably different mechanisms of selection are occurring in these cases. This uncertainty means bioprospectors need to compare a broad range of soils of differing properties, land uses, and nutrient inputs.

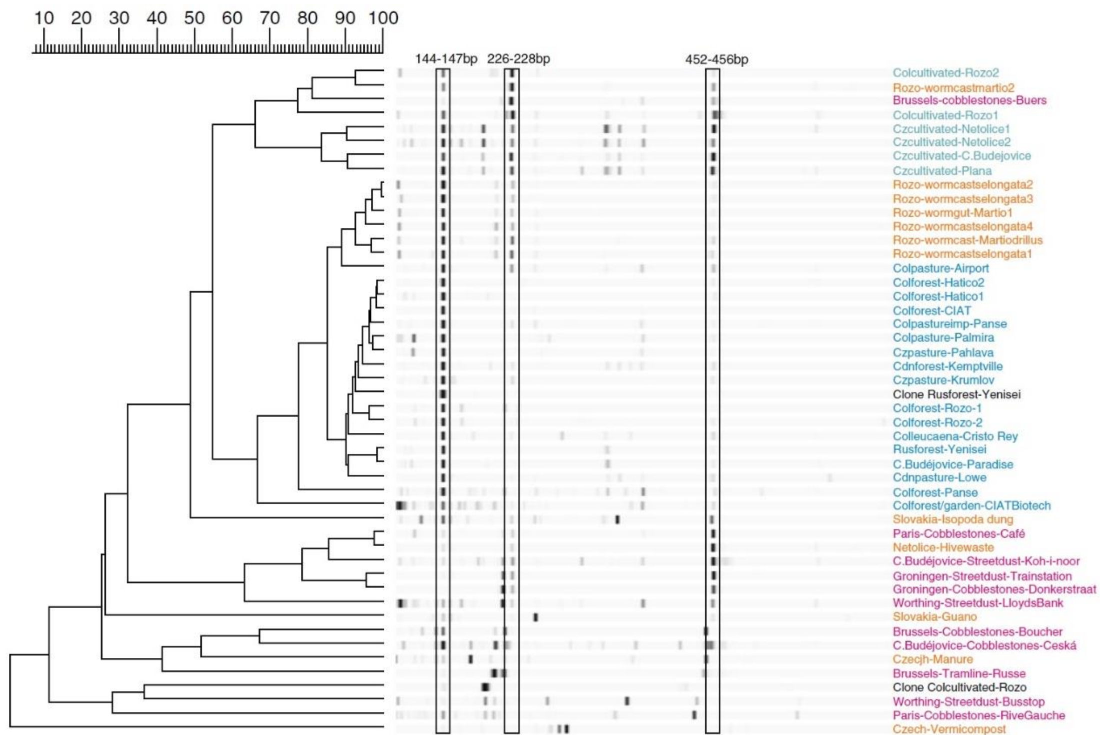


Figure 1.4. Forward T-RFLP community fingerprints of the Actinobacteria from Uncultivated soil (blue) Cultivated soils (green), animal associated sediments (orange) and street dust (red). After Hill, P.; Kištůfek, V.; Dijkhuizen, L.; Boddy, C.; Kroetsch, D.; Van Elsas, J.D. Land use intensity controls actinobacterial community structure. *Microb. Ecol.* 2011, *61*, 286–302.

2.4. Non-Actinobacterial Bacterial Polyketide Producers

While most polyketides discovered by cultivation of bacteria in the laboratory so far have come from Actinobacteria, many bacterial polyketide producers in the environment are not Actinobacteria. In the case of aromatic type II polyketides most producers are likely Actinobacteria. Genes encoding type II polyketide synthases (PKSII) amplified from the soils DNA either directly [92–95] or from metagenomic

libraries [96] appear to be entirely actinomycetal. The only example where PKSII amplicons appear to be not-actinomycetal is a set of five out of fifty-seven sequences amplified from European soils. In this case the PKSII amplicons appear to be from proteobacteria or firmicutes [96].

In contrast many studies have found environmental amplicons and clones with type I polyketide synthase (PKSI) sequences that cluster with non-actinobacterial sequences. A broad range of studies have used PKSI specific primers to amplify soil DNA followed by cloning and Sanger sequencing. For example, most PKSI sequences from soils from Malaysian rainforest [94], an Antarctic Island [97] [98], close to cucumber roots [99], and the Tibetan plateau [100] clustered with non-actinobacterial sequences when phylogenetically analysed.

Three studies have sequenced metagenomics PKSI clones from soil that do not appear to be actinomycetal. As part of the European Union funded Metacontrol project, a metagenomic library was made of environmental DNA from a French cultivated soil [101]. The 60,000 clones in the library were screened with PKSI specific primers and 139 positive clones found. Three of these clones were fully sequenced. Two PKS domains from these clones, the ketosynthase (KS) and acyltransferase (AT) domains clustered with myxobacterial KS and AT amino acid sequences. A fosmid library from a cultivated soil in the Metagenomic project at the [39] University of Wisconsin, was screened and 29 clones were identified as containing KS domains by hybridization. The inserts of five of these clones were sequenced. Their KS domain homology suggested that they were from the Proteobacteria or Cyanobacteria. However, analysis of the full insert, its GC content, and comparison with KS domain sequences from three acidobacterial isolates suggested that they were from the Acidobacteria.

As a final example, DNA was extracted from a Brazilian eucalyptus plantation soil, and made into a fosmid clone library. A clone with a PKSI hit was sequenced [102]. The start of the clone contained a single PKSI module with 20 other open reading frames (ORFs). Matches to the ORF were all non-actinomycetal, while the KS and AT domains of the PKSI module clustered away from their

actinobacterial counterparts when treed with the SEARCHPKS database [103]. This database was disproportionately actinomycetal possibly biasing the analysis. However, this result was confirmed when the sequences were reanalyzed with a broader range of non-actinomycetal PKS domains [104].

These environmental PKSI sequences were probably non-actinomycetal, but as only part of their genome was sequenced we cannot be certain which bacterial group they were from. In some environments most PKSI producers are actinomycetal.

In our work we compared PKSI amplicons of KS and AT from street sediments [104], which are enriched with Actinomycetes, and soils [46]. Many soil amplicons clustered with myxobacterial sequences. However, several clades clustered with actinomycetal sequences and were specific to samples from the streets of Ottawa (Canada), Faisalabad (Pakistan), and four European cities. The Brady group compared PKSI KS sequences in soils from New England and desert soils of the Southwestern United States [105] through pyrosequencing amplicons. KS domain diversity was greater in the actinobacterial rich desert soils.

As further evidence that non-actinomycetal source may prove important in bioprospecting, a new class of antibiotics (albeit a non-ribosomal peptide natural product) was recently discovered through the use of an in situ cultivation technique that grew a novel β-proteobacteria from soil [17]. Non-actinomycetal producing organisms may be important for future polyketide discovery.

2.5. The Question of Biogeography

Biogeographic studies of bacteria often simply find changes in the community over geographical distance. An example of this is the study that found that the actinobacterial and PKSII communities of soils in Uzbekistan and New Jersey differed [93]. In this case we have two very different environments that are distant from each other so there is no explanation of what drives these differences.

Differences in soil bacterial communities might be simply due to Uzbeki soils being more arid and alkaline to those is New Jersey. However different animals and plants are often found in similar environments that are separated by distance or a barrier such as an ocean. This difference is attributed to the communities evolving separately as Wallace proposed in the 1876 *Geographic Distribution of Animals* [106]. Biogeography can have a large effect on the distribution of large multicellular organisms. For example, marsupials are found in Australia but not in Africa, even though there are similar environments in both places.

For polyketide discovery, the more polyketide synthase gene distribution in terrestrial environments is controlled by such isolation, the more important it is to sample similar habitats that are separated from each other by distance or geographical barriers. If bacterial polyketide producers can distperse easily and quickly, it is more important to sample different habitats, even if they are close to each other. Addressing this question is complicated by the difficulty of defining bacterial, actinobacterial and polyketide habitats.

Microbes face fewer barriers to distribution than larger organisms. It has been argued that there are no barriers to organisms <20 µm because they are easily spread across the globe and their numbers are so high [107]. This is a restatement of ”*Everything is everywhere but the environment selects*”—a famous quote from the first half of the 20th century. There is disagreement as to who is being quoted and what was said [108] and some maintain that this view has roots in the 19th century [109] but it stems from the observation that the bacteria isolated depended more on the isolation method than the sample. If this is true, bioprospecting depends on developing culturing methods that favour the growth of novel microbes “*perhaps using virtually any natural sample*” [110].

In the 2000s, as molecular methods were introduced, the question of microbial biogeography was revisited [111]. This was partly because the it was shown that isolated bacteria were “weedy” (i.e.,

found everywhere and fast growing) compared to the uncultivated majority of bacteria that could only be studied with molecular methods [112].

A clear effect of distance was found for archaea [113] cyanobacteria [114] and actinobacteria [115] from volcanic hot springs that are thousands of miles apart. However, hot springs are isolated from each other so this does not mean that there are barriers to dispersion is important in a broadly distributed environments like soil. This is an argument for ensuring that cave microorganisms, which are also in highly isolated, are protected from contamination by cavers and scientists [116].

National level surveys of the soils of France [56] and Britain [55] using fingerprinting have found little or no effect of distance. This may be because it is difficult to find identical soils that are distant from each other at the national scale [55].

Several studies, which have used the higher resolution methods of next generation sequencing, have looked at the biogeography of soil bacteria. A comparison of mollisols (soils developed under prairie) of Manchuria found that environment rather than distance had the greatest effect on bacterial community structure, although distance may have had some effect [117]. An effect of distance was found for Actinobacteria in soils that were recently been exposed by receding glaciers [118]. This effect was not found for Chinese glaciers in a study that used the same methods of DNA extraction and amplification, but was observed when they were compared with literature data from other continents. As different methods of PCR were used for some of the literature samples, this result may be an artefact of the method.

A novel approach to the question of soil bacterial biogeography was to transplant soils from two sites in China around 1000 km apart, leave them for 20 years, and then use next generation sequencing to see if their bacterial communities were the same as in the soils that they had come from or their "new homes". After 20 years the bacterial communities resembled those of the soils around them; from this the authors argue against a biogeographical effect [119].

Five research groups have compared polyketide distribution in soils sampled far from each other and three have reported evidence of distance. However, in two of these cases, this may be due to sampling different habitats rather than geographical separation.

Wawrik et al. compared actinobacterial and PKSII T-RFLP fingerprints from an acid New Jersey Pine forest soils and a range of undefined soils from Uzbekistan and concluded that there was a biogeographical effect [93]. Given that the climate, soil pH, soil type, and land use of New Jersey and Uzbekistan is very different, this is not surprising.

Morlon et al. amplified part of type II αKS domains from three sites under Mediterranean vegetation in South Africa, Australia and Chile [95]. While the vegetation in all three areas may have been similar it is not clear that the soils were. The landscape varied between a flat plain (Australia), rolling hills (Chile) and along a ridge (South Africa). While they identified a large clade of αKS domains found in all three sites, overall their advice for bioprospectors was to sample broadly on the range of different continents.

The strongest arguments for the importance of distance have come from the Brady laboratory who have identified biogeographic effects between three soils in the arid deserts of the Southwestern United States [120], between 96 samples from these deserts and New England [105] (Figure 1.5) and among a collection of these 96 American samples and an additional 96 samples from China, Brazil, Alaska, Hawaii, Costa Rica, Ecuador, the Dominican Republic, Australia, Tanzania and South Africa [121].

In this last publication samples were characterized by next generation sequencing of Adenylation (AD) domains from non-ribosomal polyketide (NRP) pathways and KS domains from modular PKSI pathways. There was strong evidence of location being most important factor in determining sequence composition. The samples that were most similar to each other were geographically close to each other, even though they were often from different biomes. Similar biomes on different continents had less than three percent of sequences in common. Within samples that were close to each other, samples of

similar biomes were most similar to each other. Certain pathways were much more common in particular samples. Two areas, the deserts of the American South West and the Brazilian Atlantic forest, had soils where there was an order of magnitude more KS and AD diversity than other soils, making them attractive targets for bioprospecting.

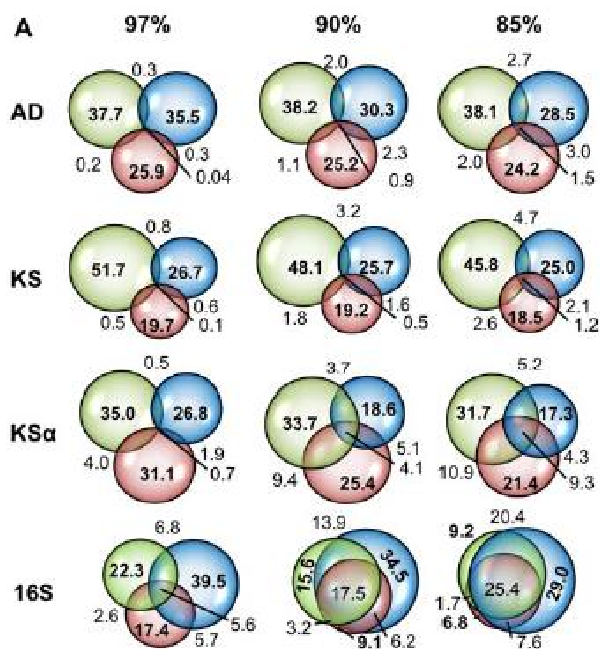


Figure 1.5. Overlap of sequences amplified from three soils from the desert sites in Arizona (green), Utah (blue) and California (red). Sequence are from non-ribosomal polypeptide (NRP) Adenylation (AD) domains, type I PKS ketosynthase domains (KS), type II PKS alpha ketosynthase domains (KSα) and 16S ribosomal sub units. After Reddy, B.V.B.; Kallifidas, D.; Kim, J.H.; Charlop-Powers, Z.; Feng, Z.; Brady, S.F. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* 2012, *78*, 3744–3752.

Two papers present evidence against distance controlling distribution of polyketide synthase genes. As part of a study that describes a screening method for PKSII and NRP sequences in metagenomics libraries, libraries of environmental DNA from Antarctica, Cuba and Europe were screened and clones sequenced [96]. No evidence was found of clustering by area.

Our laboratory extracted DNA from soils and street sediments in the Americas and Europe. DNA was amplified with PKSI specific primers before cloning and sanger sequencing [104]. Sequencing was extremely shallow and broad, with between three and 18 sequences for 21 samples. Several clusters of

sequences were found in many sites thousands of miles from each other. In some cases these clusters were from street samples (e.g., Ottawa and Budapest) where human activity might be expected to carry and mix Actinomycetes from around the world. However, a group of sequences that appeared to be myxobacterial was found in soils from the Canadian Arctic and Europe. While there were differences in sequences from site to site, these differences were comparable to those found between highly similar known pathways. Even if they are biogeographical effects in sequence composition, they may not be enough to affect polyketide function.

It does not matter to bioprospectors whether or not the biogeography of antibiotic production is controlled by distance. What matters is if or how much it inconveniences sampling. Two recent studies of the microbiology of New York parks suggest that location may not be an obstacle to bioprospecting. These studies used next generation sequencing to characterize either ribosomal or PKSI amplicons.

The ribosomal study, conducted by the Fierer group, extensively sampled Central Park (40,000 sequences/sample, 596 samples) [122]. Bacterial community structure was found to be controlled by pH, and varied greatly over the park, probably because of a wide range of soil management. Sequences from 52 of these samples were randomly picked and compared with 16S sequences from 52 soils (also 40,000 sequences/sample) taken from a broad range of environments in Canada, the United States, Peru, Argentina and Antarctica [72]. There was considerable overlap between the two datasets. 94% of the most common sequences were found in both and the diversity (as measured in rarefaction curves) was only slightly less in the central park dataset. The few sequences that were not found in Central Park were from deserts. The authors concluded that distance and climate is relatively unimportant in determining bacterial community structure.

A study carried out by the Brady group took 275 soil samples from 41 parks in the New York City area, amplified extracted DNA with of PKSI and non-ribosomal peptide primers, and sequence the amplicon with next generation sequencing. NRP amplicons were compared to similar amplicons from 96

soil samples from four areas of the continental United States, and they clustered apart, suggesting that they were New York City specific sequences. However, sequences encoding the biosynthesis of 11 known polyketide and non-ribosomal peptide medicines that had been initially discovered all over the world, could be found in one or more of the New York City samples [123]. The Brady group suggested that the same might be true of many samples and suggest that bioprospectors look deeply in a few samples rather than "*scratching the surface*" of many. They note that in some cases natural products originally found in marine microorganisms, can be found in terrestrial environments.

It could be argued that these results do not disprove the importance of barriers to dispersion in bacterial polyketde biogeography, as a large city will have many introduced bacteria from around the world and human activity will provide a broader range of substrates for them to grow. These effects have been found for earthworms in Australian cities [124] and fungi in Vienna, Austria [125]. In a practical sense, for polyketide bioprospecting, this biogeography will not matter as sampling large cities is relatively easy compared to sampling many soils from around the world. Biogeography nevertheless remains an important unsettled question in bioprospecting. A recent (2017) review of metagenomic antibiotic bioprospecting in soil considered that biogeographic studies would probably be essential to finding new antibiotics [126].

3. Bioprospecting for New Polyketide Discovery

Uncertainty about how polyketide synthase genes are distributed in the environment has not stopped bioprospecting. Most previous polyketide discovery has been through isolating Actinomycetes from soil. As fewer new polyketide are discovered this way, bioprospectors have begun to look at alternatives ways of sampling soil as well as alternatives to soil. The most common alternative to terrestrial soil is marine environments, a topic covered by many other reviews (e.g., [127]). Here we will

review recent advances in bioprospecting from terrestrial samples, including eukaryotic associated bacteria, desert soils, and disease suppressive soils.

3.1. Eukaryotic Associated Bacteria

One of the more promising environments for bioprospecting is the bacteria that are associated with or symbiotic to terrestrial eukaryotes. Jensen and Fenical in 1996 recommended bioprospecting from bacteria that live in close association with marine plants and animals, a strategy that many in natural product discovery community have since followed. Jensen and Fenical's rationale applies equally well to terrestrial macrobes, such as insects. Bacteria that are sheltered or in association with multicellular organisms inhabit environments that vary both between and within individual eukaryotic species, leading to innumerable, highly specific microenvironments. As the bacteria-host interaction becomes more complex, there is more chance of finding bacteria that have adapted to the specific environment. Furthermore the probability of finding that adaptation outside of the interaction is highly unlikely. Thus these bacteria-host interactions offer a highly unique environment with distinct and unique bacteria taxa. If these environmental adaptations affect polyketide production, then these bacteria represent an enormous potential for new polyketide discovery [128]. As well as this ecological argument for exploiting eukaryotically associated bacteria, we now know that many secondary metabolites once thought to be produced by eukaryotes are now known to be made by their associated bacteria and that the bacteria and their eukaryotic hosts have coevolved [129,130]. In particular, *Streptomycetes* are proposed as a phylogenetic group with a long evolutionary history of symbiosis [131].

3.2. Insect-Associated Bacteria

Insects are heavily colonized by microorganisms with between 1% and 10% of their biomass being microbial so it is not surprising that insects and microbes have developed mutualistic interactions. Kaltenpoth reviews mutualism between Actinomycetes and insects [132]. Insects have mutualistic

relationships with bacteria for nutrition but Actinobacteria are rarely involved in this (see [133]). Instead actinobacterial mutualism is usually defensive. These defensive interactions can be difficult to study. Defensive bacteria may not be on insects in large numbers for their full life cycle. Often they are found when the insects are in an immobile life stage such as eggs or larvae and so vulnerable to infection or predation. Nutritional mutualistic bacteria are vertically transmitted (i.e., transmitted from parent to offspring), however, defensive bacteria may be acquired from the environment [134].

There may be many undiscovered non-obligatory symbioses, particularly among fungus farming insects and insects that live in soil or rotting wood where there is continual contact with bacteria [104]. Both ants and bark beetles farm fungi, and Actinomycetes protect the farmed fungus from infection (discussed below). Actinomycetes make good defensive symbionts not only because the produce secondary metabolites but also because they can use a wide variety of substrates for growth including eukaryote excretions. In addition actinomycetal spores are easily transmitted between insects. Fungus farming bees have also recently been discovered [135] and unidentified Actinomyctes and their natural products may similarly protect them. In this section we will review social and non-social insect associated bacteria and their polyketides. Several other publications review this topic more extensively [136–139].

3.3. Social Insects as Sources for Bacterial Polyketide Discovery

Social insects are a promising source of new polyketides. Lombardo suggest that social behaviour is driven by the need to keep and coevolve with microbial symbionts [140]. Crowding together helps symbionts that can defend their host pass between insects. The best known social insects with antibiotic producing Actinobacteria are the attine ants (Figure 1.6). Debates over the relationship between attine ants and their Actinobacteria illustrate issues involved in bioprospecting bacterial symbionts.

Many ants cultivate fungus for food. In primitive forms of ant fungal agriculture, ants collect debris to feed fungi. The most developed form of this symbiosis is that of leaf cutting ants who grow fungus (genus Leucoagaricus) on leaf fragments that the ants provide. This fungus grows in underground gardens that the ants maintain (Figure 1.6B). These fungal gardens can be infected by pathogens, such as the fungus *Escovopsis* (Ascomycota: anamorphic Hypocreales). The ants, the fungus and the fungal pathogen appear to have coevolved [141].

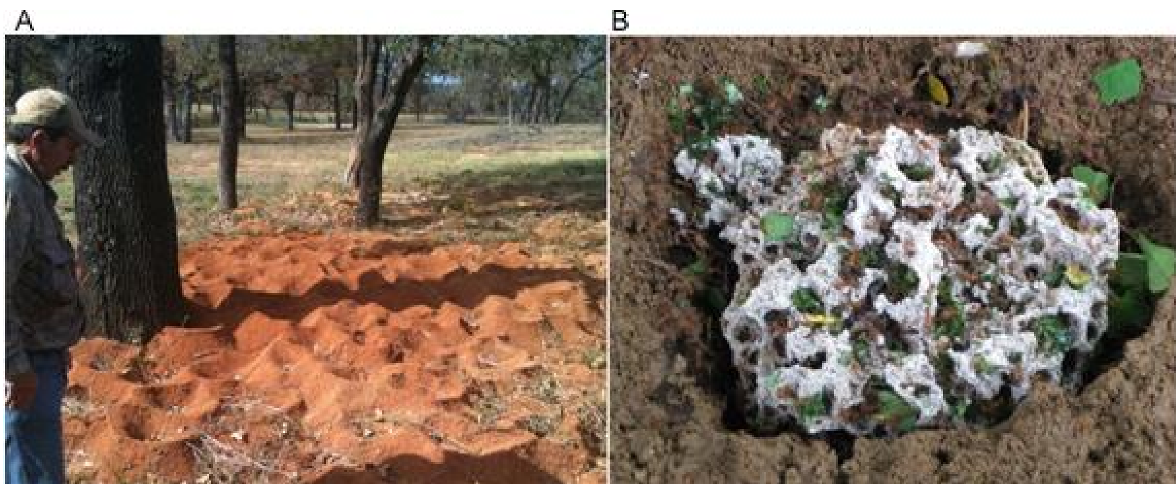


Figure 1.6. (A) Leaf cutter ant colonies in Texas. Photo courtesy of Texas A&M AgriLife Extension/Josh Blanek; (B) Leaf cutting ant nest in Costa Rica. In this case the nest was exposed when a rain barrel was moved, normally they are found at greater depth. Photograph courtesy of Herster Barres, *Reforest the Tropics*.

In 1999 Currie et al. [142] described a fourth part of this system. The bodies of these ants are covered in a whitish gray material originally thought to be exuded by the ants themselves. In a survey of 22 species of leaf cutting ants from Panama and Ecuador, Currie et al. identified this white material as Actinomycete biomass. This Streptomyces was vertically transmitted (i.e., down generations) and found in all species of attine ants. These Actinomycetes were tested for antagonism against several strains of fungi. There was little inhibitory effect on generalist fungi, but they did inhibit *Escovopsis*, which infects attine fungal gardens. The Actinomycetes strain also encouraged the growth of the fungi that the ants were farming. The system was a complex quadripartite symbiosis between ant, fungus, pathogen and

Actinomycete, in which the ant, fungus and Actinomycete are in an arms race against the fungal pathogen [141].

Later work showed that the Actinomycete in question was *Pseudonocardia* rather than a *Streptomycetes*. More importantly, evidence emerged that suggested that *Pseudonocardia* were picked up from the environment rather than passed down through generations. When leaf cutting ants were raised in several laboratories, 16S sequences from *Pseudonorcardia* on them clustered by laboratory rather than ant species [143]. Secretions from the *Pseudonocardia* were more antagonistic to the farmed fungi than the infecting *Escovopsis*. It was suggested that the *Pseudonocardia's* role may not be to protect the farmed fungi but the ants themselves or the ant nest and that there may not even be a mutualistic relationship [144]. Since then more evidence has been found suggesting that *Pseudonocardia* have coevolved with attine ants and *Escovopsis*. A comparison of colonies in Costa Rica and Panama found that pseudonocardial distribution followed that of *Escovopsis* [145]. Recently, evidence of lateral transfer of polyketide synthase genes between strains of *Pseudonocardia* has been found [146].

It is beyond the scope of this review to say whether leaf cutting ants get their Actinomycetes from their ancestors, other leaf cutting ants, or the environment (see [116–119,147]). There are implications for sampling that apply to all cases where there are these alternatives. If polyketide producing Actinomycetes in eukaryotes are acquired vertically, bioprospectors should sample by eukaryote phylogeny. If polyketide producing Actinomycetes are continually acquired from the environment, bioprospectors should view ant nests as an environment that is enriched in bacterial secondary metabolite producers.

Most of the work on antibiotic production by attine associated bacteria has been done on Actinomycetes (e.g., [148]) although β-proteobacteria have also been found in fungal gardens with antifungal activity specific to fungal pathogens [149]. Both known and novel polyketides have been

found to be produced by attine ant associated Actinomycetes (Figure 1.7). Dentigerumycin, **1**, is a mixed non-ribosomal peptide type I polyketide with a novel core structure that was produced by a *Pseudonocardia*. The pseudonocardial strain was isolated from the cuticle of the attine ant species *Apterostigma. dentigerum* from Panama. Dentigerumycin has [150] antifungal activity against *Escovopsis* and also inhibits several strains of Candida.

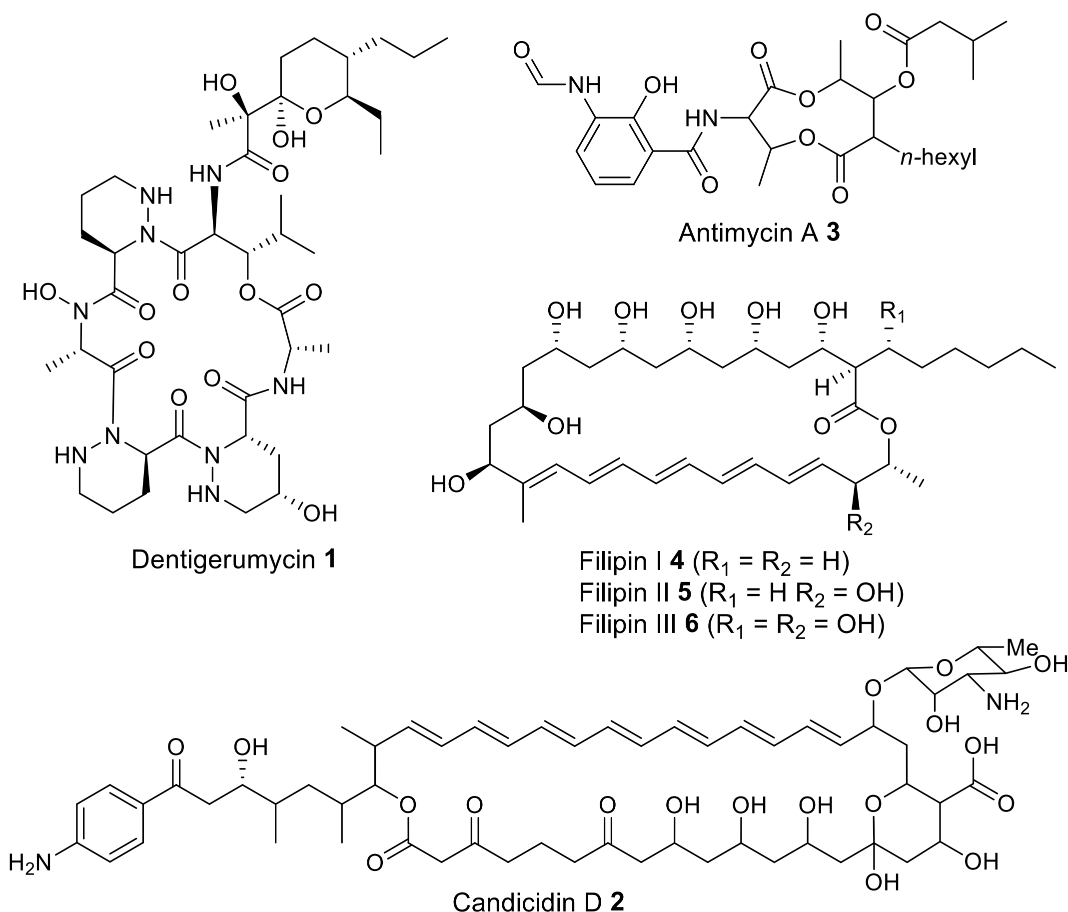


Figure 1.7. Polyketide and non-ribosomal peptide natural products isolated from bacteria associated with social ants.

Pseudonocardia is not the only ant-associated bacterium producing polyketides. A *Streptomyces* from a fungal garden of a Panamanian ants nest produced candicidin (**2**) [148]. The *Streptomycetes* strain that produces this polyene antibiotic has also been also been isolated from ants' nests in Trinidad [151].

The study that identified the candicidin producing *Streptomyces* in a Trinidadian ant nest was part of a pair of publications. The first isolated a number of *Streptomycetes* and *Pseudonocardia* from the nests and found that the *Streptomyces* (strain S4) produced candicidin. The *Pseudonocardia* was shown to produce an analog of nystatin called nystatin P1. While the complete structure of nystatin P1 was not solved, both the biosynthetic machinery, deduced from the genome sequence, and liquid chromatography–mass spectroscopy (LC–MS) analysis of the culture media are consistent with a nystatin A_1 derivative possessing an additional dideoxy amino hexose [123].

A following study sequenced the genome of the *Streptomyces* S4 strain [152]. The biosynthetic gene cluster encoding candicidin was identified in the genome as well as a number of other uncharacterized biosynthetic gene clusters. The authors were able to assign one of these gene clusters to the known mixed polyketide-non-ribosomal peptide antimycin (**3**). When the candicidin and antamycin gene clusters were disrupted in *Streptomyces* S4, it was still active against *Escovopsis*. This suggests that additional natural products encoded by the many uncharacterized gene clusters in this strain may be responsible for the antibacterial or antifungal activity. Isolation and characterization of these compounds is still unaddressed.

Recently a polyene was discovered on two strains of *Pseudonocardia* in the La Selva biological station in Costa Rica. Selvamicin was found to have less antifungal potency but was more soluble than the similar nystatin. The genes for selvamicin production were found on the genome of one pseudonocardial strain and in a plasmid on the other. The genes were more similar to each other than they were to the genomes of either *Pseudonocardia*, suggesting lateral gene transfer [146].

Social ants do not only use fungi for food. Gao et al. describe the use of fungi by *Allomerus* ants to build traps for insects on plants [153] and found using activity guided fractionation and LCMS-based comparisons to authentic standards that Streptomycetal isolates from *Allomerus* ant bodies produced the filipins **4**–**6**, a family of type I polyketides also produced by *S. avermitilis*.

While fungus growing attine ants are the main insect herbivore in the tropics of the Americas, in tropical sub-Saharan Africa and Southeast Asia this role is taken by fungus growing termites. Their ecological roles are similar. Unlike the attine ants, the fungi that termites eat (a Basidiomycetes of the genus *Termitomyces*) appear to spread laterally rather than vertically [154]. As with the attine ants, unwanted fungi can compete with the farmed fungi. In particular, the Ascomycete *Pseudolxylaria* appears to compete with *Termitomyces* for food. As with the ants, Actinobacteria likely play a role in combating these pathogens. A study which cultured Actinobacteria from 30 South African termite nests of three genera of termites found that while the Actinobacteria (*Streptomycetes, Micromonospora*, and *Actinomadura*) often had anti-fungal activity, they were more likely to inhibit the *Termitomyces* than the *Pseudolxylaria*. There was no evidence of Actinobacteria clustering by termite genera and their closest 16S matches were often to Actinomyetes isolated from other insects such as the Southern Pine beetle [155], suggesting that the Actinobacteria in the termite four way relationship were acquired from the environment rather than vertically transmitted.

Potent and often structurally unique polyketides can be produced under these conditions. A Streptomycete strain isolated from South African termites was cultivated and found to produce two related hybrid non-ribosomal peptide polyketide derived compounds, microtermolides A and B (**7** and **8**, respectively, Figure 1.8) [127].

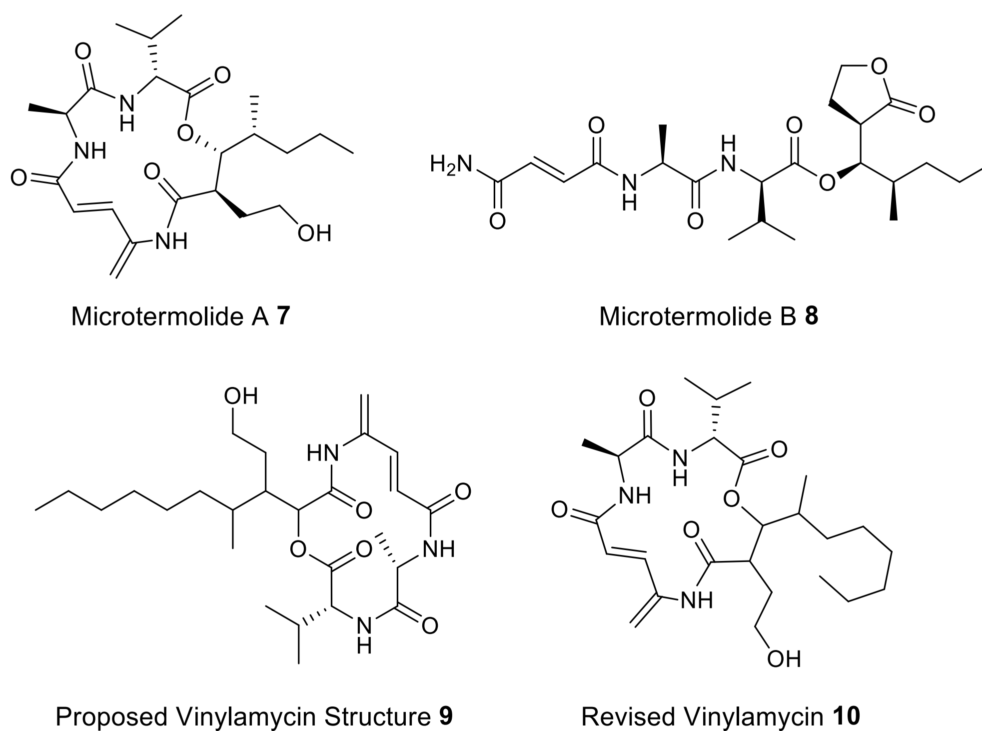


Figure 1.8. Microtermolides, produced by termite-associated microbes, and the originally proposed and revised structure of the related NRP vinylamycin.

The structure and spectroscopic data of microtermolide A (**7**) was similar to that of vinylamycin (**9**, Figure 1.8), isolated in 1999 from a soil derived Streptomyces strain and prompted the structural reassignment of vinylamicin to the correct structure shown (**10**) [156] Microtermolide B **8** is a related acyclic nor derivative of **7** and is one of the only known linear depsipeptides produced by a Streptomycetes. [157] A second Streptomyces isolate M56, which is closely related to *Streptomyces malaysiensis* 1160, was isolated from the fungal comb of a South African termites nest by the same research group. This strain produced a new and highly unusual geldanamycin (**11**) analog, possessing an unprecedented [6.4.0]-bicyclic core, named natalamycin (**12**, Figure 1.9**)**, as well as a number of related analogs **13**–**19** [158].

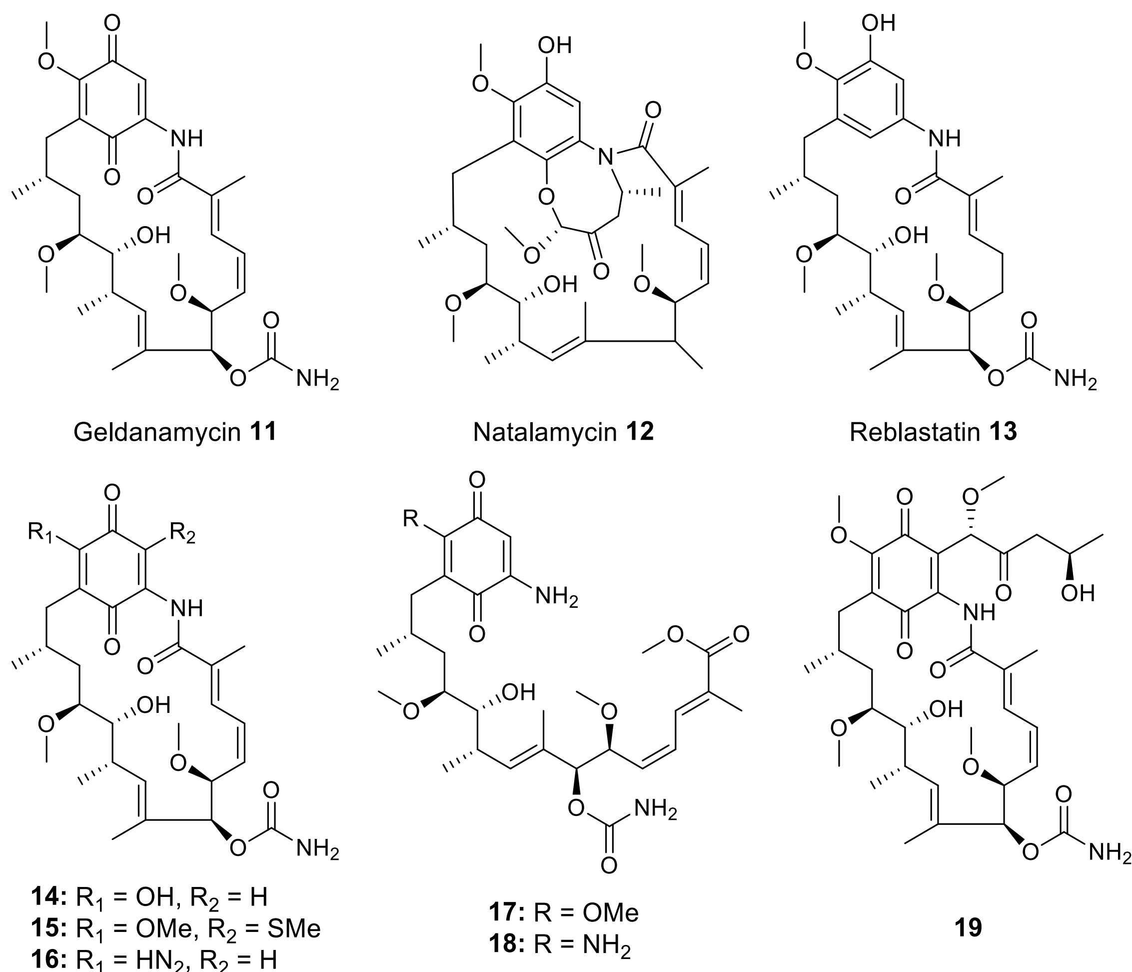


Figure 1.9. The known antibiotic geldanamycin and the related analogs discovered from a termite associate *Streptomyces*.

While the bulk of polyketide producing strains from termite colonies that are reported in the literature are Actinomycetes, there are some examples of non-actinobacterial producers. For example, two *Bacillus subtilis* strains were isolated from a termite colony in South Africa produced the known compound bacillaene A [159].

A third group of fungus cultivating insects is the Southern pine bark beetle, *Dendroctonus frontalis*. This beetle, a major pest of pine trees in the South Eastern United States, cultivates the fungus *Entomocorticium* for its larvae in the pine phloem. This fungus in turn suffers from the parasite *Ophiostoma minus*. Two closely related strains of *Streptomycetes* were isolated from the beetles. One

was a red colony, SPB74, the other a white colony, SPB78. The SPB74 strain produced the novel polyene mycangimycin (**20**, Figure 1.10) which was twenty times more active against the pathogenic *Ophiostoma minus* than the farmed *Entomocorticium* [160].

After testing different cultivation conditions with Streptomyces SPB78, this strain was also shown to produce antifungal compounds. The frontalamides **21**–**25**, are mixed polyketide-non-ribosomal peptides containing a tetramic acid group and are active against *Ophiostoma minus*. The frontalamides are related to a number of known natural products **26**–**31**, including dihydromaltophilin (**26**) whose biosynthetic pathways has been previously identified the genome of *Lysobacter enzymogenes* [161]. Primers were designed based on the dihydromaltophilin gene cluster and it was confirmed that both SPB74 and SPB78 possessed similar sequences. Ultimately LCMS analysis provided support for production of a frontalamide like compound from SPB74. Dihydromaltophilin gene cluster primers also gave products from soil isolates [162].

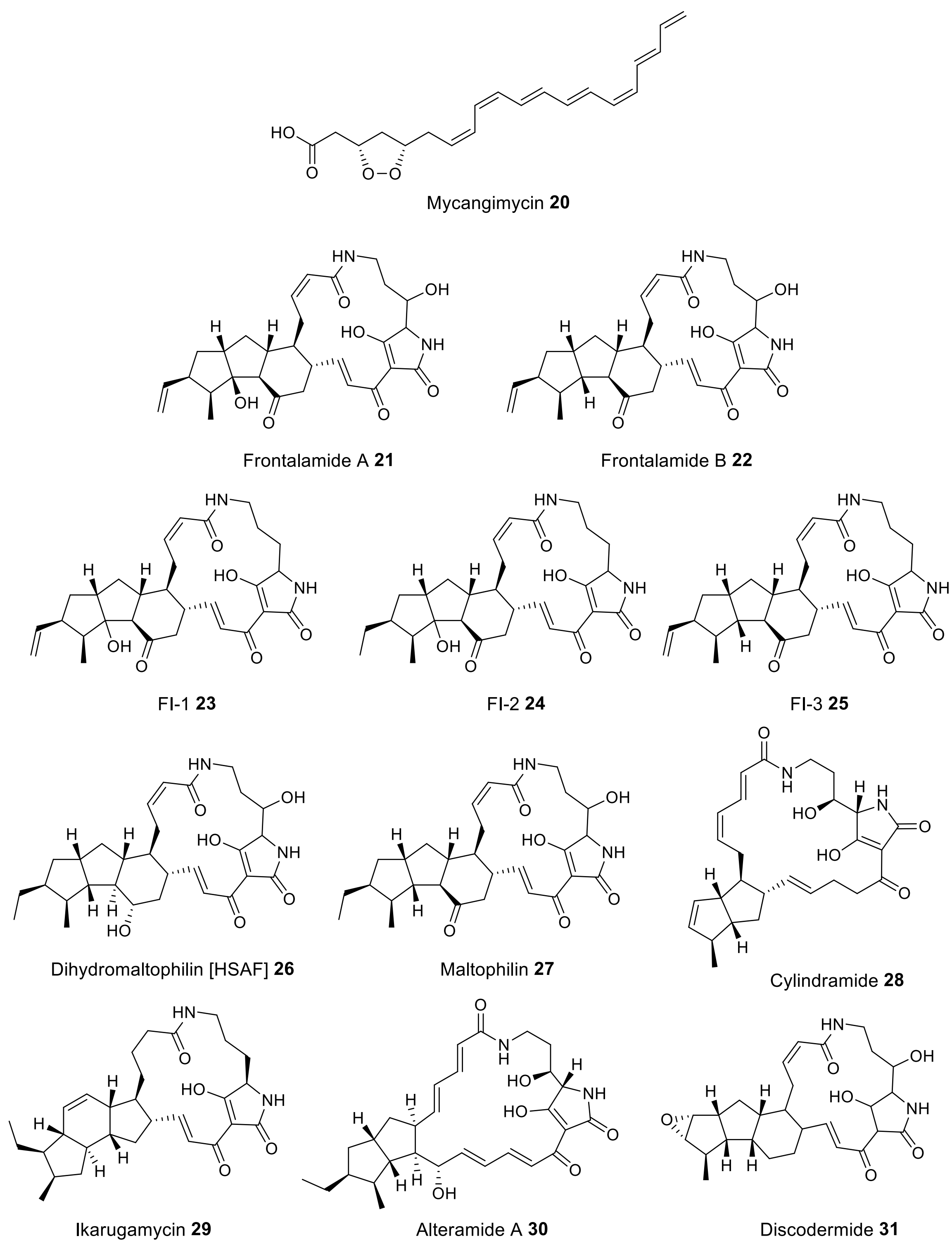


Figure 1.10. Polyketides discovered from beetle associated bacteria and known related compounds.

Social insects that do not cultivate fungi may also have polyketide producing bacteria in their nests. *Coptotermes formosanus*, the Formosan termite is originally from Southern China. It is an invasive pest that has now spread to Taiwan, Japan, Hawaii and much of the South Eastern United States. It eats wood and builds large underground nests with foraging galleries that can be over 100 metres long. The centres of nests are reinforced with a mixture of chewed wood and faeces known as "carton material". Streptomyces with antagonistic activity towards termite pathogens were isolated from this carton material. These Streptomyces appeared to reduce the death rate of *Coptotermes formosanus* termites when they were exposed to fungal (*Metarhizium anisopliae*) pathogens [163]. Similarly, Actinomycetes with antagonistic activity against *Pseudomonas aeruginosa*, *Escherichia coli*, *Staphylococcus aureus*, *Serratia marcescens*, and *B. subtilis* were isolated from the nest of the paper wasp *Polistes dominulus*, an invasive European wasp [164]. However, no compound responsible for these activities has been isolated.

3.4. Non-Social Insects as Sources for Polyketides

Polyketide producing Actinomycetes have also been found associated with non-social insects. The larvae of some wasps are carnivorous and are laid in the bodies of paralyzed insects. In the case of the European Beewolf wasp (Figure 1.11), each larva is left in an underground brood cell with one to six paralyzed honeybees. Fungal infection from the surrounding soil is a frequent problem [165]. A white biomass produced in the antennal glands of the female beewolf is secreted into the brood cells which contains *Streptomyces philanthi* spores [166]. When the spores were removed many of the larvae died [167,168].

S. philanthi can either be vertically transmitted or acquired from the environment. *S. philanthi* strains cultured from different genera of beewolves from Europe, Asia, Africa and South America, are not resistant to antibiotics, cannot take up inorganic nitrogen, and are associated with a particular

beewolf species. Because of this, they are likely vertically transmitted. In contrast the *S. philanthi* cultivated from North American beewolves are resistant to a broader range of antibiotics and share the antenna glands of their beewolves with other Actinomycetes. Because of this, they appear to have been acquired from the environment [169].

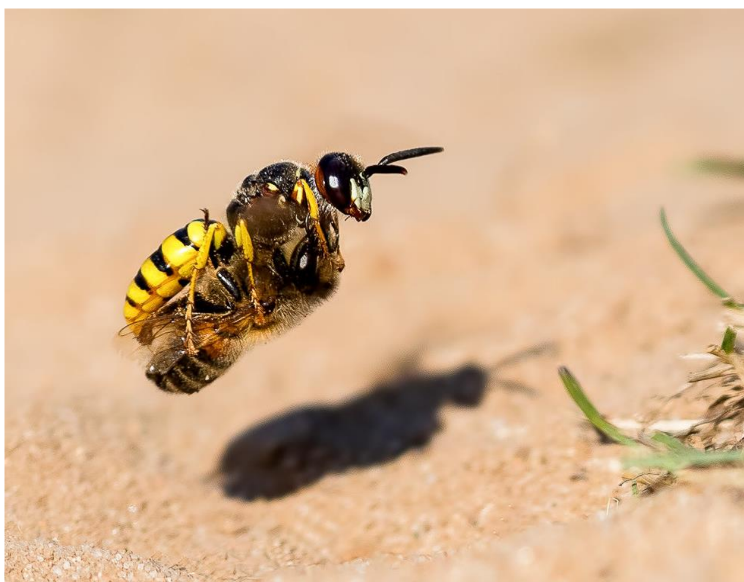


Figure 1.11. Beewolf with honey bee prey. Courtesy of Simon Jenkins http://www.simon-jenkins.photography.

Beewolf larvae are protected by several potent antibiotics, shown in Figure 1.12. Extracts of European *Philanthus* beewolf larval cocoons were analyzed and a cocktail of eight known antibiotics were detected. This included streptochlorin (**32**, Figure 1.12) and seven variants of piericidin **33**–**40**. Streptochlorin was previously isolated from the culture broth of a marine Streptomyces [170] and piericidin from the culture broth of *Streptomyces mobaraensis* [171]. This combination of eight compounds was antagonistic against a broad range of fungi and bacteria [172].

Mud dauber wasps are solitary predatory wasps that build nests from mud. Actinomycetes were isolated from two species of mud dauber wasp, *Sceliphron caementarium* and *Chalybion californicum*, and tested for production of secondary metabolites. LC–MS analyses of culture medium extracts showed the presence of a number of active compounds. While a number of these were known

compounds, including the highly toxic vacuolar type H+ ATPase inhibitors bafilomycins A1 and A2 (**41** and **42**, respectively), and mycangimycin (**20)**, which was also found in Southern Pine Bark beetles [173], a new polyene macrolactam, sceliphrolactam (**43**), was also discovered. This compound has potent antifungal properties against amphotericin B-resistant *Candida albicans* [174].

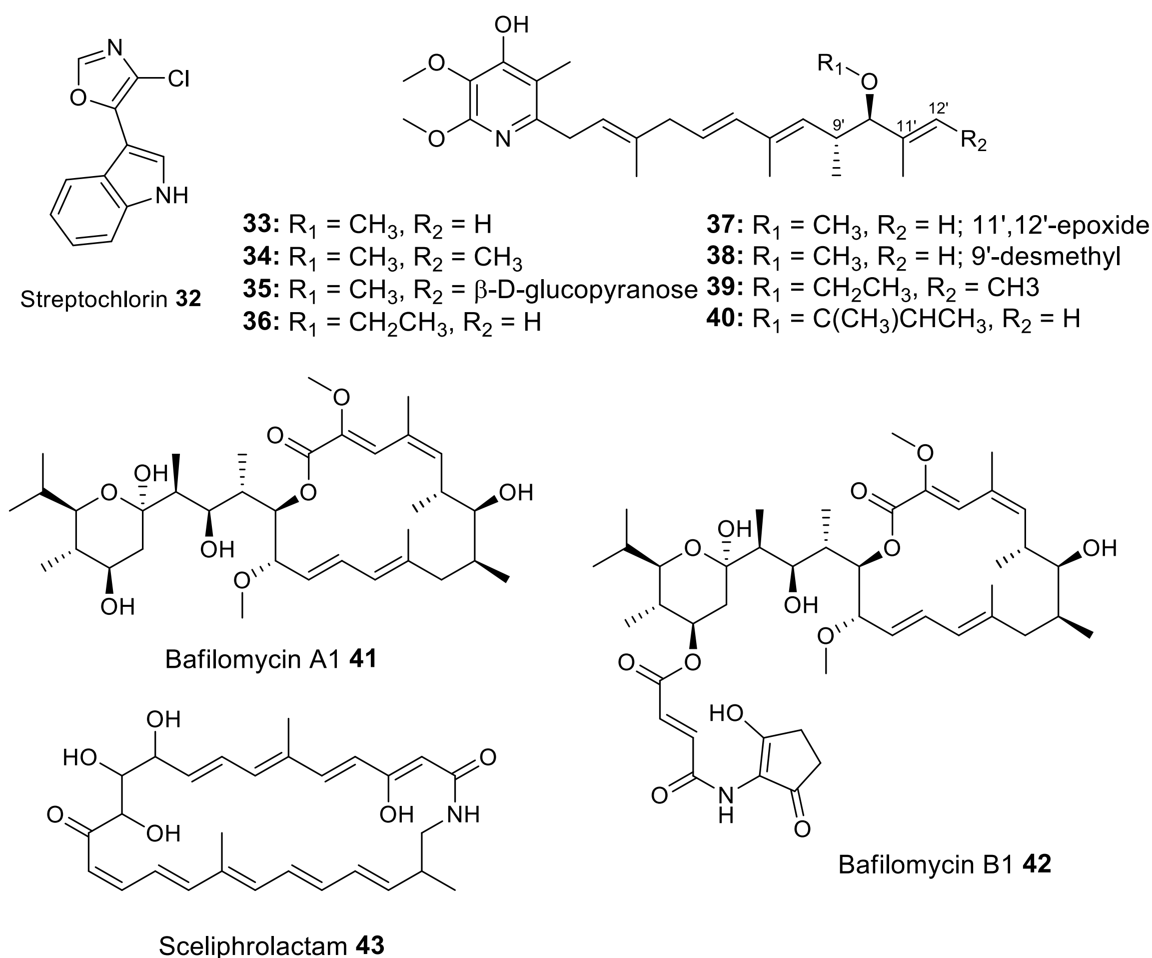


Figure 1.12. Bacterial secondary metabolites associated with non-social wasp species.

The Dung beetle lays it's larvae in a pellet of faeces, another environment with potential for infection. A Streptomyces (*Streptomyces* strain, SNA112) from a pellet used by *Copris tripartitus*, the Korean dung beetle, produces tripartilactam (**44**, Figure 1.13), a novel tricyclic lactam. The compound is

inactive against all bacteria, fungi and cancer cells that it was tested against, but does inhibit Na^+/K^+ ATPase [175].

The best known case of a bacteria living on a non-social insect producing a polyketide (in this case a mixed NRP type I PKS) is the case of pederin (**45**) found on Paederus and Paederidus beetles. Pederin causes dermatitis on the human skin and appears to deter predators from eating the beetles, and has strong anti-tumour activity. Pederin is produced by a *Pseudomonas* (a gamma proteobacteria) that is transferred vertically through female beetles. Both Pederin itself and its intermediates are similar to mycalamide A (**46**) made by a bacteria that inhabits marine sponges [176]. Recently a compound highly similar to pederin (diaphorin (**47**)) was found to be produced by a β-proteobacteria which grew on the Asian citrus psyllid *Diaphorina citri* [177].

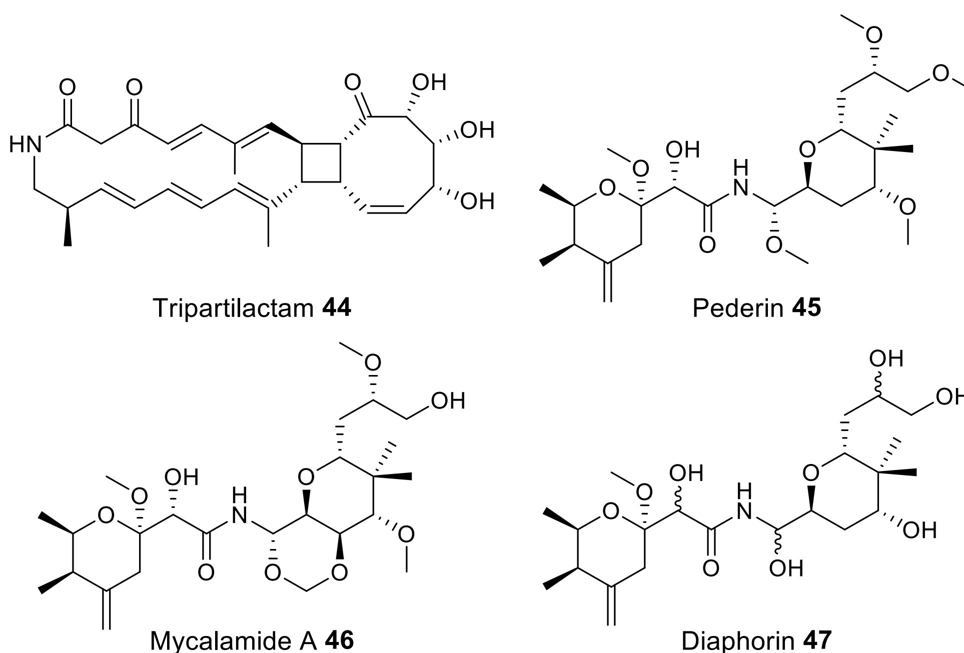


Figure 1.13. Tripartilactam and pederin from non-social beetles and related natural products from non-insect associated microbes.

Insect-associated bacteria are the most studied of the environments that we review in this article. A large literature is available to help pick where to sample. The main question in picking samples is

whether the kind of insect or the environment that the insect produces is more important. As some of the social insects discussed here are pests (attine ants, Southern Pine Bark beetles, the Formosan termite), plant protection or pest control offices may be able to help find sampling sites.

3.5. Desert Soils

Deserts are areas with less than 200 mm of precipitation a year and they can be found from the equator to the poles. Deserts may have arid or hyper arid climates. An arid climate has a ratio of mean annual rainfall (MAR) to mean annual evaporation (MAE) of 0.2–0.03 and precipitation of 100–300 mm/year. In hyper arid areas, where there is no vegetation, the MAR/MAE ratio is less than 0.03 and there is less than 100 mm of precipitation/year [178].

Several research groups are bioprospecting desert soils for pharmaceutically useful secondary metabolites as desert soils are often enriched in Actinobacteria. Actinobacteria have been found to be the most common group in an Antarctic desert soil [179] and Atacama desert soils [180,181] and the second most common group in a Saharan soil [182]. This is not the case for all studies though, their relative percentage in the bacterial community has been found to vary with season in the Negev [183]. Furthermore in a study of soils of the Gobi and Taklamakan, the Actinobacteria were found to be relatively unimportant compared to the Bacilli [184].

The most reliable assessment of desert soils for actinobacterial enrichment and the presence of polyketide synthase genes are studies that compare a range of desert soils with other soils using molecular methods. This minimizes bias due to primers and sample processing. Fierer et al. in 2012 [72] compared soil DNA from three hot arid deserts from the Southwestern United States with soil DNA from six Antarctic cold hyper-arid deserts and seven non desert soils (arctic tundra (1), prairie (1) and tropical (2), temperate (2) and boreal forest (1)). DNA was characterized by next generation sequencing of 16S amplicons and shotgun metagenomics sequencing. According to both metagenomic shotgun sequencing

and amplicon sequencing, the percentage of Actinobacteria was highest in cold desert soils, followed by hot desert soils. Fierer et al. ascribed this to the high pH and dryness of these environments. The bacterial communities of the non-desert soils were relatively similar to each other. The bacterial communities of cold and hot desert soils differed both from each other and non-desert soils. 16S sequence results found that hot deserts had comparable diversity to other non-desert soils, while hyper arid Antarctic cold deserts were less diverse.

Xu et al. 2014 [185] reanalyzed Fierer et al.'s [72] results with literature data from 17 other sites, including 12 English grassland and a Brazilian Mangrove soil. Again this study showed that desert soils were enriched in Actinobacteria, although most were not *Streptomyce*s, but rather *Bifidobacteriaceae*, *Mycobacteriaceae*, and *Frankiaceae*.

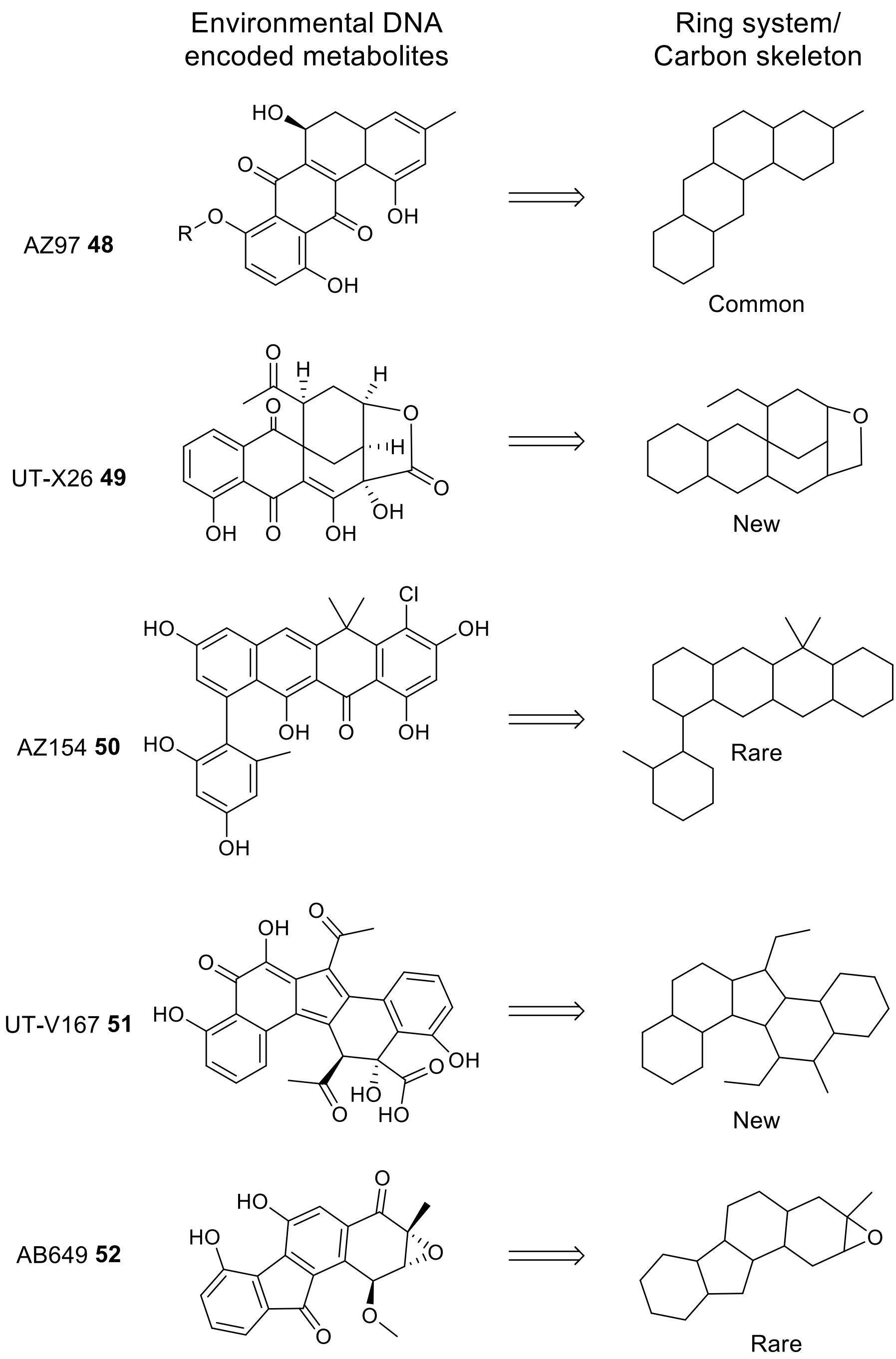


Figure 1.14. Type II polyketide molecules coded for by desert soil metagenomic clones. After Feng, Z.; Kallifidas, D.; Brady, S.F. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. USA.* 2011, *108*, 12629–12634.

One could argue the relatively large actinobacterial fraction of the bacterial community means that these soils are a good source of new polyketide natural products. In contrast, Fierer et al. [72] found

that fewer of the metagenome shotgun sequences in his study were implicated in virulence and defence, including antibiotic resistance, compared to other soils. They suggested that the harsh desert conditions are less conducive to bacterial growth, reducing microbe-microbe competition, leading to less antibiotic production and resistance.

The Brady laboratory used environmental DNA from hot desert soils from the same area to bioprospect for natural products and came to the opposite conclusion. Powers et al. [105] compared soil DNA from desert, forest, farmland, grassland and salt-water marsh soils by amplifying with 16S, NRP A domain, and PKSII KSα specific primers. Sampling sites were in the US Southwest and New England. The KSα diversity, (from type II aromatic polyketide biosynthesis) correlated well with the 16S actinobacterial fraction of the soil bacterial community. Most known PKSII producers are Actinobacteria. The study found that arid soils contained the most diverse KSα domains and that New England forest and salt march soils contained the least.

The environmental DNA from these desert soils were then used to make metagenomic libraries that were screened with primers for the KS domains of type I [186]and type II [187] Polyketide synthase pathways. This work discovered of several novel aromatic type II polyketides. Three environmental clone libraries from California [188], Arizona, and Utah [187] were screened with Type II specific primers and five Type II pathways were identified and heterologously expressed in Streptomyces. One of the pathways produced a compound similar to the known polyketide landomycin E, an angucycline (**48**, Figure 1.14) [188]. The four remaining pathways produced far more unique compounds **49**–**52**. Two possessed new ring systems, a novel naphaquinone called erdacin (**51**) and a new aureolic acid (**49**). The structures of the last two compound were rare variants of the angucyclines (**52**) and pentangular polyphenol **50** [187]. Since this [187], the Brady group has screened desert soil metagenomics libraries for more pentangular polyphenols [19,189] and have found arixanthomycins, calixanthromycin A (**53**,

Figure 1.15), and arenimycins C (**56**) and D (**57**). Desert soils are promising for aromatic type II polyketide discovery.

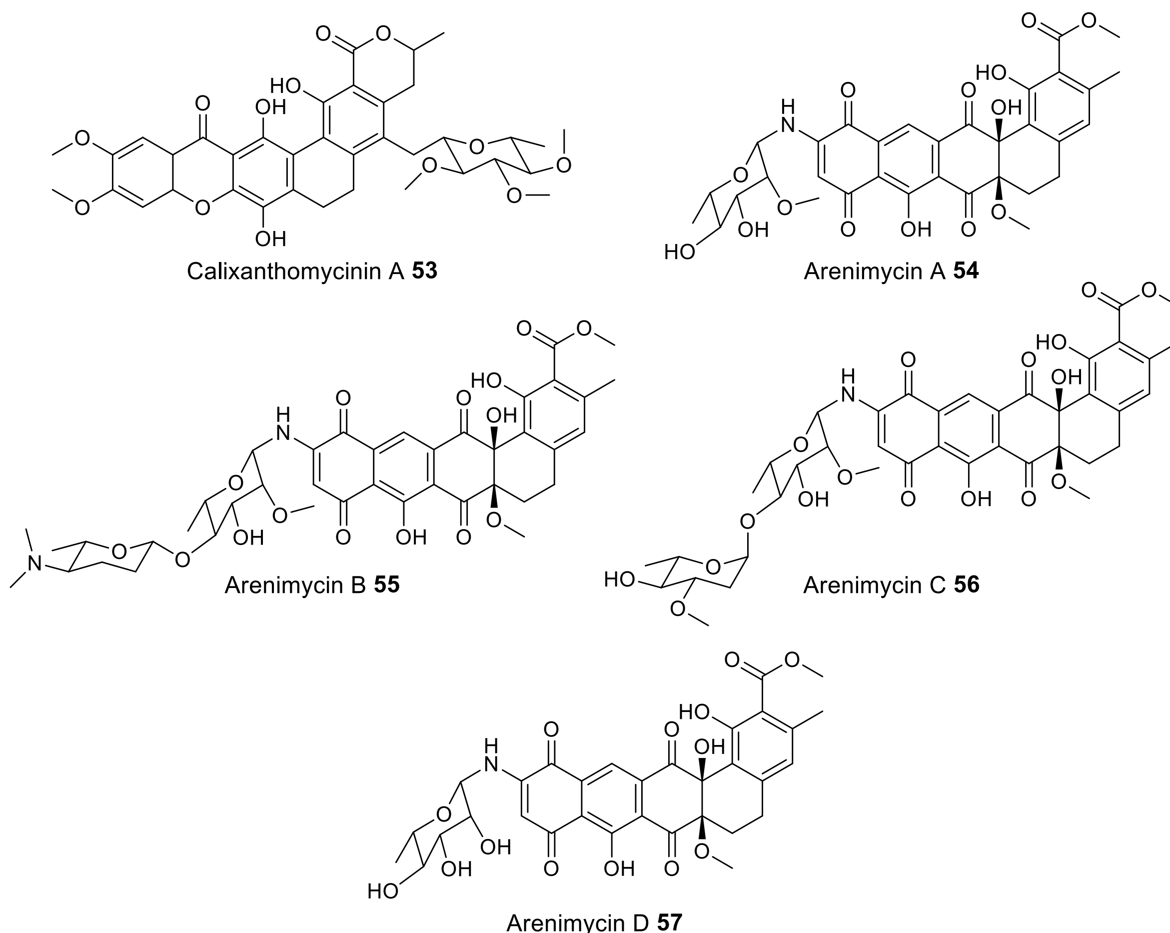


Figure 1.15. Pentangular polyphenols discovered from desert soil derived metagenomic libraries.

Several other groups have found polyketides in desert soils by culturing Actinomyces from them. Several non-streptomycetal actinomycetal isolates from the Algerian Sahara have been found to produce antibiotics. An *Actinomadura* isolate was cultured and made an antifungal agent, however, the final structure of the compound was not determined [190]. Similarly a *Streptosporangium* isolate produced several of presumed glycosylated aromatics natural products with antibacterial activity against Gram-positive bacteria [191]. Again the structure of these compounds was not fully determined. Lastly a

Saccharothrix isolate also from the Algerian Sahara made two similar highly novel polyketides (**58** and **59**, Figure 1.16) with potent antifungal effect on some filamentous fungi but not yeasts [192].

The Atacama Desert in Northern Chile is the driest and oldest non-polar desert in the world. Most of the Atacama Desert is hyper-arid and the most significant source of water is sea fog that condenses and supports limited plant life. There is an extremely hyper arid valley near the Yungay region where the MAR/MAE is less than 0.002. Fog does not reach this area because of a mountain barrier between it and the Pacific Ocean. This region has no vegetation and soils have very little organic matter. It is often impossible to detect bacteria in these soils which are often described as Mars like [193].

Five studies have used molecular methods to characterize the bacterial community phylogeny of Atacama Desert soils and found that the bacterial communities of these soils are very different from other desert soils. When Fierer et al. compared the bacterial communities of deserts and other biomes they found that most hot desert soils were as or more diverse than soils from forests and grass lands (8000–12,000 OTUs). Hyper-arid Antarctic soils could have half this number of OTUs [72]. Soils from the Atacama Desert are orders of magnitude less diverse than this. Drees et al. [194] compared the bacterial community of soils in the core arid region with more humid areas closer to the ocean and further inland at higher altitudes using a 16S community fingerprinting method. There were two distinct clusters of fingerprints, one from the core arid region and a second from a coastal sample and samples from higher up in the mountains. This same research group from the University of Arizona, later used next generation sequencing of 16S amplicons to characterize three samples taken from the edge of the Mars-like core of extreme hyper-aridity. In these samples 70 and 77% of sequences were actinomycetal. However, the community structure was far less diverse than that of most soils in Fierer's study with 123 and 142 OTUs found in 2200 and 3560 sequences, respectively. [180].

Two studies using cloning and sequencing of 16S amplicons have found that the bacterial community of soils from the hyper arid core were dominated by a single clone. In one case the

community was almost entirely actinomycetal, with sequences ranging from 91% to 95% similarity to a known Frankia genus making up 94% of sequences [181]. A second study [195] however, only found bacilli in 244 clones from a single sample. 154 of these 244 amplicons were a single sequence.

A recent report describes three even more arid areas in the Atacama Desert than the Yungay region. The authors report that the bacterial community as determined by 16S community fingerprinting as being actinobacterial with sequences from Actinobacterium, Aciditerrimonas and Geodermatophilus genera, Proteobacteria, Firmicutes and Acidobacteria [196].

In November of 2004 Bull, from the University of Kent, sampled three sites in the Atacama desert in arid, hyper-arid, and extremely hyper arid regions [197]. These samples were used to culture Actinobacteria using selective media and 46 strains were found from the *Amycolatopsis, Lechevaliera* and *Streptomyces* genera. Two Streptomyces strains from the hyper arid site (C34, C38), a salt flat known as the Salar de Atacama were closely related to *Streptomyces leeuwenhoekii*. Each of these strains has been a source of a novel type I polyketide family, the atacamycins and the chaxamyins.

The atacamycins are a family of 22 membered macrolactones [198] produced by strain C38. There are three closely related forms of atacamycin A, B and C (**60**–**62**, respectively, Figure 1.16) that vary by oxidation of a common macrolactone ring. These compounds were weakly inhibitory to *Ralstonia solanacearum* but atacamycin A and B showed cytotoxic activity against several human tumour cell lines. Strain C34 produced four members of the chaxamycin family, compounds A–D (**63**–**66**). Chaxamycins are napthalene ansamycins. Chaxamycin D (**66**) showed activity against of methicillin-sensitive and resistant *S. aureus* (MRSA) although it was not as active as rifampicin. Chaxamycins A–C were inhibitory toward ATPase activity of human Hsp90, a test of anti-tumour activity for ansamycins [199].

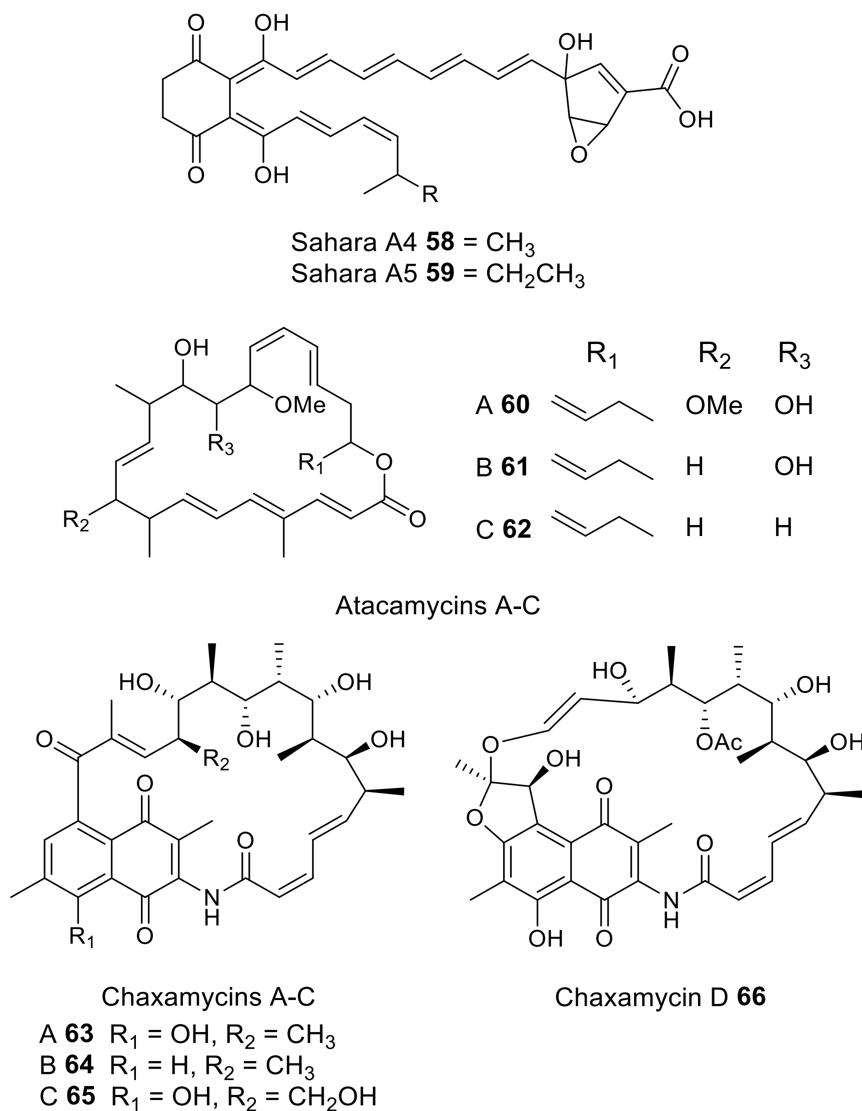


Figure 1.16. Natural products from desert isolates.

Both metagenomics and culturing are thus finding new actinobacterial polyketides from desert soils. The soils of the hyper arid Atacama Desert have a distinct bacterial community to that of deserts such as the Sahara or the Mohave and it is not yet certain how this affects the polyketides that their bacterial communities produce.

3.6. Disease Suppressive Sediments

Soils that suppress disease have been a target for antibiotic polyketide discovery. In the most direct sense disease suppression can mean using soil to cure human disease, such as red soils from the Mediterranean climate region of north western Jordan. These soils are used to treat skin infections in traditional medicine. When inoculated with *Micrococcus luteus* and *Staphylococcus aureus*, bacterial isolates from these soils are more likely to produce antibacterial compounds including actinomycin [200]. Usually though disease suppressive soils refer to soils on which soil borne plant diseases, particularly fungal pathogens, are less likely to infect crops.

Disease suppression alone does not mean that a soil is a good target for bioprospecting. Suppression can be caused by a soil property such as pH. Certain crops can also produce chemicals which suppress disease. Even when suppression is caused by the soil microbial community this does not necessarily mean that antibiotics are being produced. Any soil that is sterilized will become much easier for a pathogen to colonize as it does not have to compete with microbes already there. The suppressivity of the non-sterile soil is called natural suppression and is caused by the whole soil microbial community [201]. In contrast induced suppression occurs when certain crops are grown, certain crop growing sequences are used, or the crops are managed a certain way [202]. Induced suppression can be due to antibiotic production.

The different roles of natural and induced suppression are reviewed by Janvier et al. [203]. The test of microbial induced suppression is if sub-samples of this soil can be mixed with non-suppressive soils making it suppressive to a particular pathogen and if the suppressive effect of a soil is destroyed by autoclaving. If this is the case, the microbial community of these soils is probably causing induced suppression.

Two publications suggest suppressive soils as a good target for antibiotic bioprospecting [204] [205] but recommend very different kinds of soil. The EU funded METACONTROL project focused on

bioprospecting from disease suppressive soils to find new antibiotics. Five suppressive soils from across west Europe were used. The most suppressive of them was a grassland that had recently been converted to arable land [206]. Researchers ascribed this suppressiveness to the biodiversity of the soil plant cover and microbes [207]. In contrast other studies have found that soil suppressiveness increases under monoculture, when a single crop is grown at the same site for a long time [202,208]. These soils were also suggested as a target for bioprospecting for antibiotics [205].

The suppressive effect of these soils under long term monoculture has been ascribed to the effect of coevolution [209]. If a range of plants are grown on a soil, there are many different kinds of food for the microbial community. Ecological niches can develop as certain microbes specialize in certain substrates. If a single crop is grown over a long period there is a much narrower range of substrates for microorganisms and there is more competition and inducement to use antibiotics. As these conditions persist, there is an "arms race" between bacteria as they use more of their genome to produce more diverse antibiotics.

Kinkel et al. [209] list other factors that encourage the development of soil suppressivity. The soil should receive a lot of readily available substrate for the microbial community to consume. Readily available substrates ensure that the microbial community can produce the metabolically expensive secondary metabolites. Additionally readily available substrate ensures high population densities, encouraging antagonism. Finally moderate ploughing will distribute antibiotic producing bacteria through the soil where they can outcompete other microorganisms that specialize in particular soil niches. However, too much ploughing can homogenize soil microbial communities to the point where there is less coevolution and thus fewer new antibiotics.

Soils can be made suppressive by adding organic matter to the soil such as manures, green manures (plants that are grown and then ploughed into the soil), peat, and compost [210,211]. The best known and most effective of these organic materials is compost [212]. The suppressive effects of compost in

soil vary depending on the stage of compost development and the materials used [213] [214]. Antibiotic production is thought to be one several ways that compost microorganisms suppress disease, with the others including competition for nutrients and parasitism [212].

Hadar and Papadopoulou [212] review coevolution and suggest that this probably does not explain the suppresive effect of compost because the bacteria in compost are not in the soil long enough to coevolve with the plants and microbes in the soil. Instead they suggest that as raw organic matter develops into compost certain microbial groups are selected for. The actinobacterial fraction of the soil bacterial can increase after adding manure [215], although others recommend the addition of organic amendments as they provide the conditions for suppressive bacteria to grow [216]. A review of over 2000 studies of the suppressiveness of organic amendments found that while actinobacterial prevalence had an effect on suppressivness, it had less of an effect that total bacterial and fungal prevalence [211] suggesting that some of the suppresive effects of compost microbes is due to natural suppression.

There is evidence that composts can produce antibiotics. Sometimes instead of adding compost directly, it is soaked in water for several hours or days and the supernatant, called compost tea is sprayed on the plant leaves to prevent plant disease (see [217] for full discussion). This compost tea can retain much of its antifungal effect after pasturization [218] or passage through 0.1 or 0.22 µM filtres [219,220], suggesting it may be small molecule based. The authors were careful to point out that their result was caused by the growth of anaerobic microbes [219] and that other compost teas made with different ingredients were rendered ineffective by filtration and/or sterilization [194,219]. Thus, these result may be due to induced systemic resistance rather than antibiotic production [218].

Composts go through several phases as they mature. There is an initial hot phase (thermophilic), after which the compost cools (mesophilic) and matures. It is generally thought the actinobacterial fraction of the bacterial community increases during the later stages of compost development [213] [214,221]. A recent study that used next generation sequencing of 16S amplicons to characterize the

bacterial community of three forms of compost, found that the actinobacterial community fraction decreased during the thermophilic phase and increased during the mesophilic phase [43]. Fatty acid–based community profiling has also shown that the actinobacterial community changes composition as compost develops [187]. Compost may be a good target for polyketide bioprospecting but its bacterial community is continually changing.

Several studies have identified antibiotics, including polyketides, produced by bacterial isolates from compost. A strain of *Pseudomonas aeruginosa* supressed a fungal pathogen *Fusarium oxysporum* infection of cucumber [222]. Suppression was due to 2,4-diacetylphloroglucinol (**67**, Figure 1.17), a previously known type III polyketide [223]. Pseudomonads are important antibiotic producers in both suppressive soils and compost [221,224].

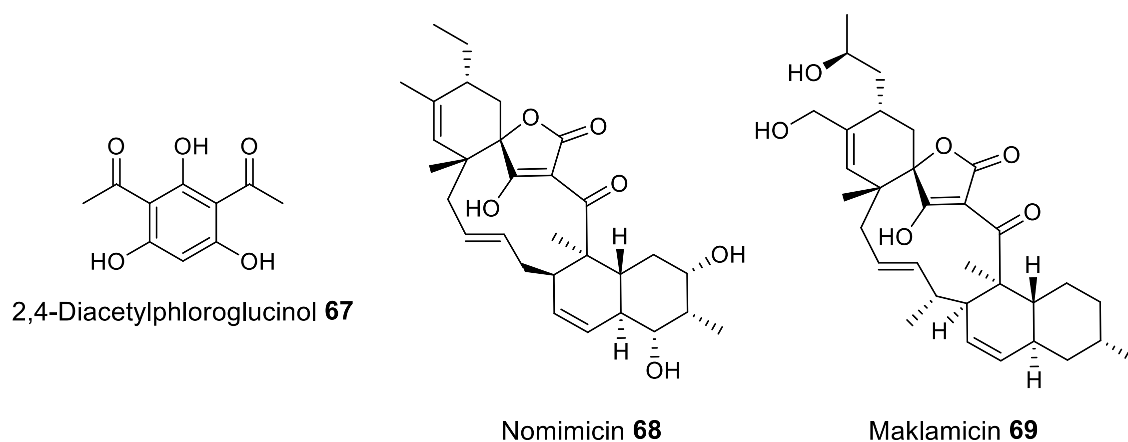


Figure 1.17. Natural products from disease suppressive soils.

As well as the spirotetronate type I polyketide, nomimicin (**68**) was discovered from a non-streptomycetal Actinomycete (*Actinomadura*) that was isolated from compost [225]. Nomimicin is similar to the already known maklamicin (**69**). The genomes of two Streptomyces isolated from compost are also being sequenced. Both contain a number of interesting polyketide synthase pathways suggesting these strains may produce a number of interesting new compounds [226,227].

Little polyketide bioprospecting has been done in disease suppressive environments compared to others listed here. Disease suppressive sediments include a broad range of environments, natural soils, intensively cultivated soils, and composts. They can thus have many different bacterial communities. Determining if a sediment is disease suppressive and whether this due to bacterial production of antibiotics requires either greenhouse testing and/or knowledge of plant disease history on a particular piece of land. Bioprospectors will need to work with agriculturalists to identify suppressive soils and sediments.

3.7. Other Unique Environments

3.7.1. Caves

There are several reasons for bioprospecting for polyketides in caves. Caves represent isolated and stable environments where bacteria can evolve independently of life on the surface [228], and also develop long term chemical “arms races” between each other [205]. For example, bat guano can build up over hundreds of years [229] providing a stable and nutrient rich environment in which insects and streptomycetes can flourish [230]. However, most caves are nutrient poor (oligotrophic) [231] and this lack of nutrients may encourage cooperation and interaction between microbes in caves rather than competition [232].

Usually the proteobacteria are the largest phylum in cave bacterial communities, but cave isolates are often actinobacterial [233]. This is not always the case though, some cave bacterial communities are highly actinobacterial [234,235] and the majority of these Actinomycetes are often Pseudonocardia. Tomczyk-Zak and Zielenkiewicz [236] reviewed the distribution of different bacterial phyla in caves. Communities dominated by Actinobacteria tend to be found on cave walls and in crystal structures such as stalactites and stalagmites. Between 2000 and 2009, 34 new species of Actinomycetes were isolated

from caves, including several new genera [233]. Caves may contain many novel actinomycetes that may also produce novel polyketides [237].

Several different processes can lead to cave formation. Most caves are found in karst landscapes that cover around a fifth of the worlds land area [231]. Karst landscapes develop when slightly acidic rainwater gradually dissolves carbonate rock such as limestone. Two novel aromatic type II polyketides have been characterized from Actinomycetes isolated from karst caves.

The Groto de Cervi is an organic matter-rich cave in Southern Italy with extensive bat guano deposits. Bacterial isolates from throughout the cave are often Actinomycetal. The cave also has 5000 year old Neolithic paintings in ochre and guano [238]. An isolate similar to *Streptomyces rochei* was isolated from a guano painting. This isolate was found to produce four type II polyketides (Cervimycins A–D, **70**–**73**, Figure 1.18) with activity against multidrug resistant *Staphylococcus aureus* and vancomycin resistant *Enterococcus faecalis*. All of the cervimycins contained a central four ring structure similar to the tetracyclines that is bis-glycosylated [239,240].

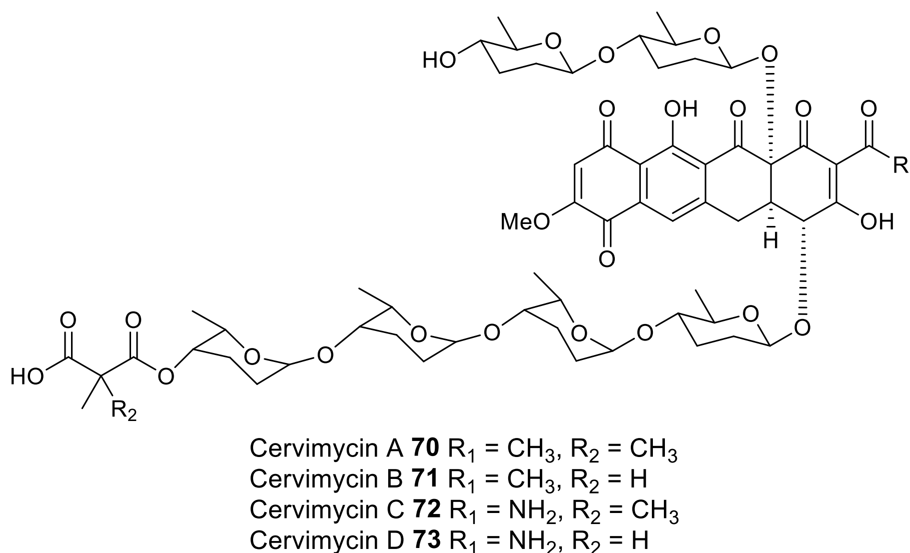


Figure 1.18. ***Cont.***

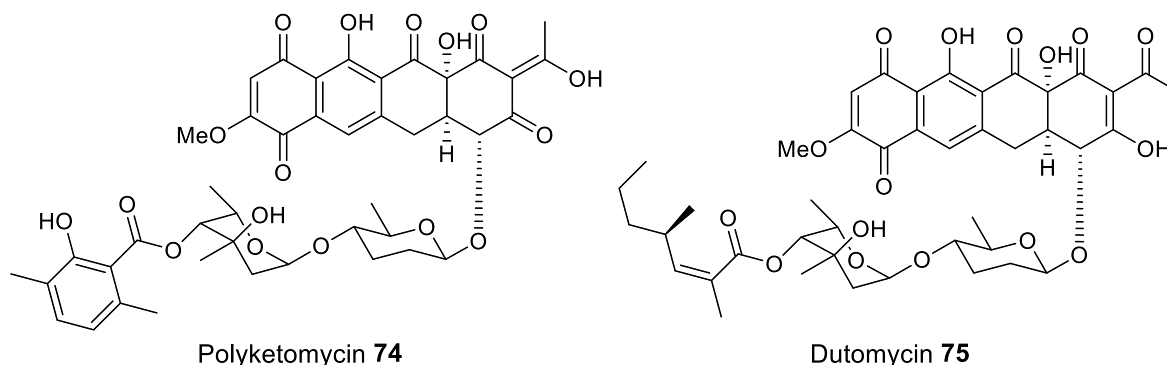
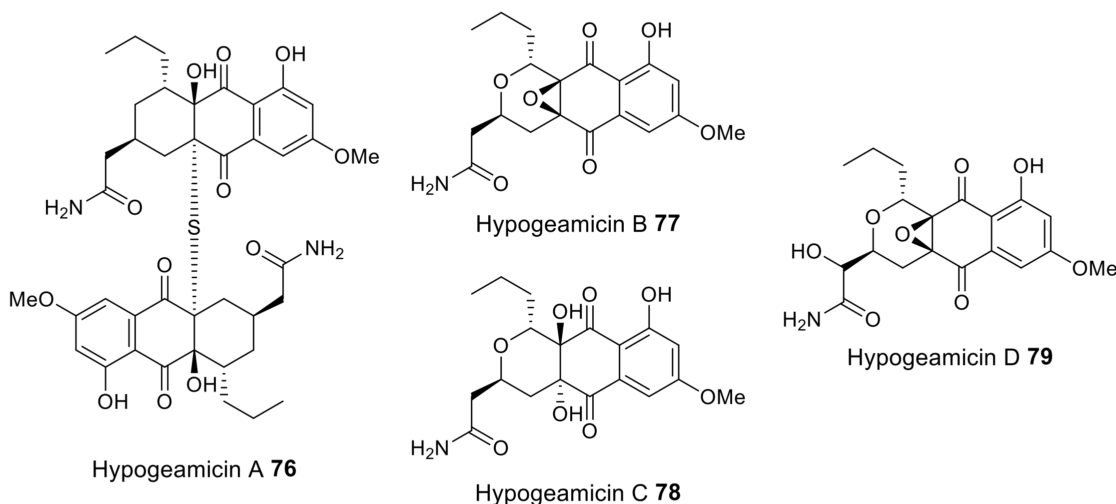


Figure 1.18. Cervimycins, isolated from cave derived Actinomycetes, are related to known tetracycline antibiotics 74 and 75.

Hardin's cave is a relatively small organic matter rich karst cave in Tennessee, also with a large bat population. An Actinomycete, *Nonomuraea specus*, was isolated from a piece of decomposed bark in the cave. *Nonomuraea specus* produces a sulphur-bridged dimeric pyronaphthoquinone, called hypogeamicin A (**76**, Figure 1.19). The non-dimeric precursors **77**–**79** were also isolated and were weakly toxic to *Bacillus subtillis*. The dimeric product **76** is toxic to TCT-1 colon cancer cell line. This is similar to the cytotoxicity of the related sulphur bridged dimer BE-52440 series (half maximal inhibitory concentration IC_{50} = 6.4–12.8 μM) but significantly lower than that of paclitaxel [241]



.

Figure 1.19 Type II polyketides isolated from an Actinomycetes isolate from Hardin's cave.

"Moon milk" is often found in karst caves (AKA mondmilch, Figure 1.20A). Moon milk is one of several kinds of cave deposits that is formed by the dissolution and reprecipitation of carbonates. Bacteria are thought to be involved in moon milk precipitation. Moon milk is made of calcium or magnesium carbonate and its texture can range from paste to powder [242]. There is a long history of moon milk use in medicine that dates from at least 1555, which led to it being exhaustively mined from some European caves [243]. A Russian expedition to a large karst cave formation in Siberia isolated *Streptomyces* and *Nocardia* from cave moonmilk that produced of antibacterial and antifungal compounds. One of these compounds chaxamycin B (**64**), was previously found in the Atacama desert (discussed earlier in this review) [244].

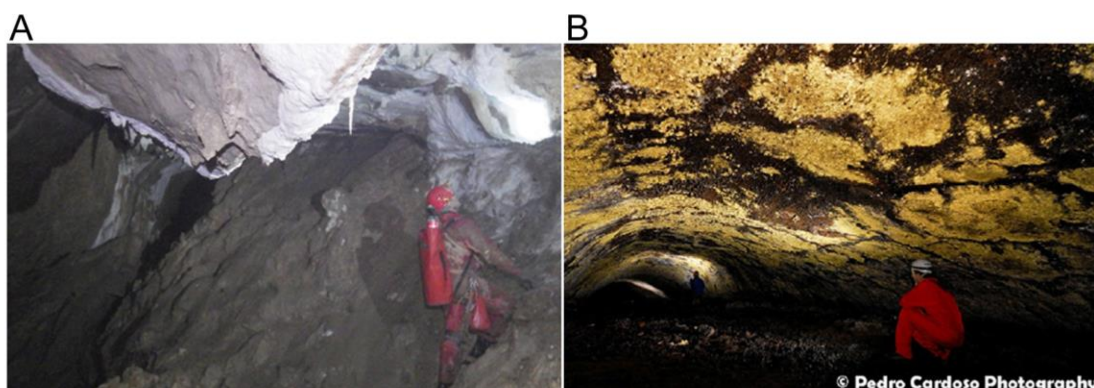


Figure 1.20. (A) An example of Moonmilk (above) in Goatherds Chasm in Switzerland. Photograph provided by Olivier Gallois of the Groupe Spéléologique Archéologique Mandeure; (B) Yellow Microbial mats from the volcanic cave Gruta de Terra Mole in the Azores. Photo courtesy of Pedro Cardoso. Figure as originally published in Riquelme, C.; Marshall Hathaway, J.J.; Enes Dapkevicius, M. de L.N.; Miller, A.Z.; Kooser, A.; Northup, D.E.; Jurado, V.; Fernandez, O.; Saiz-Jimenez, C.; Cheeptham, N. Actinobacterial Diversity in Volcanic Caves and Associated Geomicrobiological Interactions. *Front. Microbiol.*2015,*6*, 1–16. 01342/full.

A rarer and shallower form of caves are lava tubes, which form in volcanic areas when the surface of lava solidifies and the underlying molten lava continues to flow [231]. These caves are usually in basalt and can receive organic material from the surface through tree roots. Several researchers have bioprospected for antibiotic producing Actinomycetes in volcanic lava tubes in the Azores [245] and British Columbia, Canada [246] and found isolates with antibacterial activity. Coloured microbial mats

are often found on the surface of volcanic cave walls (Figure 1.20B). 16S amplicons from DNA of coloured microbial mats from volcanic caves in the Azores and Hawaii were sequenced and the data showed that Actinobacteria were one of the major phyla present [247]. A comparison of the actinobacterial fraction from volcanic caves from the Azores, Hawaii and New Mexico found that most (74%) of the 16S sequences were from five OTUs, with the two most common 16S OTUs (59%) being from *Pseudonorcardiaceae*. However, most OTUs (71%) were a single sequence. The authors interpret this to indicate that much of the actinobacterial richness in caves is derived from species unique to those particular environments [248], suggesting that sampling of many different caves is likely to yield new Actinobacteria and new polyketide natural products.

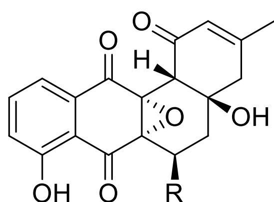
3.7.2. Extremophiles

The main argument for looking for bacterial polyketides in environments with extremes of pH, salinity and heat is that while soil has been sampled to the point of rediscovery, extreme environments are poorly studied. Differences in the environment will mean that there are differences in the secondary metabolites produces by the bacteria found in these environments [249].

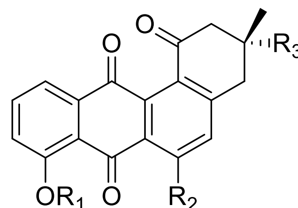
Extreme pH environments can often be found in abandoned mines, such as an abandoned coal mine in South Korea where sulphides are oxidized on contact with the atmosphere. Here the mine drainage is at pH 3. A Streptomycete isolate from this acid mine drainage produced eight type II aromatic polyketide angucyclinones (**80**–**87**, Figure 1.21) [250]. These compounds were tested against several bacteria and found to have antimicrobial activity against the Actinomycete *Micrococcus luteus* and Firmicutes *Enterococcus hirae* and MRSA.

Mines can also be highly alkaline. An Actinomycete from the genus *Nocardiopsis* was isolated from a tin mine tailings in southern China with a pH of 10 and was shown to produce the structurally unprecedented compound naphthospironone A (**88**). This highly unusual spiro[bicyclo[3.2.1]octene-

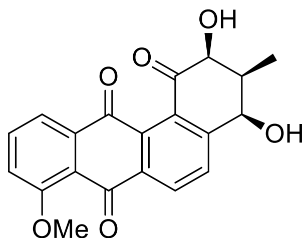
pyran]dione ring containing compound was moderately active against a small panel of cancer cell lines, as well as several Gram positive and negative bacteria [251].



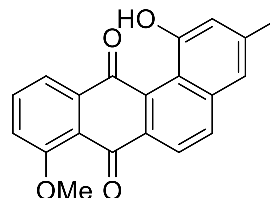
Angumycinone A **80** R = H
Angumycinone B **81** R = OH



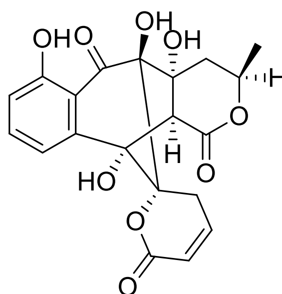
MM 47755 **82** R_1 = CH_3, R_2 = H R_3 = OH
(+)-Rubiginone B_2 **83** R_1 = CH_3, R_2 = H, R_3 = H
(+)-Ochromycinone **84** R_1 = H, R_2 = H, R_3 = H
(+)-Hatomarubigin A **85** R_1 = CH_3, R_2 = OH, R_3 = H



(+)-Rubiginone D_2 **86**



X-14881 E **87**



Naphthospirone A **88**

Figure 1.21. Polyketides discovered from cultured Actinomycete strains isolated from highly acidic mine drainage and highly basic mine tailings.

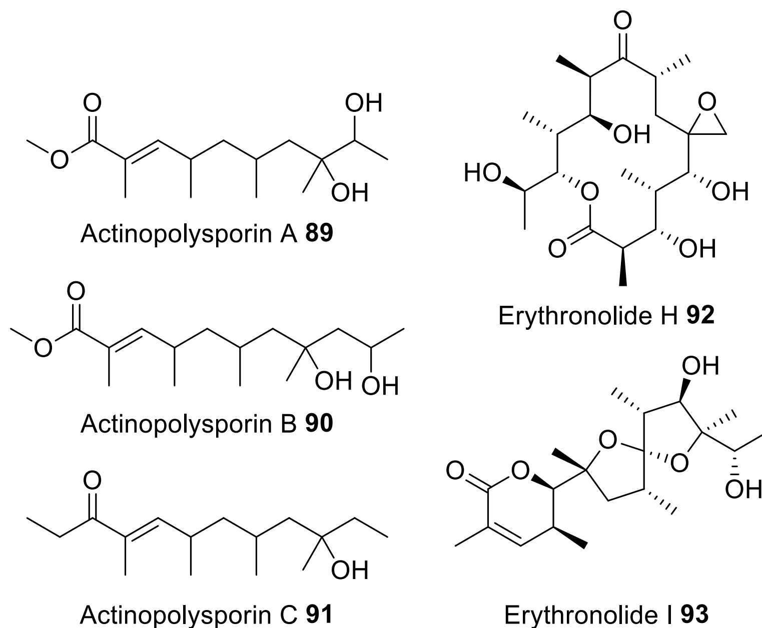


Figure 1.22. Polyketides identified from *Actinopolyspora erythraea*.

Jose and Jebakumar briefly review bioprospecting the Actinomycetes of hypersaline environments [249]. Two new species *Actinopolyspora alba* sp. nov. and *Actinopolyspora erythraea* sp. nov. were isolated from the Baicheng salt field in Xinjiang province, China [252]. *Actinopolyspora erythraea* produces several interesting polyketides. Actinopolysporins A–C (**90**–**91**, Figure 1.22) are novel polyketides; unfortunately with no detectable biological activity [253]. *Actinopolyspora erythraea* also produces two congeners of erythromycin, erythronolide H and I (**92** and **93**, respectively).

3.8. Potential New Environments for Bioprospecting

We have reviewed some well cited terrestrial environments that have produced new polyketides. However, this is by no means a definitive list. We will end this section by describing two environments that have been suggested for polyketide bioprospecting as they are enriched in Actinobacteria but have not yet yielded any new polyketides.

3.8.1. Cities

There are several arguments for bioprospecting in cities. Soils in cities may contain a broad range of bacteria as they are continually introduced by goods and travelers and are managed in a wide range of ways. Cities may also select for Actinobacteria, particularly in environments such as street dust or stone surfaces.

Two recent studies (discussed earlier in the section on biogeography) from well-known research groups, characterized Park soils of New York City. A next generation sequencing study of 16S amplicons by the Fierer laboratory found that most of the 16S amplicons present in a broad range of soils from many climates could be found in Central Park soil [122]. A very recent next generation sequencing study of type I PKS ketosynthase domain amplicons, by the Brady laboratory, from several New York parks found that while there were park soil specific communities, synthase sequences for a range of secondary metabolites that were originally isolated from across the world could be found in New York parks. The Brady laboratory suggests, that this is not a property of city soils but of most samples and that it might be more useful to screen a few samples deeply rather than shallowly screen many [123].

While cities are usually less biologically diverse than rural areas, this is not true for all organisms at every scale. Plant communities of suburban areas can be more diverse than rural or downtown areas [254,255]. Several studies have found that some eukaryotic organisms are more diverse in cities. This has been ascribed to continuous introduction by humans as seen with earthworms in Australia [124], a broader diversity of environments as seen with Clitellate worms in Stockholm [256], and a greater range of substrates as seen with fungi colonization of a stone surface in Vienna [125].

It is also possible that cities are also enriched in Actinobacteria. The Central Park New York 16S sequencing study, found that the microbial community in the soils of Central Park was enriched with Actinobacteria compared to a broad sampling of soils from across a wide range of eukaryotic biomes [122]. Our comparison of the bacterial communities of forest, cultivated soils and street dust found

using cloning and sequencing of 16S amplicons showed that street dusts were enriched in Actinobacteria [46]. When a subset of these samples was amplified with PKSI specific primers, PKSI pathways that were actinobacterial seemed to be selected for in street dust while non-actinobacterial pathways were selected for in soil [104]. Actinobacteria are known to be selected for by nitrogen, and urban environments receive more nitrogen from the atmosphere than the countryside [257], which may affect biological processes in urban soils [258]. Our laboratory is currently working on a Streptomycetal isolate from a bus stop. Its genome has 35 secondary metabolite pathways. [259].

3.8.2. Airborne Bacteria

Weber and Worth used 16S next generation sequencing to compare the bacterial communities of soils with the bacteria from the air above them (1.5–18.0 m). The sample site was in a small city in Idaho. More airborne bacteria were actinomycetal compared to the top 2 cm of soil (12% versus 38–69%) and while half of soil Actinomycetes were *Streptomycetes* most airborne Actinomycetes were from other groups (Figure 1.23). The authors suggest that the airborne bacterial community has many Actinomycetes as their spores are easily carried by air (spore forming Firmicutes made up most of the rest of the bacterial community) [260]. As *Streptomycetes* have already been extensively exploited for polyketide discovery [237], collecting bacteria from the air could be an simple mechanism for selecting novel non-streptomycetal Actinomycetes for bioprospecting.

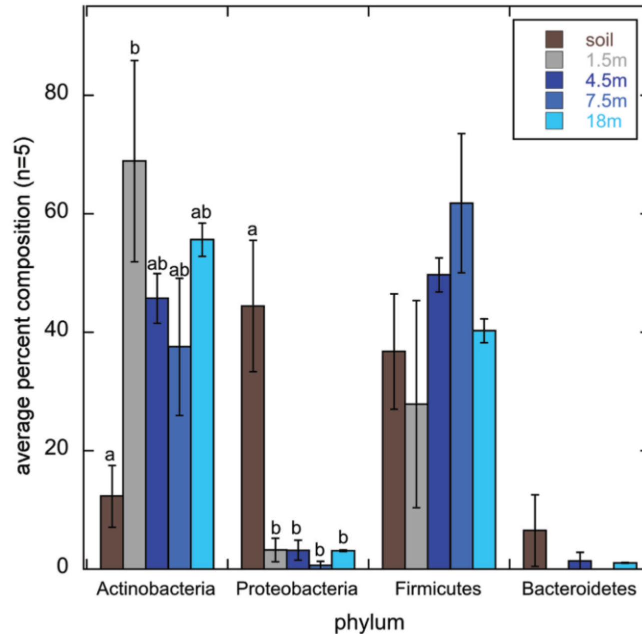


Figure 1.23 Average percent composition at the phylum-level classification of bacterial communities at the soil surface or at 1.5, 4.5, 7.5, or 18 m above the surface. Figure as originally published in Weber, C.F.; Werth, J.T. Is the lower atmosphere a readily accessible reservoir of culturable, antimicrobial compound-producing Actinomycetales? *Front. Microbiol.*2015,*6*,1–6.http://journal.frontiersin.org/article/10.3389/fmicb.2015.00802/full.

4. Conclusions

4.1. Taking Samples

Finding new polyketides in the environments is essential as antibiotic resistance increases. Sampling is the first step in this. For reasons of cost and convenience most polyketide bioprospecting will be on land [260]. Understanding the questions that Microbial ecology can and cannot answer about bacterial and polyketide distribution will help natural product chemists sample. The first section of this review covers these uncertainties which make it difficult to decide how to sample for polyketides in terrestrial environments.

The second section of this review lists several of the environments that have been prospected for new bacterial polyketides, either through isolation or metagenomics methods. The research groups that do this usually specialize in a single environment that they sample extensively, whether insect

associated bacteria, caves, or deserts soils. This may limit discovery. We have few comparisons of bacterial populations or polyketide synthases from very different environments such as soil versus compost or cave sediment, until we do it is not wise to invest heavily in a single environment. It is still unclear if most bacterial or polyketide synthase distribution is controlled by barriers to dispersion. If they are not, then there is little value to sampling a habitat more than once.

Several studies have found the similar polyketide synthases in very different environments, such as sea sponges and insects [176] or deserts and caves [244] so it is possible that everything is everywhere in an absolute sense. Comparisons of polyketides between different soils [59,93,121] and soils/street dust/vermicompost [104] have found that at least certain polyketides are more common in certain environments.

An efficient way to access many new polyketides quickly would be to sample (isolates and/or metagenomics clones) at a low level many environments with contrasting polyketide synthases. This will mean taking many samples and choosing a subset of them that have plentiful and contrasting polyketide synthases. The first step is to pick samples well. This will often mean talking to people who know more about the environment to be sampled than the bioprospectors themselves, such as plant protectionists, cavers, pedologists, compost producers, or street sweepers. For soils, at the local level the best sources of information are soil surveys and surveyors. A good overview of how soil properties vary at higher levels is http://soilgrids.org/, which predicts soil properties such as pH at a one square kilometre resolution. Several studies have mapped the distribution of bacterial communities in soil at the regional, national or continental scale [55,56,261]. The Earth microbiome project is comparing the bacterial communities of contrasting environments [37].

4.2. Characterising Samples

The second step is to characterize samples. Properties such as pH, texture, organic carbon and nitrogen can be measured easily compared to the bacterial community structure and so can be carried

out for all samples. After reducing the number of samples that have similar properties, the bacterial and/or polyketide synthase communities can be measured by molecular methods.

There are now a broad range of molecular methods that can be used to characterize bacterial communities. In applying these, it is important to use the same DNA extraction and analysis methods so that results from sample to sample are fully comparable. As sequencing is a rapidly developing technology, methods that were once widely used (e.g., 454 pyrosequencing) are now no longer offered [262]. Next generation sequencing can provide a much more detailed view of the bacterial community than earlier methods, through sequencing of tens of thousands of amplicons per sample, or shotgun cloning and sequencing of the metagenome. However, these methods often mean that all samples have to be pooled before they are sent for sequencing to keep costs low, which can slow the decision cycle time in sampling/characterizing/resampling interesting habitats.

Community fingerprinting may not be able to provide the same level of detail as next generation sequencing, but can quickly pick bacterial communities that are outliers and worth further sampling or more detailed characterization through next generation sequencing. In our experience restriction enzyme fingerprintints such as T-RFLP are difficult to interpret when eubacterial 16S primers are used as many cut sites are the same for different phylogenetic groups, so fingerprints are similar. Using primers that amplify the verrucomicrobia and actinobacteria [263] or the actinobacteria alone [264] (Figure 1.24) gave community fingerprints that differed from each other [46]. T-RFLP has been successfully used on PKSII sequences from soil [93]. The microbial communities of a subset of these samples can be determined to pick which samples will be used for growing isolates or making metagenomic libraries.

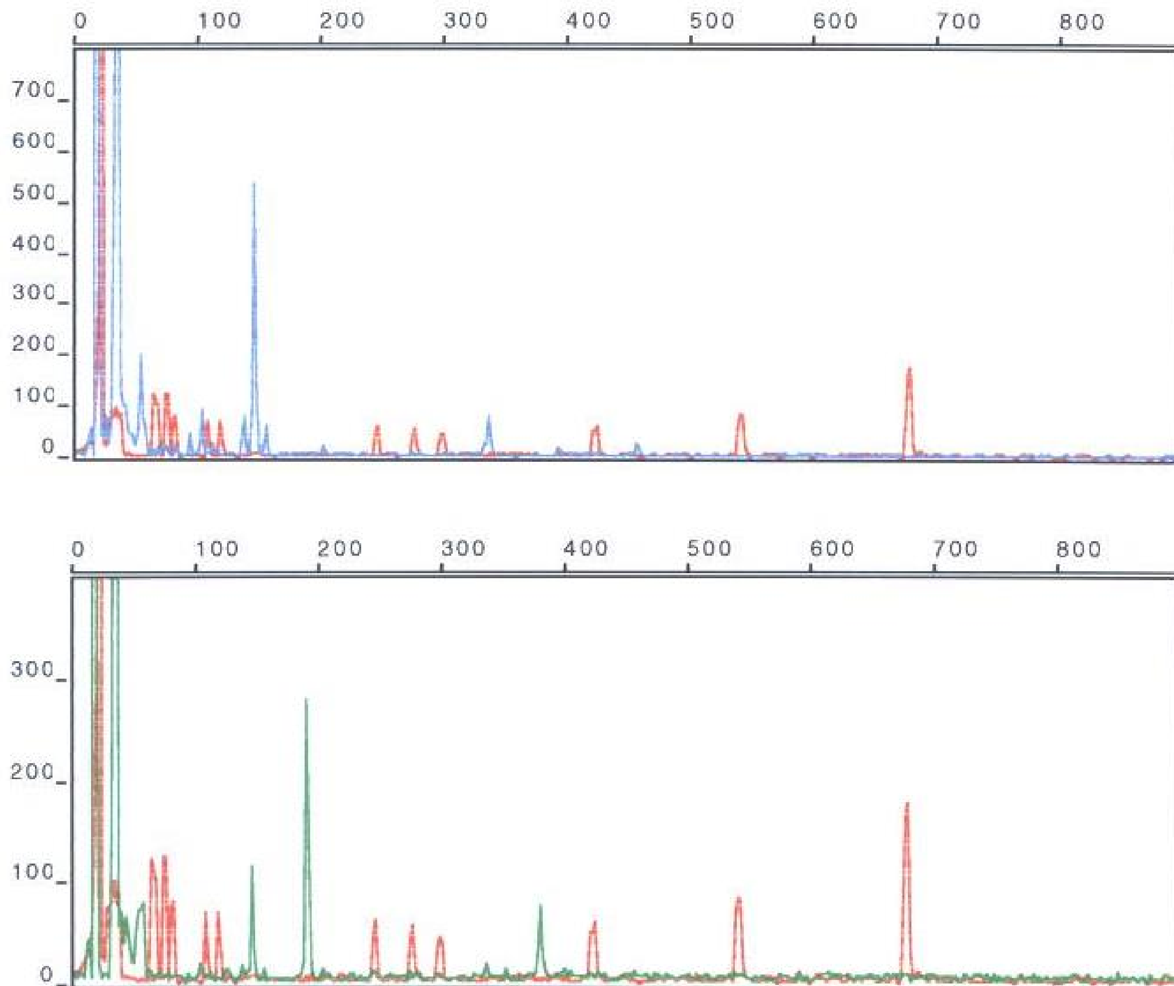


Figure 1.24. An example of actinobacterial 16S fingerprinting. Forward and reverse Terminal Restriction Fragment Length Polymorphism (T-RFLP) from a Sanger sequencer for a Russian taiga forest soil (Rusforest-Yenisei). PCR products of 16S actinobacterial specific primers were labelled with the dyes hexachloro-6-carboxyfluorescein (blue, forward) and carboxyfluorescein (green, reverse). Red peaks are the ROX 1000 size standards. Size is shown on the X axis, fluorescence on the Y axis. The size range 81–677 bp was used for clustering analysis of forward T-RFLP patterns from a range of samples (shown in Figure 1.4).

4.3. Non Actinomycetal Polyketide Producers

Characterizing environmental DNA with polyketide specific rather than 16S specific primers will give a better view and environments potential for polyketide discovery. 16S specific primers can show how diverse Actinobacteria are and if they are a large fraction of the soil bacterial community. Evidence from

type I polyketide specific primers suggests that most PKSI producers in many environments are not actinobacterial.

Whatever their relative importance in the environment, there is a strong argument for focusing on polyketides that appears to be from non-actinomycetal bacteria. Since the mid-1940s pharmaceutical companies have isolated millions of strains of Actinomycetes and tested their secondary metabolites for activity to discover new natural products [36]. Only a very small fraction of these strains has had their genomes or polyketide producing gene clusters sequenced. Thus even novel polyketide sequences from the environment which appears to be actinobacterial may be from an already known polyketide. Non-actinomycetal polyketides on the other hand have not been heavily screened for polyketide production. Since sequencing bacterial genomes has become cheaper, genome mining has found many polyketide pathways in non Actinobacteria. Non actinobacterial genomes may be a better source of genuinely novel polyketides.

Acknowledgments: We thank Josh D. Neufeld (University of Waterloo), David Yohalem (Mycological Innovations), Elisangela Gomes and Eliana Gertrudes de Macedo Lemos (Universidade Estadual Paulista "Julio de Mesquita Filho") and Lubomir Kovac (P.J. Šafárik University) for advice on cited literature.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Newman, D.J.; Cragg, G.M. Natural Products as Sources of New Drugs over the 30 Years. *J. Nat. Prod.* **2012**, *75*, 311–335.
2. Staunton, J.; Weissman, K.J. Polyketide biosynthesis: A millennium review. *Nat. Prod. Rep.* **2001**, *18*, 380–416.
3. Hertweck, C. The biosynthetic logic of polyketide diversity. *Angew. Chem. Int. Ed.* **2009**, *48*, 4688–4716.
4. Helfrich, E.J.N.; Piel, J. Biosynthesis of polyketides by trans-AT polyketide synthases. *Nat. Prod. Rep.* **2016**, *33*, 231–316.
5. Horsman, M.E.; Hari, T.P.A.; Boddy, C.N. Polyketide synthase and non-ribosomal peptide synthetase thioesterase selectivity: Logic gate or a victim of fate? *Nat. Prod. Rep.* **2016**, *33*, 183–202.
6. Hertweck, C.; Luzhetskyy, A.; Rebets, Y.; Bechthold, A. Type II polyketide synthases: Gaining a deeper insight into enzymatic teamwork. *Nat. Prod. Rep.* **2007**, *24*, 162–190.
7. Walsh, C.T. Polyketide and nonribosomal peptide antibiotics: Modularity and versatility. *Science* **2004**, *303*, 1805–1810.
8. Finking, R.; Marahiel, M.A. Biosynthesis of nonribosomal peptides. *Annu. Rev. Microbiol.* **2004**, *58*, 453–488.
9. Felnagle, E.A.; Jackson, E.E.; Chan, Y.A.; Podevels, M.A.; Berti, A.D.; Mcmahon, M.D.; Thomas, M.G.; Podevels, A.M. Nonribosomal Peptide Synthetases Involved in the Production of Medically Relevant Natural Products. *Mol. Pharm.* **2008**, *5*, 191–211.
10. Baltz, R.H. Marcel Faber Roundtable: Is our antibiotic pipeline unproductive because of starvation, constipation or lack of inspiration? *J. Ind. Microbiol. Biotechnol.* **2006**, *33*, 507–513.

11. Bode, H.B.; Bethe, B.; Höfs, R.; Zeeck, A. Big effects from small changes: Possible ways to explore nature's chemical diversity. *ChemBioChem* **2002**, *3*, 619–627.
12. Okada, B.K.; Wu, Y.; Mao, D.; Bushin, L.B.; Seyedsayamdost, M.R. Mapping the Trimethoprim-Induced Secondary Metabolome of Burkholderia thailandensis. *ACS Chem. Biol.* **2016**, *11*, 2124–2130.
13. Li, S.; Li, Y.; Lu, C.; Zhang, J.; Zhu, J.; Wang, H.; Shen, Y. Activating a Cryptic Ansamycin Biosynthetic Gene Cluster to Produce Three New Naphthalenic Octaketide Ansamycins with n-Pentyl and *n*-Butyl Side Chains. *Org. Lett.* **2015**, *17*, 3706–3709.
14. Gomez-Escribano, J.P.; Song, L.; Fox, D.J.; Yeo, V.; Bibb, M.J.; Challis, G.L. Structure and biosynthesis of the unusual polyketide alkaloid coelimycin P1, a metabolic product of the cpk gene cluster of Streptomyces coelicolor M145. *Chem. Sci.* **2012**, *3*, 2716.
15. Stevens, D.C.; Hari, T.P.A.; Boddy, C.N. The role of transcription in heterologous expression of polyketides in bacterial hosts. *Nat. Prod. Rep.* **2013**, *30*, 1391–1411.
16. Cole Stevens, D.; Henry, M.R.; Murphy, K.A.; Boddy, C.N. Heterologous expression of the oxytetracycline biosynthetic pathway in *Myxococcus xanthas*. *Appl. Environ. Microbiol.* **2010**, *76*, 2681–2683.
17. Ling, L.L.; Schneider, T.; Peoples, A.J.; Spoering, A.L.; Engels, I.; Conlon, B.P.; Mueller, A.; Hughes, D.E.; Epstein, S.; Jones, M.; et al. A new antibiotic kills pathogens without detectable resistance. *Nature* **2015**, *517*, 455–459.
18. Wang, G.Y.; Graziani, E.; Waters, B.; Pan, W.; Li, X.; McDermott, J.; Meurer, G.; Saxena, G.; Andersen, R.J.; Davies, J. Novel natural products from soil DNA libraries in a streptomycete host. *Org. Lett.* **2000**, *2*, 2401–2404.
19. Kang, H.S.; Brady, S.F. Arixanthomycins A–C: Phylogeny-guided discovery of biologically active eDNA-derived pentangular polyphenols. *ACS Chem. Biol.* **2014**, *9*, 1267–1272.
20. Jensen, P.R. Natural Products and the Gene Cluster Revolution. *Trends Microbiol.* **2016**, *24*, 968–977.
21. Ikeda, H. Natural products discovery from micro-organisms in the post-genome era. *Biosci. Biotechnol. Biochem.* **2017**, *81*, 13–22.
22. Lewis, K. New Approaches to Antimicrobial Discovery. *Biochem. Pharmacol.* **2016**, doi: 10.1016/j.bcp.2016.11.002.
23. Pawar, S.V.; Ho, J.C.H.; Yadav, G.D.; Yadav, V.G. The Impending Renaissance in Discovery & Development of Natural Products. *Curr. Top. Med. Chem.* **2017**, *17*, 251–267.
24. Zhang, M.M.; Qiao, Y.; Ang, E.L.; Zhao, H. Using natural products for drug discovery: The impact of the genomics era. *Expert Opin. Drug Discov.* **2017**, *12*, 1–13.
25. Ren, H.; Wang, B.; Zhao, H. Breaking the silence: New strategies for discovering novel natural products. *Curr. Opin. Biotechnol.* **2017**, *48*, 21–27.
26. Methodik, Z.; Lhlung, D.B. Zur Methodik der Bakterienz~Lhlung. *Z. Hyg*. *Infect.* **1898**, *29*, 75–93.
27. Torsvik, V.; Goksøyr, J.; Daae, F.L.; Torsvik, V.; Goksyr, J.; Daae, F.L. High diversity in DNA of soil bacteria. High Diversity in DNA of Soil Bacteria. *Appl. Environ. Microbiol.* **1990**, *56*, 782–787.
28. Ueda, T. Suga, Y.; Matsuguchi, M. Molecular phylogenetic analysis of a soil microbial community in a soybean field. *Eur. J. Soil Sci.* **1995**, *46*, 415–421.
29. Muyzer, G.; De Waal, E.C.; Uitterlinden, A.G. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of polymerase chain reaction-amplified genes coding for 16S rRNA. *Appl. Environ. Microbiol.* **1993**, *59*, 695–700.
30. Liu, W.T.; Marsh, T.L.; Cheng, H.; Forney, L.J. Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Appl. Environ. Microbiol.* **1997**, *63*, 4516–4522.
31. Roesch, L.; Fulthorpe, R.; Riva, A.; Casella, G.; Hadwin, A.; Kent, A.; Daroub, S.; Camargo, F.; Farmerie, W.; Triplett, E. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J.* **2007**, *1*, 283–290.
32. Meyer, F.; Paarmann, D.; D'Souza, M.; Olson, R.; Glass, E.; Kubal, M.; Paczian, T.; Rodriguez, A.; Stevens, R.; Wilke, A.; et al. The metagenomics RAST server—A public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinform.* **2008**, *9*, 386.
33. Tunlid, A.; Hoitink, H.A.; Low, C.; White, D.C. Characterization of bacteria that suppress rhizoctonia damping-off in bark compost media by analysis of Fatty Acid biomarkers. *Appl. Environ. Microbiol.* **1989**, *55*, 1368–1374.
34. Frostegård, Å.; Tunlid, A.; Bååth, E. Use and misuse of PLFA measurements in soils. *Soil Biol. Biochem.* **2011**, *43*, 1621–1625.
35. Kaur, A.; Chaudhary, A.; Kaur, A.; Choudhary, R.; Kaushik, R. Phospholipid fatty acid—A bioindicator of environment monitoring and assessment in soil ecosystem. *Curr. Sci.* **2005**, *89*, 1103–1112.
36. Bérdy, J. Bioactive microbial metabolites. *J. Antibiot.* **2005**, *58*, 1–26.
37. Gilbert, J.A.; Jansson, J.K.; Knight, R. The Earth Microbiome project: Successes and aspirations. *BMC Biol.* **2014**, *12*, 69.
38. Chu, H.; Fierer, N.; Lauber, C.L.; Caporaso, J.G.; Knight, R.; Grogan, P. Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes. *Environ. Microbiol.* **2010**, *12*, 2998–3006.
39. Parsley, L.C.; Linneman, J.; Goode, A.M.; Becklund, K.; George, I.; Goodman, R.M.; Lopanik, N.B.; Liles, M.R. Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. *FEMS Microbiol. Ecol.* **2011**, *78*, 176–187.

40. Quaiser, A.; Ochsenreiter, T.; Lanz, C.; Schuster, S.C.; Treusch, A.H.; Eck, J.; Schleper, C. Acidobacteria form a coherent but highly diverse group within the bacterial domain: Evidence from environmental genomics. *Mol. Microbiol.* **2003**, *50*, 563–575.
41. Tripathi, B.M.; Kim, M.; Singh, D.; Lee-Cruz, L.; Lai-Hoe, A.; Ainuddin, A.N.; Go, R.; Rahim, R.A.; Husni, M.H.A.; Chun, J.; et al. Tropical Soil Bacterial Communities in Malaysia: PH Dominates in the Equatorial Tropics Too. *Microb. Ecol.* **2012**, *64*, 474–484.
42. Mathew, R.P.; Feng, Y.; Githinji, L.; Ankumah, R.; Balkcom, K.S. Impact of No-tillage and conventional tillage systems on soil microbial communities. *Appl. Environ. Soil Sci.* **2012**, *2012*, 548620.
43. Neher, D.A.; Weicht, T.R.; Bates, S.T.; Leff, J.W.; Fierer, N. Changes in bacterial and fungal communities across compost recipes, preparation methods, and composting times. *PLoS ONE* **2013**, *8*, e79512.
44. Delmont, T.O.; Robe, P.; Cecillon, S.; Clark, I.M.; Constancias, F.; Simonet, P.; Hirsch, P.R.; Vogel, T.M. Accessing the soil metagenome for studies of microbial diversity. *Appl. Environ. Microbiol.* **2011**, *77*, 1315–1324.
45. Hong, S.; Bunge, J.; Leslin, C.; Jeon, S.; Epstein, S.S. Polymerase chain reaction primers miss half of rRNA microbial diversity. *ISME J.* **2009**, *3*, 1365–1373.
46. Hill, P.; Krištůfek, V.; Dijkhuizen, L.; Boddy, C.; Kroetsch, D.; Van Elsas, J.D. Land use intensity controls actinobacterial community structure. *Microb. Ecol.* **2011**, *61*, 286–302.
47. Pan, Y.; Bodrossy, L.; Frenzel, P.; Hestnes, A.G.; Krause, S.; Lüke, C.; Meima-Franke, M.; Siljanen, H.; Svenning, M.M.; Bodelier, P.L.E. Impacts of inter- and intralaboratory variations on the reproducibility of microbial community analyses. *Appl. Environ. Microbiol.* **2010**, *76*, 7451–7458.
48. Cary, S.C.; Fierer, N. The importance of sample archiving in microbial ecology. *Nat. Rev. Microbiol.* **2014**, *12*, 789–790.
49. Vogel, T.M.; Simonet, P.; Jansson, J.K.; Hirsch, P.R.; Tiedje, J.M.; Elsas, V.; Dirk, J.; Bailey, M.J.; Nalin, R.; Philippot, L. TerraGenome: A consortium for the sequencing of a soil metagenome. *Nat. Rev. Microbiol.* **2009**, *7*, 2009.
50. China Soil Microbiome Initiative Launched. Available online: http://english.issas.cas.cn/ns/es/201407/t20140702_123686.html .
51. Pylro, V.S.; Roesch, L.F.W.; Ortega, J.M.; do Amaral, A.M.; Tótola, M.R.; Hirsch, P.R.; Rosado, A.S.; Góes-Neto, A.; da Costa da Silva, A.L.; Rosa, C.A.; et al. Brazilian Microbiome Project: Revealing the Unexplored Microbial Diversity-Challenges and Prospects. *Microb. Ecol.* **2014**, *67*, 237–241.
52. Blaser, M.J.; Cardon, Z.G.; Cho, M.K.; Dangl, J.L.; Donohue, T.J.; Green, J.L.; Knight, R.; Editor, S.; Maxon, M.E.; Northen, T.R.; et al. Toward a Predictive Understanding of Earth's Microbiomes to Address 21st Century Challenges. *MBio* **2016**, *7*, doi:10.1128/mBio.00714-16.
53. Sharpton, T.J. An introduction to the analysis of shotgun metagenomic data. *Front. Plant Sci.* **2014**, *5*, 209.
54. Singh, B.K.; Campbell, C.D.; Sorenson, S.J.; Zhou, J. Soil genomics. *Nat. Rev. Microbiol.* **2009**, *7*, 756.
55. Griffiths, R.I.; Thomson, B.C.; James, P.; Bell, T.; Bailey, M.; Whiteley, A.S. The bacterial biogeography of British soils. *Environ. Microbiol.* **2011**, *13*, 1642–1654.
56. Dequiedt, S.; Thioulouse, J.; Jolivet, C.; Saby, N.P.A.; Lelievre, M.; Maron, P.A.; Martin, M.P.; Prévost-Bouré, N.C.; Toutain, B.; Arrouays, D.; et al. Biogeographical patterns of soil bacterial communities. *Environ. Microbiol. Rep.* **2009**, *1*, 251–255.
57. Methé, B.A.; Nelson, K.E.; Pop, M.; Creasy, H.H.; Giglio, M.G.; Huttenhower, C.; Gevers, D.; Petrosino, J.F.; Abubucker, S.; Badger, J.H.; et al. A framework for human microbiome research. *Nature* **2012**, *486*, 215–221.
58. McDonald, D.; Birmingham, A.; Knight, R. Context and the human microbiome. *Microbiome* **2015**, *3*, 52.
59. Reddy, B.; Milshteyn, A.; Charlop-Powers, Z.; Brady, S. ESNaPD: A Versatile, Web-Based Bioinformatics Platform for Surveying and Mining Natural Product Biosynthetic Diversity from Metagenomes. *Chem. Biol.* **2014**, *21*, 1023–1033.
60. Willig, M.; Kaufman, D.; Stevens, R. Latitudinal gradients of biodiversity: Pattern, process, scale, and synthesis. *Annu. Rev. Ecol. Evol. Syst.* **2003**, *34*, 273–309.
61. Fujisaka, S.; Escobar, G.; Veneklaas, E. Plant community diversity relative to human land uses in an Amazon forest colony. *Biodivers. Conserv.* **1998**, *7*, 41–57.
62. Tedersoo, L.; Bahram, M.; Polme, S.; Koljalg, U.; Yorou, S.; Wardle, D.A.; Lindahl, B.D. Disentangling global soil fungal diversity. *Science* **2014**, *346*, 1052–1053.
63. Gelsomino, A.; Keijzer-Wolters, A.C.; Cacco, G.; van Elsas, J.D. Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gradient gel electrophoresis. *J. Microbiol. Methods* **1999**, *38*, 1–15.
64. Girvan, M.S.; Bullimore, J.; Pretty, J.N.; Mark, A.; Ball, A.S.; Osborn, A.M. Soil Type Is the Primary Determinant of the Composition of the Total and Active Bacterial Communities in Arable Soils Soil Type Is the Primary Determinant of the Composition of the Total and Active Bacterial Communities in Arable Soils. *Appl. Environ. Microbiol.* **2003**, *69*, 1800–1809.
65. Bossio, D.A.; Girvan, M.S.; Verchot, L.; Bullimore, J.; Borelli, T.; Albrecht, A.; Scow, K.M.; Ball, A.S.; Pretty, J.N.; Osborn, A.M. Soil microbial community response to land use change in an agricultural landscape of western Kenya. *Microb. Ecol.* **2005**, *49*, 50–62.
66. Fierer, N.; Jackson, R.B. The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 626–31.

67. Mendes, L.W.; de Lima Brossi, M.J.; Kuramae, E.E.; Tsai, S.M. Land-use system shapes soil bacterial communities in Southeastern Amazon region. *Appl. Soil Ecol.* **2015**, *95*, 151–160.
68. Kuramae, E.E.; Yergeau, E.; Wong, L.C.; Pijl, A.S.; Van Veen, J.A.; Kowalchuk, G.A. Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS Microbiol. Ecol.* **2012**, *79*, 12–24.
69. Ulrich, A.; Becker, R. Soil parent material is a key determinant of the bacterial community structure in arable soils. *FEMS Microbiol. Ecol.* **2006**, *56*, 430–443.
70. Fierer, N.; Lennon, J.T. The generation and maintenance of diversity in microbial communities. *Am. J. Bot.* **2011**, *98*, 439–448.
71. Lauber, C.L.; Hamady, M.; Knight, R.; Fierer, N. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl. Environ. Microbiol.* **2009**, *75*, 5111–5120.
72. Fierer, N.; Leff, J.W.; Adams, B.J.; Nielsen, U.N.; Bates, S.T.; Lauber, C.L.; Owens, S.; Gilbert, J.A.; Wall, D.H.; Caporaso, J.G. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 21390–21395.
73. Rousk, J.; Bååth, E.; Brookes, P.C.; Lauber, C.L.; Lozupone, C.; Caporaso, J.G.; Knight, R.; Fierer, N. Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* **2010**, *4*, 1340–1351.
74. Zhalnina, K.; Dias, R.; de Quadros, P.D.; Davis-Richardson, A.; Camargo, F.A.O.; Clark, I.M.; McGrath, S.P.; Hirsch, P.R.; Triplett, E.W. Soil pH Determines Microbial Diversity and Composition in the Park Grass Experiment. *Microb. Ecol.* **2014**, *69*, 395–406.
75. Jenkins, S.N.; Waite, I.S.; Blackburn, A.; Husband, R.; Rushton, S.P.; Manning, D.C.; O'Donnell, A.G. Actinobacterial community dynamics in long term managed grasslands. *Antonie Van Leeuwenhoek* **2009**, *95*, 319–334.
76. Bartram, A.K.; Jiang, X.; Lynch, M.D.J.; Masella, A.P.; Nicol, G.W.; Dushoff, J.; Neufeld, J.D. Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiol. Ecol.* **2014**, *87*, 403–415.
77. Lauber, C.L.; Strickland, M.S.; Bradford, M.A.; Fierer, N. The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil Biol. Biochem.* **2008**, *40*, 2407–2415.
78. Navarrete, A.A.; Tsai, S.M.; Mendes, L.W.; Faust, K.; De Hollander, M.; Cassman, N.A.; Raes, J.; Van Veen, J.A.; Kuramae, E.E. Soil microbiome responses to the short-term effects of Amazonian deforestation. *Mol. Ecol.* **2015**, *24*, 2433–2448.
79. Montecchia, M.S.; Tosi, M.; Soria, M.A.; Vogrig, J.A.; Sydorenko, O.; Correa, O.S. Pyrosequencing reveals changes in soil bacterial communities after conversion of Yungas forests to agriculture. *PLoS ONE* **2015**, *10*, e0119426.
80. Waldrop, M.; Balser, T.; Firestone, M. Linking microbial community composition to function in a tropical soil. *Soil Biol. Biochem.* **2000**, *32*, 1837–1846.
81. Shange, R.S.; Ankumah, R.O.; Ibekwe, A.M.; Zabawa, R.; Dowd, S.E. Distinct soil bacterial communities revealed under a diversely managed agroecosystem. *PLoS ONE* **2012**, *7*, e40338.
82. Lee-Cruz, L.; Edwards, D.P.; Tripathi, B.M.; Adams, J.M. Impact of logging and forest conversion to oil palm plantations on soil bacterial communities in borneo. *Appl. Environ. Microbiol.* **2013**, *79*, 7290–7297.
83. Sul, W.J.; Asuming-Brempong, S.; Wang, Q.; Tourlousse, D.M.; Penton, C.R.; Deng, Y.; Rodrigues, J.L.M.; Adiku, S.G.K.; Jones, J.W.; Zhou, J.; et al. Tropical agricultural land management influences on soil microbial communities through its effect on soil organic carbon. *Soil Biol. Biochem.* **2013**, *65*, 33–38.
84. Fierer, N.; Bradford, M.A.; Jackson, R.B. Toward an Ecological Classification of Soil Bacteria. *Ecology* **2007**, *88*, 1354–1364.
85. Ramirez, K.S.; Craine, J.M.; Fierer, N. Consistent effects of nitrogen amendments on soil microbial communities and processes across biomes. *Glob. Chang. Biol.* **2012**, *18*, 1918–1927.
86. Ramirez, K.S.; Lauber, C.L.; Knight, R.; Bradford, M.A.; Fierer, N. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology* **2010**, *91*, 3463–3470.
87. Fierer, N.; Lauber, C.L.; Ramirez, K.S.; Zaneveld, J.; Bradford, M.A.; Knight, R. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J.* **2012**, *6*, 1007–1017.
88. Pan, Y.; Cassman, N.; de Hollander, M.; Mendes, L.W.; Korevaar, H.; Geerts, R.H.E.M.; van Veen, J.A.; Kuramae, E.E. Impact of long-term N, P, K, and NPK fertilization on the composition and potential functions of the bacterial community in grassland soil. *FEMS Microbiol. Ecol.* **2014**, *90*, 195–205.
89. Zhao, C.; Fu, S.; Mathew, R.P. Soil microbial community structure and activity in a 100-year-old fertilization and crop rotation experiment. *J. Plant Ecol.* **2015**, 1–10.
90. Warnecke, F.; Amann, R.; Pernthaler, J. Actinobacterial 16S rRNA genes from freshwater habitats cluster in four distinct lineages. *Environ. Microbiol.* **2004**, *6*, 242–253.
91. Humbert, J.F.; Dorigo, U.; Cecchi, P.; Le Berre, B.; Debroas, D.; Bouvy, M. Comparison of the structure and composition of bacterial communities from temperate and tropical freshwater ecosystems. *Environ. Microbiol.* **2009**, *11*, 2339–2350.
92. Wawrik, B.; Kerkhof, L.; Zylstra, G.J.; Jerome, J.; Kukor, J.J. Identification of Unique Type II Polyketide Synthase Genes in Soil Identification of Unique Type II Polyketide Synthase Genes in Soil. *Appl. Environ. Microbiol.* **2005**, *71*, 2232–2238.

93. Wawrik, B.; Kutliev, B.; Abdivasievna, U.A.; Kukor, J.J.; Zylstra, G.J.; Kerkhof, L. Biogeography of actinomycete communities and type II polyketide synthase genes in soils collected in New Jersey and Central Asia. *Appl. Environ. Microbiol.* **2007**, *73*, 2982–2989.
94. Pang, M.-F.; Tan, G.-Y.A.; Abdullah, N.; Lee, C.-W.; Ng, C.-C. Phylogenetic analysis of type I and type II polyketide synthase from tropical forest soil. *Biotechnology* **2008**, *7*, 660–668.
95. Morlon, H.; O'Connor, T.K.; Bryant, J.A.; Charkoudian, L.K.; Docherty, K.M.; Jones, E.; Kembel, S.W.; Green, J.L.; Bohannan, B.J.M. The biogeography of putative microbial antibiotic production. *PLoS ONE* **2015**, *10*, e0130659.
96. Amos, G.C.A.; Borsetto, C.; Laskaris, P.; Krsek, M.; Berry, A.E.; Newsham, K.K.; Calvo-Bado, L.; Pearce, D.A.; Vallin, C.; Wellington, E.M.H. Designing and implementing an assay for the detection of rare and divergent NRPS and PKS clones in European, Antarctic and Cuban soils. *PLoS ONE* **2015**, *10*, e0138327.
97. Zhao, J.; Yang, N.; Zeng, R. Phylogenetic analysis of type I polyketide synthase and nonribosomal peptide synthetase genes in Antarctic sediment. *Extremophiles* **2008**, *12*, 97–105.
98. Zhao, J.; Yang, N.; Chen, X.; Jiang, Q.; Zeng, R. Phylogenetic diversity of Type I polyketide synthase genes from sediments of Ardley Island in Antarctica. *Acta Oceanol. Sin.* **2011**, *30*, 104–111.
99. Zhao, B.; Gao, Z.; Shao, Y.; Yan, J.; Hu, Y.; Yu, J.; Liu, Q.; Chen, F. Diversity analysis of type I ketosynthase in rhizosphere soil of cucumber. *J. Basic Microbiol.* **2012**, *52*, 224–231.
100. Luo, K.; Du, G.-P.; Zhao, Z.-X.; Xie, B.; Li, D.-J. Phylogenetic analysis of type I polyketide synthase and non-ribosomal peptide synthase genes from Mila Mountain in Tibet Plateau. *J. Hunan Agric. Univ.* **2010**, *36*, 506–511.
101. Ginolhac, A.; Jarrin, C.; Gillet, B.; Robe, P.; Pujic, P.; Tuphile, K.; Bertrand, H.; Vogel, T.M.; Perrière, G.; Simonet, P.; et al. Phylogenetic analysis of polyketide synthase I domains from soil metagenomic libraries allows selection of promising clones. *Appl. Environ. Microbiol.* **2004**, *70*, 5522–5527.
102. Gomes, E.S.; Schuch, V.; De Macedo Lemos, E.G. Biotechnology of polyketides: New breath of life for the novel antibiotic genetic pathways discovery through metagenomics. *Braz. J. Microbiol.* **2013**, *44*, 1007–1034.
103. Yadav, G.; Gokhale, R.S.; Mohanty, D. SEARCHPKS: A program for detection and analysis of polyketide synthase domains. *Nucleic Acids Res.* **2003**, *31*, 3654–3658.
104. Hill, P.; Piel, J.; Aris-Brosou, S.; Krištůfek, V.; Boddy, C.N.; Dijkhuizen, L. Habitat-specific type I polyketide synthases in soils and street sediments. *J. Ind. Microbiol. Biotechnol.* **2014**, *41*, 75–85.
105. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Brady, S.F. Chemical-biogeographic survey of secondary metabolism in soil. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 3757–3762.
106. Wallace, A.R. *The Geographical Distribution of Animals. With a Study of the Relations of Living and Extinct Faunas as Elucidating the Past Changes of the Earth's Surface*; Harper & Brothers: New York, NY, USA, 1876.
107. Finlay, B.J. Global dispersal of free-living microbial eukaryote species. *Science* **2002**, *296*, 1061–1063.
108. De Wit, R.; Bouvier, T. "Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say? *Environ. Microbiol.* **2006**, *8*, 755–758.
109. O'Malley, M.A. The nineteenth century roots of "everything is everywhere". *Nat. Rev. Microbiol.* **2007**, *5*, 647–651.
110. Bland, J.F.; Esteban, G.F. Ubiquitous Dispersal of Free-Living Microorganisms. In *Microbial Diversitiy and Bioprospecting*; Bull, A.T., Ed.; ASM Press: Washington, DC, USA, 2004; pp. 216–224.
111. Martiny, J.B.H.; Bohannan, B.J.M.; Brown, J.H.; Colwell, R.K.; Fuhrman, J.A.; Green, J.L.; Horner-Devine, M.C.; Kane, M.; Krumins, J.A.; Kuske, C.R.; et al. Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* **2006**, *4*, 102–112.
112. Lozupone, C.; Knight, R. UniFrac: A New Phylogenetic Method for Comparing Microbial Communities UniFrac: A New Phylogenetic Method for Comparing Microbial Communities. *Appl. Environ. Microbiol.* **2005**, *71*, 8228–8235.
113. Whitaker, R.J.; Grogan, D.W.; Taylor, J.W. Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science* **2003**, *301*, 2002–2004.
114. Papke, R.T.; Ramsing, N.B.; Bateson, M.M.; Ward, D.M. Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **2003**, *5*, 650–659.
115. Valverde, A.; Tuffin, M.; Cowan, D.A. Biogeography of bacterial communities in hot springs: A focus on the actinobacteria. *Extremophiles* **2012**, *16*, 669–679.
116. Barton, L.L.; Mandl, M.; Loy, A. *Geomicrobiology: Molecular and Environmental Perspective*; 2010; pp. 1–437.
117. Liu, J.; Sui, Y.; Yu, Z.; Shi, Y.; Chu, H.; Jin, J.; Liu, X.; Wang, G. High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. *Soil Biol. Biochem.* **2014**, *70*, 113–122.
118. Zhang, B.; Wu, X.; Zhang, G.; Li, S.; Zhang, W.; Chen, X.; Sun, L.; Zhang, B.; Liu, G.; Chen, T. The diversity and biogeography of the communities of Actinobacteria in the forelands of glaciers at a continental scale. *Environ. Res. Lett.* **2016**, *11*, 54012.
119. Sun, B.; Wang, F.; Jiang, Y.; Li, Y.; Dong, Z.; Li, Z.; Zhang, X.X. A long-term field experiment of soil transplantation demonstrating the role of contemporary geographic separation in shaping soil microbial community structure. *Ecol. Evol.* **2014**, *4*, 1073–1087.

120. Reddy, B.V.B.; Kallifidas, D.; Kim, J.H.; Charlop-Powers, Z.; Feng, Z.; Brady, S.F. Natural product biosynthetic gene diversity in geographically distinct soil microbiomes. *Appl. Environ. Microbiol.* **2012**, *78*, 3744–3752.
121. Charlop-Powers, Z.; Owen, J.G.; Reddy, B.V.B.; Ternei, M.; Guimaraes, D.O.; De Frias, U.A.; Pupo, M.T.; Seepe, P.; Feng, Z.; Brady, S.F. Global biogeographic sampling of bacterial secondary metabolism. *eLIFE* **2015**, *2015*, 1–10.
122. Ramirez, K.S.; Leff, J.W.; Barberán, A.; Bates, S.T.; Betley, J.; Crowther, T.W.; Kelly, E.F.; Oldfield, E.E.; Shaw, E.A.; Steenbock, C.; et al. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B* **2014**, *281*, 20141988.
123. Charlop-Powers, Z.; Pregitzer, C.C.; Lemetre, C.; Ternei, M.A.; Maniko, J.; Hover, B.M.; Calle, P.Y.; McGuire, K.L.; Garbarino, J.; Forgione, H.M.; et al. Urban park soil microbiomes are a rich reservoir of natural product biosynthetic diversity. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 201615581.
124. Baker, G.H.; Thumlert, T.A.; Meisel, L.S.; Carter, P.J.; Kilpin, G.P. "Earthworms downunder": A survey of the earthworm fauna of urban and agricultural soils in Australia. *Soil Biol. Biochem.* **1997**, *29*, 589–597.
125. Sterflinger, K.; Prillinger, H. Molecular taxonomy and biodiversity of rock fungal communities in an urban environment (Vienna, Austria). *Antonie Van Leeuwenhoek* **2001**, *80*, 275–286.
126. Borsetto, C.; Wellington, E.M. Bioprospecting Soil Metagenomes for Antibiotics. In *Bioprospecting Success, Potential and Constraints*; Paterson, R., Lima, N., Eds.; Springer: Cham, Switzerland, 2017; pp. 113–136.
127. Gerwick, W.H.; Fenner, A.M. Drug Discovery from Marine Microbes. *Microb. Ecol.* **2013**, *65*, 800–806.
128. Jensen, P.; Fenical, W. Marine bacterial diversity as a resource for novel microbial products. *J. Ind. Microbiol.* **1996**, *17*, 346–351.
129. Crawford, J.M.; Clardy, J. Bacterial symbionts and natural products. *Chem. Commun.* **2011**, *47*, 7559–7566.
130. Newman, D.J. Predominately Uncultured Microbes as Sources of Bioactive Agents. *Front. Microbiol.* **2016**, *7*, 1–15.
131. Seipke, R.F.; Kaltenpoth, M.; Hutchings, M.I. Streptomyces as symbionts: An emerging and widespread theme? *FEMS Microbiol. Rev.* **2012**, *36*, 862–876.
132. Kaltenpoth, M. Actinobacteria as mutualists: General healthcare for insects? *Trends Microbiol.* **2009**, *17*, 529–535.
133. Aylward, F.O.; Suen, G.; Biedermann, P.H.W.; Adams, A.S.; Scott, J.J.; Malfatti, S.A.; Glavina, T.; Tringe, S.G.; Poulsen, M.; Raffa, K.F.; et al. Convergent Bacterial Microbiotas in the Fungal Agricultural Systems of Insects. *MBio* **2014**, *5*, e02077.
134. Kaltenpoth, M.; Biedermann, P.H.W.; Engl, T.; Kaltenpoth, M. Themed issue: Chemical Ecology. **2015**, *32*.
135. Menezes, C.; Vollet-Neto, A.; Marsaioli, A.J.; Zampieri, D.; Fontoura, I.C.; Luchessi, A.D.; Imperatriz-Fonseca, V.L. A Brazilian social bee must cultivate fungus to survive. *Curr. Biol.* **2015**, *25*, 2851–2855.
136. Beemelmanns, C.; Guo, H.; Rischer, M.; Poulsen, M. Natural products from microbes associated with insects. *Beilstein J. Org. Chem.* **2016**, *12*, 314–327.
137. O'Brien, J.; Wright, G.D. An ecological perspective of microbial secondary metabolism. *Curr. Opin. Biotechnol.* **2011**, *22*, 552–558.
138. Cantley, A.M.; Clardy, J. Animals in a bacterial world: Opportunities for chemical ecology. *Nat. Prod. Rep.* **2015**, *32*, 882–892.
139. Klassen, J.L. Microbial secondary metabolites and their impacts on insect symbioses. *Curr. Opin. Insect Sci.* **2014**, *4*, 15–22.
140. Lombardo, M.P. Access to mutualistic endosymbiotic microbes: An underappreciated benefit of group living. *Behav. Ecol. Sociobiol.* **2008**, *62*, 479–497.
141. Currie, C.R.; Wong, B.; Stuart, A.E.; Schultz, T.R.; Rehner, S.A.; Mueller, U.G.; Sung, G.-H.; Spatafora, J.W.; Straus, N. A Ancient tripartite coevolution in the attine ant-microbe symbiosis. *Science* **2003**, *299*, 386–388.
142. Currie, C.R.; Scott, J.A.; Summerbell, R.C.; Malloch, D. Fungus-growing ants use antibiotic-producing bacteria to control garden parasites. *Nature* **1999**, *398*, 701–704.
143. Mueller, U.G.; Dash, D.; Rabeling, C.; Rodrigues, A. Coevolution between attine ants and actinomycete bacteria: A reevaluation. *Evolution* **2008**, *62*, 2894–2912.
144. Sen, R.; Ishak, H.D.; Estrada, D.; Dowd, S.E.; Hong, E.; Mueller, U.G. Generalized antifungal activity and 454-screening of Pseudonocardia and Amycolatopsis bacteria in nests of fungus-growing ants. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 17805–17810.
145. Caldera, E.J.; Currie, C.R. The Population Structure of Antibiotic-Producing Bacterial Symbionts of Apterostigma dentigerum Ants: Impacts of Coevolution and Multipartite Symbiosis. *Am. Nat.* **2012**, *180*, 604–617.
146. Van Arnam, E.B.; Ruzzini, A.C.; Sit, C.S.; Horn, H.; Pinto-Tomás, A.A.; Currie, C.R.; Clardy, J. Selvamicin, an atypical antifungal polyene from two alternative genomic contexts. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 12940–12945.
147. Kost, C.; Lakatos, T.; Böttcher, I.; Arendholz, W.R.; Redenbach, M.; Wirth, R. Non-specific association between filamentous bacteria and fungus-growing ants. *Naturwissenschaften* **2007**, *94*, 821–828.
148. Haeder, S.; Wirth, R.; Herz, H.; Spiteller, D. Candicidin-producing Streptomyces support leaf-cutting ants to protect their fungus garden against the pathogenic fungus Escovopsis. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 4742–4746.
149. Santos, A.V.; Dillon, R.J.; Dillon, V.M.; Reynolds, S.E.; Samuels, R.I. Ocurrence of the antibiotic producing bacterium *Burkholderia* sp. in colonies of the leaf-cutting ant Atta sexdens rubropilosa. *FEMS Microbiol. Lett.* **2004**, *239*, 319–323.

150. Oh, D.-C.; Poulsen, M.; Currie, C.R.; Clardy, J. Dentigerumycin: A bacterial mediator of an ant-fungus symbiosis. *Nat. Chem. Biol.* **2009**, *5*, 391–393.
151. Barke, J.; Seipke, R.F.; Grüschow, S.; Heavens, D.; Drou, N.; Bibb, M.J.; Goss, R.J.M.; Yu, D.W.; Hutchings, M.I. A mixed community of actinomycetes produce multiple antibiotics for the fungus farming ant Acromyrmex octospinosus. *BMC Biol.* **2010**, *8*, 109.
152. Seipke, R.F.; Barke, J.; Brearley, C.; Hill, L.; Yu, D.W.; Goss, R.J.M.; Hutchings, M.I. A single Streptomyces symbiont makes multiple antifungals to support the fungus farming ant acromyrmex octospinosus. *PLoS ONE* **2011**, *6*, e22028.
153. Gao, H.; Grüschow, S.; Barke, J.; Seipke, R.F.; Hill, L.M.; Orivel, J.; Yu, D.W.; Hutchings, M.; Goss, R.J.M. Filipins: The first antifungal "weed killers" identified from bacteria isolated from the trap-ant. *RSC Adv.* **2014**, *4*, 57267–57270.
154. Aanen, D.K.; Eggleton, P.; Rouland-Lefevre, C.; Guldberg-Froslev, T.; Rosendahl, S.; Boomsma, J.J. The evolution of fungus-growing termites and their mutualistic fungal symbionts. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14887–14892.
155. Visser, A.A.; Nobre, T.; Currie, C.R.; Aanen, D.K.; Poulsen, M. Exploring the Potential for Actinobacteria as Defensive Symbionts in Fungus-Growing Termites. *Microb. Ecol.* **2012**, *63*, 975–985.
156. Igarashi, M.; Shida, T.; Sasaki, Y.; Kinoshita, N.; Naganawa, H.; Hamada, M.; Takeuchi, T. Vinylamycin, a new depsipeptide antibiotic, from *Streptomyces* sp. *J. Antibiot.* **1999**, *52*, 873–879.
157. Carr, G.; Poulsen, M.; Klassen, J.L.; Hou, Y.; Wyche, T.P.; Bugni, T.S.; Currie, C.R.; Clardy, J. Microtermolides A and B from Termite-associated actinomyctes. *Org. Lett.* **2012**, *14*, 2822–2825.
158. Kim, K.H.; Ramadhar, T.R.; Beemelmanns, C.; Cao, S.; Poulsen, M.; Currie, C.R.; Clardy, J. Natalamycin A, an Ansamycin from a Termite-Associated *Streptomyces* sp. *Chem. Sci.* **2014**, *5*, 4333–4338.
159. Um, S.; Fraimout, A.; Sapountzis, P.; Oh, D.-C.; Poulsen, M. The fungus-growing termite Macrotermes natalensis harbors bacillaene-producing *Bacillus* sp. that inhibit potentially antagonistic fungi. *Sci. Rep.* **2013**, *3*, 3250.
160. Scott, J.J.; Oh, D.C.; Yuceer, M.C.; Klepzig, K.D.; Clardy, J. Bacterial Protection of Beetle-Fungus Mutualism. *Science* **2008**, *322*, 2008.
161. Yu, F.; Zaleta-Rivera, K.; Zhu, X.; Huffman, J.; Millet, J.C.; Harris, S.D.; Yuen, G.; Li, X.C.; Du, L. Structure and biosynthesis of heat-stable antifungal factor (HSAF), a broad-spectrum antimycotic with a novel mode of action. *Antimicrob. Agents Chemother.* **2007**, *51*, 64–72.
162. Blodgett, J.A.; Oh, D.C.; Cao, S.; Currie, C.R.; Kolter, R.; Clardy, J. Common biosynthetic origins for polycyclic tetramate macrolactams from phylogenetically diverse bacteria. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 11692–11697.
163. Chouvenc, T.; Efstathion, C.A.; Elliott, M.L.; Su, N.-Y. Extended disease resistance emerging from the faecal nest of a subterranean termite. *Proc. Biol. Sci. R. Soc.* **2013**, *280*, 20131885.
164. Madden, A.A.; Grassetti, A.; Soriano, J.N.; Starks, P.T. Actinomycetes with Antimicrobial Activity Isolated from Paper Wasp (Hymenoptera: Vespidae: Polistinae) Nests. *Environ. Entomol.* **2013**, *42*, 703–710.
165. Herzner, G.; Engl, T.; Strohm, E. Cryptic combat against competing microbes is a costly component of parental care in a digger wasp. *Anim. Behav.* **2011**, *82*, 321–328.
166. Kaltenpoth, M.; Gottler, W.; Herzner, G.; Strohm, E. Symbiotic Bacteria Protect Wasp Larvae from Fungal Infestation Martin. *Curr. Biol.* **2005**, *15*, 475–479.
167. Kaltenpoth, M.; Schmitt, T.; Polidori, C.; Koedam, D.; Strohm, E. Symbiotic streptomycetes in antennal glands of the South American digger wasp genus Trachypus (Hymenoptera, Crabronidae). *Physiol. Entomol.* **2010**, *35*, 196–200.
168. Kaltenpoth, M.; Yildirim, E.; Gürbüz, M.F.; Herzner, G.; Strohm, E. Refining the roots of the beewolf-streptomyces symbiosis: Antennal symbionts in the rare genus Philanthinus (Hymenoptera, Crabronidae). *Appl. Environ. Microbiol.* **2012**, *78*, 822–827.
169. Nechitaylo, T.Y.; Westermann, M.; Kaltenpoth, M. Cultivation reveals physiological diversity among defensive "Streptomyces philanthi" symbionts of beewolf digger wasps (Hymenoptera, Crabronidae). *BMC Microbiol.* **2014**, *14*, 202.
170. Shin, H.J.; Jeong, H.S.; Lee, H.S.; Park, S.K.; Kim, H.M.; Kwon, H.J. Isolation and structure determination of streptochlorin, an antiproliferative agent from a marine-derived *Streptomyces* sp. 04DH110. *J. Microbiol. Biotechnol.* **2007**, *17*, 1403–1406.
171. Takahashi, N.; Suzuki, A.; Tamura, S. Structure of Piericidin A. *J. Am. Chem. Soc* **1965**, *87*, 2066–2068.
172. Kroiss, J.; Kaltenpoth, M.; Schneider, B.; Schwinger, M.-G.; Hertweck, C.; Maddula, R.K.; Strohm, E.; Svatos, A. Symbiotic Streptomycetes provide antibiotic combination prophylaxis for wasp offspring. *Nat. Chem. Biol.* **2010**, *6*, 261–263.
173. Poulsen, M.; Oh, D.C.; Clardy, J.; Currie, C.R. Chemical analyses of wasp-associated Streptomyces bacteria reveal a prolific potential for natural products discovery. *PLoS ONE* **2011**, *6*, e16763.
174. Oh, D.C.; Poulsen, M.; Currie, C.R.; Clardy, J. Sceliphrolactam, a polyene macrocyclic lactam from a wasp-associated *Streptomyces* sp. *Org. Lett.* **2011**, *13*, 752–755.
175. Park, S.H.; Moon, K.; Bang, H.S.; Kim, S.H.; Kim, D.G.; Oh, K.B.; Shin, J.; Oh, D.C. Tripartilactam, a cyclobutane-bearing tricyclic lactam from a *Streptomyces* sp. in a dung beetle's brood ball. *Org. Lett.* **2012**, *14*, 1258–1261.
176. Piel, J. A polyketide synthase-peptide synthetase gene cluster from an uncultured bacterial symbiont of Paederus beetles. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14002–14007.

177. Nakabachi, A.; Ueoka, R.; Oshima, K.; Teta, R.; Mangoni, A.; Gurgui, M.; Oldham, N.J.; Van Echten-Deckert, G.; Okamura, K.; Yamamoto, K.; et al. Defensive bacteriome symbiont with a drastically reduced genome. *Curr. Biol.* **2013**, *23*, 1478–1484.
178. Chapter I. The Arid Environments. Available online: http://www.fao.org/docrep/t0122e/t0122e03.htm .
179. Pointing, S.B.; Chan, Y.; Lacap, D.C.; Lau, M.C.Y.; Jurgens, J.A.; Farrell, R.L. Highly specialized microbial diversity in hyper-arid polar desert. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 19964–19969.
180. Neilson, J.W.; Quade, J.; Ortiz, M.; Nelson, W.M.; Legatzki, A.; Tian, F.; LaComb, M.; Betancourt, J.L.; Wing, R.A.; Soderlund, C.A.; et al. Life at the hyperarid margin: Novel bacterial diversity in arid soils of the Atacama Desert, Chile. *Extremophiles* **2012**, *16*, 553–566.
181. Connon, S.A.; Lester, E.D.; Shafaat, H.S.; Obenhuber, D.C.; Ponce, A. Bacterial diversity in hyperarid atacama desert soils. *J. Geophys. Res. Biogeosci.* **2007**, *112*, 1–9.
182. Chanal, A.; Chapon, V.; Benzerara, K.; Barakat, M.; Christen, R.; Achouak, W.; Barras, F.; Heulin, T. The desert of Tataouine: An extreme environment that hosts a wide diversity of microorganisms and radiotolerant bacteria. *Environ. Microbiol.* **2006**, *8*, 514–525.
183. Saul-Tcherkas, V.; Steinberger, Y. Soil Microbial Diversity in the Vicinity of a Negev Desert Shrub-Reaumuria negevensis. *Microb. Ecol.* **2011**, *61*, 64–81.
184. An, S.; Couteau, C.; Luo, F.; Neveu, J.; DuBow, M.S. Bacterial Diversity of Surface Sand Samples from the Gobi and Taklamaken Deserts. *Microb. Ecol.* **2013**, *66*, 850–860.
185. Xu, Z.; Hansen, M.A.; Hansen, L.H.; Jacquiod, S.; Sørensen, S.J. Bioinformatic approaches reveal metagenomic characterization of soil microbial community. *PLoS ONE* **2014**, *9*, e93445.
186. Owen, J.G.; Reddy, B.V.B.; Ternei, M.A.; Charlop-Powers, Z.; Calle, P.Y.; Kim, J.H.; Brady, S.F. Mapping gene clusters within arrayed metagenomic libraries to expand the structural diversity of biomedically relevant natural products. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 11797–11802.
187. Feng, Z.; Kallifidas, D.; Brady, S.F. Functional analysis of environmental DNA-derived type II polyketide synthases reveals structurally diverse secondary metabolites. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12629–12634.
188. Feng, Z.; Kim, J.H.; Brady, S.F. Fluostatins produced by the heterologous expression of a TAR reassembled environmental DNA derived type II PKS gene cluster. *J. Am. Chem. Soc.* **2010**, *132*, 11902–11903.
189. Kang, H.S.; Brady, S.F. Mining soil metagenomes to better understand the evolution of natural product structural diversity: Pentangular polyphenols as a case study. *J. Am. Chem. Soc.* **2014**, *136*, 18111–18119.
190. Lahoum, A.; Aouiche, A.; Bouras, N.; Verheecke, C.; Klenk, H.-P.; Sabaou, N.; Mathieu, F. Antifungal activity of a Saharan strain of *Actinomadura* sp. ACD1 against toxigenic fungi and other pathogenic microorganisms. *J. Mycol. Méd.* **2016**, *26*, 193–200.
191. Boudjella, H.; Bouti, K.; Zitouni, A.; Mathieu, F.; Lebrihi, A.; Sabaou, N. Taxonomy and chemical characterization of antibiotics of Streptosporangium Sg 10 isolated from a Saharan soil. *Microbiol. Res.* **2006**, *161*, 288–298.
192. Boubetra, D.; Sabaou, N.; Zitouni, A.; Bijani, C.; Lebrihi, A.; Mathieu, F. Taxonomy and chemical characterization of new antibiotics produced by Saccharothrix SA198 isolated from a Saharan soil. *Microbiol. Res.* **2013**, *168*, 223–230.
193. Gómez-Silva, B.; Rainey, F.A.; Warren-Rhodes, K.A.; Mckay, C.P.; Navarro-González, R. Atacama Desert Soil Microbiology. In *Microbiology of Extreme Soils*; **2008**; 117–132.
194. Drees, K.P.; Neilson, J.W.; Betancourt, J.L.; Quade, J.; Henderson, D.A.; Pryor, B.M.; Maier, R.M. Bacterial community structure in the hyperarid core of the Atacama Desert, Chile. *Appl. Environ. Microbiol.* **2006**, *72*, 7902–7908.
195. Piubeli, F.; de Lourdes Moreno, M.; Kishi, L.T.; Henrique-Silva, F.; García, M.T.; Mellado, E. Phylogenetic Profiling and Diversity of Bacterial Communities in the Death Valley, an Extreme Habitat in the Atacama Desert. *Indian J. Microbiol.* **2015**, *55*, 392–399.
196. Azua-Bustos, A.; Caro-Lara, L.; Vicuña, R. Discovery and microbial content of the driest site of the hyperarid Atacama Desert, Chile. *Environ. Microbiol. Rep.* **2015**, *7*, 388–394.
197. Okoro, C.K.; Brown, R.; Jones, A.L.; Andrews, B.A.; Asenjo, J.A.; Goodfellow, M.; Bull, A.T. Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. *Antonie Van Leeuwenhoek* **2009**, *95*, 121–133.
198. Nachtigall, J.; Kulik, A.; Helaly, S.; Bull, A.T.; Goodfellow, M.; Asenjo, J.A.; Maier, A.; Wiese, J.; Imhoff, J.F.; Sussmuth, R.D.; et al. Atacamycins A–C, 22-membered antitumor macrolactones produced by *Streptomyces* sp. C38*. *J. Antibiot* **2011**, *64*, 775–780.
199. Rateb, M.E.; Houssen, W.E.; Arnold, M.; Abdelrahman, M.H.; Deng, H.; Harrison, W.T.A.; Okoro, C.K.; Asenjo, J.A.; Andrews, B.A.; Ferguson, G.; et al. Chaxamycins A–D, bioactive ansamycins from a hyper-arid desert *Streptomyces* sp. *J. Nat. Prod.* **2011**, *74*, 1491–1499.
200. Joseph, I.O.F.; Wall, T.E.; Tanner, J.R.; Tawaha, K.; Alali, F.Q.; Li, C.; Oberlies, N.H. Proliferation of antibiotic-producing bacteria and concomitant antibiotic production as the basis for the antibiotic activity of Jordan's red soils. *Appl. Environ. Microbiol.* **2009**, *75*, 2735–2741.
201. Mazzola, M. Assessment and Management of Soil Microbial Community Structure for Disease Suppression 1. *Annu. Rev. Phytopathol.* **2004**, *42*, 35–59.

202. Mazzola, M. Mechanisms of natural soil suppressiveness to soilborne diseases. *Antonie Van Leeuwenhoek* **2002**, *81*, 557–564.
203. Janvier, C.; Villeneuve, F.; Alabouvette, C.; Edel-Hermann, V.; Mateille, T.; Steinberg, C. Soil health through soil disease suppression: Which strategy from descriptors to indicators? *Soil Biol. Biochem.* **2007**, *39*, 1–23.
204. Van Elsas, J.D.; Costa, R.; Jansson, J.; Sjöling, S.; Bailey, M.; Nalin, R.; Vogel, T.M.; van Overbeek, L. The metagenomics of disease-suppressive soils—Experiences from the METACONTROL project. *Trends Biotechnol.* **2008**, *26*, 591–601.
205. Smanski, M.J.; Schlatter, D.C.; Kinkel, L.L. Leveraging ecological theory to guide natural product discovery. *J. Ind. Microbiol. Biotechnol.* **2016**, *43*, 115–128.
206. Adesina, M.F.; Lembke, A.; Costa, R.; Speksnijder, A.; Smalla, K. Screening of bacterial isolates from various European soils for in vitro antagonistic activity towards Rhizoctonia solani and Fusarium oxysporum: Site-dependent composition and diversity revealed. *Soil Biol. Biochem.* **2007**, *39*, 2818–2828.
207. Garbeva, P.; Postma, J.; Van Veen, J.A.; Van Elsas, J.D. Effect of above-ground plant species on soil microbial community structure and its impact on suppression of Rhizoctonia solani AG3. *Environ. Microbiol.* **2006**, *8*, 233–246.
208. Bakker, M.G.; Otto-Hanson, L.; Lange, A.J.; Bradeen, J.M.; Kinkel, L.L. Plant monocultures produce more antagonistic soil Streptomyces communities than high-diversity plant communities. *Soil Biol. Biochem.* **2013**, *65*, 304–312.
209. Kinkel, L.L.; Bakker, M.G.; Schlatter, D.C. A coevolutionary framework for managing disease-suppressive soils. *Annu. Rev. Phytopathol.* **2011**, *49*, 47–67.
210. Bailey, K.L.; Lazarovits, G. Suppressing soil-borne diseases with residue management and organic amendments. **2003**, *72*, 169–180.
211. Bonanomi, G.; Antignani, V.; Capodilupo, M.; Scala, F. Soil Biology & Biochemistry Identifying the characteristics of organic soil amendments that suppress soilborne plant diseases. *Soil Biol. Biochem.* **2010**, *42*, 136–144.
212. Hadar, Y.; Papadopoulou, K.K. Suppressive composts: Microbial ecology links between abiotic environments and healthy plants. *Annu. Rev. Phytopathol.* **2012**, *50*, 133–153.
213. Boulter, J.I.; Trevors, J.T.; Boland, G.J. Microbial studies of compost: Bacterial identification, and their potential for turfgrass pathogen suppression. *World J. Microbiol. Biotechnol.* **2002**, *18*, 661–671.
214. Boulter-Bitzer, J.I.; Trevors, J.T.; Boland, G.J. A polyphasic approach for assessing maturity and stability in compost intended for suppression of plant pathogens. *Appl. Soil Ecol.* **2006**, *34*, 65–81.
215. Bossio, D.A. Scow, K.M. Gunapala, N. Graham, K.J. Determinants of Soil Microbial Communities: Effects of Agricultural Management, Season, and Soil Type on Phospholipid Fatty Acid Profiles. *Microb. Ecol.* **1998**, *36*, 1–12.
216. Raaijmakers, J.M.; Mazzola, M. Insights perspectives. *Science* **2016**, *352*, 1392–1393.
217. Martin, C.C.G. Compost Tea. In *Organic Amendments and Soil Suppressiveness in Plant Disease Management*; 2015; pp. 25–49.
218. Elad, Y.; Shtienberg, D. Effect of compost water extracts on grey mould (*Botrytis cinerea*). *J. Crop Prot.* **1994**, *13*, 109–112.
219. Cronin, M.J. Yohalem, D.S. Harris, R.F.; Andrew, J.H. Putative Mechanism and Dynamics of Inhibition of the Apple Scab Pathogen Venturia Inaequalis by compost extracts. *Soil Biol. Biochem.* **1996**, *28*, 1241–1249.
220. Sang, M.K.; Kim, J.; Kim, K.D. Biocontrol Activity and Induction of Systemic Resistance in Pepper by Compost Water Extracts Against Phytophthora capsici. *Phytopathology* **2010**, *100*, 774–783.
221. Mehta, C.M.; Palni, U.; Franke-Whittle, I.H.; Sharma, A.K. Compost: Its role, mechanism and impact on reducing soil-borne plant diseases. *Waste Manag.* **2014**, *34*, 607–622.
222. Bradley, G.G.; Punja, Z.K. Composts containing fluorescent pseudomonads suppress fusarium root and stem rot development on greenhouse cucumber. *Can. J. Microbiol.* **2010**, *1995*, 896–905.
223. Thomashow, L.S. Identification and Characterization of a Gene Cluster for Synthesis of the Polyketide Antibiotic 2,4-Diacetylphloroglucinol from Pseudomonas fluorescens Q2–87. *J. Bacteriol.* **1999**, *181*, 3155–3163.
224. Haas, D.; Défago, G. Biological Control of Soil-Borne Pathogens by Fluorescent Pseudomonads. *Nat. Rev. Microbiol.* **2005**, *3*, 307–319.
225. Igarashi, Y.; Iida, T.; Oku, N.; Watanabe, H.; Furihata, K.; Miyanouchi, K. Nomimicin, a new spirotetronate-class polyketide from an actinomycete of the genus Actinomadura. *J. Antibiot.* **2012**, *65*, 355–359.
226. Sun, X.; Meng, J.; Liu, S.; Zhang, H; Wang, L. Draft Genome Sequence of *Streptomyces* sp. F-3. *Genome Announc.* **2013**, *4*, e00780-16.
227. Komaki, H.; Ichikawa, N.; Hosoyama, A.; Fujita, N.; Harunari, E.; Igarashi, Y. Draft Genome Sequence of an Anthracimycin Producer, *Streptomyces* sp. TP-A0875. *Genome Announc.* **2015**, *3*, e01149-15.
228. Bhullar, K.; Waglechner, N.; Pawlowski, A.; Koteva, K.; Banks, E.D.; Johnston, M.D.; Barton, H.A.; Wright, G.D. Antibiotic resistance is prevalent in an isolated cave microbiome. *PLoS ONE* **2012**, *7*, e34953.
229. Žák, K.; Světlík, I.V.; Krištůfek, D.; Elhottová, Ľ.; Kováč, A. Chroňáková, K. Žák, I. Světlík: The age of bat guano heap in Domica Cave (Slovak Karst NP) and electron microscopy of bat excrements. **2008**, 163–170.
230. Nováková, A.; Elhottová, D.; Krištůfek, V.; Lukešová, A.; Hill, P.; Kováč, Ľ.; Mock, A.; Ľuptáčik, P. Feeding sources of invertebrates in theArdovská Cave and Domica Cave systems—Preliminary results. *Contrib. Soil Zool. Cent. Eur. I* **2005**, 107–112.

231. Jones, D.S.; Macalady, J.L. *Their World: A Diversity of Microbial Environments*; 2016; Volume 1, pp. 203–224.
232. Barton, H.A.; Jurado, V.; Barton, H.A. What's Up Down There? Microbial Diversity in Caves. **2007**, *2*, 132–138.
233. Jurado, V.; Laiz, L.; Rodriguez-Nava, V.; Boiron, P.; Hermosin, B.; Sanchez-Moral, S.; Saiz-Jimenez, C. Pathogenic and opportunistic microorganisms in caves. *Int. J. Speleol.* **2010**, *39*, 15–24.
234. Schabereiter-Gurtner, C.; Saiz-Jimenez, C.; Piñar, G.; Lubitz, W.; Rölleke, S. Phylogenetic diversity of bacteria associated with Paleolithic paintings and surrounding rock walls in two Spanish caves (Llonín and La Garma). *FEMS Microbiol. Ecol.* **2004**, *47*, 235–247.
235. Barton, H.A.; Taylor, N.M.; Kreate, M.P.; Springer, A.C.; Oehrle, S.A.; Bertog, J.L. The impact of host rock geochemistry on bacterial community structure in oligotrophic cave environments. *Int. J. Speleol.* **2007**, *36*, 93–104.
236. Tomczyk-Żak, K.; Zielenkiewicz, U. Microbial diversity in caves. *Geomicrobiol. J.* **2016**, *33*, 20–38.
237. Tiwari, K.; Gupta, R.K. Rare actinomycetes: A potential storehouse for novel antibiotics. *Crit. Rev. Biotechnol.* **2012**, *32*, 108–132.
238. Groth, I.; Schumann, P.; Laiz, L.; Sanchez-Moral, S.; Cañaveras, J.C.; Saiz-Jimenez, C. Geomicrobiological Study of the Grotta dei Cervi, Porto Badisco, Italy. *Geomicrobiol. J.* **2001**, *18*, 241–258.
239. Herold, K.; Gollmick, F.A.; Groth, I.; Roth, M.; Menzel, K.D.; Möllmann, U.; Gräfe, U.; Hertweck, C. Cervimycin A–D: A polyketide glycoside complex from a cave bacterium can defeat vancomycin resistance. *Chemistry* **2005**, *11*, 5523–5530.
240. Herold, K.; Xu, Z.; Gollmick, F.A; Grafe, U.; Hertweck, C. Biosynthesis of cervimycin C, an aromatic polyketide antibiotic bearing an unusual dimethylmalonyl moiety. *Org. Biomol. Chem.* **2004**, *2*, 2411–2414.
241. Derewacz, D.K.; McNees, C.R.; Scalmani, G.; Covington, C.L.; Shanmugam, G.; Marnett, L.J.; Polavarapu, P.L.; Bachmann, B.O. Structure and stereochemical determination of hypogeamicins from a cave-derived actinomycete. *J. Nat. Prod.* **2014**, *77*, 1759–1763.
242. Barton, H.A.; Northup, D.E. Geomicrobiology in cave enviroments: Past, current and future prespectives. *J. Cave Karst Stud.* **2007**, *69*, 163–178.
243. Reinbacher, W.R. Is it gnome, is it berg, is it mont, is it mond? An updated view of the origin and etymology of moonmilk. *Bull. Natl. Speleol. Soc.* **1994**, *56*, 1–13.
244. Axenov-Gibanov, D.V.; Voytsekhovskaya, I.V.; Tokovenko, B.T.; Protasov, E.S.; Gamaiunov, S.V.; Rebets, Y.V.; Luzhetskyy, A.N.; Timofeyev, M.A. Actinobacteria Isolated from an Underground Lake and Moonmilk Speleothem from the Biggest Conglomeratic Karstic Cave in Siberia as Sources of Novel Biologically Active Compounds. *PLoS ONE* **2016**, *11*, e0149216.
245. Varela, A.R.; Dapkevicius, M.L.N.E.; Northup, D.E. Microorganisms isolated from Azorean lava tubes have antimicrobial activity towards food-borne pathogens. *Actas do 9° Encontro Química dos Aliment.* In Proceedings of the 9th Food Chemistry Meeting of the Portuguese Society for Chemistry, 2009; p. 146.
246. Cheeham, N.; Sadoway, T.; Rule, D.; Watson, K.; Moote, P.; Soliman, L.C.; Azad, N.; Donkor, K.K.; Horne, D. Cure from the cave: Volcanic cave actinomycetes and their potential in drug discovery. *Int. J. Speleol.* **2013**, *42*, 35–47.
247. Hathaway, J.J.M.; Garcia, M.G.; Balasch, M.M.; Spilde, M.N.; Stone, F.D.; De Lurdes, M.; Dapkevicius, N.E.; Amorim, D.E.I.R.; Gabriel, R.; Borges, P.A.V.; et al. Comparison of Bacterial Diversity in Azorean and Hawai'ian Lava Cave Microbial Mats. *Geomicrobiol. J.* **2014**, *313*, 205–220.
248. Riquelme, C.; Marshall Hathaway, J.J.; Enes Dapkevicius, M. de L.N.; Miller, A.Z.; Kooser, A.; Northup, D.E.; Jurado, V.; Fernandez, O.; Saiz-Jimenez, C.; Cheeptham, N. Actinobacterial Diversity in Volcanic Caves and Associated Geomicrobiological Interactions. *Front. Microbiol.* **2015**, *6*, 1–16.
249. Jose, P.A.; Robinson, S.; Jebakumar, D. Unexplored hypersaline habitats are sources of novel actinomycetes. *Front. Microbiol* **2014**, *5*, 242.
250. Park, H.B.; Lee, J.K.; Lee, K.R.; Kwon, H.C. Angumycinones A and B, two new angucyclic quinones from *Streptomyces* sp. KMC004 isolated from acidic mine drainage. *Tetrahedron Lett.* **2014**, *55*, 63–66.
251. Ding, Z.G.; Li, M.G.; Zhao, J.Y.; Ren, J.; Huang, R.; Xie, M.J.; Cui, X.L.; Zhu, H.J.; Wen, M.L. Naphthospironone A: An unprecedented and highly functionalized polycyclic metabolite from an alkaline mine waste extremophile. *Chemistry* **2010**, *16*, 3902–3905.
252. Tang, S.K.; Wang, Y.; Klenk, H.P.; Shi, R.; Lou, K.; Zhang, Y.J.; Chen, C.; Ruan, J.S.; Li, W.J. *Actinopolyspora alba* sp. nov. and *Actinopolyspora erythraea* sp. nov., isolated from a salt field, and reclassification of Actinopolyspora iraqiensis Ruan et al. 1994 as a heterotypic synonym of *Saccharomonospora halophila*. *Int. J. Syst. Evol. Microbiol.* **2011**, *61*, 1693–1698.
253. Zhao, L.X.; Huang, S.X.; Tang, S.K.; Jiang, C.L.; Duan, Y.; Beutler, J.A.; Henrich, C.J.; McMahon, J.B.; Schmid, T.; Blees, J.S.; et al. Actinopolysporins A–C and tubercidin as a pdcd4 stabilizer from the halophilic actinomycete Actinopolyspora erythraea YIM 90600. *J. Nat. Prod.* **2011**, *74*, 1990–1995.
254. McKinney, M.L. Effects of urbanization on species richness: A review of plants and animals. *Urban Ecosyst.* **2008**, *11*, 161–176.
255. Faeth, S.H.; Bang, C.; Saari, S. Urban biodiversity: Patterns and mechanisms. *Ann. N. Y. Acad. Sci.* **2011**, *1223*, 69–81.
256. Erséus, C.; Grimm, R.; Healy, B.; Lundberg, S.; Rota, E.; Timm, T. Clitellate diversity in Nationalstadsparken, an urban national park in Stockholm, Sweden. *Hydrobiologia* **1999**, *406*, 101–110.

257. Fang, Y.; Yoh, M.; Koba, K.; Zhu, W.; Takebayashi, Y.; Xiao, Y.; Lei, C.; Mo, J.; Zhang, W.; Lu, X. Nitrogen deposition and forest nitrogen cycling along an urban-rural transect in southern China. *Glob. Chang. Biol.* **2011**, *17*, 872–885.
258. McDonnell, MarkPickett, S.; Pickett, S.; Groffman, P.; Groffman, P.; Bohlen, P.; Bohlen, P.; Pouyat, R.; Pouyat, R.; Zipperer, W.; Zipperer, W.; et al. Ecosystem processes along an urban to rural gradient. *Urban Ecosyst.* **1997**, *1*, 21–36.
259. Gosse, J.T.; Hill, P.; Dowd, S.E.; Boddy, C.N. Draft Genome Sequence of *Streptomyces* sp. Strain PBH53, Isolated from an Urban Environment. *Genome Announc.* **2015**, *3*, e00859-15.
260. Weber, C.F.; Werth, J.T. Is the lower atmosphere a readily accessible reservoir of culturable, antimicrobial compound-producing Actinomycetales? *Front. Microbiol.* **2015**, *6*, 802.
261. Griffiths, R.I.; Thomson, B.C.; Plassart, P.; Gweon, H.S.; Stone, D.; Creamer, R.E.; Lemanceau, P.; Bailey, M.J. Mapping and validating predictions of soil bacterial biodiversity using European and national scale datasets. *Appl. Soil Ecol.* **2016**, *97*, 61–68.
262. Levy, S.E.; Myers, R.M. Advancements in Next-Generation Sequencing. *Annu. Rev. Genom. Hum. Genet.* **2016**, *17*, 95–115.
263. Heuer, H.; Krsek, M.; Baker, P.; Smalla, K.; Wellington, E.M.H. Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. *Appl. Environ. Microbiol.* **1997**, *63*, 3233–3241.
264. Monciardini, P.; Sosio, M.; Cavaletti, L.; Chiocchini, C.; Stefano, D. New PCR primers for the selective amplication of 16S rDNA from different groups of actinomycetes. *FEMS Microbiol. Ecol.* **2002**, *42*, 419–429.

Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

Under review as: A comparison of hard and soft direct methods for DNA extraction from soil. Patrick Hill, Mathieu F Dextraze, David Kroetsch, Christopher N Boddy *Microbiology Spectrum.*

BioRxiv **doi:** https://doi.org/10.1101/2022.03.07.483395

There are now thousands of published studies of soil bacterial populations that bioprospectors can use as a guide to pick samples to build metagenomics libraries from or use to cultivate isolates. Most of these studies use next generation sequencing of 16S PCR amplicons from directly extracted DNA. Many different primer pairs are used to generate these amplicons. Amplicon sequencing is being replaced by shotgun metagenomic sequencing. Both amplicon and metagenomics next generation sequencing use a range of extraction methods to get DNA from the environment. This raises two questions:

How comparable are results from amplicon sequencing and metagenomics studies given the range of extraction methods and primers that they use ?

How much does the fraction of the soil bacterial community that amplicon and metagenomics shotgun sequencing covers differ from that that is available for metagenomics libraries for expression of biosynthetic operons or the cultivation of isolates ?

A full exploration of these questions would be a thesis in itself. This study looks at an extreme case.

Most amplicon sequencing and shotgun metagenomic sequencing studies use direct DNA extraction methods, which use bead beating to lyse bacteria. This study compares the most commonly used direct extraction method, the Mobio kit, with an extraction method that uses enzymatic lysis.

Samples were taken from contrasting soils < 2km apart on the Central Experimental Farm in Ottawa as well as from nearby wooded areas, distant samples (Vancouver, San Diego) as well as compost. DNA was amplified using a range of 16S PCR primers.

Conclusions

16S amplicon sequencing results from contrasting soils and environments can be compared, even if different direct extraction methods or 16S primers were used. It may however, be difficult to compare the bacterial communities of similar soils. Even if the same extraction methods and primers are used on all samples, they may completely miss certain bacterial subgroups.

Results differed more by primer choice than extraction method. Several bacterial groups appeared and disappeared from results depending on the primers used and in one case the PCR conditions and/or Sequencing laboratory. Based on these findings we recommend sending samples for analysis to different sequencing services that use different primer pairs.

Role of Ph.D. candidate in Chapter 2

The Ph.D. candidate carried out:

-sampling all non-soil samples as well as the two woodland soil samples from Ottawa.

-All laboratory work.

-All data analysis except for writing the python code for reformatting Qime assigned taxonomy files.

-writing the first version of the manuscript.

A comparison of hard and soft direct methods for DNA extraction from soil.

Patrick Hill, Mathieu F Dextraze, David Kroetsch, Christopher N Boddy*

*Corresponding Author:cboddy@uottawa.ca *Department of Chemistry and Biomolecular Sciences, Centre for Chemical and Synthetic Biology, University of Ottawa, 10 Marie-Curie, Ottawa, K1N 6N5, Canada.*

Keywords: Soil DNA extraction, JG30-KF-CM45, *Pseudomonadales*, *Desulfuromonadales*,

Enterobacteriales, *Alteromonadales*

Abstract

Nucleic acid extraction is the first step in molecular biology studies of soil bacterial communities. The most common used soil DNA extraction method is the direct, hard extraction Mobio method, which uses bead beating to lyse bacteria. In this study we compared the Mobio method with a soft, enzymatic lysis extraction method. Next generation sequencing (Illumina and Pyrosequencing) of amplicons generated from four 16S primer pairs and DNA from 12 soils and 3 composts was used to compare the two extraction methods.

Four bacterial orders, the delta proteobacterial *Desulfuromonadales* and gamma proteobacterial *Pseudomonadales*, *Enterobacteriales*, and *Alteromonadales* were more common in amplicons from soft extracted DNA, sometimes by two orders of magnitude. These groups can be a significant fraction of the bacterial population. For example the *Pseudomonadales* made up to 16 % and *Enterobacteriales* 10% of amplicons from Soft extracted DNA. The JG30-KF-CM45 order was under extracted by the enzymatic lysis extraction method. Results differed more by primer choice than extraction method and the phylogenetic resolution of differences between extraction methods changed with primer choice.

Given how often Mobio extraction is used, these proteobacterial orders are probably under-represented in the studies of soil bacteria that use nucleic acid methods. Further improvements in soil DNA extraction are needed. Amplicons sequencing studies should use a range of different primers to confirm the phylogenetic resolution of their results.

Importance

Several large scale studies of soil bacteria that compare thousands of soil samples across continents have used the Mobio method for DNA extraction. Large scale studies like these are increasing with the recent establishment of the Global Soil Biodiversity Observation Network (Soil BON), which also uses the Mobio method. The results of this work will be used to make policy decisions about how to manage the soil and may be a guide for bioprospectors. As the Mobio method is so widely used, it is important to know its limitations. Studies that use the Mobio method underestimate the fraction of several proteobacterial groups. Most notably the Enterobacteria and Pseudomonas can be under extracted by 10-100 fold. The degree of under extraction varies with different soils.

Introduction

Most soil bacteria cannot be grown in the laboratory (Torsvik et al., 1990). To study these bacteria without cultivating them, their DNA is extracted from soil and the DNA or PCR products amplified from it are sequenced (Delmont et al., 2011). Bacterial DNA can be extracted from soil indirectly or directly. In indirect extraction, bacteria are taken out of soil before being lysed, in direct extraction bacteria are lysed in place in the soil.

Indirect extraction is used when DNA quality is important. As bacteria are not in the soil when lysed, the extracted DNA is cleaner. Usually “soft” chemical methods such as enzymes and/or detergent are used for lysis so DNA fragments are relatively long (> 50 000 bp). DNA of this length can be useful for

several reasons. Longer fragments are more likely to contain the full length functional units, such as natural product biosynthetic gene clusters, which can then be heterologously expressed (Gabor, Alkema, & Janssen, 2004)., As well as this, the longer fragments are, the more easily they can be assembled into genomes. This may be useful as newer long sequencing methods are used in metagenomic sequencing (Haro-Moreno, López-Pérez, & Rodríguez-Valera, 2020)(Goethem et al., 2021) (Waschulin, James, Newsham, Donadio, & Corre, 2022).

Direct extraction is faster and extracts more of the whole bacterial population than indirect extraction (Delmont et al., 2011) and is used in ecological studies that compare the bacterial populations of soils. These extractions typically use “hard” methods that physically lyse bacteria and are faster than soft lysis. The hard lysis shears extracted DNA into short fragments (< 20 000 bp) (Yeates, Gillings, Davison, Altavilla, & Veal, 1998) (Cullen & Hirsch, 1998) (Bremen, Miller, Bryant, Madsen, & Ghiorse, 1999) (Ínceošlu, Hoogwout, Hill, & Van Elsas, 2010). However, these studies often use PCR to amplify much shorter lengths (< 1500 bp) of ribosomal genes for sequencing, so DNA length does not affect the results.

Since the early 2000s direct extraction is usually done with kits containing premixed reagents and well tested, standardized methods. Kit extraction usually begins by weighing less than a gram of soil or sediment into a tube with beads. The tube is then shaken for a few minutes so that the beads lyse bacterial membranes and cell walls. An extract containing the DNA of the lysed bacteria is then cleaned and concentrated by precipitating out impurities and binding the DNA to a stationary phase in a column before washing and elution.

The most popular of these kits is the Mobio Power soil kit (modified and renamed the DNeasy PowerSoil Kit). It is used by several large scale projects that compare bacterial populations from thousands of soils across large areas (e.g Australia, (Bissett et al., 2016) the European Union (Orgiazzi,

Ballabio, Panagos, Jones, & Fernandez-Ugalde, 2018)) and the Earth Microbiome Project (Thompson et al., 2017). The Mobio Powersoil kit is also the method that will be used in the recently established global Soil Biodiversity Observation Network (Guerra et al., 2021). In this study we refer to this extraction method as the Mobio method.

We compare the Mobio method with a method that is a compromise between direct and indirect methods. It directly extracts DNA from soil using soft lysis (enzymes and SDS). Zhou et al's (1996) soil DNA extraction method uses enzymes, and detergent to lyse bacteria (Zhou, Bruns, & Tiedje, 1996). When used on clayey bamboo forest soils from the Cauca flood plain near Cali, Colombia, this method yielded thick black liquid that could not be read on spectrophotometers nor amplified with PCR. After several modifications such as washing before extraction and altering extraction and cleaning steps, it yielded long fragments of clean DNA from many soils (Peña-Venegas, Cardona, Arguelles, & Arcos, 2007) (Hill et al., 2011) and sediments (Nováková et al., 2005). We call this method the Soft method.

We extracted DNA from soils and compost using both methods, PCR amplified the 16S ribosomal genes using four different primer pairs, and sequenced the amplicons with next generation sequencing. Mobio extracted DNA was more alpha diverse than Soft extracted DNA. However, there were often 10-100 × more sequenced amplicons from several gamma proteobacterial orders (*Enterobacterales* and *Pseudomonadales*) in the soft extracted DNA. These bacterial groups may be under-represented in the many studies that use the Mobio method.

Results and Discussion

This study examines how extraction methods affect our view of soil bacterial populations. Therefore we sampled a broad range of soils only once rather than fewer soils in triplicate as is done to characterize the bacterial community of a particular soil. Twelve contrasting soils and three compost

samples were sampled once. DNA was extracted in triplicate from one of these samples (Exp1B) and in duplicate from a second (Exp2A) to make sure that results were reproducible for technical replicates.

Eight soils were from Ottawa, Canada, two were acid forest podzols from the Mattawa plain (Ontario) and Vancouver, Canada and one each was from San Diego, California and North-East Kansas U.S.A. (Table 2.1). The three compost samples were from the top and bottom of a compost pile receiving garden and household waste (80F and 80F Top) and the top of a compost pile receiving garden waste in a long-term care home (Glebe).

DNA was extracted from the samples using the MoBio and Soft DNA extraction methods. The raw uncleaned and cleaned final DNA yields were measured for each extraction method. Four different primers pairs for differing regions of the 16S rRNA gene were used to generate amplicons from the cleaned final DNA samples. Amplicons were sequenced with several methods of next generation sequencing (Table 2.1).

Table 2.1 Sample description and amplicon sequencing plan

											Pyrosequencing			Illumina	
											EL1	EL2	EL3	EL4	
Sample	**Land Use**	**pH**	**%O.M.**	**%sand**	**%silt**	**%clay**	**Raw DNA Soft/Mobio** µg DNA/g soil	**DNA yields Soft/Mobio** µg DNA/g soil	**Site**	**Comments**	**V4V6**		**V1V3**	**V4V5**	**V6V8**
Exp1B	Soil-Grassland	6.8	10.1	10.6	66.9	22.4	57.8/63.3	1.7/15.2		North Gower Clay Loam	+	(+[a])	+	+	+
Exp2A	Soil-Cultivated	7.2	4.2	49.1	44.3	6.7	19.3/9.6	1.3/7.8		Kars Sandy Loam	+	+		+	+
Exp2B	Soil-Grassland	7.5	4.3	48.6	43.5	7.9	16.7/15.5	1.4/5.0	Contrasting soils, Central Experiment Farm. Ottawa		+			+	
Exp3A	Soil-Cultivated	7.5	6.1	44.5	46.9	7.6	45.9/67.3	2.3/11.2		Grenville Sandy Loam				+	
Exp3B	Soil-Grassland	7.3	7.4	62.6	33.6	3.9	57.8/62.8	1.7/12.1						+	
Exp4A	Soil-Cultivated	7.5	39.1	22.1	69.6	8.4	nd	24.0/8.6		Peat Soil		+		([b])	
BillBrdg	Soil-Forest	6.8	9.9	20.8	61.5	17.7	83.8/51.2	4.6/16.2	Billings Bridge Ottawa		+			+	
RACentre	Soil-Forest	6.3	26.3	18.7	61.4	19.9	88.3/74.1	3.3/29.3	Primary forest RA Centre Ottawa					+	
Vancouver	Soil-Forest	4.3	88.7	nd	nd	nd	82.9/40.0	3.2/11.8	Capilano forest	O horizon,Podzol	+	+		+	+
Mattawa	Soil-Forest	3.6	9.7	66.1	33.9	0	43.1/71.3	2.2/8.3	Mattawa plain	Ae horizon Podzol				+	
Ellsworth	Soil-Grassland	7.5	15.1	35.2	63.1	1.8	nd	4.1/9.5	Ellsworth, Kansas	Loess				+	
Diamond	Soil-Grassland	7.4	5.7	81.4	18.6	0	nd	2.4/6.6	Park, San Diego	Park Turfgrass				+	
80Ftop	Compost	6.7	59.6	nd	nd	nd	50.0/187.8	8.8/25.2	Garden compost, Ottawa South	Top and centre of compost pile.				+	
80F	Compost	7.7	37.9	nd	nd	nd	75.4/106.9	8.3/19.0						+	
Glebe	Compost	6.7	47.6	nd	nd	nd	nd	4.5/13.6	Glebe Centre.	Seniors Home compost.				+	

a-Exp1B EL2 sequencing of Mobio and soft extracted DNA was done for three biological replicates that were sequenced on separate runs. b-Exp4A DNA failed to amplify with V4V5 primers.

Minor effects of extraction method on Alpha and Beta diversity compared to primers and sample.

Amplicons from the five soil samples Exp1B, Exp2A, Exp2B, BillingsBridge, and Vancouver were generated with V4V5 and V4V6 primers from DNA extracted using both methods. V4V6 amplicons were sequenced by pyrosequencing and the V4V5 amplicons by Illumina. Amplicons from V4V5 primers were more alpha diverse than those from V4V6 ($p<0.002$ Chao, $p<0.014$ Faiths PD Supplemental Table 2). There was no statistically significant effect for extraction method. There was a statistically significant effect of extraction method on alpha diversity on results of Illumina V4V5 amplicon sequences from all 11 soils and 3 composts. Mobio amplicons were more alpha diverse than the Soft extracted amplicons ($p<0.013$ Chao, Supplemental Table1).

For the five soils with both V4V6 and V4V5 results (Exp1B, Exp2A, Exp2B, BillingsBridge, Vancouver), primers had a statistically significant effect on beta diversity (Figure 2.1, UniFrac weighted $p<0.001$, UniFrac Unweighted $p<0.001$, Bray-Curtis $p<0.005$ Supplemental Table 3) while extraction method did not. However, weighted UniFrac analysis of V4V5 amplicons from seven soils from Ottawa found that extraction affected clustering ($p<0.017$, Supplemental Table 4).

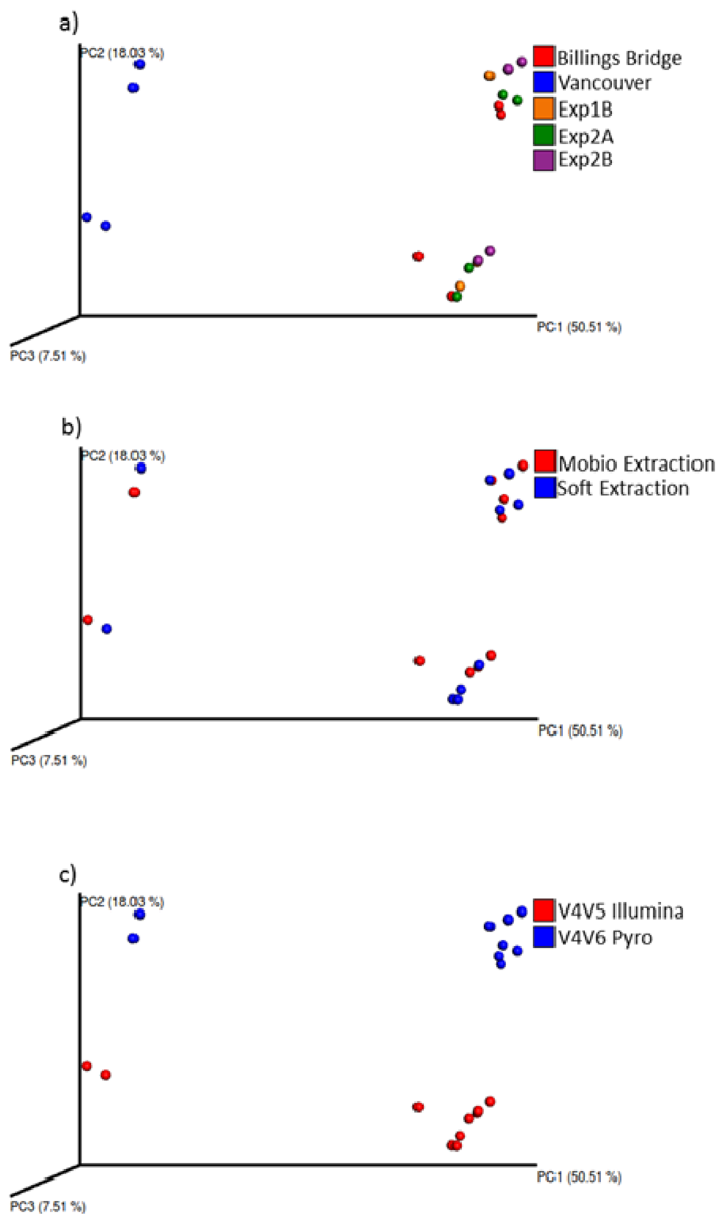


Figure 2.1. Primers affected the apparent beta diversity of soils more than extraction method. V4V6 Pyrosequencing and V4V5 Illumina amplicon sequencing results for the soils: Vancouver, Billings Bridge, Exp1B, Exp2A, Exp2B. Results are presented through weighted UniFrac Principal co ordinate analysis. Samples are colour coded by a-Soil b-DNA extraction method c-Primer Sequencing method.

There was no evidence that using any combination of primers or extraction method could fundamentally alter understanding of how bacterial communities vary between soils or between soils and other biomes. We compared all of our V4V5 and V4V6 amplicons from soil DNA extracted with both soft and mobio extraction with V4V6 amplicons from non-soil samples: compost, faeces, beach sand,

and street dust. UniFrac cluster analysis found that soils clustered together when compared with other environments (Figure 2.2A). There were the same soil outliers using both extraction methods, an urban park soil from San Diego, Diamond and the two acidic podzols (Vancouver, Mattawa). This is not surprising, many previous studies have found that pH strongly affects bacterial populations in soil (e.g.Fierer & Jackson, 2006).

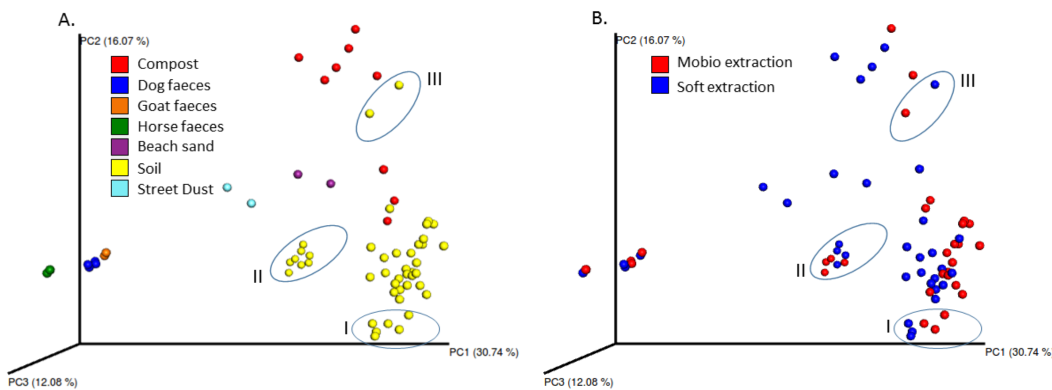


Figure 2.2. Varying primers (A) and extraction method (B) does not alter overall community differences between soils and other biomes. Pyrosequencing and Illumina sequencing results from this study were combined with pyrosequencing results from other studies from faeces, beach sand and street sediments. Parameters were set to the lowest quality reads in the whole batch (faecal sample amplicons. Minimum length was set at 250 bp and sequences were trimmed with SEED2. Results are presented through weighted UniFrac Principal co ordinate analysis. Samples are colour coded by a-Environment b-DNA extraction method. Three clusters of results are circled. I are the results of V4V6 amplicon pyrosequencing of triplicate samples of DNA from Exp1B extracted with Mobio and soft extraction II are results from two acidic podzols, Mattawa and Vancouver. III are results from Diamond soil from San Diego, USA.

These result match those of earlier comparisons of extraction methods across different biomes and contrasting soils. When several bead beating direct extraction methods were used to compare the bacterial communities of sea water, deep sea sediment, soil, mattress dust, the human mouth, and faeces, samples clustered by environment (Marotz, Amir, Humphrey, Gogul, & Knight, 2017). Similarly, when different direct extraction bead beating methods are used on different soils, samples cluster by

soil rather than extraction method (Terrat et al., 2012)(Terrat et al., 2015)(Soliman, Yang, Yamazaki, & Jenke-Kodama, 2017).

However, extraction method can affect results if the samples are very similar. For example, when soil samples were taken from the same field at different depths and DNA extracted using several different direct and indirect methods, extraction method strongly affected the results (Delmont et al., 2011).

Similarly 16S DGGE fingerprints from three direct extraction methods that included both soft extraction and Mobio extraction of DNA from three Dutch potato soils showed that the two sandy soils of similar pH clustered by extraction method (Ínceošlu et al., 2010). Lastly, Roldan et al compared the Mobio method to a second direct bead beating method using Illumina 16S amplicon sequencing on three land uses on the same clayey soil and showed that extraction method affected how samples clustered by weighted Unifrac principle coordinate analysis (Roldan, Junca, & Arbeli, 2019).

We found that extraction method was statistically significant on weighted UniFrac clustering of seven Ottawa mineral soils found within five km of each other with a pH range of 6.3-7.5 (Supplemental Table 4). Thus the choice of DNA extraction method may change the conclusions of studies that compare similar soils or treatments on the same soil. If different biomes are compared though the differences between them outweight those of extraction method.

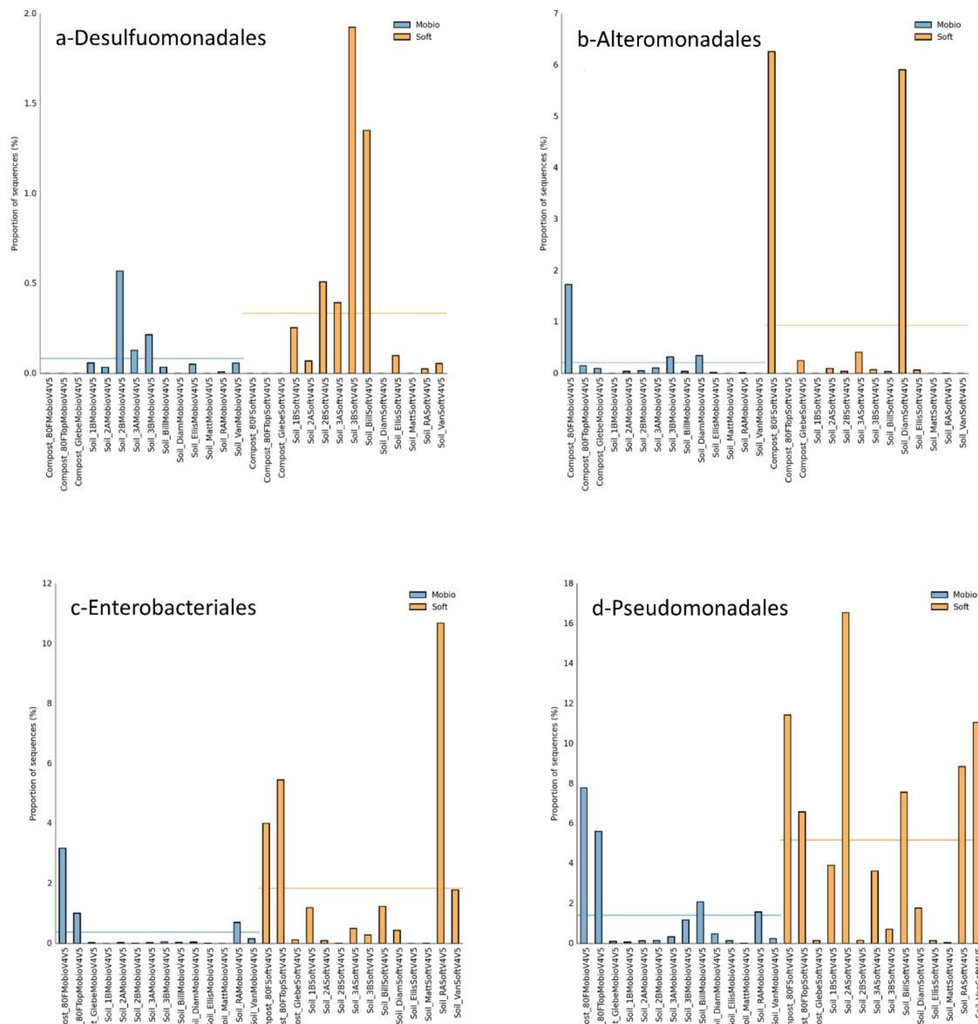
STAMP analysis shows differential extraction of *Enterobacteriales* and *Pseudomonades*

STAMP (Statistical analysis of taxonomic and functional profiles)(Parks, Tyson, Hugenholtz, & Beiko, 2014) group analysis found many statistically significant differences at a range of phylogenetic levels for both extraction method and primers. However, while many differences were statistically significant, they were also small and so would not alter the overall view of the bacterial community (e.g. Supplemental Figure S2). When a cutoff for an effect size of at least four-fold change in abundance or greater was used to refine the STAMP output, there were still significant differences for both primers and extraction method.

The largest of these differences was for primers (V4V5 vs V4V6) for the *planctomycete* phylum, which was preferentially amplified by V4V5 primers (Supplemental figure S1). It is uncertain if the V4V6 primers were under amplifying or V4V5 primers over amplifying *planctomycetes*. Mantri et al compared 16S amplicon sequencing and shotgun sequencing of DNA from four forest soils and found that the *planctomycetes* were overrepresented in the amplicons sequences by several orders of magnitude (Mantri et al., 2021), suggesting that some primers can over amplify *planctomycetes*. However, this study used V3V4 amplicons and so cannot be directly compared to ours.

Extraction method had smaller effects on the abundance of some phylogenetic groups but this effect was found using several primer pairs and thus is unlikely to be an artifact of amplification. There was no statistically significant difference of extraction method for group level analysis of the five soils with V4V5 and V4V6 amplicon data. There was however, for the eleven soils with V4V5 data. Five orders were more common in amplicons from one or the other extraction method (Figure 2.3). Three gamma proteobacterial orders, the *Pseudomonades*, *Enterobacteriales*, *Alteromonadales* and the delta proteobacterial order *Desulfuromonadales* were more abundant in amplicons from Soft extracted DNA. The order JG30-KF-CM45 from the *Thermomicrobia* was more abundant in amplicons from Mobio

extracted DNA. (Supplemental tables 2 and 3). The largest effect for differential extraction was observed for the *Pseudomonadales a*nd *Enterobacteriales*. These orders were often a significant fraction of the 16S amplicons. For example the *Pseudomonadales* were 16.5 % of the amplicon from the Soft extraction of Exp2A soil sample and the *Enterobacteriales* were 10.7 % of the amplicon in the Soft extraction of RACentre soil. *Pseudomonadales* and *Enterobacteriales* amplicons were over 100 times more common in Soft extracted than Mobio extracted DNA from Exp1B and Exp2A samples (Supplemental Table 5).



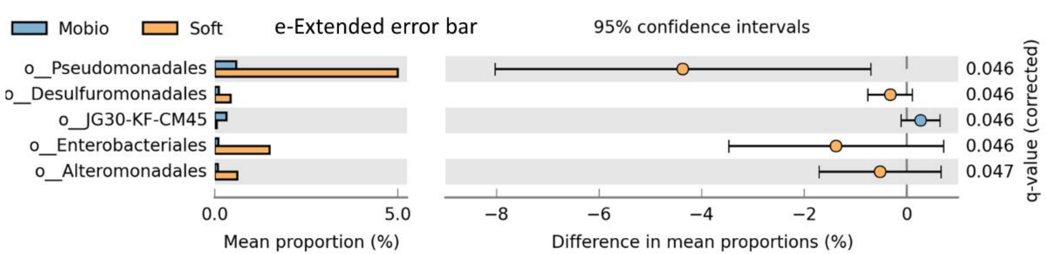


Figure 2.3. Extraction method affects abundance of different phylogenetic groups from the V4-V5 primer datasets for 11 soils and three compost samples. Bar graphs of the percentages of the orders a-*Pseudomonadales*, b-*Enterobacteriales*, c-*Alternomonadales*, d-*Desulfuromonadales* in soils and compost. e-STAMP Extended error bar plots at the order level for group analysis of soils (Soft versus Mobio extraction). Only results where differences were at least 4 fold and in at least one of the samples the order was more than 1% of sequences are shown.

Differential extraction of *Pseudomonadales* and *Enterobacteriales* was confirmed with a range of amplicon sequencing methods. All of the primer pairs used found preferential extraction of the *Enterobacteriales* with the Soft extraction method, but the genera amplified changed with primer and PCR conditions. V4V5 and V6V8 Enterobacterial amplicons were overwhelmingly unclassified *Enterobacteriaceae*. V4V6 Enterobacterial amplicons sequenced by EL1 were often *Pantoea* while the V4V6 amplicons sequenced by EL2 were often *Yersinia*. This suggests that Soft extraction preferentially extracts a broad range of *Enterobacteriales.* Similar Pseudomonal genera were also sequenced from amplicons generated from all primers except the V4V6 primers (Supplemental Tables 5). An example of the differences in results with different primers is shown in figure 2.4 for Mobio and Soft extracted DNA for sample Exp1B.

Other bacterial groups were under-extracted by the Mobio method although results were only found for one or two samples with a single set of primers. V4V6 amplicon sequencing found that amplicons from the *Bacillales* order and the genus *Stenotrophomonas* (a gamma proteobacterial group in the *Xanthomonadales* order), were overwhelmingly more common in Soft extracted DNA from two soils (Figure 2.5, Supplemental Table 6,7) although the *Bacillales* were overall more common in Mobio extracted DNA (Supplemental Figure 2). The largest fraction of the bacterial community preferentially extracted by either method in any one sample was the *Streptophytes* (from Chloroplasts) by Soft extraction in 80F top compost sample. These made up 35% of V4V5 amplicons from Soft extracted DNA, 261 fold more than Mobio DNA. (Figure 2.6, Supplemental Table 7).

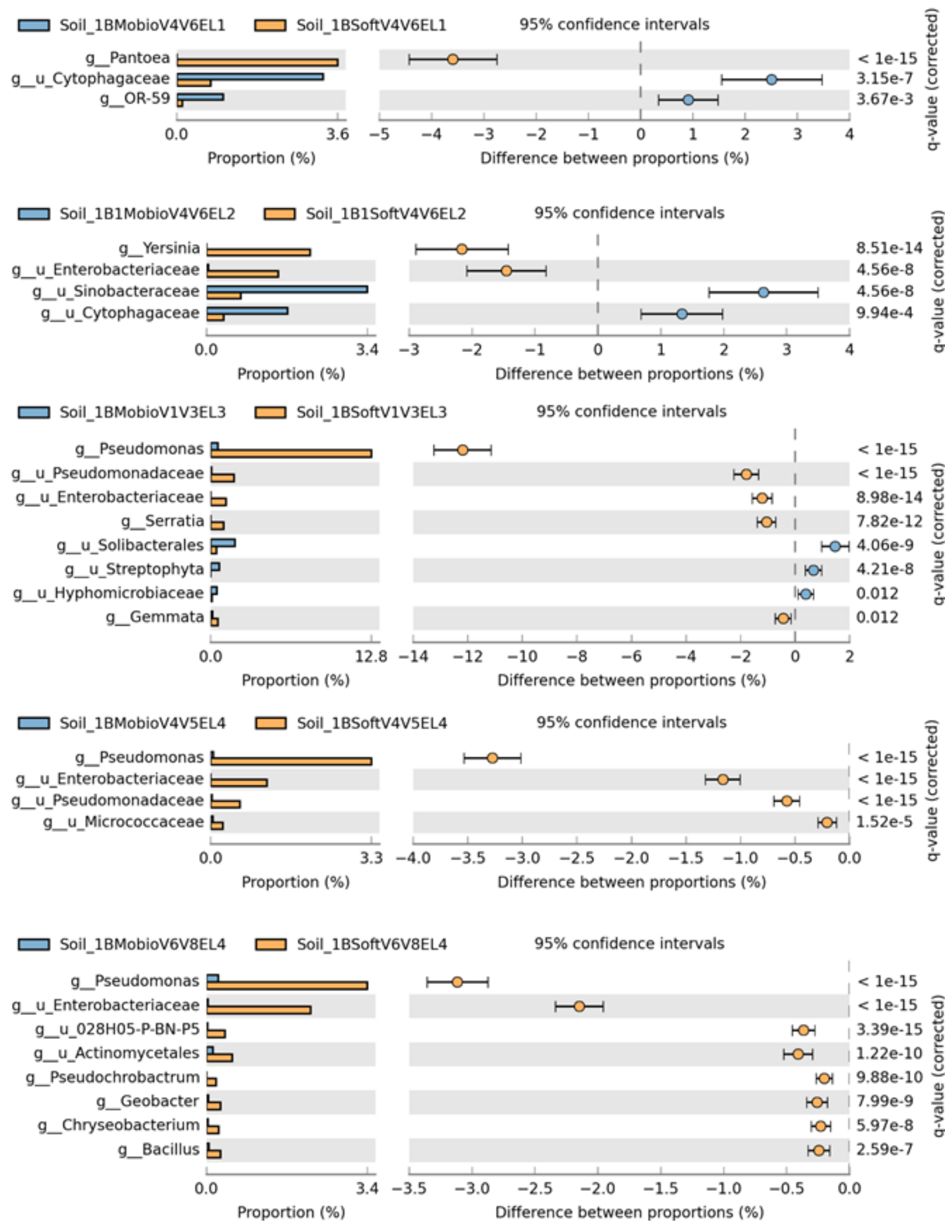


Figure 2.4. Comparisons of extraction method give different results with different primers and PCR conditions. Examples of STAMP Extended error bar plots at the genus level of soil Exp1B comparing extraction methods from four external sequencing laboratories and primer pairs. Results are from a-EL1/V4V6 primers, b-EL2/V4V6 primers, c-EL3/V1V3 primers, d-EL4/V4V5 primers, e EL4/V6V8 primers. Only results that are four-fold or more different between extraction method are shown. Multiple test correction with Storey FDR.

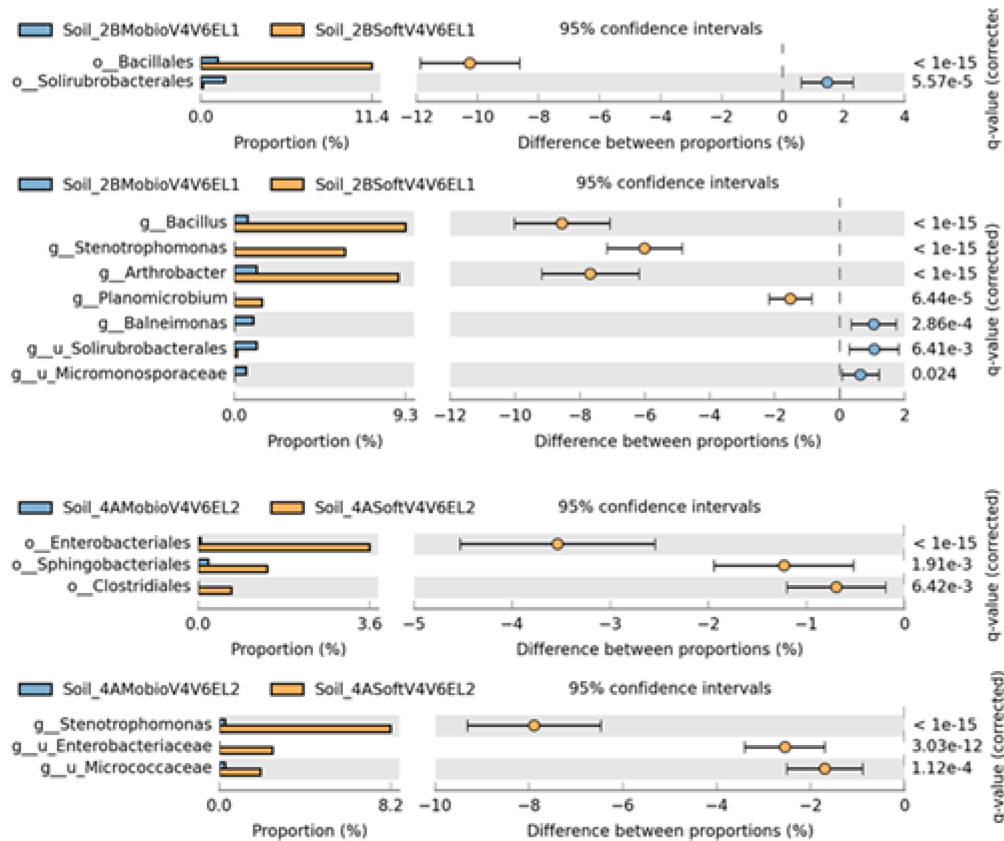


Figure 2.5. The *Bacillales* order and *Stenotrophomonas* genus were preferentially extracted by Soft extraction in two soils. STAMP Extended error bar plots at the genus and order levels of soils Exp2B and Exp4A. Results are from two external sequencing laboratories, using the V4-V6 primers (EL1, EL2), EL1 for Ex2B, EL2 for Exp4A. Sequences from EL2 and EL3 were trimmed to 450bp. Only results that are four-fold or more shown. Multiple test correction with Storey FDR.

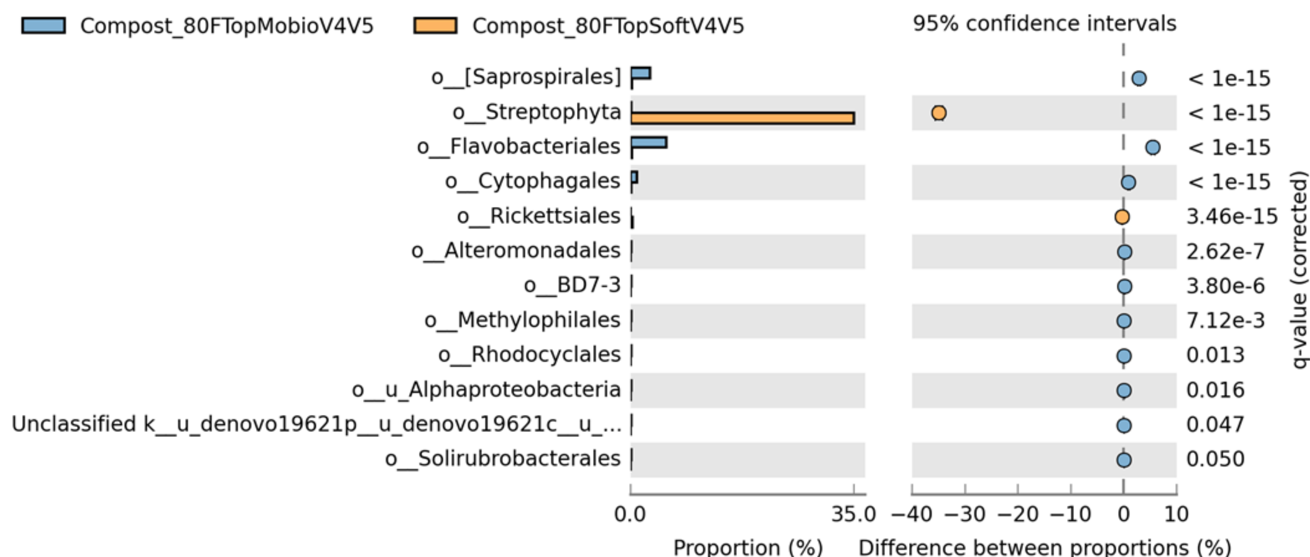


Figure 2.6. Preferential extraction of the *Streptophyta* order in compost. STAMP Extended error bar plots at the order level from the top layer of garden compost 80F. Results are from EL4, using the V4-V5 primers. Only results that are ten-fold or more shown. Multiple test correction with Storey FDR.

Extraction efficiency

We found thatseveral phylogenetic groups were a much larger fraction of amplicons from soft extracted DNA than Mobio extracted DNA. A possible reason for this is that the Soft extraction method did not extract more of these groups but rather a lot less of everything else. For this to be true, the Soft extraction would have to extract tens to hundreds times less DNA than the Mobio method from the whole bacterial population in some soils and similar amounts in others. We tested this by measuring how much DNA was extracted by each method in eleven of our samples. Most of the DNA extracted from soil is bacterial (Fierer et al., 2012) (Frisli, Haverkamp, Jakobsen, Stenseth, & Rudi, 2013) (Kerfahi et al., 2019). Final DNA yield is strongly affected by the cleaning steps that different extraction methods use. Raw DNA yield, i.e. the yield before any DNA is lost during cleaning is a reasonable estimate of how well the bacterial community is sampled.

Raw DNA yields were similar for both extraction methods (Table 2.1) and comparable to raw DNA yield from soils in other studies (Zhou et al., 1996)(Bürgmann, Pesaro, Widmer, & Zeyer, 2001)

(Carrigg, Rice, Kavanagh, Collins, & O'Flaherty, 2007) (Marstorp & Witter, 1999) (Taylor, Wilson, Mills, & Burns, 2002)(Yokoyama et al., 2017). Thus our STAMP results were not caused by the Soft method under extracting DNA.

DNA extraction efficiency is often measured by apparent alpha diversity of the sample, most simply measured as species richness (how many species are extracted) (Ínceošlu et al., 2010) (Terrat et al., 2012) (Hermans, Buckley, & Lear, 2018) (Zielińska et al., 2017). Efficiency can also be measured by how well the method extracts from "mock communities" of known mixtures of different bacteria (Hallmaier-Wacker, Lueert, Roos, & Knauf, 2018) (Hermans et al., 2018) (Morgan, Darling, & Eisen, 2010).

Both of these strategies are a good way of comparing extraction methods if what limits extraction is the diversity of cell walls and membranes of bacteria. Some of our results suggest that this is so. Bead beating is better than enzymatic extraction in lysing bacterial spores and we found that the Mobio kit extracted more of the *Actinomycetales* and *Bacillales* orders than Soft extraction. This was also the case for the JG30-KF-CM45 order and Alpha proteobacteria (Supplemental Table 2).

Others results from our study do not match this assumption. It is unlikely that Soft extraction worked better than Mobio extraction on the cell walls and membranes of such different groups as the gammaproteobacterial Pseudomonadales, Enterobacteriales, and Alteromonadales, and the deltaproteobacterial Desulfuromonadales. As well, if the differences in the ability of the Mobio and soft extraction methods to degrade cell walls and membranes were responsible for these differences, the ratio for each sample should be similar. Instead they varied by orders of magnitude. For example with the Ilumina V4V5 Pseudomonadales amplicons the Soft/Mobio ratio ranges from 118.3 in Exp2A to 0.6 in Exp3B (Supplemental Tables 5, 6, 7.).

DNA extraction from soil may be limited by the soil itself rather than the membranes and cell walls of soil bacteria. Our Soft lysis method is unusual in that it includes a washing step with Tris and EDTA. Washing soil before cell lysis has several effects. It reduces extraction of humic materials, adjusts pH, and removes contaminants all of which help lytic enzymes work. (LaMontagne, Michel, Holden, & Reddy, 2002) (Fortin, Beaumier, Lee, & Greer, 2004) (Stoeva et al., 2014). Washing may also remove relic DNA, the DNA remaining in soil after cell death. Removing relic DNA can reduce DNA yield and apparent alpha bacterial diversity (Carini et al., 2016)(Carini et al., 2020). This may partly explain the lower apparent alpha diversity and some lower DNA yields for soft extracted DNA.

In contrast, He *et al* found that soil washing increased DNA yield (He, Xu, & Hughes, 2005). He *et al* ascribe this to washing reducing DNA adsorbtion to clay and increasing soil dispersion. While DNA adsorbtion is unlikely to cause preferential extraction of bacterial orders, soil dispersion might.

Soil dispersion is the breaking up soil aggregates into their individual sand, silt and clay particles. Large aggregates (> 200 µm), are easily disrupted. Smaller aggregates are more resistant to physical disruption. Aggregates between 2 and 20 µm can be very resistant to physical disruption and are often made of bacteria surrounded by polysaccharides and clay particles (Tisdall & Oades, 1982). These bacterial microaggregates are bound together by surface charges and cation bridging between organic matter and clay particles and can be very stable, lasting even after bacteria die (Totsche et al., 2018). As they are resistant to mechanical degradation, the microaggregates surrounding bacteria may protect them from Mobio bead beating. Washing with EDTA could remove the stabilizing cations, releasing bacteria from the microaggregates. This could explain the preferential extraction of *Pseudomonadales* and *Enterobacteriale*s by soft extraction.

As seen in our results, different soils would release different numbers of bacteria. Not all soils have as many or as stable aggregates. Extremely sandy soils (Arenosols) are structureless and aggregation is depends on clay content and mineralogy (Totsche et al., 2018).

As also seen in our results, the released bacteria would also be from different groups than those of the bulk soil. The bacterial populations in small aggregates differ from the community in the whole soil (Ranjard & Poly, 2000) (Sessitsch, Weilharter, Gerzabek, Kirchmann, & Kandeler, 2001)(Mummey, Holben, Six, & Stahl, 2006). Three recent studies have used bead beating DNA extraction and next generation sequencing of 16S amplicons to characterize the bacterial communities of soil aggregates to phylogenetic levels at or below the order level (Bach, Williams, Hargreaves, Yang, & Hofmockel, 2018)(Fox et al., 2018)(Biesgen, Frindte, Maarastawi, & Knief, 2020). All found differences in statistically significant differences in bacterial populations of microaggregates compared to other aggregate sizes and the bulk soil.

Conclusion

This study compared two very different direct soil DNA extraction methods. The Mobio method is fast, convenient, and scalable and like most methods used in Microbial Ecology, uses bead beating to lyse bacteria. The Soft method is much slower and less convenient as it is uses enzymes and detergents for lysis. Results from both methods gave the same broad view of how soil bacterial populations differ from other environments and each other. We show that the soil extraction methods can under extract certain groups of bacteria and thus misrepresent how similar soils differ from each other but do not distort the overall picture of soil bacterial diversity.

We found that that primers and PCR conditions can affected 16S amplicon sequencing results (Figure 2.4 Supplemental Tables 5,6,7), as have many other earlier studies (Wu et al., 2010) (Ahn, Kim, Song, & Weon, 2012) (Rintala et al., 2017) (Hallmaier-Wacker et al., 2018)(Sze & Schloss, 2019) (Lerma

et al., 2020). The sequencing platform used (Pyrosequencing vs Illumina) may have also affected results; but probably less so than primer choice (Tremblay et al 2015). We also observed that results for different primers all confirmed that the soft method preferentially extracted Enterobacteria, even though they amplified different Enterobacterial genera. Based on this, we suggest that important results be confirmed with a range of PCR primers.

Extraction method biases may be less than primer/PCR biases but may be more important. Many different primers and PCR conditions are used in amplicon sequencing studies, which limit the potential for systemic bias. In contrast only a few extraction methods are used. Almost all of these method use bead beating methods similar to the Mobio method to extract DNA. Any biases common to bead beating extraction may cause systematic biases across the scientific literature. Furthermore, as direct sequencing of soil DNA is used, primer biases will become less important while extraction biases will remain.

We show that the Mobio method under-extracts several bacterial orders, particularly the *Pseudomonadales* and *Enterobacteriale*s, sometimes by orders of magnitude as compared to the Soft extraction method. As well as this, the degree of under-extraction varies from soil to soil. This suggests that these groups may be systematically under-represented in the literature. They also may be thought to be common or rare in soils where they are easier or more difficult to extract. When different studies or results from large scale studies are compared this could give a false picture of their distribution.

Under-extraction of the *Pseudomonadales* may mean that some of the secondary metabolites in soil are missed by metagenomics studies. Pseudomonads are one of most common bacterial producers of antibiotics (Bérdy, 2005) and secondary metabolite production by *Pseudomonadales* is important in soil suppressivity (Haas & Défago, 2005), the ability of the soil microbial community to prevent plant diseases.

Methods.

Sample sites.

Twelve soils were sampled. Seven were from Ottawa, Canada, two were acid forest podzols from the Mattawa plain (Ontario) and Vancouver, Canada and one each was from San Diego, California and North-East Kansas U.S.A. (Table 2.1). The three compost samples were also sampled. They were from the top and bottom of a compost pile receiving garden and household waste (80F and 80F Top) and the top of a compost pile receiving garden waste in an old peoples home (Glebe).

Sample collection and storage.

Soil samples were taken by driving a metal core 5 cm into the surface of the soil five to seven times in a meter square area. For forest soils organic litter was cleared to get to the surface of the mineral layer of soil, for the Mattawa soil the mineral soil was dug away until the Ae horizon could be seen. Compost samples were taken from smaller areas on the scale of centimetres. Samples were mixed by hand in a plastic bag and a two to three gram sub-sample was taken. Samples for the two extraction methods (Mobio/Soft extraction) were taken from this sub-sample. Samples were frozen within 3 hours at −20 °C > 12 hours before extraction.

DNA Extraction

DNA was extracted using the Mobio and soft extraction method once for 10 of the soils and all three compost samples. For two samples DNA was extracted either twice (Exp2A) and three times (Exp1B) using both methods.

Mobio DNA extraction.

This method is similar to the Earth Microbiomes version of the Mobio extraction method (as the manufacturer's instructions but with an initial 10 min incubation at 65 °C). We used bead beating tubes rather than 96 well plates that the Earth Microbiome does (http://www.earthmicrobiome.org/emp-standard-protocols/dna-extraction-protocol/).

0.25 g of sample was weighed into plastic tubes with glass beads from the Mobio Powersoil kit, the tubes were incubated in a water bath for 10 min at 65 °C and then shaken with the Mobio adaptor fitted onto a Scientific Industries Vortex-Genie 2 benchtop vortexer. All steps after this were as in the manual provided with the Mobio Powersoil kit.

As the enzymatic extraction method extracted 10-100 times more gamma proteobacterial DNA than the Mobio method from several soils, bead beating was intensified for Ottawa sample Exp1B and Exp2A. The bead beating time was doubled from 10 to 20 min (1BExp20mins, 2AExp20mins). Instead of using a benchtop vortex a more powerful Retsch GmbH - Mixer Mill MM 301 grinder was used for three 30 s periods at maximum speed and the tube was cooled on ice water between bead beating (1BExp30sec) DNA.

Soft DNA extraction.

This method uses enzymatic and SDS lysis. As several previous descriptions of this method were incomplete, it is described in full detail in the supplementary material (Supplemental Data1-Methods).

1.0- 2.0 g of sample is weighed into a 50 mL Falcon tube. 45 mL of washing buffer (50 mM EDTA/50 mM Tris/HCl pH 8.3) is added and tubes are then centrifuged at 2643 × g (3500 rpm on a

Sorvall Legend RT+ centrifuge) at 4 °C for 60 min. The supernatant is poured off, before storing at −20 °C for > 12 hours.

To begin extraction 2.5 mL of enzymatic extraction buffer (500 mM NaCl, 50 mM Tris, 50 mM EDTA, pH 8.0) is added to pellets for digestion with 12.5 mg of lysozyme. Tubes are incubated for 1 h. 140 µL of 20% SDS, is then added and mixed before adding 1 mg of proteinase K for a further hour of digestion.

Five millilitres of SDS extraction buffer (500 mM NaCl, 300 mM succinic acid, 10 mM ETDA, pH 5.7) is then added, followed by 700 µL of 20% SDS before a 45-min incubation at 65 °C. This buffer is added to lower the pH and raise the salinity, reducing the humic material that goes into solution at 65 °C.

Samples are centrifuged at 2643 × g for an hour at 4 °C. 8 mL of the supernatant is transferred to a 15 mL falcon tube. 1 mL of NaCl (5M) is added to each tube and mixed before 1 mL of 10% cetyl trimethyl ammonium bromide (CTAB) is added and incubated for 30 min at 65 °C. The tubes are then cooled to 4 °C on ice and 8 mL of chloroform is added. The tubes are gently mixed and centrifuged at 4000 × g in a centrifuge at 4 °C. The supernatant is removed. A half volume of 0.5 by weight PEG 6000 in water is added to the supernatant, mixed and left overnight at 4 °C in a 15 mL falcon tube to precipitate the DNA.

Samples are transferred to a 50 mL falcon tube, centrifuged at 4500 × g at 4 °C for 3 hours and the supernatant discarded. The pellets are washed with cold 70% ethanol before being dissolved in 400 µL of TE and transferred to a 1.5 mL Eppendorf tube. The samples are then cleaned with an equal volume of phenol/chloroform/isoamylalcohol (25:24:1) followed by one volume of chloroform/isoamylalcohol (24:1) at 4 °C.

5 µL of $CaCl_2$ (0.57 M) is then added for each 100 µL of the supernatant left over after previous steps and the solution is mixed and incubated at 65 °C for 30 min. The tubes are then centrifuged at 14 000 ×g in a bench top mini centrifuge for 10 min.

Finally the supernatant from the $CaCl_2$ cleaning is transferred to a new Eppendorf tube and DNA precipitated with 1/10 volume of sodium acetate (3 M, pH 5.7) and 1 volume of isopropanol at room temperature for 10 min. The tubes are then centrifuged at 14,000 × g with a bench top mini centrifuge and the supernatant removed. Pellets are then washed with cold 70% ethanol and dissolved in 50 µL of TE.

Measuring DNA concentration

Final DNA concentrations were measured using the PicoGreen double stranded DNA dye and the Tecan infinite F200 fluorescence microplate reader with Fluorescence Top Reading and excitation 485 nm/ emission 535 nm. To measure DNA concentration, 1-2 µL of the final DNA sample was brought to 100 µL with TE and mixed with 100 µL of 1:200 dilution of PicoGreen dye in TE on a 96 well plate and read on the plate reader. Standards between 25-1000 ng/mL of λ DNA were used to generate a standard curve for quantification.

For all but four samples (Glebe compost and the soils Exp4A, Ellsworth, Diamond) the DNA concentration was also measured immediately after the final lysis step (Mobio bead beating or the Soft hot SDS pH 5.7 step). To prevent interference, serial dilutions, which gave the equivalent of 0.02-0.10 µL of lysate for were used. Standards were between 0.025-25 ng/mL.

Sequencing overview.

Samples were sent to four different external laboratories (EL) for pyrosequencing or Illumina sequencing. Pyrosequencing was done at EL1 and EL2 for the V4/V6 regions and EL3 for the V1/V3 regions. Illumina sequencing for the V4/V5 and V6/V9 regions was done at EL4. A summary of sequencing is shown in table 2.1.

Pyrosequencing results (V4/V6 EL1) from Soft extracted DNA from beach sand, faeces and street dust samples are also included in Figure 2.2. These samples were included to show the relative effect of the two extraction methods and primers on Principal Coordinate Analysis compared to the differences between completely different environments.

Pyrosequencing

Three versions of pyrosequencing were carried out, two with primers for the V4 to V6 variable regions of the 16S ribosomal subunit, a third with primers for the V1 to V3 variable region.

V4-V6 pyrosequencing (EL1 and EL2).

25-50 ng of each sample DNA was used as a template for three independent PCR reactions in 25 µL PCR reaction. The 16S Eubacterial primers 530F (GTG CCA GCM GCN GCG G) and 1100R (GGG TTN CGN TCG TTG) were at 0.5 µM concentrations. The reaction cocktail also included 10x New England Buffer, 0.16 mM Bovine Serine Albumen (New England), 0.5 mM d NTP and 1.25 U of New England Taq.

Cycling steps were 95 °C for 2 min followed by 25 cycles of 95 °C for 1 min; 60 °C for 45 s and 68 °C for 45 s and a final elongation step at 68° C for 5 min.

PCR products were either then sent to an external laboratory for bar coded PCR (EL2) or amplified in house (EL1).

If amplified in house, three independent reactions were again run for each PCR. PCR template was 1 µL of the first PCR, with reaction conditions as before but without Bovine Serine Albumen and for only ten cycles. Bar coded 530-F primers were used in the place of 530-F. PCR product concentration for each sample was measured with Pico Green and the samples combined in equal amounts for a single batch of pyrosequencing. These samples were sent for sequencing on an eighth of a pyrosequencing PCR plate to External Laboratory 1 (EL1).

If sent to external laboratory 2 (EL2) for bar coded PCR, cycling conditions were 94 °C for 3 min followed by 28 cycles of 94 °C for 30 s, 53 °C for 40 s and 72 °C for 1 min and a final elongation step of 72 °C for 5 min. The HotStarTaq Plus Master Mix Kit (Qiagen, Valencia, CA) was used in PCR reactions. PCR products from different studies were combined at this laboratory. Samples sent to EL2 , were processed in three different runs, each including an Exp1B sample.

V1-V3 pyrosequencing (EL3)

Mobio and Soft extracted DNA from 1BExp was sent to a third sequencing laboratory (EL3) for pyrosequencing the V1 to V3 region, with the 27-F (CAC ATG TGA CGA GTT TGA TCM TGG CTCA G) and 518-R primers (CAC ATG TGA CGA GTT TGA TCM TGG CTC AG).

Illumina Sequencing

Samples were sent to fourth external sequencing laboratory (EL4) for Illumina sequencing. There a single PCR amplification was carried out with either the 515-F (GTG YCA GCM GCC GCG GTA A) and 962-R (CCG YCA ATT YMT TTR AGT TT) amplicons (V4-V5 sequencing) or the B969-F (ACG CGH NRA ACC TTA CC) and BA1406-R (ACG GGC RGT GWG TRCA A) amplicons (V6-V8 sequencing) using labelled primers for Mi Seq paired end sequencing.

Sequence Processing.

All sequences were edited for quality and length using the Initial processing step of the Ribosomal Database Program (Cole et al., 2014). Pyrosequencing Q values were set at 25 for soil results from EL1 and EL3 or 20 for all other results. Illumina sequencing Q values were set at 27.

Sequences from EL1 included the full length between the 530-F and 1100-R primers and so both primers were used for processing. Sequences from EL2 and EL3 were shorter, so only the forward primer was used and minimum and maximum lengths were set at 450 and 600 bp. These sequences were then trimmed to the minimum 450 bp length using the SEED2 program (Větrovský, Baldrian, & Morais, 2018). When sequences from EL1 and EL2 were analysed together, all sequences were trimmed to 450 bp.

After this initial processing and trimming, all sequences were aligned with the PyNAST program using the default reference alignment in Qiime1. PyNAST rejects sequences that do not align with the reference alignment. These sequences were usually from non-ribosomal genes (Caporaso et al., 2010). The output of PyNAST was then aligned using the Ribosomal Database Program to check that all sequences aligned to the correct region of the 16S subunit. Sequences were then chimera cleaned with UCHIME in the Ribosomal Database (Edgar, Haas, Clemente, Quince, & Knight, 2011).

Further processing was done with Qiime1. Operational taxonomic units were picked de novo with the pick_otus.py (Edgar, 2010), their representative sequences with pick_rep_set.py and their taxonomy assigned with assign_taxonomy.py (DeSantis et al., 2006). All three of these scripts were run with default values. For the next step align_seqs.py, alignment was with MUSCLE (Edgar, 2004). After filtering (filter_alignment.py) sequences were treed with fasttree (Price, Dehal, & Arkin, 2010) and an OTU table made (make_otu_table.py).

Alpha and Beta diversity analysis with Qiime.

Rarefaction curves were made with the alpha_rarefaction.py script. Alpha rarefaction was compared with using the multiple_rarefactions.py, alpha_diversity.py, collate_alpha.py and compare_alphadiverisity.py scripts. Beta diversity was measured using the beta_diversity_through_plots.py script to run unweighted and weighted UniFrac principal coordinate analysis (Lozupone & Knight, 2005). EMPeror and jackknifed UPGMA clustering was used to present the results of this work (Vázquez-Baeza, Pirrung, Gonzalez, & Knight, 2013). To test if the differences that were seen on the PCoA plots and UPGMA trees was statistically significant, Anosim analysis was run using the script compare_categories.py .

16S library comparison with Library compare/STAMP

The 16S sequences that were extracted from using each method for each sample were compared pairwise with the Library Compare function of the Ribosomal Database (Cole et al., 2014). Library Compare compares two bacterial communities across the phylum to genus levels of phylogeny however, the significance of each difference is not corrected for multiple tests (if enough phylogenetic levels are tested some will randomly be significant) or sample effect (differences may be statistically significant but small).

STAMP (Statistical analysis of taxonomic and functional profiles) is a program that only compares bacterial populations at a single phylogenetic level at a time, but it corrects results for multiple tests and sample effect (Parks et al., 2014). STAMP was used to confirm and present the results of Library Compare.

To make STAMP files the Qiime1 representative sequence file was assigned taxonomy with uclust. The assigned taxonomy file was then reformatted with a python script. This script fills in missing column entries where there is no taxonomic information (SupplementalDataMethods-2). This

reformatting is necessary for further processing of the data. The modified taxonomy file and a list of operation taxonomic units (from the pick_otu.py script) was used to make a biom file (make_otu_table.py). The biom file was then converted to json format and used in STAMP to create a STAMP (.spf) file. The species and OTU columns were removed in the LibreOffice Calc spreadsheet.

This spf file was then checked with the check hierarchy script available at http://kiwi.cs.dal.ca/Software/STAMP, and corrected in LibreOffice Calc before being uploaded into STAMP with a mapping file.

Sequence results from each laboratory were compared using two group analysis (Mobio vs Soft extraction) and laboratory results from each sample using different extraction method were compared using two sample comparison. Comparisons were carried out with the recommended settings, for two group analysis Welch's t-test/Welch's inverted/Storey's FDR, for two sample analysis Fisher's exact test/DP: Newcombe-Wilson/Storey's FDR. Filtres were set with a p value of 0.05 and to identify groups where differences between extraction methods were not only statistically significant but large. For two group analysis this was with an effect size for rations of 4 (differences less than 4 × would not be reported) for two sample analysis the effect size for ratios set at 10. A maximum value screen of 240 was set for Illumina data as they had many more sequences (at least one samples of the must have at least 240 sequences from the group).

Accession Numbers

Raw sequence data for this paper can be found in the NCBI SRA database as BioProject PRJNA482847.

Acknowledgements.

The broad sampling that this study used means that we owe thanks to many who helped us. The beach sample from Cornwallis Island was collected by Debbie Iqaluk and Alexandre Poulain; the sampling was made possible thanks to an NSERC discovery grant awarded to Alexandre Poulain, a Research license from the Nunavut Research Institute (license 02-109-11R), the logistical support of PCSP and the consent of the Resolute Bay Hunter's and Trapper's

Association. Dr. Rashid Nazir provided the Faisalabad Clock tower street dust sample, Panos Argyropoulos the Vancouver forest soil sample, Keavin M. Stanford-Finnerty the beach sample from Prince Edward Island. An employee of the Glebe Centre showed us a compost heap that was in use in a long-term care facility. Douglas Millar set up collaboration with Kara-Lee Golota, allowing us to take dung samples at her farm the Gathering Place Sanctuary. Matt Dextraze was supported by an Ontario Graduate Scholarship. Texture analysis of samples was done with the help of Jean Bjornson of the Geography Dept of the University of Ottawa. Fred Meyer provided the 80F and 80F Top compost samples, was interested in the work, but did not live to see the results.

References

Ahn, J. H., Kim, B. Y., Song, J., & Weon, H. Y. (2012). Effects of PCR cycle number and DNA polymerase type on the 16S rRNA gene pyrosequencing analysis of bacterial communities. *Journal of Microbiology*, *50*(6), 1071–1074. http://doi.org/10.1007/s12275-012-2642-z

Bach, E. M., Williams, R. J., Hargreaves, S. K., Yang, F., & Hofmockel, K. S. (2018). Greatest soil microbial diversity found in micro-habitats. *Soil Biology and Biochemistry*, *118*(January), 217–226. http://doi.org/10.1016/j.soilbio.2017.12.018

Bérdy, J. (2005). Bioactive microbial metabolites. *The Journal of Antibiotics*, *58*(1), 1–26. http://doi.org/10.1038/ja.2005.1

Biesgen, D., Frindte, K., Maarastawi, S., & Knief, C. (2020). Geoderma Clay content modulates differences in bacterial community structure in soil aggregates of different size. *Geoderma*, *376*(June), 114544. http://doi.org/10.1016/j.geoderma.2020.114544

Bissett, A., Fitzgerald, A., Meintjes, T., Mele, P. M., Reith, F., Dennis, P. G., … Ragan, M. A. (2016). Introducing BASE : the Biomes of Australian Soil Environments soil microbial diversity database. *GigaScience*, *5*(21). http://doi.org/10.1186/s13742-016-0126-5

Bremen, S. U., Miller, D. N., Bryant, J. E., Madsen, E. L., & Ghiorse, W. C. (1999). Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples Evaluation and Optimization of DNA Extraction and Purification Procedures for Soil and Sediment Samples. *Applied and Environmental Microbiology*, *65*(11), 4715–4724.

Bürgmann, H., Pesaro, M., Widmer, F., & Zeyer, J. (2001). A strategy for optimizing quality and quantity of DNA extracted from soil. *Journal of Microbiological Methods*, *45*(1), 7–20. http://doi.org/10.1016/S0167-7012(01)00213-5

Caporaso, J. G., Bittinger, K., Bushman, F. D., Desantis, T. Z., Andersen, G. L., & Knight, R. (2010). PyNAST: A flexible tool for aligning sequences to a template alignment. *Bioinformatics*, *26*(2), 266–267. http://doi.org/10.1093/bioinformatics/btp636

Carini, P., Delgado-baquerizo, M., Hinckley, E. S., Brewer, T. E., Rue, G., Vanderburgh, C., … Fierer, N. (2020). Effects of Spatial Variability and Relic DNA Removal on the Detection of Temporal Dynamics in Soil Microbial. *MBio*, *11*(1), e02776-19.

Carini, P., Marsden, P. J., Leff, J. W., Morgan, E. E., Strickland, M. S., & Fierer, N. (2016). Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nature Microbiology*, *16242*(December). http://doi.org/10.1038/nmicrobiol.2016.242

Carrigg, C., Rice, O., Kavanagh, S., Collins, G., & O'Flaherty, V. (2007). DNA extraction method affects microbial community profiles from soils and sediment. *Applied Microbiology and Biotechnology*, *77*(4), 955–964. http://doi.org/10.1007/s00253-007-1219-y

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., McGarrell, D. M., Sun, Y., … Tiedje, J. M. (2014). Ribosomal Database Project: Data and tools for high throughput rRNA analysis. *Nucleic Acids Research*, *42*(D1), 633–642. http://doi.org/10.1093/nar/gkt1244

Cullen, D. W., & Hirsch, P. R. (1998). Simple and rapid method for direct extraction of microbial DNA from soil for PCR. *Soil Biology and Biochemistry*, *30*(8–9), 983–993. http://doi.org/10.1016/S0038-0717(98)00001-7

Delmont, T. O., Robe, P., Cecillon, S., Clark, I. M., Constancias, F., Simonet, P., … Vogel, T. M. (2011). Accessing the soil metagenome for studies of microbial diversity. *Applied and Environmental Microbiology*, *77*(4), 1315–1324. http://doi.org/10.1128/AEM.01526-10

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., … Andersen, G. L. (2006). Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Applied and Environmental Microbiology*, *72*(7), 5069–5072. http://doi.org/10.1128/AEM.03006-05

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. http://doi.org/10.1093/nar/gkh340

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, *26*(19), 2460–2461. http://doi.org/10.1093/bioinformatics/btq461

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C., & Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics*, *27*(16), 2194–2200. http://doi.org/10.1093/bioinformatics/btr381

Fierer, N., & Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A*, *103*(3), 626–31. http://doi.org/10.1073/pnas.0507535103

Fierer, N., Leff, J. W., Adams, B. J., Nielsen, U. N., Bates, S. T., Lauber, C. L., … Caporaso, J. G. (2012). Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. *Proceedings of the National Academy of Sciences*, *109*(52), 21390–21395. http://doi.org/10.1073/pnas.1215210110

Fortin, N., Beaumier, D., Lee, K., & Greer, C. W. (2004). Soil washing improves the recovery of total community DNA from polluted and high organic content sediments. *Journal of Microbiological Methods*, *56*(2), 181–191. http://doi.org/10.1016/j.mimet.2003.10.006

Fox, A., Ikoyi, I., Torres-sallan, G., Lanigan, G., Schmalenberger, A., Wakelin, S., & Creamer, R. (2018). The in fl uence of aggregate size fraction and horizon position on microbial

community composition. *Applied Soil Ecology*, *127*(October 2017), 19–29. http://doi.org/10.1016/j.apsoil.2018.02.023

Frisli, T., Haverkamp, T. H. A., Jakobsen, K. S., Stenseth, N. C., & Rudi, K. (2013). Estimation of metagenome size and structure in an experimental soil microbiota from low coverage next-generation sequence data. *Journal of Applied Microbiology*, *114*(1), 141–151. http://doi.org/10.1111/jam.12035

Gabor, E. M., Alkema, W. B. L., & Janssen, D. B. (2004). Quantifying the accessibility of the metagenome by random expression cloning techniques. *Environmental Microbiology*, *6*(9), 879–886. http://doi.org/10.1111/j.1462-2920.2004.00640.x

Goethem, M. W. Van, Osborn, A. R., Bowen, B. P., Andeer, P. F., Swenson, T. L., Clum, A., … Northen, T. R. (2021). Long-read metagenomics of soil communities reveals phylum-specific secondary metabolite dynamics. *Commun Biol*, *18*(4), 1302. http://doi.org/10.1038/s42003-021-02809-4

Guerra, B. C. A., Bardgett, R. D., Caon, L., Crowther, T. W., Montanarella, L., Navarro, L. M., … Eisenhauer, N. (2021). Tracking, targeting, and conserving soil biodiversity. *Science*, *37*(6526), 239–242.

Haas, D., & Défago, G. (2005). Biological Control of Soil-Borne Pathogens by Fluorescent Pseudomonads. *Nature Reviews Microbiology*, (March). http://doi.org/10.1038/nrmicro1129

Hallmaier-Wacker, L. K., Lueert, S., Roos, C., & Knauf, S. (2018). The impact of storage buffer, DNA extraction method, and polymerase on microbial analysis. *Scientific Reports*, *8*(1), 1–9. http://doi.org/10.1038/s41598-018-24573-y

Haro-Moreno, J. M., López-Pérez, M., & Rodríguez-Valera, F. (2020). Long read metagenomics, the next step? *BioRxiv*, 2020.11.11.378109. http://doi.org/10.1101/2020.11.11.378109

He, J., Xu, Z., & Hughes, J. (2005). Pre-lysis washing improves DNA extraction from a forest soil. *Soil Biology and Biochemistry*, *37*(12), 2337–2341. http://doi.org/10.1016/j.soilbio.2005.04.016

Hermans, S. M., Buckley, H. L., & Lear, G. (2018). Optimal extraction methods for the simultaneous analysis of DNA from diverse organisms and sample types. *Molecular Ecology Resources*, *18*(3), 557–569. http://doi.org/10.1111/1755-0998.12762

Hill, P., Krištůfek, V., Dijkhuizen, L., Boddy, C., Kroetsch, D., & Van Elsas, J. D. (2011). Land use intensity controls actinobacterial community structure. *Microbial Ecology*, *61*(2), 286–302. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20924760

Ínceošlu, Ö., Hoogwout, E. F., Hill, P., & Van Elsas, J. D. (2010). Effect of DNA extraction method on the apparent microbial diversity of soil. *Applied and Environmental Microbiology*, *76*(10), 3378–3382. http://doi.org/10.1128/AEM.02715-09

Kerfahi, D., Tripathi, B. M., Dong, K., Kim, M., Kim, H., Slik, J. W. F., … Adams, J. M. (2019). From

the High Arctic to the Equator : Do Soil Metagenomes Differ According to Our Expectations ? *Microbial Ecology*, *77*(1), 168–185.

LaMontagne, M. G., Michel, F. C., Holden, P. A., & Reddy, C. A. (2002). Evaluation of extraction and purification methods for obtaining PCR-amplifiable DNA from compost for microbial community analysis. *Journal of Microbiological Methods*, *49*(3), 255–264. http://doi.org/10.1016/S0167-7012(01)00377-3

Lerma, A. S., Carrasco, V. P., Marañón, M. S., González, M. O., Martín, V. S., Gijón, J., … Soriano, M. (2020). Influence of 16S rRNA target region on the outcome of microbiome studies in soil and saliva samples. *Scientific Reports*, 1–13. http://doi.org/10.1038/s41598-020-70141-8

Lozupone, C., & Knight, R. (2005). UniFrac : a New Phylogenetic Method for Comparing Microbial Communities UniFrac : a New Phylogenetic Method for Comparing Microbial Communities. *Applied and Environmental Microbiology*, *71*(12), 8228–8235. http://doi.org/10.1128/AEM.71.12.8228

Mantri, S. S., Negri, T., Sales-ortells, H., Angelov, A., Peter, S., Neidhardt, H., … Ziemert, N. (2021). Metagenomic Sequencing of Multiple Soil Horizons and Sites in Close Vicinity Revealed Novel Secondary Metabolite Diversity. *MSystems*, *6*(5), e01018-21.

Marotz, C., Amir, A., Humphrey, G., Gogul, G., & Knight, R. (2017). DNA extraction for streamlined metagenomics of diverse environmental samples. *Bio Techniques*, *293*(June), 290–293. http://doi.org/10.2144/000114559

Marstorp, H., & Witter, E. (1999). Extractable dsDNA and product formation as measures of microbial growth in soil upon substrate addition. *Soil Biology and Biochemistry*, *31*(10), 1443–1453. http://doi.org/10.1016/S0038-0717(99)00065-6

Morgan, J. L., Darling, A. E., & Eisen, J. A. (2010). Metagenomic Sequencing of an In Vitro-Simulated Microbial Community. *Plos One*, *5*(4), 1–10. http://doi.org/10.1371/journal.pone.0010209

Mummey, D., Holben, W., Six, J., & Stahl, P. (2006). Spatial Stratification of Soil Bacterial Populations in Aggregates of Diverse Soils. *Microbial Ecology*, *51*, 404–411. http://doi.org/10.1007/s00248-006-9020-5

Nováková, A., Elhottová, D., Krištůfek, V., Lukešová, A., Hill, P., Kováč, Ľ., … Ľuptáčik, P. (2005). Feeding sources of invertebrates in theArdovská Cave and Domica Cave systems – preliminary results. *Contributions to Soil Zoology in Central Europe I*, 107–112.

Orgiazzi, A., Ballabio, C., Panagos, P., Jones, A., & Fernandez-Ugalde, O. (2018). LUCAS Soil , the largest expandable soil dataset for Europe : a review. *European Journal of Soil Science*, (January), 140–153. http://doi.org/10.1111/ejss.12499

Parks, D. H., Tyson, G. W., Hugenholtz, P., & Beiko, R. G. (2014). STAMP: Statistical analysis of taxonomic and functional profiles. *Bioinformatics*, *30*(21), 3123–3124. http://doi.org/10.1093/bioinformatics/btu494

Peña-Venegas, C. P., Cardona, G. I., Arguelles, J. H., & Arcos, A. L. (2007). Micorrizas Arbusculares del Sur de la Amazonia Colombiana y su Relación con Algunos Factores Fisicoquímicos y Biológicos del Suelo. *Instituto Amazónico de Investigaciones Científicas Sinchi*, *37 (3)*(373), 327–326. http://doi.org/10.1590/S0044-59672007000300003

Price, M. N., Dehal, P. S., & Arkin, A. P. (2010). FastTree 2 - Approximately maximum-likelihood trees for large alignments. *PLoS ONE*, *5*(3). http://doi.org/10.1371/journal.pone.0009490

Ranjard, L., & Poly, F. (2000). Heterogeneous Cell Density and Genetic Structure of Bacterial Pools Associated with Various Soil Microenvironments as Determined by Enumeration and DNA Fingerprinting Approach (RIS... *Microbial Ecology*, (June 2000), 263–272. http://doi.org/10.1007/s002480000032

Rintala, A., Pietilä, S., Munukka, E., Eerola, E., Pursiheimo, J. P., Laiho, A., … Huovinen, P. (2017). Gut microbiota analysis results are highly dependent on the 16s rRNA gene target region, whereas the impact of DNA extraction is minor. *Journal of Biomolecular Techniques*, *28*(1), 19–30. http://doi.org/10.7171/jbt.17-2801-003

Roldan, F., Junca, H., & Arbeli, Z. (2019). Effect of the extraction and purification of soil DNA and pooling of PCR amplification products on the description of bacterial and archaeal communities. *Journal of Applied Microbiology*, *126*, 1464–1467. http://doi.org/10.1111/jam.14231

Sessitsch, A., Weilharter, A., Gerzabek, M. H., Kirchmann, H., & Kandeler, E. (2001). Microbial Population Structures in Soil Particle Size Fractions of a Long-Term Fertilizer Field Experiment. *Applied and Environmental Microbiology*, *67*(9), 4215–4224. http://doi.org/10.1128/AEM.67.9.4215

Soliman, T., Yang, S.-Y., Yamazaki, T., & Jenke-Kodama, H. (2017). Profiling soil microbial communities with next-generation sequencing: the influence of DNA kit selection and technician technical expertise. *PeerJ*, *5*, e4178. http://doi.org/10.7717/peerj.4178

Stoeva, M. K., Aris-Brosou, S., Chételat, J., Hintelmann, H., Pelletier, P., & Poulain, A. J. (2014). Microbial community structure in lake and wetland sediments from a high arctic polar desert revealed by targeted transcriptomics. *PLoS ONE*, *9*(3), 1–12. http://doi.org/10.1371/journal.pone.0089531

Sze, M. A., & Schloss, P. D. (2019). The Impact of DNA Polymerase and Number of Rounds of Amplification in PCR on 16S rRNA Gene Sequence Data. *MSphere*, (May/June), 1–13.

Taylor, J. P., Wilson, B., Mills, M. S., & Burns, R. G. (2002). Comparison of microbial numbers and enzymatic activities in surface soils and subsoils using various techniques. *Soil Biology and Biochemistry*, *34*(3), 387–401. http://doi.org/10.1016/S0038-0717(01)00199-7

Terrat, S., Christen, R., Dequiedt, S., Leli??vre, M., Nowak, V., Regnier, T., … Ranjard, L. (2012). Molecular biomass and MetaTaxogenomic assessment of soil microbial communities as influenced by soil DNA extraction procedure. *Microbial Biotechnology*, *5*(1), 135–141. http://doi.org/10.1111/j.1751-7915.2011.00307.x

Terrat, S., Plassart, P., Bourgeois, E., Ferreira, S., Dequiedt, S., Adele-Dit-De-Renseville, N., … Ranjard, L. (2015). Meta-barcoded evaluation of the ISO standard 11063 DNA extraction procedure to characterize soil bacterial and fungal community diversity and composition. *Microbial Biotechnology*, *8*(1), 131–142. http://doi.org/10.1111/1751-7915.12162

Thompson, L. R., Sanders, J. G., McDonald, D., Amir, A., Ladau, J., Locey, K. J., … Zhao, H. (2017). A communal catalogue reveals Earth's multiscale microbial diversity. *Nature*. http://doi.org/10.1038/nature24621

Tisdall, J. M., & Oades, J. M. (1982). Organic matter and water-stable aggregates in soils. *Journal of Soil Science*, *33*, 141–163.

Torsvik, V., Goksøyr, J., Daae, F. L., Torsvik, V., Goksyr, J., & Daae, F. L. (1990). High diversity in DNA of soil bacteria . High Diversity in DNA of Soil Bacteria. *Applied and Environmental Microbiology*, *56*(3), 782–787.

Totsche, K. U., Amelung, W., Gerzabek, M. H., Guggenberger, G., Klumpp, E., Knief, C., … Ko, I. (2018). Microaggregates in soils. *Journal of Plant Nutrition and Soil Science*, *181*, 104–136. http://doi.org/10.1002/jpln.201600451

.Tremblay J, Singh K, Fern A, Kirton ES, He S, Woyke T, Lee J, Chen F, Dangl JL, Tringe SG (2015) Primer and platform effects on 16S rRNA tag sequencing. Front. Microbiol. 6:771. doi: 10.3389/fmicb.2015.00771

Vázquez-Baeza, Y., Pirrung, M., Gonzalez, A., & Knight, R. (2013). EMPeror: a tool for visualizing high-throughput microbial community data. *GigaScience*, *2*(1), 16. http://doi.org/10.1186/2047-217X-2-16

Větrovský, T., Baldrian, P., & Morais, D. (2018). SEED 2: a user-friendly platform for amplicon high-throughput sequencing data analyses. *Bioinformatics*, (February), 0–0. http://doi.org/10.1093/bioinformatics/xxxxx

Waschulin, V., James, R., Newsham, K. K., Donadio, S., & Corre, C. (2022). Biosynthetic potential of uncultured Antarctic soil bacteria revealed through long-read metagenomic sequencing. *ISME Journal*, *16*(July), 101–111. http://doi.org/10.1038/s41396-021-01052-3

Wu, J. Y., Jiang, X. T., Jiang, Y. X., Lu, S. Y., Zou, F., & Zhou, H. W. (2010). Effects of polymerase, template dilution and cycle number on PCR based 16 S rRNA diversity analysis using the deep sequencing method. *BMC Microbiology*, *10*(1), 255. http://doi.org/10.1186/1471-2180-10-255

Yeates, C., Gillings, M. R., Davison, A. D., Altavilla, N., & Veal, D. A. (1998). Methods for microbial DNA extraction from soil for PCR amplification. *Biological Procedures Online*, *1*(1), 40–47. http://doi.org/10.1251/bpo6

Yokoyama, S., Yuri, K., Nomi, T., Komine, M., Nakamura, S. ichi, Hattori, H., & Rai, H. (2017). The high correlation between DNA and chloroform-labile N in various types of soil. *Applied Soil Ecology*, *117–118*(September 2016), 1–9. http://doi.org/10.1016/j.apsoil.2017.04.002

Zhou, J., Bruns, M. A., & Tiedje, J. M. (1996). DNA recovery from soils of diverse composition. *Applied and Environmental Microbiology*, *62*(2), 316–322. http://doi.org/D - NLM: PMC167800 EDAT- 1996/02/01 MHDA- 1996/02/01 00:01 CRDT- 1996/02/01 00:00 PST - ppublish

Zielińska, S., Radkowski, P., Blendowska, A., Ludwig-Gałęzowska, A., Łoś, J. M., & Łoś, M. (2017). The choice of the DNA extraction method may influence the outcome of the soil microbial community structure analysis. *MicrobiologyOpen*, *6*(4), 1–11. http://doi.org/10.1002/mbo3.453

Chapter 3. Investigation of how the Actinobacterial community structure changes across land use.

Previously published as: Patrick Hill, Václav Krištůfek, Lubbert Dijkhuizen, Christopher N Boddy, David Kroetsch, Jan Dirk van Elsas, Land Use Intensity Controls Actinobacterial Community Structure.
Microbial Ecology, Volume 61, Issue 2, February 2011, Pages 286–302

The most common antibiotic producer is the order Actinomycetales in the Phylum Actinobacteria. This study compared Actinobacterial distribution across three contrasting kinds of environments, soil street dust and animal affected soil. Contrasting soil samples of were taken across several continents; street dust samples were taken across Western Europe. The “Animal affected samples” were a mix of samples from worm casts, bee hive waste, guano, isopoda dung and compost. DNA was extracted using the “Soft method” described in Chapter 2 and characterized with 16S fingerprinting.

As in Chapter 2, several 16S primer pairs were used to amplify environmental DNA. In chapter 2 Next Generation Sequencing was used to sequence amplicons from a few samples (12 soils and 3 composts) and a subgroup of these samples were sequenced with a range of 16S primers. In chapter 3 two sets of Actinobacterial specific primers were used to fingerprints a broader range of samples (22 soils, 12 street sediments, 12 animal affected environments) and Sanger sequencing was used to interpret the results of community fingerprinting. Finally, Sanger sequencing of bacterial amplicons was used for a subset of samples (6 soils and 6 streets sediments).

Conclusions

As previously shown, results varied with primer. However, there were common themes to the results from the three combinations of 16S primers that were used. Actinobacterial community structure was controlled by land use rather than geography. Actinomycetes were selected for in street dust.

Role of Ph.D. candidate in Chapter 3

The Ph.D. candidate carried out:

-All sampling with the exception of the Vermicompost and Siberian soil sample.

-All laboratory work.

-All data analysis.

-writing the first version of the manuscript.

Land Use Intensity Controls Actinobacterial Community Structure

Patrick Hill & Václav Krištůfek & Lubbert Dijkhuizen & Christopher Boddy & David Kroetsch & Jan Dirk van Elsas

Abstract

Actinobacteria are major producers of secondary metabolites; however, it is unclear how they are distributed in the environment. DNA was extracted from forest, pasture and cultivated soils, street sediments (dust and material in place), and sediments affected by animal activity (e.g. guano, vermicompost) and characterised with two actinobacterial and a bacterial-specific 16S rDNA primer set. Amplicons (140/156) generated with the two actinobacterial-specific and amplicons (471) generated with bacterial-specific primers were analysed. Amplicons from actinobacterial-specific primer were disproportionately actinomycetal from animal affected (soil) samples and street sediments and either verrucomicrobial (i.e. non-actinobacterial) and from a novel non-actinomycetal actinobacterial group for soils. Actinobacterial amplified ribosomal DNA restriction analysis and terminal restriction fragment length polymorphism fingerprints clustered by land use, with cultivated soils clustering apart from uncultivated soils. Actinobacterial amplicons generated with eubacterial primers were overwhelmingly from (116/126) street sediments; acidobacterial amplicons from soils (74/75). In two street samples, >90% of clones were actinomycetal. Actinomycetes are selected in terrestrial soils and sediments by cultivation, urbanisation and animal activity.

Introduction

Most known antibiotics from bacteria are produced by Actinobacteria. Within this phylum, members of the order Actinomycetes are the major producers of antibiotics and in some cases

anticancer agents and immunosuppressants. The identification of new Actinomycetes thus can play an important role in drug discovery. While Actinobacteria are generally considered as soil-dwelling bacteria, it remains unclear what determines their community structure. Understanding actinobacterial distribution in the environment is important in understanding their ecological role and for pharmaceutical bioprospecting.

The distribution of soil bacterial communities does not correspond to eukaryotic biomes such as rainforest, prairie or tundra [32]. There is no consensus on how terrestrial microbial populations are distributed. Several studies have found that soil bacterial community structure is more affected by the soil properties than the vegetation that the soil supports [3, 11]. Soil texture [10], pH [8] or parent material [43] have all been suggested as the controlling factors for soil bacterial community structure.

Within the bacterial community, bacterial subgroups can be controlled by factors other than those that influence the entire community structure. For example, the proteobacterial Psychrobacter and firmicute Exiguobacterium genera are more common in soil samples from cold climates [35]. Additionally, the community structure of soil ammonium oxidising bacteria is also controlled by climate [9]. Identifying factors which control actinobacterial community structure will thus play an important role in identifying sampling sites for bioprospecting.

Two culture-independent studies have found contrasting factors influencing actinobacterial community structure in soil. Wawrik et al. [42] compared actinobacterial-specific terminal restriction fragment length polymorphism (T-RFLP) patterns from Uzbekistan and New Jersey, USA, and found a strong biogeographic effect. Based on these results, he recommended that bioprospectors sample geographically separated sites. Lauber et al. [26], using pyrosequencing, determined that pH, rather than location, controlled both actinobacterial and total bacterial community structures. These results suggest that sampling should be directed by soil pH.

Molecular methods give a culture-independent view of the Actinobacteria and have shown that non-actinomycetal actinobacterial groups such as the Rubrobacteria and Acidimicrobia are as common as the Actinomycetes in soil [17, 20]. However, the results of molecular methods are highly dependent on primer choice and extraction method. A recent study suggests that there is little overlap in the 16S sequences generated using different primers and extraction methods and that these limitations cannot be overcome by more extensive sequencing[18]. Therefore, the use of several primer systems for both fingerprinting and sequencing may give a more reliable view of community composition. In this context, sequencing is used to interpret the results of community fingerprinting rather than to characterise community structure alone.

To investigate the controlling factor for actinobacterial community structure, we applied two actinobacterial primer systems to a broad range of environmental DNA samples and applied a eubacterial primer system to a subset of soil and street sediment DNA samples. The primer sets used were the system of Heuer et al. [15], which uses a forward actinobacterial-specific 243-F primer and a eubacterialspecific reverse primer (F-Act/R-Bact); that of Monciardini et al. [30], which replaces the reverse primer with the actinobacterial-specific A3 primer (F-Act/R-Act); and the eubacterial primer system of Marchesi et al. [28], which uses 63-F and 1387-R primers. Environmental DNA was obtained from sites selected to cover a broad range of sample diversity. Samples from pasture, forest and cultivated soils on contrasting parent material (Table 3.1) were included, as were samples of non-soil sediments associated with humans and animals: street sediments, earthworm casts, vermicompost, beehive waste, isopoda dung and guano (Table 3.2). Sampling sites were located in Colombia, Canada, the Czech Republic and Siberia. Together, these samples cover a wide range of parent materials, climates and land use in diverse geographical locations. The results from our community fingerprint analyses coupled with sequencing show that overall actinobacterial community structure appears to be controlled by land use and that street sediments are enriched with Actinomycetes.

Methods

Sampling for Characterisation with Actinobacterial-Specific Primers

Soils from contrasting parent materials were sampled from areas within 20–50 km of Cali, (Colombia), Ottawa, (Canada) and České Budějovice (Czech Republic; Table 3.1). A single soil was sampled from the Yenisei Valley, Siberia, Russia. A medieval church garden soil from the centre of České Budějovice (garden of paradise) was also sampled. Samples were taken by driving a 5-cm-deep ring into the surface of the mineral layer of soil. Most samples were composites of five cores taken randomly from a 1-m^2 area. Exceptions were the Yenisei site (Rusforest-Yenisei) and a site near České Budějovice (Czcultivated-ČeskeBudéjovice) where samples were composites of 20 cores taken from a 20×20-m area. As the object of sampling was to sample as many contrasting soils as possible, most sites were only sampled once. At four sites, two samples were taken about 50 m from each other (Colcultivated-Rozo 1 and 2, Colforest-Rozo 1 and 2, Colforest-Hatico 1 and 2, Czcultivated-Netolice 1 and 2).

Samples were also taken from a range of non-soil animal associated sediments. These sediments were: earthworm casts taken from a microcosm experiment [16], bat cave guano that was heavily colonised by isopoda (Slovakia-Guano) and isopoda dung (Slovakia-Isopodadung) from the Domica cave system in Slovakia [33], manure (Czech-Manure) and vermicompost (Czech-Vermicompost) from a composting system, and soil that had received dead bees (i.e. chitin enriched anthropogenic environment) from a hive that had been left at least 15 years on a site in the Czech Republic (Netolice-Hivewaste).

Street sediment samples were either dust from street surfaces where there might be continual transport of both sediment and bacteria or sediment taken from spaces between cobblestones or an abandoned tramline where the presence of debris (heterogeneous pieces of broken glass, screws, marbles and pottery fragments) suggested that the material had been in place for years. Street

sediments were sampled in five urban areas: the suburb Goring of Worthing (UK) and the centres of České Budějovice, the Czech Republic, Groningen, the Netherlands, Paris and Brussels (Table 3.2).

Samples for Characterisation with Eubacterial Primers

To confirm the results obtained with actinobacterial specific primers, two environments that gave contrasting results with actinobacterial-specific primers, soil and street sediments, were characterised with eubacterial primers. Soils were collected in either open cultivated fields (three) or under tree canopy (three). Street sediments were either street dust (three) or sediments that had collected between cobblestones or a pavement crack (three). Four soil and four street sediment DNA samples were used from the actinobacterial study. Four additional DNA extractions were performed: (1) on an acid forest soil from near České Budějovice (Czforest-Kolny) and (2) three samples from two new sampling areas: soil under woodland (Hunforest-Citadel) and sediment from a railway station pavement crack (Budapest-trainstationfissure) were sampled in Budapest, Hungary, and street dust was sampled in the centre of Faisalabad, Pakistan (Faisalabad-streetdust-clock) by a collaborator. This additional sampling was to ensure that earlier results were not an artefact of extraction or handling.

All samples with the exception of Rusforest-Yenisei were stored at 4°C for no more than 3 days before a 1.5-2.5-g subsample was mixed with 30 mL of 50 mM EDTA/50 mM Tris/HCl (pH 8.3) buffer, centrifuged at 6,000×g at 4°C for 30 min and the supernatant discarded before storage at −20°C. This washing step was carried out to homogenise extraction conditions and remove contaminants that might interfere with PCR. The Yenisei sample was stored for several months at −80°C before washing and DNA extraction.

Texture, pH and organic carbon content of samples were measured either in the soil analysis laboratories of the International Centre for Tropical Agriculture (CIAT) near Palmira, Colombia, the BC

ASCR, v. v. i. Institute of Soil Biology, in České Budějovice, the Czech Republic, or Wageningen Agricultural University, the Netherlands (Tables 3.1 and 3.2).

Table 3.1 Soil samples and their sequenced clones.

Area	Land use	Parent material	Identifier[a]	pH	% organic matter	% Clay	Season Sampled
Ottawa, Canada	Pine plantation	Glacial moraine	Cdnforest-Kemptville AC-10/AY-2 (A-1/V-2/AD-0)	5.8	2.8	9.9	Summer
	Pasture	Marine clay	Cdnpasture-Lowe	5.6	7.0	25.1	Summer
Cali, Colombia	Garden soil	Artificial soil	Colforest/garden-CIATBiotech[c] AC-4/AY-4 (A-1/V-2/AD-0)	6.7	13.2	30.4	Dry season
	Tobacco/mustard rotation	Cauca River floodplain	Colcultivated-Rozo1 AC-7/AY-7 (A-2/V-4/AD-0)	7.1	2.3	31.3	Dry season
			Colcultivated-Rozo2	6.6	3.2	44.6	Dry season
	Bamboo forest		Colforest-Rozo1[c,d] AC-16/AY-3 (A-0/V-6/AD-0)	6.7	6.1	40.7	Dry season
			Colforest-Rozo2	6.7	6.5	37.0	Dry season
	Primary forest		Colforest-Hatico1	7.7	7.0	25.0	Wet Season
			Colforest-Hatico2	6.8	8.2	36.0	Wet Season
	Bamboo forest		Colforest-CIAT	6.3	5.7	33.5	Wet Season
	Leucaena plantation		Colleucaena-CIAT[b]	7.6	5.8	53.2	Wet Season
	Pasture		Colpasture-airport	7.3	10.1	26.1	Wet Season
	Pasture		Colpasture-Palmira	6.3	12.4	34.7	Wet Season
	Leucaena plantation	Gabbro	Colleucaena-CristoRey (A-5/V-6/AD-0)	6.6	7.3	37.2	Wet Season
	Secondary forest	Inactive alluvial fan	Colforest-Panse (A-0/V-10/AD-1)	4.8	13.6	48.0	Dry season
	Improved pasture		Colpasture-Panse (A-1/V-1/AD-2)	4.3	1.1	18.3	Dry season
Č.Budějovice, Czech Republic	Church garden	Unknown	České.Budějovice–Paradise (A-5/V-3/AD-0)	7.5	2.8	3.0	Spring
	Maize	Vltava Floodplain	Czcultivated-Plana[c]	6.1	2.9	15.0	Summer
	Potato	Granite	Czcultivated-České.Budějovice[c] (A-0/V-3/AD-0)	5.8	1.9	11.0	Summer
	Pasture	Limestone	Czpasture-Krumlov[c] AC-9/AY-3 (A-1/V-6/AD-0)	5.9	6.9	19.0	Winter
	Native grassland	Limestone	Czpasture-Pahlava	7.2	5.0	20.0	Winter
	Winter wheat	Migmatite	Czcultivated-Netolice1 (A-1/V-9/AD-0)	6.0	2.09	19.0	Spring
	Winter wheat		Czcultivated-Netolice2	6.0	2.1	15.0	Spring
	Pine forest	Marine sediments	Czforest-Kolny[e]	3.0	26.1	8.0	Winter
Yenisei Valley, Siberia	Taiga forest	Yenisei Floodplain	Rusforest-Yenisei AC12/AY-0 (A-0/V-11/AD-1)	3.8	4.0	4.2	Summer
Buda, Hungary	Scrub forest	Limestone	Hung-forest-Citadel[e]	7.3	18.6	5.0	Winter

[a] Abbreviations for the different primer pairs for sequenced clones are: Monciardini et al. [30]—*AC* non-actinomycetal actinobacterial, *AY* actinomycetal; Heuer et al. [15]—*A* actinobacterial, *V* verrucomicrobial, *AD* AD-3 group

[b] Not used for forward T-RFLP

[c] Not used for reverse T-RFLP

[d] Not used for for ARDRA

[e] Only used for bacterial sequencing

Table 3.2 Insect, Earthworm and urban sediments and their sequenced clones.

Area	Sediment	Identifier[a]	pH	% Organic Matter	% Clay	Season sampled
Central Brussels, Belgium	Tramline, Rue de Russe	Brussels-tramline-Russe	8.1	2.2	2.0	Winter
	Cobblestones, Stockexchange	Brussels-cobblestones-Buers	7.4	4.0	2.0	Winter
	Cobblestones, Rue Boucher	Brussels-cobblestones-Boucher AC-0/AY-11	7.0	2.9	4.0	Winter
Central České.Budějovice, Czech Republic	Cobblestones, Česka Street	České.Budějovice-cobblestones-Česka AC-1/AY-14 (A-5/V-0/AD-0)	7.5	3.0	3.0	Spring
	Street dust Koh-i-noor factory	České.Budějovice-streetdust-Koh-i-Noor (A-6/V-2/AD-0)	7.3	3.3	2.0	Spring
Groningen, the Netherlands	Dust Platform 3, Train station	Groningen-streetdust-trainstation	7.3	0.7	2.0	Winter
	Cobblestones Donkerstraat	Groningencobblestones-Donkerstraat-	6.3	4.5	5.0	Winter
Left Bank, Paris, France	Cobblestones, Rive Gauche Bar	Pariscobblestones-RiveGauche	7.3	9.1	4.0	Summer
	Cobblestones, Café Procope	Pariscobblestones-café	7.0	10.0	6.0	Summer
Goring, Worthing, United Kingdom	Street sediment,Lloyds Bank, Goring	Worthingstreetdust-LloydsBank	7.2	10.6	16.0	Summer
	Streetsediment, Busstop, Strand	Worthingstreetdust-Busstop	6.0	21.4	13.0	Summer
Budapest, Hungary	Cobblestones Trainstation	Budapest-Trainstation-Fissure[c]	7.9	10.7	5.7	Winter
Faisalabad, Pakistan	Street sediment Clocktower	Streetdust-Clock-Faisalabad[c]	6.8	12.0	2.0	Dry season
Domica cave system, Slovakia	Bat Guano pile	Slovakia-Guano (A-9/V-0/AD-0)	3.4	42.7	n/d	n/a
	Isopoda dung from cave floor	Slovakia-Isopodadung	6.3	15.5	n/d	n/a
Vermicomposting system, Czech Republic	Raw manure	Czech-Manure[c] (A-6/V-7/AD-0)	7.2	33.5	n/d	n/a
	Vermicomposed manure	Czech-Vermicompost (A-2/V-2/AD-0)	7.2	32.1	n/d	n/a
Netolice, Czech Republic	Soil receiving dead bees >15 years	Netolice-Hivewaste AC-1/AY-23 (A-7/V-1/AD-0)	5.4	13.5	13.0	Spring
Earthworm microcosms Soil used was Colcultivated-Rozo, worms (*Martiodrillus heterostichon*, *Polypheretima elongata*) were from the Colforest-Rozo soil.	Gut contents *M. heterostichon*	Rozo-wormgut-Martiodrillus AC-8/AY-19 (A-8/V-1/AD-0)	Samples too small for analysis			Dry season
	Surface casts *M. heterostichon*	Rozo-wormcastMartio1[d]				Dry season
	Gallery casts *M. heterostichon*	Rozo-wormcastMartio2[d]				Dry season
	Surface casts *P. elongata*	Rozo-wormcastElongata1 (A-6/V-4/AD-0)				Dry season
	Surface casts *P. elongata*	Rozo-wormcastElongata2				Dry season
	Gallery casts *P. elongata*	Rozo-wormcastElongata3[b]				Dry season
	Gallery casts *P. elongata*	Rozo-wormcastElongata4[b]				Dry season

[a] Abbreviations for the different primer pairs for sequenced clones are: Monciardini et al. [30]—*AC* non-actinomycetal actinobacterial, *AY* actinomycetal; Heuer et al. [15]—*A* actinobacterial, *V* verrucomicrobial, *AD* AD-3 group

[b] Not used for reverse T-RFLP

[c] Not used for ARDRA

[d] Only used for bacterial sequencing

DNA Extraction from Soils and Sediments

DNA was extracted from soil using a modified version of Zhou and Bruns' [45] direct extraction method. Briefly, 2.5 mL of a 500 mM NaCl/50 mM Tris/50 mM EDTA buffer (pH 8.0) was added to soil pellets for digestion with lysozyme and proteinase K at 37°C. Lysozyme was first added to a final concentration of 5 mg/mL and samples were incubated for 1 h. Of 20% SDS, 140 µL was then added along with 1 mg of proteinase K for a further hour of digestion. Five millilitres of a 500 mM NaCl/300

mM succinic acid/10 mM ETDA (pH 5.7) buffer was then added, followed by 700 µL of 20% SDS before a 30-min incubation at 65°C. After centrifugation, the extract was CTAB cleaned before precipitation with PEG 8000. Pellets were resuspended in 240 µL TE in Eppendorf tubes before being cleaned with a 25:24:1 phenol/ chloroform/isoamyl alcohol mixture and subsequently with a 24:1 chloroform/isoamyl alcohol mixture. Remaining humic acids were precipitated by adding $CaCl_2 2H_2O$ to a final concentration of 0.4% (w/v) and incubation at 65°C for 1 h before centrifugation. Samples were then cleaned and precipitated with 3 M potassium acetate, washed with 70% ethanol and dissolved in 20 µL TE. DNA concentration was measured with a Hoefer DyNA Quant 200 Fluorometer. Final DNA yields were between 2.3 and 30.0 µg/g soil or sediment (air dry weight).

16S rDNA Primers

In the F-Act/R-Bact amplification system of Heuer et al. [15], a single forward actinobacterial-specific primer 226-243-F (5-GGATGAGCCCGCGGCCTA-3) is used in combination with the reverse bacterial primer R-1378 CGGTCTCTA CAAGGCCCGGGAACG). Monciardini et al. [30] combined the 226-243-F primer used by Heuer et al. [15] with an actinobacterial-specific reverse primer, i.e. (A3R) 14141430-R (5-CCAGCCCCACCTTCGAC-3). The bacterial primers of Marchesi et al. [28] consist of the forward 63-F CAGGCCTAACAC ATGCAAGTC and the reverse 1387-R GGGCGGWGTGTACAAGGC primers.

All three primer systems were used for cloning and sequencing. Products from the primers of Heuer et al. were used for amplified ribosomal DNA restriction analysis (ARDRA); products from the system of Monciardini et al. were used for T-RFLP.

PCR Amplification-F-Act/R-Bact

Cycling conditions for the primers of Heuer et al. [15] were: 35 cycles of 94°C (1 min)/64°C (1 min)/72°C (2 min), followed by a final extension of 72°C for 10 min. Amplification was carried out on 10–20 ng of environmental DNA in a total volume of 50 µL. Two units of in-house Taq polymerase produced

at the CIAT biotechnology laboratories was used with Sigma PCR buffer P-2192 with a final concentration of 10 mM Tris– HCl (pH 8.3), 1.5 mM $MgCl_2$ and 0.001% gelatin. PCR took place with 0.2 mM of each dNTP and 100 nM >of each primer. PCR products were run on a 0.8% agarose gel to check for successful amplification. Three samples did not give amplicons with the F-Act/R-Bact primers: ColforestRozo1, Rozo-wormcastMartio1, Rozo-wormcastMartio2.

PCR Amplification—F-Act/R-Act

Cycling conditions for the primers of Monciardini et al. [30] were: 30 cycles of 94°C (30 s)/68°C (2 min)/72°C (1 min), followed by a final extension of 72°C for 10 min. Amplification was carried out on a template of 10–20 ng of environmental DNA in a total volume of 50 µL. In addition to 2 U of Roche Taq and 10× buffer (1×concentration 10 mM Tris–HCl/50 mM KCl/1.5 mM $MgCl_2$, pH 8.3), the PCR cocktail had the following concentrations: 2% DMSO, 0.4 mM each dNTP,0.2 mg/mL bovine serum albumin and 500 nM of each primer. The 10× Roche buffer gives a final $MgCl_2$ concentration of 1.5 mM; this was raised to 1.825 mM. PCR products were run on a 1% agarose gel to check for successful amplification.

PCR Amplification—Bacterial Primers

Cycling conditions for the bacterial primers of Marchesi et al. [28] were: 30 cycles of 95°C (1 min)/55°C (1 min)/72°C (1.5 min), followed by a final extension of 72°C for 5 min. Amplification was carried out on 10–20 ng of environmental DNA in a total volume of 50 µL. In addition to 2 U of Roche Taq and 10× buffer (1×concentration 10 mM Tris– HCl/50 mM KCl/1.5 mM $MgCl_2$, pH 8.3), the PCR cocktail had the following concentrations: 2% DMSO, 0.2 mM each dNTP,0.2 mg/mL bovine serum albumin and 200 nM of each primer.

Cloning and Sequencing from F-Act/R-Bact Primer Amplicons

PCR products from 20 environmental DNA samples (11 soils and 9 animal-affected and urban sediments; Tables 3.1 and 3.2) were used for cloning with the Invitrogen TOPO TA cloning system. White colonies were picked and used for colony PCR using the T7 and SP-6 vector primers. PCR products were sent to the University of South Bohemia, the Czech Republic, for simultaneous bidirectional sequencing using a Li-Cor 4200 L automated DNA sequencer. All of the 1,100-bp amplicons of the 155 clones were sequenced.

Cloning and Sequencing from F-Act/R-Act and Bacterial Primers Amplicons

F-Act/R-Act PCR products from ten environmental DNA samples (six soils and four other sediments) and bacterial PCR products from six soils and six street sediments (Tables 1 and 2) were picked for cloning and sequencing with 16S rDNA gene-specific primers. PCR products were cleaned with QIA quick PCR purification columns before cloning using the Promega pGEM-T easy cloning system directly. Colonies were plated on X-gal and white colonies picked and used for colony PCR using the T7 and SP-6 vector primers. Sequencing reactions were carried out on the PCR products using either the universal bacterial 1406 R primer (F-Act/R-Act PCR products) or 765-F primer (bacterial PCR products) and the ABI prism Big Dye terminator sequence reaction. Sequence reaction products were precipitated and cleaned with 75% isopropanol, dried at 50°C and dissolved in 1.0 µL of a mixture of 0.72 deionised formamide /0.28 µL loading dye. The products were denatured at 95°C for 3 min, placed on ice and loaded on an ABI377 sequencer. Sequences were extracted and analysed with GeneScanTM analysis package of programmes. Five hundred to 700 bp were sequenced from each clone. Base calling was carried out in Chromas, version 2.1. 155 (F-Act/R-Act) and 520 (Bacterial) clones were sequenced at the Centre for Evolutionary and Ecological studies sequencing facility in Haren, the Netherlands. All F-Act/R-

Act sequences were used. After the exclusion of chimerical and sequences that were too short to align, 471 bacterial sequences were used for further analysis.

Sequence Data analysis and Accession Numbers

All 16S rRNA gene sequences were screened for chimeras using the CHECKCHIMERA programme of the Ribosomal Database programme, version 8.0 [6]. Fifteen sequences generated with the F-Act/R-Bact primer system and 49 sequences generated with eubacterial primers were rejected for further analysis as being either chimerical or too short for analysis. Five sequences generated with the F-Act/R-Bact primer system were either non-chimerical or chimaeras from closely related Verrucomicrobia; these sequences were used for analysis but not submitted to Genebank (CdnforestKemptH1, ColCIATH1, CdnforestKemptH2, ColforestPanceH10, NetoliceHiveH5). Sequences from both 16S rDNA actinobacterial and primer systems were aligned and phylogenetic trees were constructed using the MEGA, version 3.0 (Molecular Evolutionary Genetics Analysis) programme [23]. Clustering was carried out using the neighbour joining algorithm bootstrapping 1,000 times. ARDRA and T-RFLP bands for BLAST matches and clone sequences were simulated either using the TAP-TRFLP programme of the ribosomal database project (Release 8.0) or Bioedit.

Three Newick files using sequences from the F-Act/R Bact, F-Act/R-Act and bacterial primers were submitted to the Unfrac site (http://bmf2.colorado.edu/unifrac/index.psp) [27] for lineage-specific analysis.

For nucleotide sequence accession numbers, sequences were deposited in Genbank, accession numbers: F-Act/RBact HM444620-HM444755, F-Act/R-Act GQ494151GQ494304 and Bacterial HM444148-HM444619.

ARDRA F-Act/R-Bact

PCR products from soil/sediment DNA using the F-Act/R Bact primers were precipitated with ethanol and a final concentration of 0.3M potassium acetate. Pellets were washed with 70% ethanol and resuspended in 20 µL of restriction buffer E before digestion with 1µL of Taq I (Promega). 3–5µL of the restriction digest, was then run at 0.5V/cm for 2 h on a 0.7% agarose/1.5% Synergel (Diversified Biotech) gel. Gels were stained with ethidium bromide and images captured on a GenoSmart UV transilluminator. Band sizes were estimated using the Genosoft software package (VWR International, West Chester, PA).

T-RFLP F-Act/R-Act

PCR was carried out in 50 µL reaction volumes as described above using hexachloro-6-carboxyfluorescein and carboxyfluorescin-labelled forward and reverse primers, respectively. Actinobacterial PCR products were cleaned and concentrated in 20 µL of restriction buffer using Qiagen QIAquick PCR purification columns before digestion with 2 µL of HhaI (Promega). Digested PCR product (1.1–4 µL) was dried at 50°C for 5–10 min, redissolved in 1.1 µL of 55 deionised formamide/11 ROX 1000 marker (Applied Biosystems)/11 loading dye (Applied Biosystems), denatured at 95°C for 3 min, immediately placed on ice and then run on an Applied Biosystems ABI377 sequencer to generate T-RFLP patterns. T-RFLP patterns were analysed using peaks in the 81- 677-bp range. All T-RFLP patterns used had cumulative peak heights >10,000 units [2]; for eight samples, reverse patterns and for one sample the forward pattern did not meet this criterion and were not used for analysis.

Analysis of ARDRA and T-RFLP Patterns

Patterns were imported into GELCOMPAR (Applied Maths, http://www.applied-maths.com) and compared using unweighted pair group method with arithmetic mean (UPGMA) cluster analysis.

Redundancy analysis (RDA) of patterns from samples for which pH, organic matter and clay content were measured (i.e. excluding most animal-associated sediments where high organic matter contents made it impractical to determine texture) was carried out using canonical community ordination (CANOCO) as provided by Plant Research International BV, Wageningen, the Netherlands. Community similarities were shown on ordination tri-plots, with scaling focussed on inter-sample differences and species (i.e. bands) data removed for ease of presentation. A Monte Carlo permutation test based on 499 random permutations was carried out, with the null hypothesis that patterns were unrelated to land use. Land use was treated as a nominal variable with declining values for increasing intensity of use (i.e.uncultivated soils, 3; cultivated soils, 2; street sediments, 1). Organic matter, pH and clay content were treated as supplementary variables. Organic matter and clay content were log-transformed before analysis as they were expressed as percentages; species data (i.e. band information), was square root-transformed.

Results

Sequences Generated Using the Single 16S rDNA Actinobacterial-Specific Primer System (F-Act/R-Bact)

Sequencing results show that the actinobacterial-specific primers F-Act/R-Bact amplify actinobacterial and verrucomicrobial 16S rDNA. One hundred and forty 16S rDNA amplicons from the F-Act/R-Bact primer system were sequenced and more than half of these sequences (75/140) were not actinobacterial. The majority of the non-actinobacterial clones sequenced were Verrucomicrobial (71/75). These clones are from two Verrucomicrobial subdivisions: subdivision 2, renamed Spartobacteria [39], and subdivision 3. These subdivisions are the dominant Verrucomicrobia in soil [40]. Of the 71 Verrucomicrobia sequences identified, 53 were from soil samples (Supplemental data 1). Lineage-specific analysis with UniFrac confirmed that there was selection for the Verrucomicrobia in soils (P=0.0001, Fig. 1).

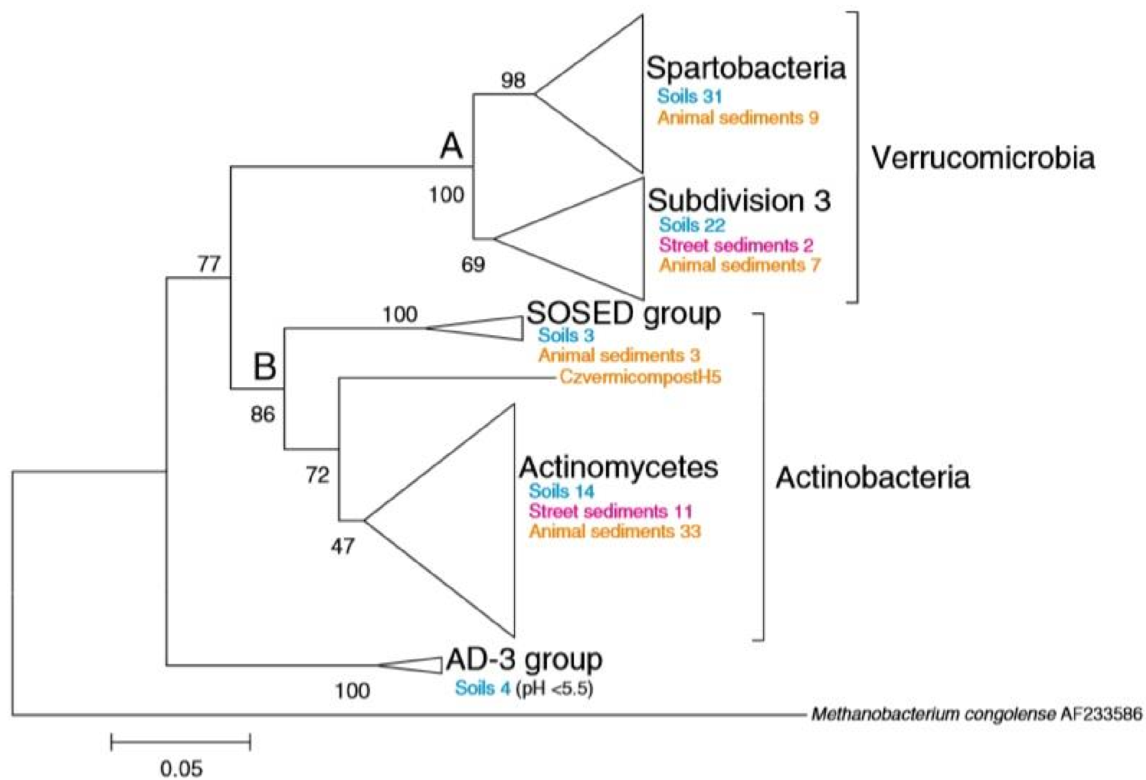


Figure 3.1 Neighbour-joining tree of 16S sequences generated with the F-Act/R-Bact primers of Heueretal. [15]. Bootstrap values below 50 are not shown. Samples are colour coded as: forest/pasture soils (blue), insect- and earthworm-associated sediments (orange), street sediments (red). Unifrac lineage analysis was carried out on nodes A and B. For node A the P value was <0.0002 and observed/expected occurrences were for soils 53/37.7, street sediments 2/6.6 and animal affected sediments 16/26.7. For node B the P value was <0.0000 and observed/expected occurrences were for soils 17/35.1, street sediments 11/6.1 and animal affected sediments 37/28.8

The remaining four non-actinobacterial sequences grouped with the AD-3 candidate division previously found in a sandy acid soil from eastern USA [44]. All four clones were from soils with pH ≤ 5.0, and all of the closest BLAST matches were from soils and sediments with pH ≤ 5.5 (ESM Supplemental data 2).

The remaining 65 clones generated with the F-Act/R-Bact primers were actinobacterial. Actinobacterial clones were from Streptomycineae, Micrococcineae, Frankineae, Propionibacterineae, Pseudonocardineae and Corynebacterineae. Six clones from soil and worm gut were from a novel, non actinomycetal Actinobacterial group. This soil and sediment group,(SOSED) is further discussed below.

Actinobacterial clones were predominantly identified (48 of 65) in samples from animal-affected and urban sediments (Fig. 1), with a smaller fraction of clones (17 of 65) identified in soils. Lineage-specific analysis with UniFrac confirmed that there was selection for Actinobacteria in non-soil sediments (Supplemental data 3).

Sequences Generated Using the 16S rDNA Double Actinobacterial-Specific Primers (F-Act/R-Act)

Given the poor specificity of the F-Act/R-Bact primers, samples were characterised with the more specific F-Act/RAct primers. All clones generated with the F-Act/R-Act primers were actinobacterial. Two phylogenetic groups, a novel group (the SOSED group) and the Nocardioidaceae family constituted 79% of the clones of the library generated with the F-Act/R-Act primers (61/156 and 62/ 156, respectively). The SOSED group made up the majority of soil clones and the Nocardioidaceae the majority of street and animal sediment clones.

There was considerable difference in the Actinobacteria identified using the F-Act/R-Bact and F-Act/R-Act primer sets. The SOSED group and Nocardioidaceae were substantially less represented in the actinobacterial clones generated using the F-Act/R-Bact primers (6/65 and 12/ 65, respectively.). In addition, the Frankineae and non mycobacterial Corynebacterineae found with the F-Act/R-Bact primer set were not detected with the F-Act/R-Act primers.

A large number of clones from the new SOSED group were identified. The closest BLAST matches to the SOSED group are from environmental clones from soils in Australia [17], Hawaii [12], Germany [13], England [7], California [24] and the Netherlands [19], as well as deep sea sediments [34], freshwater lake sediments [31], uranium mining waste [38] and a hot spring [22] (Supplemental data 4). A preliminary designation of this group by İnceoğlu et al. [19] was environmental group 2; we refer to it as the SOSED (soil/sediment) group as the BLAST matches and our sequences were from marine and terrestrial soils and sediments. The vast majority of SOSED amplicons were from soil samples (58/61),

and of the 79 soil amplicons, 58 were from the SOSED group. Lineage specific analysis with UniFrac confirmed that there was selection for the SOSED group in soils (P=0.0000, Fig. 2).

Amplicons from samples with increased intensity of land use were largely actinomycetal. Sixty-seven of 77 clones from cultivated (Colcultivated-Rozo), animal-affected (wormgutsRozo, hivewaste-Netolice) and urban (cobblestonesBrusselsBoucher, Č.Budéjovice-streetsediment-Česká) DNA samples were from the Actinomycetes. Lineage-specific analysis with UniFrac confirmed that there was selection for the Actinomycetes in non-soil sediments (P=0.0002, Fig. 2).

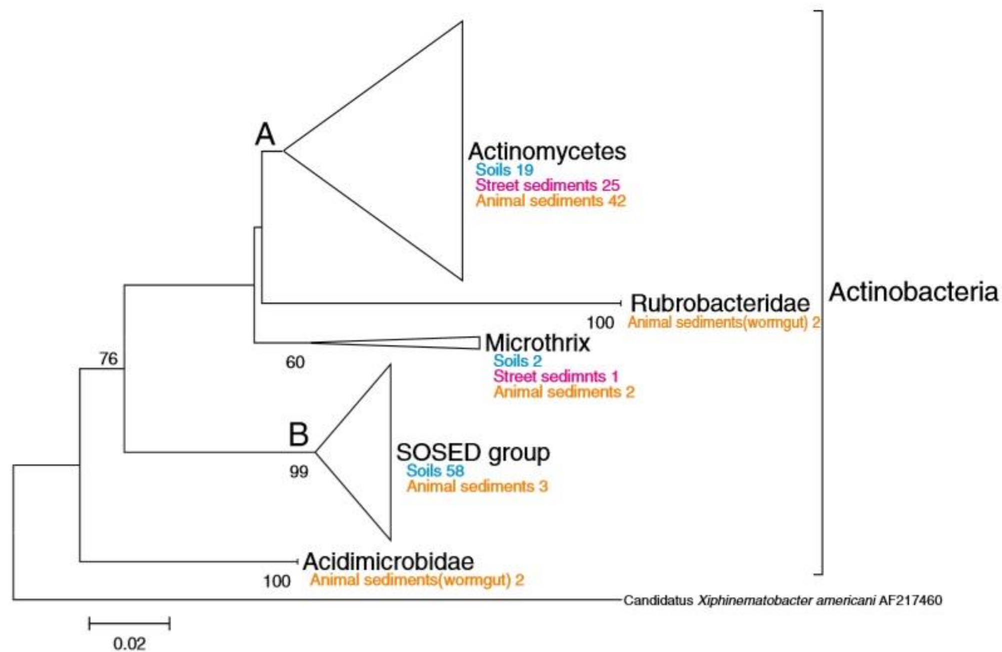
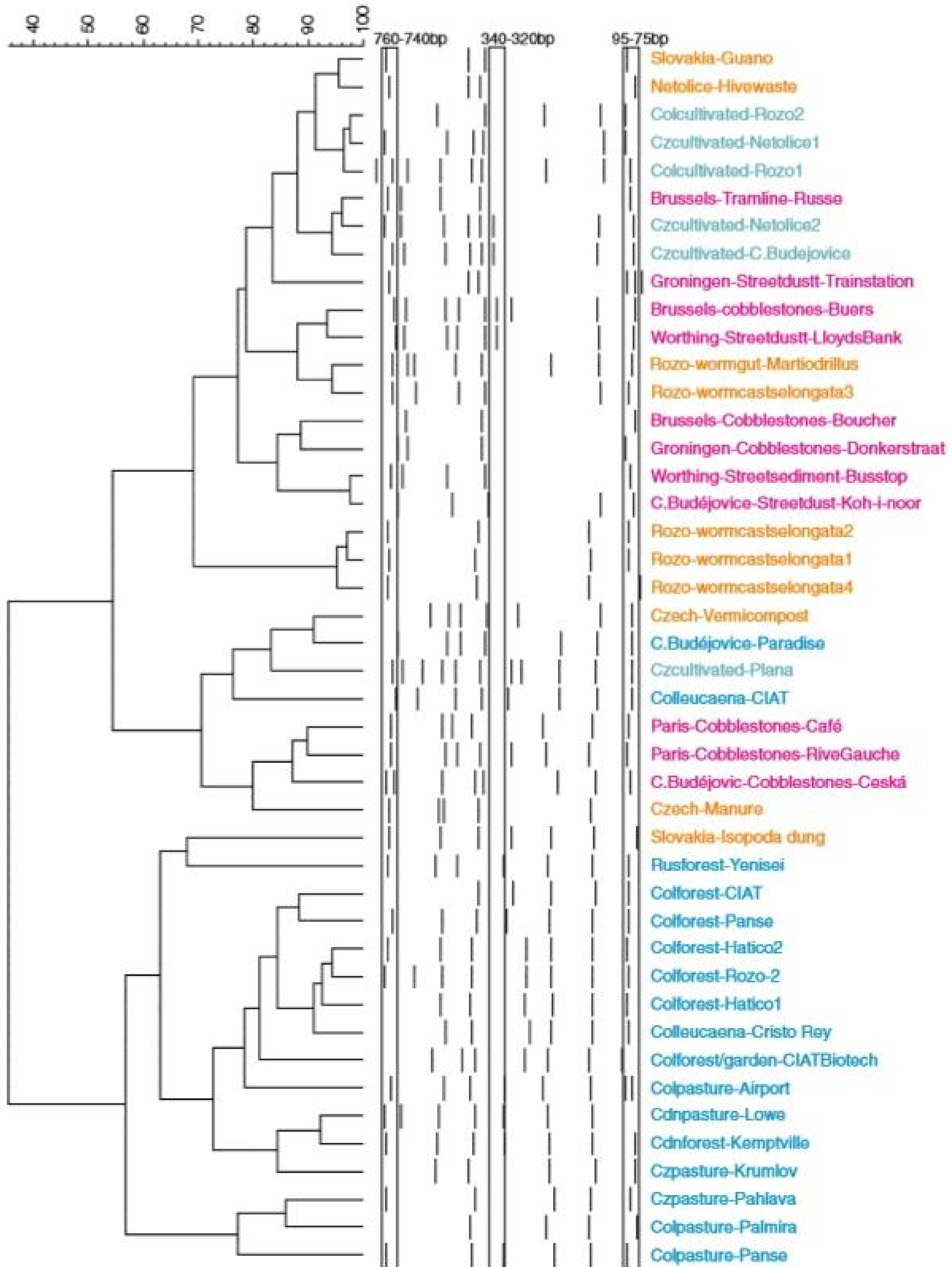


Figure 3.2 Neighbour-joining tree of 16S clones generated with the FAct/R-Act primers of Monciardini et al. [30]. Bootstrap valules below 50 are not shown. Simulated T-RFLP values are shown for all matches were sequence length allows. Samples are colour coded as: soils (blue), insect- and earthworm-associated sediments (orange), street sediments (red). Unifrac lineage analysis was carried out on nodes A and B. For node A the P value was <0.0000 and observed/expected occurrences were for soils 19/43, street sediments 25/14.5 and animal affected sediments 42/28.5. For node B the P value was <0.0000 and observed/expected occurrences were for soils 58/29, street sediments 0/9.9 and animal affected sediments 3/19.1

ARDRA Patterns Generated Using the 16S rDNA Single Actinobacterial-Specific Primer System (F-Act/R-Bact)

Having gained information into the actinobacterial community composition and phylogenetic relationships from sequencing, finger printing analysis of the 16S rDNA sequences was carried out to enable comparison of the many sample sites. TaqI ARDRA was used as the actinobacterial and verrucomicrobial amplicons possessing markedly different banding patterns. Simulated Taq I restriction of 65 actinobacterial sequences found that for all but 12 of the sequences, there were nobands between 94 and 32 0bp.Six of the sequences with bands in this range were non-actinomycetal Actinobacteria and four of these were from the SOSED group. Simulated Taq I restriction of all 71 verrucomicrobial clones and their closest matches gave bands in the 95-320-bp range (e.g. 120, 131 and 181 bp; Supplemental data 1 and 3).

The ARDRA patterns of the different soils and sediments that were investigated in this study clustered into three groups. One group contained cultivated soils, street sediments and animal sediments and possessed few bands in the 94-320-bp range. A second group contained most of the uncultivated soils and possessed many bands in the 94-320-bp range. Finally, an intermediate group containing animal-affected sediments, soils and street sediments was also identified (Fig. 3). These ARDRA patterns confirmed the F-Act/R-Bact cloning and sequencing results, demonstrating that verrucomicrobial amplicons were more common in soil PCR products and that actinobacterial amplicons were more common in animal-affected and street sediment PCR products.



40
50
60
70
80
90
100
760-740bp
340-320bp
95-75bp
Slovakia-Guano
Netolice-Hivewaste
Colcultivated-Rozo2
Czcultivated-Netolice1
Colcultivated-Rozo1
Brussels-Tramline-Russe
Czcultivated-Netolice2
Czcultivated-C.Budejovice
Groningen-Streetdustt-Trainstation
Brussels-cobblestones-Buers
Worthing-Streetdustt-LloydsBank
Rozo-wormgut-Martiodrillus
Rozo-wormcastselongata3
Brussels-Cobblestones-Boucher
Groningen-Cobblestones-Donkerstraat
Worthing-Streetsediment-Busstop
C.Budéjovice-Streetdust-Koh-i-noor
Rozo-wormcastselongata2
Rozo-wormcastselongata1
Rozo-wormcastselongata4
Czech-Vermicompost
C.Budéjovice-Paradise
Czcultivated-Plana
Colleucaena-CIAT
Paris-Cobblestones-Café
Paris-Cobblestones-RiveGauche
C.Budéjovic-Cobblestones-Ceská
Czech-Manure
Slovakia-Isopoda dung
Rusforest-Yenisei
Colforest-CIAT
Colforest-Panse
Colforest-Hatico2
Colforest-Rozo-2
Colforest-Hatico1
Colleucaena-Cristo Rey
Colforest/garden-CIATBiotech
Colpasture-Airport
Cdnpasture-Lowe
Cdnforest-Kemptville
Czpasture-Krumlov
Czpasture-Pahlava
Colpasture-Palmira
Colpasture-Panse

Figure 3.3 UPGMA tree of amplified ribosomal DNA restriction analysis (ARDRA) patterns generated using the F-Act/R-Bact primers of Heuer et al. [15] and TaqI digestion. Samples are colour coded as: forest/pasture soils (blue), cultivated soils (sea green), insect- and earthwormassociated sediments (orange), street sediments (red). For sample descriptions, see Tables 1 and 2

T-RLFP Patterns Generated Using the 16S rDNA Double Actinobacterial (F-Act/R-Act) Primers

Both forward and reverse T-RFLP patterns were used to characterise the F-Act/R-Act 16S rDNA amplicons. Reverse T-RFLP patterns showed the distribution of the SOSED group, whilst forward T-RFLP patterns gave a picture of the actinobacterial community structure.

Reverse T-RFLP

Simulated Hha I restriction of all 84actinomycetal clones and their BLAST matches showed reverse T-RFLP peaks between 348 and 385 bases (ESMSupplementaldata4 and 5), with all but three peaks between 374 and 385 bases (Colforest/ garden-CIATbiotech8, 352 bases; Netolice-hivewaste8, 348 bases; and *Micromonospora aurantiaca*, 349 bases). Reverse T-RFLP peaks of 374–385 bases were also found among the non-actinomycetal actinobacterial clones, notably the five Microthrix-related clones. However, most non-actinomycetal actinobacterial clones had reverse T-RFLP peaks outside the 374- 385-base range. The most common reverse SOSED group T-RFLP peak was at 186–190 bases (42 of 60 clones), with other peaks at 148, 239–242 and 376–390 bases. For reverse T-RFLP, all peaks outside the 348- 390-base range thus represent non-actinomycetal Actinobacteria. These peaks are mainly representative of the SOSED group.

All but one of the reverse T-RFLP patterns clustered in three groups, two from soil and the third from street and animal-affected sediments. A single sample of environmental DNA from vermicompost (Czech-vermicompost) gave two bands at 262 and 272 bp, which were unrelated to any environmental clones (Fig. 4).

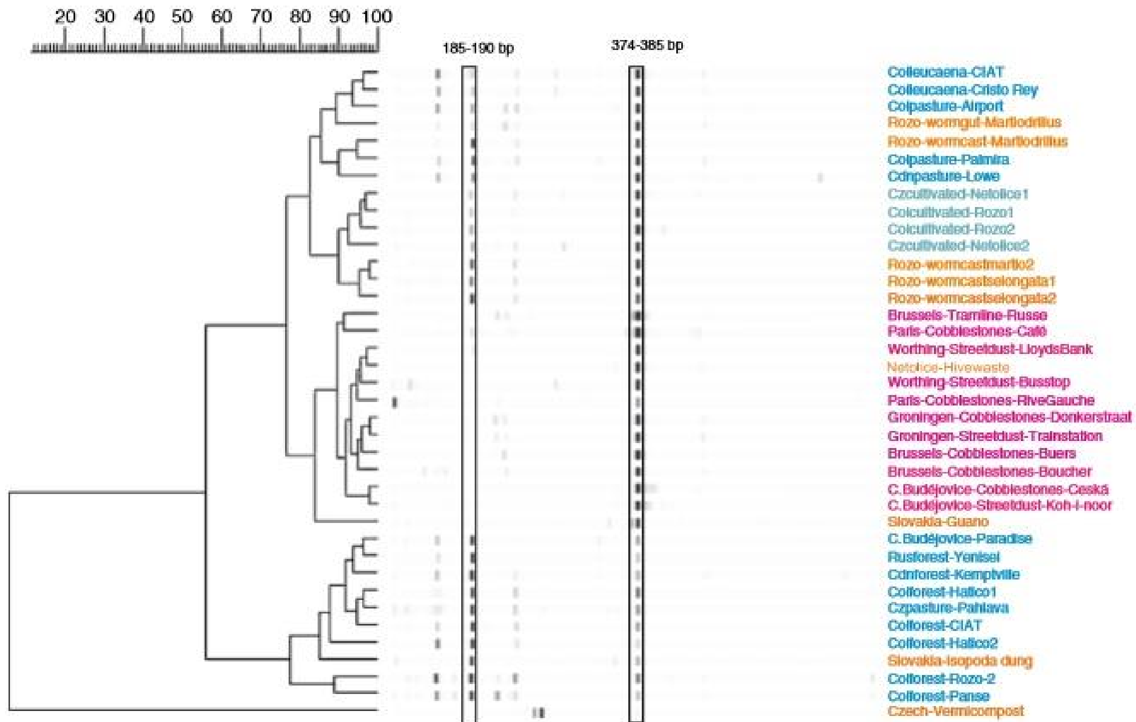


Figure 3.4 UPGMA tree of terminal restriction fragment polymorphism (T-RFLP) patterns generated using the A3R primer of Monciardini et al. [30] F-Act/R-Act primers and HhaI digestion. Samples are colour coded as: forest/pasture soils (blue), cultivated soils (sea green), insect- and earthworm-associated sediments (orange), street sediments (red). For sample descriptions, see Tables 1 and 2

Forward T-RFLP

Identical forward T-RFLP HhaI peaks were often produced by different phylogenetic groups; however, patterns clustered clearly by land use.

Simulated HhaI restriction of BLAST matches to the SOSED group and of Norcardioidaceae sequences from ribosomal database programme both gave forward T-RFLP fragments of 147 bases. A Siberian OS clone gave a peak at 145 bases.

Forward T-RFLP patterns clustered by land use with an uncultivated soil cluster, cultivated soil cluster and cluster of street and animal-affected sediments. Worm cast T-RFLP patterns grouped between cultivated and forested soils; the microcosm that they were from contained earthworms from Colforest-Rozo1 and 2, and soil from Colcultivated-Rozo2. A single street sediment (Brussels-

cobblestones-Beurs) clustered with cultivated soils; however, most T-RFLP patterns from street and animal-affected sediments not only clustered away from soils but also differed more from each other (30–60% similarity) than soil patterns (65–90% similarity). Several patterns (e.g. Czech-Vermicompost) did not cluster with any of the others (Fig. 5).

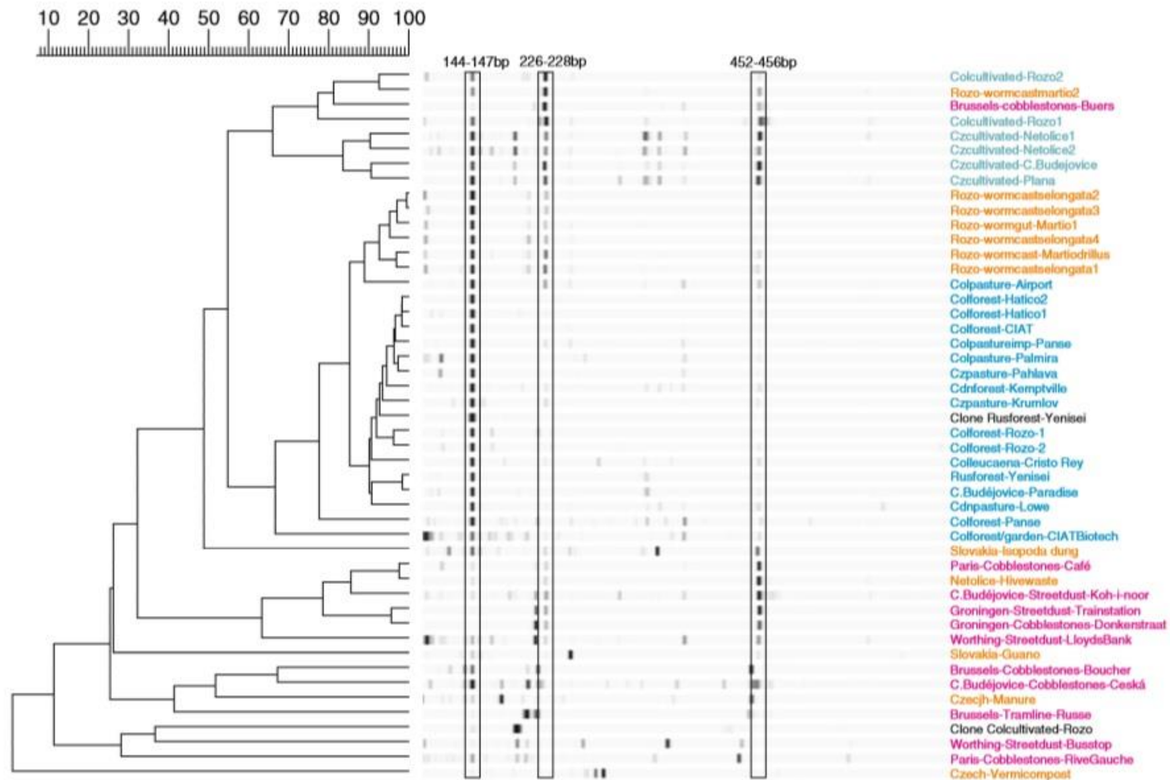


Figure 3.5 UPGMA tree of terminal restriction fragment polymorphism (T-RFLP) patterns generated using the F-243 primer of Monciardini et al's[30] F-Act/R-Act primers and HhaI digestion. Samples are colour coded as: forest/pasture soils (blue), cultivated soils (sea green), insect- and earthworm-associated sediments (orange), street sediments (red). Two sequenced clones are also included for comparison. For sample descriptions, see Tables 1 and 2

Canonical Correspondence Analysis of Fingerprints

As clustering suggested that land use controlled ARDRA and T-RFLP patterns, this was tested using a constrained analysis (RDA) with pH, log % clay and log % organic matter as supplementary variables to interpret their effect as well as with a Monte Carlo permutation test to test the significance

of land use. Tests were performed on samples for which pH, organic matter and clay content had been determined (i.e. Netolice-Hivewaste, all urban sediments and soils) as well as soils alone (Table 3.3).

Table 3.3 Gradient lengths and Monte Carlo Permutation values.

	Gradient length[a]	F ratio[b]	P[c]
ARDRA	3.061	3.6	0.002
Forward T-RFLP	3.237	32	0.002
Reverse T-RFLP	3.694	33.57	0.002
ARDRA soils only	2.892	2.42	0.008
Forward T-RFLP soils only	2.663	12.98	0.002
Reverse T-RFLP soils only	2.580	3.18	0.004

[a] Gradient lengths were determined by DCA analysis

[b] F ratio for the first axis

[c] The significance of the first axis based on a Monte Carlo permutation test (499 permutations)

RDA analysis showed that land use was correlated with ARDRA and forward and reverse T-RFLP patterns and that clay content in turn was closely correlated with land use (Fig. 6a–c), with land use explaining 30%, 48% and 55% of variance (in each case, P values were 0.002). As street sediments were unusually sandy and alkaline, an RDA analyses of only soils was carried out. Again, land use controlled T-RFLP patterns, explaining 28%, 37% and 18% of variance with P values of 0.008, 0.002 and 0.004, respectively. However, there was poor correlation to log clay content and a much stronger correlation to log organic matter (Fig. 6).

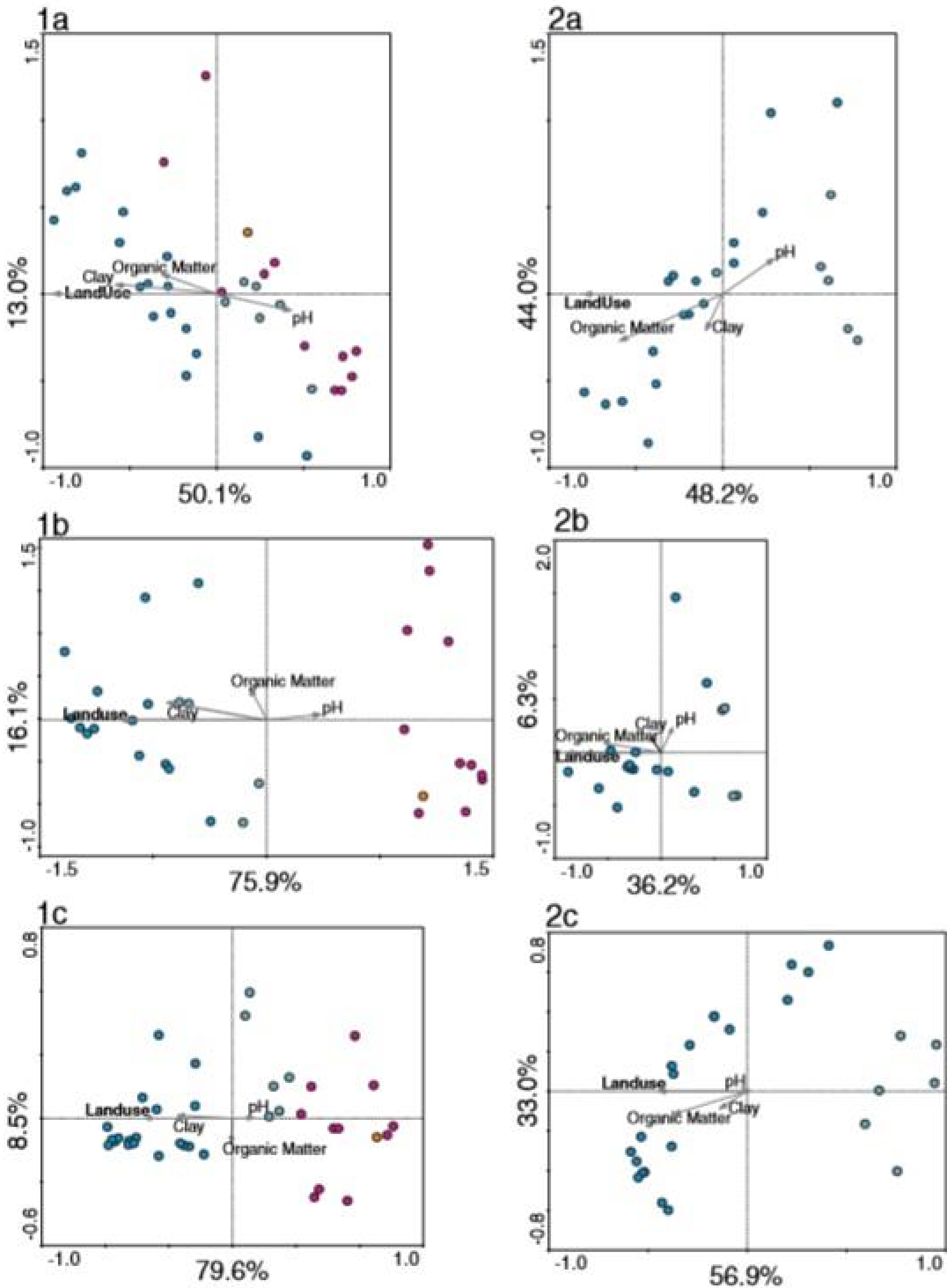


Figure 3.6 Redundancy analysis ordination plots of fingerprint patterns for all characterised samples. Samples are colour coded as: forest/pasture soils (blue), cultivated soils (sea green), insect- and earthworm associated sediments (orange), street sediments (red). 1a Amplified ribosomal DNA restriction analysis (ARDRA) patterns for all characterised samples. 1b Reverse terminal restriction fragment polymorphism (T-RFLP) patterns for all characterised samples. 1c Forward terminal restriction fragment polymorphism (T-RFLP) patterns for all characterised samples. 2a Amplified ribosomal DNA restriction analysis (ARDRA) patterns for all soils. 2b

Reverse terminal restriction fragment polymorphism (T-RFLP) patterns for all soils. 2c Forward terminal restriction fragment polymorphism (T-RFLP) patterns for all soils.Samples are colour coded as: forest/pasture soils (blue), cultivated soils (sea green).

Land use thus appeared to control the actinobacterial ARDRA and T-RFLP patterns. However, either the controlling factors varied across the gradient of land use or there was a controlling factor that was not measured.

Sequences Generated with the Eubacterial Primers of Marchesi et al. [28]

To determine if changes in the actinobacterial community structure found with actinobacterial-specific primers are accompanied by an increase in the actinobacterial population, two environments soil and street sediments were more intensively characterised with eubacterial primers. Whilst Actinobacteria were more common in street sediments than soil, the bacterial communities of street sediments varied considerably.

Four hundred seventy-one Eubacterial amplicons from six soils (three cultivated/three forest) and six street sediments (three street dust/three sediments in place) were analysed. Eighty-seven per cent of sequences were from seven orders: the Actinobacteria (126), Alpha proteobacteria (89), Acidobacteria (75), Gamma proteobacteria (59), Bacterioidetes (28), Firmicutes (20) and Delta proteobacteria (15). Whilst UniFrac lineage analysis confirmed that the Actinobacteria were specific to street sediments (P< 0.0000) and that the Acidobacteria were specific to soils (P<0.0000, Table 3.4), the gamma and delta proteobacterial clusters interleaved with each other, making lineage-specific analysis impossible. In two samples (České.Budějovicestreetdust-Koh-i-Noor, Streetsediment-ClockFaisalabad), Actinomycetes made up 95–100% of clones sequenced. When these two samples were removed from UniFrac lineage analysis, the Actinobacteria were still specific to street sediments (P<0.002). Sixty-four per cent (81 out of 127) of actinobacterial clones were from the Micrococcinae (Fig. 7), whereas only 2.7% and 0.64% of clones generated withthe F-Act/R-Bact and F-Act/R-Act primers, respectively, were from the Micrococcinae.

Table 3.4 Eubacterial clones sequenced.

	No. (percentage) of clones assigned at 80% confidence level (Cole et al. [6])												
	Samples	Clones sequenced	Acidobacteria	Actinobacterial	Proteobacteria					Bactericidetes	Firmicutes	Other[a]	Unclassified bacteria[a]
					Alpha	Beta	Delta[a]	Gamma[a]	Unclassified[a]				
Street sediments	Budapest-trainstation-fissure	39	0	7 (18)	1 (3)	0	0	18 (46)	0	9 (23)	4 (10)	0	0
	Faisalabad-streetdust-clock	50	0	49 (98)	0	0	0	0	0	0	1 (2)	0	0
	Groningen-trainstation-streetdust	42	0	11 (26)	5 (12)	0	0	16 (38)	0	9 (21)	0	1 (2)	0
	Brussels-cobblestones-Boucher	43	0	11 (26)	7 (16)	3 (7)	1 (2)	4 (9)	1 (2)	1 (2)	12 (28)	3 (7)	0
	CeskeBudějovice-streetdust-Kohinoor	39	0	35 (90)	0	0	0	0	1 (3)	0	0	0	3 (8)
	CeskeBudějovice-cobblestones-Ceska	45	1(2)	3 (7)	20 (44)	3 (7)	1 (2)	10 (22)	1 (2)	2 (4)	1 (2)	2 (4)	1 (2)
	Total for street sediments (RDP database)	258	1	116	33	6	2	48	3	21	18	6	4
	Total for street sediments (cluster analysis)[b]		1	116 (32)[c]	34	6				21	17		
	Expected value for street sediments (given even distribution)[d]		47.1	69.6 (19.0)[c]	50	6.6				15.3	10.4		
	P value[d]		0.0000	0.0000 (0.003)[c]	0.0483	1				1	0.1590		
Soils—cultivated	Colcultivated-Rozo	27	5 (19)	5 (19)	8 (30)	1 (4)	2 (7)	1 (4)	0	1 (4)	0	1 (4)	3 (11)
	Czcultivated-Netolice1	29	12 (41)	0	8 (28)	1 (3)	1 (3)	0	0	2 (7)	0	3 (10)	2 (7)
	Czcultivated-Plana	47	15 (32)	0	15 (32)	3 (6)	1 (2)	4 (9)	1 (2)	3 (6)	0	3 (6)	2 (4)
Soils—uncultivated	Hunforest-Citadel	24	6 (25)	0	5 (21)	0	4 (17)	4 (17)	0	0	0	0	5 (21)
	Colforest-Rozo	32	12 (38)	1 (3)	4 (13)	0	5 (16)	0	2 (6)	0	0	2 (6)	6 (19)
	Czforest-Kolny	54	24 (44)	4 (7)	16 (30)	1 (2)	0	2 (4)	0	1 (2)	2 (4)	0	4 (7)
	Total for soil (RDP database)	213	74	10	56	6	13	11	3	7	2	9	22
	Total for soil (cluster analysis)[b]		85	11 (11)[c]	57	6				7	2		
	Expected value for soils (given even distribution)[d]		39.9	57.4 (24.0)[c]	41	5.4				12.6	8.3		
	P value[d]		0.0000	0.0000 (0.003)[c]	0.0483	1				1	0.1590		

[a] Did not form a coherent subtree

[b] Values for 80% Ribosomal Database Project and cluster analysis did not always correspond completely

[c] Actinobacterial Unifrac lineage-specific analysis which excludes CeskeBudějovice-streetdust-Kohinoor and Faisalabad-streetdust-clock

[d] As determined by UniFrac (Lozupone et al. [27])

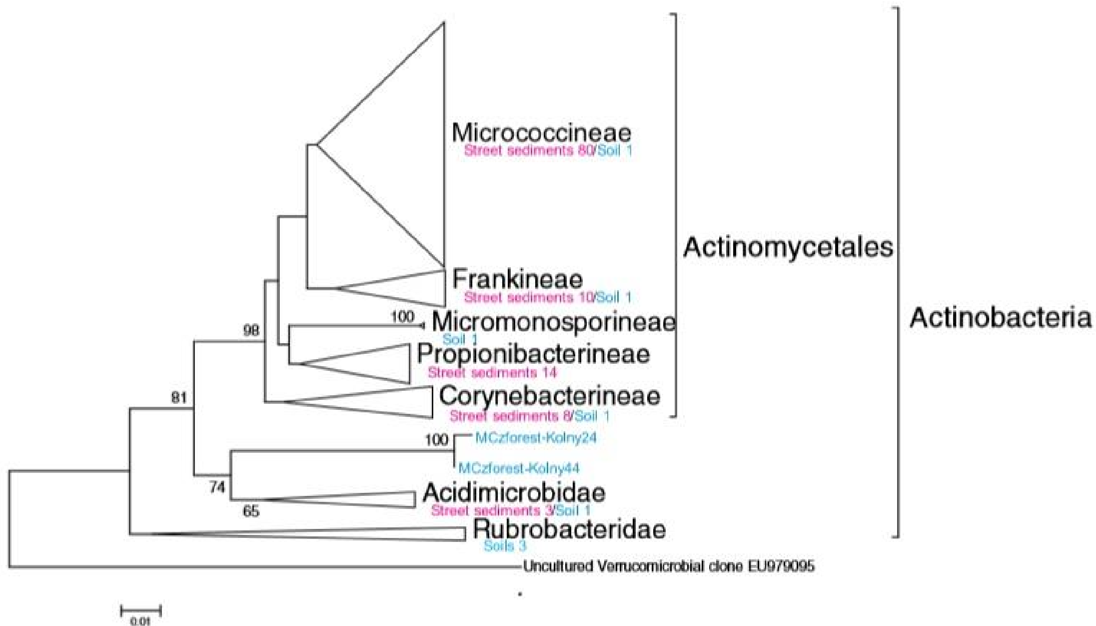


Figure 3.7 Neighbour-joining tree of Actinobacterial 16S clones generated with the F-Act/R-Act primers of Marchesi et al.[28]. Bootstrap values below 50 are not shown. Samples are colour coded as: soils (blue), insect- and earthworm-associated sediments (orange), street sediments (red).

UniFrac Jackknife clustering of the 12 clone libraries found that the three cultivated soils clustered together. The two actinomycetal-rich street dusts clustered together; however, the dust sample from Groningen railway station clustered with the sediment that appeared to be settled from the Budapest railway station. The two railway station communities were dominated by the Bacterioidetes and Gamma proteobacteria (Pseudomonas and in particular the Psychrobacter, Supplemental data 6).The two cobblestone samples differed from each other and all other samples. Brussels-Cobblestones-Boucher contained many Firmicutes (12/42), whilst České.Budějovice-cobblestonesČeska resembled the soil communities (Fig. 8).

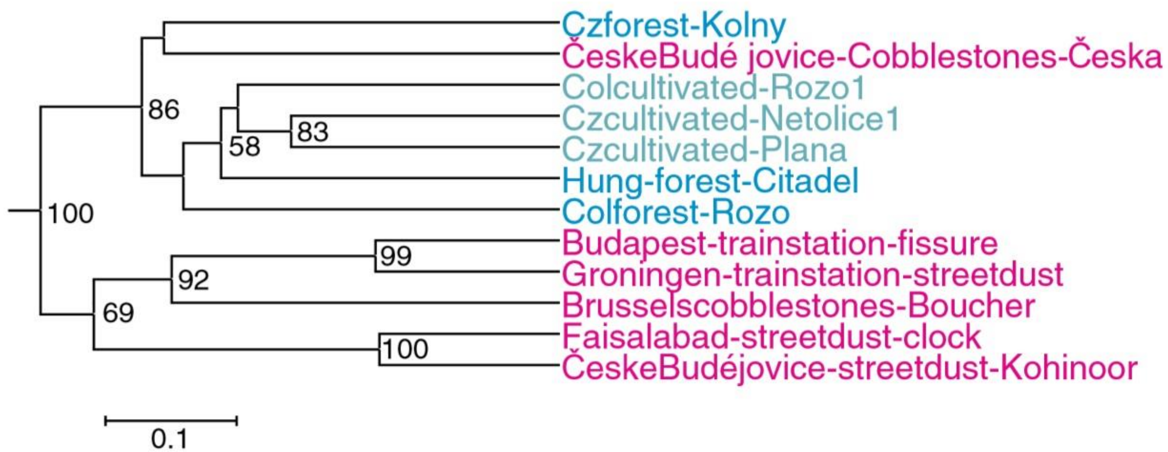


Figure 3.8 Unifrac Jackknife environmental clustering of the 12 eubacterial clone libraries generated with the primers of Marchesi et al. [28] using the weighted Unifrac algorithm. Jackknife values below 50 are not shown. Samples are colour coded as: forest soils (blue), cultivated soils (sea green), street sediments (red)

Discussion

The results from the three primer sets each tell a different phylogenetic story; however, they give a consistent ecological message. The F-Act/R-Bact primers show a switch from the Verrucomicrobia to the Actinobacteria, increasing from [Uncultivated soils][Cultivated soils][Street sediments/Animal affected sediments]; the F-Act/R-Act primers show a switch from the actinobacterial (but non-actinomycetal) SOSED group to the actinomycetal Norcardioides, increasing from [Uncultivated

soils][Cultivated soils][Street sediments/ Animal-affected sediments]; and sequences generated with bacterial primers from street sediments were enriched with the actinomycetal Micrococcinae compared to soils.

These results are supported by publications that have looked at a narrower range of soils and land uses. Buckley and Schmidt [4] found that the Verrucomicrobial population of soils decreased with cultivation. Whilst our results show that cultivation changes the actinobacterial community composition, there was no statistically significant evidence that cultivation increases the actinobacterial fraction of the bacterial community. In contrast, Lauber et al. [25] compared four land uses in Georgia, USA—cultivated, pasture, pine plantation and mixed wood forest—characterising the entire bacterial community using Q-PCR and found that the Actinobacteria were more common in pasture and cultivated soils. Waldrop et al. [41] found that an actinomycete-specific PLFA more than doubled on the conversion of Tahitian forest to pineapple plantation, whilst pH decreased slightly. Burke et al. [5] compared the bacterial communities of forest, pasture and sugarcane in Hawaii, Brazil and Ecuador. Actinomycetes were a larger fraction of the microbial community in agricultural than forest soils. Whilst we know of no other studies that characterise the Actinobacteria of street sediments, molecular and cultivation methods have found that Actinomycetes are common in urban environments such as masonry [37], stained glass [36] and wall paintings [14].

The common theme of these results is that the actinobacterial community is controlled and encouraged by human and animal activities: cultivation, urbanisation and digestion. In this study, urbanisation also had a strong selective effect on the Actinobacteria that occurred in different bacterial communities, suggesting that actinobacterial selection is independent of overall bacterial community structure (Fig 3.3). What is not clear is whether there is a single controlling factor for all three of these activities which controls actinobacterial community structure or if cultivation, urbanisation and

digestion each change the community structure through different mechanisms to develop distinct actinobacterial habitats.

If there is a single factor controlling actinobacterial community structure in these samples, it is not one of the three that we measured: pH, clay content or organic matter. Ordination plots showed that organic matter controlled actinobacterial community structure when fingerprints from soils were compared; pH appeared to control actinobacterial structure when soils and street sediments were compared.

Actinobacterial fingerprints from acidic animal waste samples clustered with street sediments. We can also safely exclude climate and parent material as overall controlling factors as there was no evidence of clustering by sampling area.

Actinobacterial distribution appears to differ from the distribution of the total bacterial community which is controlled by soil properties such as pH [9] and texture [10] and resemble that of the fungi, which is also increases with the conversion of forest to agriculture [5, 41]. It is possible that similarities in role, such as the development of multicellular structures such as filaments/hyphae, mean that there is a common factor controlling actinobacterial and fungal communities. Urbanisation has also been shown to have a profound effect on fungal community structure, with soils from abandoned Russian medieval cities containing similar fungal populations to modern cities [29]. Marfenina et al. also found that urban fungal communities were more beta diverse and ascribed this to the mosaic of environments and substrates found in cities.

The results of this study also have implications for experimental strategies in applying PCR to environmental DNA from soils and sediments.

These results were generated with a relatively low level of sequencing using three sets of primers; clone libraries of at least 400 clones are needed to characterise the soil bacterial community

[20]. This is because the object of our sequencing was not to characterise the bacterial communities of samples but either to interpret ARDRA and T-RFLP fingerprinting or to determine if the Actinobacteria were more common in certain samples through UniFrac lineage specific analysis. It could be argued that the resources spent on fingerprinting would have been better spent on concentrated sequencing to capture more of the rarer fraction (see [1] for both sides of this debate). However, sequences that are rare in clone libraries generated using one primer set may be common in libraries generated with a different primer set [18]. The actinobacterial Micrococcinae were rare in amplicons generated with actinobacterial primers whilst common in our eubacterial amplicons, whereas the Verrucomicrobia, previously found to be common in soil [40], were half of the sequences generated with F-Act/RBact primers and absent from sequences generated with eubacterial primers. Relatively shallow sequencing with a range of primers may give a better view of the bacterial community structure than intensive sequencing with a single primer set. Amplification of non-target sequences may yield interesting, if unpredictable, information about the distribution of non-actinobacterial groups. In our case, we found evidence that the AD-3 candidate division is selected for in acid soils, that the Acidobacteria are absent from street sediments (possibly due to alkalinity, they are selected for in acid soils) [21] and that the genus Psychrobacter, known to be common in extreme environments [35], also made up a large fraction of the bacterial community of two railway station samples.

Previous studies of actinobacterial distribution have found that location [42] and pH [26] control community structure in soil, an environment which provides many important ecosystem services. The goal of Actinobacterial bioprospecting is to identify environments that are enriched in Actinobacteria, so these studies should include a range of environments which are unimportant for ecosystem services (e.g. in this study, street dust, hive waste). As the results of any PCR reaction are a product of the sample, the DNA extraction method and the primer system used, sampling many contrasting samples

including commonly characterised environments such as soil also ensures that one can determine how much results are sample-specific and how much a product of extraction or PCR methods.

Acknowledgements

We thank Dr. P. Šmilauer, Pablo Hardoim and Dr. J. Salles for assistance with CANOCO analysis, Dr. A. Poulain for useful criticism of the manuscript, Dr. H. Šantrůčková for providing the Siberian soil sample, Ota Rauch for advice on sampling in the Czech Republic and Rashid Nazir for providing the Faisalabad Clock tower sample. We are indebted to the Slovak Caves Administration in Liptovský Mikuláš (Slovakia) for the sampling and research permit and to Daniël van Middelkoop and Sarina de Roos who carried out the bacterial cloning and sequencing as part of a work experience programme for the School for Biomedical Laboratory Techniques. Funding for this work was provided by the NATO Programme Security Through Science ESP.EAP.CLG 981785, the Ministry of Education, Youth and Sports of the Czech Republic (LC06066, 2B06154) and the Research Plan of the Institute of Soil Biology (AV 0Z 6066 0521).

References

1. Bent SJ, Pierson JD, Forney LJ (2007) Measuring species richness based on microbial community fingerprints: the emperor has no clothes. Appl Environ Microbiol 73:2399–2401

2. Blackwood C, Marsh T, Kim S-H, Paul EA(2003)Terminal restriction fragment length polymorphism data analysis for quantitative comparison of microbial communities. Appl Environ Microbiol 69:926–932

3. Bossio DA, Girvan MS, Verchot L, Bullimore J, Borelli T, Albrecht A, Scow KM, Ball AS, Pretty JN, Osborn AM (2005) Soil microbial community response to land use change in an agricultural landscape of Western Kenya. Microb Ecol 49:50–62

4. Buckley DH, Schmidt TM (2001) Environmental factors influencing the distribution of rRNA from Verrucomicrobia in soil. FEMS Microbiol Ecol 35:105 –112

5. Burke RA, Molina M, Cox JE, Osher LJ, Piccolo MC (2003) Stable carbon isotope ratio and composition of microbial fatty acids in tropical soils. J Environ Qual 32:198–206

6. Cole JR, Chai B, Farris RJ, Wang Q, Kulam SA, McGarrell DM, Garrity GM, Tiedje JM (2005) The Ribosomal Database Project (RDP-II): sequences and tools for high-throughput rRNA analysis. Nucleic Acids Res 33:294–296

7. Ellis RJ, Morgan P, Weightman AJ, Fry JC (2003) Cultivation dependent and -independent approaches for determining bacterial diversity in heavy-metal-contaminated soil. Appl Environ Microbiol 69:3223–3230

8. Fierer N, Jackson RB (2006) The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci USA 103:626–631

9. Fierer N, Carney KM, Horner-Devine MC, Megonigal JP (2009) The biogeography of ammonia-oxidizing bacterial communities in soil. Microb Ecol 58:435–445

10. Gelsomino A, Keijzer-Wolters AC, Cacco G, van Elsas JD (1999) Assessment of bacterial community structure in soil by polymerase chain reaction and denaturing gel electrophoresis. J Microbiol Methods 38:1–15

11. Girvan MS, Bullimore J, Pretty JN, Osborn AM, Ball AS (2003) Soil type is the primary determinant of the composition of the total and active bacterial communities in arable soils. Appl Environ Microbiol 69:1800–1809

12. Gomez-Alvarez V, King G, Nuesslein K (2007) Comparative bacterial diversity in recent Hawaiian volcanic deposits of different ages. FEMS Microbiol Ecol 60:60–73

13. Graff A, Conrad R (2005) Impact of flooding on soil bacterial communities associated with poplar (Populus sp.) trees. FEMS Microbiol Ecol 53:401–415

14. Gurtner C, Piñar G, Lubitz W, Swings J, Rölleke S (2000) Comparative analyses of the bacterial diversity on two different biodeteriorated wall paintings by DGGE and 16S rDNA sequence analysis. Int Biodeterior Biodegrad 46:229–239

15. Heuer H, Krsek M, Baker P, Smalla K, Wellington EM (1997) Analysis of actinomycete communities by specific amplification of genes encoding 16S rRNA and gel-electrophoretic separation in denaturing gradients. Appl Environ Microbiol 63:3233–3241

16. Hill P, Krištůfek V, Martinez AF, Gallego G (2005) The characterization of wormcasts through of amplified ribosomal DNA (ARDRA) and 16S ribosomal sequencing. In: Tajovský K, Schalghamerský J, Pižl V (eds) Proceedings of the 7th Central European Workshop on Soil Zoology, 14–15 April 2003. Institute of Soil Biology ASCR, České Budějovice, pp 25–29

17. Holmes AJ, Bowyer J, Holley MP, O'Donoghue M, Montgomery M, Gillings MR (2000) Diverse, yet-to-be-cultured members of the Rubrobacter subdivision of the Actinobacteria are widespread in Australian arid soils. FEMS Microbiol Ecol 33:111–120

18. Hong SH, Bunge J, Leslin C, Jeon S, Epstein SS (2009) Polymerase chain reaction primers miss half of rRNA microbial diversity. ISME J 12:1365–1373

19. İnceoğlu Ö, Hoogwout EF, Hill P, van Elsas JD (2010) Effect of DNA extraction method on the apparent microbial diversity of soil. Appl Environ Microbiol 76:3378–3382

20. Janssen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol 72:1719–1728

21. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N (2009)A comprehensive survey of soil acido bacterial diversity using pyrosequencing and clone library analysis. ISME J 3:442–453

22. Kanokratana P, Chanapan S, Pootanakit K, Eurwilaichitr L (2004) Diversity and abundance of bacteria and archaea in the Bor Khlueng Hot Spring in Thailand. J Basic Microbiol 44:430–444

23. Kumar S, Tamura K, Nei M (2004) MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

24. LaMontagne MG, Schimel JP, Holden PA (2003) Comparison of subsurface and surface soil bacterial communities in California grassland as assessed by terminal restriction fragment length polymorphisms of PCR-amplified 16S rRNA genes. Microb Ecol 46:216–227

25. Lauber CL, Strickland MS, Bradford MA, Fierer N (2008) The influence of soil properties on the structure of bacterial and fungal communities across land use types. Soil Biol Biochem 40:2407–2415

26. Lauber CL, Hamady M, Knight R, Fierer N (2009) Pyrosequencing based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale.Appl Environ Microbiol 75:5111–5120

27. Lozupone C, Knight R (2005) UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–8235

28. Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, Hiom SJ, WadeWG(1998)Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. Appl Environ Microbiol 64:795–799

29. Marfenina OE, Ivanova AE, Kislova EE, Sacharov (2008) The mycological properties of medieval culture layers as a form of soil 'biological memory'about urbanization .J Soils Sediments 8:340–348

30. Monciardini P, Sosio M, Cavaletti L, Chiocchini C, Donadio S (2002) New PCR primers for the selective amplification of 16S rDNA from different groups of actinomycetes. FEMS Microbiol Ecol 43:419–429

31. Nercessian O, Noyes E, Kalyuzhnaya MG, Lidstrom ME, Chistoserdova L (2005) Bacterial populations active in metabolism of C1 compounds in the sediment of Lake Washington, a freshwater lake. Appl Environ Microbiol 71:6885–6899

32. Neufeld JD, Mohn WW (2005) Unexpectedly high bacterial diversity in arctic tundra relative to boreal forest soils, revealed by serial analysis of ribosomal sequence tags. Appl Environ Microbiol 71:5710–5718

33. Nováková A, Elhottová D, Krištůfek V, Lukešová A, Hill P, Kováč L, Mochk A, Lùptáčik P (2005) Feeding sources of invertebrates in Ardovská and Domica cave systems. In: Tajovský K, Schalghamerský J, Pižl V (eds) Proceedings of the 7th Central European Workshop on Soil Zoology, 14–15 April 2003. Institute of Soil Biology ASCR, České Budějovice, pp 107–112

34. Reed DW, FujitaY DME, Blackwelder DB, Sheridan P, Uchida T, Colwell FS (2002) Microbial communities from methane hydrate bearing deep marine sediments in a Forearc Basin. Appl Environ Microbiol 68:3759–3770

35. Rodrigues DF, da Jesus EC, Ayala-del-Rio HL, Pellizari VH, Gilichinsky D, Sepulveda-Torres L, Tiedje JT (2009) Biogeography of two cold-adapted genera: Psychrobacter and Exiguobacterium. ISME J 3:658–665

36. Rölleke S, Gurtner C, Drewello D, Lubitz W, Weissmann R (2000) Analysis of bacterial communities on historical glass by denaturing gradient gel electrophoresis of PCR-amplified gene fragments coding for 16S rRNA. J Microbiol Methods 36:107–114

37. Schabereiter-Gurtner C, Piñar G, Vybiral D, Lubitz W, Rölleke S (2001) Rubrobacter-related bacteria associated with rosy discolouration of masonry and lime wall paintings. ArchMicrobiol 176:347–354

38. Selenska-Pobell S, Kampf G, Flemming K, Radevaand G, Satchanska G (2001) Bacterial diversity in soil samples from two uranium waste piles as determined by rep-APD, RISA and 16S rDNA retrieval. Antonie Leeuwenhoek 79:149–161

39. Sangwan P, Chen X, Hugenholtz P, Janssen P H(2004) Chthoniobacter flavus gen. nov., sp. nov., the first pure-culture representative of subdivision two, Spartobacteria classis nov. of the phylum Verrucomicrobia. Appl Environ Microbiol 70:5875–5881

40. Sangwan P, Kovacs S, Davis KER, Sait M, Janssen PH (2005) Detection and cultivation of soil Verrucomicrobia. Appl Environ Microbiol 71:8402–8410 41. Waldrop MP, Balser TC, Firestone MK (2000) Linking microbial community composition to function in a tropical soil. Soil Biol Biochem 32:1837–1846

42. Wawrik B, Kutliev D, Abdivasievna UA, Kukor JJ, Zylstra GJ, Kerkhof L (2007) Biogeography of actinomycete communities and type II polyketide synthase genes insoilcollected in NewJersey and Central Asia. Appl Environ Microbiol 73:2982–2989

43. Ulrich A, Becker R (2006) Soil parent material is a key determinant of the bacterial community structure in arable soils. FEMS Microbiol Ecol 56:430–443

44. Zhou J, Beicheng X, Huang H, Treves DS, Hauser LJ, Mural RJ, Palumbo A, Tiedje JM (2003) Bacterial phylogenetic diversity and a novel candidate division of two humid region sandy surface soils. Soil Biol Biochem 3:915–924

45. Zhou J, Bruns M (1996) DNA recovery from soils of diverse composition. Appl Environ Microbiol 62:316–322

Chapter 4 Comparing the Type I polyketide synthase pathways of soil and street sediments.

Previously published as: Patrick Hill, Jörn Piel, Stéphane Aris-Brosou, Václav Krištůfek, Christopher N Boddy, Lubbert Dijkhuizen, Habitat-specific type I polyketide synthases in soils and street sediments, *Journal of Industrial Microbiology and Biotechnology*, Volume 41, Issue 1, 1 January 2014, Pages 75–85.

Actinomycetal isolates from soil have been a source of medicinally useful PKSI producers. Results from Chapter 3 showed that Actinomycetes are a larger fraction of the total bacterial population in street dust compared to soil. This study attempted to see how this difference in Actinomycetal distribution affects the distribution of PKSI operons in the metagenome.

Soil samples were from a smaller set of samples than chapter three but which ranged from Tropical to Arctic climates. Street sediments were from Canada, Europe and Pakistan. DNA was extracted using the "Soft method" described in Chapter 2 and amplified with PKSI specific primers. These primers amplified two domains of PKSI synthase genes. The Ketosynthase domain has been shown to cluster phylogenetically, the Acyl Transferase domain by substrate specificity.

Conclusions

This study found several PKSI sequences that were specific to either soil or streets and cosmopolitan (i.e. present in sites that were distant from each other). The two sequences found in street dust both appeared to by Actinomycetal. Half of sequences found in soil came from a cluster that could not be identified unambiguously but was most similar to Myxobacterial PKSI operons. These sequences were similar enough to have come from a single PKSI pathway. Several other sequence clusters were found that appeared to be from other groups than the Actinomycetes. These non Actinomycetal sequences appeared to be more novel that the Actinomycetal sequences with novel motifs and more distant matches to GenBank.

In spite of the relatively low fraction of Actinomycetes in soil bacterial populations soil might be a more promising target for bioprospecting than street dust.

Role of Ph.D. candidate in Chapter 4

The Ph.D. candidate carried out:

-All sampling with the exception of the Resolute Bay, Chelton Beach and Pahlava soils, the Budapest street dust and Vermicompost samples.

-All laboratory work.

-All data analysis.

-writing the first version of the manuscript.

Habitat-specific type I polyketide synthases in soils and street sediments

Patrick Hill · Jörn Piel · Stéphane Aris-Brosou · Václav Krištufek · Christopher N. Boddy · Lubbert Dijkhuizen

Abstract

Actinomycetes produce many pharmaceutically useful compounds through type I polyketide biosynthetic pathways. Soil has traditionally been an important source for these actinomycete-derived pharmaceuticals. As the rate of antibiotic discovery has decreased and the incidence of antibiotic resistance has increased, researchers have looked for alternatives to soil for bioprospecting. Street sediment, where actinomycetes make up a larger fraction of the bacterial population than in soil, is one such alternative environment. To determine if these differences in actinomycetal community structure are reflected in type I polyketide synthases (PKSI) distribution, environmental DNA from soils and street sediments was characterized by sequencing amplicons of PKSI-specific PCR primers. Amplicons covered two domains: the last 80 amino acids of the ketosynthase (KS) domain and the first 240 amino acids of the acyltransferase (AT) domain. One hundred and ninety clones from ten contrasting soils from six regions and nine street sediments from six cities were sequenced. Twenty five clones from two earthworm-affected samples were also sequenced. UniFrac lineage-specific analysis identified two clades that clustered with actinomycetal GenBank matches that were street sediment-specific, one similar to the PKSI segment of the mycobactin siderophore involved in mycobacterial virulence. A clade of soil-specific sequences clustered with GenBank matches from the ambruticin and jerangolid pathways of *Sorangium cellulosum*. All three of these clades were found in sites >700 km apart. Street sediments are enriched in actinomycetal PKSIs. non-actinomycetal PKSI pathways may be more chemically diverse than actinomycetal PKSIs. Common soil and street sediment PKIs are globally distributed.

Keywords: natural product discovery · Polyketides · Urban microbiology · Bioprospecting · Mycobactin

Introduction

Type I polyketide synthases (PKSIs) produce many and varied bioactive microbial secondary metabolites [11]. Many of these metabolites have been developed into pharmaceuticals including the macrolide antibiotics and the newly approved anti-cancer drug Ixabepilone [22]. Most of these polyketide antibiotics were discovered through culturing actinomycetes, a subclass of the Actinobacteria from soils. This approach worked well from the 1950s until the late 1980s, however, since then, antibiotic discovery has slowed, with serious implications for drug development [21, 38].

Several authors have suggested soil of the hyper-arid Atacama Desert as a new source of actinomycetal secondary metabolites. The actinomycetal fraction of 16S libraries from soils is normally between 0 and 19 % [18]; in soils of the Atacama Desert, actinomycetes can make up over 90 % of the bacterial community [6] and these actinomycetes have been the source of novel secondary metabolites such as the chaxamycins [26, 31]. Dust found on the street surfaces or that collects in between cobblestones may be another new source of actinomycetal PKSIs. Street sediments support a broad range of bacterial communities [19] and are enriched in actinomycetes compared to soils, with some 16S clone libraries being more than 90 % actinomycetal [16]. Even if the bacterial communities of street sediments are largely actinomycetal, this may not mean that street sediments are a good alternative to soil for PKSI bioprospecting. The PKSI genes in the actinomycetal-rich street sediment communities may also be in soil.

Actinomycetal distribution may not be the best measure of PKSI distribution. Sequencing of amplicons from the conserved beta ketosynthase (KS) domain of PKSI genes of environmental DNA from soil has found novel PKSI KS diversity ascribed to the Proteobacteria, Firmicutes, and Chloroflexi [7, 24, 27]. We compared PKSI distribution in soils and street sediments by lightly sampling (215 sequences) a

broad range of samples (n = 21) to determine if there are PKIs genes that are specific to soil and street sediment. Previous work has characterized PKSIs in the environment using the KS domain, which clusters by phylogeny rather than function [12, 17]. We used a primer pair, which amplifies two PKSI domains, the last 20 % of the KS domain and the first 80 % of the acyltransferase (AT) domain, which clusters by substrate specificity [3]. These primers have previously been used to characterize soil environmental DNA [9].

Methods

Sampling and DNA extraction

Ten soil samples were taken from forest, grassland, and cultivated soil in arctic tundra to tropical savannah climates. Nine street samples were from six cities in Canada (n = 1), Europe (n = 4), and Pakistan (n = 1). Street and soil samples are named using the country or city sampled, land use, and site. Thus, Czpasture-Palava was taken from a grassland soil in the Palava protected area of the Czech republic; Brussels-cobblestones-Boucher is material from between cobblestones in Boucher Street, Brussels.

To eliminate the possibility of a sampling bias towards sequences present in all animal affected environments but absent from soils, we also sampled two earthworm-affected sites: the gut contents of *Martiodrilus heterostichon* from Rozo Colombia (Rozoworm gut-Martiodrilus) and vermicomposted cattle manure (Cz-vermicompost). Sampling sites are described in tables 4.1 and 4.2. Sampling method and DNA extraction followed Hill et al. [16].

Table 4.1 Soil samples

Area	Land use	Parent material	Identifier	pH	Organic matter (%)	Clay (%)	Season	No. of seq
Ottawa, Canada	Pine plantation	Glacial moraine	Cdnforest-Kemptville*	5.8	2.8	9.9	Summer	3
Chelton Beach Park, P.E.I. Canada	Scattered grassland	Beach sand	Cdnbeach-PEI	4.7	3.7	4.5	Summer	8
Resolute Bay, Nunavut, Canada	Tundra	Dolomite	Cdntundra-Resolute	8.3	4.3	7.3	Summer	18
Česke, Budéjovice, Czech Republic	Maize	Vltava floodplain	Czcultivated-Plana*	6.1	2.9	15.0	Summer	11
	Church garden	Unknown	CeskeBudejovice-Paradise*	7.5	2.8	3.0	Winter	8
	Native grassland	Limestone	Czpasture-Palava*	7.2	5.0	20.0	Summer	9
Cali, Colombia	Native grassland	Inactive alluvial fan	Colpasture-Pance*	4.3	1.1	18.3	Dry season	10
	Primary forest	Cauca River floodplain	Colforest-Hatico*	7.7	7.0	25.0	Dry season	17
	Bamboo forest		Colforest-CIAT*	6.3	5.7	33.5	Wet season	7
Budapest, Hungary	Scrub forest, Citadel Park	Limestone	Hungforest-Citadel*	7.3	18.6	5.0	Winter	12

* Samples previously used in Hill et al. [16]

Table 4.2 Street sediment and earthworm affected samples

Area	Sediment	Identifier	pH	Organic matter (%)	Clay (%)	Season	No. of seq
Brussels, Belgium	Cobblestones Rue, Boucher	Brussels-cobblestones-Boucher*	7.0	2.9	4.0	Winter	9
Byward market/Rideau centre, Ottawa, Canada	Street bricks, William Street by Dubliner pub.	Ottawa-cobblestones-Dubliner	7.3	18.5	0.0	Winter	16
	Pedestrian crosswalk, Rideau Centre	Ottawa Street dust-Rideau	7.7	7.5	0.1	Winter	7
Central Česke Budéjovice, Czech Republic	Cobblestones, Česka Street	Česke Budéjovice cobblestones-Česka *	7.5	3.0	3.0	Spring	10
	Street dust near Koh-i-noor pencil factory	Česke Budéjovicestreetdust-Koh-i-Noor*	7.3	3.3	2.0	Spring	4
Budapest, Hungary	Pavement crack, Terez Korut, near West Railway station	Budapest pavement-Terez Korut	6.8	13.8	0.0	Summer	13
Left Bank, Paris, France	Cobblestones, Rue du Sabot, Rive Gauche nightclub	Paris cobblestones-Rive Gauche*	7.3	9.1	4.0	Summer	10
	Cobblestones, Café Procope	Paris cobblestones-café*	7.0	10.0	6.0	Summer	6
Faisalabad, Pakistan	Street dust, clock tower, central Faisalabad	Faisalabad Street dust-Clock*	6.8	12.0	2.0	Dry season	12
Cattle manure, vermicomposting system, Czech Republic	Vermicomposted (*Eisenia andrei*) manure	Czech-Vermicompost*	7.2	32.1	n/d	n/a	15
Gut contents of Colombian earthworm from microcosm	*Martiodrilus heterostichon* gut contents	Rozoworm gut-Martiodrilus*	Sample too small for analysis				10

n/d not determined as too much organic carbon in sample to be practicable

PCR amplification

The degenerate PKSI-specific primers and reaction conditions of Ayuso–Sacido and Genilloud [3] were used (K1F 5′-TSAAGTCSAACATCGGBCA-3′ M6R 5′-CGCAGGTTSCSGTACCAGTA-3′) with an annealing temperature of 58 °C.

PCR gave two amplification products, a 1,100–1,390-bp band, which included sections of the AT and KS domains, and a 650–700-bp non-PKSI band (data not shown). PCR products were run on a 1 % agarose gel and the larger band was excised and extracted using the QIAquick gel Extraction Kit (Hilden, Germany).

Cloning and Sequencing

Gel-purified PCR products were cloned directly into the Promega pGEM-T easy cloning system. Ligation reactions were plated on X-gal and white colonies picked and used in colony PCR using the T7/SP-6 vector primers. Colony PCR products were sent to gatc Biotech (Konstanz, Germany) or Beckman Coulter genomics (Danvers, Ma, USA) for double-ended sequencing. The entire PCR product of 215 clones from 21 samples were sequenced (see tables S1 and S2 for amplicon details). Nucleotide sequences were deposited in GenBank with the following accession numbers: Mbt sequence from street sediments KF764484-KF764498, other street sediment sequences KF781465-KF781488, KF826392-KF826435, soil sequences KF781438-KF781464, KF826317-KF826391, Earth worm affected sequences KF826436-KF826458, AT domain only sequences KF826459-KF826465.

Sequence Analysis

For six of the 215 sequences (Rozoworm gut-MartiodrillusP5, Czcultivated-PlanaP7, České Budejovice cobblestones-ceska P6, P7, P8, and Paris cobblestones-rive gaucheP1) poor chromatogram quality at the start of the sequence meant that the KS domain could not be analyzed.

Sequences were aligned using Muscle [8] in MEGA, version 5.00 [37] with default parameters. AT domain alignments were curated with GBLOCKS [4] with the minimum length of a block reduced from 10 to 6 and with gap positions allowed. Model selection was performed with Prottest based on the Akaike Information criterion [1], which identified Wag as the best substitution model. Phylogenetic analysis was performed with PhyML with the subtree pruning and regrafting (SPr) tree-searching algorithm [13]. Branch support was estimated using approximate likelihood ratio test (aLRT; [2]).

The AT and KS amino acid trees were analyzed with UniFrac with the significance and lineage-specific analyses (www.bmf2.colorado.edu/UniFrac, [23]). Soil, street sediment, and earthworm-affected environments were defined as three separate environments. UniFrac significance tests were performed within each environment with 100 permutations.

UniFrac lineage statistically significant soils and street sediment clades included sequences found between 700 and 10,000 km apart. Five other soil and street sediment clades, which were not UniFrac lineage significant, also included sequences from distant sites. To determine if these clades were narrow enough to represent a single PKSI pathway, their maximum pairwise and overall mean distance was determined with MEGA, version 5.00 [37] using the whole uncurated amplified region of the AT and KS domains. These values were compared with similar values for the methylmalonyl AT domains and their associated upstream KS domains picked from ten known PKSI pathways for the same regions of the KS and at domains.

Individual AT and KS trees were used to define highly supported clades (A5, A6, A8-1, A8-3, and Mbt, Figs. 4.1 and 4.2). AT/KS cophylogenies were plotted with the APE library in R [28]. Tree mismatch was estimated by the Robinson and Foulds [34] distance (RFD) and its significance was computed from the distribution of all-against-all RFDs of bootstrapped trees.

Full-length AT and KS domain amino acid sequences were compared with GenBank using protein BLAST. AT domains were compared with all acyl transferase (Pfam family PF00698) domains downloaded from the Protein FaMily database on April 18, 2012 (pfam.sanger.ac.uk, [30]) containing 10,266 sequences from 2,781 species of Bacteria and 366 Eukaryotes.

Results

Sequence length and GC ratios

A total of 215 amplicons were sequenced from 21 different environmental DNA samples. GC content of environmental sequences ranged from 61 to 75 %. these values were in the range of proteobacterial (58–76 %) and actinomycetal (66–77 %) matches but exclude Firmicutes (35 %) and Cyanobacteria (43–51 %) as likely sources of these sequences (table S1).

Amplicon sequence lengths ranged from 1,048 to 1,392 bp. PKSI primers [3] amplified the C-terminal 80 amino acids of the 400–460 amino acid long KS domain, an interspacer region, and the first 240 amino acids of the 280–310 amino acid-long AT domain. The three shortest sequences Brussels-cobblestones-Boucher P2, 4, and 9 had truncated at domains of 473 to 574 bp (versus >700 bp of most sequences). KS (Fig. 4.2) and AT (Fig. 4.1) domains were analyzed separately as they have been shown to cluster differently. Interspacer regions were too variable in length and composition for phylogenetic analysis (data not shown).

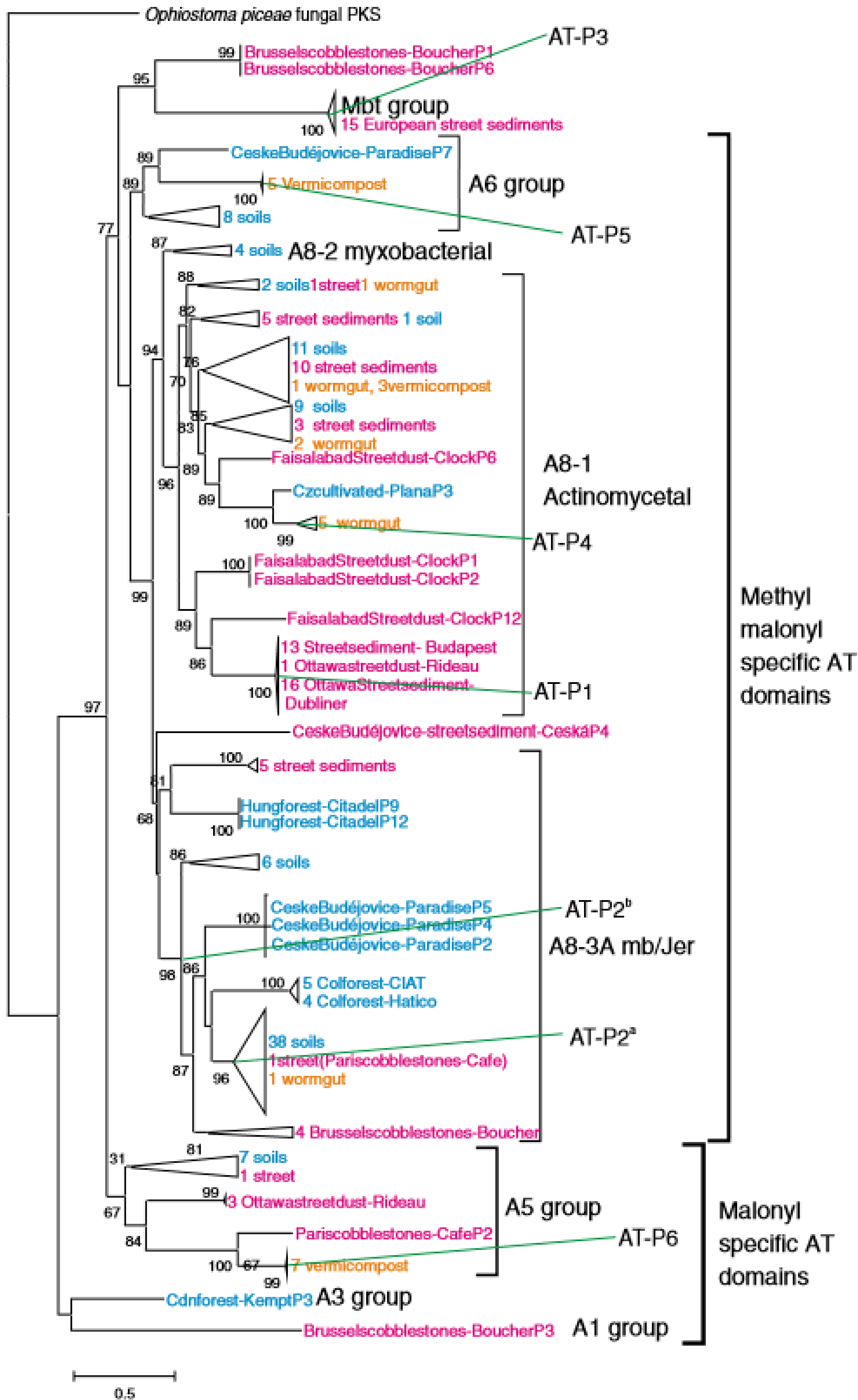
AT and KS tree topology

Clusters in the AT and KS trees are classified using Jenke–Kodama et al.'s [17] system. Ginholac et al. [12] found four at domain clusters: a malonyl-coa, a methylmalonylcoa selective cluster, and two unresolved clusters (set A and B). Jenke–Kodama et al. [17] further refined this classification into eight AT groups: A1-A8. Groups A1-A5 are malonyl-coa-specific, groups A6 and A8 are methylmalonyl-coa

specific, and the A7 group contains sequences for mycobacterial cell wall lipids. Sets A and B are groups A5 and A6.

Most (173) of the 215 sequenced amplicons had methylmalonyl-coa-specific AT domains from the A8 (159) and A6 (14) groups (Fig. 4.1). A8 AT domains consisted of three subgroups: the A8-1 (86), A8-2 (6) clusters clustered with actinobacterial and myxobacterial GenBank sequences, respectively (Figs. S1 and S3) and the A8-3 (69) subgroup, which clustered with matches from the ambruticin (amb) and jerangolid (jer) pathways from *Sorangium cellulosum* [20] (Fig. S2). Twenty-one malonyl-coa-specific at domains were from the A1 (1), A3 (1), and A5 (19, Fig. S4) groups. Fifteen sequences clustered with PKSI in the mixed NRPS/PKS pathway of the mycobacterial siderophore mycobactin (Mbt, Fig. S5) [5].

Fig. 4.1 Rooted maximum likelihood phylogenetic amino acid tree of AT domains. Samples are colour coded as: soils (blue), earthworm-associated sediments (orange), street sediments (red). WAG substitution model used. Branch support values at nodes are approximate likelihood-ratio test for branches (aLRT) values. The scale bar represents 0.5 amino acid substitutions per position. Mbt refers to the mycobactin siderophore. Nodes identified as significant by UniFrac lineage-specific analysis are P1, P2, P2, P5, P6, and P7. p values and actual/expected values are: AT-P1 $p = 6.39 \times 10^{-11}$ Soils-0/14.37 Street sediments-30/12.14 Earthworm sediments-0/3.49, AT-P2A $p = 8.83 \times 10^{-9}$ Soils-38/19.16 Street sediments-1/13.19 Earthworm sediments-1/4.65, AT-P2B $p = 2.22 \times 10^{-3}$ Soils-18/8.17 Street sediments-4/10.81 Earthworm sediments-0/3.02 (note this analysis was run with descendants from P2A removed), AT-P3 $p = 2.96 \times 10^{-4}$ Soils-0/7.19 Street sediments-15/6.07 Earthworm sediments-0/1.74, AT-P4 $p = 4.92 \times 10^{-3}$ Soils-0/2.40 Street sediments-0/2.02 Earthworm sediments-5/0.58, AT-P5 $p = 6.23 \times 10^{-3}$ Soils-0/2.40 Street sediments-0/2.02 Earthworm sediments-5/0.58, AT-tP6 $p = 4.36 \times 10^{-5}$ Soils-0/3.35 Street sediments-0/2.83 Earthworm sediments-7/0.81


Ophiostoma piceae fungal PKS
Brusselscobblestones-BoucherP1
Brusselscobblestones-BoucherP6
AT-P3
Mbt group
15 European street sediments
CeskeBudéjovice-ParadiseP7
5 Vermicompost
8 soils
A6 group
AT-P5
4 soils
A8-2 myxobacterial
2 soils1street1 wormgut
5 street sediments 1 soil
11 soils
10 street sediments
1 wormgut, 3vermicompost
9 soils
3 street sediments
2 wormgut
FaisalabadStreetdust-ClockP6
Czcultivated-PlanaP3
5 wormgut
A8-1
Actinomycetal
AT-P4
FaisalabadStreetdust-ClockP1
FaisalabadStreetdust-ClockP2
FaisalabadStreetdust-ClockP12
13 Streetsediment- Budapest
1 Ottawastreetdust-Rideau
16 OttawaStreetsediment-
Dubliner
AT-P1
Methyl
malonyl
specific AT
domains
CeskeBudéjovice-streetsediment-CeskáP4
5 street sediments
Hungforest-CitadelP9
Hungforest-CitadelP12
6 soils
CeskeBudéjovice-ParadiseP5
CeskeBudéjovice-ParadiseP4
CeskeBudéjovice-ParadiseP2
AT-P2b
A8-3A mb/Jer
5 Colforest-CIAT
4 Colforest-Hatico
38 soils
1street(Pariscobblestones-Cafe)
1 wormgut
AT-P2a
4 Brusselscobblestones-Boucher
7 soils
1 street
3 Ottawastreetdust-Rideau
Pariscobblestones-CafeP2
7 vermicompost
A5 group
AT-P6
Malonyl
specific AT
domains
Cdnforest-KemptP3
A3 group
Brusselscobblestones-BoucherP3
A1 group
0.5

Jenke–Kodama [17] classified KS domains into the K1 to K7 groups. K1 is a mixed non-ribosomal peptide synthase (NRPS)/polyketide synthase (PKS) group and K2 contains loading KS domains. The remaining groups did not cluster by KS function but by phylogeny, the K3 and K5 groups being Cyanobacterial, the K4 and K6 groups mixed myxobacterial and actinomycetal, and the K7 group actinomycetal.

The actinomycetal Mbt and A8-1 subgroups both corresponded to a single KS domain clade, an Mbt cluster and the K7-1 KS cluster. The A8-3, A5, and A6 at domains each corresponded to several different KS domain clusters (Supporting data S1–S5). These clusters were often not found in Jenke–Kodama et al.'s classification and in several cases did not include any GenBank sequences. These new KS clusters (K-A to K-G) are shown in Fig. 4.2. Most (48) of the environmental sequences from the A8-3 group clustered in a K7-2 subgroup, which did not contain any sequences from GenBank, nine were also found in the K-G subgroup. Lastly, neither the AT nor KS domains from Brusselscobblestones-Boucher P1 and P6 clustered with GenBank sequences.

Fig. 4.2 Rooted maximum likelihood phylogenetic amino acid tree of KS domains. Samples are colour coded as: soils (blue), earthworm-associated sediments (orange), street sediments (red). WAG substitution model used. Branch support values at nodes are aLRT values. The scale bar represents 0.1 amino acid substitutions per position. Clusters are labeled with the JenkeKodama classification of the KS domain except in the case of Mbt which refers to the mycobactin siderophore. When KS domains do not match the Jenke-Kodama classification and are not Mbt they are labeled with letters. Clusters with no GenBank members annotated with unfilled triangle. Nodes identified as significant by UniFrac lineage-specific analysis are P1, P2c, P2d, P, P6, and P7. p values and actual/expected values are:

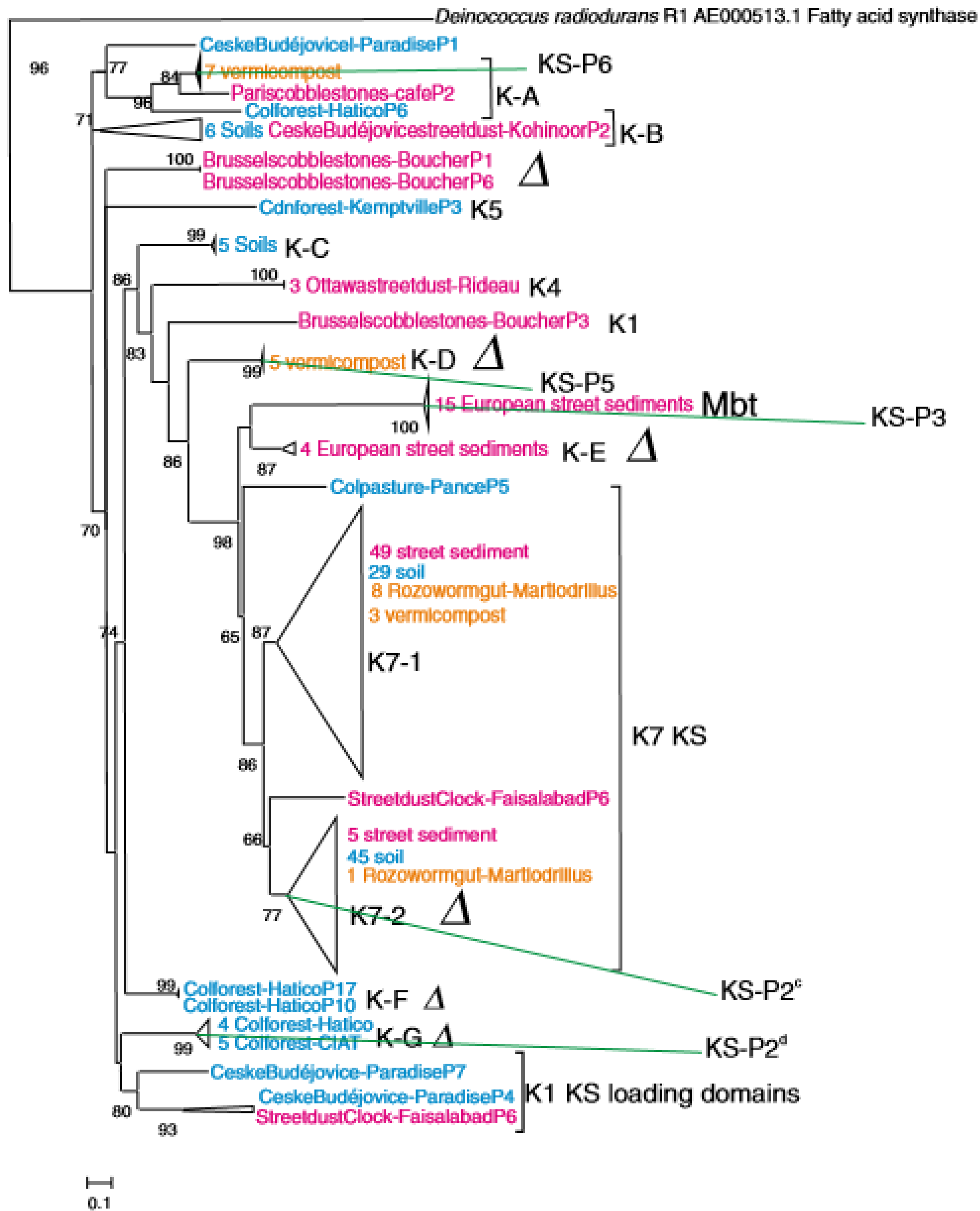
KS-P1 Found within A8-1/A8-2 AT not shown. $p = 1.47 \times 10^{-11}$ Soils-0/14.64 Street sediments-30/11.91 Earthworm sediments-0/3.44,

KS-P2c $p = 4.70 \times 10^{-8}$ Soils45/24.89 Street sediments-5/20.25 Earthworm sediments-1/5.86, KSP2d $p = 8.41 \times 10^{-2}$ Soils-9/4.39 Street sediments-0/3.57 Earthworm sediments-0/1.03,

KS-P3 $p = 1.00 \times 10^{-4}$ Soils-0/7.32 Street sediments-15/5.96 Earthworm sediments-0/1.72,

KS-P5 $p = 1.25 \times 10^{-3}$ Soils-0/2.44 Street sediments-0/1.98 Earthworm sediments-5/0.57,

KS-P6 $p = 2.76 \times 10^{-5}$ Soils-0/3.42 Street sediments-0/2.78 Earthworm sediments-7/0.80



Soil and street sediments are distinct PKSI habitats

UniFrac analyses showed that the soil at and KS trees were significantly different from non-soil trees (at and KS; $p < 0.01$) and that street sediment trees were significantly different from non-street sediment trees (at and KS; $p < 0.01$). The earthworm-affected sediment trees, however, could not be differentiated from non-earthworm affect sediment trees (at and KS; p values = 0.20 and 0.24, respectively).

Soil and street sediment-specific clades

UniFrac lineage-specific analysis identified two actinomycetal nodes that were specific to street sediments. The first, within the A8-1 subgroup, contained 30 sequences from three street sediment samples in Ottawa and Budapest (Fig. 4.1. AT-P1; $p = 6.39 \times 10^{-11}$). The KS sequences corresponding to these AT sequences were found in the K7-1 subgroup and were also statistically significant (Fig. 4.2, KS-P1; $p = 1.47 \times 10^{-11}$). The closest matches to these sequences were all actinomycetal (Figs. S1 and S2).

A second clade contained 15 sequences from European cities and was significant for both AT (Fig. 4.1, AT-P3; $p = 2.96 \times 10^{-4}$) and KS domains (Fig. 4.2, KS-P3; $p = 1.00 \times 10^{-4}$). These sequences clustered with proteins from the mixed NRPS/PKS pathway that produces a siderophore, Mbt, needed for virulence in a variety of mycobacterial and Nocardia strains [5]. Unlike other PKSI KS and AT sequences, the Mbt KS and at domains were separated by a stop codon.

AT domains from the A8-3 group were soil-specific at several levels. One of the largest clusters of A8-3 sequences (40 of 69 sequences) is significant for soil (AT-P2a; $p = 1.02 \times 10^{-8}$). Even with this cluster removed from the UniFrac environmental file, a node remains statistically significant for soil (AT-P2b; $p = 1.88 \times 10^{-3}$).

KS domains associated with A8-3 AT domains were also strongly soil-specific. Fifty-one sequences with A8-3 sequences had KS domains in the K7-2 group. The K7-2 group was significantly selected for in soils (Fig. 4.2. KS-P2C; $p = 4.70 \times 10^{-8}$). A second cluster of nine KS domains (KS-G) associated with A8-3 AT domains from two Colombian soils was selected for in soils with a less significant p value (Fig. 4.2. KS-P2D; $p = 8.41 \times 10^{-2}$). Neither of these sequences clustered with any sequences from GenBank (Fig. S2).

The phylogeny of the A8-3 group is unclear. BlASTP matches for the 69 A8-3 sequences show that the closest matches to their at domains are myxobacterial, either from the first at domains in the

amb and jer biosynthetic pathways from two strains of *Sorangium cellulosum* [20] or from a single *Myxococcus xanthus* sequence. However, a single unpublished sequence from an actinomycete isolated from mangrove soil (AEE69401) also clusters in the A8-3 group (supplemental data Fig. S2). KS domains from all of these matches clustered in the K7-1 group, while most KS domains from environmental A8-3 sequences were from the K7-2 and K-G clusters (Fig. 4.2).

Both soil and street sediment-specific clades were found in samples from distant sites. The P1 and AT-P2A/KS-P2C clades that UniFrac lineage analysis identified as statistically significant for soils or street sediments included sequences from sites that were from different continents. The P3 clade included sequences from cities at least 700 km apart in Europe. Five other small clades of either soil or street sediment sequences (clusters C1–C5, table S4; Figs. S1, S2, S3), while not significant for UniFrac lineage-specific analysis, included sequences from either soil or street sediments sites that were from different continents (C3, C4, C5) or from European cities that were at least 700 km apart (C1, C2).

The sequences within each of these widely geographically distributed (cosmopolitan) clades are similar to known PKSI proteins. The overall pairwise divergence within these cosmopolitan clades is between 0.012 and 0.117 for at sequences and 0.00 and 0.167 for KS sequences (table S4). In contrast, for ten known PKSI pathways, the overall pairwise divergence is between 0.00 and 0.614 for at sequences and 0.064 and 0.528 for KS sequences (table S4). Thus, these clades could represent either dissimilar pathways that share similar modules or similar pathways that produce identical or related polyketides products. In all of these cases, at least a single step in the PKSI assembly process is widely distributed over the landscape.

Earthworm-specific clades

UniFrac lineage-specific analysis identified three clades as earthworm-specific. AS each clade was only found in a single sample, it is uncertain whether they were specific to earthworm species

(*Martiodrilus heterostichon*, *Eisenia andrei*, environment (Worm gut, Vermicompost) or some other factor.

Five sequences from Rozoworm gut-Martiodrilus formed a cluster within the actinomycetal A8-1 subgroup that is UniFrac lineage significant (Fig. 4.1, AT-P4; $p = 4.92 \times 10^{-3}$). As the KS domain from one of these sequences was not included in analysis (Rozoworm gut-MartiodrillusP5), this node was not significant for the KS tree.

Two clusters were specific to Cz-vermicompost. The first was in the A6 group (Fig. 4.1, AT-P5; $p = 6.23 \times 10^{-3}$). The corresponding KS sequences were also selected for in vermicompost (Fig. 4.2, KS-P5 $p = 1.25 \times 10^{-3}$). A second cluster in the A5 group was selected for in the AT tree (Fig. 4.1, AT-P6; $p = 4.36 \times 10^{-5}$) and KS tree (Fig. 4.2, KS-P6; $p = 2.76 \times 10^{-5}$).

Sequence comparisons

Amino acid sequences were compared to the GenBank database by BlastP. Two amino acid sequences, Faisalabad street dust-clock P11 and P12, had 89 and 98 % identity to GenBank matches and the percentage identity of the 15 Mbt group sequences to GenBank was between 75 and 80 %. Most (180/209, table S2) full-length amino acid sequences had 65 % or less identity to GenBank. Sequences with A5 AT domains were the most dissimilar (42–53 %).

AT domains catalyze a highly selective self-acylation by malonyl-coa, methylmalonyl-coa, or other malonyl-coa derivatives followed by transfer to the acyl carrier protein (ACP). The specificity of the AT domain is controlled by a set of motifs identified by Haydock et al. [15], modified by Reeves et al. [33] and Yadav et al. [39] and most recently summarized by Smith and Tsai [36]. These motifs are at the 63, 92 and 201 amino acid positions of the AT domain (number based on the AT domain of FabD in Escherichia coli table 3).

Individual motifs of all 215 AT sequences and 99 comparable GenBank sequences were compared to the motif code in table 4.3 (see table S1 for full results). Few of the GenBank hits (30/82) and environmental clones (62/215) matched known, characterized motifs at all three positions (63, 92, and 201). Most AT domains with motifs that corresponded to the Reeves model at all three positions were A8 AT domains (21 GenBank matches and 57 environmental matches). Sequences that did not match known motifs, most often varied at position 63. Sequences from the Mbt, A1, A3, A5, and A6 groups generally did not match the known model for at least one of three motifs; this was also the case for their GenBank matches.

Table 4.3 Reeve's signature model for AT domain motifs.

E. coli FabD residue position	63[a]	92	201
Malonyl specific	ZTX$(T/A)(Q/E0)	GHS(L/V/I)GE	H(A/G)FH or (H/T/V/Y)AFH
Methylmalonyl specific	R(V/A/I)(D/E)VVQ or (R/Q/S/E/D)V(D/E)VVQ	GHS(Q/M)GE	(Y/V/W)ASH

X any amino acid, *Z* any hydrophilic amino acid, *$* any aromatic amino acid

[a] Consensus patterns given with all possible variations in *parentheses*. After Reeves et al. [33], Ginolhac et al. [12], Smith and Tsai [36]

Motifs from environmental clones were compared with the Pfam database. Forty-three environmental sequences contained Reeve's motifs not found in the database; 34 of these were from street sediments or earthworm-affected samples.

Discussion

Our study compared 215 amplicons encoding part of the KS and AT domains of type I PKSI from 21 samples. Previous studies of PKSIs in the environment have amplified and sequenced the KS domain. Having both the AT and KS domains allowed cophylogenetic analysis, which showed discrepancies of clustering between domains. The length of the amplicons meant that next-generation sequencing could not be used, limiting the sequence coverage of any one sample. However, UniFrac lineage-specific analysis can identify habitat-specific sequences at a level of sequencing that does not fully characterize

any one sample. Shallow sequencing across many samples can identify domains common to environments. Strategies that give the most information for the least sequencing will become important if sequencing costs stabilize [14].

Distinct PKSI synthases are specific to soil and street sediments

UniFrac analyses showed that soil and street sediment are different PKSI habitats. This is probably as street sediments are enriched in Actinomycetes [16]. UniFrac lineage analysis identified two clades that clustered with actinobacterial sequences from street sediments. Thirty sequences with A8-1 AT and K7-1 KS domains were found in Ottawa and Budapest (AT-P1, KS-P1). A second cluster of 15 sequences related to the Mbt group of siderophores for mycobacterial virulence was found in European streets (AT-P3, KS-P3).

Actinomycetal distribution may not be a good guide to PKSI distribution. One hundred and eight sequences (50.2 %) with A5, A6, A8-2, and A8-3 at domains do not cluster with actinomycetal GenBank matches. These sequences are probably not from actinomycetes but we cannot say which taxon they are from.

The largest of these clusters of uncertain origin is the A8-3 at domains that is soil-specific (AT-P2) and make up 56 % (58/103) of the amplicons from soil. The A8-3 at domains cluster with a *Myxococcus xanthus* sequence, domains from the start of the amb and jer pathways from two strains of *Sorangium cellulosum* (myxobacteria) and a single actinobacterial isolate from a Chinese mangrove soil. However, most KS domains from this group do not cluster with any GenBank matches.

Most GenBank sequences that cluster with the A6 and A8-2 sequences are myxobacterial. A5 sequences cluster with GenBank sequences from the Firmicutes, cyanobacteria, Gamma proteobacteria, myxobacteria, Planctomycetes, and soil metagenomic clones. Comparison of GC ratios shows that the A5 sequences are unlikely to be gammaproteobacterial, cyanobacterial, or firmicute (tables S1, S3; Fig.

S4). Even when PKSI domains consistently cluster with a single bacterial phylum with a similar GC content it is not certain that they are from that phylum. For example, Parsley et al. [29] found no consistent homology to any one bacterial phyla over the insert length of several soil metagenomic PKSI clones.

These putatively non-actinomycetal sequences may produce more chemically diverse compounds than better characterized actinomycetal PKSIs. Sequences with the actinomycetal A8-1 AT domain/K7-1 KS domain sequences were more likely to have known reeve's motifs. Sequences from the A8-3, A5, and A6 at domains contained wider ranges of KS domains, some of which did not cluster with any matches from GenBank. The A5 and A6 groups also contained many novel and non-standard Reeve's motifs (table S1), and especially for the A5 group, lower percentage identities to GenBank (table S2).

Previous studies using PKSI KS domain-specific primers on soil DNA ascribed most of the PKSI diversity to non-actinomycetes in a rhizosphere [40], a marine sediment [41] and a broad range of soils [7, 24, 27]. Two studies of soil metagenomic libraries have also found non actinomycetal PKSI synthases in soil [12, 29]; in one case, these clones appeared to be acidobacterial.

In contrast, Reddy et al. [32], using pyrosequencing of amplicons from PKSI KS primers found that most (89 %) domains amplified from desert soils from Arizona, California, and Utah were actinomycetal and few (0.02 %) were proteobacterial. Desert soils may have bacterial communities that are more actinomycetal than other soils [6, 10, 25, 35], like street sediments and unlike most soils they contain few acidobacteria [16, 25].

Many common environmental PKSIs are habitat-specific but widely distributed

Several soil or street sediment-specific clades included sequences from sites hundreds to thousands of kilometers from each other. The best example of this was a sub-clade of the A8-3 group (AT-P2a; Fig. 4.1) with sequences from the Canadian arctic, the Czech Republic, and Colombia. In total,

96 of the 190 soil and street sediments sequences were in clades with sequences from two or more sites that were over 700 km from each other. Overall, the mean distance of these clades is low enough for each of them to be from a single PKSI pathway (table S4).

This suggests that many common polyketides are specific to particular environments such as soil or street sediments but cosmopolitan (i.e., found in a particular environment wherever that environment may be). If this is true, efficient polyketide discovery will depend on identifying environments that have their own specific PKSIs rather than extensively sampling a single environment such as soil. Identifying these environments may not be straightforward; the two earthworm-affected samples contained different sequences while a single PKSI sequence predominated in street samples from Ottawa and Budapest but was not found in other street samples. Finding new polyketides will mean surveying anthropogenic habitats such as street sediment that are rarely studied as they neither correspond to biodiverse eukaryotic habitats nor provide ecosystem services.

Again, our results differ from those of Reddy et al. [32] who found that even at 85 % similarity, 90 % of PKSI sequences were only found in a single sample. Reddy et al. compared three soils far closer and more similar to each other than our sites were. This difference could be due to methods and/or the nature of desert soils.

Acknowledgments

Dr. Meng Wang was kind enough to translate the paper "Screening identifying and functional analysis of Polyketide Synthase I domains from soil and seawater of Yangshan Harbour" into English for us. We thank Dr. A. Poulain for providing the sample from resolute Bay, canada, Dr. Rashid Nazir for providing the Faisalabad clock tower sample, Dr. Declan Hill for providing the street sample from Budapest, Hungary, and Keavin M. Stanford-Finnerty for the beach sand sample from Prince Edward Island, Canada. Our laboratory work was helped by Jean Bjornson of the Geography Dept of the University of Ottawa who assisted with soil analysis and Daniël van Middelkoop and Sarina de Roos who carried out cloning and sequencing as part of a work experience program for the School for Biomedical laboratory techniques in groningen, the netherlands. Funding for this work was provided by the NATO Programme Security through Science ESP.EAP.CLG 981785, the Ministry of Education, Youth and Sports of the Czech Republic (LC06066, 2B06154), the Research Plan of the Institute of Soil Biology (AV 0Z 6066 0521), the

Natural Sciences and Engineering Council of Canada, the Ministry of Research and Innovation, and the University of Ottawa.

References

1. Abascal F, Zardoya R, Posada D (2005) Prottest: selection of best-fit models of protein evolution. Bioinforma 21: 2104–2105

2. Anisimova M, Gascuel O (2006) approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol 55:539–552

3. Ayuso–Sacido A, Genilloud O (2005) New PCR primers for the screening of NRPS and PKS-I systems in Actinomycetes: detection and distribution of these biosynthetic gene sequences in major taxonomic groups. Microb Ecol 49:24–49

4. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17:540–552

5. Chavadi SS, Stirrett Kl, Edupuganti UR, Vergnolle O, Sadhanandan G, Marchiano E, Martin C, Qiu WG, Soll CE, Quadri lE (2011) Mutational and phylogenetic analyses of the mycobacterial mbt gene cluster. J Bacteriol 193:5905–5913

6. Connon S A, Lester ED, Shafaat HS, Obenhuber DC, Ponce A (2007) Bacterial diversity in Hyperarid Atacama Desert soils. J Geophys Res 112: g04S17, doi: 10.1029/2006Jg000311

7. Dong XY, Wang lH, Sun MJ, Zong Y, Jiao YL, Jiao BH (2008) Screening, identifying and function analysis of polyketide synthase I domains from soil and seawater of Yangshan Harbor. Microbiology 35:1359–1366

8. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

9. Faizal I, Lestari R, Kurnia F, Latif A, Hadianto D, Kusumawati N, Lachmawati I, Marwoto B, Purbowasito W (2008) Polymorphism analysis of polyketide synthase gene from Actinomycetes genome DNA of Taman Nasional Gunung Halimun soil by using metagenome method. J Biotechnology Res Tropical Reg 1: biotechindonesia.org/journal/jbr/jbr-2008-00-01/jrb-1-08-2.pdf

10. Fierer N, Leff JW, Adams BJ, Nielsen UN, Bates St, Lauber CL, Owens S, Gilbert JA, Wall DA, Caporaso JG (2012) Cross-biome Metagenomic analyses of soil microbial communities and their functional attributes. PNAS 109:21390–21395

11. Fischbach Ma, Walsh CT (2006) Assembly-line enzymology for polyketide and non-ribosomal peptide antibiotics: logic, machinery, and mechanisms. Chem Rev 106:3468–3496

12. Ginolhac A, Jarrin C, Gillet B, Robe P, Pujic P, Tuphile K, Bertrand H, Vogel TM, Perrière G, Simonet P, nalin R (2004) Phylogenetic analysis of polyketide I domains from soil metagenomic libraries allows selection of promising clones. Appl Environ Microbiol 70:5522–5527

13. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O (2010) new algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyMl 3.0. Syst Biol 59:307–321

14. Hall N (2013) After the gold rush. Genome Biol 14:115–117

15. Haydock SF, Aparicio JF, Molnár I, Schwecke T, Khaw lE, König A, Marsden AFA, Galloway IS, Staunton J, Leadlay PF (1995) Divergent sequence motifs correlated with the substrate specificity of (methlyl) malonyl-coa: acyl carrier protein during transcyclase domains in modular polyketide synthases. FEBS Lett 374:246–248

16. Hill P, Krištufek V, Dijkhuizen l, Boddy C, Kroetsch D, van Elsas D (2011) Land use intensity controls actinobacterial community structure. Microb Ecol 61:286–302

17. Jenke-Kodama H, Sandmann A, Muller R, Dittmann E (2005) Evolutionary implications of bacterial polyketide synthases. Mol Biol Evol 22:2027–2039

18. Jensen PH (2006) Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. Appl Environ Microbiol 72:1719–1728

19. Johnson MJ, Lee KY, Scow KM (2003) DNA fingerprinting reveals links among agricultural crops, soil properties, and the composition of soil microbial communities. Geoderma 114:279–303 2

20. Julien B, Tian ZQ, Reid R, Reeves CD (2006) analysis of the ambruticin and jerangolid gene clusters of *Sorangium cellulosum* reveals unusual mechanisms of polyketide biosynthesis. Chem Biol 13:1277–1286

21. Lam KS (2007) New aspects of natural products in drug discovery. Trends Microbiol 15:279–289

22. Lee FYF, Borzilleri D, Fairchild CR, Kamath A, Smykla R, Kramer R, Vite G (2008) Preclinical discovery of Ixabepilone, a highly active antineoplastic agent. Cancer Chemother Pharmacol 63:157–166

23. Lozupone C, Hamady M, Knight R (2006) UniFrac: An online tool for comparing microbial community diversity in a phylogenetic context. BMC Bioinformatics 7:371

24. Luo K, Du G-P, Zhao Z-X, Xie BY, Li D-J (2010) Phylogenetic analysis of type I polyketide synthase and non-ribosomal peptide synthase genes from Mila Mountain in Tibet plateau. J Hunan Agric Uni (Nat Sci) 36:506–511

25. Neilson JW, Quade J, Ortiz M, Nelson WM, Legatzki A, Tian F, Lacomb M, Betancourt JL, Wing RA, Soderlund CA, Maier RM (2012) Life at the hyperarid margin: novel bacterial diversity in arid soils of the Atacama Desert, Chile. Extremophiles 16:553–566

26. Okoro CK, Brown R, Jones AL, Andrews BA, Asenjo JA, Goodfellow M, Bull at (2009) Diversity of culturable actinomycetes in hyper-arid soils of the Atacama Desert, Chile. Antonie Leeuwenhoek 95:121–133

27. Pang MF, Tan G-YA, Abdullah N, Lee C-W, Ng C-C (2008) Phylogenetic analysis of type I and type II polyketide synthase from tropical forest soil. Biotechnology 7:660–668

28. Paradis A (2006) Analysis of phylogenetics and evolution. Springer, Berlin Heidelberg New York

29. Parsley LC, Linneman J, Goode AM, Becklund K, George I, Goodman RM, Lopanik NB, Liles MR (2011) Polyketide synthase pathways identified from a metagenomic library are derived from soil Acidobacteria. FEMS Microbiol Ecol 78:176–187

30. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm l, Sonnhammer Ell, Eddy SR, Bateman A, Finn RD (2012) the Pfam protein families database. Nucleic Acids Res 40:D290–D301

31. Rateb ME, Houssen WE, Arnold M, Abdelrahman MH, Deng H, Harrison Wta, Okoro CK, Asenjo JA, Andrews BA, Ferguson G, Bull A, Goodfellow M, Ebel R, Jaspars M (2011) Chaxamycins A-D, bioactive Ansamycins from a hyper-arid desert Streptomyces sp. J Nat Prod 74:1491–1499

32. Reddy BB, Kallifidas D, Kim JH, Charlop-Powers Z, Feng Z, Brady SF (2012) natural product biosynthetic gene diversity in geographically distinct soil micro biomes. Appl Environ Microb 78:3744–3752

33. Reeves CD, Murli S, Ashley GW, Piagentini M, Hutchinson CR, McDaniel R (2001) Alteration of the substrate specificity of a modular polyketide synthase acyltransferase domain through site specific mutations. Biochemistry 40:15464–15470

34. Robinson DF, Foulds LR (1981) Comparison of phylogenetic trees. Math Biosci 53:131–147

35. Saul-Tcherkas V, Steinberger Y (2011) Soil microbial diversity in the vicinity of a Negev Desert shrub–Reaumuria negevensis. Microb Ecol 61:64–81

36. Smith S, Tsai S-C (2007) The type I fatty acid and polyketide synthases: A tale of two megasynthases. Nat Prod Rep 24:1041–1072

37. Tamura K, Peterson D, Stecher G, Nei M, Kumar S (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

38. Watve Mg, Tickoo R, Jog MM, Bhole BD (2001) How many antibiotics are produced by the genus Streptomyces? Arch Microbiol 176:386–390

39. Yadav G, Gokhale RS, Mohanty D (2003) Computational approach for prediction of domain organization and substrate specificity of modular polyketide synthases. J Mol Biol 328:335–363

40. Zhao B, Gao Z, Shao Y, Yan J, Hu Y, Yu J, Liu Q, Chen F (2012) Diversity analysis of type I ketosynthase in rhizosphere soil of cucumber. J Basic Microbiol 52:224–231

41. Zhao J, Yang N, Zeng R (2008) Phylogenetic analysis of type I polyketide synthase and non-ribosomal peptide synthetase genes in Antarctic sediment. Extremophiles 12:97–105

Chapter 5 Conclusions.

The literature review in chapter one lists five questions in how to sample for polyketides in terrestrial environments. These questions have not been resolved and the question of sampling remains *"among the most understudied topics in the field"* of antibiotic discovery (Hernandez & Murphy, 2021). In this chapter the experimental results of the chapters two to four are organized around the five questions of sampling raised in the first chapter. Some recent experimental papers which appeared after the first chapter was written are also briefly discussed.

1-Most of the available data about bacterial communities in the environment come from molecular biology studies. The results of these studies depend on the methods that they use. This means that it is uncertain if we can compare results between studies making the scientific literature an unreliable guide for bioprospectors.

Most molecular biology studies of bacteria in the environment use bead beating to lyse bacteria when extracting DNA and amplify this DNA with a single primer pair. In chapter two we compared bead beating and an enzymatic lysis extraction using a range of primers, so the effects of method are larger than most comparisons of studies. There was no evidence that either extraction or primers and any effect when we compared bacterial biomes (e.g. faeces vs soil, street dust vs faeces). Similar results were found in a paper uploaded to BioRxiv the same day as ours. (Shaffer et al., 2022) compared six bead beating extraction methods extractions from a range of environments (e.g. soil, faeces, skin) and found that the effect of extraction method was orders of magnitude less than sample.

While extraction and primer effects may not distort comparisons of different bacterial biomes, they may change our view of differences between bacterial communities in similar samples. The effects of method on amplicon sequencing studies was recently discussed by Alteio et al (Alteio et al., 2021). As well as discussing the primers and extraction method, Alteio et al recommend spiking samples with DNA standards before extraction. This is done to determine the total amount of bacterial DNA extracted.

Without this, it is impossible to know if changes in the fraction of a bacterial group are absolute (it is growing or dying) or relative (other groups are growing or dying more).

Primer effects may be particularly important in the new field of network analysis (Guseva et al., 2022). Co-occurance and absence of groups are used in network analysis to determine keystone species. However, we found in both chapters 2 and 3 that while different 16S primer pairs may find similar changes in bacterial communities they will differ in the phylogenetic level and group where they find it. A bacterial genus, OTU or ASV that network analysis of amplicons identifies as a keystone may be the one that a primer pair preferentially amplifies in a keystone bacterial order or family.

Extraction/primer and other method artifacts should not limit bioprospecting for antibiotics though. From the literature review in the first chapter we saw that the most important task to be done in bioprospecting is comparing contrasting biomes. The problem of determining the absolute and relative changes in bacterial community discussed by Alteio et al (2021) are less important for bioprospecting than microbial ecology. For bioprospectors the relative fraction of bacterial group in the whole population is more important than the absolute size of the population, as the object is to reduce screening of isolates or clones.

Next Generation amplicon sequencing is relatively time consuming and the data that is provides needs a fair amount of time to be turned into useful information. This information is both very detailed and unreliable given primer bias. Amplifiying DNA with a range of primers and then using the amplicons for community fingerprinting is cheaper and faster. Results are less detailed but more reliable (as they come from several primer pairs) and can be quickly interpreted. Information should be good enough to make decisions about further sampling and antibiotic bioprospecting. An example of this approach is given in Chapter 3.

2-While there is over a hundred years of study of how macrobial eukaryotic communities are distributed across the earth, bacterial communities may not follow the same patterns. Environments such as tropical rainforest may have a diverse plant and insect populations, but this does not mean that they will also have diverse bacterial populations.

This thesis presents little direct evidence either way for this. The second chapter which compared extraction methods and primers did find that a ploughed soil was the most diverse soil and acid forest soils the least diverse.

There have been many studies since the first chapter was published which have confirmed that eukaryotic diversity is not a good guide to bacterial diversity. Recently Id and Martiny (Id & Martiny, 2020) compared 16S amplicon sequencing results from the Earth Microbiome project for soils, sediments, marine and inland water, air and biofilms. Bacterial diversity did not match any pattern that eukaryotic diversity would predict.

While soils were the most alpha diverse habitat, agricultural soils were the most diverse, more so than forest or shrubland soils. Bacterial diversity in soils peaked at a neutral pH and a mean annual temperature of 10 °C. However, beta bacterial diversity in soils was less than that of sediments and inland waters. Beta bacterial diversity in agricultural soils was as large as that in other soils.

A sampling plan to capture bacterial diversity based these results would not look much like one that aimed for plant or insect diversity. It would not concentrate on the tropics, it would sample as many or more agricultural soils as ones under native vegetation and would have as many or more water samples as soil samples.

3-Actinomycetes have been the source for most bacterial antibiotics. Actinomycetes distribution may not follow the same patterns as the whole bacterial community. Diverse bacterial communities may not be equally diverse in Actinomycetes.

The sampling design for Chapters 3 and 4 was unusual for their time. Bioprospecting studies usually began with assumptions about which environments were promising, often through analogy with eukaryotic ecology and sampled this environment intensively without comparing it to other

environments (e.g. van Elsas et al., 2008, Visser, Nobre, Currie, Aanen, & Poulsen, 2012). Ecological theories were tested on a single site, and only altered a single variable e.g. pH for Actinomycetes (Jenkins et al., 2009).

Our sampling was over a broad range of soils and non-soil environments. No assumptions were made that actinomycetal distribution would follow patterns of eukaryote or bacterial diversity. Our results support this approach. While pH has been shown to control bacterial community structure in soil, our results from chapter 3 show that actinobacterial populations are controlled by land use. As our soil samples varied in pH, parent material and were from different continents we can be confident that the effect that we found is broadly significant.

Street dust, an environment that is unimportant for ecosystem services and eukaryotic ecology has actinobacterial communities that are enriched and diverse compared to those of soil, the environment where most actinomycetal bioprospecting has been carried out so far.

There is no clear evidence from subsequent literature that either supports or refutes these findings. While there are many studies of soil bacterial communities that cite our study, none of them present evidence that supports our results. Nor are there any studies which compare the actinobacterial communities of a broad range of soils across different land uses.

Since the publication of chapter 3 there has been considerable interest in the microbiology of cities (The MetaSUB International Consortium, 2016). However, most of these studies have looked at urban microbiology from the view of human health and so have sampled surfaces and air which are continually in contact with humans. Subway systems have been particularly well researched. There have been a few studies which found that soils in cities have a larger Actinomycetal fraction than rural soil (Chen, Martinez, Cleavenger, Rudolph, & Barber, 2021)

4-Non Actinomycetal bacteria may also be source of many novel antibiotics. Actinomycetal distribution may not be a reliable guide to Antibiotic production genes in the terrestrial environment.

The PKSI paper in chapter four amplified PKSI amplicons from many different soils that appeared to be from bacteria that are not Actinomycetes. Earlier papers had found this using Sanger sequencing of amplicons from individual soils e.g. (Pang, Tan, Abdullah, Lee, & Ng, 2008), (Luo, Du, Zhao, Xie, & Li, 2010). Later publications using Next Generation Sequencing of amplicons (Benaud et al., 2019) or Metagenomic genomes (Sharrar et al., 2020) have shown that many of the PKSI sequences in soil metagenomes are from other groups such as the Acidiobacteria and Chloroflexi.

In the results of chapter 4 non Actinomycetal PKSI sequences seemed more novel than Actinomycetal PKS sequences. Many have suggested that bioprospectors concentrate on these non Actinomycetal producers e.g (Loureiro, Medema, Oost, & Sipkema, 2018). However, a recent analysis of sequenced genomes found that the Streptomyces, the most exploited order of the Actinomycetes still had a great deal of potential for undiscovered biosynthetic operons (Gavriilidou et al., 2022).

5-It is unclear what controls the biogeography of antibiotic genes. If barriers to dispersion are important then similar environments that are geographically distant will have different antibiotic genes. If not, similar environments will have similar antibiotics wherever they are.

None of the 3 experimental chapters found evidence for distance affecting either 16S or polyketide synthase distribution. Actinobacterial 16S fingerprints clustered by land use with no evidence of geographical clustering even though samples were taken from sites thousands of kilometres from each other. This does not necessarily mean that distance does not have any effect on Actinobacterial community structure, but if there is, it is much less than that of land use.

Chapter 4 identified cosmopolitan PKSI sequences that were found in soils and street dust in sites on different continents. These results can be criticized as it used Sanger Sequencing of amplicons rather than Next Generation Sequencing of environmental DNA (as used in many later studies). Primer and cloning bias as well as very limited sequencing meant that only a small fraction of the PKSI operons

were identified. However, there is no inherent bias for endemic or cosmopolitan sequences in PCR or cloning, so cosmopolitan PKSI pathways are probably common.

It is beyond the scope of this thesis to establish if there are significant barriers to dispersion for polyketide operons in the environment. Answering this question definitively needs deep sequencing as polyketides that appear to be endemic to one set of samples may be present in others at low levels. Many samples would be also needed as it is uncertain what constitutes bacterial habitats. What can be said though is that given the limited resources available to most bioprospectors, it is more efficient to sample many environments in a small area rather than try to sample a single environment exhaustively over a broad scale.

A recent review (Loureiro et al., 2018) pointed out that similar polyketides are found in terrestrial insects and marine sponges and say that for natural products “everything is everywhere”. A comparison of marine and terrestrial natural products found many in both environments (Carroll, Campbell, & Carroll, 2022). Because of this Loureiro et al say that there is little point in governments restricting access for microbial natural product bioprospectors.

Serendipitous Results

The Pseudomonas are an order of natural product producers of secondary importance (Gavriilidou et al., 2022). In chapter 2 it was shown that they are systematically under extracted by the most common DNA extraction method in many soils. The results of chapter 2 may be used to better understand their distribution, and point to soil microaggregates as a target for Pseudomonas bioprospecting.

The AD-3 candidate division was first identified in Sanger sequenced PCR amplicons from a study in of sandy forest soils in New Jersey {Formatting Citation} (Zhou et al 2003). In the Actinobacterial study in Chapter 2, AD-3 sequences were amplified from two acid soils (pH <5) one in Colombia and a

second in Siberia. These four clones clustered with NCBI sequences from soils and sediments with pH <5 so this chapter suggests that they are characteristic of acid soils and sediments.

Since then genomes from this candidate group have been assembled by whole metagenomic sequencing and it has been renamed the Dormibacteraeota (Ji et al., 2017) and confirmed to be acidiophile (Montgomery et al., 2021). Whole metagenomics sequencing of Dormibacteraeota in soil has also shown them to be potential secondary metabolite producers (Sharrar et al., 2020).

References

Alteio, L. V, Joana, S., Canarini, A., Angel, R., Jansa, J., Guseva, K., … Schmidt, H. (2021). A critical perspective on interpreting amplicon sequencing data in soil ecological research. *Soil Biology and Biochemistry*, *160*(April). http://doi.org/10.1016/j.soilbio.2021.108357

Benaud, N., Zhang, E., van Durst, J., Kalaitzis, J. A., Neilan, B. A., & Ferrari, B. C. (2019). Harnessing Long-Read Amplicon Sequencing to Uncover NRPS and Type I PKS Gene Sequence Diversity in Polar Desert Soils Harnessing long-read amplicon sequencing to uncover NRPS and Type I PKS gene sequence diversity in polar desert soils. *FEMS Microbiology Ecology*, *95*(March (4)). http://doi.org/10.1093/femsec/fiz031

Carroll, A. R., Campbell, M. D., & Carroll, A. R. (2022). Natural Product Reports. *Natural Product Reports*, *39*(1), 7–19. http://doi.org/10.1039/d1np00051a

Chen, Y., Martinez, A., Cleavenger, S., Rudolph, J., & Barber, A. (2021). Changes in Soil Microbial Communities across an Urbanization Gradient : A Local-Scale Temporal Study in the Arid Southwestern USA. *Microorganisms*, *9*, 1470.

Gavriilidou, A., Kautsar, S. A., Zaburannyi, N., Krug, D., Müller, R., Medema, M. H., & Ziemert, N. (2022). Compendium of specialized metabolite biosynthetic diversity encoded in bacterial genomes. *Nat. Microbiol*, *7*(May). http://doi.org/10.1038/s41564-022-01110-2

Guseva, K., Darcy, S., Simon, E., Alteio, L. V, Montesinos-navarro, A., & Kaiser, C. (2022). From diversity to complexity : Microbial networks in soils. *Soil Biology and Biochemistry*, *169*(February), 108604. http://doi.org/10.1016/j.soilbio.2022.108604

Hernandez, A., & Murphy, B. T. (2021). Natural Product Reports The need to innovate sample collection and library generation in microbial drug discovery : a focus on academia. *Natural Product Reports*, *38*, 292–300. http://doi.org/10.1039/d0np00029a

Id, K. E. W., & Martiny, J. B. H. (2020). Alpha-, beta-, and gamma- diversity of bacteria varies across habitats. *PLoS ONE*, *15*(9), e0233872. http://doi.org/10.1371/journal.pone.0233872

Jenkins, S. N., Waite, I. S., Blackburn, A., Husband, R., Rushton, S. P., Manning, D. C., & O'Donnell, A. G. (2009). Actinobacterial community dynamics in long term managed grasslands. *Antonie van Leeuwenhoek*, *95*(4), 319–334. http://doi.org/10.1007/s10482-009-9317-8

Ji, M., Greening, C., Vanwonterghem, I., Carere, C. R., Bay, S. K., Steen, J. A., … Ferrari, B. C. (2017). Atmospheric trace gases support primary. *Nature Publishing Group*, *552*, 400–405. http://doi.org/10.1038/nature25014

Loureiro, C., Medema, M. H., Oost, J. Van Der, & Sipkema, D. (2018). ScienceDirect Exploration and exploitation of the environment for novel specialized metabolites. *Current Opinion in Biotechnology*, *50*, 206–213. http://doi.org/10.1016/j.copbio.2018.01.017

Luo, K., Du, G.-P., Zhao, Z.-X., Xie, B., & Li, D.-J. (2010). Phylogenetic analysis of type I polyketide synthase and non-ribosomal peptide synthase genes from Mila Mountain in Tibet plateau. *J Hunan Agric Uni*, *36*, 506–511. http://doi.org/10.3724/SP.J.1238.2010.00506

Montgomery, K., Williams, T. J., Brettle, M., Berengut, J. F., Zhang, E., Zaugg, J., … Sydney, U. (2021). Persistence and resistance : survival mechanisms of Candidatus Dormibacterota from nutrient-poor Antarctic soils. *Environmental Microbiology*, *23*, 4276–4294. http://doi.org/10.1111/1462-2920.15610

Pang, M.-F., Tan, G.-Y. A., Abdullah, N., Lee, C.-W., & Ng, C.-C. (2008). Phylogenetic analysis of type I and type II polyketide synthase from tropical forest soil. *Biotechnology*. Retrieved from http://www.docsdrive.com/pdfs/ansinet/biotech/2008/660-668.pdf

Shaffer, J. P., Carpenter, C. S., Martino, C., Salido, R. A., Minich, J., Bryant, M., … Diego, S. (2022). A comparison of six DNA extraction protocols for 16S , ITS , and shotgun metagenomic sequencing of microbial communities. *BioRxiv*, 1–20.

Sharrar, A. M., Crits-Christoph, A., Meheust, R., Diamond, S., Starr, E. P., & Banfield, J. F. (2020). Bacterial Secondary Metabolite Biosynthetic Potential in Soil varies with Phylum, Depth and Vegetation Type. *MBio*, *11*(3), 1–17.

The MetaSUB International Consortium. (2016). The Metagenomics and Metadesign of the Subways and Urban Biomes (MetaSUB) International Consortium inaugural meeting report. *Microbiome*, *4*(24). http://doi.org/10.1186/s40168-016-0168-z

van Elsas, J. D., Costa, R., Jansson, J., Sjöling, S., Bailey, M., Nalin, R., … van Overbeek, L. (2008). The metagenomics of disease-suppressive soils - experiences from the METACONTROL project. *Trends in Biotechnology*, *26*(11), 591–601. http://doi.org/10.1016/j.tibtech.2008.07.004

Visser, A. A., Nobre, T., Currie, C. R., Aanen, D. K., & Poulsen, M. (2012). Exploring the Potential for Actinobacteria as Defensive Symbionts in Fungus-Growing Termites. *Microbial Ecology*, *63*(4), 975–985. http://doi.org/10.1007/s00248-011-9987-4

Zhou J, Beicheng X, Huang H, Treves DS, Hauser LJ, Mural RJ, Palumbo A, Tiedje JM (2003) Bacterial

phylogenetic diversity and a novel candidate division of two humid region sandy surface soils. Soil Biol Biochem 3:915–924

Appendix 1 Supplemental Data for Chapter 2 Assessing the importance of primers and extraction method on molecular ecology studies.

A comparison of hard and soft direct methods for DNA extraction from soil.

Patrick Hill, Mathieu F Dextraze, David Kroetsch, Christopher N Boddy

Supplemental Data1-Methods

1-Soft lysis DNA extraction method.

The description below explains the method in detail and should be usable for inexperienced laboratory workers.

Reagents

<u>Sample washing</u>

Washing buffer 50 mM Tris/50 mM EDTA pH 8.3

<u>Lysis.</u>

Enzymatic extraction buffer 500 mM NaCl/50 mM Tris/50 mM EDTA pH8.3

SDS extraction buffer 500 mM NaCl/300 mM Succinic Acid/10 mM EDTA pH 5.7

20 % SDS. (Sodium dodecyl sulfate)

Lysozyme

Proteinase-K

<u>Cleaning</u>

10 % CTAB (Cetyl trimethyl ammonium bromide)

5M NaCl

Chloroform

PEG Polyethylene Glycol (6000 or 8000)

T.E. Buffer (10 mM Tris, 1 mM EDTA; pH 8.0)

Phenol/chloroform/isoamylalcohol (25:24:1).

Calcium chloride.

3 M Sodium Acetate (pH-5.2)

70 % ethanol.

Also needed-Fume hood. Bench top and swing rotor centrifuges that can be cooled.

Method

1.1-Sample washing

1 to 2.5-g of soil/sediment are mixed with 45 mL of washing buffer (50 mM EDT/50 mM Tris buffer pH 8.3) in a 50 mL Falcon tube. This is best done by half filling the tubes with buffer, shaking briefly and then adding the rest of the buffer, so that the soil disperses. The tubes are then centrifuged at 2643 g (3500 rpm on a Sorvall Legend RT + centrifuge) at 4 C for one hour.

Enzymes will not work well at pHs below 8 and the buffering power of soils varies from soil to soil. This step is designed to make sure that extraction conditions are as similar from sample to sample and to removes a lot of humic material and metal cations. Vary the amount of sample depending on how dirty (i.e. humic rich) and how much DNA you expect to get out.

Pour off the supernatant and store overnight at -20 C *(do this even if you have time to continue the extraction to make sure that all samples are treated the same way.)*.

1.2-Lysing the bacteria.

1.2.1 Enzymatic lysis at 37 C, pH 8

For each sample that you are extracting add 2.5 mL of Enzymatic extraction buffer (500 mM NaCl, 50 mM Tris, 50 mM EDTA pH 8.3) to a Falcon tube. Then add lysozyme to get a final concentration of 5 mg/ml. Add 2.5 ml of the Enzymatic buffer/lysozyme mix to each sample and incubate for 1 h at 37 °C.

(The Tris and EDTA are to keep the pH at 8, the high NaCl concentration keeps humics from going into solution).

Add 140 µl of 20% SDS to each tube and shake gently. Add 100 µl of Enzymatic buffer for each sample to an Eppendorf tube. Then add proteinase-K to a final concentration of 10 mg/mL. Add 100 µl of the Enzymatic buffer/proteinase-K mix to each tube. Incubate for a second hour at 37 C.

(Proteinase-K works better with SDS)

1.2.2 SDS lysis at 65 C, pH 5.7.

Add 5 ml of an SDS buffer of 500 mM NaCl, 300 mM succinic acid, 10 mM EDTA (pH 5.7) then 700 µl of 20% SDS to each tube. Incubate for 45 minutes at 65°C, gently inverting the tubes twice. Centrifuge at 2643 g for an hour at 4 C, save the supernatant and discard the soil pellet.

Humics go into solution when the pH and temperature are high. At this stage we no longer need the high pH for the enzymes to work and we are raising the temperature so we lower the pH. The low temperature during centrifugation means that a lot of the SDS goes with the pellet. This is a good stopping point as if you leave your supernatant in the fridge at 4C overnight more of the SDS and often some humic impurities settle out.

Up to here it is important that the procedure is the same from sample to sample. Below is how to clean and concentrate the supernatant. The method was developed to get DNA >30 kb from the A horizons of forest soils. If samples are cleaner or DNA size is less important these steps could probably be simplified.

1.3-Cleaning the Supernatant.

1.3.1 CTAB cleaning (in fume hood)

Add 8 ml of sample supernatant to a 50-mL Falcon tube. Then add 1 ml of 5 M NaCl and gently mix. Then add 1 ml of 10% cetyl trimethyl ammonium bromide (CTAB) and gently mix again. Incubate for 30 minutes at 65 ºC.

The order of adding NaCl first before the CTAB is important. At low salinities CTAB binds to DNA and you could lose it in the next step.

Then place the falcon tubes in a box of crushed ice to cool to 4 ºC. Then add around 5 mL of cold (4 C) chloroform to each tube. Shake gently until the mixture looks milky *(for low organic matter sediment)* to chocolate milky (*for organic rich soils*).

Spin the tubes down on a Sorval centrifuge that has previously been cooled to 4 ºC. Usually half an hour at 4 ºC and 2000 g will work, but what matters is that you get a clear separation between the upper

water phase (with your DNA) and the lower chloroform humic phase. Occasionally you will have to take off a messy upper layer and mix it with chloroform a second time before a second centrifugation.

1.3.2 PEG precipitation

Gently pipette off the supernatant after spinning. Transfer it to a 15 mL falcon tube with a volume scale on the side. Use this scale to add half the volume of your supernatant of 50% polyethylene glycol (PEG 8000 or 6000). Mix gently but thoroughly, then store overnight at 4 C. Transfer the supernatant/PEG mix to a 50 mL falcon tube and centrifuge 2-3 hours at 4ºC, (maximum safe speed on a Sorvall centrifuge). Pour off the supernatant, carefully rinse with cold 70% ethanol *(it is easy to lose some of the pellet at this stage)*.

Let the pellet dry then add 400-600 µl TE buffer (10 mM Tris, 1 mM EDTA; pH 8.0) and leave, ideally overnight, to dissolve.

We transfer the PEG to a 50 mL tube before centrifugation as there may be problems of viscosity with the PEG. If it is at the bottom of a 50 mL tube there is less distance to travel for the DNA to pellet on the bottom.

The amount of TE to dissolve the pellet in varies as you want to cover the entire "pellet" which is often a thin layer which covers a lot of the bottom of the Falcon tube.

1.4 To the final sample.

1.4.1 Phenol cleaning (in fume hood)

Transfer the DNA dissolved in TE to a 1.5 mL Eppendorf tube. Cool the tube on ice and then add one volume of cold (4 C) phenol/chloroform/isoamylalcohol (25:24:1). Shake gently until the mixture is milky (like CTAB cleaning above) and then spin on a chilled benchtop centrifuge at maximum speed until there is a clear boundary between the top (water) and bottom (phenol/chloroform/isoamylalcohol) phases. This usually takes 10-15 minutes at most.

Pipette off the top phase into a fresh 1.5 mL Eppendorf tube. Occasionally you will have a sample where the boundary between the two phases is messy and you will have to repeat this step. These samples are usually DNA rich, but they also have a lot of protein.

Add an equal volume of chloroform/isoamylalcohol, spin again and transfer to a fresh 1.5 mL Eppendorf tube again.

Phenol/chloroform/isoamlyalcohol needs to be buffered to pH 8. This is difficult to do, so we buy ours prebuffered from Sigma (P3803-100ML). We have used chloroform rather than chloroform/isoamylalcohol for the second step with no problems.

1.4.2 $CaCl_2$ precipitation

The next step is optional. If your supernatant is brown/dark at this stage you can precipitate out humic material with $CaCl_2$ precipitation. Otherwise move on to the next step.

For each 200 µl of your supernatant (you will have lost some during phenol/chloroform/isoamylalcohol cleaning) add 10 µl of 0.57 M $CaCl_2$.

Incubate at 65 C for 30 minutes. Spin down for 10 minutes at 14 000 g on a bench top centrifuge. Transfer the supernatant to a fresh 1.5 mL Eppendorf tube.

Checking the volume of the supernatant is important as if too much $CaCl_2$ is added you can lose a lot of DNA. Best to take off a known rounded volume from each sample so the ratio of supernatant to $CaCl_2$ is exact.

1.4.3 Sodium Acetate precipitation.

Add 1 tenth of a volume of 3 M sodium acetate to the 1.5 mL Eppendorf tube and mix gently. Add an equal volume of isopropanol at room temperature to the mixture and mix gently. After 10 minutes spin down at 14 000 g on a bench top centrifuge.

It is important that the above step is done at room temperature, otherwise a lot of salts can get into your sample.

Gently pipette off the supernatant, add 1 mL of cold 70 % ethanol, and spin down again. Pipette off the 70% ethanol, allow pellet to nearly dry. Add 20-50 µl of T.E. and let the pellet dissolve over a couple of hours.

2-Python script for RDP Qiime1 taxonomy file.

```
import textio as tio
import numpy as np
import re

re_1 = re.compile('([a-z]\_\_)')
re_k = re.compile('(k\_\_)')
re_p = re.compile('(p\_\_)')
re_c = re.compile('(c\_\_)')
re_o = re.compile('(o\_\_)')
re_f = re.compile('(f\_\_)')
re_g = re.compile('(g\_\_)')
re_s = re.compile('(s\_\_)')

k = 'k__'
p = 'p__'
c = 'c__'
o = 'o__'
f = 'f__'
g = 'g__'
s = 's__'

ll = [k,p,c,o,f,g,s]
ii = [2,4,6,8,10,12,14]
```

```
class ss_obj(tio.txt_obj):

    def read_it(z):
        z.read_tsv()

    def make_l(z):
        lg = len(z.s_obj)
        z.lg2 = int(len(z.s_obj[0][2:]))
        z.l = np.zeros(lg*(15+z.lg2),dtype='<U200').reshape(lg,(15+z.lg2))
        z.l[:,0] = z.s_obj[:,0]
        for x in range(1,z.lg2+1):
            z.l[:,-x] = z.s_obj[:,-x]
        z.l[:,1] = k
        z.l[:,3] = p
        z.l[:,5] = c
        z.l[:,7] = o
        z.l[:,9] = f
        z.l[:,11] = g
        z.l[:,13] = s

    def get_names(z):

        for x in range(len(z.s_obj)):
            g = re.split('(;)',z.s_obj[x][1])
            for y in range(len(ll)):
                for f in g:
                    if ll[y] in f:
                        if ' ' in f[0]:
                            f = f[1:]
                        z.l[x][ii[y]] = f[3:]+';'

    def fill_spaces(z):

        for x in range(len(z.l)):
            for y in range(len(z.l[x])):
                if z.l[x][y] == '':
                    z.l[x][y] = ';'

    def fill_names(z):

        for x in range(len(z.l)):
            for y in range(len(z.l[x])):
                if z.l[x][y] == ';':
                    if 'u_' in z.l[x][y-2][:2]:
                        z.l[x][y] = z.l[x][y-2]
                    elif 'u_' not in z.l[x][y-2][:2]:
                        z.l[x][y] = 'u_'+z.l[x][y-2]
```

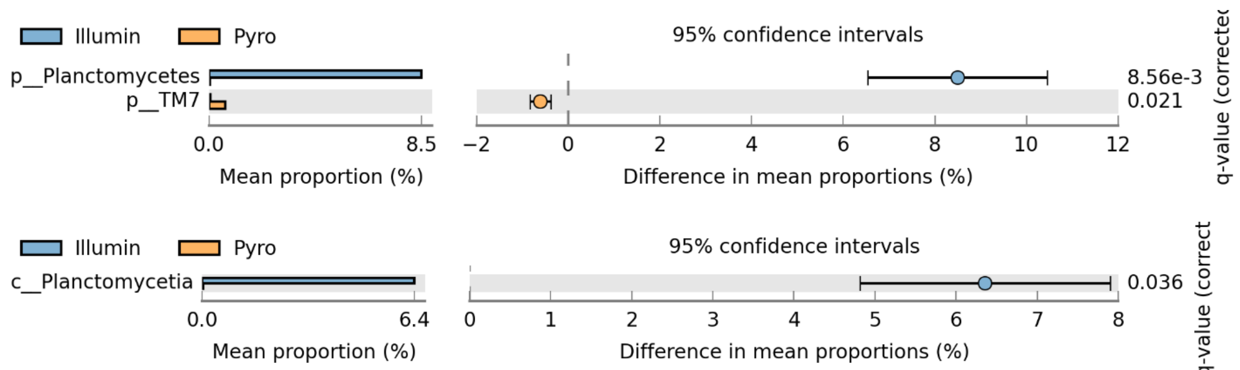
```
def b_t(z):

    z.n = np.zeros(len(z.l)*(8+z.lg2),dtype='<U200').reshape(len(z.l),(8+z.lg2))
    z.n[:,0] = z.l[:,0]
    for x in range(1,z.lg2+1):
        z.n[:,-x] = z.l[:,-x]

    for x in range(len(z.l)):

        z.n[x][1] = ''.join(z.l[x][1:3])
        z.n[x][2] = ''.join(z.l[x][3:5])
        z.n[x][3] = ''.join(z.l[x][5:7])
        z.n[x][4] = ''.join(z.l[x][7:9])
        z.n[x][5] = ''.join(z.l[x][9:11])
        z.n[x][6] = ''.join(z.l[x][11:13])
        z.n[x][7] = ''.join(z.l[x][13:15])

    z.n2 = np.zeros(len(z.n)*(2+z.lg2),dtype='<U200').reshape(len(z.n),(2+z.lg2))
    z.n2[:,0] = z.n[:,0]
    for x in range(1,z.lg2+1):
        z.n2[:,-x] = z.n[:,-x]

    for x in range(len(z.n)):

        z.n2[x][1] = ''.join(z.n[x][1:8])

def do_it(z):
    z.read_it()
    z.make_l()
    z.get_names()
    z.fill_spaces()
    z.fill_names()
    z.b_t()

def write_it(z):
    z.select(z.n2)
    z.write_tsv('renamed_'+z.txtobj)
```

A comparison of hard and soft direct methods for DNA extraction from soil.

Patrick Hill, Mathieu F Dextraze, David Kroetsch, Christopher N Boddy

Supplemental Data2-Results Figures

a-Mobio Extracted DNA



b-Soft extracted DNA

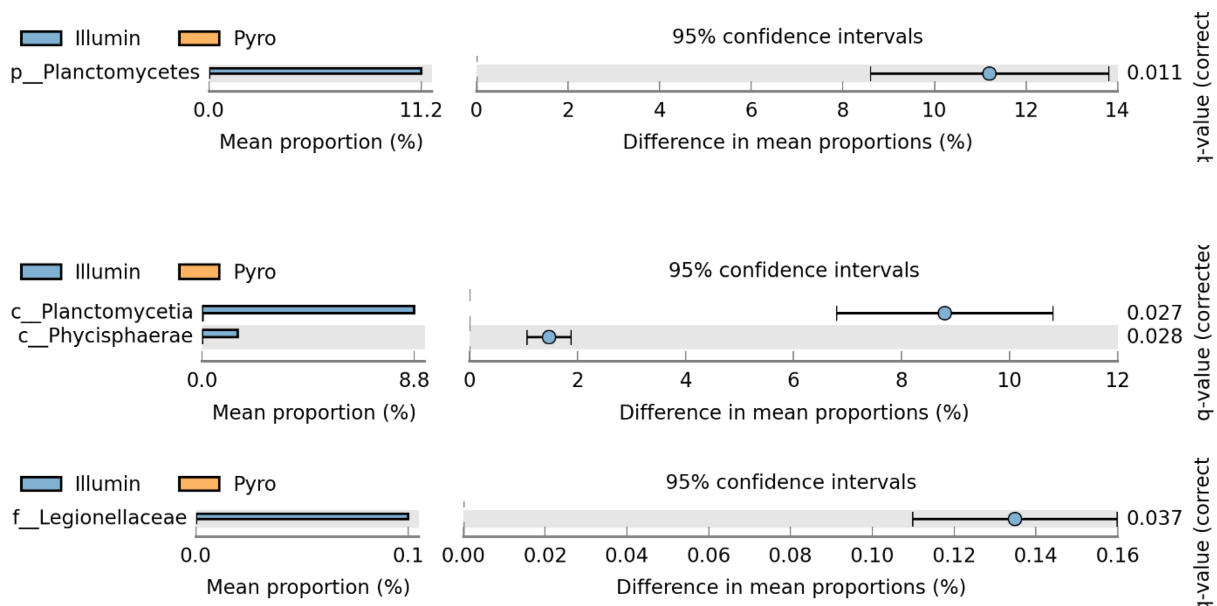
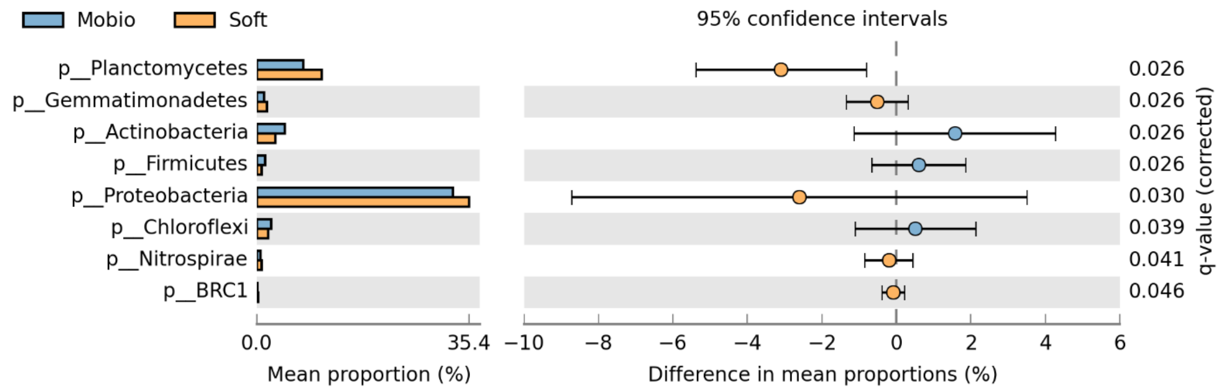
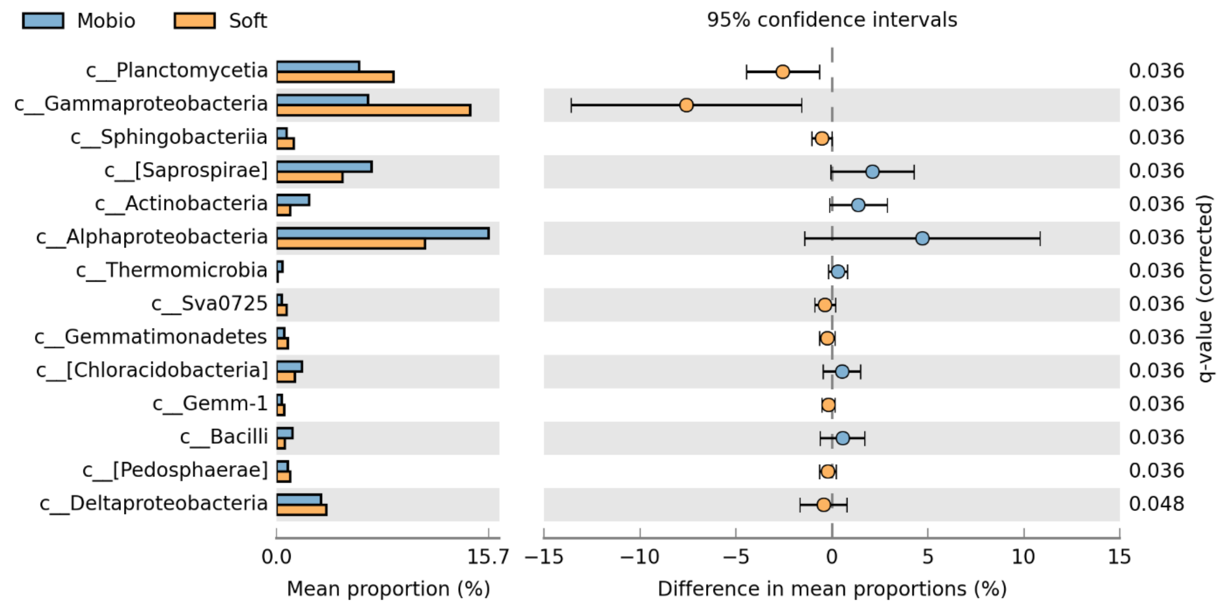


Figure S1 STAMP Extended error bar plots at the a-phylum and class level and b- phylum, class and family level for group analysis of Illumina V4V5 EL4 versus Mobio V4V6 EL1 sequencing for the soils Vancouver, Exp1B, Exp2A, Exp2B and Billings Bridge. For each phylum, class or family bar was at least more than 1% of sequences in at least one sample and differences were at least four fold. Multiple test correction with Storey FDR.


Mobio
Soft
95% confidence intervals
p__Planctomycetes
p__Gemmatimonadetes
p__Actinobacteria
p__Firmicutes
p__Proteobacteria
p__Chloroflexi
p__Nitrospirae
p__BRC1
0.026
0.026
0.026
0.026
0.030
0.039
0.041
0.046
q-value (corrected)
0.0
35.4
−10
−8
−6
−4
−2
0
2
4
6
Mean proportion (%)
Difference in mean proportions (%)


Mobio
Soft
95% confidence intervals
c__Planctomycetia
c__Gammaproteobacteria
c__Sphingobacteriia
c__[Saprospirae]
c__Actinobacteria
c__Alphaproteobacteria
c__Thermomicrobia
c__Sva0725
c__Gemmatimonadetes
c__[Chloracidobacteria]
c__Gemm-1
c__Bacilli
c__[Pedosphaerae]
c__Deltaproteobacteria
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.036
0.048
q-value (corrected)
0.0
15.7
−15
−10
−5
0
5
10
15
Mean proportion (%)
Difference in mean proportions (%)

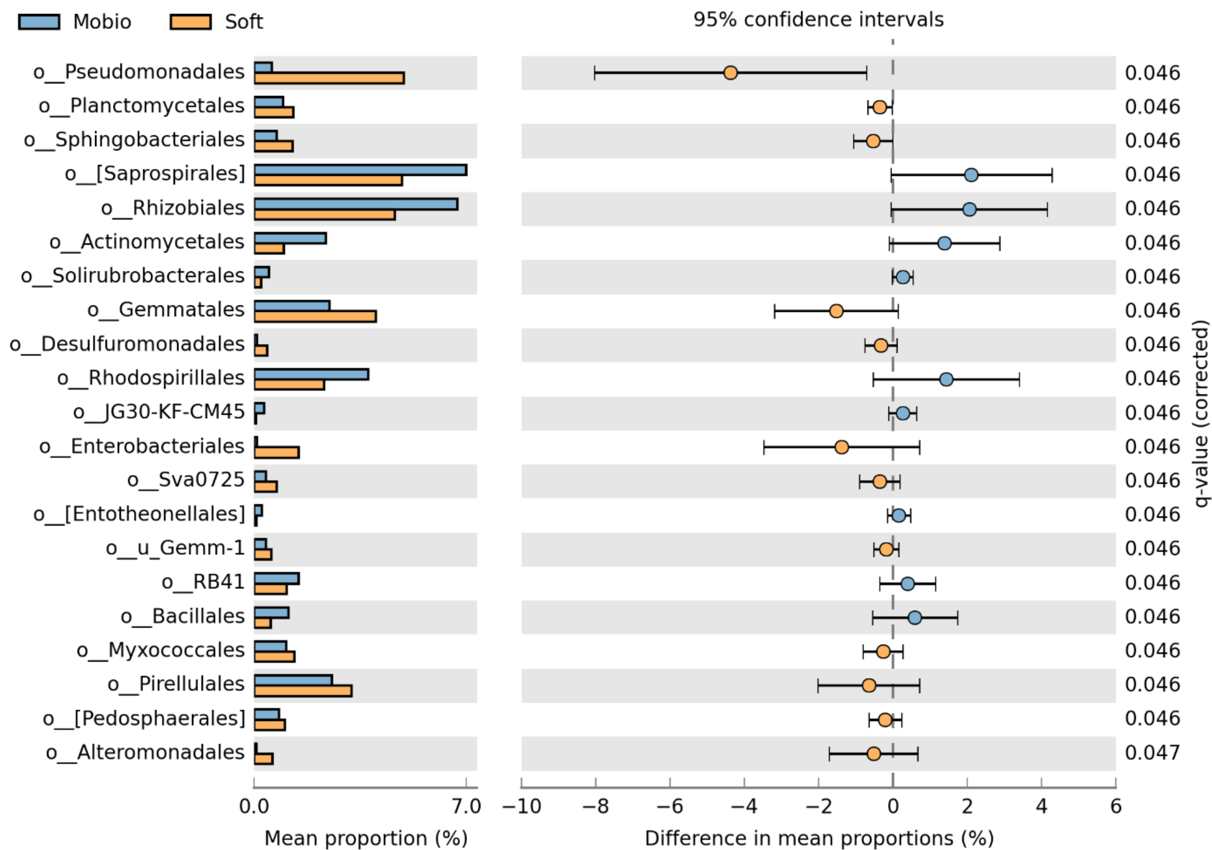
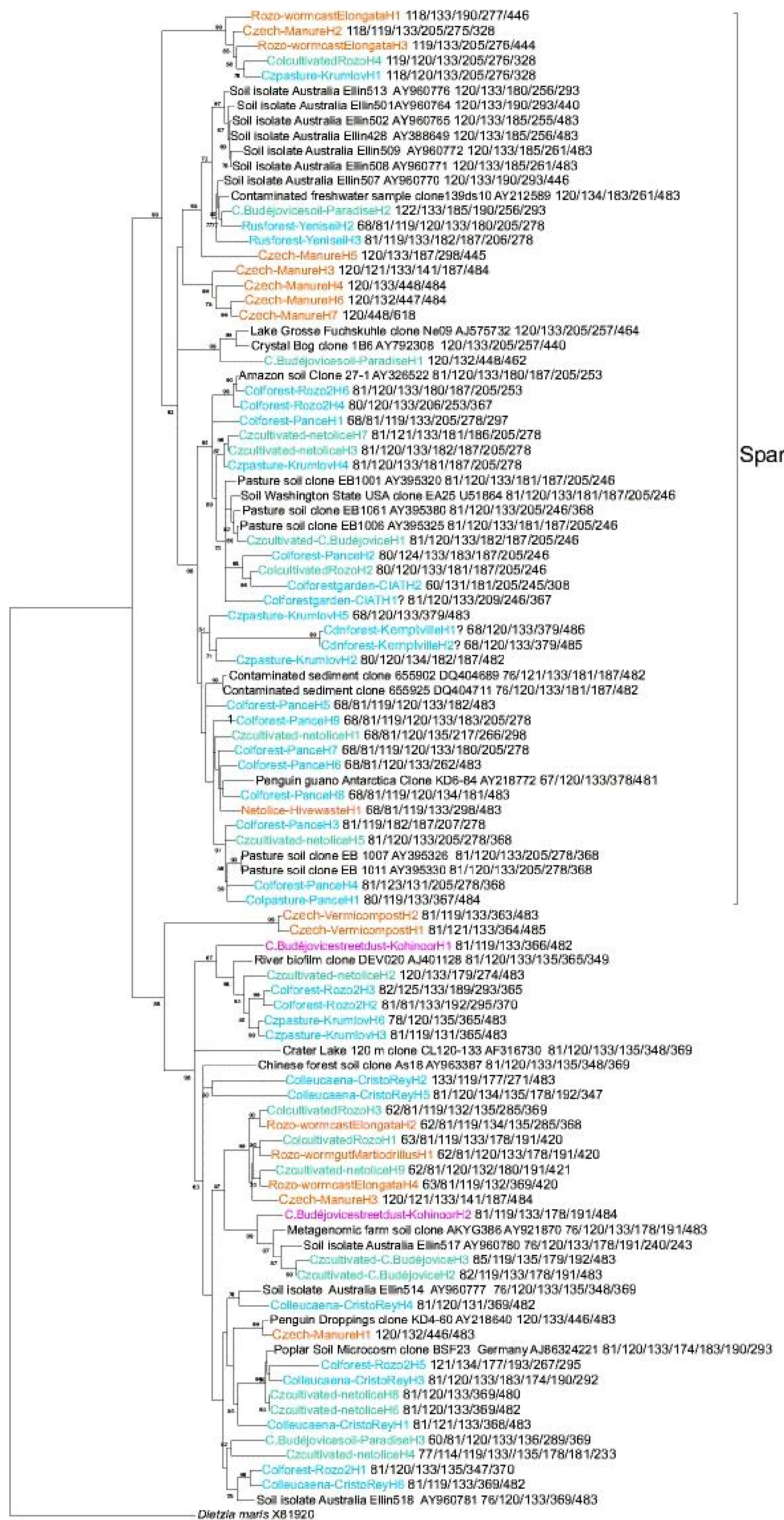


Figure S2 STAMP Extended error bar plots at the phylum, class and order level for group analysis (Soft versus Mobio extraction) from V4-V5 primers (EL4) of 11 soils. Each phylum, class or order bar represented was at least more than 1% of sequences in at least one sample. Multiple test correction with Storey FDR.

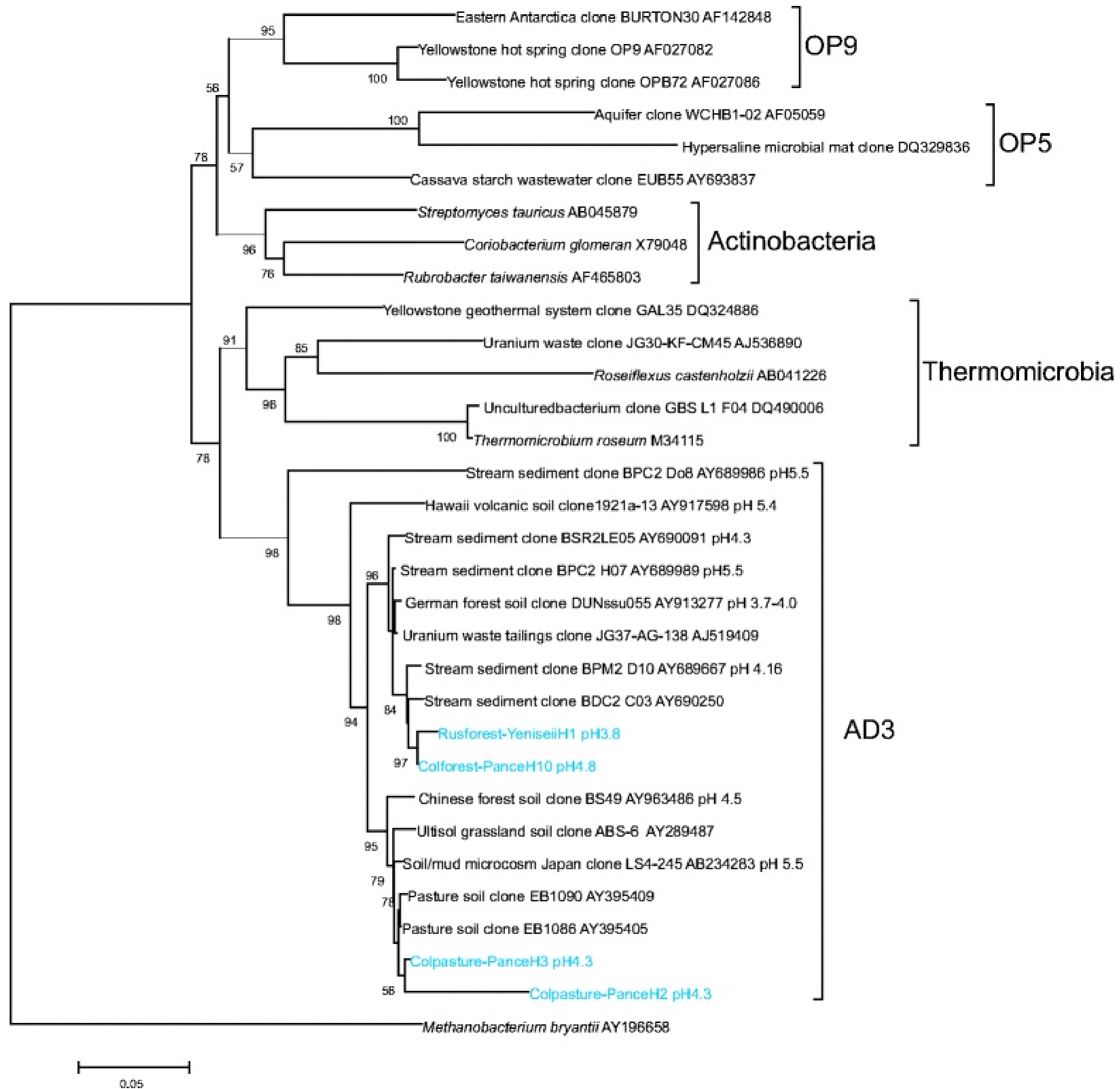
Appendix 2 Supplemental Data for Chapter 3 Investigation of how the Actinobacterial community structure changes across land use.

Land Use Intensity Controls Actinobacterial Community Structure

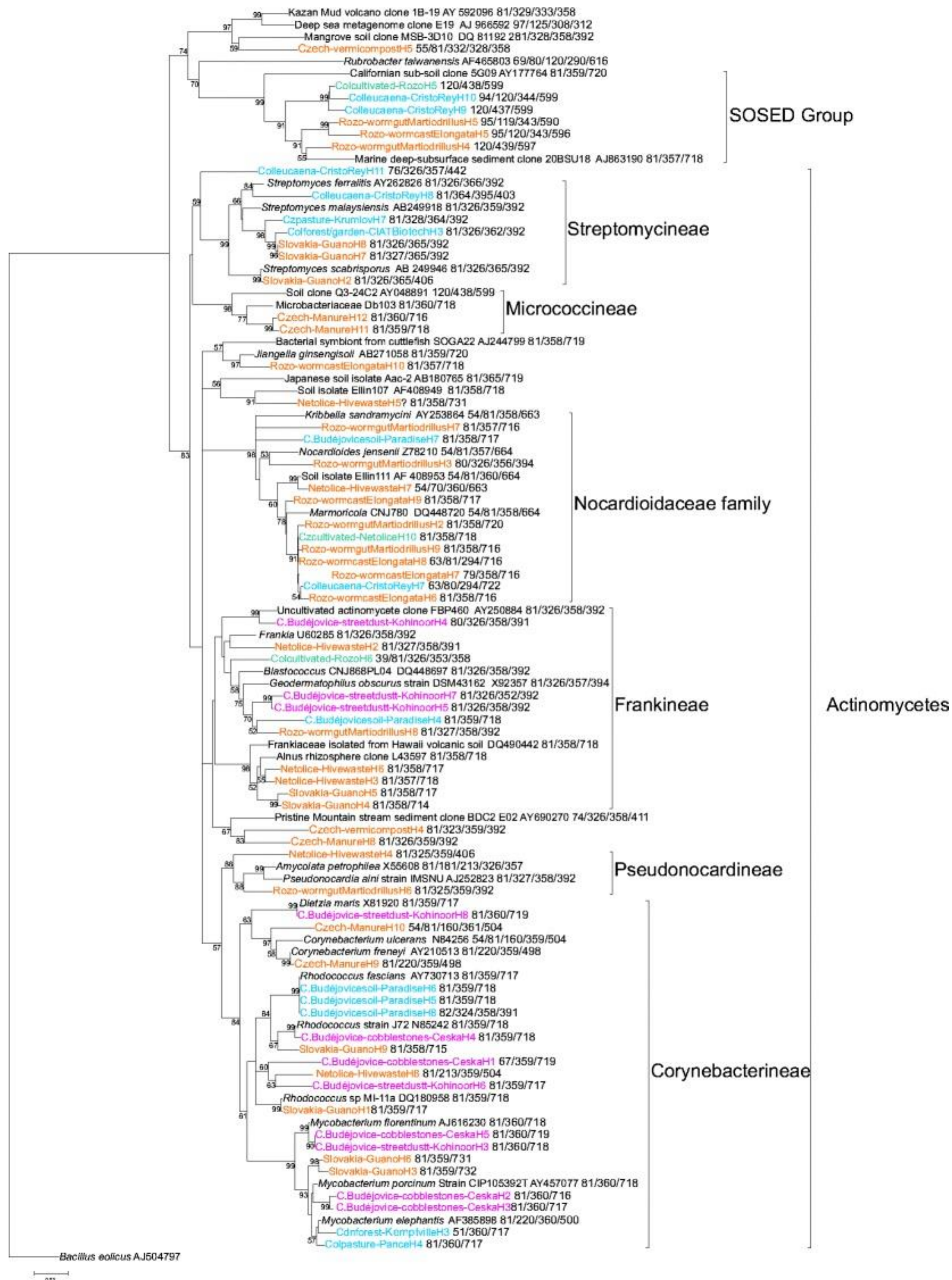


Rozo-wormcastElongataH1 118/133/190/277/446
Czech-ManureH2 118/119/133/205/275/328
Rozo-wormcastElongataH3 119/133/205/276/444
ColcultivatedRozoH4 119/120/133/205/276/328
Czpasture-KrumlovH1 118/120/133/205/276/328
Soil isolate Australia Ellin513 AY960776 120/133/180/256/293
Soil isolate Australia Ellin501AY960764 120/133/190/293/440
Soil isolate Australia Ellin502 AY960765 120/133/185/255/483
Soil isolate Australia Ellin428 AY388649 120/133/185/256/483
Soil isolate Australia Ellin509 AY960772 120/133/185/261/483
Soil isolate Australia Ellin508 AY960771 120/133/185/261/483
Soil isolate Australia Ellin507 AY960770 120/133/190/293/446
Contaminated freshwater sample clone139ds10 AY212589 120/134/183/261/483
C.Budějovicesoil-ParadiseH2 122/133/185/190/256/293
Rusforest-YeniseiH2 68/81/119/120/133/180/205/278
Rusforest-YeniseiH3 81/119/133/182/187/206/278
Czech-ManureH5 120/133/187/298/445
Czech-ManureH3 120/121/133/141/187/484
Czech-ManureH4 120/133/448/484
Czech-ManureH6 120/132/447/484
Czech-ManureH7 120/448/618
Lake Grosse Fuchskuhle clone Ne09 AJ575732 120/133/205/257/464
Crystal Bog clone 1B6 AY792308 120/133/205/257/440
C.Budějovicesoil-ParadiseH1 120/132/448/462
Amazon soil Clone 27-1 AY326522 81/120/133/180/187/205/253
Colforest-Rozo2H6 81/120/133/180/187/205/253
Colforest-Rozo2H4 80/120/133/206/253/367
Colforest-PanceH1 68/81/119/133/205/278/297
Czcultivated-netoliceH7 81/121/133/181/186/205/278
Czcultivated-netoliceH3 81/120/133/182/187/205/278
Czpasture-KrumlovH4 81/120/133/181/187/205/278
Pasture soil clone EB1001 AY395320 81/120/133/181/187/205/246
Soil Washington State USA clone EA25 U51864 81/120/133/181/187/205/246
Pasture soil clone EB1061 AY395380 81/120/133/205/246/368
Pasture soil clone EB1006 AY395325 81/120/133/181/187/205/246
Czcultivated-C.BudejoviceH1 81/120/133/182/187/205/246
Colforest-PanceH2 80/124/133/183/187/205/246
ColcultivatedRozoH2 80/120/133/181/187/205/246
Colforestgarden-CIATH2 60/131/181/205/245/308
Colforestgarden-CIATH1? 81/120/133/209/246/367
Czpasture-KrumlovH5 68/120/133/379/483
Cdnforest-KemptvilleH1? 68/120/133/379/486
Cdnforest-KemptvilleH2? 68/120/133/379/485
Czpasture-KrumlovH2 80/120/134/182/187/482
Contaminated sediment clone 655902 DQ404689 76/121/133/181/187/482
Contaminated sediment clone 655925 DQ404711 76/120/133/181/187/482
Colforest-PanceH5 68/81/119/120/133/182/483
Colforest-PanceH9 68/81/119/120/133/183/205/278
Czcultivated-netoliceH1 68/81/120/135/217/266/298
Colforest-PanceH7 68/81/119/120/133/180/205/278
Colforest-PanceH6 68/81/120/133/262/483
Penguin guano Antarctica Clone KD6-84 AY218772 67/120/133/378/481
Colforest-PanceH8 68/81/119/120/134/181/483
Netolice-HivewasteH1 68/81/119/133/298/483
Colforest-PanceH3 81/119/182/187/207/278
Czcultivated-netoliceH5 81/120/133/205/278/368
Pasture soil clone EB 1007 AY395326 81/120/133/205/278/368
Pasture soil clone EB 1011 AY395330 81/120/133/205/278/368
Colforest-PanceH4 81/123/131/205/278/368
Colpasture-PanceH1 80/119/133/367/484
Spartobacteria
Czech-VermicompostH2 81/119/133/363/483
Czech-VermicompostH1 81/121/133/364/485
C.Budějovicestreetdust-KohinoorH1 81/119/133/366/482
River biofilm clone DEV020 AJ401128 81/120/133/135/365/349
Czcultivated-netoliceH2 120/133/179/274/483
Colforest-Rozo2H3 82/125/133/189/293/365
Colforest-Rozo2H2 81/81/133/192/295/370
Czpasture-KrumlovH6 78/120/135/365/483
Czpasture-KrumlovH3 81/119/131/365/483
Crater Lake 120 m clone CL120-133 AF316730 81/120/133/135/348/369
Chinese forest soil clone As18 AY963387 81/120/133/135/348/369
Colleucaena-CristoReyH2 133/119/177/271/483
Colleucaena-CristoReyH5 81/120/134/135/178/192/347
ColcultivatedRozoH3 62/81/119/132/135/285/369
Rozo-wormcastElongataH2 62/81/119/134/135/285/368
ColcultivatedRozoH1 63/81/119/133/178/191/420
Rozo-wormgutMartiodrillusH1 62/81/120/133/178/191/420
Czcultivated-netoliceH9 62/81/120/132/180/191/421
Rozo-wormcastElongataH4 63/81/119/132/369/420
Czech-ManureH3 120/121/133/141/187/484
C.Budějovicestreetdust-KohinoorH2 81/119/133/178/191/484
Metagenomic farm soil clone AKYG386 AY921870 76/120/133/178/191/483
Soil isolate Australia Ellin517 AY960780 76/120/133/178/191/240/243
Czcultivated-C.BudějoviceH3 85/119/135/179/192/483
Czcultivated-C.BudějoviceH2 82/119/133/178/191/483
Soil isolate Australia Ellin514 AY960777 76/120/133/135/348/369
Colleucaena-CristoReyH4 81/120/131/369/482
Penguin Droppings clone KD4-60 AY218640 120/133/446/483
Czech-ManureH1 120/132/446/483
Poplar Soil Microcosm clone BSF23 Germany AJ86324221 81/120/133/174/183/190/293
Colforest-Rozo2H5 121/134/177/193/267/295
Colleucaena-CristoReyH3 81/120/133/183/174/190/292
Czcultivated-netoliceH8 81/120/133/369/480
Czcultivated-netoliceH6 81/120/133/369/482
Colleucaena-CristoReyH1 81/121/133/368/483
C.Budějovicesoil-ParadiseH3 60/81/120/133/136/289/369
Czcultivated-netoliceH4 77/114/119/133//135/178/181/233
Colforest-Rozo2H1 81/120/133/135/347/370
Colleucaena-CristoReyH6 81/119/133/369/482
Soil isolate Australia Ellin518 AY960781 76/120/133/369/483
Subdivision 3
Dietzia maris X81920

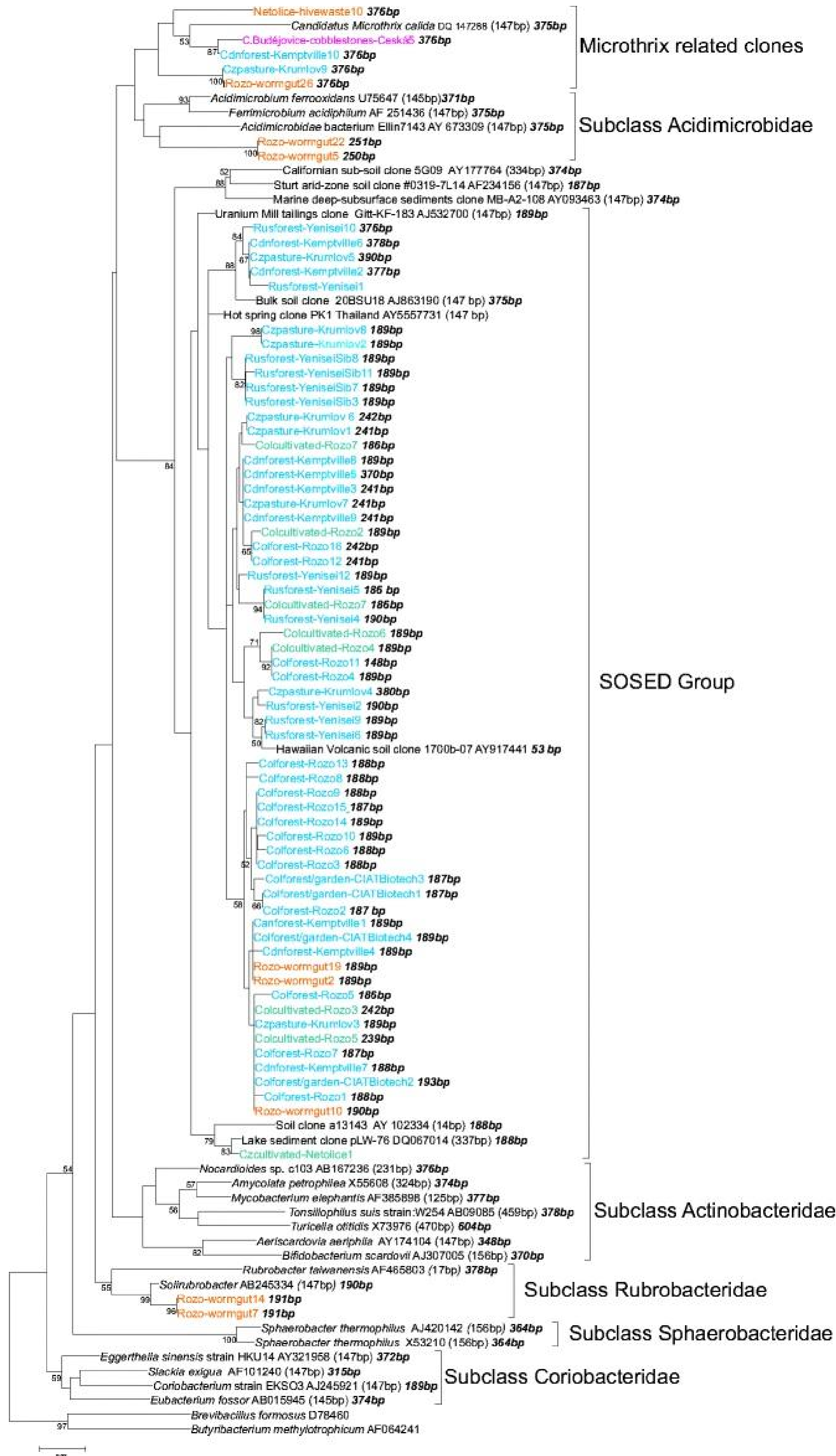
Supplemental Data 1: Neighbour-joining tree of Verrucomicrobial 16S sequences generated with the F-Act/R-Bact primers of Heuer et al. [15]. Bootstrap values below 50 are not shown. Samples are colour coded as: forest/pasture soils (*blue*), cultivated soils (*sea green*), insect- and earthworm-associated sediments (*orange*), street sediments (*red*). Simulated TaqI ARDRA band positions are shown for each sequence (e.g. Czcultivated-netoliceH5 81/120/133/205/278/368). For sample descriptions, see Tables 1 and 2.



Supplemental Data 2: Neighbour-joining tree of candidate division AD3 16S clones generated with the F-Act/R-Bact primers of Heuer et al. [15]; pH values of closest BLAST matches are given when available. For sample descriptions, see Table 3.1.

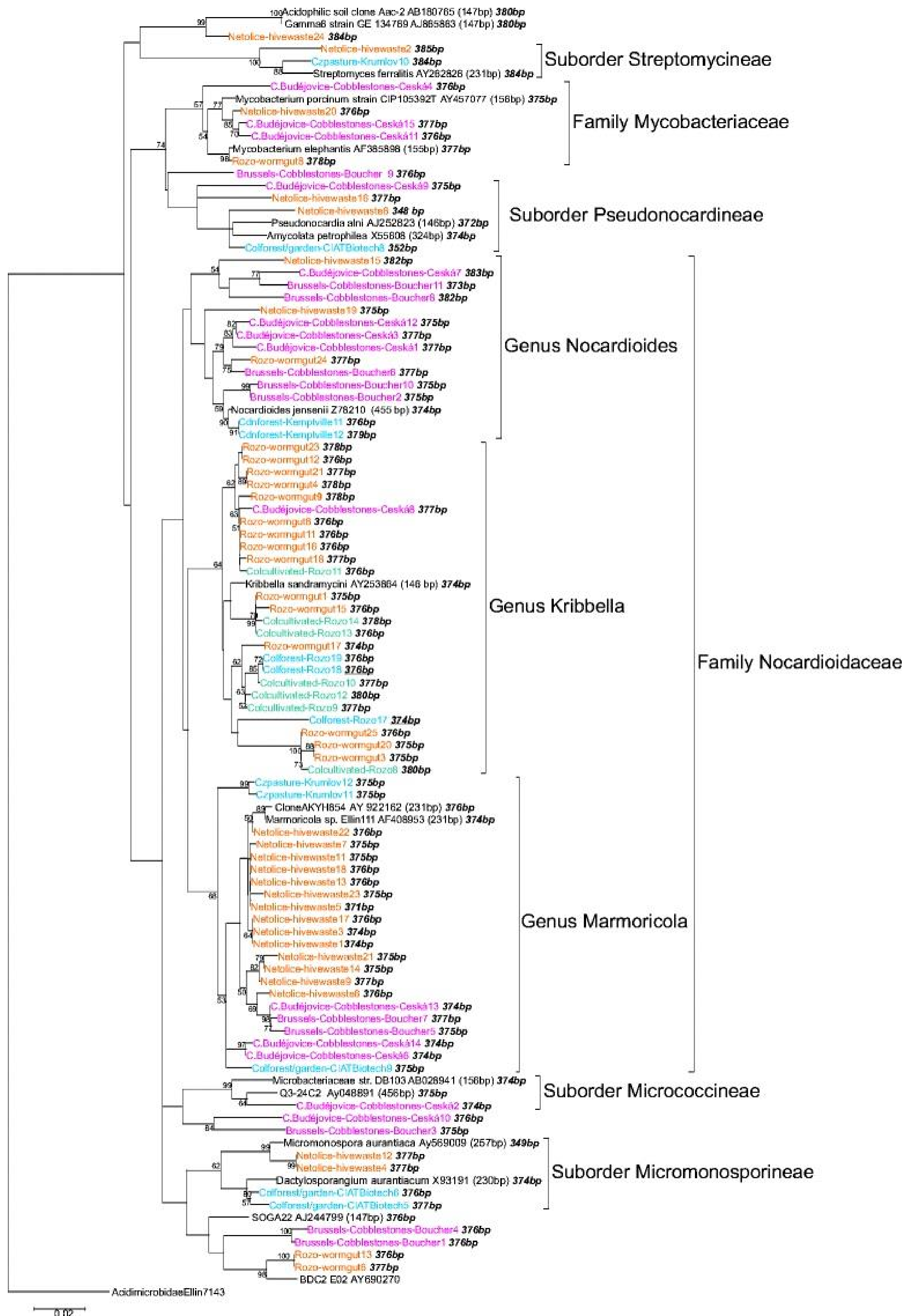


Supplemental Data 3: Neighbour-joining tree of actinobacterial 16S sequences generated with the F-Act/R-Bact primers of Heuer et al. [15]. Bootstrap values below 50 are not shown. Samples are colour coded as: forest/pasture soils (*blue*), cultivated soils (*sea green*), insect- and earthworm0associated sediments (*orange*), street sediments (*red*). Simulated TaqI amplified ribosomal DNA restriction analysis (*ARDRA*) band positions are shown for each sequence (e.g. Czpasture-KrumlovH7 81/328/364/392). For sample abbreviations, see Tables 1 and 2.

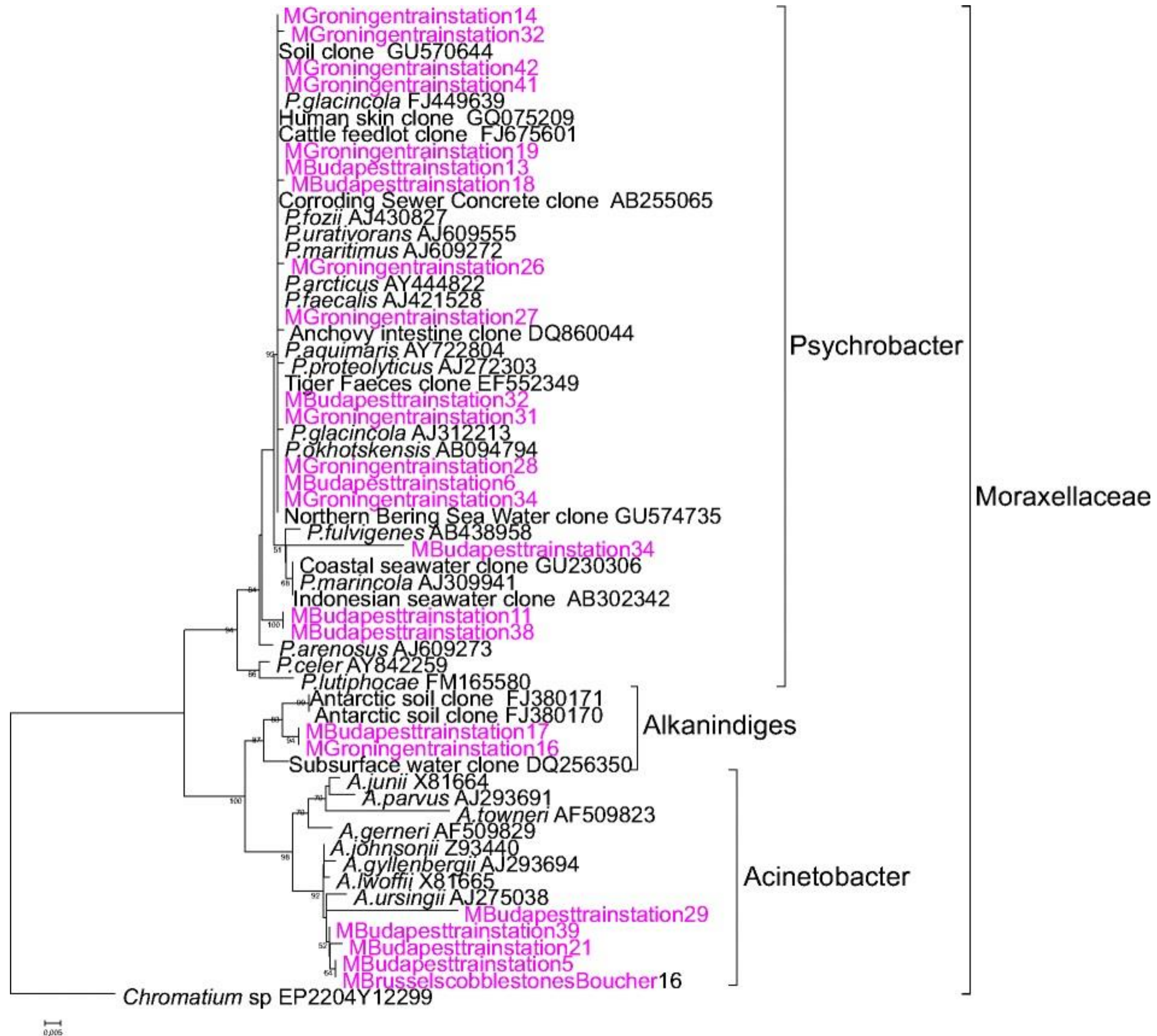


Netolice-hivewaste10 376bp
Candidatus Microthrix calida DQ 147288 (147bp) 375bp
C.Budějovice-cobblestones-Ceská5 376bp
Cdnforest-Kemptville10 376bp
Czpasture-Krumlov9 376bp
Rozo-wormgut26 376bp
Microthrix related clones
Acidimicrobium ferrooxidans U75647 (145bp)371bp
Ferrimicrobium acidiphilum AF 251436 (147bp) 375bp
Acidimicrobidae bacterium Ellin7143 AY 673309 (147bp) 375bp
Rozo-wormgut22 251bp
Rozo-wormgut5 250bp
Subclass Acidimicrobidae
Californian sub-soil clone 5G09 AY177764 (334bp) 374bp
Sturt arid-zone soil clone #0319-7L14 AF234156 (147bp) 187bp
Marine deep-subsurface sediments clone MB-A2-108 AY093463 (147bp) 374bp
Uranium Mill tailings clone Gitt-KF-183 AJ532700 (147bp) 189bp
Rusforest-Yenisei10 376bp
Cdnforest-Kemptville6 378bp
Czpasture-Krumlov5 390bp
Cdnforest-Kemptville2 377bp
Rusforest-Yenisei1
Bulk soil clone 20BSU18 AJ863190 (147 bp) 375bp
Hot spring clone PK1 Thailand AY5557731 (147 bp)
Czpasture-Krumlov8 189bp
Czpasture-Krumlov2 189bp
Rusforest-YeniseiSib8 189bp
Rusforest-YeniseiSib11 189bp
Rusforest-YeniseiSib7 189bp
Rusforest-YeniseiSib3 189bp
Czpasture-Krumlov 6 242bp
Czpasture-Krumlov1 241bp
Colcultivated-Rozo7 186bp
Cdnforest-Kemptville8 189bp
Cdnforest-Kemptville5 370bp
Cdnforest-Kemptville3 241bp
Czpasture-Krumlov7 241bp
Cdnforest-Kemptville9 241bp
Colcultivated-Rozo2 189bp
Colforest-Rozo16 242bp
Colforest-Rozo12 241bp
Rusforest-Yenisei12 189bp
Rusforest-Yenisei5 186 bp
Colcultivated-Rozo7 186bp
Rusforest-Yenisei4 190bp
Colcultivated-Rozo6 189bp
Colcultivated-Rozo4 189bp
Colforest-Rozo11 148bp
Colforest-Rozo4 189bp
Czpasture-Krumlov4 380bp
Rusforest-Yenisei2 190bp
Rusforest-Yenisei9 189bp
Rusforest-Yenisei6 189bp
Hawaiian Volcanic soil clone 1700b-07 AY917441 53 bp
Colforest-Rozo13 188bp
Colforest-Rozo8 188bp
Colforest-Rozo9 188bp
Colforest-Rozo15 187bp
Colforest-Rozo14 189bp
Colforest-Rozo10 189bp
Colforest-Rozo6 188bp
Colforest-Rozo3 188bp
Colforest/garden-CIATBiotech3 187bp
Colforest/garden-CIATBiotech1 187bp
Colforest-Rozo2 187 bp
Canforest-Kemptville1 189bp
Colforest/garden-CIATBiotech4 189bp
Cdnforest-Kemptville4 189bp
Rozo-wormgut19 189bp
Rozo-wormgut2 189bp
Colforest-Rozo5 186bp
Colcultivated-Rozo3 242bp
Czpasture-Krumlov3 189bp
Colcultivated-Rozo5 239bp
Colforest-Rozo7 187bp
Cdnforest-Kemptville7 188bp
Colforest/garden-CIATBiotech2 193bp
Colforest-Rozo1 188bp
Rozo-wormgut10 190bp
Soil clone a13143 AY 102334 (14bp) 188bp
Lake sediment clone pLW-76 DQ067014 (337bp) 188bp
Czcultivated-Netolice1
SOSED Group
Nocardioides sp. c103 AB167236 (231bp) 376bp
Amycolata petrophilea X55608 (324bp) 374bp
Mycobacterium elephantis AF385898 (125bp) 377bp
Tonsillophilus suis strain:W254 AB09085 (459bp) 378bp
Turicella otitidis X73976 (470bp) 604bp
Aeriscardovia aeriphila AY174104 (147bp) 348bp
Bifidobacterium scardovii AJ307005 (156bp) 370bp
Subclass Actinobacteridae
Rubrobacter taiwanensis AF465803 (17bp) 378bp
Solirubrobacter AB245334 (147bp) 190bp
Rozo-wormgut14 191bp
Rozo-wormgut7 191bp
Subclass Rubrobacteridae
Sphaerobacter thermophilus AJ420142 (156bp) 364bp
Sphaerobacter thermophilus X53210 (156bp) 364bp
Subclass Sphaerobacteridae
Eggerthella sinensis strain HKU14 AY321958 (147bp) 372bp
Slackia exigua AF101240 (147bp) 315bp
Coriobacterium strain EKSO3 AJ245921 (147bp) 189bp
Eubacterium fossor AB015945 (145bp) 374bp
Subclass Coriobacteridae
Brevibacillus formosus D78460
Butyribacterium methylotrophicum AF064241

Supplemental Data 4: Neighbour-joining tree of non-actinomycetal actinobacterial 16S clones generated with the F-Act/R-Act primers of Monciardini et al. (2003). Bootstrap values below 50 are not shown. Simulated T-RFLP values are shown for all matches where sequence length allows; forward values are in *parentheses* e.g. (231 bp), reverse values in *bold italics*, e.g. *375 bp*. Samples are colour coded as: forest/pasture soils (*blue*), cultivated soils (*sea green*), insect- and earthworm-associated sediments (*orange*), street sediments (*red*). For sample abbreviations, see Tables 1 and 2.


Acidophilic soil clone Aac-2 AB180765 (147bp) 380bp
Gamma6 strain GE 134769 AJ865863 (147bp) 380bp
Netolice-hivewaste24 384bp
Netolice-hivewaste2 385bp
Czpasture-Krumlov10 384bp
Streptomyces ferralitis AY262826 (231bp) 384bp
Suborder Streptomycineae
C.Budějovice-Cobblestones-Ceská4 376bp
Mycobacterium porcinum strain CIP105392T AY457077 (156bp) 375bp
Netolice-hivewaste20 376bp
C.Budějovice-Cobblestones-Ceská15 377bp
C.Budějovice-Cobblestones-Ceská11 376bp
Mycobacterium elephantis AF385898 (155bp) 377bp
Rozo-wormgut8 378bp
Family Mycobacteriaceae
Brussels-Cobblestones-Boucher 9 376bp
C.Budějovice-Cobblestones-Ceská9 375bp
Netolice-hivewaste16 377bp
Netolice-hivewaste6 348 bp
Pseudonocardia alni AJ252823 (146bp) 372bp
Amycolata petrophilea X55608 (324bp) 374bp
Colforest/garden-CIATBiotech8 352bp
Suborder Pseudonocardineae
Netolice-hivewaste15 382bp
C.Budějovice-Cobblestones-Ceská7 383bp
Brussels-Cobblestones-Boucher11 373bp
Brussels-Cobblestones-Boucher8 382bp
Netolice-hivewaste19 375bp
C.Budějovice-Cobblestones-Ceská12 375bp
C.Budějovice-Cobblestones-Ceská3 377bp
C.Budějovice-Cobblestones-Ceská1 377bp
Rozo-wormgut24 377bp
Brussels-Cobblestones-Boucher6 377bp
Brussels-Cobblestones-Boucher10 375bp
Brussels-Cobblestones-Boucher2 375bp
Nocardioides jensenii Z78210 (455 bp) 374bp
Cdnforest-Kemptville11 376bp
Cdnforest-Kemptville12 379bp
Genus Nocardioides
Rozo-wormgut23 378bp
Rozo-wormgut12 376bp
Rozo-wormgut21 377bp
Rozo-wormgut4 378bp
Rozo-wormgut9 378bp
C.Budějovice-Cobblestones-Ceská8 377bp
Rozo-wormgut6 376bp
Rozo-wormgut11 376bp
Rozo-wormgut16 376bp
Rozo-wormgut18 377bp
Colcultivated-Rozo11 376bp
Kribbella sandramycini AY253864 (146 bp) 374bp
Rozo-wormgut1 375bp
Rozo-wormgut15 376bp
Colcultivated-Rozo14 378bp
Colcultivated-Rozo13 376bp
Rozo-wormgut17 374bp
Colforest-Rozo19 376bp
Colforest-Rozo18 376bp
Colcultivated-Rozo10 377bp
Colcultivated-Rozo12 380bp
Colcultivated-Rozo9 377bp
Colforest-Rozo17 374bp
Rozo-wormgut25 376bp
Rozo-wormgut20 375bp
Rozo-wormgut3 375bp
Colcultivated-Rozo8 380bp
Genus Kribbella
Family Nocardioidaceae
Czpasture-Krumlov12 375bp
Czpasture-Krumlov11 375bp
CloneAKYH854 AY 922162 (231bp) 376bp
Marmoricola sp. Ellin111 AF408953 (231bp) 374bp
Netolice-hivewaste22 376bp
Netolice-hivewaste7 375bp
Netolice-hivewaste11 375bp
Netolice-hivewaste18 376bp
Netolice-hivewaste13 376bp
Netolice-hivewaste23 375bp
Netolice-hivewaste5 371bp
Netolice-hivewaste17 376bp
Netolice-hivewaste3 374bp
Netolice-hivewaste1 374bp
Netolice-hivewaste21 375bp
Netolice-hivewaste14 375bp
Netolice-hivewaste9 377bp
Netolice-hivewaste8 376bp
C.Budějovice-Cobblestones-Ceská13 374bp
Brussels-Cobblestones-Boucher7 377bp
Brussels-Cobblestones-Boucher5 375bp
C.Budějovice-Cobblestones-Ceská14 374bp
C.Budějovice-Cobblestones-Ceská6 374bp
Colforest/garden-CIATBiotech9 375bp
Genus Marmoricola
Microbacteriaceae str. DB103 AB028941 (156bp) 374bp
Q3-24C2 Ay048891 (456bp) 375bp
C.Budějovice-Cobblestones-Ceská2 374bp
C.Budějovice-Cobblestones-Ceská10 376bp
Brussels-Cobblestones-Boucher3 375bp
Suborder Micrococcineae
Micromonospora aurantiaca Ay569009 (257bp) 349bp
Netolice-hivewaste12 377bp
Netolice-hivewaste4 377bp
Dactylosporangium aurantiacum X93191 (230bp) 374bp
Colforest/garden-CIATBiotech6 376bp
Colforest/garden-CIATBiotech5 377bp
Suborder Micromonosporineae
SOGA22 AJ244799 (147bp) 376bp
Brussels-Cobblestones-Boucher4 376bp
Brussels-Cobblestones-Boucher1 376bp
Rozo-wormgut13 376bp
Rozo-wormgut6 377bp
BDC2 E02 AY690270
AcidimicrobidaeEllin7143
0.02

Supplemental Data 5: Neighbour-joining tree of actinomycetal actinobacterial 16S clones generated with the F-Act/R-Act primers of Monciardini et al. (2003). Bootstrap values below 50 are not shown. Simulated T-RFLP values are shown for all matches were sequence length allows, forward values are in *parentheses*, e.g. (231 bp), reverse values in *bold italics*, e.g. *375 bp*. Samples are colour coded as: forest/pasture soils (*blue*), cultivated soils (*sea green*), insect- and earthworm-associated sediments (*orange*), street sediments (*red*). For sample abbreviations, see Tables 1 and 2



Supplemental Data 6: Neighbour-joining tree of Psychrobacter 16S clones generated with the eubacterial primers of Marchesi et al. (1998). Bootstrap values below 50 are not shown. For sample abbreviations, see Table 3.2.

Appendix three. Supplementary data for Chapter Four. Comparing the Type I polyketide synthase pathways of soil and street sediments.

Habitat-specific type I polyketide synthases in soils and street sediments,

Figure S1 AT A8-1/KS K7-1 cophylogeny. **Sequences in blue are from soil, in red from street sediments and in orange from earthworm-affected environments**

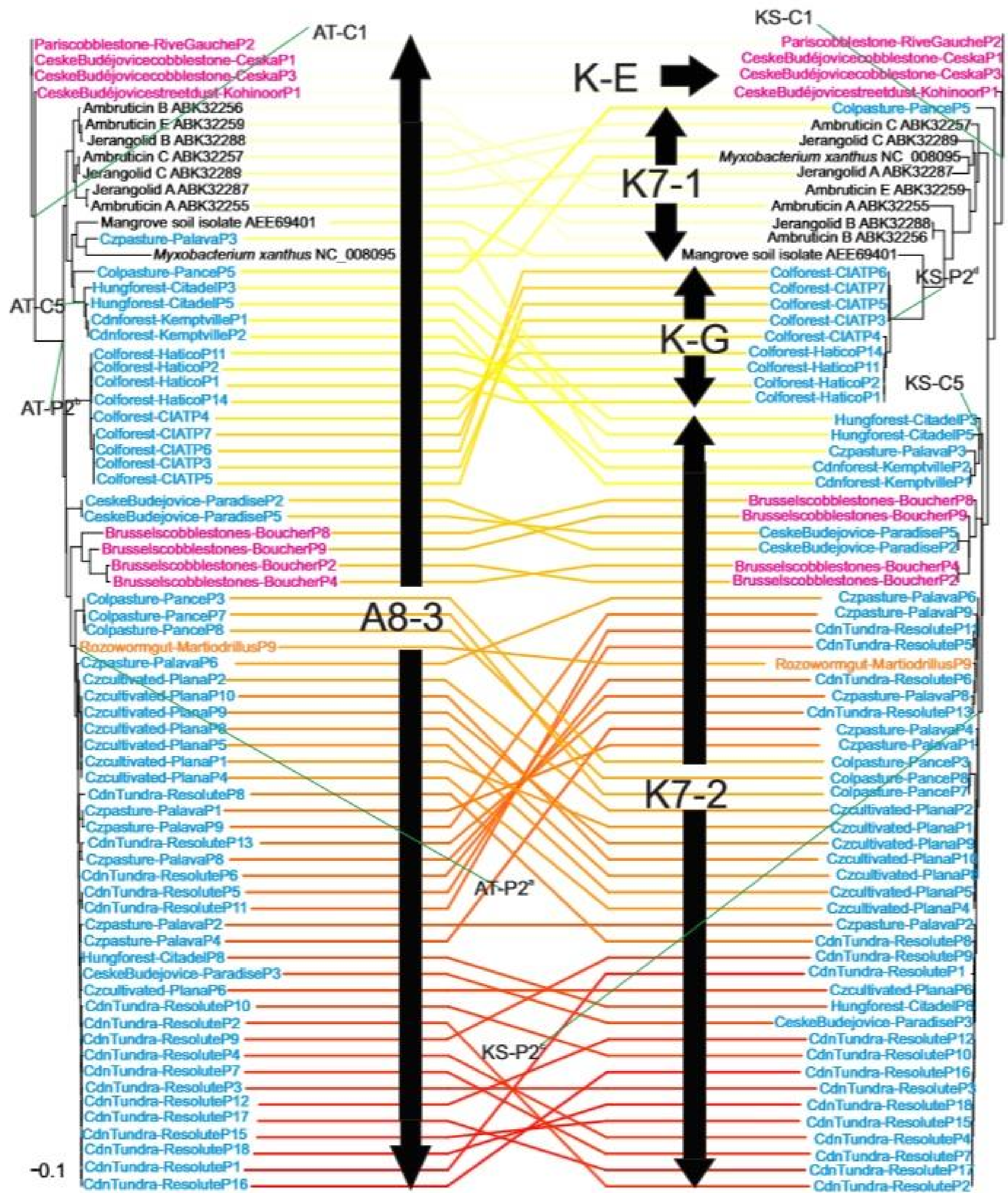


Figure S2 AT A8-3/KS K7-2 cophylogeny. Sequences in blue are from soil, in red from street sediments and in orange from earthworm-affected environments.

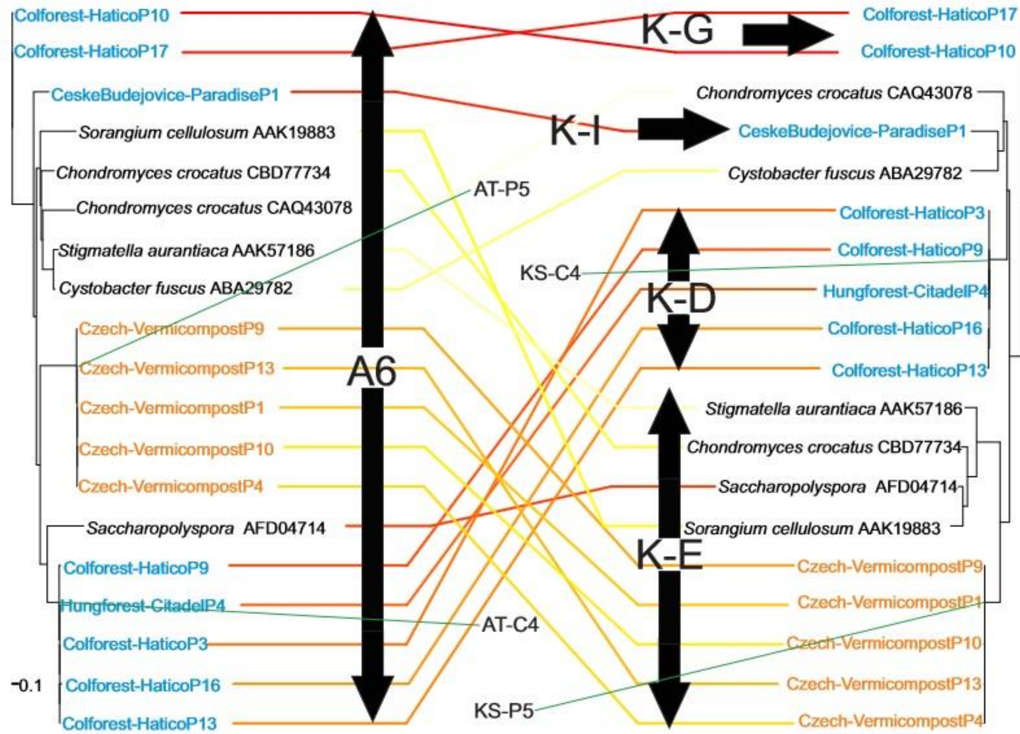


Figure S3 AT A6/KS-K7 D,E,G,I cophylogeny. Sequences in blue are from soil, in red from street sediments and in orange from earthworm-affected environments.

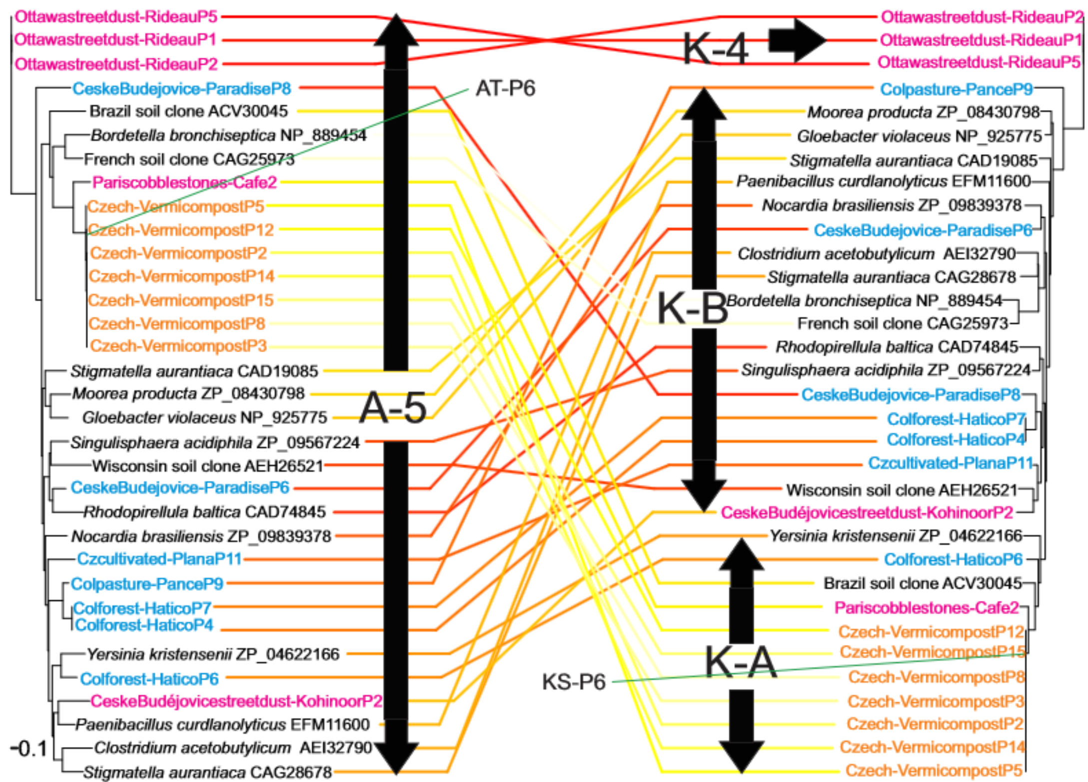


Figure S4 AT A5/KS-K4 A, B cophylogeny. **Sequences in blue are from soil, in red from street sediments and in orange from earthworm-affected environments.**

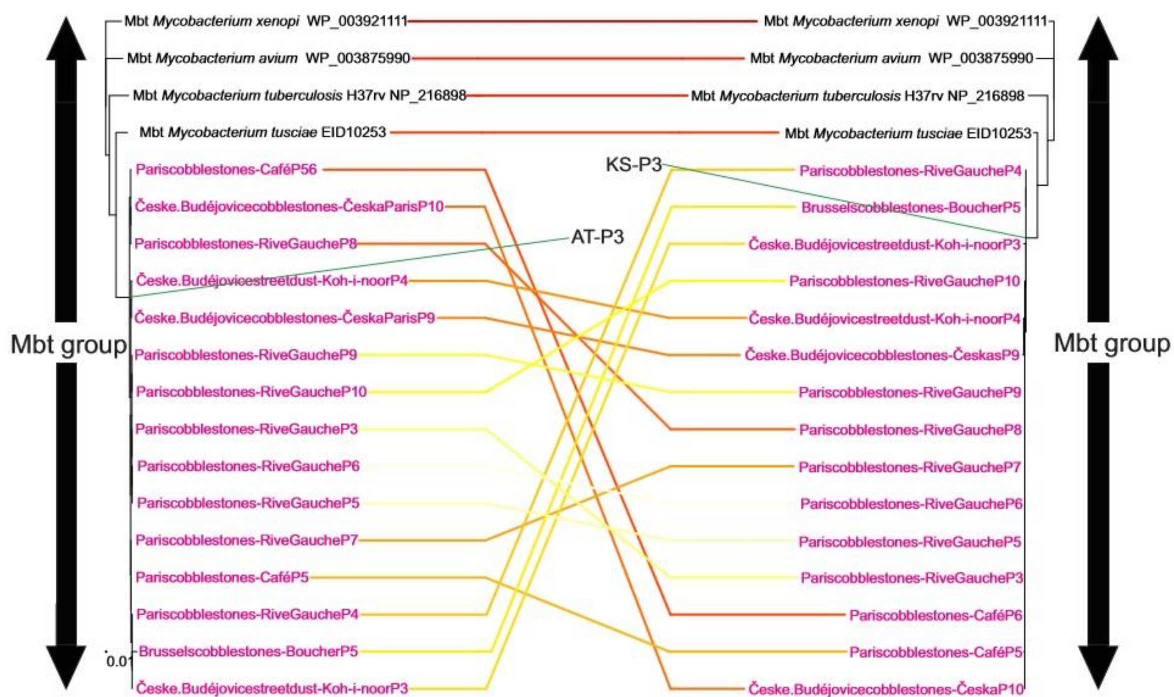


Figure S5 Mbt group cophylogeny. **Sequences in blue are from soil, in red from street sediments and in orange from earthworm-affected environments.**

Appendix 4 Supplementary Tables Chapter 2

Supplemental Figure 1 Non parametric comparison of alpha diversity from Illumina V4V5 sequencing from EL4. Samples are 11 soils (all except Exp4A) and 3 compost samples. Values represent the mean and standard deviation.			
	Soft	**Mobio**	**p-value**
Chao	3794 ± 1291	5005 ± 1110	0.013
Observed OTUs	1673 ± 481	1885 ± 395	0.245
Shannon	8.5 ± 1.1	8.9 ± 0.8	0.361
Simpson	0.98 ± 0.03	0.991 ± 0.006	0.277
Faiths PD	141 ± 40	155 ± 38	0.399

Supplemental Table 2

Non parametric comparison of alpha diversity from Soft and Mobio extraction and Illumina V4V5 sequencing from EL4 and Pyrosequencing V4V6 sequencing from EL1.

Samples are the soils Exp1B, Exp2A, Exp2B, BillingsBridge, Vancouver.

	Group1	**Group2**	**Group1 (mean ± std)**	**Group2 (mean ± std)**	**p-value**
Chao	Soft	Mobio	1307 ± 330	1360 ± 390	0.762
Chao	Pyrosequencing V4V6	Illumina V4V5	1092 ± 279	1575 ± 262	0.002
Observed OTUs	Soft	Mobio	550 ± 87	549 ± 98	0.99
Observed OTUs	Pyrosequencing V4V6	Illumina V4V5	525 ± 101	573 ± 76	0.26
Shannon	Soft	Mobio	8.05 ± 0.45	8.16 ± 0.50	0.652
Shannon	Pyrosequencing V4V6	Illumina V4V5	8.12 ± 0.52	8.10 ± 0.44	0.923
Simpson	Soft	Mobio	0.988 ± 0.007	0.992 ± 0.004	0.16
Simpson	Pyrosequencing V4V6	Illumina V4V5	0.991 ± 0.005	0.988 ± 0.006	0.433
Faiths PD	Soft	Mobio	45 ± 9	45 ± 9	0.964
Faiths PD	Pyrosequencing V4V6	Illumina V4V5	40 ± 8	51 ± 7	0.014

Supplemental Table 3				
Anosim analysis of Beta diversity using 999 permutations and weighted and unweighted UniFrac and Bray Curtis clustering of Illumina V4V5 sequencing from EL4 and Pyrosequencing V4V6 sequencing from EL1. Samples are the soils Exp1B, Exp2A, Exp2B, BillingsBridge, Van. 999 permutations were used.				
Clustering	**Number of Samples**	**Number of Groups**	**Groups**	**p-value**
UniFrac Weighted	20	2	Mobio vs Soft Extraction	0.414
UniFrac Weighted	20	2	V4V6Pyrosequencing vs V4V5Illumina	0.001
UniFrac Unweighted	20	2	Mobio vs Soft Extraction	0.878
UniFrac Unweighted	20	2	V4V6Pyrosequencing vs V4V5Illumina	0.001
Bray-Curtis	20	2	Mobio vs Soft Extraction	0.4
Bray-Curtis	20	2	V4V6Pyrosequencing vs V4V5Illumina	0.005

Supplemental Table 4

Anosim analysis of Beta diversity using 999 permutations and weighted and unweighted UniFrac and Bray Curtis clustering of Illumina V4V5 sequencing from EL4 using Soft and Mobio extraction for all soils and compost samples (n=28) all soils (n=22) and Ottawa soils (n=14).

Clustering	Number of Samples	Number of Groups	Groups	p-value
UniFrac Weighted	28	2	Mobio vs Soft Extraction	0.407
UniFrac Weighted	22	2	Mobio vs Soft Extraction	0.125
UniFrac Weighted	14	2	Mobio vs Soft Extraction	0.017
UniFrac Unweighted	28	2	Mobio vs Soft Extraction	0.832
UniFrac Unweighted	22	2	Mobio vs Soft Extraction	0.816
UniFrac Unweighted	14	2	Mobio vs Soft Extraction	0.088
Bray-Curtis	28	2	Mobio vs Soft Extraction	0.466
Bray-Curtis	22	2	Mobio vs Soft Extraction	0.41
Bray-Curtis	14	2	Mobio vs Soft Extraction	0.072

Supplemental Data Table 5-Relative distribution of the orders a) Pseudomonas b) Enterobacteria c) Alteromonadales d)Desulfuromonadales, e)JG30-KF-CM45 in amplicons generated in this study.

Note: Soil_Exp1B-2 and Soil_Exp2A-2 refers to a comparison between the soft method and a doubled Mobio bead beating time. Soil_Exp1B3 refers to a comparison between soft extraction and Retsch GmbH beadbeating.

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Pseudomonas count	% Pseudomonas Mobio	Soft total count	Soft Pseudomonas	% Pseudomonas soft	p-values	p-values corrected
Soil_Exp1B	EL1	Pyrosequencing	V4V6	>0.81	1925	0	0.00	2367	1	0.04	1.00	0.95
SoilExp1B-1	EL2	Pyrosequencing	V4V6	undefined	2654	0	0.00	1945	0	0.00	1.00	0.92
SoilExp1B-2	EL2	Pyrosequencing	V4V6	0.57	2343	2	0.09	2052	1	0.05	0.50	0.93
SoilExp1B-3	EL2	Pyrosequencing	V4V6	>2.4	2267	0	0.00	1873	2	0.11	0.20	0.90
Soil_Exp1B	EL3	Pyrosequencing	V1V3	22.40	4114	27	0.66	4510	663	14.70	0.00	0.00
Soil_Exp1B	EL4	Illumina	V4V5	52.13	19930	15	0.08	19983	784	3.92	1.19E-212	1.33E-209
Soil_Exp1B	EL4	Illumina	V6V8	13.19	16829	43	0.26	25925	874	3.37	0.00	0.00
Soil_Exp2A	EL1	Pyrosequencing	V4V6	undefined	1947	0	0.00	2174	0	0.00	1.00	0.94
Soil_Exp2A-1	EL2	Pyrosequencing	V4V6	>1.66	2011	0	0.00	2413	2	0.08	0.50	0.93
Soil_Exp2A-2	EL2	Pyrosequencing	V4V6	>2.08	2517	0	0.00	2413	2	0.08	0.24	0.79
Soil_Exp2A	EL4	Illumina	V4V5	118.26	19987	28	0.14	19998	3313	16.57	0.00	0.00
Soil_Exp2A	EL4	Illumina	V6V8	111.84	16331	27	0.17	15597	2884	18.49	0.00	0.00
Soil_Vancouver	EL1	Pyrosequencing	V4V6	1.24	1899	3	0.16	3566	7	0.20	1.00	0.98
Soil_Vancouver	EL2	Pyrosequencing	V4V6	1.49	1856	2	0.11	2494	4	0.16	1.00	0.94
Soil_Vancouver	EL4	Illumina	V4V5	43.15	20655	53	0.26	19871	2200	11.07	0.00	0.00
Soil_Vancouver	EL4	Illumina	V6V8	28.91	14956	64	0.43	19140	2368	12.37	0.00	0.00
Soil_Exp2B	EL1	Pyrosequencing	V4V6	>0.61	1240	0	0.00	2048	1	0.05	1.00	0.95
Soil_Exp2B	EL4	Illumina	V4V5	1.05	20136	30	0.15	20379	32	0.16	0.94	1.00
Soil_BillingsBridge	EL1	Pyrosequencing	V4V6	undefined	1323	0	0.00	1821	0	0.00	1.00	0.98
Soil_BillingsBridge	EL1	Illumina	V4V5	3.62	20073	420	2.09	20189	1529	7.57	0.00	0.00
Soil_Exp4A	EL2	Pyrosequencing	V4V6	>1.06	1836	1	0.05	1727	0	0.00	1.00	0.97
Soil_Exp3A	EL4	Illumina	V4V5	10.69	18541	63	0.34	21097	766	3.63	0.00	0.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Pseudomonas count	% Pseudomonas Mobio	Soft total count	Soft Pseudomonas	% Pseudomonas soft	p-values	p-values corrected
Soil_Exp3B	EL4	Illumina	V4V5	0.60	19915	235	1.18	20181	144	0.71	1.62E-06	0.00
Soil_RACentre	EL4	Illumina	V4V5	5.62	20039	316	1.58	7911	701	8.86	0.00	0.00
Soil_Mattawa	EL4	Illumina	V4V5	4.24	19838	2	0.01	14050	6	0.04	0.07	1.00
Soil_Ellisworth	EL4	Illumina	V4V5	1.03	19650	28	0.14	21073	31	0.15	0.99	1.00
Soil_Diamond	EL4	Illumina	V4V5	3.63	20279	99	0.49	20582	365	1.77	0.00	0.00
Compost_80F	EL4	Illumina	V4V5	1.47	20038	1561	7.79	12557	1438	11.45	0.00	0.00
Compost_80FTop	EL4	Illumina	V4V5	1.17	20117	1131	5.62	19869	1310	6.59	5.44E-05	0.01
Compost_GlebeCentre	EL4	Illumina	V4V5	1.40	20086	22	0.11	20206	31	0.15	0.28	1.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Enterobacterial count	% Enterobacterial	Soft total count	Soft Enterobacterial	% Enterobacterial soft	p-values	p-values (corrected)
Soil_Exp1B	EL1	Pyrosequencing	V4V6	>91.90	1925	0	0.00	2367	113	4.77	2.71E-30	5.16E-28
SoilExp1B-1	EL2	Pyrosequencing	V4V6	53.08	2654	2	0.08	1945	82	4.00	1.01E-28	1.87E-26
SoilExp1B-2	EL2	Pyrosequencing	V4V6	>116.46	2343	0	0.00	2052	102	4.97	4.62E-35	8.61E-33
SoilExp1B-3	EL2	Pyrosequencing	V4V6	>90.77	2267	0	0.00	1873	75	4.00	6.40E-27	1.21E-24
Soil_Exp1B	EL3	Pyrosequencing	V1V3	139.57	4114	1	0.02	4510	153	3.39	2.08E-42	1.41E-40
Soil_Exp1B	EL4	Illumina	V4V5	>239.36	19930	0	0.00	19983	240	1.20	5.78E-73	5.39E-70
Soil_Exp1B	EL4	Illumina	V6V8	75.95	16829	5	0.03	25925	585	2.26	4.08E-119	7.96E-117
Soil_Exp2A	EL1	Pyrosequencing	V4V6	>1.79	1947	0	0.00	2174	2	0.09	0.50	0.94
Soil_Exp2A-1	EL2	Pyrosequencing	V4V6	>7.50	2011	0	0.00	2413	9	0.37	0.01	0.09
Soil_Exp2A-2	EL2	Pyrosequencing	V4V6	3.13	2517	3	0.12	2413	9	0.37	0.09	0.61
Soil_Exp2A	EL4	Illumina	V4V5	3.16	19987	6	0.03	19998	19	0.10	0.01	1.00
Soil_Exp2A	EL4	Illumina	V6V8	>20.94	16331	0	0.00	15597	20	0.13	5.95E-07	5.49E-05
Soil_Vancouver	EL1	Pyrosequencing	V4V6	12.35	1899	5	0.26	3566	116	3.25	2.32E-34	4.37E-32
Soil_Vancouver	EL2	Pyrosequencing	V4V6	15.13	1856	9	0.48	2494	183	7.34	3.94E-16	7.75E-14
Soil_Vancouver	EL4	Illumina	V4V5	10.85	20655	34	0.16	19871	355	1.79	0	0
Soil_Vancouver	EL4	Illumina	V6V8	13.35	14956	36	0.24	19140	615	3.21	0	0
Soil_Exp2B	EL1	Pyrosequencing	V4V6	>1.82	1240	0	0.00	2048	3	0.15	0.30	0.95
Soil_Exp2B	EL4	Illumina	V4V5	1.32	20136	3	0.01	20379	4	0.02	1.00	1.00
Soil_BillingsBridge	EL1	Pyrosequencing	V4V6	24.70	1323	2	0.15	1821	68	3.73	2.35E-14	2.31E-12
Soil_BillingsBridge	EL1	Illumina	V4V5	31.07	20073	8	0.04	20189	250	1.24	1.18E-63	4.39E-61
Soil_Exp4A	EL2	Pyrosequencing	V4V6	65.91	1836	1	0.05	1727	62	3.59	1.02E-18	1.99E-16
Soil_Exp3A	EL4	Illumina	V4V5	18.81	18541	5	0.03	21097	107	0.51	1.58E-23	1.27E-20
Soil_Exp3B	EL4	Illumina	V4V5	4.77	19915	12	0.06	20181	58	0.29	2.29E-08	7.55E-06
Soil_RACentre	EL4	Illumina	V4V5	15.09	20039	142	0.71	7911	846	10.69	0.00	0.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Enterobacterial count	% Enterobacterial	Soft total count	Soft Enterobacterial	% Enterobacterial soft	p-values	p-values (corrected)
Soil_Mattawa	EL4	Illumina	V4V5	>2.82	19838	0	0.00	14050	2	0.01	0.17	1.00
Soil_Ellisworth	EL4	Illumina	V4V5	<0.23	19650	4	0.02	21073	0	0.00	0.05	1.00
Soil_Diamond	EL4	Illumina	V4V5	8.87	20279	10	0.05	20582	90	0.44	3.34E-17	9.84E-15
Compost_80F	EL4	Illumina	V4V5	1.26	20038	637	3.18	12557	504	4.01	8.51E-05	0.014851723
Compost_80FTop	EL4	Illumina	V4V5	5.41	20117	203	1.01	19869	1085	5.46	0	0
Compost_GlebeCentre	EL4	Illumina	V4V5	4.14	20086	6	0.03	20206	25	0.12	0.00	0.14

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Alteromonadales count	%Alteromonadales Mobio	Soft total count	Soft Alteromonadales	% Alteromonadales	p-values	p-values (corrected)
Soil_Exp1B	EL1	Pyrosequencing	V4V6	undefined	1925	0	0.00	2367	0	0.00	1.00	1.00
SoilExp1B-1	EL2	Pyrosequencing	V4V6	undefined	2654	0	0.00	1945	0	0.00	1.00	1.00
SoilExp1B-2	EL2	Pyrosequencing	V4V6	undefined	2343	0	0.00	2052	0	0.00	1.00	1.00
SoilExp1B-3	EL2	Pyrosequencing	V4V6	undefined	2267	0	0.00	1873	0	0.00	1.00	1.00
Soil_Exp1B	EL3	Pyrosequencing	V1V3	1.37	4114	2	0.05	4510	3	0.07	1.00	0.85
Soil_Exp1B	EL4	Illumina	V4V5	undefined	19930	0	0.00	19983	0	0.00	1.00	1.00
Soil_Exp1B	EL4	Illumina	V6V8	0.11	16829	6	0.04	25925	1	0.00	0.02	0.45
Soil_Exp2A	EL1	Pyrosequencing	V4V6	undefined	1947	0	0.00	2174	0	0.00	1.00	1.00
Soil_Exp2A-1	EL2	Pyrosequencing	V4V6	undefined	2011	0	0.00	2413	0	0.00	0.50	0.93
Soil_Exp2A-2	EL2	Pyrosequencing	V4V6	undefined	2517	0	0.00	2413	0	0.00	1.00	1.00
Soil_Exp2A	EL4	Illumina	V4V5	2.22	19987	9	0.05	19998	20	0.10	0.06	1.00
Soil_Exp2A	EL4	Illumina	V6V8	1.62	16331	11	0.07	15597	17	0.11	0.26	1.00
Soil_Vancouver	EL1	Pyrosequencing	V4V6	undefined	1899	0	0.00	3566	0	0.00	1.00	0.98
Soil_Vancouver	EL2	Pyrosequencing	V4V6	undefined	1856	0	0.00	2494	0	0.00	1.00	0.94
Soil_Vancouver	EL4	Illumina	V4V5	undefined	20655	0	0.00	19871	0	0.00	1.00	1.00
Soil_Vancouver	EL4	Illumina	V6V8	0.52	14956	3	0.02	19140	2	0.01	0.66	0.99
Soil_Exp2B	EL1	Pyrosequencing	V4V6	undefined	1240	0	0.00	2048	0	0.00	0.02	0.27
Soil_Exp2B	EL4	Illumina	V4V5	0.81	20136	11	0.05	20379	9	0.04	0.66	1.00
Soil_BillingsBridge	EL1	Pyrosequencing	V4V6	undefined	1323	0	0.00	1821	0	0.00	0.94	0.00
Soil_BillingsBridge	EL1	Illumina	V4V5	0.88	20073	9	0.04	20189	8	0.04	0.81	1.00
Soil_Exp4A	EL2	Pyrosequencing	V4V6	undefined	1836	0	0.00	1727	0.00	0.00	1.00	0.94
Soil_Exp3A	EL4	Illumina	V4V5	3.87	18541	20	0.11	21097	88	0.42	1.42E-09	6.11E-07
Soil_Exp3B	EL4	Illumina	V4V5	0.23	19915	65	0.33	20181	15	0.07	7.03E-09	2.46E-06
Soil_RACentre	EL4	Illumina	V4V5	0.63	20039	4	0.02	7911	1	0.01	1.00	1.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Alteromonadales count	%Alteromonadales Mobio	Soft total count	Soft Alteromonadales	% Alteromonadales	p-values	p-values (corrected)
Soil_Mattawa	EL4	Illumina	V4V5	undefined	19838	0	0.00	14050	0	0.00	1.00	1.00
Soil_Ellisworth	EL4	Illumina	V4V5	2.61	19650	5	0.03	21073	14	0.07	0.07	1.00
Soil_Diamond	EL4	Illumina	V4V5	16.89	20279	71	0.35	20582	1217	5.91	0.00	0.00
Compost_80F	EL4	Illumina	V4V5	3.62	20038	347	1.73	12557	787	6.27	0.00	0.00
Compost_80FTop	EL4	Illumina	V4V5	<0.03	20117	31	0.15	19869	0	0.00	9.37E-10	2.62E-07
Compost_GlebeCentre	EL4	Illumina	V4V5	2.53	20086	20	0.10	20206	51	0.25	0.00	0.06

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Desulfuromonadales	%Desulfuromonadales Mobio	Soft total count	Soft Desulfuromonadale	% Desulfuromonadale	p-values	p-values (corrected)
Soil_Exp1B	EL1	Pyrosequencing	V4V6	2.44	1925	2	0.10	2367	6	0.25	0.31	0.95
SoilExp1B-1	EL2	Pyrosequencing	V4V6	undefined	2654	0	0.00	1945	0	0.00	1.00	0.92
SoilExp1B-2	EL2	Pyrosequencing	V4V6	7.99	2343	1	0.04	2052	7	0.34	0.03	0.26
SoilExp1B-3	EL2	Pyrosequencing	V4V6	>6.05	2267	0	0.00	1873	5	0.27	0.02	0.21
Soil_Exp1B	EL3	Pyrosequencing	V1V3	2.74	4114	1	0.02	4510	3	0.07	0.63	0.81
Soil_Exp1B	EL4	Illumina	V4V5	4.24	19930	12	0.06	19983	51	0.26	7.33E-07	2.05E-04
Soil_Exp1B	EL4	Illumina	V6V8	10.17	16829	6	0.04	25925	94	0.36	3.67E-14	5.57E-12
Soil_Exp2A	EL1	Pyrosequencing	V4V6	>5.37	1947	0	0.00	2174	6	0.28	0.03	0.34
Soil_Exp2A-1	EL2	Pyrosequencing	V4V6	>2.09	2517	0	0.00	2413	2	0.08	0.79	0.08
Soil_Exp2A-2	EL2	Pyrosequencing	V4V6	>1.67	2011	0	0.00	2413	2	0.08	0.24	0.79
Soil_Exp2A	EL4	Illumina	V4V5	2.00	19987	7	0.04	19998	14	0.07	0.19	1.00
Soil_Exp2A	EL4	Illumina	V6V8	2.62	16331	8	0.05	15597	20	0.13	0.02	0.56
Soil_Vancouver	EL1	Pyrosequencing	V4V6	2.13	1899	1	0.05	3566	4	0.11	0.66	0.98
Soil_Vancouver	EL2	Pyrosequencing	V4V6	undefined	1856	0	0.00	2494	0	1	0.94	0.98
Soil_Vancouver	EL4	Illumina	V4V5	0.95	20655	12	0.06	19871	11	0.06	1.00	1.00
Soil_Vancouver	EL4	Illumina	V6V8	1.66	14956	8	0.05	19140	17	0.09	0.31	0.99
Soil_Exp2B	EL1	Pyrosequencing	V4V6	1.85	1240	18	1.45	2048	55	2.69	0.02	0.27
Soil_Exp2B	EL4	Illumina	V4V5	0.89	20136	115	0.57	20379	104	0.51	0.44	1.00
Soil_BillingsBridge	EL1	Pyrosequencing	V4V6	3.78	1323	5	0.38	1821	26	1.43	0.00	0.10
Soil_BillingsBridge	EL1	Illumina	V4V5	38.78	20073	7	0.03	20189	273	1.35	1.42E-71	5.69E-69
Soil_Exp4A	EL2	Pyrosequencing	V4V6	3.72	1836	2	0.11	1727	7	0.41	0.10	0.86
Soil_Exp3A	EL4	Illumina	V4V5	3.04	18541	24	0.13	21097	83	0.39	3.05E-07	9.49E-05
Soil_Exp3B	EL4	Illumina	V4V5	8.93	19915	43	0.22	20181	389	1.93	0.00	0.00
Soil_RACentre	EL4	Illumina	V4V5	2.53	20039	2	0.01	7911	2	0.03	0.32	1.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio Desulfuromonadales	%Desulfuromonadales Mobio	Soft total count	Soft Desulfuromonadale	% Desulfuromonadale	p-values	p-values (corrected)
Soil_Mattawa	EL4	Illumina	V4V5	undefined	19838	0	0.00	14050	0	0.00	1.00	1.00
Soil_Ellisworth	EL4	Illumina	V4V5	1.96	19650	10	0.05	21073	21	0.10	0.10	1.00
Soil_Diamond	EL4	Illumina	V4V5	undefined	20279	0	0.00	20582	0	0.00	1.00	1.00
Compost_80F	EL4	Illumina	V4V5	undefined	20038	0	0.00	12557	0	0.00	1.00	1.00
Compost_80FTop	EL4	Illumina	V4V5	undefined	20117	0	0.00	19869	0	0.00	1.00	1.00
Compost_GlebeCentre	EL4	Illumina	V4V5	undefined	20086	0	0.00	20206	0	0.00	1.00	1.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio JG30-KF-CM45 count	% JG30-KF-CM45 Mobio	Soft total count	Soft JG30-KF-CM45 count	% JG30-KF-CM45 soft	p-values	p-values (corrected)
Soil_Exp1B	EL1	Pyrosequencing	V4V6	<0.20	1925	4	0.21	2367	0	0.00	0.04	0.38
SoilExp1B-1	EL2	Pyrosequencing	V4V6	2.73	2654	2	0.08	1945	4	0.21	0.25	0.92
SoilExp1B-2	EL2	Pyrosequencing	V4V6	3.43	2343	1	0.04	2052	3	0.15	0.35	0.93
SoilExp1B-3	EL2	Pyrosequencing	V4V6	1.21	2267	1	0.04	1873	1	0.05	1.00	0.94
Soil_Exp1B	EL3	Pyrosequencing	V1V3	<0.91	4114	1	0.02	4510	0	0.00	0.48	0.70
Soil_Exp1B	EL4	Illumina	V4V5	0.72	19930	29	0.15	19983	21	0.11	0.32	1.00
Soil_Exp1B	EL4	Illumina	V6V8	1.02	16829	53	0.31	25925	83	0.32	1.00	1.00
Soil_Exp2A	EL1	Pyrosequencing	V4V6	0.90	1947	2	0.10	2174	2	0.09	1.00	0.94
Soil_Exp2A-1	EL2	Pyrosequencing	V4V6	0.42	2011	2	0.10	2413	1	0.04	0.59	0.93
Soil_Exp2A-2	EL2	Pyrosequencing	V4V6	0.26	2517	4	0.16	2413	1	0.04	0.38	0.91
Soil_Exp2A	EL4	Illumina	V4V5	0.80	19987	20	0.10	19998	16	0.08	0.51	1.00
Soil_Exp2A	EL4	Illumina	V6V8	0.68	16331	51	0.31	15597	33	0.21	0.10	1.00
Soil_Vancouver	EL1	Pyrosequencing	V4V6	undefined	1899	0	0.00	3566	0	0.00	1.00	0.94
Soil_Vancouver	EL2	Pyrosequencing	V4V6	undefined	1856	0	0.00	2494	0	0.00	1.00	0.98
Soil_Vancouver	EL4	Illumina	V4V5	undefined	20655	0	0.00	19871	0	0.00	1.00	1.00
Soil_Vancouver	EL4	Illumina	V6V8	undefined	14956	0	0.00	19140	0	0.00	1.00	0.99
Soil_Exp2B	EL1	Pyrosequencing	V4V6	0.09	1240	7	0.56	2048	1	0.05	0.01	0.11
Soil_Exp2B	EL4	Illumina	V4V5	1.04	20136	35	0.17	20379	37	0.18	0.95	1.00
Soil_BillingsBridge	EL1	Pyrosequencing	V4V6	undefined	1323	0	0.00	1821	0	0.00	1.00	0.98
Soil_BillingsBridge	EL1	Illumina	V4V5	0.40	20073	10	0.05	20189	4	0.02	0.12	1.00
Soil_Exp4A	EL2	Pyrosequencing	V4V6	0.27	1836	4	0.22	1727	1	0.06	0.38	0.97
Soil_Exp3A	EL4	Illumina	V4V5	0.39	18541	27	0.15	21097	12	0.06	0.01	0.54
Soil_Exp3B	EL4	Illumina	V4V5	1.48	19915	12	0.06	20181	18	0.09	0.36	1.00
Soil_RACentre	EL4	Illumina	V4V5	>0.18	20039	14	0.07	7911	0	0.00	0.01	1.00

Sample	Lab	Sequencing method	Primers	Soft/Mobio ratio	Mobio total count	Mobio JG30-KF-CM45 count	% JG30-KF-CM45 Mobio	Soft total count	Soft JG30-KF-CM45 count	% JG30-KF-CM45 soft	p-values	p-values (corrected)
Soil_Mattawa	EL4	Illumina	V4V5	undefined	19838	0	0.00	14050	0	0.00	1.00	1.00
Soil_Ellisworth	EL4	Illumina	V4V5	0.02	19650	217	1.10	21073	5	0.02	2.26E-61	5.26E-59
Soil_Diamond	EL4	Illumina	V4V5	<0.003	20279	353	1.74	20582	0	0	1.02E-84	3.36E-82
Compost_80F	EL4	Illumina	V4V5	0.05	20038	31	0.15	12557	1	0.01	4.56E-06	8.50E-04
Compost_80FTop	EL4	Illumina	V4V5	<0.34	20117	3	0.01	19869	0	0.00	0.25	1.00
Compost_GlebeCentre	EL4	Illumina	V4V5	0.17	20086	78	0.39	20206	13	0.06	8.94E-13	3.85E-10

Supplemental Data Table 6-Relative distribution of the orders a) Pseudomonas b) Enterobacteria c) Alteromonadales d)Desulfuromonadales, e)JG30-KF-CM45 f)Bacillales in amplicons generated in this study. Results are divided into genera. Results are not shown where these orders were not found.

Note: Soil_Exp1B-2 and Soil_Exp2A-2 refers to a comparison between the soft method and a doubled Mobio bead beating time. Soil_Exp1B3 refers to a comparison between soft extraction and Retsch GmbH beadbeating.

Sample	Sequencing method	Lab/ Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	1925	0	0.00	2367	1	0.04	1.00	0.98
Exp1B2	Pyrosequencing	EL2V4V6	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	2343	2	0.09	2052	0	0.00	0.50	0.50
Exp1B3	Pyrosequencing	EL2V4V6	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	2267	0	0.00	1873	1	0.05	0.45	0.45
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	2267	0	0.00	1873	1	0.05	0.45	0.45
Exp1B	Pyrosequencing	EL3V1V3	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	4114	24	0.58	4510	576	12.77	2.75E-135	9.40E-133
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	4114	3	0.07	4510	84	1.86	2.36E-20	4.04E-18
			o__Pseudomonadales	f__u_Pseudomonadales	g__u_Pseudomonadales	4114	0	0.00	4510	3	0.07	0.25	0.62
Exp1B	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	19930	10	0.05	19983	664	3.32	7.91E-184	4.90E-180
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	19930	5	0.03	19983	120	0.60	9.86E-30	1.53E-26
Exp1B	Illumina	EL4V6V8	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	16829	41	0.24	25925	871	3.36	4.42E-137	7.52E-134
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	16829	1	0.01	25925	0	0.00	0.39	0.98
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	16829	0	0.00	25925	2	0.01	0.52	0.98
			o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	16829	1	0.01	25925	1	0.00	1.00	0.98
Exp2A1	Pyrosequencing	EL2V4V6	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	2011	0	0.00	2413	2	0.08	0.50	0.50
Exp2A	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	19987	13	0.07	19998	3250	16.25	0.00	0.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	19987	14	0.07	19998	63	0.32	1.37E-08	6.98E-06
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	19987	1	0.01	19998	0	0.00	0.50	1.00
Exp2A	Illumina	EL4V6V8	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	16331	27	0.17	15597	2863	18.36	0.00	0.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	16331	0	0.00	15597	17	0.11	5.11E-06	0.000335
			o__Pseudomonadales	f__u_Pseudomonadales	g__u_Pseudomonadales	16331	0	0.00	15597	4	0.03	0.06	0.78
Vancouver	Pyrosequencing	EL1V4V6	o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	1899	2	0.11	3566	4	0.11	1.00	1.00
			o__Pseudomonadales	f__Moraxellaceae	g__Perlucidibaca	1899	1	0.05	3566	3	0.08	1.00	1.00

Table S6-2

Sample	Sequencing method	Lab/ Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Vancouver	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20655	11	0.05	19871	1999	10.06	0.00	0.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20655	20	0.10	19871	186	0.94	1.51E-36	1.56E-33
			o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	20655	16	0.08	19871	6	0.03	0.05	1.00
			o__Pseudomonadales	f__Moraxellaceae	g__Perlucidibaca	20655	6	0.03	19871	9	0.05	0.45	1.00
Vancouver	Illumina	EL4V6V8	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	14956	28	0.19	19140	2314	12.09	0.00	0.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	14956	2	0.01	19140	39	0.20	3.35E-08	3.87E-06
			o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	14956	23	0.15	19140	6	0.03	0.00	0.01
			o__Pseudomonadales	f__Moraxellaceae	g__Perlucidibaca	14956	11	0.07	19140	2	0.01	0.00	0.19
			o__Pseudomonadales	f__u_Pseudomonadales	g__u_Pseudomonadales	14956	0	0.00	19140	7	0.04	0.02	0.80
Billings B	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20073	259	1.29	20189	1465	7.26	9.85E-211	6.09E-207
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20073	161	0.80	20189	64	0.32	4.48E-11	7.30E-09
Exp2B	Pyrosequencing	EL1V4V6	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	1240	0	0.00	2048	1	0.05	1.00	0.97
Exp2B	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20136	9	0.04	20379	17	0.08	0.17	1.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20136	19	0.09	20379	14	0.07	0.39	1.00
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	20136	2	0.01	20379	1	0.00	0.62	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	1836	1	0.05	1727	0	0.00	1.00	1.00
Exp3A	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	18541	40	0.22	21097	361	1.71	1.39E-57	4.30E-54
			o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	18541	0	0.00	21097	1	0.00	1.00	1.00
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	18541	0	0.00	21097	1	0.00	1.00	1.00
Exp3B	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	19915	114	0.57	20181	62	0.31	5.71E-05	0.008222
			o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	19915	115	0.58	20181	79	0.39	0.01	0.52
			o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	19915	6	0.03	20181	3	0.01	0.34	1.00
RACentre	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20039	127	0.63	7911	418	5.28	0.00	0.00
			o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20039	189	0.94	7911	283	3.58	0.00	0.00

Table S6-3

Sample	Sequencing method	Lab/ Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20279	34	0.17	20582	236	1.15	4.02E-38	1.38E-35
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20279	59	0.29	20582	103	0.50	0.00	0.05
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	20279	0	0.00	20582	7	0.03	0.02	0.66
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	20279	5	0.02	20582	12	0.06	0.14	1.00
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Alkanindiges	20279	1	0.00	20582	4	0.02	0.38	1.00
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Psychrobacter	20279	0	0.00	20582	2	0.01	0.50	1.00
Diamond	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Perlucidibaca	20279	0	0.00	20582	1	0.00	1.00	1.00
Glebe	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20086	15	0.07	20206	18	0.09	0.73	1.00
Glebe	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20086	7	0.03	20206	12	0.06	0.36	1.00
Glebe	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Perlucidibaca	20086	0	0.00	20206	1	0.00	1.00	1.00
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20038	706	3.52	12557	763	6.08	0.00	0.00
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	20038	162	0.81	12557	210	1.67	3.10E-12	6.38E-10
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__u_Moraxellaceae	20038	15	0.07	12557	2	0.02	0.02	1.00
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20038	678	3.38	12557	461	3.67	0.18	1.00
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__u_Pseudomonadales	g__u_Pseudomonadales	20038	0	0.00	12557	1	0.01	0.39	1.00
80F	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Serpens	20038	0	0.00	12557	1	0.01	0.39	1.00
80FTop	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__Pseudomonas	20117	175	0.87	19869	913	4.60	0.00	0.00
80FTop	Illumina	EL4V4V5	o__Pseudomonadales	f__Pseudomonadaceae	g__u_Pseudomonadaceae	20117	676	3.36	19869	303	1.52	0.00	0.00
80FTop	Illumina	EL4V4V5	o__Pseudomonadales	f__Moraxellaceae	g__Acinetobacter	20117	279	1.39	19869	94	0.47	0.00	0.00
80FTop	Illumina	EL4V4V5	o__Pseudomonadales	f__u_Pseudomonadales	g__u_Pseudomonadales	20117	1	0.00	19869	0	0.00	1.00	1.00

Table S6-4													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	1925	0	0.00	2367	85	3.59	5.89E-23	2.82E-20
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1925	0	0.00	2367	19	0.80	1.75E-05	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	1925	0	0.00	2367	7	0.30	0.02	0.40
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	1925	0	0.00	2367	1	0.04	1.00	0.98
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	1925	0	0.00	2367	1	0.04	1.00	0.98
TableS 6-3													
Exp1B1	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	2654	0	0.00	1945	42	2.16	1.55E-16	1.55E-16
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	2654	1	0.04	1945	29	1.49	2.29E-10	2.29E-10
			o__Enterobacteriales	f__Enterobacteriaceae	g__Rahnella	2654	0	0.00	1945	3	0.15	0.08	0.08
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	2654	0	0.00	1945	3	0.15	0.08	0.08
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	2654	0	0.00	1945	2	0.10	0.18	0.18
			o__Enterobacteriales	f__Enterobacteriaceae	g__Morganella	2654	0	0.00	1945	1	0.05	0.42	0.42
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	2654	1	0.04	1945	2	0.10	0.58	0.58
Exp1B2	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	2343	0	0.00	2052	51	2.49	9.65E-18	9.65E-18
			o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	2343	0	0.00	2052	41	2.00	2.21E-14	2.21E-14
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	2343	0	0.00	2052	5	0.24	0.02	0.02
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	2343	0	0.00	2052	3	0.15	0.10	0.10
			o__Enterobacteriales	f__Enterobacteriaceae	g__Rahnella	2343	0	0.00	2052	1	0.05	0.47	0.47
			o__Enterobacteriales	f__Enterobacteriaceae	g__Morganella	2343	0	0.00	2052	1	0.05	0.47	0.47
Exp1B3	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	2267	0	0.00	1873	36	1.92	3.30E-13	3.30E-13
			o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	2267	0	0.00	1873	25	1.33	2.24E-09	2.24E-09
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	2267	0	0.00	1873	7	0.37	0.00	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Salmonella	2267	0	0.00	1873	3	0.16	0.09	0.09
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	2267	0	0.00	1873	2	0.11	0.20	0.20
			o__Enterobacteriales	f__Enterobacteriaceae	g__Rahnella	2267	0	0.00	1873	1	0.05	0.45	0.45
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	2267	0	0.00	1873	1	0.05	0.45	0.45

Table S6-5													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL3V1V3	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	4114	0	0.00	4510	54	1.20	7.61E-16	8.68E-14
			o__Enterobacteriales	f__Enterobacteriaceae	g__Serratia	4114	0	0.00	4510	47	1.04	8.83E-14	7.56E-12
			o__Enterobacteriales	f__Enterobacteriaceae	g__Rahnella	4114	0	0.00	4510	10	0.22	0.00	0.03
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	4114	0	0.00	4510	10	0.22	0.00	0.03
			o__Enterobacteriales	f__Enterobacteriaceae	g__Ewingella	4114	0	0.00	4510	10	0.22	0.00	0.03
			o__Enterobacteriales	f__Enterobacteriaceae	g__Gluconacetobacter	4114	0	0.00	4510	8	0.18	0.01	0.07
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	4114	0	0.00	4510	6	0.13	0.03	0.20
			o__Enterobacteriales	f__Enterobacteriaceae	g__Trabulsiella	4114	0	0.00	4510	4	0.09	0.13	0.48
			o__Enterobacteriales	f__Enterobacteriaceae	g__Samsonia	4114	1	0.02	4510	0	0.00	0.48	0.67
			o__Enterobacteriales	f__Enterobacteriaceae	g__Hafnia	4114	0	0.00	4510	2	0.04	0.50	0.67
			o__Enterobacteriales	f__Enterobacteriaceae	g__Providencia	4114	0	0.00	4510	1	0.02	1.00	0.82
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pragia	4114	0	0.00	4510	1	0.02	1.00	0.82
Exp1B	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	19930	0	0.00	19983	232	1.16	1.55E-70	4.78E-67
			o__Enterobacteriales	f__Enterobacteriaceae	g__Providencia	19930	0	0.00	19983	7	0.04	0.02	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	19930	0	0.00	19983	1	0.01	1.00	1.00
Exp1B	Illumina	EL4V6V8	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	16829	5	0.03	25925	564	2.18	2.41E-114	2.05E-111
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	16829	0	0.00	25925	21	0.08	4.42E-05	0.001748
Exp2A	Pyrosequencing	EL1V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1947	0	0.00	2174	1	0.05	1.00	0.97
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	1947	0	0.00	2174	1	0.05	1.00	0.97
Exp2A1	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	2011	0	0.00	2413	2	0.08	0.50	0.50
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	2011	0	0.00	2413	2	0.08	0.50	0.50
			o__Enterobacteriales	f__Enterobacteriaceae	g__Shigella	2011	0	0.00	2413	2	0.08	0.50	0.50
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	2011	0	0.00	2413	2	0.08	0.50	0.50
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	2011	0	0.00	2413	1	0.04	1.00	1.00
Exp2A2	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	2011	0	0.00	2517	2	0.08	0.51	0.51
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	2011	0	0.00	2517	1	0.04	1.00	1.00
Exp2A	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	19987	6	0.03	19998	19	0.10	0.01	1.00
Exp2A	Illumina	EL4V6V8	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	16331	0	0.00	15597	19	0.12	1.22E-06	9.43E-05
			o__Enterobacteriales	f__Enterobacteriaceae	g__Plesiomonas	16331	0	0.00	15597	1	0.01	0.49	0.98

Table S6-6													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Vancouver	Pyrosequencing	EL1V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1899	3	0.16	3566	111	3.11	2.81E-17	1.37E-14
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	1899	2	0.11	3566	1	0.03	0.28	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	1899	0	0.00	3566	3	0.08	0.56	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	1899	0	0.00	3566	1	0.03	1.00	1.00
Vancouver	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	1856	9	0.48	2494	135	5.41	3.22E-23	3.22E-23
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1856	0	0.00	2494	28	1.12	1.90E-07	1.90E-07
			o__Enterobacteriales	f__Enterobacteriaceae	g__Rahnella	1856	0	0.00	2494	13	0.52	0.00	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Serratia	1856	0	0.00	2494	3	0.12	0.27	0.27
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	1856	0	0.00	2494	3	0.12	0.27	0.27
			o__Enterobacteriales	f__Enterobacteriaceae	g__Ewingella	1856	0	0.00	2494	1	0.04	1.00	1.00
Vancouver	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20655	31	0.15	19871	355	1.79	1.83E-74	3.77E-71
			o__Enterobacteriales	f__Enterobacteriaceae	g__Sodalis	20655	3	0.01	19871	0	0.00	0.25	1.00
Vancouver	Illumina	EL4V6V8	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	14956	36	0.24	19140	600	3.13	1.33E-106	1.15E-103
			o__Enterobacteriales	f__Enterobacteriaceae	g__Gluconacetobacter	14956	0	0.00	19140	13	0.07	0.00	0.06
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	14956	0	0.00	19140	2	0.01	0.51	1.00
Billings B	Pyrosequencing	EL1V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	1323	0	0.00	1821	60	3.29	4.33E-15	2.11E-12
			o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1323	0	0.00	1821	4	0.22	0.14	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	1323	2	0.15	1821	4	0.22	1.00	1.00
Billings B	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20073	8	0.04	20189	250	1.24	1.18E-63	1.46E-60
Exp2B	Pyrosequencing	EL1V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1240	0	0.00	2048	3	0.15	0.30	0.97
Exp2B	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20136	3	0.01	20379	4	0.02	1.00	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	1836	0	0.00	1727	44	2.55	1.09E-14	1.09E-14
			o__Enterobacteriales	f__Enterobacteriaceae	g__Yersinia	1836	1	0.05	1727	13	0.75	0.00	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Enterobacter	1836	0	0.00	1727	2	0.12	0.23	0.23
			o__Enterobacteriales	f__Enterobacteriaceae	g__Raoultella	1836	0	0.00	1727	1	0.06	0.48	0.48
			o__Enterobacteriales	f__Enterobacteriaceae	g__Escherichia	1836	0	0.00	1727	1	0.06	0.48	0.48
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	1836	0	0.00	1727	1	0.06	0.48	0.48

Table S6-7													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp3A	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	18541	5	0.03	21097	106	0.50	2.88E-23	2.23E-20
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	18541	0	0.00	21097	1	0.00	1.00	1.00
Exp3B	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	19915	10	0.05	20181	58	0.29	2.41E-09	8.29E-07
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	19915	2	0.01	20181	0	0.00	0.25	1.00
RACentre	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20039	139	0.69	7911	842	10.64	0.00	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	20039	3	0.01	7911	4	0.05	0.11	1.00
Mattawa	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	19838	0	0.00	14050	2	0.01	0.17	1.00
Ellsworth	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	19650	4	0.02	21073	0	0.00	0.05	1.00
Diamond	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20279	8	0.04	20582	86	0.42	1.34E-17	2.38E-15
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	20279	2	0.01	20582	4	0.02	0.69	1.00
Glebe	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20086	6	0.03	20206	24	0.12	0.00	0.13
			o__Enterobacteriales	f__Enterobacteriaceae	g__Providencia	20086	0	0.00	20206	1	0.00	1.00	1.00
80F	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20038	612	3.05	12557	493	3.93	3.08E-05	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	20038	20	0.10	12557	2	0.02	0.00	0.20
			o__Enterobacteriales	f__Enterobacteriaceae	g__Providencia	20038	2	0.01	12557	7	0.06	0.03	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Serratia	20038	1	0.00	12557	2	0.02	0.56	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Sodalis	20038	1	0.00	12557	0	0.00	1.00	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Pantoea	20038	1	0.00	12557	0	0.00	1.00	1.00
80FTop	Illumina	EL4V4V5	o__Enterobacteriales	f__Enterobacteriaceae	g__u_Enterobacteriaceae	20117	201	1.00	19869	1081	5.44	0.00	0.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Erwinia	20117	1	0.00	19869	3	0.02	0.37	1.00
			o__Enterobacteriales	f__Enterobacteriaceae	g__Providencia	20117	1	0.00	19869	1	0.01	1.00	1.00

Table S6-8

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL3V1V3	o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	4114	1	0.02	4510	0	0.00	0.48	0.67
			o__Alteromonadales	f__OM60	g__u_OM60	4114	1	0.02	4510	3	0.07	0.63	0.76
Exp1B	Illumina	EL4V6V8	o__Alteromonadales	f__211ds20	g__u_211ds20	16829	2	0.01	25925	0	0.00	0.15	0.98
			o__Alteromonadales	f__Colwelliaceae	g__u_Colwelliaceae	16829	1	0.01	25925	0	0.00	0.39	0.98
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	16829	1	0.01	25925	0	0.00	0.39	0.98
			o__Alteromonadales	f__OM60	g__u_OM60	16829	2	0.01	25925	1	0.00	0.57	0.98
Exp2A	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	19987	5	0.03	19998	15	0.08	0.04	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	19987	4	0.02	19998	2	0.01	0.45	1.00
			o__Alteromonadales	f__211ds20	g__u_211ds20	19987	0	0.00	19998	2	0.01	0.50	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__Marinobacter	19987	0	0.00	19998	1	0.01	1.00	1.00
Exp2A	Illumina	EL4V6V8	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	16331	4	0.02	15597	10	0.06	0.11	0.98
			o__Alteromonadales	f__OM60	g__u_OM60	16331	2	0.01	15597	6	0.04	0.17	0.98
			o__Alteromonadales	f__211ds20	g__u_211ds20	16331	2	0.01	15597	0	0.00	0.50	0.98
			o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	16331	3	0.02	15597	1	0.01	0.63	0.98
Vancouver	Illumina	EL4V6V8	o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	14956	3	0.02	19140	2	0.01	0.66	1.00
Billings B	Illumina	EL4V4V5	o__Alteromonadales	f__211ds20	g__u_211ds20	20073	2	0.01	20189	1	0.00	0.62	1.00
			o__Alteromonadales	f__125ds10	g__u_125ds10	20073	7	0.03	20189	6	0.03	0.79	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	20073	0	0.00	20189	1	0.00	1.00	1.00
Exp2B	Illumina	EL4V4V5	o__Alteromonadales	f__211ds20	g__u_211ds20	20136	10	0.05	20379	3	0.01	0.06	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20136	1	0.00	20379	4	0.02	0.38	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	20136	0	0.00	20379	1	0.00	1.00	1.00
			o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	20136	0	0.00	20379	1	0.00	1.00	1.00
			o__Anaerolineales	f__Anaerolinaceae	g__Anaerolinea	20136	5	0.02	20379	1	0.00	0.12	1.00

Table S6-9													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp3A	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	18541	10	0.05	21097	83	0.39	1.02E-13	6.32E-11
			o__Alteromonadales	f__211ds20	g__u_211ds20	18541	9	0.05	21097	5	0.02	0.28	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	18541	1	0.01	21097	0	0.00	0.47	1.00
Exp3B	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	19915	55	0.28	20181	6	0.03	2.26E-11	1.08E-08
			o__Alteromonadales	f__125ds10	g__u_125ds10	19915	2	0.01	20181	0	0.00	0.25	1.00
			o__Alteromonadales	f__211ds20	g__u_211ds20	19915	2	0.01	20181	5	0.02	0.45	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	19915	1	0.01	20181	0	0.00	0.50	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	19915	5	0.03	20181	4	0.02	0.75	1.00
RACentre	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20039	0	0.00	7911	1	0.01	0.28	1.00
			o__Alteromonadales	f__211ds20	g__u_211ds20	20039	1	0.00	7911	0	0.00	1.00	1.00
			o__Alteromonadales	f__125ds10	g__u_125ds10	20039	1	0.00	7911	0	0.00	1.00	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	20039	2	0.01	7911	0	0.00	1.00	1.00
Ellsworth	Illumina	EL4V4V5	o__Alteromonadales	f__211ds20	g__u_211ds20	19650	2	0.01	21073	5	0.02	0.46	1.00
			o__Alteromonadales	f__125ds10	g__u_125ds10	19650	0	0.00	21073	2	0.01	0.50	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	19650	3	0.02	21073	6	0.03	0.51	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	19650	0	0.00	21073	1	0.00	1.00	1.00
Diamond	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20279	26	0.13	20582	801	3.89	3.23E-201	1.00E-197
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	20279	40	0.20	20582	387	1.88	4.50E-72	2.78E-69
			o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	20279	1	0.00	20582	13	0.06	0.00	0.11
			o__Alteromonadales	f__[Chromatiaceae]	g__u_[Chromatiaceae]	20279	0	0.00	20582	4	0.02	0.13	1.00
			o__Alteromonadales	f__HTCC2188	g__HTCC	20279	1	0.00	20582	6	0.03	0.13	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__Marinobacter	20279	1	0.00	20582	5	0.02	0.22	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	20279	1	0.00	20582	0	0.00	0.50	1.00
			o__Alteromonadales	f__211ds20	g__u_211ds20	20279	1	0.00	20582	1	0.00	1.00	1.00
Glebe	Illumina	EL4V4V5	o__Alteromonadales	f__211ds20	g__u_211ds20	20086	12	0.06	20206	34	0.17	0.00	0.14
			o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20086	5	0.02	20206	9	0.04	0.42	1.00
			o__Alteromonadales	f__OM60	g__u_OM60	20086	2	0.01	20206	5	0.02	0.45	1.00
			o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Alteromonadales	f__125ds10	g__u_125ds10	20086	0	0.00	20206	2	0.01	0.50	1.00
			o__Alteromonadales	f__Idiomarinaceae	g__u_Idiomarinaceae	20086	0	0.00	20206	1	0.00	1.00	1.00

Table S6-10													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
80F	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20038	193	0.96	12557	562	4.48	0.00	0.00
			o__Alteromonadales	f__Shewanellaceae	g__Shewanella	20038	1	0.00	12557	64	0.51	1.12E-25	3.47E-23
			o__Alteromonadales	f__[Chromatiaceae]	g__u_[Chromatiaceae]	20038	6	0.03	12557	38	0.30	7.41E-11	1.27E-08
			o__Alteromonadales	f__Alteromonadaceae	g__BD2-13	20038	5	0.02	12557	23	0.18	4.10E-06	3.67E-04
			o__Alteromonadales	f__u_Alteromonadales	g__u_Alteromonadales	20038	0	0.00	12557	7	0.06	1.26E-03	0.08
			o__Alteromonadales	f__211ds20	g__u_211ds20	20038	32	0.16	12557	6	0.05	3.95E-03	0.22
			o__Alteromonadales	f__Alteromonadaceae	g__u_Alteromonadaceae	20038	88	0.44	12557	82	0.65	0.01	0.63
			o__Alteromonadales	f__OM60	g__u_OM60	20038	7	0.03	12557	1	0.01	0.16	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__Microbulbifer	20038	3	0.01	12557	0	0.00	0.29	1.00
			o__Alteromonadales	f__HTCC2188	g__HTCC	20038	11	0.05	12557	4	0.03	0.43	1.00
			o__Alteromonadales	f__Alteromonadaceae	g__HTCC2207	20038	1	0.00	12557	0	0.00	1.00	1.00
80FTop	Illumina	EL4V4V5	o__Alteromonadales	f__Alteromonadaceae	g__Cellvibrio	20117	28	0.14	19869	0	0.00	7.49E-09	9.85E-07
			o__Alteromonadales	f__Shewanellaceae	g__Shewanella	20117	1	0.00	19869	0	0.00	1.00	1.00
			o__Alteromonadales	f__211ds20	g__u_211ds20	20117	1	0.00	19869	0	0.00	1.00	1.00
			o__Alteromonadales	f__HTCC2188	g__HTCC	20117	1	0.00	19869	0	0.00	1.00	1.00

Table S6-11													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1925	2	0.10	2367	5	0.21	0.47	0.98
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	1925	0	0.00	2367	1	0.04	1.00	0.98
Exp1B2	Pyrosequencing	EL2V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	2343	1	0.04	2052	6	0.29	0.06	0.06
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	2343	0	0.00	2052	1	0.05	0.47	0.47
Exp1B3	Pyrosequencing	EL2V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	2267	0	0.00	1873	5	0.27	0.02	0.02
Exp1B	Pyrosequencing	EL3V1V3	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	4114	1	0.02	4510	3	0.07	0.63	0.76
Exp1B	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	19930	11	0.06	19983	46	0.23	3.27E-06	0.00
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	19930	1	0.01	19983	5	0.03	0.22	1.00
Exp1B	Illumina	EL4V6V8	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	16829	6	0.04	25925	76	0.29	7.04E-11	7.99E-09
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	16829	0	0.00	25925	15	0.06	0.00	0.02
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	16829	0	0.00	25925	3	0.01	0.28	0.98
Exp2A	Pyrosequencing	EL1V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1947	0	0.00	2174	6	0.28	0.03	0.53
Exp2A1	Pyrosequencing	EL2V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	2011	0	0.00	2413	2	0.08	0.50	0.50
Exp2A	Pyrosequencing	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	19987	7	0.04	19998	13	0.07	0.26	1.00
			o__Desulfuromonadales	f__Pelobacteraceae	g__u_Pelobacteraceae	19987	0	0.00	19998	1	0.01	1.00	1.00
Exp2A	Illumina	EL4V6V8	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	16331	6	0.04	15597	16	0.10	0.03	0.51
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	16331	0	0.00	15597	1	0.01	0.49	0.98
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	16331	2	0.01	15597	3	0.02	0.68	0.98
Vancouver	Pyrosequencing	EL1V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1899	1	0.05	3566	4	0.11	0.66	1.00
Vancouver	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	20655	12	0.06	19871	11	0.06	1.00	1.00
Vancouver	Illumina	EL4V6V8	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	14956	6	0.04	19140	13	0.07	0.36	1.00
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	14956	2	0.01	19140	4	0.02	0.70	1.00

Table S6-12

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Billings B	Pyrosequencing	EL1V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1323	5	0.38	1821	25	1.37	0.00	0.32
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	1323	0	0.00	1821	1	0.05	1.00	1.00
Billings B	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	20073	7	0.03	20189	252	1.25	1.90E-65	2.93E-62
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	20073	0	0.00	20189	21	0.10	9.50E-07	9.48E-05
Exp2B	Pyrosequencing	EL1V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1240	16	1.29	2048	48	2.34	0.04	0.53
			o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	1240	2	0.16	2048	7	0.34	0.50	0.97
Exp2B	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__u_Geobacteraceae	20136	0	0.00	20379	3	0.01	0.25	1.00
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	20136	31	0.15	20379	25	0.12	0.42	1.00
			o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	20136	84	0.42	20379	75	0.37	0.47	1.00
			o__Desulfuromonadales	f__Pelobacteraceae	g__u_Pelobacteraceae	20136	0	0.00	20379	1	0.00	1.00	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	1836	2	0.11	1727	7	0.41	0.10	0.10
Exp3A	Illumina	EL4V4V5	o__Desulfobacterales	f__Desulfobacteraceae	g__Desulfococcus	18541	2	0.01	21097	0	0.00	0.22	1.00
			o__Desulfobacterales	f__Desulfobulbaceae	g__Desulfobulbus	18541	0	0.00	21097	3	0.01	0.25	1.00
Exp3B	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	19915	42	0.21	20181	383	1.90	1.17E-69	3.63E-66
			o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	19915	1	0.01	20181	6	0.03	0.13	1.00
RACentre	Illumina	EL4V4V5	o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	20039	2	0.01	7911	2	0.03	0.32	1.00
Ellsworth	Illumina	EL4V4V5	o__Desulfuromonadales	f__u_Desulfuromonadale	g__u_Desulfuromonadales	19650	4	0.02	21073	11	0.05	0.12	1.00
			o__Desulfuromonadales	f__Geobacteraceae	g__Geobacter	19650	6	0.03	21073	10	0.05	0.46	1.00

Table S6-13

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	1925	4	0.21	2367	0	0.00	0.04	0.59
Exp1B1	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	2654	2	0.08	1945	4	0.21	0.25	0.25
Exp1B2	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	2343	1	0.04	2052	3	0.15	0.35	0.35
Exp1B3	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	2267	1	0.04	1873	1	0.05	1.00	1.00
Exp1B	Pyrosequencing	EL3V1V3	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	4114	1	0.02	4510	0	0.00	0.48	0.67
Exp1B	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	19930	29	0.15	19983	21	0.11	0.26	1.00
Exp1B	Illumina	EL4V6V8	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	16829	53	0.31	25925	83	0.32	1.00	0.98
Exp2A	Pyrosequencing	EL1V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	1947	2	0.10	2174	2	0.09	1.00	0.97
Exp2A1	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	2011	2	0.10	2413	1	0.04	0.59	0.59
Exp2A2	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	2011	2	0.10	2517	4	0.16	0.70	0.70
Exp2A	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	19987	20	0.10	19998	16	0.08	0.51	1.00
Exp2A	Illumina	EL4V6V8	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	16331	51	0.31	15597	33	0.21	0.08	0.96
Billings B	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20073	10	0.05	20189	4	0.02	0.12	1.00
Exp2B	Pyrosequencing	EL1V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	1240	7	0.56	2048	1	0.05	0.01	0.17
Exp2B	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20136	35	0.17	20379	37	0.18	0.91	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	1836	4	0.22	1727	1	0.06	0.38	0.38
Exp3A	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	18541	27	0.15	21097	12	0.06	0.01	0.37
Exp3B	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	19915	12	0.06	20181	18	0.09	0.36	1.00

Table S6-13													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
RACentre	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20039	14	0.07	7911	0	0.00	0.01	1.00
Ellsworth	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	19650	217	1.10	21073	5	0.02	2.26E-61	1.55E-58
Diamond	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20279	353	1.74	20582	17	0.08	1.02E-84	7.90E-82
Glebe	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20086	78	0.39	20206	13	0.06	8.94E-13	3.45E-10
80F	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20038	31	0.15	12557	1	0.01	4.56E-06	0.000397
80FTop	Illumina	EL4V4V5	o__JG30-KF-CM45	f__u_JG30-KF-CM45	g__u_JG30-KF-CM45	20117	3	0.01	19869	0	0.00	0.25	1.00

Table S6-14													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1925	4	0.21	2367	14	0.59	0.06	0.77
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	1925	1	0.05	2367	0	0.00	0.45	0.98
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	1925	1	0.05	2367	0	0.00	0.45	0.98
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	1925	0	0.00	2367	1	0.04	1.00	0.98
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	1925	0	0.00	2367	1	0.04	1.00	0.98
Exp1B1	Pyrosequencing	EL2V4V6	o__Bacillales	f__Planococcaceae	g__Sporosarcina	2654	0	0.00	1945	7	0.36	0.00	0.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	2654	0	0.00	1945	1	0.05	0.42	0.42
			o__Bacillales	f__Bacillaceae	g__Bacillus	2654	1	0.04	1945	2	0.10	0.58	0.58
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	2654	3	0.11	1945	1	0.05	0.64	0.64
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	2654	1	0.04	1945	1	0.05	1.00	1.00
Exp1B2	Pyrosequencing	EL2V4V6	o__Bacillales	f__Planococcaceae	g__Sporosarcina	2343	1	0.04	2052	4	0.19	0.19	0.19
			o__Bacillales	f__Planococcaceae	g__Paenisporosarcina	2343	3	0.13	2052	0	0.00	0.25	0.25
			o__Bacillales	f__Bacillaceae	g__Bacillus	2343	5	0.21	2052	2	0.10	0.46	0.46
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	2343	0	0.00	2052	1	0.05	0.47	0.47
			o__Bacillales	f__Planococcaceae	g__Planococcus	2343	0	0.00	2052	1	0.05	0.47	0.47
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	2343	1	0.04	2052	2	0.10	0.60	0.60
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	2343	1	0.04	2052	0	0.00	1.00	1.00
Exp1B3	Pyrosequencing	EL2V4V6	o__Bacillales	f__Planococcaceae	g__Sporosarcina	2267	0	0.00	1873	8	0.43	0.00	0.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	2267	0	0.00	1873	3	0.16	0.09	0.09
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	2267	0	0.00	1873	2	0.11	0.20	0.20
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	2267	1	0.04	1873	0	0.00	1.00	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	2267	1	0.04	1873	0	0.00	1.00	1.00
Exp1B	Pyrosequencing	EL3V1V3	o__Bacillales	f__Bacillaceae	g__Bacillus	4114	7	0.17	4510	15	0.33	0.20	0.60
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	4114	2	0.05	4510	0	0.00	0.23	0.60
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	4114	1	0.02	4510	0	0.00	0.48	0.67
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	4114	1	0.02	4510	0	0.00	0.48	0.67
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	4114	1	0.02	4510	1	0.02	1.00	0.82

5Table S6-15													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Illumina	EL4V4V5	o__Bacillales	f__Bacillaceae	g__Bacillus	19930	20	0.10	19983	69	0.35	1.73E-07	4.65E-05
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	19930	3	0.02	19983	13	0.07	0.02	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	19930	2	0.01	19983	0	0.00	0.25	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	19930	3	0.02	19983	1	0.01	0.37	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	19930	1	0.01	19983	0	0.00	0.50	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Shimazuella	19930	1	0.01	19983	0	0.00	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	19930	1	0.01	19983	0	0.00	0.50	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	19930	1	0.01	19983	0	0.00	0.50	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	19930	1	0.01	19983	3	0.02	0.62	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	19930	1	0.01	19983	2	0.01	1.00	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	19930	7	0.04	19983	7	0.04	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	19930	0	0.00	19983	1	0.01	1.00	1.00
Exp1B	Illumina	EL4V6V8	o__Bacillales	f__Bacillaceae	g__Bacillus	16829	9	0.05	25925	77	0.30	3.04E-09	2.59E-07
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	16829	5	0.03	25925	0	0.00	0.01	0.20
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	16829	1	0.01	25925	9	0.03	0.10	0.98
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	16829	2	0.01	25925	0	0.00	0.15	0.98
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	16829	0	0.00	25925	4	0.02	0.16	0.98
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	16829	0	0.00	25925	2	0.01	0.52	0.98
			o__Bacillales	f__Bacillaceae	g__Anoxybacillus	16829	0	0.00	25925	1	0.00	1.00	0.98
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	16829	0	0.00	25925	1	0.00	1.00	0.98
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	16829	0	0.00	25925	1	0.00	1.00	0.98
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	16829	0	0.00	25925	1	0.00	1.00	0.98
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	16829	1	0.01	25925	1	0.00	1.00	0.98

5Table S6-16													
Sample	**Sequencing method**	**Lab/Primers**	**Order**	**Family**	**Genus**	**Mobio total count**	**Mobio Genus count**	**% Genus Mobio**	**Soft total count**	**Soft Genus count**	**% Genus soft**	**p-values**	**corrected p-values**
Exp2A	Pyrosequencing	EL1V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1947	14	0.72	2174	2	0.09	0.00	0.09
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	1947	7	0.36	2174	2	0.09	0.09	0.82
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	1947	2	0.10	2174	0	0.00	0.22	0.97
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	1947	2	0.10	2174	5	0.23	0.46	0.97
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	1947	1	0.05	2174	0	0.00	0.47	0.97
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	1947	1	0.05	2174	0	0.00	0.47	0.97
			o__Bacillales	f__Planococcaceae	g__Ureibacillus	1947	1	0.05	2174	0	0.00	0.47	0.97
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	1947	1	0.05	2174	1	0.05	1.00	0.97
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	1947	1	0.05	2174	1	0.05	1.00	0.97
			o__Bacillales	f__Planococcaceae	g__Planomicrobium	1947	0	0.00	2174	1	0.05	1.00	0.97
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	1947	0	0.00	2174	1	0.05	1.00	0.97
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	1947	7	0.36	2174	7	0.32	1.00	0.97
Exp2A1	Pyrosequencing	EL2V4V6	o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	2011	4	0.20	2413	0	0.00	0.04	0.04
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	2011	4	0.20	2413	0	0.00	0.04	0.04
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	2011	2	0.10	2413	0	0.00	0.21	0.21
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	2011	1	0.05	2413	5	0.21	0.23	0.23
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	2011	1	0.05	2413	0	0.00	0.45	0.45
			o__Bacillales	f__Bacillaceae	g__Bacillus	2011	2	0.10	2413	2	0.08	1.00	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	2011	0	0.00	2413	1	0.04	1.00	1.00
Exp2A2	Pyrosequencing	EL2V4V6	o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	2011	4	0.20	2517	0	0.00	0.04	0.04
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	2011	2	0.10	2517	9	0.36	0.13	0.13
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	2011	4	0.20	2517	1	0.04	0.18	0.18
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	2011	0	0.00	2517	2	0.08	0.51	0.51
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	2011	0	0.00	2517	2	0.08	0.51	0.51
			o__Bacillales	f__Bacillaceae	g__Bacillus	2011	2	0.10	2517	1	0.04	0.59	0.59
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	2011	1	0.05	2517	3	0.12	0.63	0.63
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	2011	1	0.05	2517	1	0.04	1.00	1.00

5Table S6-17													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp2A	Illumina	EL4V4V5	o__Bacillales	f__Bacillaceae	g__Bacillus	19987	43	0.22	19998	20	0.10	0.00	0.40
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	19987	10	0.05	19998	2	0.01	0.02	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	19987	13	0.07	19998	4	0.02	0.03	1.00
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	19987	2	0.01	19998	0	0.00	0.25	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	19987	7	0.04	19998	13	0.07	0.26	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	19987	3	0.02	19998	1	0.01	0.37	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	19987	3	0.02	19998	1	0.01	0.37	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	19987	3	0.02	19998	1	0.01	0.37	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	19987	3	0.02	19998	1	0.01	0.37	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	19987	1	0.01	19998	0	0.00	0.50	1.00
			o__Bacillales	f__Staphylococcaceae	g__u_Staphylococcaceae	19987	0	0.00	19998	2	0.01	0.50	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	19987	0	0.00	19998	2	0.01	0.50	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	19987	2	0.01	19998	4	0.02	0.69	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	19987	54	0.27	19998	52	0.26	0.85	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	19987	2	0.01	19998	3	0.02	1.00	1.00
Exp2A	Illumina	EL4V6V8	o__Bacillales	f__Bacillaceae	g__Virgibacillus	16331	0	0.00	15597	7	0.04	0.01	0.16
			o__Bacillales	f__Bacillaceae	g__Bacillus	16331	44	0.27	15597	21	0.13	0.01	0.19
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	16331	4	0.02	15597	0	0.00	0.13	0.98
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	16331	4	0.02	15597	0	0.00	0.13	0.98
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	16331	10	0.06	15597	5	0.03	0.30	0.98
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	16331	0	0.00	15597	1	0.01	0.49	0.98
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	16331	3	0.02	15597	1	0.01	0.63	0.98
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	16331	28	0.17	15597	30	0.19	0.69	0.98
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	16331	4	0.02	15597	4	0.03	1.00	0.98
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	16331	4	0.02	15597	4	0.03	1.00	0.98
Vancouver	Pyrosequencing	EL1V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1899	0	0.00	3566	4	0.11	0.31	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	1899	1	0.05	3566	0	0.00	0.35	1.00
Vancouver	Illumina	EL4V4V5	o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	20655	0	0.00	19871	3	0.02	0.12	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20655	0	0.00	19871	1	0.01	0.49	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20655	0	0.00	19871	1	0.01	0.49	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20655	0	0.00	19871	1	0.01	0.49	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20655	0	0.00	19871	1	0.01	0.49	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20655	2	0.01	19871	0	0.00	0.50	1.00
			o__Bacillales	f__Sporolactobacillaceae	g__u_Sporolactobacillaceae	20655	1	0.00	19871	0	0.00	1.00	1.00

5Table S6-18													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Vancouver	Illumina	EL4V6V8	o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	14956	0	0.00	19140	3	0.02	0.26	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	14956	1	0.01	19140	0	0.00	0.44	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	14956	0	0.00	19140	2	0.01	0.51	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	14956	0	0.00	19140	1	0.01	1.00	1.00
Billings B	Pyrosequencing	EL1V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1323	2	0.15	1821	16	0.88	0.01	0.34
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	1323	0	0.00	1821	1	0.05	1.00	1.00
Billings B	Illumina	EL4V4V5	o__Bacillales	f__Bacillaceae	g__Bacillus	20073	7	0.03	20189	68	0.34	1.13E-13	2.26E-11
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20073	0	0.00	20189	6	0.03	0.03	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20073	8	0.04	20189	2	0.01	0.06	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20073	1	0.00	20189	0	0.00	0.50	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20073	0	0.00	20189	1	0.00	1.00	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20073	0	0.00	20189	1	0.00	1.00	1.00
Exp2B	Pyrosequencing	EL1V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1240	9	0.73	2048	190	9.28	1.05E-29	4.99E-27
			o__Bacillales	f__Planococcaceae	g__Planomicrobium	1240	0	0.00	2048	31	1.51	5.41E-07	6.44E-05
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	1240	2	0.16	2048	0	0.00	0.14	0.97
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	1240	2	0.16	2048	5	0.24	0.72	0.97
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	1240	0	0.00	2048	1	0.05	1.00	0.97
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	1240	0	0.00	2048	1	0.05	1.00	0.97
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	1240	0	0.00	2048	1	0.05	1.00	0.97
			o__Bacillales	f__Bacillaceae	g__Gracilibacillus	1240	0	0.00	2048	1	0.05	1.00	0.97
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	1240	0	0.00	2048	1	0.05	1.00	0.97
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	1240	1	0.08	2048	2	0.10	1.00	0.97

5Table S6-19

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp2B	Illumina	EL4V4V5	o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20136	51	0.25	20379	28	0.14	0.01	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20136	70	0.35	20379	46	0.23	0.03	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20136	15	0.07	20379	5	0.02	0.03	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20136	36	0.18	20379	24	0.12	0.12	1.00
			o__Bacillales	f__Bacillaceae	g__Anoxybacillus	20136	2	0.01	20379	0	0.00	0.25	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20136	0	0.00	20379	3	0.01	0.25	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20136	0	0.00	20379	3	0.01	0.25	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	20136	5	0.02	20379	2	0.01	0.29	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	20136	1	0.00	20379	0	0.00	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__Marinibacillus	20136	1	0.00	20379	0	0.00	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20136	0	0.00	20379	2	0.01	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	20136	0	0.00	20379	2	0.01	0.50	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20136	2	0.01	20379	1	0.00	0.62	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	20136	1	0.00	20379	1	0.00	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20136	1	0.00	20379	1	0.00	1.00	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	20136	1	0.00	20379	2	0.01	1.00	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__Bacillales	f__Bacillaceae	g__Bacillus	1836	0	0.00	1727	1	0.06	0.48	0.48
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	1836	1	0.05	1727	2	0.12	0.61	0.61
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	1836	1	0.05	1727	1	0.06	1.00	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	1836	1	0.05	1727	1	0.06	1.00	1.00
Exp3A	Illumina	EL4V4V5	o__Bacillales	f__Planococcaceae	g__Sporosarcina	18541	39	0.21	21097	5	0.02	7.58E-09	2.04E-06
			o__Bacillales	f__Bacillaceae	g__Bacillus	18541	98	0.53	21097	42	0.20	3.35E-08	8.28E-06
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	18541	20	0.11	21097	4	0.02	0.00	0.03
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	18541	14	0.08	21097	2	0.01	0.00	0.12
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	18541	3	0.02	21097	0	0.00	0.10	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	18541	3	0.02	21097	0	0.00	0.10	1.00
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	18541	2	0.01	21097	0	0.00	0.22	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	18541	9	0.05	21097	6	0.03	0.32	1.00
			o__Bacillales	f__Bacillaceae	g__Anaerobacillus	18541	1	0.01	21097	0	0.00	0.47	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	18541	1	0.01	21097	0	0.00	0.47	1.00
			o__Bacillales	f__Planococcaceae	g__Planomicrobium	18541	1	0.01	21097	0	0.00	0.47	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	18541	1	0.01	21097	0	0.00	0.47	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	18541	1	0.01	21097	1	0.00	1.00	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	18541	1	0.01	21097	2	0.01	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	18541	1	0.01	21097	2	0.01	1.00	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	18541	2	0.01	21097	2	0.01	1.00	1.00

5Table S6-20													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp3B	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Cohnella	19915	4	0.02	20181	0	0.00	0.06	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	19915	14	0.07	20181	24	0.12	0.14	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	19915	3	0.02	20181	8	0.04	0.23	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	19915	1	0.01	20181	0	0.00	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	19915	1	0.01	20181	0	0.00	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	19915	2	0.01	20181	1	0.00	0.62	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	19915	3	0.02	20181	2	0.01	0.69	1.00
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	19915	7	0.04	20181	9	0.04	0.80	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	19915	1	0.01	20181	2	0.01	1.00	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	19915	1	0.01	20181	1	0.00	1.00	1.00
RACentre	Illumina	EL4V4V5	o__Bacillales	f__Bacillaceae	g__Bacillus	20039	78	0.39	7911	182	2.30	0.00	0.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20039	4	0.02	7911	13	0.16	5.25E-05	0.01
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	20039	0	0.00	7911	4	0.05	0.01	0.50
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20039	3	0.01	7911	6	0.08	0.02	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20039	0	0.00	7911	2	0.03	0.08	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20039	2	0.01	7911	2	0.03	0.32	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20039	1	0.00	7911	0	0.00	1.00	1.00
Mattawa	Illumina	EL4V4V5	o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	19838	28	0.14	14050	1	0.01	4.71E-06	0.001533
			o__Bacillales	f__Alicyclobacillaceae	g__u_Alicyclobacillaceae	19838	19	0.10	14050	1	0.01	4.19E-04	0.11
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	19838	8	0.04	14050	1	0.01	0.09	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	19838	4	0.02	14050	1	0.01	0.41	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	19838	0	0.00	14050	1	0.01	0.41	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	19838	1	0.01	14050	0	0.00	1.00	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	19838	1	0.01	14050	0	0.00	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	19838	1	0.01	14050	0	0.00	1.00	1.00

5Table S6-21													
Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Ellsworth	Illumina	EL4V4V5	o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	19650	320	1.63	21073	1	0.00	2.31E-100	2.86E-97
			o__Bacillales	f__Bacillaceae	g__Bacillus	19650	317	1.61	21073	1	0.00	2.09E-99	2.16E-96
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	19650	65	0.33	21073	1	0.00	8.82E-20	1.52E-17
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	19650	50	0.25	21073	3	0.01	5.34E-13	6.74E-11
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	19650	13	0.07	21073	0	0.00	7.67E-05	4.31E-03
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	19650	12	0.06	21073	0	0.00	1.59E-04	0.01
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	19650	5	0.03	21073	0	0.00	0.03	0.83
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	19650	4	0.02	21073	0	0.00	0.05	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	19650	4	0.02	21073	0	0.00	0.05	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	19650	3	0.02	21073	0	0.00	0.11	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Shimazuella	19650	1	0.01	21073	0	0.00	0.48	1.00
			o__Bacillales	f__Planococcaceae	g__Rummeliibacillus	19650	1	0.01	21073	0	0.00	0.48	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	19650	1	0.01	21073	0	0.00	0.48	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	19650	1	0.01	21073	0	0.00	0.48	1.00
Diamond	Illumina	EL4V4V5	o__Bacillales	f__Bacillaceae	g__Bacillus	20279	653	3.22	20582	87	0.42	6.09E-112	9.41E-109
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20279	125	0.62	20582	11	0.05	3.29E-26	8.47E-24
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20279	51	0.25	20582	2	0.01	1.16E-13	1.83E-11
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	20279	29	0.14	20582	2	0.01	2.06E-07	2.16E-05
			o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20279	10	0.05	20582	43	0.21	5.64E-06	4.53E-04
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	20279	12	0.06	20582	0	0.00	2.23E-04	0.01
			o__Bacillales	f__Planococcaceae	g__Planomicrobium	20279	18	0.09	20582	4	0.02	2.43E-03	0.13
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20279	8	0.04	20582	0	0.00	3.68E-03	0.18
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20279	8	0.04	20582	0	0.00	3.68E-03	0.18
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20279	5	0.02	20582	0	0.00	0.03	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20279	0	0.00	20582	3	0.01	0.25	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20279	1	0.00	20582	0	0.00	0.50	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	20279	1	0.00	20582	0	0.00	0.50	1.00
			o__Bacillales	f__Staphylococcaceae	g__Salinicoccus	20279	1	0.00	20582	0	0.00	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	20279	1	0.00	20582	0	0.00	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20279	2	0.01	20582	4	0.02	0.69	1.00
			o__Bacillales	f__Bacillaceae	g__Marinococcus	20279	1	0.00	20582	2	0.01	1.00	1.00

5Table S6-22													
Sample	**Sequencing method**	**Lab/Primers**	**Order**	**Family**	**Genus**	**Mobio total count**	**Mobio Genus count**	**% Genus Mobio**	**Soft total count**	**Soft Genus count**	**% Genus soft**	**p-values**	**corrected p-values**
Glebe	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20086	34	0.17	20206	2	0.01	9.30E-09	2.39E-06
			o__Bacillales	f__Bacillaceae	g__Bacillus	20086	24	0.12	20206	2	0.01	5.31E-06	0.000763
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20086	6	0.03	20206	0	0.00	0.02	0.93
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20086	6	0.03	20206	0	0.00	0.02	0.93
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20086	2	0.01	20206	0	0.00	0.25	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20086	0	0.00	20206	2	0.01	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20086	2	0.01	20206	1	0.00	0.62	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20086	1	0.00	20206	3	0.01	0.62	1.00
			o__Bacillales	f__Bacillaceae	g__Virgibacillus	20086	0	0.00	20206	1	0.00	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20086	0	0.00	20206	1	0.00	1.00	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	20086	1	0.00	20206	1	0.00	1.00	1.00
80F	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20038	17	0.08	12557	111	0.88	1.41E-29	4.59E-27
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20038	0	0.00	12557	35	0.28	3.08E-15	7.03E-13
			o__Bacillales	f__Listeriaceae	g__u_Listeriaceae	20038	0	0.00	12557	2	0.02	0.15	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20038	3	0.01	12557	0	0.00	0.29	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20038	2	0.01	12557	0	0.00	0.53	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20038	1	0.00	12557	2	0.02	0.56	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20038	2	0.01	12557	2	0.02	0.64	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20038	4	0.02	12557	3	0.02	1.00	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20038	1	0.00	12557	0	0.00	1.00	1.00
80FTop	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20117	27	0.13	19869	2	0.01	1.63E-06	1.86E-04
			o__Bacillales	f__Listeriaceae	g__u_Listeriaceae	20117	3	0.01	19869	0	0.00	0.25	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20117	11	0.05	19869	7	0.04	0.48	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20117	0	0.00	19869	1	0.01	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	20117	0	0.00	19869	1	0.01	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	20117	2	0.01	19869	0	0.00	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20117	20	0.10	19869	24	0.12	0.62	1.00

5Table S6-23													
Sample	**Sequencing method**	**Lab/Primers**	**Order**	**Family**	**Genus**	**Mobio total count**	**Mobio Genus count**	**% Genus Mobio**	**Soft total count**	**Soft Genus count**	**% Genus soft**	**p-values**	**corrected p-values**
Glebe	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20086	34	0.17	20206	2	0.01	9.30E-09	2.39E-06
			o__Bacillales	f__Bacillaceae	g__Bacillus	20086	24	0.12	20206	2	0.01	5.31E-06	0.000763
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20086	6	0.03	20206	0	0.00	0.02	0.93
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20086	6	0.03	20206	0	0.00	0.02	0.93
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20086	2	0.01	20206	0	0.00	0.25	1.00
			o__Bacillales	f__Thermoactinomycetac	g__Planifilum	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Brevibacillus	20086	1	0.00	20206	0	0.00	0.50	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20086	0	0.00	20206	2	0.01	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__u_Bacillaceae	20086	2	0.01	20206	1	0.00	0.62	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20086	1	0.00	20206	3	0.01	0.62	1.00
			o__Bacillales	f__Bacillaceae	g__Virgibacillus	20086	0	0.00	20206	1	0.00	1.00	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20086	0	0.00	20206	1	0.00	1.00	1.00
			o__Bacillales	f__Thermoactinomycetac	g__u_Thermoactinomyceta	20086	1	0.00	20206	1	0.00	1.00	1.00
80F	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20038	17	0.08	12557	111	0.88	1.41E-29	4.59E-27
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20038	0	0.00	12557	35	0.28	3.08E-15	7.03E-13
			o__Bacillales	f__Listeriaceae	g__u_Listeriaceae	20038	0	0.00	12557	2	0.02	0.15	1.00
			o__Bacillales	f__Staphylococcaceae	g__Staphylococcus	20038	3	0.01	12557	0	0.00	0.29	1.00
			o__Bacillales	f__u_Bacillales	g__u_Bacillales	20038	2	0.01	12557	0	0.00	0.53	1.00
			o__Bacillales	f__Paenibacillaceae	g__u_Paenibacillaceae	20038	1	0.00	12557	2	0.02	0.56	1.00
			o__Bacillales	f__Paenibacillaceae	g__Cohnella	20038	2	0.01	12557	2	0.02	0.64	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20038	4	0.02	12557	3	0.02	1.00	1.00
			o__Bacillales	f__Alicyclobacillaceae	g__Alicyclobacillus	20038	1	0.00	12557	0	0.00	1.00	1.00
80FTop	Illumina	EL4V4V5	o__Bacillales	f__Paenibacillaceae	g__Paenibacillus	20117	27	0.13	19869	2	0.01	1.63E-06	1.86E-04
			o__Bacillales	f__Listeriaceae	g__u_Listeriaceae	20117	3	0.01	19869	0	0.00	0.25	1.00
			o__Bacillales	f__Planococcaceae	g__Lysinibacillus	20117	11	0.05	19869	7	0.04	0.48	1.00
			o__Bacillales	f__Planococcaceae	g__Sporosarcina	20117	0	0.00	19869	1	0.01	0.50	1.00
			o__Bacillales	f__Paenibacillaceae	g__Ammoniphilus	20117	0	0.00	19869	1	0.01	0.50	1.00
			o__Bacillales	f__Planococcaceae	g__u_Planococcaceae	20117	2	0.01	19869	0	0.00	0.50	1.00
			o__Bacillales	f__Bacillaceae	g__Bacillus	20117	20	0.10	19869	24	0.12	0.62	1.00

Supplemental Data Table 7-Relative distribution of the genus a) *Stenotrophomonas* and b) order Streptophyta in amplicons generated in this study. Results are not shown where the *Stenotrophomonas* and Streptophyta were not found.

Note: Soil_Exp1B-2 refers to a comparison between the soft method and a doubled Mobio bead beating time. Soil_Exp1B3 refers to a comparison between soft extraction and Retsch GmbH beadbeat

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B	Pyrosequencing	EL1V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	1925	0	0.00	2367	6	0.25	0.04	0.58
Exp1B1	Pyrosequencing	EL2V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	2654	0	0.00	1945	2	0.10	0.18	0.18
Exp1B2	Pyrosequencing	EL2V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	2343	0	0.00	2052	8	0.39	2.24E-03	2.24E-03
Exp1B3	Pyrosequencing	EL2V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	2267	0	0.00	1873	4	0.21	0.04	0.04
Exp1B	Pyrosequencing	EL3V1V3	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	4114	0	0.00	4510	10	0.22	0.00	0.03
Exp1B	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	19930	0	0.00	19983	7	0.04	0.02	1.00
Exp2A	Pyrosequencing	EL1V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	1947	0	0.00	2174	2	0.09	0.50	0.97
Billings B	Pyrosequencing	EL1V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	1323	0	0.00	1821	2	0.11	0.51	1.00
Exp2B	Pyrosequencing	EL1V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	1240	0	0.00	2048	123	6.01	2.46E-26	5.85E-24
Exp2B	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	20136	0	0.00	20379	1	0.00	1.00	1.00
Exp4A	Pyrosequencing	EL2V4V6	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	1836	5	0.27	1727	141	8.16	5.47E-39	5.47E-39
Exp3A	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	18541	2	0.01	21097	19	0.09006	0.00	0.05
Exp3B	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	19915	1	0.01	20181	0	0	0.50	1.00
RACentre	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	20039	0	0.00	7911	13	0.16433	7.43E-08	2.09E-05
80F	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	20038	6	0.03	12557	5	0.03982	0.76	1.00
80FTop	Illumina	EL4V4V5	o__Xanthomonadales	f__Xanthomonadaceae	g__Stenotrophomonas	20117	31	0.15	19869	4	0.02013	3.48E-06	3.91E-04

Supplemental Data Table 7 b) order Streptophyta in amplicons generated in this study

Sample	Sequencing method	Lab/Primers	Order	Family	Genus	Mobio total count	Mobio Genus count	% Genus Mobio	Soft total count	Soft Genus count	% Genus soft	p-values	corrected p-values
Exp1B1	Pyrosequencing	EL2V4V6	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	2654	0	0.00	1945	1	0.05	0.42	0.42
Exp1B	Pyrosequencing	EL3V1V3	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	4114	28	0.68	4510	0	0.00	9.52E-10	4.07E-08
Exp1B	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	19930	0	0.00	19983	10	0.05	0.00	0.21
Exp2A	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	19987	3	0.02	19998	6	0.03	0.51	1.00
Exp2A	Illumina	EL4V6V8	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	16331	1	0.01	15597	0	0.00	1.00	0.98
Vancouver	Pyrosequencing	EL1V4V6	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	1899	1	0.05	3566	0	0.00	0.35	1.00
Vancouver	Pyrosequencing	EL2V4V6	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	1856	3	0.16	2494	0	0.00	0.08	0.08
Billings B	Pyrosequencing	EL1V4V6	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	1323	0	0.00	1821	1	0.05	1.00	1.00
Billings B	Pyrosequencing	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20073	0	0.00	20189	1	0.00	1.00	1.00
Exp2B	Pyrosequencing	EL1V4V6	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	1240	0	0.00	2048	2	0.10	0.53	0.97
Exp2B	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20136	9	0.04	20379	2	0.01	0.04	1.00
Exp3A	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	18541	3	0.02	21097	13	0.06	0.03	1.00
Exp3B	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	19915	0	0.00	20181	1	0.00	1.00	1.00
RACentre	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20039	15	0.07	7911	1	0.01	0.05	1.00
Mattawa	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	19838	3	0.02	14050	1	0.01	0.65	1.00
Ellsworth	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	19650	3	0.02	21073	4	0.02	1.00	1.00
Diamond	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20279	2	0.01	20582	0	0.00	0.25	1.00
Glebe	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20086	31	0.15	20206	49	0.24	0.06	1.00
80F	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20038	17	0.08	12557	16	0.13	0.28	1.00
80FTop	Illumina	EL4V4V5	o__Streptophyta	f__u_Streptophyta	g__u_Streptophyta	20117	27	0.13	19869	6964	35.05	0.00	0.00

Appendix 5 Supplementary Tables Chapter 4

Table S1 Sequenced clones-1		**length**	**GC content**	**AT group**	**63 Amino Acid motif**	*** if motif**	**92 Amino Acid motif**	*** if motif**	**201 Amino Acid motif**
	#	**bp**	**%**	**classification of**	**and predicted substrate.**	**not found**	**and predicted substrate.**	**not found**	**and predicted substrate.**
				Jenke-Kodama		**on PFAM**		**on PFAM**	
				et al (2005)		**database**		**database**	
Cdnforest-KemptvilleP	1	1311	68.88	A8-3	RVDAVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnforest-KemptvilleP	2	1311	68.88	A8-3	RVDAVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnforest-KemptvilleP	3	1308	65.06	A3	QTLYTQ		GHSLGEmethylmalonyl		HAFH-malonyl
Colpasture-PanceP	1	1306	68.84	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YAAH
Colpasture-PanceP	2	1301	71.41	A8-2	RVDVVQ methylmalonyl				YASH-methylmalonyl
Colpasture-PanceP	3	1341	67.64	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YAAH
Colpasture-PanceP	4	1305	69.12	A8-1	PTEVIQ		GHSQGE methylmalonyl		YAAH
Colpasture-PanceP	5	1293	63.88	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Colpasture-PanceP	6	1303	71.14	A8-2	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Colpasture-PanceP	7	1341	67.86	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YAAH
Colpasture-PanceP	8	1342	67.73	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Colpasture-PanceP	9	1305	68.51	A5	DTEVAQ		GHSVGE malonyl		VPYH
Colpasture-PanceP	10	1320	69.39	A8-1	RIDVVQ-methylmalonyl		GHSQGE methylmalonyl		CAGH
Colforest-HaticoP	1	1338	68.91	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-HaticoP	2	1346	68.72	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-HaticoP	3	1302	69.35	A6	EVDVVQ methylmalonyl		GHSMGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	4	1308	63.99	A5	ETQVAQ	*	GHSVGE malonyl		VPYH
Colforest-HaticoP	5	1303	67.54	A8-2	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	6	1304	66.79	A5	DTNIAQ	*	GHSIGE malonyl		VPFH
Colforest-HaticoP	7	1315	63.88	A5	ETQVAQ	*	GHSVGE malonyl		VPYH
Colforest-HaticoP	8	1317	63.71	A8-2	EGHVVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	9	1305	69.12	A6	EVDVVQ methylmalonyl		GHSMGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	10	1318	70.11	A6	DVDVIQ	*	GHSMGE methylmalonyl		VASH-methylmalonyl
Colforest-HaticoP	11	1338	68.91	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-HaticoP	12	1289	71.06	A8-1	RIDVVQ-methylmalonyl		ATRRGE	*	YASH-methylmalonyl
Colforest-HaticoP	13	1306	69.22	A6	EVDVVQ methylmalonyl		GHSMGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	14	1343	68.8	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-HaticoP	15	1276	67.87	A8-1	RTEIIQ		GHSQGE methylmalonyl		YPSH
Colforest-HaticoP	16	1310	69.08	A6	EVDVVQ methylmalonyl		GHSMGE methylmalonyl		YASH-methylmalonyl
Colforest-HaticoP	17	1313	70.3	A6	DVDVIQ	*	GHSMGE methylmalonyl		VASH-methylmalonyl

Table S1 Sequenced clones-2 	# 	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Colforest-CIATP	1	1305	67.05	A8-3	RMDVVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Colforest-CIATP	2	1350	68.59	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YAAH
Colforest-CIATP	3	1326	69.76	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-CIATP	4	1326	69	A8-3	RVDVVQ methylmalonyl		GYSQGK		YASH-methylmalonyl
Colforest-CIATP	5	1326	69.61	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-CIATP	6	1326	69.53	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Colforest-CIATP	7	1326	69.53	A8-3	RVDVVQ methylmalonyl		GYSQGE	*	YASH-methylmalonyl
Czcultivated-PlanaP	1	1317	68.34	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	2	1322	69.06	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	3	1244	68.09	A8-1	RVEVVX	*	GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	4	1316	68.31	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	5	1388	68.31	A8-3	RVDVIQ		GHSQGE methylmalonyl		YAGH
Czcultivated-PlanaP	6	1338	67.79	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	7	1305	68.66	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	8	1319	68.31	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	9	1316	68.54	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	10	1319	68.08	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czcultivated-PlanaP	11	1300	67	A5			GHSAGE		
Cbudejovicesoil-ParadiseP	1	1306	66.46	A6	DISVVQ	*	GHSMGE methylmalonyl		VASH-methylmalonyl
Cbudejovicesoil-ParadiseP	2	1310	68.02	A8-3	RIDVVQ-methylmalonyl		GHSQGE methylmalonyl		VAGH
Cbudejovicesoil-ParadiseP	3	1340	67.91	A8-3	RVDVIQ		GHSQGE methylmalonyl		YTSH
Cbudejovicesoil-ParadiseP	4	1305	67.89	A8-3	RIDVVQ-methylmalonyl		GHSQGE methylmalonyl		VAGH
Cbudejovicesoil-ParadiseP	5	1302	67.9	A8-3	RIDVVQ-methylmalonyl		GHSQGE methylmalonyl		VAGH
Cbudejovicesoil-ParadiseP	6	1306	63.55	A5	ETAVAQ	*	GHSVGE malonyl		YAFH-malonyl
Cbudejovicesoil-ParadiseP	7	1323	64.1	A6	ELDVVQ		GHSMGK		YASH-methylmalonyl
Cbudejovicesoil-ParadiseP	8	1316	61.17	A5	ETHVAQ	*	GHSVGE malonyl		HAFH-malonyl

Table S1 Sequenced clones-3	#	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Czpasture-PalavaP	1	1311	66.74	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czpasture-PalavaP	2	1341	68.61	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czpasture-PalavaP	3	1272	66.27	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Czpasture-PalavaP	4	1315	67.53	A8-3	RVDVIQ		GHSQGE methylmalonyl		YTSH
Czpasture-PalavaP	5	1303	68.23	A8-1	RLEVVQ		GHSQGE methylmalonyl		YAAH
Czpasture-PalavaP	6	1339	69.08	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Czpasture-PalavaP	7	1272	66.59	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Czpasture-PalavaP	8	1324	67.52	A8-3	RVDVIQ		GHSQ*EI		YASH-methylmalonyl
Czpasture-PalavaP	9	1308	65.06	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	1	1321	68.05	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	2	1339	67.66	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	3	1321	67.75	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	4	1339	67.59	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	5	1342	67.96	A8-3	RVDVIQ		GHSQRE		YASH-methylmalonyl
Cdntundra-ResoluteP	6	1346	67.98	A8-3	RVDVIQ		GHSQGE methylmalonyl		YAGH
Cdntundra-ResoluteP	7	1339	67.89	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASR
Cdntundra-ResoluteP	8	1340	67.84	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	9	1321	67.83	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	10	1321	68.05	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	11	1342	67.96	A8-3	RVDVIQ		GHSQRE		YASH-methylmalonyl
Cdntundra-ResoluteP	12	1321	67.68	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	13	1321	67.98	A8-3	RVSVIQ		GHSQGE methylmalonyl		YAPH
Cdntundra-ResoluteP	14	1306	70.14	A8-1	RVDVVQ methylmalonyl		RPLAGE	*	YASH-methylmalonyl
Cdntundra-ResoluteP	15	1339	67.66	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	16	1321	67.52	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	17	1339	67.74	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdntundra-ResoluteP	18	1339	67.89	A8-3	RVSVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl

Table S1 Sequenced clones-4	#	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Cdnbeach-PEIP	1	1317	69.32	A8-1	RTEVIQ		GHSQDE		YPSH
Cdnbeach-PEIP	2	1354	70.75	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnbeach-PEIP	3	1354	69.79	A8-1	QPH*GH		GHSQGE methylmalonyl		YPSH
Cdnbeach-PEIP	4	1319	71.87	A8-1	RVEVAQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnbeach-PEIP	5	1309	70.97	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnbeach-PEIP	6	1354	69.79	A8-1	RTEVIQ		GHSQGE methylmalonyl		YPSH
Cdnbeach-PEIP	7	1309	70.97	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Cdnbeach-PEIP	8	1324	72.36	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	1	1261	67.8	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	2	1224	66.09	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	3	1306	67.84	A8-3	RVDAVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	4	1306	68.99	A6	EVDVVQ methylmalonyl		GHSMGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	5	1315	69.05	A8-3	RVDAVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	6	1315	68.37	Outlier	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		CASH
Hungforest-CitadelP	7	1222	66.86	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	8	1342	67.66	A8-3	RVDVIQ		GHSQGG		YASH-methylmalonyl
Hungforest-CitadelP	9	1315	68.37	Outlier	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		CASH
Hungforest-CitadelP	10	1223	66.97	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Hungforest-CitadelP	11	1237	69.2	A8-1	RVDVVQ methylmalonyl		GHSQGG		YASH-methylmalonyl
Hungforest-CitadelP	12	1252	73.08	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Brusselscobblestones-BoucherP	1	1392	63.43	Outlier	YLEINQ	*	GHSFGE		GAGH
Brusselscobblestones-BoucherP	2	1048	64.69	A8-3	RIDIAE	*	GHGQGE		blank_____
Brusselscobblestones-BoucherP	3	1276	62.38	A1	KPSVLQ	*	GHSMGE methylmalonyl		IAAH
Brusselscobblestones-BoucherP	4	1057	65.09	A8-3	RIDIAE	*	GHGQGE		blank_____
Brusselscobblestones-BoucherP	5	1289	67.65	Mbt	QVEI_Q		_RHHGS		YPGH
Brusselscobblestones-BoucherP	6	1389	63.5	Outlier	YLEINQ	*	GHSFGE		GAGH
Brusselscobblestones-BoucherP	7	1308	67.28	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Brusselscobblestones-BoucherP	8	1290	63.64	A8-3	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		HASH
Brusselscobblestones-BoucherP	9	1149	66.49	A8-3	RVEVVQ methylmalonyl		GHSQGE methylmalonyl		__SH

Table S1 Sequenced clones-5	#	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Ottawacobblestones-DublinerP	1	1297	66.77	A8-1	RVDVVP		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	2	1297	67	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	3	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	4	1297	66.62	A8-1	RVDVV_		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	5	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	6	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	7	1296	66.9	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	8	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHTQGE		YASH-methylmalonyl
Ottawacobblestones-DublinerP	9	1297	66.69	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	10	1297	66.85	A8-1	RLDVVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	11	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	12	1297	66.69	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	13	1297	66.92	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawacobblestones-DublinerP	14	1298	67.18	A8-1	RVDVVQ methylmalonyl		GHTQGE		YASH-methylmalonyl
Ottawacobblestones-DublinerP	15	1297	66.77	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		FASH
Ottawacobblestones-DublinerP	16	1297	67.15	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawastreetdust-RideauP	1	1302	72.81	A5	QTEVAQ		GHSVGE malonyl		YASH-methylmalonyl
Ottawastreetdust-RideauP	2	1303	72.68	A5	QTEVAQ		GHSVGE malonyl		YAFH-malonyl
Ottawastreetdust-RideauP	3	1285	73.15	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawastreetdust-RideauP	4	1297	68	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawastreetdust-RideauP	5	1303	72.6	A5	QTEVAQ		GHSVGE malonyl		YAFH-malonyl
Ottawastreetdust-RideauP	6	1324	75.3	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Ottawastreetdust-RideauP	7	1324	74.09	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	1	1321	68.13	A8-3	RIEVVQ-methylmalonyl		GHSQGE methylmalonyl		VASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	2	1302	65.44	A8-1	RVDIVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	3	1324	68.58	A8-3	RIEVVQ-methylmalonyl		GHSQGE methylmalonyl		VASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	4	1294	64.61	A8-1	TVDIVQ		GHSQGE methylmalonyl		GAGH
Česke.Budejovicecobblestones-ČeskaP	5	1312	67.23	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	6	1198	66.94	A8-1	RVDIVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	7	1264	65.43	A8-1	RVDIVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	8	1247	65.28	A8-1	RVDIVQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Česke.Budejovicecobblestones-ČeskaP	9	1286	67.26	Mbt	QVEI-G		GHSLGEmethylmalonyl		YPGH
Česke.Budejovicecobblestones-ČeskaP	10	1287	67.21	Mbt	QVEI-G		GHSLGEmethylmalonyl		YPGH

Table S1 Sequenced clones-6	#	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Česke.Budéjovicestreetdust-KohiNoorP	1	1316	68.24	A8	RIEVVQ-methylmalonyl		GHSQGE methylmalonyl		VASH-methylmalonyl
Česke.Budéjovicestreetdust-KohiNoorP	2	1317	66.14	A5	ETEVAQ		GHSAGE		VPYH
Česke.Budéjovicestreetdust-KohiNoorP	3	1287	67.21	Mbt	QVEI-G		GHSLGEmethylmalonyl		YPHG
Česke.Budéjovicestreetdust-KohiNoorP	4	1286	67.26	Mbt	QVEI-G		GHSLGEmethylmalonyl		YPHG
Budapestpavement-TerezKorutP	1	1297	66.62	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	2	1297	67	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	3	1297	67.08	A8-1	RVDVVR		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	4	1297	67.08	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	5	1297	67	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	6	1288	66.77	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	7	1297	66.69	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	8	1297	67.15	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	9	1300	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	10	1297	66.85	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	11	1297	67.31	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	12	1297	66.92	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Budapestpavement-TerezKorutP	13	1297	66.92	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Pariscobblestones-RiveGaucheP	1	1171	69.26	A8	RIEVVQ-methylmalonyl		GHSQGE methylmalonyl		VASH-methylmalonyl
Pariscobblestones-RiveGaucheP	2	1317	68.26	A8	RIEVVQ-methylmalonyl		GHSQGE methylmalonyl		VASH-methylmalonyl
Pariscobblestones-RiveGaucheP	3	1288	67.31	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	4	1283	67.58	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	5	1289	67.34	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	6	1283	67.5	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	7	1283	67.58	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	8	1283	67.03	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	9	1295	67.26	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-RiveGaucheP	10	1291	67.39	Mbt	QVE-IQ		GHSLGEmethylmalonyl		YPGH

Table S1 Sequenced clones-7 	# 	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Pariscobblestones-CafeP	1	1302	63.06	A8-1	RLDIVQ		GHSQGE methylmalonyl		YAAH
Pariscobblestones-CafeP	2	1287	68.92	A5	QSDIGQ	*	GHSFGE		YAFH-malonyl
Pariscobblestones-CafeP	3	1287	70.01	A8-1	RIDVIQ		GHSQGE methylmalonyl		APVH
Pariscobblestones-CafeP	4	1337	67.91	A8-3	RVDVIQ		GHSQGE methylmalonyl		YTSH
Pariscobblestones-CafeP	5	1284	67.52	Mbt	QVEI_Q		GHSLGEmethylmalonyl		YPGH
Pariscobblestones-CafeP	6	1285	67	Mbt	HVEI_Q		GHSLGEmethylmalonyl		YPGH
FaisalabadStreetdust-ClockP	1	1321	69.11	A8-1	RVDVVQ methylmalonyl		GLSQGE	*	YASH-methylmalonyl
FaisalabadStreetdust-ClockP	2	1326	68.93	A8-1	RVDVVQ methylmalonyl		GLSQGE	*	YASH-methylmalonyl
FaisalabadStreetdust-ClockP	3	1329	73.74	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	4	1324	73.87	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	5	1291	74.36	A8-1	RADVVQ-methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	6	1270	70.08	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	7	1285	72.06	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	8	1297	72.4	A8-1	RVEVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	9	1324	74.02	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	10	1324	73.87	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	11	1312	74.54	A8-1	RADVVQ-methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
FaisalabadStreetdust-ClockP	12	1303	73.37	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASL

Table S1 Sequenced clones-8	#	length bp	GC content %	AT group classification of Jenke-Kodama et al (2005)	63 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	92 Amino Acid motif and predicted substrate.	* if motif not found on PFAM database	201 Amino Acid motif and predicted substrate.
Czech-VermicompostP	1	1318	69.12	A6	EQEIVQ	*	GHSAGE		VAGH
Czech-VermicompostP	2	1288	67	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	3	1290	67.13	A5	RSDVGG		GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	4	1316	69.15	A6	EQEIVQ	*	GHSAGE		VAGH
Czech-VermicompostP	5	1291	67.31	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	6	1350	64.96	A8-1	RLDIVQ		GHSQGE methylmalonyl		YAAH
Czech-VermicompostP	7	1351	64.99	A8-1	RLDIVQ		GHSQGE methylmalonyl		YAAH
Czech-VermicompostP	8	1294	67.23	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	9	1317	69.17	A6	EQEIVQ	*	GHSAGE		VAGH
Czech-VermicompostP	10	1326	69.38	A8-1	EQEIAQ	*	GHSAGE		VAGH
Czech-VermicompostP	11	1357	65.14	A6	RLDIVQ		GHSQGE methylmalonyl		YAAH
Czech-VermicompostP	12	1285	67.32	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	13	1315	69.13	A6	EQEIVQ	*	GHSAGE		VAGH
Czech-VermicompostP	14	1293	67.21	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Czech-VermicompostP	15	1294	67	A5	RSDVGQ	*	GHSIGE malonyl		YAFH-malonyl
Rozowormgut-Martiodrillus P	1	1315	67.91	A8-1	QVDKVQ	*	GHSQGE methylmalonyl		YASH-methylmalonyl
Rozowormgut-Martiodrillus P	2	1275	70.12	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Rozowormgut-Martiodrillus P	3	1302	65.59	A8-1	SLPVVQ		GHSQGE methylmalonyl		AAGH
Rozowormgut-Martiodrillus P	4	1314	67.88	A8-1	QVDKVQ	*	GHSQGE methylmalonyl		YASH-methylmalonyl
Rozowormgut-Martiodrillus P	5	1154	68.28	A8-1	HADRVQ	*	GHSQGE methylmalonyl		CASH
Rozowormgut-Martiodrillus P	6	1326	69.31	A8-1	QVDVVQ methylmalonyl		GHSQGE methylmalonyl		YASH-methylmalonyl
Rozowormgut-Martiodrillus P	7	1276	64.34	A8-1	RVDVVQ methylmalonyl		GHSQGE methylmalonyl		YAAH
Rozowormgut-Martiodrillus P	8	1206	68.49	A8-1	HADRVQ	*	GHSQGE methylmalonyl		CASH
Rozowormgut-Martiodrillus P	9	1309	70.74	A8-3	RVDVIQ		GHSQGE methylmalonyl		YASH-methylmalonyl
Rozowormgut-Martiodrillus P	10	1205	68.3	A8-1	HADRVQ	*	GHSQGE methylmalonyl		CASH

Table S2-1											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Cdnforest-Kemptville	1	A8-3	AmbE [Sorangium cellulosum]	62%	ABK32259.1	JerC [Sorangium cellulosum]	66%	ABK32289.1	laidlomycin Streptomyces sp. CS684	74%	AFL48528.1
			AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	65%	ABK32257.1	laidlomycin Streptomyces sp. CS684	73%	AFL48533.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbE [Sorangium cellulosum]	65%	ABK32259.1	Streptomyces mobaraensis NBRC 13819	73%	WP_004954442.1
Cdnforest-Kemptville	2	A8-3	AmbE [Sorangium cellulosum]	62%	ABK32259.1	JerC [Sorangium cellulosum]	66%	ABK32289.1	laidlomycin Streptomyces sp. CS684	74%	AFL48528.1
			AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	65%	ABK32257.1	laidlomycin Streptomyces sp. CS684	73%	AFL48533.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbE [Sorangium cellulosum]	65%	ABK32259.1	Streptomyces mobaraensis NBRC 13819	73%	WP_004954442.1
Cdnforest-Kemptville	3	A3	Chondromyces crocatus	51%	CAQ18829.1	StiF [Stigmatella aurantiaca]	55%	CAD19090.1	Chondromyces crocatus	69%	CAQ18829.1
			Scytonema hofmanni	50%	WP_017746870.1	Scytonema hofmanni	55%	WP_017746870.1	Streptomyces sp. K01-0509	67%	AFU82617.1
			StiH protein [Stigmatella aurantiaca]	50%	CAD19092.1	Chlorogloeopsis sp. PCC 9212	54%	WP_016877275.1	Nostoc punctiforme PCC 73102	66%	YP_001866564.1
Colpasture-Pance	1	A8-1	Saccharomonospora xinjiangensis	61%	WP_006236889.1	Streptomyces himastatinicus ATCC 5	66%	WP_009715048.1	Streptomyces violaceusniger Tu 4113	83%	YP_004813900.1
			Streptomyces sp. CNQ766	61%	WP_018840958.1	Saccharopolyspora erythraea NRRL 2	66%	YP_001106335.1	Mycocerosate [Streptomyces griseus XylebK	82%	WP_003970497.1
			Streptomyces thioluteus	60%	CAE02602.1	Streptomyces sp. CNQ766	65%	WP_018840958.1	NigAV [Streptomyces violaceusniger]	82%	ABC84460.1
Colpasture-Pance	2	A8-2	Soraphen B [Sorangium cellulosum]	65%	AAA79984.2	Soraphen B [Sorangium cellulosum]	72%	AAA79984.2	TgaA [Sorangium cellulosum]	80%	ADH04639.1
			TgaA [Sorangium cellulosum]	64%	ADH04639.1	TgaA [Sorangium cellulosum]	70%	ADH04639.1	polyketide synthase [Sorangium cellulosum	78%	CAL58681.1
			Polyangium cellulosum	62%	CAD43448.1	Polyangium cellulosum	69%	BAG69055.1	AngAV [Streptomyces eurythermus]	77%	ABY21542.1
Colpasture-Pance	3	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerC [Sorangium cellulosum]	61%	ABK32289.1	laidlomycin Streptomyces sp. CS684	73%	AFL48528.1
			AmbE [Sorangium cellulosum]	58%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Streptomyces mobaraensis NBRC 13819	73%	WP_004954442.1
Colpasture-Pance	4	A8-1	Streptomyces rapamycinicus	53%	WP_020874122.1	Streptomyces sp. FxanaC1	62%	WP_018091595.1	Streptomyces bingchenggensis BCW-1	74%	YP_004966668.1
			Streptomyces rapamycinicus	52%	WP_020874116.1	Streptomyces himastatinicus ATCC 5	60%	WP_009715048.1	Nocardiopsis potens	68%	WP_017590936.1
			Streptomyces sp. FxanaC1	52%	WP_018091595.1	Streptomyces rapamycinicus	61%	WP_020874122.1	Nocardiopsis potens	71%	WP_017595632.1
Colpasture-Pance	5	A8-3	JerA [Sorangium cellulosum]	57%	ABK32287.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	megalomicin [Micromonospora megalomice	68%	AAG13917.1
			AmbA [Sorangium cellulosum]	56%	ABK32255.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Saccharopolyspora erythraea NRRL 2338	68%	YP_001106335.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces scabiei 87.22	67%	YP_003493862.1
Colpasture-Pance	6	A8-2	soraphen B [Sorangium cellulosum]	66%	AAA79984.2	soraphen B [Sorangium cellulosum]	71%	AAA79984.2	TgaA [Sorangium cellulosum]	81%	ADH04639.1
			TgaA [Sorangium cellulosum]	64%	ADH04639.1	TgaA [Sorangium cellulosum]	69%	ADH04639.1	Sorangium cellulosum	80%	CAL58681.1
			Polyangium cellulosum	63%	CAD43448.1	Sorangium cellulosum	69%	BAG69055.1	AngAV [Streptomyces eurythermus]	79%	ABY21542.1

Table S2-2											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Colpasture-Pance	7	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	laidlomycin [Streptomyces sp. CS684]	73%	AFL48528.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerC [Sorangium cellulosum]	61%	ABK32289.1	Streptomyces mobaraensis NBRC 13819	73%	WP_004954442.1
			AmbE [Sorangium cellulosum]	58%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Actinoalloteichus spitiensis	73%	WP_016701026.
Colpasture-Pance	8	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	laidlomycin Streptomyces sp. CS684	73%	AFL48528.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerC [Sorangium cellulosum]	61%	ABK32289.1	Streptomyces mobaraensis NBRC 13819	73%	WP_004954442.1
			AmbE [Sorangium cellulosum]	57%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Actinoalloteichus spitiensis	73%	WP_016701026.1
Colpasture-Pance	9	A5	Gloeobacter violaceus PCC 7421	52%	NP_925775.1	Gloeobacter violaceus PCC 7421	52%	NP_925775.1	Gloeobacter violaceus PCC 7421	65%	NP_925775.1
			Sorangium cellulosum So0157	52%	YP_008154278.1	Sorangium cellulosum So0157-2	52%	YP_008154278.1	Sorangium cellulosum So0157-2	64%	YP_008154278.1
			Saccharothrix espanaensis DSM 44229	51%	YP_007039643.1	Saccharothrix espanaensis DSM 4422	51%	YP_007039643.1	Saccharothrix espanaensis DSM 44229	64%	YP_007039643.1
Colpasture-Pance	10	A8-1	Streptomyces globosus	58%	ACL97710.1	Streptomyces rapamycinicus NRRL 54	66%	WP_020871231.1	Salinispora pacifica	82%	WP_018223798.1
			Saccharomonospora xinjiangensis	58%	WP_006236889.1	Streptomyces bingchenggensis BCW-	65%	YP_004959792.1	Streptomyces tsukubaensis NRRL18488	82%	WP_006344682.1
			Streptomyces globosus	58%	ACL97712.1	Streptomyces sp. Mg1	65%	WP_008742159.1	Streptomyces ornatus	81%	ACL97795.1
Colforest-Hatico	1	A8-3	AmbA [Sorangium cellulosum]	53%	ABK32255.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			AmbE [Sorangium cellulosum]	53%	ABK32259.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
			JerA [Sorangium cellulosum]	53%	ABK32287.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Chondromyces crocatus	65%	CAQ18838.1
Colforest-Hatico	2	A8-3	JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			AmbA [Sorangium cellulosum]	54%	ABK32255.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
			AmbE [Sorangium cellulosum]	54%	ABK32259.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Chondromyces crocatus	65%	CAQ18838.1
Colforest-Hatico	3	A6	Myxococcus stipitatus DSM 146751467	53%	YP_007361308.1	Myxococcus xanthus DK 1622	61%	YP_632698.1	Cystobacter fuscus DSM 2262	75%	WP_002631289.1
			Myxococcus xanthus DK 1622	53%	YP_632697.1	Myxococcus xanthus	61%	WP_020477870.1	Stigmatella aurantiaca DW4/3-1	72%	YP_003950451.1
			Sorangium cellulosum So0157-2	52%	YP_008154098.1	Myxococcus stipitatus DSM 14675	61%	YP_007361308.1	Chlorogloeopsis sp. PCC 9212	71%	WP_016872856.1
Colforest-Hatico	4	A5	Sorangium cellulosum So0157-2	51%	YP_008154278.1	StiA protein [Stigmatella aurantiaca]	55%	CAD19085.1	Saccharothrix espanaensis DSM 44229	72%	YP_007039643.1
			Saccharothrix espanaensis DSM 44229	50%	YP_007039643.1	Methylocystis sp. SC2	53%	YP_006591528.1	[Streptomyces sp. FxanaD5	67%	WP_020700051.1
			Hydrogenophaga sp. PB	49%	WP_009517179.1	Streptomyces griseus XylebKG-1	53%	WP_003964706.1	Herpetosiphon aurantiacus DSM 785	65%	YP_001545180.1
Colforest-Hatico	5	A8-2	Sorangium cellulosum	61%	CAL58682.1	TgaC [Sorangium cellulosum]	66%	ADH04641.1	Sorangium cellulosum	81%	CAL58681.1
			soraphenB [Sorangium cellulosum]	59%	AAA79984.2	soraphen B [Sorangium cellulosum]	65%	AAA79984.2	Streptomyces sp. TOR3209	80%	WP_019331156.1
			Sorangium cellulosum	59%	CAL58684.1	TgaA [Sorangium cellulosum]	65%	ADH04639.1	Saccharomonospora azurea NA-128	80%	WP_005441172.1
Colforest-Hatico	6	A5	Myxococcus xanthus DK 1622	47%	YP_632255.1	Nocardia brasiliensis ATCC 700358	55%	YP_006808695.1	uncultured bacterium B7P37metaSE	63%	ACV30045.1
			Myxococcus xanthus	47%	WP_020478336.1	Yersinia rohdei ATCC 43380	55%	WP_004716672.1	Yersinia kristensenii ATCC 33638]	59%	WP_004388876.1
			Nocardia brasiliensis ATCC 700358	47%	YP_006808695.1	Myxococcus xanthus DK 1622	53%	YP_632255.1	Streptomyces griseus	59%	ACL97773.1

Table S2-3												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Colforest-Hatico	7	A5	Sorangium cellulosum So0157-2	51%	YP_008154278.1	StiA protein [Stigmatella aurantiaca]	55%	CAD19085.1	Saccharothrix espanaensis DSM 44229	72%	YP_007039643.1
			Saccharothrix espanaensis DSM 44229	50%	YP_007039643.1	Methylocystis sp. SC2	53%	YP_006591528.1	Streptomyces sp. FxanaD5	67%	WP_020700051.1
			Hydrogenophaga sp. PBC	49%	WP_009517179.1	Streptomyces griseus XylebKG-1	53%	WP_003964706.1	Herpetosiphon aurantiacus DSM 785	65%	YP_001545180.1
Colforest-Hatico	8	A8-2	TgaA [Sorangium cellulosum]	61%	ADH04639.1	TgaA [Sorangium cellulosum]	67%	ADH04639.1	Streptomyces bingchenggensis BCW-1]	81%	YP_004958978.1
			soraphen B [Sorangium cellulosum]	60%	AAA79984.2	JerE [Sorangium cellulosum]	66%	ABK32291.1	Streptomyces spiroverticillatus	81%	ABW96542.1
			JerE [Sorangium cellulosum]	59%	ABK32291.1	TugD [Chondromyces crocatus]	63%	ADH04660.1	Streptomyces sp. TOR3209	80%	WP_019331156.1
Colforest-Hatico	9	A6	Myxococcus stipitatus DSM 14675	53%	YP_007361308.1	polyketide synthase [Myxococcus sti	60%	YP_007361308.1	Cystobacter fuscus DSM 2262	75%	WP_002631289.1
			Myxococcus xanthus DK 1622	53%	YP_632697.1	polyketide synthase [Myxococcus xa	60%	YP_632698.1	Stigmatella aurantiaca DW4/3-1	72%	YP_003950451.1
			Chondromyces crocatus	52%	CAQ18832.1	polyketide synthase [Myxococcus xa	60%	WP_020477870.1	Chlorogloeopsis sp. PCC 9212	71%	WP_016872856.1
Colforest-Hatico	10	A6	Chondromyces crocatus	55%	CBD77734.1	Sorangium cellulosum So0157-2	61%	YP_008149810.1	Chondromyces crocatus	73%	CBD77732.1
			Myxococcus stipitatus DSM 146751467	54%	YP_007361308.1	Cystobacter fuscus DSM 2262	60%	WP_002624105.1	Mycobacterium avium	73%	WP_019735124.1
			Sorangium cellulosum So0157-2	54%	YP_008149810.1	MmxB [Cystobacter fuscus]	59%	ABA29782.1	Moorea producens	72%	WP_008179436.1
Colforest-Hatico	11	A8-3	AmbA [Sorangium cellulosum]	54%	ABK32255.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Scytonema hofmanni	65%	WP_017745372.1
			AmbE [Sorangium cellulosum]	54%	ABK32259.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Chondromyces crocatus	65%	CAQ18838.1
Colforest-Hatico	12	A8-1	Salinispora arenicola	61%	WP_018802054.1	Streptomyces rapamycinicus NRRL 54	65%	WP_020865106.1	Nocardiopsis sp. FU40	81%	AEP40936.1
			Salinispora arenicola	61%	WP_019030802.1	Streptomyces rapamycinicus NRRL 54	63%	WP_020874117.1	Frankia sp. BCU110501	80%	WP_020572514.1
			Salinispora arenicola	60%	WP_018588345.	Streptomyces violaceusniger Tu 4113	63%	YP_004814091.1	Streptomyces griseus NBRC 13350]	80%	YP_001827691.1
Colforest-Hatico	13	A6	Myxococcus stipitatus DSM 14675	53%	YP_007361308.1	Myxococcus xanthus DK 1622	61%	YP_632698.1	Cystobacter fuscus DSM 2262	73%	WP_002631289.1
			Myxococcus xanthus DK 1622	53%	YP_632697.1	Myxococcus xanthus	61%	WP_020477870.1	Stigmatella aurantiaca DW4/3-1	71%	YP_003950451.1
			Sorangium cellulosum So0157-2	52%	YP_008154098.	Myxococcus stipitatus DSM 14675	61%	YP_007361308.1	Chlorogloeopsis sp. PCC 9212	70%	WP_016872856.1
Colforest-Hatico	14	A8-3	AmbA [Sorangium cellulosum]	54%	ABK32255.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
			AmbE [Sorangium cellulosum]	53%	ABK32259.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Chondromyces crocatus	65%	CAQ18838.1
Colforest-Hatico	15	A8-1	Amycolatopsis orientalis HCCB10007	55%	YP_008011939.1	Streptomyces sp. HPH0547	56%	WP_016467328.1	ChlA3 [Streptomyces antibioticus]	80%	AAZ77696.1
			Streptomyces bingchenggensis BCW-	53%	YP_004966650.1	Streptomyces rapamycinicus NRRL 54	56%	WP_020874116.1	ChlA1 [Streptomyces antibioticus]	80%	AAZ77693.1
			Streptomyces sp. LaPpAH-202	53%	WP_018471165.	Herc [Streptomyces chromofuscus]	56%	AEZ64504.1	Streptomyces sp. CNY243	79%	WP_018851822.1
Colforest-Hatico	16	A6	Myxococcus stipitatus DSM 14675	52%	YP_007361308.1	Myxococcus xanthus	58%	WP_020477870.1	Cystobacter fuscus DSM 2262	73%	WP_002631289.1
			Myxococcus xanthus DK 1622	51%	YP_632697.1	Myxococcus xanthus DK 1622	58%	YP_632698.1	Stigmatella aurantiaca DW4/3-1	71%	YP_003950451.1
			Sorangium cellulosum So0157-2	51%	YP_008154098.1	Myxococcus stipitatus DSM 14675	58%	YP_007361308.1	Chlorogloeopsis sp. PCC 9212	70%	WP_016872856.

Table S2-4												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Colforest-Hatico	17	A6	Myxococcus stipitatus DSM 14675	54%	YP_007361308.1	Sorangium cellulosum So0157-2	61%	YP_008149810.1	Chondromyces crocatus	71%	CBD77732.1
			Sorangium cellulosum So0157-2	54%	YP_008149810.1	Cystobacter fuscus DSM 2262	60%	WP_002624105.1	Mycobacterium avium	71%	WP_019735124.1
			Chondromyces crocatus	54%	CBD77734.1	MmxB [Cystobacter fuscus]	60%	ABA29782.1	Moorea producens 3L	71%	WP_008179436.1
Colforest-CIAT	1	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces sp. HPH0547	73%	WP_016467321.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces viridochromogenes Tue57	72%	WP_004004070.1
			AmbB [Sorangium cellulosum]	58%	ABK32256.1	AmbE [Sorangium cellulosum]	61%	ABK32259.1	Streptomyces sp. ScaeMP-e10	71%	WP_018515076.1
Colforest-CIAT	2	A8-1	Streptomyces rapamycinicus	61%	WP_020866324.1	Streptomyces rapamycinicus NRRL 54	69%	WP_020866324.1	Streptomyces rapamycinicus NRRL 5491	79%	WP_020872959.1
			Streptomyces sp. CNQ766	60%	WP_018840958.1	Streptomyces himastatinicus ATCC 5	68%	WP_009719940.1	PimS2 [Streptomyces natalensis]	79%	CAC20921.1
			Streptomyces rapamycinicus	60%	WP_020866323.1	Streptomyces sp. CNQ766	67%	WP_018840958.1	TugB [Chondromyces crocatus]	79%	ADH04658.1
Colforest-CIAT	3	A8-3	AmbA [Sorangium cellulosum]	55%	ABK32255.1	AmbC [Sorangium cellulosum]	58%	ABK32257.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			AmbC [Sorangium cellulosum]	54%	ABK32257.1	AmbE [Sorangium cellulosum]	58%	ABK32259.1	Streptomyces albus	65%	CCD31890.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Streptomyces sp. JS360	64%	CBW54671.1
Colforest-CIAT	4	A8-3	JerA [Sorangium cellulosum]	55%	ABK32287.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Gloeobacter violaceus PCC 7421	67%	NP_925775.1
			AmbA [Sorangium cellulosum]	54%	ABK32255.1	AmbE [Sorangium cellulosum]	58%	ABK32259.1	Scytonema hofmanni	65%	WP_017745372.1
			AmbC [Sorangium cellulosum]	54%	ABK32257.1	JerB [Sorangium cellulosum]	58%	ABK32288.1	Chondromyces crocatus	65%	CAQ18838.1
Colforest-CIAT	5	A8-3	AmbA [Sorangium cellulosum]	55%	ABK32255.1	AmbC [Sorangium cellulosum]	58%	ABK32257.1	Gloeobacter violaceus PCC 7421	68%	NP_925775.1
			AmbC [Sorangium cellulosum]	54%	ABK32257.1	AmbE [Sorangium cellulosum]	58%	ABK32259.1	Pelosinus fermentans JBW45	65%	WP_007954884.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
Colforest-CIAT	6	A8-3	AmbA [Sorangium cellulosum]	55%	ABK32255.1	AmbC [Sorangium cellulosum]	58%	ABK32257.1	Gloeobacter violaceus PCC 7421	68%	NP_925775.1
			AmbC [Sorangium cellulosum]	54%	ABK32257.1	AmbE [Sorangium cellulosum]	58%	ABK32259.1	Pelosinus fermentans JBW45	65%	WP_007954884.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
Colforest-CIAT	7	A8-3	AmbA [Sorangium cellulosum]	55%	ABK32255.1	AmbC [Sorangium cellulosum]	58%	ABK32257.1	Gloeobacter violaceus PCC 7421	68%	NP_925775.1
			AmbC [Sorangium cellulosum]	54%	ABK32257.1	AmbE [Sorangium cellulosum]	58%	ABK32259.1	Pelosinus fermentans JBW45	65%	WP_007954884.1
			JerA [Sorangium cellulosum]	54%	ABK32287.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Scytonema hofmanni	65%	WP_017745372.1
Czcultivated-Plana	1	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces sp. CcalMP-8W	74%	WP_018491143.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. Wigar10	73%	WP_019765969.1
Czcultivated-Plana	2	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	61%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. HPH0547	73%	WP_016467321.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1

Table S2-5												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Czcultivated-Plana	3	A8-1	Streptomyces sp. NRRL 30748	55%	ABC87511.1	Streptomyces violaceusniger Tu 4113	62%	YP_004815232.1	Streptomyces globosus	87%	ACL97710.1
			Amycolatopsis vancoresmycina	55%	WP_004560666.1	Streptomyces sp. NRRL 30748	61%	ABC87511.1	Streptomyces halstedii	87%	BAF02925.1
			Streptomyces violaceusniger Tu 4113	55%	YP_004815233.1	MerA [Streptomyces violaceusniger]	61%	ABJ97437.1	[Nocardiopsis chromatogenes	86%	WP_017627467.1
Czcultivated-Plana	4	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	QmnA2 [Amycolatopsis orientalis]	73%	AFI57006.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
Czcultivated-Plana	5	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	QmnA2 [Amycolatopsis orientalis]	73%	AFI57006.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
Czcultivated-Plana	6	A8-3	AmbC [Sorangium cellulosum]	57%	ABK32257.1	JerC [Sorangium cellulosum]	57%	ABK32289.1	Streptomyces acidiscabies	74%	WP_010356044.1
			JerC [Sorangium cellulosum]	56%	ABK32289.1	AmbC [Sorangium cellulosum]	57%	ABK32257.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	55%	ABK32259.1	AmbE [Sorangium cellulosum]	54%	ABK32259.1	Streptomyces viridochromogenes Tue57	72%	WP_004004070.1
Czcultivated-Plana	7	A8-3	Poor chromatogram quality.				64%	ABK32255.1	Poor chromatogram quality.		
							64%	ABK32257.1			
							63%	ABK32287.1			
Czcultivated-Plana	8	A8-3	AmbC [Sorangium cellulosum]	62%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces sp. CcalMP-8W	74%	WP_018491143.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. Wigar10	73%	WP_019765969.1
Czcultivated-Plana	9	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	fosC [Streptomyces pulveraceus]	74%	AEC13069.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Frankia sp. BMG5.12	73%	WP_018639361.1
Czcultivated-Plana	10	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	64%	ABK32255.1	Streptomyces sp. CcalMP-8W]	74%	WP_018491143.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces sp. Wigar10	73%	WP_019765969.1
Czcultivated-Plana	11	A5	Nocardiopsis halophila	51%	WP_017542010.1	Streptomyces sp. ATexAB-D23	53%	WP_018552688.1	Tistrella mobilis KA081020-065	69%	YP_006369910.1
			Nocardiopsis baichengensis	51%	WP_017559898.1	Streptomyces sp. ScaeMP-e10	52%	WP_018513160.1	Singulisphaera acidiphila DSM 18658	68%	YP_007202109.1
			Sorangium cellulosum So0157-2	50%	YP_008154278.1	Streptomyces sp. KhCrAH-244	52%	WP_018519914.1	Plesiocystis pacifica SIR-1	68%	WP_006977101.1
Česke.Budéjovicesoil	1	A6	Myxococcus xanthus	59%	WP_020477870.1	EpoD [synthetic construct]	64%	ABB92693.1	MelF [Melittangium lichenicola]	65%	CAD89777.1
Paradise			Myxococcus xanthus	59%	YP_632698.1	epoD [Sorangium cellulosum]	64%	AAF62883.1	Sorangium cellulosum So0157-2	65%	YP_008154278.1
			Sorangium cellulosum So0157-2	55%	YP_008154098.1	Sorangium cellulosum So0157-2	64%	ACB46195.1	Gloeocapsa sp. PCC 73106	63%	WP_006529839.1

Table S2-6

Sample	#	AT group	BLASTP Full length	% identity	Accession number	BLASTP AT domain	% identity	Accession number	BLASTP KS domain	% identit	Accession number
České.Budéjovicesoil	2	A8-3	AmbE [Sorangium cellulosum]	56%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Actinoplanes sp. SE50/110	69%	YP_006269033.1
Paradise			AmbC [Sorangium cellulosum]	56%	ABK32257.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Mycocerosate [Tsukamurella paurometabola	69%	YP_003648400.1
			JerB [Sorangium cellulosum]	55%	ABK32288.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Streptomyces avermitilis MA-4680	69%	NP_822727.1
České.Budéjovicesoil	3	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
Paradise			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces sp. Wigar10	72%	WP_019765969.1
České.Budéjovicesoil	4	A8-3	AmbE [Sorangium cellulosum]	56%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Streptomyces avermitilis MA-4680	72%	NP_822727.1
Paradise			AmbC [Sorangium cellulosum]	55%	ABK32257.1	AmbB [Sorangium cellulosum]	58%	ABK32256.1	Mycocerosate synthase [Tsukamurella pauro	71%	YP_003648400.1
			JerB [Sorangium cellulosum]	55%	ABK32288.1	JerB [Sorangium cellulosum]	58%	ABK32288.1	Actinoplanes sp. SE50/110	71%	YP_006269033.1
České.Budéjovicesoil	5	A8-3	AmbE [Sorangium cellulosum]	56%	ABK32259.1	AmbE [Sorangium cellulosum]	59%	ABK32259.1	Streptomyces avermitilis MA-4680	72%	NP_822727.1
Paradise			JerB [Sorangium cellulosum]	55%	ABK32288.1	AmbB [Sorangium cellulosum]	59%	ABK32256.1	Mycocerosate[Tsukamurella paurometabola	71%	YP_003648400.1
			AmbC [Sorangium cellulosum]	55%	ABK32257.1	JerB [Sorangium cellulosum]	59%	ABK32288.1	Actinoplanes sp. SE50/110	71%	YP_006269033.1
České.Budéjovicesoil	6	A5	Chlorogloeopsis sp. PCC 9212	53%	WP_016872856.1	Singulisphaera acidiphila DSM 18658	57%	YP_007202109.1	Salinispora arenicola	71%	WP_019032757.1
Paradise			Singulisphaera acidiphila DSM 18658	52%	YP_007202109.1	Chlorogloeopsis sp. PCC 9212	55%	WP_016872856.1	Saccharopolyspora erythraea NRRL 2338	67%	YP_001104830.1
			Pelosinus fermentans	50%	WP_007954884.1	Pelosinus fermentans JBW45	55%	WP_007954884.1	[Saccharothrix espanaensis DSM 44229	67%	YP_007039643.1
České.Budéjovicesoil	7	A6	Mycobacterium tusciae	68%	WP_006245903.1	Mycobacterium tusciae JS617	72%	WP_006245903.1	Mycobacterium tusciae JS617	75%	WP_006245903.1
Paradise			MmxB [Cystobacter fuscus]	51%	ABA29782.1	Myxococcus stipitatus	56%	BAG69001.1	Moorea producens 3L	68%	WP_008179436.1
			Ktedonobacter racemifer	50%	WP_007919320.1	Myxococcus virescens	56%	BAG68991.1	Stigmatella aurantiaca DW4/3-1	68%	YP_003950451.1
České.Budéjovicesoil	8	A5	Singulisphaera acidiphila DSM 18658	46%	YP_007202109.1	Sorangium cellulosum So0157-2	48%	YP_008154098.1	Scytonema hofmanni	67%	WP_017745372.1
Paradise			Sorangium cellulosum So0157-2	44%	YP_008154098.1	Singulisphaera acidiphila DSM 18658	48%	YP_007202109.1	Moorea producens 3L	66%	WP_008179436.1
			Gloeobacter violaceus PCC 742	45%	NP_924900.1	Chondromyces crocatus	47%	CAQ18838.1	Cystobacter fuscus DSM 2262	63%	WP_002631289.1
Czpasture-Palava	1	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	62%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
			AmbE [Sorangium cellulosum]	61%	ABK32259.1	AmbE [Sorangium cellulosum]	62%	ABK32259.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces sp. HPH0547	73%	WP_016467321.1
Czpasture-Palava	2	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbC [Sorangium cellulosum]	61%	ABK32257.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerC [Sorangium cellulosum]	60%	ABK32289.1	Streptomyces sp. Wigar10	72%	WP_019765969.1
			AmbE [Sorangium cellulosum]	58%	ABK32259.1	AmbA [Sorangium cellulosum]	60%	ABK32255.1	Streptomyces sp. ScaeMP-e10	71%	WP_018515076.1
Czpasture-Palava	3	A8-3	AmbE [Sorangium cellulosum]	57%	ABK32259.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Micromonospora megalomicea subsp. Nigra	76%	AAG13918.1
			AmbC [Sorangium cellulosum]	57%	ABK32257.1	AmbE [Sorangium cellulosum]	64%	ABK32259.1	Saccharomonospora paurometabolica YIM 9(	75%	WP_007027132.1
			JerC [Sorangium cellulosum]	56%	ABK32289.1	AmbB [Sorangium cellulosum]	64%	ABK32256.1	Streptomyces rapamycinicus NRRL 5491	75%	WP_020866165.1

Table S2-7												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Czpasture-Palava	4	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	62%	ABK32257.1	Streptomyces sp. CcalMP-8W	74%	WP_018491143.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces sp. HPH0547	74%	WP_016467321.1
			AmbE [Sorangium cellulosum]	60%	ABK32259.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	Streptomyces acidiscabies	74%	WP_010356044.1
Czpasture-Palava	5	A8-1	Streptomyces sp. CNQ766	55%	WP_018840958.1	Streptomyces sp. Mg1	56%	WP_008742159.1	Streptomyces sp. TOR3209	77%	WP_019331156.1
			Streptomyces sp. CNS335	55%	WP_018842118.1	Streptomyces sp. Mg1	56%	WP_008742168.1	Streptomyces tsukubaensis NRRL18488	77%	WP_006344599.1
			Streptomyces violaceusniger Tu 4113	55%	YP_004817599.1	Streptomyces sp. FxanaC1	55%	WP_018091595.	Streptomyces cattleya NRRL 8057	77%	YP_004912779.1
Czpasture-Palava	6	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbA [Sorangium cellulosum]	62%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	JerA [Sorangium cellulosum]	62%	ABK32287.1	laidlomycin [Streptomyces sp. CS684]	73%	AFL48528.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbC [Sorangium cellulosum]	62%	ABK32257.1	[Streptomyces viridochromogenes Tue57	72%	WP_004004070.1
Czpasture-Palava	7	A8-1	Streptomyces bingchenggensis BCW-	68%	YP_004960235.1	Streptomyces rapamycinicus NRRL 5	73%	WP_020872449.1	BFAS4 [Kitasatospora setae KM-6054]	82%	YP_004909058.1
			Amycolatopsis benzoatilytica	67%	WP_020659282.1	Amycolatopsis benzoatilytica	72%	WP_020659280.1	Streptomyces mobaraensis NBRC 13819	82%	WP_004951568.1
			oleandomycin [Streptomyces himast	66%	WP_009715040.1	Streptomyces bingchenggensis BCW-	72%	YP_004960235.1	Streptomyces sp. R1-NS-10	81%	WP_019075974.1
Czpasture-Palava	8	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	Streptomyces avermitilis MA-4680	69%	NP_822727.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	AmbC [Sorangium cellulosum]	61%	ABK32257.1	Haliangium ochraceum DSM 14365[	68%	YP_003267362.1
			AmbA [Sorangium cellulosum]	58%	ABK32255.1	JerA [Sorangium cellulosum]	60%	ABK32287.1	Haliangium ochraceum DSM 14365]	68%	YP_003267363.1
Czpasture-Palava	9	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbE [Sorangium cellulosum]	62%	ABK32259.1	Streptomyces sp. HPH0547	71%	WP_016467321.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbC [Sorangium cellulosum]	61%	ABK32257.1	Streptomyces viridochromogenes Tue57	70%	WP_004004070.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	Streptomyces sp. ATexAB-D23	69%	WP_018555451.1
CdnTundra-Resolute	1	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces acidiscabies	71%	WP_010356044.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. CcalMP-8W	68%	WP_018491143.1
			AmbA [Sorangium cellulosum]	58%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	laidlomycin [Streptomyces sp. CS684]	68%	AFL48528.1
CdnTundra-Resolute	2	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	[Streptomyces acidiscabies]	73%	WP_010356044.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
CdnTundra-Resolute	3	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	75%	WP_010356044.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	laidlomycin [Streptomyces sp. CS684]	72%	AFL48533.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces viridochromogenes Tue57	71%	WP_004004070.1
CdnTundra-Resolute	4	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1

Table S2-8											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
CdnTundra-Resolute	5	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	laidlomycin [Streptomyces sp. CS684]	72%	AFL48528.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerA [Sorangium cellulosum]	61%	ABK32287.1	Streptomyces viridochromogenes Tue57	71%	WP_004004070.1
			AmbE [Sorangium cellulosum]	57%	ABK32259.1	AmbC [Sorangium cellulosum]	60%	ABK32257.1	Nocardiopsis sp. CNS639	71%	WP_019606713.1
CdnTundra-Resolute	6	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	62%	ABK32257.1	Streptomyces viridochromogenes Tue57	70%	WP_004004070.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces sp. ATexAB-D23	69%	WP_018555451.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	Streptomyces acidiscabies	69%	WP_010356044.1
CdnTundra-Resolute	7	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
CdnTundra-Resolute	8	A8-3	AmbC [Sorangium cellulosum]	58%	ABK32257.1	AmbC [Sorangium cellulosum]	60%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1
			AmbE [Sorangium cellulosum]	58%	ABK32259.1	AmbA [Sorangium cellulosum]	59%	ABK32255.1	Streptomyces sp. Wigar10	71%	WP_019765969.1
			AmbA [Sorangium cellulosum]	57%	ABK32255.1	JerA [Sorangium cellulosum]	59%	ABK32287.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
CdnTundra-Resolute	9	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. Wigar10	72%	WP_019765969.1
CdnTundra-Resolute	10	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	75%	WP_010356044.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	laidlomycin [Streptomyces sp. CS684]	73%	AFL48533.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces viridochromogenes Tue57	72%	WP_004004070.1
CdnTundra-Resolute	11	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbA [Sorangium cellulosum]	61%	ABK32255.1	laidlomycin [Streptomyces sp. CS684]	72%	AFL48528.1
			JerC [Sorangium cellulosum]	58%	ABK32289.1	JerA [Sorangium cellulosum]	61%	ABK32287.1	Streptomyces viridochromogenes Tue57	71%	WP_004004070.1
			AmbE [Sorangium cellulosum]	57%	ABK32259.1	AmbC [Sorangium cellulosum]	60%	ABK32257.1	Nocardiopsis sp. CNS639	71%	WP_019606713.1
CdnTundra-Resolute	12	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. Wigar10	72%	WP_019765969.1
CdnTundra-Resolute	13	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbA [Sorangium cellulosum]	60%	ABK32255.1	Streptomyces sp. ATexAB-D23	69%	WP_018555451.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	60%	ABK32289.1	Streptomyces acidiscabies	69%	WP_010356044.1
			AmbE [Sorangium cellulosum]	58%	ABK32259.1	AmbC [Sorangium cellulosum]	60%	ABK32257.1	modularStreptomyces sp. S4	68%	WP_010638173.1
CdnTundra-Resolute	14	A8-1	Streptomyces violaceusniger Tu 4113	60%	YP_004817599.1	Streptomyces sp. Mg1	64%	WP_008743187.1	Streptomyces halstedii	88%	BAD08373.1
			Streptomyces rapamycinicus NRRL 54	60%	WP_020871229.1	Streptomyces rimosus ATCC 10970]	63%	WP_003982116.1	Nocardiopsis gilva	88%	WP_017620963.1
			Streptomyces violaceusniger Tu 4113	60%	YP_004814097.1	Streptomyces sp. PsTaAH-124	63%	WP_018565205.1	Streptomyces rapamycinicus NRRL 5491	87%	WP_020869284.1

Table S2-9												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
CdnTundra-Resolute	15	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
CdnTundra-Resolute	16	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	laidlomycin [Streptomyces sp. CS684]	71%	AFL48533.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces viridochromogenes Tue57	70%	WP_004004070.1
CdnTundra-Resolute	17	A8-3	AmbC [Sorangium cellulosum]	59%	ABK32257.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces acidiscabies	73%	WP_010356044.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces sp. CcalMP-8W	72%	WP_018491143.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	Streptomyces viridochromogenes Tue57	71%	WP_004004070.1
CdnTundra-Resolute	18	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces acidiscabies	73%	WP_010356044.1
			AmbA [Sorangium cellulosum]	59%	ABK32255.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
CdnBeach-P.E.I.	1	A8-1	Streptomyces clavuligerus	53%	WP_003959186.1	Amycolatopsis vancoresmycina DSM	56%	WP_004560200.1	CppC [Pseudonocardia autotrophica]	80%	ABV83230.1
			Streptomyces clavuligerus	53%	WP_009995036.1	Amycolatopsis vancoresmycina DSM	56%	WP_004559079.1	Streptomyces gancidicus BKS 13-15	79%	WP_006129567.1
			Amycolatopsis orientalis HCCB10007	53%	YP_008011939.1	Streptomyces scabiei 87.22	56%	YP_003493864.1	Streptomyces hygroscopicus	78%	AAR01274.1
CdnBeach-P.E.I.	2	A8-1	Streptomyces prunicolor	64%	WP_019065464.1	Streptomyces prunicolor	66%	WP_019065464.1	Streptomyces prunicolor	79%	WP_019065464.1
			Micromonospora sp. ATCC 39149	60%	WP_007071854.1	Streptomyces rapamycinicus NRRL 54	65%	WP_020874117.1	Nocardia asteroides	79%	WP_019048597.1
			Streptomyces himastatinicus	60%	WP_009717842.1	Streptomyces mobaraensis NBRC 138	64%	WP_004940991.1	Micromonospora sp. ATCC 39149	79%	WP_007071850.1
CdnBeach-P.E.I.	3	A8-1	Streptomyces bingchenggensis BCW-	48%	YP_004966650.1	Streptomyces bingchenggensis BCW-	50%	YP_004966650.1	CppC [Pseudonocardia autotrophica]	79%	ABV83230.1
			Streptomyces thioluteus	48%	CAE02606.1	Streptomyces scabiei 87.22	49%	YP_003493864.1	Streptomyces gancidicus BKS 13-15	77%	WP_006129567.1
			NanA11 [Streptomyces nanchangensi	48%	AAP42873.1	Streptomyces olivoviridis	49%	ACL97716.1	Streptomyces hygroscopicus	76%	AAR01274.1
CdnBeach-P.E.I.	4	A8-1	Actinokineospora enzanensis	67%	WP_018680913.1	Actinokineospora enzanensis	67%	WP_018680913.1	Streptomyces sp. S4	80%	WP_010638160.1
			FscB [Streptomyces albus J1074]	62%	YP_007749163.1	Amycolatopsis balhimycina	65%	WP_020638041.1	Streptomyces halstedii	80%	BAF02926.1
			Streptomyces sp. S4	62%	WP_010639290.1	Streptomyces violaceusniger Tu 4113	65%	YP_004814091.1	Streptomyces violaceusniger Tu 4113]	80%	YP_004815232.1
CdnBeach-P.E.I.	5	A8-1	Herc [Streptomyces chromofuscus]	63%	AEZ64504.1	Actinokineospora enzanensis	65%	WP_018680913.1	Streptomyces rapamycinicus NRRL 5491	87%	WP_020869284.1
			QmnA1 [Amycolatopsis orientalis]	62%	AFI57005.1	QmnA1 [Amycolatopsis orientalis]	64%	AFI57005.1	Streptomyces aureofaciens	85%	BAH67283.1
			Streptomyces violaceusniger Tu 4113	62%	YP_004813739.1	Streptomyces platensis	64%	BAH02270.1	Saccharomonospora xinjiangensis XJ-54	85%	WP_006236439.1
CdnBeach-P.E.I.	6	A8-1	Amycolatopsis orientalis HCCB10007	53%	YP_008011939.1	Streptomyces sp. M-02	57%	ACM62767.1	CppC [Pseudonocardia autotrophica]	80%	ABV83230.1
			Streptomyces clavuligerus	52%	WP_009995036.1	Nocardiopsis sp. FU40	56%	AEP40932.1	Streptomyces gancidicus BKS 13-15	79%	WP_006129567.1
			Streptomyces clavuligerus	52%	WP_003959186.1	Streptomyces scabiei 87.22	56%	YP_003493864.1	Streptomyces hygroscopicus	78%	AAR01274.1

Table S2-10												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
CdnBeach-P.E.I.	7	A8-1	Herc [Streptomyces chromofuscus]	63%	AEZ64504.1	Actinokineospora enzanensis	65%	WP_018680913.1	Streptomyces rapamycinicus NRRL 5491	87%	WP_020869284.1
			QmnA1 [Amycolatopsis orientalis]	62%	AFI57005.1	mnA1 [Amycolatopsis orientalis]	64%	AFI57005.1	Streptomyces aureofaciens	85%	BAH67283.1
			Streptomyces violaceusniger Tu 4113	62%	YP_004813739.1	Streptomyces platensis	64%	BAH02270.1	Saccharomonospora xinjiangensis XJ-54	85%	WP_006236439.1
CdnBeach-P.E.I.	8	A8-1	Streptomyces mobaraensis	60%	WP_004951693.1	Streptomyces mobaraensis NBRC 138	67%	WP_004951693.1	Streptomyces flaveolus	82%	ACY06287.1
			Streptomyces violaceusniger Tu 4113	59%	YP_004817597.1	Streptomyces auratus AGR0001	67%	WP_006608277.1	Streptomyces flaveolus	81%	ACY06289.1
			Streptomyces flaveolus	59%	ACY06287.1	Streptomyces mobaraensis NBRC 138	67%	WP_004951699.1	Streptomyces sp. R1-NS-10	80%	WP_019075974.1
Hungforest-Citadel	1	A8-1	Streptomyces bingchenggensis BCW-	66%	YP_004960235.1	Amycolatopsis benzoatilytica	74%	WP_020659280.1	[Streptomyces halstedii	83%	BAD38978.1
			Amycolatopsis benzoatilytica	66%	WP_020659280.1	Streptomyces bingchenggensis BCW-	73%	YP_004960235.1	Streptomyces sp. CNY243	82%	WP_018851822.1
			AmphI [Streptomyces nodosus]	65%	AAK73501.1	Streptomyces bingchenggensis BCW-	73%	YP_004960237.1	Streptomyces sp. CNQ766	82%	WP_018840991.1
Hungforest-Citadel	2	A8-1	Streptomyces acidiscabies	57%	WP_010360984.1	Streptomyces hygroscopicus	63%	CCF23200.1	Streptomyces violaceusniger Tu 4113	82%	YP_004815233.1
			Streptomyces sp. Wigar10	57%	WP_019766151.1	[Streptomyces rapamycinicus NRRL 5	62%	WP_020865106.1	Streptomyces mobaraensis NBRC 13819	82%	WP_004951568.1
			AVES 4 [Streptomyces mobaraensis]	57%	WP_004951568.1	[Streptomyces rapamycinicus NRRL 5	62%	WP_020874116.1	Nocardiopsis chromatogenes	82%	WP_017627467.1
Hungforest-Citadel	3	A8-3	AmbE [Sorangium cellulosum]	59%	ABK32259.1	AmbB [Sorangium cellulosum]	64%	ABK32256.1	laidlomycin [Streptomyces sp. CS684]	71%	AFL48533.1
			JerB [Sorangium cellulosum]	57%	ABK32288.1	AmbE [Sorangium cellulosum]	64%	ABK32259.1	QmnA2 [Amycolatopsis orientalis]	70%	AFI57006.1
			AmbC [Sorangium cellulosum]	57%	ABK32257.1	JerC [Sorangium cellulosum]	63%	ABK32289.1	laidlomycin [Streptomyces sp. CS684]	69%	AFL48528.1
Hungforest-Citadel	4	A6	Myxococcus xanthus DK 1622	55%	YP_632698.1	Myxococcus xanthus DK 1622	61%	YP_632698.1	Cystobacter fuscus DSM 2262	72%	WP_002631289.1
			Myxococcus xanthus	55%	WP_020477870.1	Myxococcus xanthus	61%	WP_020477870.1	Stigmatella aurantiaca DW4/3-1	70%	YP_003950451.1
			Myxococcus stipitatus DSM 14675	53%	YP_007361308.1	Myxococcus stipitatus DSM 14675	61%	YP_007361308.1	Chlorogloeopsis sp. PCC 9212	69%	WP_016872856.1
Hungforest-Citadel	5	A8-3	AmbE [Sorangium cellulosum]	61%	ABK32259.1	JerC [Sorangium cellulosum]	64%	ABK32289.1	laidlomycin [Streptomyces sp. CS684]	71%	AFL48533.1
			JerC [Sorangium cellulosum]	61%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	69%	WP_018555451.1
			AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	laidlomycin [Streptomyces sp. CS684]	69%	AFL48528.1
Hungforest-Citadel	6	Outlier	Sorangium cellulosum	56%	CAL58682.1	TgaA [Sorangium cellulosum]	59%	ADH04639.1	Streptomyces acidiscabies	77%	WP_010356044.1
			TgaA [Sorangium cellulosum]	56%	ADH04639.1	JerE [Sorangium cellulosum]	58%	ABK32291.1	Streptomyces sp. CS684	76%	AFL48527.1
			soraphen B [Sorangium cellulosum]	56%	AAA79984.2	polyketide synthase [Sorangium cell	58%	BAG69055.1	Streptomyces nanchangensis	75%	ADC45535.1
Hungforest-Citadel	7	A8-1	Streptomyces acidiscabies	59%	WP_010360984.1	Streptomyces hygroscopicus	67%	CCF23200.1	Streptomyces violaceusniger Tu 4113	82%	YP_004815233.1
			Streptomyces sp. Wigar10	59%	WP_019766151.1	Streptomyces rapamycinicus NRRL 54	66%	WP_020865106.1	Streptomyces mobaraensis NBRC 13819	82%	WP_004951568.1
			AVES 4 Streptomyces mobaraensis]	59%	WP_004951568.1	Streptomyces rapamycinicus NRRL 54	66%	WP_020874116.1	Nocardiopsis chromatogenes	82%	WP_017627467.1
Hungforest-Citadel	8	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces acidiscabies	73%	WP_010356044.1
			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces sp. CcalMP-8W	71%	WP_018491143.1
			AmbA [Sorangium cellulosum]	58%	ABK32255.1	JerA [Sorangium cellulosum]	63%	ABK32287.1	laidlomycin [Streptomyces sp. CS684]	71%	AFL48528.1

Table S2-11											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Hungforest-Citadel	9	Outlier	Sorangium cellulosum	56%	CAL58682.1	TgaA [Sorangium cellulosum]	59%	ADH04639.1	Streptomyces acidiscabies	77%	WP_010356044.1
			TgaA [Sorangium cellulosum]	56%	ADH04639.1	JerE [Sorangium cellulosum]	58%	ABK32291.1	laidlomycin [Streptomyces sp. CS684]	76%	AFL48527.1
			soraphen B [Sorangium cellulosum]	56%	AAA79984.2	Sorangium cellulosum	58%	BAG69055.1	Streptomyces nanchangensis	75%	ADC45535.1
Hungforest-Citadel	10	A8-1	Streptomyces acidiscabies	59%	WP_010360984.1	Streptomyces hygroscopicus	67%	CCF23200.1	Streptomyces sp. CNB091	83%	WP_018962025.1
			AVES 4 [Streptomyces mobaraensis]	59%	WP_004951568.1	Streptomyces rapamycinicus NRRL 54	66%	WP_020865106.1	Streptomyces sp. Wigar10	82%	WP_019761156.1
			Streptomyces violaceusniger Tu 411	57%	YP_004816248.1	Streptomyces rapamycinicus NRRL 54	65%	WP_020874116.1	Streptomyces violaceusniger Tu 4113	82%	YP_004815233.1
Hungforest-Citadel	11	A8-1	MerA [Streptomyces violaceusniger]	63%	ABJ97437.1	MerA [Streptomyces violaceusniger]	69%	ABJ97437.1	Streptomyces sp. ScaeMP-e10	82%	WP_018514182.1
			Streptomyces violaceusniger Tu 4113	63%	YP_004815233.1	Streptomyces rapamycinicus NRRL 54	68%	WP_020874116.1	Kitasatospora setae KM-6054	81%	YP_004909056.1
			Streptomyces sp. NRRL 30748	62%	ABC87509.1	MerC [Streptomyces violaceusniger]	68%	ABJ97439.1	Streptomyces sp. C	81%	WP_007269165.1
Hungforest-Citadel	12	A8-1	Streptomyces acidiscabies	64%	WP_010360984.1	Streptomyces hygroscopicus	71%	CCF23202.1	Streptomyces tsukubaensis NRRL18488	82%	WP_006344990.1
			Streptomyces sp. Wigar10	59%	WP_019766151.1	Streptomyces hygroscopicus	71%	CCF23203.1	Streptomyces avermitilis MA-4680	81%	NP_824073.1
			Streptomyces sp. NRRL 30748	66%	ABC87511.1	Streptomyces violaceusniger Tu 4113	71%	YP_004815231.1	Streptomyces sp. Wigar10	81%	WP_019762301.1
Brusselscobblestones	1	Outlier	Chlorogloeopsis sp. PCC 9212	45%	WP_016872856.1	Gloeobacter violaceus PCC 7421	49%	NP_925775.1	CurK [Moorea producens 3L]	67%	WP_008191793.1
Boucher			Pelosinus fermentans JBW45	45%	WP_007954884.1	Pelosinus fermentans JBW45	48%	WP_007954884.1	Stigmatella aurantiaca DW4/3-1	67%	YP_003950451.1
			Gloeobacter violaceus PCC 7421	44%	NP_925775.1	Chlorogloeopsis sp. PCC 9212	48%	WP_016872856.1	Salinispora arenicola	66%	WP_019032757.1
Brusselscobblestones	2	A8-3	Streptomyces viridochromogene	51%	WP_004004070.1	Streptomyces neyagawaensis	58%	AAZ94390.1	Streptomyces viridochromogenes Tue57	70%	WP_004004070.1
Boucher			soraphen B [Sorangium cellulosum]	51%	AAA79984.2	Streptomyces scabiei 87.22	58%	YP_003493865.1	Streptomyces griseus XylebKG-1	69%	WP_003971098.1
			AmbC [Sorangium cellulosum]	50%	ABK32257.1	CppI [Pseudonocardia autotrophica]	57%	ABV83221.1	Streptomyces griseus subsp. griseus NBRC 1	69%	YP_001828290.1
Brusselscobblestones	3	A1	Hyphomicrobium zavarzinii	60%	WP_020084358.1	Hyphomicrobium zavarzinii	62%	WP_020084358.1	[Hyphomicrobium zavarzinii	77%	WP_020084358.1
Boucher			Hoeflea sp. 108	58%	WP_018430828.1	Rhizobium sp. CF080	59%	WP_007753446.1	Rhizobium sp. CF080	75%	WP_007753446.1
			Rhizobium sp. CF080	57%	WP_007753446.1	Labrenzia alexandrii DFL-11	59%	WP_008191204.1	hypothetical protein [Hoeflea sp. 108]	74%	WP_018430828.1
Brusselscobblestones	4	A8-3	soraphen B [Sorangium cellulosum]	51%	AAA79984.2	Streptomyces neyagawaensis	55%	AAZ94390.1	Nocardiopsis sp. CNS639	71%	WP_019606713.1
Boucher			Nocardiopsis sp. CNS639	50%	WP_019606713.1	Streptomyces scabiei 87.22	55%	YP_003493865.1	Streptomyces bingchenggensis BCW-1	69%	YP_004958981.1
			Streptomyces griseus	50%	ACL97692.1	Streptomyces mediolani	55%	ACL97837.1	Streptomyces viridochromogenes Tue57	69%	WP_004004070.1
Brusselscobblestones	5	Mbt	Mycobacterium tusciae JS617	75%	WP_006241084.1	Mycobacterium tusciae JS617	77%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
Boucher			Mycobacterium phlei RIVM601174	62%	WP_003890403.1	Mycobacterium phlei RIVM601174	63%	WP_003890403.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium hassiacum DSM 44199	60%	WP_005628540.	Mycobacterium hassiacum DSM 4419	61%	WP_005628540.1	Mycobacterium hassiacum DSM 44199	80%	WP_005628542.1
Brusselscobblestones	6	Outlier	Chlorogloeopsis sp. PCC 9212	45%	WP_016872856.1	Gloeobacter violaceus PCC 742	49%	NP_925775.1	Stigmatella aurantiaca DW4/3-1	68%	YP_003950451.1
Boucher			Pelosinus fermentans JBW45	45%	WP_007954884.1	Pelosinus fermentans JBW45	49%	WP_007954884.1	CurK [Moorea producens 3L]	67%	WP_008191793.1
			Gloeobacter violaceus PCC 742	44%	NP_925775.1	Chlorogloeopsis sp. PCC 9212	48%	WP_016872856.1	Cystobacter fuscus DSM 2262	67%	WP_002631289.1

Table S2-12												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Brusselscobblestones-	7	A8-1	Streptomyces prunicolor	64%	WP_019065464.1	Streptomyces prunicolor	69%	WP_019065464.1	laidlomycin [Streptomyces sp. CS684]	81%	AFL48525.1
Boucher			Micromonospora sp. ATCC 39149	62%	WP_007071854.1	Streptomyces scabiei 87.22	68%	YP_003493865.1	Streptomyces mediolani	80%	ACL97839.1
			Mycobacterium avium subsp. hominis	61%	BAN31758.1	Streptomyces neyagawaensis	68%	AAZ94390.1	Streptomyces mediolani	80%	ACL97838.1
Brusselscobblestones	8	A8-3	AmbC [Sorangium cellulosum]	52%	ABK32257.1	Streptomyces clavuligerus	54%	WP_009995034.1	laidlomycin [Streptomyces sp. CS684]	75%	AFL48533.1
Boucher			AmbE [Sorangium cellulosum]	52%	ABK32259.1	Streptomyces clavuligerus	54%	WP_003959184.1	Streptomyces cyaneogriseus subsp. Noncyar	74%	BAF85843.1
			JerC [Sorangium cellulosum]	51%	ABK32289.1	AmbC [Sorangium cellulosum]	54%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
Brusselscobblestones	9	A8-3	Streptomyces griseus	51%	ACL97693.1	Streptomyces ornatus	62%	ACL97800.1	Streptomyces bingchenggensis BCW-1]	75%	YP_004958978.1
Boucher			Streptomyces griseus	51%	ACL97775.1	soraphen B [Sorangium cellulosum]	46%	AAA79984.2	Streptomyces cyaneogriseus subsp. Noncyar	74%	BAF85843.1
			Streptomyces griseus	51%	ACL97758.1	Streptomyces himastatinicus ATCC 5	59%	WP_009712634.1	Streptomyces sp. ATexAB-D23	74%	WP_018555451.1
Ottawacobblestones	1	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Dubliner			Streptomyces zinciresistens	57%	WP_007491078.1	Streptomyces sp. FxanaC1	58%	WP_018091595.1	Streptomyces sp. CNS335	77%	WP_018844987.1
			Mycobacterium xenopi RIVM700367	56%	WP_003918930.1	Streptomyces scabiei 87.22	58%	YP_003493865.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
Ottawacobblestones	2	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zinciresistens	57%	WP_007491078.1	Streptomyces violaceusniger Tu 4113	59%	YP_004814091.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces hygroscopicus	59%	CCF23200.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	3	A8-1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Dubliner			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces scabiei 87.22	58%	YP_003493865.1	Streptomyces violaceusniger Tu 4113	77%	YP_004815231.1
			Streptomyces tsukubaensis NRRL1848	56%	WP_006344990.1	Streptomyces griseus	58%	ACL97804.1	Streptomyces aizunensis	77%	AAX98187.1
Ottawacobblestones	4	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
Dubliner			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. CNS335	78%	WP_018844987.1
			Mycobacterium xenopi RIVM700367	57%	WP_003918930.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces violaceusniger Tu 4113	77%	YP_004815231.1
Ottawacobblestones	5	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces hygroscopicus	59%	CCF23200.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces sp. CNS615	59%	WP_017950169.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	6	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	7	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. CNS615	59%	WP_017950169.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. CNS335	78%	WP_018844987.1

Table S2-13											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Ottawacobblestones	8	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces tsukubaensis NRRL1848	57%	WP_006344990.1	Streptomyces neyagawaensis	59%	AAZ94390.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	9	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces violaceusniger Tu 4113]	77%	YP_004815231.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces hygroscopicus subsp. Azalomy	77%	BAH67095.1
Ottawacobblestones	10	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces tsukubaensis NRRL1848	57%	WP_006344990.1	Streptomyces scabiei 87.22	59%	WP_006608270.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	11	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces hygroscopicus	61%	CCF23200.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces sp. FxanaC1	60%	WP_018091595.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	12	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces rapamycinicus NRRL 54	60%	WP_020874116.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces hygroscopicus	60%	CCF23200.1	Streptomyces sp. CNS335	78%	WP_018844987.1
			Streptomyces zincresistens K42	56%	WP_007491077.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Ottawacobblestones	13	A8-1	Streptomyces violaceusniger Tu 4113	59%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
Dubliner			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces violaceusniger Tu 4113	59%	YP_004814091.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces sp. CNS615	59%	WP_017950169.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Ottawacobblestones	14	A8-1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces filamentosus	76%	WP_010070082.1
Dubliner			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	rifamycin [Streptomyces roseosporus NRRL	76%	WP_006123015.1
			Streptomyces zincresistens K42	56%	WP_007491078.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. NRRL 30748	76%	ABC87511.1
Ottawacobblestones	15	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces sp. FxanaC1	58%	WP_018091595.1	Streptomyces sp. CNS335	77%	WP_018844987.1
Ottawacobblestones	16	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
Dubliner			Streptomyces zincresistens K42	57%	WP_007491078.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. CNS335	78%	WP_018844987.1
			Mycobacterium xenopi RIVM700367	57%	WP_003918930.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Ottawastreetdust	1	A5	Hydrogenophaga sp. PBC	51%	WP_009517179.1	Hydrogenophaga sp. PBC	56%	WP_009517179.1	Sorangium cellulosum So0157-2	67%	YP_008154098.1
Rideau			Sorangium cellulosum So0157-2	50%	YP_008149089.1	Sorangium cellulosum So0157-2	53%	YP_008149089.1	Chondromyces crocatus	67%	CAJ46689.1
			MmxC [Cystobacter fuscus]	50%	ABA29781.1	Gloeobacter violaceus PCC 7421	52%	NP_924900.1	Chondromyces crocatus	65%	CAJ46690.1

Table S2-14

Sample	#	AT group	BLASTP Full length	% identity	Accession number	BLASTP AT domain	% identity	Accession number	BLASTP KS domain	% identit	Accession number
Ottawastreetdust	2	A5	Hydrogenophaga sp. PBC	52%	WP_009517179.1	Hydrogenophaga sp. PBC	57%	WP_009517179.1	Sorangium cellulosum So0157-2	67%	YP_008154098.1
Rideau			Sorangium cellulosum So0157-2	51%	YP_008149089.1	Sorangium cellulosum So0157-2	54%	YP_008149089.1	Chondromyces crocatus	67%	CAJ46689.1
			MxaC [Stigmatella aurantiaca]	51%	AAK57187.1	Gloeobacter violaceus PCC 7421	53%	NP_924900.1	Chondromyces crocatus	65%	CAJ46690.1
Ottawastreetdust	3	A8-1	Streptomyces sp. Wigar10	59%	WP_019766151.1	BafAI [Streptomyces lohii]	62%	ADC79616.1	Frankia sp. BCU110501	81%	WP_018504769.1
Rideau			BafAI [Streptomyces lohii]	59%	ADC79616.1	Streptomyces sp. Wigar10	61%	WP_019766151.1	Streptomyces sp. Eco86	87%	ABB88533.1
			Herc [Streptomyces chromofuscus]	58%	AEZ64504.	Streptomyces tsukubaensis NRRL184	60%	WP_006344989.1	Saccharomonospora xinjiangensis XJ-54	84%	WP_006236439.1
Ottawastreetdust	4	A8-1	Streptomyces mobaraensis NBRC 138	56%	WP_004951568.1	Streptomyces flaveolus	56%	ACY06289.1	Streptomyces sp. NRRL 30748	80%	ABC87511.1
Rideau			Streptomyces violaceusniger Tu 4113	56%	YP_004817599.1	Streptomyces flaveolus	56%	ACY06287.1	Streptomyces sp. CNQ766	80%	WP_018837728.1
			NorA' [Streptomyces orinoci]	56%	CAO85896.1	Streptomyces griseus	56%	ACL97749.1	Streptomyces sp. CNS335	80%	WP_018844987.1
Ottawastreetdust	5	A5	Hydrogenophaga sp. PBC	52%	WP_009517179.1	Hydrogenophaga sp. PBC	57%	WP_009517179.1	Sorangium cellulosum So0157-2	67%	YP_008154098.1
Rideau			MxaC [Stigmatella aurantiaca]	51%	AAK57187.1	Sorangium cellulosum So0157-2	53%	YP_008149089.1	Chondromyces crocatus	67%	CAJ46689.1
			MmxC [Cystobacter fuscus]	50%	ABA29781.1	Gloeobacter violaceus PCC 7421]	52%	NP_924900.1	Chondromyces crocatus	65%	CAJ46690.1
Ottawastreetdust	6	A8-1	Streptomyces prunicolor	65%	WP_019065464.1	Streptomyces mobaraensis NBRC 13	70%	WP_004951699.1	Streptomyces sp. ID05-A0024	81%	BAH67561.1
Rideau			Micromonospora sp. ATCC 39149	64%	WP_007071854.1	Streptomyces auratus AGR0001	69%	WP_006608277.1	uncultured bacterium	81%	AFP23856.1
			Streptomyces mobaraensis NBRC 138	63%	WP_004951699.1	Nocardiopsis sp. FU40	69%	AEP40936.1	rifamycin Amycolatopsis mediterranei S699	81%	YP_003762843.1
Ottawastreetdust	7	A8-1	Streptomyces cattleya NRRL 8057	72%	YP_004912779.1	Streptomyces cattleya NRRL 8057	81%	YP_004912779.1	Salinispora pacifica	82%	WP_018223798.1
Rideau			Streptomyces violaceusniger Tu 4113	63%	YP_004817597.1	ChlA1 [Streptomyces antibioticus]	70%	AAZ77693.1	monensin [Streptomyces cinnamonensis]	82%	AAO65796.1
			ChlA1 [Streptomyces antibioticus]	63%	AAZ77693.1	Streptomyces violaceusniger Tu 411	70%	YP_004817597.1	fosB [Streptomyces pulveraceus]	82%	AEC13080.1
Českе.Budéjovice	1	A8-3	Streptomyces viridochromogenes Tu	54%	WP_004004070.1	TgaC [Sorangium cellulosum]	55%	ADH04641.1	Streptomyces sp. Amel2xE9	73%	WP_019985563.1
cobblestones- Česka			soraphen polyketide synthase B [Sora	53%	AAA79984.2	Haliangium ochraceum DSM 14365	55%	YP_003267379.1	Streptomyces flaveolus	73%	ACY06286.1
			TgaC [Sorangium cellulosum]	52%	ADH04641.1	soraphen B [Sorangium cellulosum]	54%	AAA79984.2	Streptomyces sp. ATexAB-D23	73%	WP_018554535.1
Česke.Budéjovice	2	A8-1	Streptomyces prunicolor	64%	WP_019065464.1	Streptomyces himastatinicus ATCC 5	66%	WP_009715048.1	Streptomyces ambofaciens ATCC 23877	81%	CAJ88186.1
cobblestones- Česka			Micromonospora sp. ATCC 39149	62%	WP_007071854.1	Streptomyces bingchenggensis BCW	65%	YP_004959792.1	Streptomyces hygroscopicus	80%	AAQ20778.1
			Saccharomonospora xinjiangensis XJ-	60%	WP_006236889.1	Streptomyces bingchenggensis BCW	65%	YP_004959796.1	Streptomyces rapamycinicus NRRL 5491	80%	WP_020873897.1
Česke.Budéjovice	3	A8-3	Streptomyces viridochromogenes Tu	54%	WP_004004070.1	TgaC [Sorangium cellulosum]	55%	ADH04641.1	Streptomyces sp. Amel2xE9	73%	WP_019985563.1
cobblestones- Česka			TgaA [Sorangium cellulosum]	53%	ADH04639.1	Haliangium ochraceum DSM 14365	55%	YP_003267379.1	Streptomyces flaveolus	73%	ACY06286.1
			soraphen B [Sorangium cellulosum]	53%	AAA79984.2	soraphen B [Sorangium cellulosum]	54%	AAA79984.2	Streptomyces sp. ATexAB-D23	73%	WP_018554535.1
Česke.Budéjovice	4	A8-1	Streptomyces violaceusniger Tu 4113	52%	YP_004814096.1	Streptomyces scabrisporus	54%	WP_020556935.1	Saccharopolyspora erythraea	72%	WP_009951540.1
cobblestones- Česka			Streptomyces bingchenggensis BCW-	52%	YP_004960235.1	Streptomyces violaceusniger Tu 411	54%	YP_004814097.1	Streptomyces griseus NBRC 13350]	71%	YP_001827692.1
			Streptomyces violaceusniger Tu 4113	52%	YP_004814093.1	Streptomyces auratus AGR0001	54%	WP_006608270.1	Saccharopolyspora erythraea NRRL 2338	71%	YP_001106332.

Table S2-15											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
České.Budéjovice	5	A8-1	Streptomyces prunicolor	63%	WP_019065464.1	Streptomyces prunicolor	68%	WP_019065464.1	laidlomycin [Streptomyces sp. CS684]	81%	AFL48525.1
cobblestones- Česka			Micromonospora sp. ATCC 39149	62%	WP_007071854.1	Streptomyces scabiei 87.22	68%	YP_003493865.1	Streptomyces mediolani	80%	ACL97839.1
			erythronolide [Mycobacterium avium	61%	YP_881643.1	Streptomyces neyagawaensis	67%	AAZ94390.1	Streptomyces mediolani	80%	ACL97838.1
České.Budéjovice	6	A8-1	Poor chromatogram quality.			Streptomyces himastatinicus ATCC 5	63%	WP_009715048.1	Poor chromatogram quality.		
cobblestones- Česka						Mycobacterium avium	63%	WP_019306431.1			
						Mycobacterium avium	63%	WP_019699750.1			
České.Budéjovice	7	A8-1	Poor chromatogram quality.			Streptomyces himastatinicus ATCC 5	66%	WP_009715048.1	Poor chromatogram quality.		
cobblestones- Česka						Streptomyces bingchenggensis BCW	65%	YP_004959792.1			
						Streptomyces bingchenggensis BCW	65%	YP_004959796.1			
České.Budéjovice	8	A8-1	Poor chromatogram quality.			Streptomyces bingchenggensis BCW	65%	YP_004959792.1	Poor chromatogram quality.		
cobblestones- Česka						Streptomyces bingchenggensis BCW	65%	YP_004959796.1			
						Streptomyces auratus AGR0001	64%	WP_006608277.1			
České.Budéjovice	9	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	Mycobacterium tusciae JS617	82%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
cobblestones- Česka			Mycobacterium hassiacum DSM 4419	65%	WP_005628540.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium phlei RIVM601174	84%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium hassiacum DSM 4419	66%	WP_005628540.1	Mycobacterium hassiacum DSM 44199	77%	WP_005628542.1
České.Budéjovice	10	Mbt	Mycobacterium tusciae JS617	78%	WP_006241084.1	Mycobacterium tusciae JS617	81%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
cobblestones- Česka			Mycobacterium hassiacum DSM 4419	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	68%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium phlei RIVM601174	68%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
České.Budéjovice	1	A8	Streptomyces viridochromogenes Tue	53%	WP_004004070.1	TgaC [Sorangium cellulosum]	56%	ADH04641.1	Frankia sp. BMG5.12	68%	WP_018639361.1
streetdust			TgaA [Sorangium cellulosum]	52%	ADH04639.1	TgaA [Sorangium cellulosum]	55%	ADH04639.1	Streptomyces flaveolus	68%	ACY06286.1
Koh-i-Noor			TgaC [Sorangium cellulosum]	52%	ADH04641.1	AmbE [Sorangium cellulosum]	55%	ABK32259.1	rifamycin [Salinispora arenicola]	68%	WP_020217874.1
České.Budéjovice	2	A5	PksE [Paenibacillus mucilaginosus 30	46%	YP_005313855.1	PksE [Paenibacillus mucilaginosus 30	51%	YP_004642885.1	Stanieria cyanosphaera PCC 7437	63%	YP_007131667.1
streetdust			PksE [Paenibacillus mucilaginosus 30	46%	YP_004642885.1	JamP [Paenibacillus mucilaginosus K	51%	YP_006190789.1	Streptomyces albus	62%	CCD31890.1
Koh-i-Noor			JamP [Paenibacillus mucilaginosus K	46%	YP_006190789.1	PksE [Paenibacillus mucilaginosus 30	51%	YP_005313855.1	PksE [Bacillus sonorensis L12]	62%	WP_006640280.1
České.Budéjovice	3	Mbt	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	90%	WP_006241085.1
streetdust			Mycobacterium hassiacum DSM 4419	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	84%	WP_003890404.1
Koh-i-Noor			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium phlei RIVM601174	67%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	77%	WP_005628542.1
České.Budéjovice	4	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	93%	WP_006241085.1
streetdust			Mycobacterium hassiacum DSM 4419	65%	WP_005628540.1	Mycobacterium phlei RIVM601174	67%	WP_003890403.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
Koh-i-Noor			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium hassiacum DSM 44199	77%	WP_005628542.

Table S2-16											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Budapestpavement	1	A8-1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
TerezKorut			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces sp. CNS615	58%	WP_017950169.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
			Streptomyces zinciresistens K42	56%	WP_007491078.1	Streptomyces scabiei 87.22	58%	YP_003493865.1	Streptomyces sp. CNS335]	77%	WP_018844987.1
Budapestpavement	2	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces scabiei 87.22	60%	YP_003493865.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
TerezKorut			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces hygroscopicus	60%	CCF23200.1	Streptomyces sp. CNS335	77%	WP_018844987.1
Budapestpavement	3	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces scabiei 87.22	58%	YP_003493865.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
TerezKorut			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	58%	WP_006608270.1	Streptomyces sp. CNS335	77%	WP_018844987.1
			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces violaceusniger Tu 411	58%	YP_004814091.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
Budapestpavement	4	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
TerezKorut			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces zinciresistens K42	57%	WP_007491078.1	uncultured Streptomyces sp.	58%	ADJ57302.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Budapestpavement	5	A8-1	Streptomyces violaceusniger Tu 4113	59%	YP_004817599.1	Streptomyces sp. FxanaC1	60%	WP_018091595.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
TerezKorut			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces scabiei 87.22	60%	YP_003493865.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020871223.1	Streptomyces sp. CNS615	60%	WP_017950169.1	Streptomyces sp. CNS335]	77%	WP_018844987.1
Budapestpavement	6	A8-1	[Streptomyces violaceusniger Tu 411	58%	YP_004817599.1	AVES 4 [Streptomyces mobaraensis N	59%	WP_004951568.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
TerezKorut			Streptomyces zinciresistens K42	57%	WP_007491078.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. CNQ766]	78%	WP_018837728.1
			Streptomyces mobaraensis NBRC 138	56%	WP_004951568.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Budapestpavement	7	A8-1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces violaceusniger Tu 4113]	75%	YP_004815231.1
TerezKorut			Streptomyces zinciresistens K42	56%	WP_007491078.1	Saccharomonospora paurometabolic	58%	WP_007026651.1	Streptomyces sp. NRRL 30748	75%	ABC87511.1
			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces neyagawaensis	58%	AAZ94390.1	Streptomyces sp. CNQ766	75%	WP_018837728.1
Budapestpavement	8	A8-1	Streptomyces violaceusniger Tu 4113	56%	YP_004817599.1	Streptomyces scabiei 87.22	59%	YP_003493865.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
TerezKorut			AVES 4 [Streptomyces mobaraensis N	55%	WP_004951568.1	Streptomyces auratus AGR0001	59%	WP_006608270.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			[Streptomyces sp. CNS615]	55%	WP_017950169.1	Streptomyces sp. CNS615	58%	WP_017950169.1	Streptomyces sp. CNS335	78%	WP_018844987.
Budapestpavement	9	A8-1	Streptomyces violaceusniger Tu 4113	55%	YP_004817599.1	Streptomyces rapamycinicus NRRL 5	54%	WP_020874116.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
TerezKorut			Streptomyces rapamycinicus NRRL 54	54%	WP_020874122.1	Streptomyces tsukubaensis NRRL184	54%	WP_006344988.1	Streptomyces sp. CNQ766	77%	WP_018837728.1
			NorA' [Streptomyces orinoci]	54%	CAO85896.1	Amycolatopsis vancoresmycina DSM	54%	WP_004559079.1	Streptomyces sp. CNS335	77%	WP_018844987.1
Budapestpavement	10	A8-1	Streptomyces violaceusniger Tu 4113	59%	YP_004817599.1	Streptomyces sp. FxanaC1	60%	WP_018091595.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
TerezKorut			Streptomyces rapamycinicus NRRL 54	57%	WP_020871223.1	Streptomyces scabiei 87.22	60%	YP_003493865.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Streptomyces sp. CNS615	57%	WP_017950169.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. CNS335	78%	WP_018844987.1

Table S2-17											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Budapestpavement	11	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	Streptomyces sp. FxanaC1	60%	WP_018091595.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
TerezKorut			Streptomyces zincirasistens K42	57%	WP_007491078.1	Streptomyces hygroscopicus	60%	CCF23200.1	Streptomyces sp. CNS335	78%	WP_018844987.1
			Streptomyces rapamycinicus NRRL 54	57%	WP_020874122.1	Streptomyces auratus AGR0001	60%	WP_006608270.1	Streptomyces sp. NRRL 30748	77%	ABC87511.1
Budapestpavement	12	A8-1	Streptomyces violaceusniger Tu 4113	58%	YP_004817599.1	uncultured Streptomyces sp.	58%	ADJ57302.1	Streptomyces violaceusniger Tu 4113	76%	YP_004815231.1
TerezKorut			Streptomyces zincirasistens K42	57%	WP_007491078.1	Streptomyces sp. AH7-2	57%	AFD04715.1	Streptomyces sp. CNQ766	76%	WP_018837728.1
			Streptomyces rapamycinicus NRRL 54	56%	WP_020874122.1	Streptomyces sp. FxanaC1	59%	WP_018091595.1	Streptomyces sp. CNS335	76%	WP_018844987.1
Budapestpavement	13	A8-1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	Mycobacterium marinum M	57%	YP_001851317.1	Streptomyces sp. NRRL 30748	79%	ABC87511.1
TerezKorut			Streptomyces zincirasistens K42	56%	WP_007491078.1	Streptomyces sp. FxanaC1	56%	WP_018091595.1	Streptomyces sp. CNQ766	78%	WP_018837728.1
			Mycobacterium xenopi RIVM700367	55%	WP_003918930.1	Streptomyces auratus AGR0001	56%	WP_006608270.1	Streptomyces sp. CNS335	78%	WP_018844987.1
Pariscobblestones	1	A8	Poor chromatogram quality.			Streptomyces avermitilis MA-4680	54%	NP_822118.1	Poor chromatogram quality.		
RiveGauche						soraphen B [Sorangium cellulosum]	52%	AAA79984.2			
						BFAS1 [Streptomyces tsukubaensis I	53%	WP_006344992.1			
Pariscobblestones	2	A8	soraphen B [Sorangium cellulosum]	53%	AAA79984.2	TgaC [Sorangium cellulosum]	56%	ADH04641.1	Streptomyces sp. Amel2xE9	73%	WP_019985563.1
RiveGauche			Streptomyces viridochromogenes Tue	53%	WP_004004070.1	Streptomyces bingchenggensis BCW	55%	YP_004967444.1	Streptomyces sp. ATexAB-D23	73%	WP_018554535.1
			TgaC [Sorangium cellulosum]	51%	ADH04641.1	soraphen B [Sorangium cellulosum]	55%	AAA79984.2	Streptomyces sp. E14	71%	WP_009191668.1
Pariscobblestones	3	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium hassiacum DSM 44199	65%	WP_005628540.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	4	Mbt	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium phlei RIVM601174	67%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	5	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	Mycobacterium tusciae JS617	82%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	6	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	polyketide synthase [Mycobacterium	82%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	polyketide synthase [Mycobacterium	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	polyketide synthase [Mycobacterium	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	7	Mbt	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium hassiacum DSM 44199	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1

Table S2-18											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Pariscobblestones	8	Mbt	Mycobacterium tusciae JS617	77%	WP_006241084.1	polyketMycobacterium tusciae JS617	81%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	68%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403.1	Mycobacterium phlei RIVM601174	68%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	9	Mbt	Mycobacterium tusciae JS617	79%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	10	Mbt	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
RiveGauche			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	64%	WP_003890403.1	Mycobacterium phlei RIVM601174	66%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
Pariscobblestones	1	A8-1	Streptomyces sp. CNY243	55%	WP_018853338.1	Streptomyces himastatinicus ATCC 5	58%	WP_009719940.1	QmnA1 [Amycolatopsis orientalis]	76%	AFI57005.1
RiveGauche			Streptomyces sp. CNQ766	55%	WP_018840958.1	Streptomyces violaceusniger Tu 4113	57%	YP_004817599.1	MycD4 [Streptomyces flaveolus]	78%	AFG19422.1
			Streptomyces sp. CNS335	55%	WP_018842118.1	Streptomyces rapamycinicus NRRL 54	57%	WP_020874116.1	Amycolatopsis vancoresmycina DSM 44592	77%	WP_003070830.1
Pariscobblestones	2	A5	Saccharothrix espanaensis DSM 44229	42%	YP_007037544.1	Hydrogenophaga sp. PBC	46%	WP_009517179.1	uncultured bacterium B7P37metaSE	65%	ACV30045.1
Cafe			wcbR [Streptomyces albulus CCRC 118	42%	WP_016570418.1	Nocardiopsis potens	45%	WP_017594183.1	actinomycete my02	65%	AEE69401.1
			Streptomyces albulus CCRC 11814	42%	EPY93276.1	Streptomyces ghanaensis ATCC 1467	45%	WP_004982014.1	Streptomyces griseus subsp. griseus NBRC 1	65%	YP_001828290.1
Pariscobblestones	3	A8-1	Streptomyces sp. C	56%	WP_007269157.1	Streptomyces sp. C	62%	WP_007269157.1	Streptomyces himastatinicus ATCC 53653	79%	WP_009717837.1
Cafe			Streptomyces violaceusniger Tu 4113	56%	YP_004817599.1	Streptomyces rapamycinicus NRRL 54	59%	WP_020874122.1	Streptomyces violaceusniger Tu 4113	79%	YP_004815656.1
			Streptomyces violaceusniger Tu 4113	55%	YP_004814091.1	Streptomyces rapamycinicus NRRL 54	59%	WP_020874117.1	GdmAII [Streptomyces hygroscopicus]	79%	AAO06917.1
Pariscobblestones	4	A8-3	AmbC [Sorangium cellulosum]	61%	ABK32257.1	AmbC [Sorangium cellulosum]	63%	ABK32257.1	Streptomyces acidiscabies	74%	WP_010356044.1
Cafe			JerC [Sorangium cellulosum]	60%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. CcalMP-8W	73%	WP_018491143.1
			AmbE [Sorangium cellulosum]	59%	ABK32259.1	JerC [Sorangium cellulosum]	62%	ABK32289.1	Streptomyces sp. Wigar10	72%	WP_019765969.1
Pariscobblestones	5	Mbt	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	83%	WP_006241084.1	Mycobacterium tusciae JS617	90%	WP_006241085.1
Cafe			Mycobacterium hassiacum DSM 44199	65%	WP_005628540.1	Mycobacterium phlei RIVM601174	67%	WP_003890403.1	Mycobacterium phlei RIVM601174	84%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403	Mycobacterium hassiacum DSM 4419	67%	WP_005628540.1	Mycobacterium hassiacum DSM 44199	77%	WP_005628542.1
Pariscobblestones	6	Mbt	Mycobacterium tusciae JS617	77%	WP_006241084.1	Mycobacterium tusciae JS617	80%	WP_006241084.1	Mycobacterium tusciae JS617	92%	WP_006241085.1
Cafe			Mycobacterium hassiacum DSM 44199	66%	WP_005628540.1	Mycobacterium hassiacum DSM 4419	68%	WP_005628540.1	Mycobacterium phlei RIVM601174	86%	WP_003890404.1
			Mycobacterium phlei RIVM601174	65%	WP_003890403.	Mycobacterium phlei RIVM601174	67%	WP_003890403.1	Mycobacterium hassiacum DSM 44199	78%	WP_005628542.1
FaisalabadStreetdust	1	A8-1	rifamycin [Salinispora pacifica]	73%	WP_018826046.1	Salinispora arenicola	78%	WP_020218631.1	rifamycin polyketide synthase [Salinispora p	81%	WP_018826046.1
Clock			Salinispora arenicola	73%	WP_020218631.1	Salinispora arenicola	78%	WP_020609567.1	Salinispora arenicola	81%	WP_018909491.1
			Salinispora arenicola	73%	WP_018588929.	Salinispora arenicola	78%	WP_019033993.1	Salinispora arenicola	81%	WP_018588929.1

Table S2-19											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
FaisalabadStreetdust	2	A8-1	rifamycin [Salinispora pacifica]	73%	WP_018826046.1	Salinispora arenicola	78%	WP_020218631.1	rifamycin [Salinispora pacifica]	81%	WP_018826046.1
Clock			Salinispora arenicola	73%	WP_020218631.1	Salinispora arenicola	78%	WP_020609567.1	Salinispora arenicola	81%	WP_018909491.1
			Salinispora arenicola	73%	WP_018588929.1	thioester reductase [Salinispora arer	78%	WP_019033993.1	Salinispora arenicola	81%	WP_018588929.1
FaisalabadStreetdust	3	A8-1	Streptomyces cattleya NRRL 8057	73%	YP_004912779.1	Streptomyces cattleya NRRL 8057	82%	YP_004912779.1	Salinispora pacifica	85%	WP_018223798.1
Clock			Streptomyces violaceusniger Tu 4113	64%	YP_004817597.1	ChlA1 [Streptomyces antibioticus]	71%	AAZ77693.1	fosB [Streptomyces pulveraceus]	85%	AEC13080.1
			ChlA1 [Streptomyces antibioticus]	64%	AAZ77693.1	Streptomyces violaceusniger Tu 411	71%	YP_004817597.1	fosA [Streptomyces pulveraceus]	85%	AEC13079.1
FaisalabadStreetdust	4	A8-1	Streptomyces cattleya NRRL 8057	73%	YP_004912779.1	Streptomyces cattleya NRRL 8057	82%	YP_004912779.1	Salinispora pacifica	84%	WP_018223798.1
Clock			Streptomyces violaceusniger Tu 4113	64%	YP_004817597.1	ChlA1 [Streptomyces antibioticus]	72%	AAZ77693.1	fosB [Streptomyces pulveraceus]	84%	AEC13080.1
			ChlA1 [Streptomyces antibioticus]	64%	AAZ77693.1	Streptomyces violaceusniger Tu 411	71%	YP_004817597.1	fosA [Streptomyces pulveraceus]	84%	AEC13079.1
FaisalabadStreetdust	5	A8-1	CppB [Pseudonocardia autotrophica]	61%	ABV83229.1	CppB [Pseudonocardia autotrophica]	68%	ABV83229.1	Streptomyces auratus AGR0001	77%	WP_006608272.1
Clock			Streptomyces mobaraensis NBRC 138	60%	WP_004951699.1	Pseudonocardia sp. P1	68%	WP_010235728.1	FscC, partial [Streptomyces albulus]	77%	WP_016572880.1
			Amycolatopsis benzoatilytica	60%	WP_020659290.1	NysB [Streptomyces noursei ATCC 11	67%	AAF71775.1	Streptomyces albulus	77%	WP_020930833.1
FaisalabadStreetdust	6	A8-1	Streptomyces rapamycinicus NRRL 54	60%	WP_020865107.1	Streptomyces rapamycinicus NRRL 54	66%	WP_020865107.1	Streptomyces flaveolus	75%	ACY06287.1
Clock			Nocardiopsis sp. FU40	60%	AEP40934.1	Amycolatopsis vancoresmycina DSM	65%	WP_004560200.1	Mycobacterium intracellulare	75%	WP_009956415.1
			Streptomyces mobaraensis NBRC 138	58%	WP_004951568.1	Streptomyces sp. PsTaAH-124	65%	WP_018570100.1	erythronolide [Mycobacterium intracellular	75%	YP_005337970.1
FaisalabadStreetdust	7	A8-1	Streptomyces rapamycinicus NRRL 54	64%	WP_020874116.1	Streptomyces rapamycinicus NRRL 54	70%	WP_020874116.1	Streptomyces sp. S4	86%	WP_010638160.1
Clock			Streptomyces sp. DSM 21069	64%	ACO94471.1	BFAS3 [Kitasatospora setae KM-6054	69%	YP_004909057.1	Salinispora pacifica	85%	WP_019872731.1
			polyketide synthase [Streptomyces s	64%	ABC87509.1	BFAS2 [Kitasatospora setae KM-6054	69%	YP_004909056.1	Salinispora pacifica	85%	WP_020113574.1
FaisalabadStreetdust	8	A8-1	Streptomyces rapamycinicus NRRL 54	60%	WP_020874117.1	Streptomyces rapamycinicus NRRL 54	71%	WP_020874116.1	Streptomyces rapamycinicus NRRL 5491	73%	WP_020869281.1
Clock			Streptomyces rapamycinicus NRRL 54	59%	WP_020866324.1	Streptomyces violaceusniger Tu 411	69%	YP_004814091.1	MerC [Streptomyces violaceusniger]	73%	ABJ97439.1
			Streptomyces rapamycinicus NRRL 54	59%	WP_020865106.1	Streptomyces rapamycinicus NRRL 54	69%	WP_020874117.1	Streptomyces nodosus subsp. asukaensis	72%	BAH67150.1
FaisalabadStreetdust	9	A8-1	Streptomyces cattleya NRRL 8057	73%	YP_004912779.1	Streptomyces cattleya NRRL 8057	82%	YP_004912779.1	Salinispora pacifica	84%	WP_018223798.1
Clock			Streptomyces violaceusniger Tu 4113	64%	YP_004817597.1	ChlA1 [Streptomyces antibioticus]	71%	AAZ77693.1	fosB [Streptomyces pulveraceus]	84%	AEC13080.1
			ChlA1 [Streptomyces antibioticus]	64%	AAZ77693.1	Streptomyces violaceusniger Tu 411	71%	YP_004817597.1	fosA [Streptomyces pulveraceus]	84%	AEC13079.1
FaisalabadStreetdust	10	A8-1	Streptomyces cattleya NRRL 8057	73%	YP_004912779.1	Streptomyces cattleya NRRL 8057	83%	YP_004912779.1	Salinispora pacifica	85%	WP_018223798.1
Clock			ChlA1 [Streptomyces antibioticus]	64%	AAZ77693.1	ChlA1 [Streptomyces antibioticus]	72%	AAZ77693.1	fosB [Streptomyces pulveraceus]	85%	AEC13080.1
			Streptomyces violaceusniger Tu 4113	64%	YP_004817597.1	Streptomyces violaceusniger Tu 411	72%	YP_004817597.1	fosA [Streptomyces pulveraceus]	85%	AEC13079.1
FaisalabadStreetdust	11	A8-1	Rhodococcus sp. P14	98%	WP_010595362.1	Rhodococcus sp. P14	98%	WP_010595362.1	Rhodococcus sp. P14	99%	WP_010595362.1
Clock			Rhodococcus ruber	98%	WP_017681581.1	Rhodococcus ruber	98%	WP_017681581.1	Rhodococcus ruber	99%	WP_017681581.1
			Rhodococcus sp. EsD8	94%	WP_006936521.1	Rhodococcus sp. EsD8	95%	WP_006936521.1	Rhodococcus sp. EsD8	95%	WP_006936521.1

Table S2-20

Sample	#	AT group	BLASTP Full length	% identity	Accession number	BLASTP AT domain	% identity	Accession number	BLASTP KS domain	% identit	Accession number
FaisalabadStreetdust	12	A8-1	Nocardiopsis halotolerans	89%	WP_017570020.1	Nocardiopsis halotolerans	88%	WP_017570020.1	Nocardiopsis halotolerans	90%	WP_017570020.1
Clock			Nocardiopsis lucentensis	76%	WP_017599747.1	Nocardiopsis lucentensis	77%	WP_017599747.1	Nocardiopsis xinjiangensis	80%	WP_017609638.1
			Nocardiopsis valliformis	72%	WP_017581971.1	Nocardiopsis valliformis	70%	WP_017581971.1	Nocardiopsis lucentensis	79%	WP_017599747.1
CzVermicompost	1	A6	MmxB [Cystobacter fuscus]	48%	ABA29782.1	MmxB [Cystobacter fuscus]	52%	ABA29782.1	Mycobacterium vaccae ATCC 25954	69%	WP_003933951.1
			Sorangium cellulosum So0157-2	46%	YP_008151400.1	Melittangium lichenicola	51%	BAG69021.1	Mycobacterium abscessus	65%	WP_016889012.1
			Myxococcus stipitatus DSM 14675	47%	YP_007361305.1	Myxococcus stipitatus	50%	BAG69001.1	Mycobacterium smegmatis str. MC2 155	65%	YP_884821.1
CzVermicompost	2	A5	uncultured bacterium B7P37metaSE	42%	ACV30045.1	Streptomyces viridosporus	45%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			Methylocystis rosea	42%	WP_018405832.1	Streptomyces ghanaensis ATCC 1467	45%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			Methylocystis sp. SC2	42%	YP_006591528.1	Pseudomonas sp. S13.1.2	43%	WP_019472290.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	3	A5	uncultured bacterium B7P37metaSE	42%	ACV30045.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	66%	ACV30045.1
			Methylocystis rosea	42%	WP_018405832.1	Streptomyces ghanaensis ATCC 1467	45%	WP_004982014.1	Sorangium cellulosum	64%	CCE88377.1
			Methylocystis sp. SC2	42%	YP_006591528.1	Streptomyces violaceusniger Tu 411	43%	YP_004813236.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	4	A6	MmxB [Cystobacter fuscus]	48%	ABA29782.1	MmxB [Cystobacter fuscus]	52%	ABA29782.1	Mycobacterium vaccae ATCC 25954	69%	WP_003933951.1
			Sorangium cellulosum So0157-2	47%	YP_008151400.1	Melittangium lichenicola	52%	BAG69021.1	Mycobacterium abscessus	65%	WP_016889012.1
			Myxococcus stipitatus DSM 14675	47%	YP_007361305.1	Myxococcus stipitatus	50%	BAG69001.1	Mycobacterium smegmatis str. MC2 155	65%	YP_884821.1
CzVermicompost	5	A5	uncultured bacterium B7P37metaSE	43%	ACV30045.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			Pseudomonas putida	43%	WP_019437274.1	Streptomyces ghanaensis ATCC 1467	46%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			seudomonas putida BIRD-1	42%	YP_005931616.1	Pseudomonas sp. S13.1.2	45%	WP_019472290.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	6	A8-1	Streptomyces sp. CNS335	55%	WP_018842119.1	Streptomyces sp. C	57%	WP_007269143.1	Streptomyces sp. CS	83%	ADM46356.1
			Streptomyces sp. CNY243	55%	WP_018853339.1	Streptomyces sp. C	57%	WP_007269142.1	uncultured bacterium	80%	AFP23856.1
			Streptomyces sp. CNQ766	55%	WP_018840957.1	Streptomyces violaceusniger Tu 411	57%	YP_004817602.1	Amycolatopsis vancoresmycina DSM 44592	80%	WP_003070830.1
CzVermicompost	7	A8-1	Streptomyces sp. CNS335	55%	WP_018842119.1	Streptomyces sp. C	57%	WP_007269143.1	polyketide synthase [Streptomyces sp. CS]	83%	ADM46356.1
			Streptomyces sp. CNY243	55%	WP_018853339.1	lasalocid [Streptomyces sp. C]	57%	WP_007269142.1	uncultured bacterium	80%	AFP23856.1
			Streptomyces sp. CNQ766	55%	WP_018840957.1	Streptomyces violaceusniger Tu 411	57%	YP_004817602.1	Amycolatopsis vancoresmycina DSM 44592	80%	WP_003070830.1
CzVermicompost	8	A5	uncultured bacterium B7P37metaSE	42%	ACV30045.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			Gloeobacter violaceus PCC 7421	42%	NP_925775.1	Streptomyces ghanaensis ATCC 1467	45%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			Methylocystis rosea	42%	WP_018405832.1	Pseudomonas sp. S13.1.2	43%	WP_019472290.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	9	A6	MmxB [Cystobacter fuscus]	49%	ABA29782.1	MmxB [Cystobacter fuscus]	53%	ABA29782.1	Mycobacterium vaccae ATCC 25954	69%	WP_003933951.1
			Myxococcus stipitatus DSM 14675	48%	YP_007361305.1	Melittangium lichenicola	53%	BAG69021.1	Mycobacterium abscessus	65%	WP_016889012.1
			[Sorangium cellulosum So0157-2]	47%	YP_008151400.1	Myxococcus stipitatus	50%	BAG69001.1	Mycobacterium smegmatis str. MC2 155	65%	YP_884821.1

Table S2-21											
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
CzVermicompost	10	A8-1	Streptomyces sp. CNS335	52%	WP_018842119.1	Streptomyces scabiei 87.22	53%	YP_003493863.1	Streptomyces sp. CS	82%	ADM46356.1
			Streptomyces sp. CNQ766	52%	WP_018840958.1	Streptomyces sp. C	52%	WP_007269143.1	Streptomyces sp. CS	80%	ADM46359.1
			Streptomyces sp. CNS335	52%	WP_018842118.1	Streptomyces cattleya NRRL 8057	52%	YP_004920763.1	uncultured bacterium	78%	AFP23856.1
CzVermicompost	11	A6	MmxB [Cystobacter fuscus]	47%	ABA29782.1	MmxB [Cystobacter fuscus]	51%	ABA29782.1	Mycobacterium vaccae ATCC 25954	69%	WP_003933951.1
			Chondromyces crocatus	47%	CAQ18832.1	Melittangium lichenicola	50%	BAG69021.1	Mycobacterium abscessus	65%	WP_016889012.1
			MtaE protein [Stigmatella aurantiaca	47%	YP_003953630.1	Sorangium cellulosum	50%	CCE88377.1	Mycobacterium smegmatis MKD8	65%	WP_003891728.1
CzVermicompost	12	A5	uncultured bacterium B7P37metaSE	42%	ACV30045.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			Pseudomonas putida	42%	WP_019437274.1	Streptomyces ghanaensis ATCC 1467	46%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			Pseudomonas putida BIRD-1	42%	YP_005931616.1	Pseudomonas sp. S13.1.2	45%	WP_019472290.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	13	A6	MmxB [Cystobacter fuscus]	48%	ABA29782.1	MmxB [Cystobacter fuscus]	52%	ABA29782.1	Mycobacterium vaccae ATCC 25954	68%	WP_003933951.1
			Sorangium cellulosum So0157-2	47%	YP_008151400.1	Melittangium lichenicola	52%	BAG69021.1	Mycobacterium colombiense CECT 3035	64%	WP_007770214.1
			Myxococcus stipitatus DSM 14675	47%	YP_007361305.1	Myxococcus stipitatus	50%	BAG69001.1	Mycobacterium abscessus	64%	WP_016889012.1
CzVermicompost	14	A5	uncultured bacterium B7P37metaSE	42%	ACV30045.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			Gloeobacter violaceus PCC 7421	42%	NP_925775.1	Streptomyces ghanaensis ATCC 1467	46%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			Singulisphaera acidiphila DSM 18658	42%	YP_007202109.1	Pseudomonas sp. S13.1.2	44%	WP_019472290.1	Colletotrichum orbiculare MAFF 240422	63%	ENH86451.1
CzVermicompost	15	A5	Streptomyces viridosporus	43%	WP_016824224.1	Streptomyces viridosporus	46%	WP_016824224.1	uncultured bacterium B7P37metaSE	67%	ACV30045.1
			uncultured bacterium B7P37metaSE	43%	ACV30045.1	Streptomyces ghanaensis ATCC 1467	45%	WP_004982014.1	Sorangium cellulosum	65%	CCE88377.1
			Methylocystis rosea	42%	WP_018405832.1	Pseudomonas sp. S13.1.2	43%	WP_019472290.1	Streptomyces pactum	64%	BAF92601.1
Rozowormgut-Martiod	1	A8-1	Streptomyces violaceusniger Tu 4113	51%	YP_004815231.1	Streptomyces violaceusniger Tu 411	56%	YP_004815232.1	Streptomyces sp. ID05-A0841	85%	BAH68302.1
			Streptomyces griseus XylebKG-1	50%	WP_003970387.1	MerA [Streptomyces violaceusniger]	56%	ABJ97437.1	Mycobacterium colombiense CECT 3035	85%	WP_007770899.1
			MerC [Streptomyces violaceusniger]	50%	ABJ97439.1	Streptomyces violaceusniger Tu 411	56%	YP_004815231.1	ChlA4 [Streptomyces antibioticus	84%	AAZ77697.1
Rozowormgut	2	A8-1	Streptomyces violaceusniger Tu 4113	55%	YP_004816248.1	Streptomyces scabiei 87.22	60%	YP_003493864.1	Streptomyces violaceusniger Tu 4113	80%	YP_004816258.1
Martiodrillus			Type I PKS [Streptomyces hygroscopi	55%	CCF23200.1	Streptomyces neyagawaensis	60%	AAZ94389.1	Streptomyces sp. CNS335	80%	WP_018844987.1
			PieA6 [Streptomyces piomogenus]	55%	AEZ54379.1	Kitasatospora setae KM-6054	60%	YP_004909057.1	Streptomyces sp. CNQ766	80%	WP_018837728.1
Rozowormgut	3	A8-1	Streptomyces nanchangensis	54%	ADC45535.1	Streptomyces nanchangensis	56%	ADC45535.1	AmbD [Sorangium cellulosum]	73%	ABK32258.1
Martiodrillus			soraphen B [Sorangium cellulosum]	53%	AAA79984.2	[Streptomyces bingchenggensis BCW	55%	YP_004958981.1	Streptomyces acidiscabies	73%	WP_010358757.1
			Streptomyces sp. CNQ766	52%	WP_018840958.1	Streptomyces scabiei 87.22	55%	YP_003493865.1	Saccharomonospora azurea SZMC 14600	73%	WP_005451977.
Rozowormgut	4	A8-1	Streptomyces violaceusniger Tu 4113	51%	YP_004815231.1	Streptomyces violaceusniger Tu 411	56%	YP_004815232.1	Streptomyces sp. ID05-A0841	85%	BAH68302.1
Martiodrillus			Streptomyces griseus XylebKG-1	50%	WP_003970387.1	MerA [Streptomyces violaceusniger]	56%	ABJ97437.1	Mycobacterium colombiense CECT 3035	85%	WP_007770899.1
			MerC [Streptomyces violaceusniger]	50%	ABJ97439.1	Streptomyces violaceusniger Tu 411	56%	YP_004815231.1	ChlA4 [Streptomyces antibioticus]	84%	AAZ77697.1

Table S2-22												
		AT		%	Accession		%	Accession		%	Accession
Sample	#	group	BLASTP Full length	identity	number	BLASTP AT domain	identity	number	BLASTP KS domain	identit	number
Rozowormgut	5	A8-1	Poor chromatogram quality.			Streptomyces sp. C	56%	WP_007269166.1	Poor chromatogram quality.		
Martiodrillus						Streptomyces violaceusniger Tu 4113	55%	YP_004815232.1			
						MerA [Streptomyces violaceusniger]	55%	ABJ97437.1			
Rozowormgut	6	A8-1	Streptomyces rapamycinicus NRRL 54	55%	WP_020866323.1	Streptomyces rapamycinicus NRRL 54	67%	WP_020866324.1	AVES 4 [Streptomyces mobaraensis NBRC 13	77%	WP_004951568.1
Martiodrillus			tautomycetin [Streptomyces sp. CK44	54%	ABI94379.1	Salinispora pacifica	66%	WP_019872710.1	Streptomyces avermitilis MA-4680	76%	NP_824073.1
			AVES 4 [Streptomyces mobaraensis N	54%	WP_004951568.1	Salinispora pacifica	66%	WP_020113574.1	Streptomyces sp. R1-NS-10	76%	WP_019075974.1
Rozowormgut	7	A8-1	actinomycete QD08-1	58%	AEE69404.1	Streptomyces hygroscopicus	67%	CCF23202.1	Streptomyces hygroscopicus subsp. Enhygru	86%	BAH67070.1
Martiodrillus			QmnA1 [Amycolatopsis orientalis]	58%	AFI57005.1	Salinispora pacifica	65%	WP_019872710.1	Streptomyces hyalinus	84%	BAH67331.1
			Streptomyces sp. M-02	57%	ACM62767.1	Salinispora pacifica	65%	WP_020113574.1	Streptomyces sp. ID05-A0332	83%	BAH67930.1
Rozowormgut	8	A8-1	Streptomyces sp. MA6548	54%	AAC68815.1	Streptomyces sp. C	56%	WP_007269166.1	Streptomyces sp. CNY243	88%	WP_018851822.1
Martiodrillus			Streptomyces violaceusniger Tu 4113	53%	YP_004815232.1	Streptomyces violaceusniger Tu 4113	56%	YP_004815232.1	Streptomyces sp. CNQ766	88%	WP_018840991.1
			Streptomyces sp. NRRL 30748	53%	ABC87511.1	Amycolatopsis vancoresmycina DSM	56%	WP_004560200.1	Streptomyces sp. CNS335	88%	WP_018841058.1
Rozowormgut	9	A8-3	AmbC [Sorangium cellulosum]	60%	ABK32257.1	AmbC [Sorangium cellulosum]	64%	ABK32257.1	Streptomyces viridochromogenes Tue57	73%	WP_004004070.1
Martiodrillus			AmbE [Sorangium cellulosum]	60%	ABK32259.1	JerC [Sorangium cellulosum]	64%	ABK32289.1	Streptomyces albus subsp. Albus	71%	BAH67207.1
			JerC [Sorangium cellulosum]	59%	ABK32289.1	AmbA [Sorangium cellulosum]	63%	ABK32255.1	Streptomyces sp. ATexAB-D23	71%	WP_018555451.1
Rozowormgut	10	A8-1	Streptomyces sp. MA6548	54%	AAC68815.1	Streptomyces sp. C	56%	WP_007269166.1	Streptomyces sp. CNY243	88%	WP_018851822.1
Martiodrillus			Streptomyces violaceusniger Tu 4113	53%	YP_004815232.1	Streptomyces violaceusniger Tu 4113	55%	YP_004815232.1	Streptomyces sp. CNQ766	88%	WP_018840991.1
			Streptomyces sp. NRRL 30748	53%	ABC87511.1	Amycolatopsis vancoresmycina DSM	55%	WP_004560200.1	Streptomyces sp. CNS335	88%	WP_018841058.1

Table S3-1									
GenBank protein accession number.	**Phylum**	**Organism**	**Length (bp)**	**GC ration (%)**	**AT group classification of Jenke-Kodama et al (2005)**	**Product (if known)**	**63 Amino Acid motif and predicted substrate.**	**92 Amino Acid motif and predicted substrate.**	**201 Amino Acid motif and predicted substrate.**
ZP_05113660	Alphaproteobacteria	Labrenzia alexandrii DFL-11	1295	58.76	A1		TPSLQL	GHSMGE methylmalonyl	IAAH
ZP_01878022	Alphaproteobacteria	Roseovarius sp. TM1035	1286	66.33	A1		QPSVQL	GHSMGE methylmalonyl	IAAH
ZP_08862185	Alphaproteobacteria	Ruegeria pomeroyi	1286	67.34	A1		TPSVQL	GHSMGE methylmalonyl	IAAH
YP_006561756	Alphaproteobacteria	Phaeobacter gallaeciensis 2.10	1289	63.69	A1		QPSVQL	GHSMGE methylmalonyl	IAAH
ZP_05080683	Alphaproteobacteria	Rhodobacterales bacterium Y4I	1289	65.87	A1		QPSVQL	GHSMGE methylmalonyl	IAAH
ZP_01444322	Alphaproteobacteria	Pelagibaca bermudensis HTCC2601]	1286	70.61	A1		RPSLQL	GHSMGE methylmalonyl	IAAH
ZP_05741641	Alphaproteobacteria	Silicibacter sp. TrichCH4B	1289	61.91	A1		KPSVQL	GHSMGE methylmalonyl	IAAH
ZP_10019130	Alphaproteobacteria	Citreicella sp. 357	1280	66.09	A1		RPSLQL	GHSMGE methylmalonyl	IAAH
ABA21235	Cyanobacteria	Anabaena variabilis ATCC 29413	1310	43.21	A1		QTGIAQ	AHSIGE	HAFH-malonyl
CAQ43079	Myxobacterial	Chondromyces crocatus	1352	68.42	A1	chondrochloren	RTFLAQ	GYSIGE	HAFH-malonyl
AAO65798	Actinomycetal	Streptomyces cinnamonensis	1446	73.65	A2	monensin	GTGMTQ malonyl	GHSVGE malonyl	HAFH-malonyl
AAO65800	Actinomycetal	Streptomyces cinnamonensis	1332	66.52	A2	monensin	HTMWAQ malonyl	GHSIGE	HAFH-malonyl
AAO65801	Actinomycetal	Streptomyces cinnamonensis	1332	68.99	A2	monensin	HTVWAQ malonyl	GHSIGE malonyl	HAFH-malonyl
ADC79616	Actinomycetal	Streptomyces lohii	1449	75.36	A2	bafilomycin	RTRYTQ malonyl	GHSIGE malonyl	HAFH-malonyl
ADC79617	Actinomycetal	Streptomyces lohii	1569	76.48	A2	bafilomycin	RTEYTQ malonyl	GHSVGE malonyl	HAFH-malonyl
CAE02605	Actinomycetal	Streptomyces thioluteus	1308	74.08	A2	aureothin	QTGYTQ malonyl	GHSIGE malonyl	HAFH-malonyl
AAC69329	Actinomycetal	Streptomyces venezuelae	1452	74.79	A2	Pikromycin	ETRYTQ malonyl	GHSVGE malonyl	HAFH-malonyl
CAA60460	Actinomycetal	Streptomyces rapamycinicus	1158	66.58	A2	rapamycin	ETGYAQ malonyl	GHSVGE malonyl	HAFH-malonyl
CAA60459	Actinomycetal	Streptomyces rapamycinicus	1158	67.36	A2	rapamycin	ETGYAQ malonyl	GHSVGE malonyl	HAFH-malonyl
CAA60459	Actinomycetal	Streptomyces rapamycinicus	1182	66.5	A2	rapamycin	ETGYAQ malonyl	GHSVGE malonyl	HAFH-malonyl
ZP_06967007	Chloroflexi	Ktedonobacter	1304	61.5	A3		ET?FTQ malonyl	GHSAGE	HAFH-malonyl
ABA23591	Cyanobacteria	Anabaena variabilis ATCC 29413	1319	45.72	A3		ETAYTQ malonyl	GHSLGEmethylmalonyl	HAFH-malonyl
ZP_07113428	Cyanobacteria	Oscillatoria sp. PCC 6506	1292	49.92	A3	cylindrospermopsin	ETAYTQ malonyl	GHSVGE malonyl	HAFH-malonyl
AEU11006	Cyanobacteria	Nostoc/Nostocaceae	1310	47.02	A3		QTAYTQ	GHSVGE malonyl	HAFH-malonyl
ABA24342	Cyanobacteria	Anabaena variabilis ATCC 29413	1304	44.33	A3		QTAYTQ	GHSVGE malonyl	HAFH-malonyl
ABA23708	Cyanobacteria	Anabaena variabilis ATCC 29413	1307	44.91	A3		ETAYTQ	GHSFGE	HAFH-malonyl
AEH26538	Metagenomic	soil core, 10-50 cm depth/Wisconsin	1310	67.63	A3		QTAYTQ	GHSVGQ	HAFH-malonyl
AEH26539	Metagenomic	soil core, 10-50 cm depth/Wisconsin	1313	66.87	A3		DTAFTQ malonyl	GHSVGE malonyl	HAFH-malonyl
AEH26539	Metagenomic	soil core, 10-50 cm depth/Wisconsin	1331	67.54	A3		DTAYTQ malonyl	GHSVGE malonyl	HAFH-malonyl
ADB12488	Myxobacterial	Sorangium cellulosum	1305	70.19	A3	epothilone	QTAFTQ	GHSIGE malonyl	HAFH-malonyl
CAD19086	Myxobacterial	Stigmatella aurantiaca Sg a15	1301	64.8	A3	stigmatellin	ET?FTQ malonyl	GHSVGE malonyl	HAFH-malonyl
CAD19087	Myxobacterial	Stigmatella aurantiaca Sg a15	1331	64.84	A3	stigmatellin	ET?FTQ malonyl	GHSVGE malonyl	HAFH-malonyl

Table S3-2									
GenBank protein accession number.	**Phylum**	**Organism**	**Length (bp)**	**GC ration (%)**	**AT group classification of Jenke-Kodama et al (2005)**	**Product (if known)**	**63 Amino Acid motif and predicted substrate.**	**92 Amino Acid motif and predicted substrate.**	**201 Amino Acid motif and predicted substrate.**
ZP_09839378	Actinomycetal	Nocardia brasiliensis ATCC 700358/Co	1328	69.88	A5		ETFIAQ	GHSIGE malonyl	VAFH-malonyl
EFM11600	Bacillales	Paenibacillus curdlanolyticus YK9	1337	56.47	A5		DTEVAQ	GHSAGE	VPYH
NP_889454	Betaproteobacteria	Burkholderiales;Bordetella bronchise	1307	72.15	A5		STEYAQ	GHSVGE malonyl	YAFH-malonyl
NP_925775	Cyanobacteria	Gloeobacter violaceus PCC 7421	1304	68.87	A5		ETEIAQ	GHSVGE malonyl	YAFH-malonyl
ZP_08430798	Cyanobacteria	Cyanobacteria; Oscillatoriales; Moore	1340	51.12	A5		ETEIAQ	GHSVGE malonyl	YAFN
AEI32790	Firmicutes	Clostridium acetobutylicum ATCC 824	1316	35.11	A5		ETLYAQ	GHSVGE malonyl	AAYH
ZP_04622166	Gammaproteobacteria	Yersinia kristensenii ATCC 33638	1328	41.27	A5		DTWVVQ	GHSLGEmethylmalonyl	VPYH
ACV30045	Metagenomic	Uncultured bacterial Eucalyptus plan	1283	67.73	A5		RSDVVQ	GHSVGE malonyl	FAYH
CAG25973	Metagenomic	organic soil, near Lyon, France	1298	72.03	A5		RTEIAQ	GHSVGE malonyl	YAFH-malonyl
AEH26521	Metagenomic	soil core, 10-50 cm depth/Wisconsin	1304	64.11	A5		ETAIAQ	GHSVGE malonyl	YAFH-malonyl
YP_632255	Myxobacterial	Myxococcus xanthus DK 1622	1310	72.14	A5		RTDIAQ	GHSAGE	VPYH
CAG28678.1	Myxobacterial	Stigmatella aurantiaca	1337	69.26	A5	myxochromide	RTDIAQ	GHSAGE	VPYH
ZP_09567224	Planctomycetales	Singulisphaera acidiphila DSM 18658	1319	67.78	A5		VTAIAQ	GHSVGE malonyl	YAFH-malonyl
CAD74845	Planctomycetes	Rhodopirellula baltica SH 1	1398	55.08	A5		RTSIAQ	GHSVGE malonyl	YAFH-malonyl
BAG69021.1	Myxobacterial	Cystobacter fuscus	1304	70.4	A6	mmxC	DIDVIQ	GHSMGE methylmalonyl	VASH-methylmalonyl
AAK57186.1	Myxobacterial	Stigmatella aurantiaca	1304	65.87	A6	myxalamid	DIDVIQ	GHSMGE methylmalonyl	VASH-methylmalonyl
YP_003953630	Myxobacterial	Stigmatella aurantiaca DW4/3-1	1316	67.48	A6	myxothiazol	DIDVLQ	GHSMGE methylmalonyl	VASH-methylmalonyl
CBD77734	Myxobacterial	Chondromyces crocatus	1307	67.41	A6	crocacin	EIDVIQ	GHSMGE methylmalonyl	VASH-methylmalonyl
CAQ43078	Myxobacterial	Chondromyces crocatus	1316	70.74	A6	chondrochloren	RIETLH	GHSVGE malonyl	VASH-methylmalonyl
ADB12490.1	Myxobacterial	Sorangium cellulosum	1301	58.8	A6	epothilone	RIDVVQ-gmethylmalonyl	GHSMGE methylmalonyl	VASH-methylmalonyl
CAD1901.1	Myxobacterial	Stigmatella aurantiaca Sg a15	1328	65.66	A6	stigmatellin	RIDVLQ	GHSMGE methylmalonyl	VASH-methylmalonyl
AAK19883	Myxobacterial	Sorangium cellulosum	1277	66.17	A6	soraphen	EIDVSL	GHSTGE	VAPH
CAD19085	Myxobacterial	Stigmatella aurantiaca Sg a15	1277	65.31	A6		SIEVIQ	GHSMGE methylmalonyl	VASH-methylmalonyl
AAQ20787.1	Actinomycetal	Streptomyces hygroscopus	1310	71.68	A8-1		RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ACL97711.1	Actinomycetal	Streptomyces globosus	1376	77.4	A8-1		RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ABV97151	Actinomycetal	Salinispora arenicola CNS-205	1310	73.59	A8-1		RVDVLQ	GHSQGE methylmalonyl	YASH-methylmalonyl
YP_003762843	Actinomycetal	Amycolatopsis mediterranei U32	1283	73.11	A8-1	rifamycin	RVDVLQ	GHSQGE methylmalonyl	YASH-methylmalonyl
ZP_04604648	Actinomycetal	Micromonospora sp. ATCC 39149	1304	76.92	A8-1		RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
AEZ64504.1	Actinomycetal	Streptomyces chromofuscus	1286	75.19	A8-1	herboxidiene	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
CAE02602	Actinomycetal	Streptomyces thioluteus	1307	75.44	A8-1	aureothin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ADC45535.1	Actinomycetal	Streptomyces nanchangensis	1301	73.71	A8-1	meilingmycin	RADVVQ-gmethylmalonyl	GHSQGE methylmalonyl	WASH-methylmalonyl
AFD04714	Actinomycetal	Saccharopolyspora sp. AH1-9	1265	72.17	A8-1		EVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ADC79618	Actinomycetal	Streptomyces lohii strain ATCC BAA-1	1424	75.35	A8-1	bafilomycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
BAH02270.1	Actinomycetal	Streptomyces platensis	1307	71.61	A8-1	pladienolide	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ZP_09978599	Actinomycetal	Mycobacterium xenopi RIVM700367	1301	68.18	A8-1		RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl

Table S3-3										
GenBank protein accession number.	**Phylum**	**Organism**	**Length (bp)**	**GC ration (%)**	**AT group classification of Jenke-Kodama et al (2005)**	**Product (if known)**	**63 Amino Acid motif and predicted substrate.**	**92 Amino Acid motif and predicted substrate.**	**201 Amino Acid motif and predicted substrate.**	
CAQ64687.1	Actinomycetal	Streptomyces lasaliensis	1295	69.88	A8-1	lasalocid	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAC69331.1	Actinomycetal	Streptomyces venezuelae	1292	71.52	A8-1	methymycin/pikromycir	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAO65796	Actinomycetal	Streptomyces cinnamonensis	1310	73.05	A8-1	monensin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAO65796	Actinomycetal	Streptomyces cinnamonensis	1305	73.03	A8-1	monensin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAO65806	Actinomycetal	Streptomyces cinnamonensis	1293	70.92	A8-1	monensin	RVEVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
ADC79616	Actinomycetal	Streptomyces lohii	1440	76.32	A8-1	bafilomycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YAAH	
ADC79617	Actinomycetal	Streptomyces lohii	1383	75.34	A8-1	bafilomycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	FAGH	
ADC79618	Actinomycetal	Streptomyces lohii	1419	75.33	A8-1	bafilomycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
CAE02606	Actinomycetal	Streptomyces thioluteus	1326	74.06	A8-1	aureothin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAC69329	Actinomycetal	Streptomyces venezuelae	1341	71.59	A8-1	Pikromycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
CAA60460	Actinomycetal	Streptomyces rapamycinicus	1227	69.76	A8-1	rapamycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
CAA60462	Actinomycetal	Streptomyces rapamycinicus	1221	68.06	A8-1	rapamycin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAA79984.2	Actinomycetal	Sorangium cellulosum	1313	70.07	A8-2	soraphen	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAA79984.2	Actinomycetal	Sorangium cellulosum	1322	71.33	A8-2	soraphen	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAA79984.2	Actinomycetal	Sorangium cellulosum	1322	71.63	A8-2	soraphen	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	FASH	
AAA79984.2	Actinomycetal	Sorangium cellulosum	1322	71.63	A8-2	soraphen	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl	
YP_003267360.1	Actinomycetal	Haliangium ochraceum DSM 14365	1307	66.56	A8-2		RLDVVQ	GHSQGE methylmalonyl	YASH-methylmalonyl	
AAA79984.2	Myxobacterial	Sorangium cellulosum	1316	69.45	A8-2	soraphen	RIEVVQ-gmethylmalonyl	GHSQGE methylmalonyl	VASH-methylmalonyl	

Table S3-4									
GenBank protein accession number.	**Phylum**	**Organism**	**Length (bp)**	**GC ration (%)**	**AT group classification of Jenke-Kodama et al (2005)**	**Product (if known)**	**63 Amino Acid motif and predicted substrate.**	**92 Amino Acid motif and predicted substrate.**	**201 Amino Acid motif and predicted substrate.**
ABK32287	Myxobacterial	Sorangium cellulosum	1328	74.32	A8-3	jerangolid	RVDAVQ	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32255.1	Myxobacterial	Sorangium cellulosum	1328	74.47	A8-3	ambruticin	RVDAVQ	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32257.1	Myxobacterial	Sorangium cellulosum	1325	76.08	A8-3	ambruticin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32289	Myxobacterial	Sorangium cellulosum	1325	75.62	A8-3	jerangolid	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32259	Myxobacterial	Sorangium cellulosum	1313	76.24	A8-3	ambruticin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32256.1	Myxobacterial	Sorangium cellulosum	1325	76.38	A8-3	ambruticin	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ABK32288.1	Myxobacterial	Sorangium cellulosum	1307	75.75	A8-3	jerangolid	RVDVVQ methylmalonyl	GHSQGE methylmalonyl	YASH-methylmalonyl
ZP_09680195	Actinomycetal	Mycobacterium tusciae JS617	1285	67.63	Mbt		QVEI-Q	GHSLGEmethylmalonyl	YPGH
ZP_04747690	Actinomycetal	Mycobacterium kansasii ATCC 12478	1336	69.24	Mbt		QIDI-Q	GHSLGEmethylmalonyl	YPGH
CCC44744	Actinomycetal	Mycobacterium canettii CIPT 140010(	1336	67.81	Mbt		EIEI-E	GHSLGEmethylmalonyl	FPVH
EID13613	Actinomycetal	Mycobacterium xenopi RIVM700367	1285	66.85	Mbt		QIQI-Q	GHSLGEmethylmalonyl	YPGH
EID10253	Actinomycetal	Mycobacterium phlei RIVM601174	1291	75.06	Mbt		QTEI-Q	GHSLGEmethylmalonyl	YPGH
AAS04491.1	Actinomycetal	Mycobacterium avium subsp. paratul	1294	72.18	Mbt		RPEI-Q	GHSLGEmethylmalonyl	YPGH

	Table S 4	**Amino acid distance**						
	Cluster/Pathway	**AT domain**		**KS domain**		**Length range (bp)**	**GC content %**	**Sequences**
		Maximum	**Overall mean**	**Maximum**	**Overall mean**			
								Sequence names.
	P1	0.137	0.056	0.069	0.014	1283-1295	67.00-67.65	Českе.Budéjovicecobblestones-ČeskaP9-10, Českе.Budéjovicestreetdust-KohinoorP3-4,
								Pariscobblestones-RiveGauchP3-10,Brusselscobblestones-BoucherP5,
								Pariscobblestones-CaféP5-6.
	P2	0.326	0.117	0.246	0.101	1305-1388	65.06-70.74	CdnTundra-ResoluteP1-13 P16-17, Czcultivated-PlanaP1, P2, P4-10, , P7, Pariscobblestones-CaféP4,
Unifrac								Czpasture-PalavaP1, P2, P4, P5-10, Colforest-CIATP1, Colpasture-PanceP3, Rozowormgut-MartiodrillusP9.
Specific								Českе.Budéjovice-ParadiseP3.
nodes	P3	0.117	0.047	0.079	0.023	1288-1300	66.62-68.00	Ottawacobblestones-DublinerP1-16, Budapestpavement-TerezKorutP1-13, Ottawastreetdust-RideauP3
Cosmo	C1	0.181	0.087	0.113	0.056	1316-1324	68.13-68.58	Českе.Budéjovicecobblestones-ČeskaP1 3, Českе.Budéjovicestreetdust-KohinoorP1, Pariscobblestones-RiveGaucheP2
politan	C2	0.012	0.012	0	0	1308-1312	67.23-67.28	Českе.Budéjovicecobblestones-ČeskaP5, Brusselscobblestones-BoucherP7.
nodes	C3	0.147	0.106	0.061	0.03	1276-1354	67.87-69.79	Colforest-HaticoP15, CdnBeach-P.E.I.P3 6.
	C4	0.073	0.032	0.049	0.022	1302-1310	68.99-69.35	Colforest-HaticoP3 9 13 16, Hungforest-CitadelP4
	C5	0.157	0.106	0.241	0.167	1306-1315	67.84-69.05	Cdnforest-KemptvilleP1 2, Hungforest-CitadelP3 5.
								GenBank Nucleotide accession numbers and region.
Erytho	Oleandomycin	0.074	0.045	0.206	0.104	1290-1308	66.82-67.44	AF220951.1: 4799-6091, 9284-1053, 14234-15541, 18923-20224.
mycin	Lankamycin	0.732	0.552	0.484	0.388	1299-1371	74.52-77.60	AB088224.2:57174-58490, 61689-62990, 68186-69556, 72689-73987, 77456-78766, 81938-83281
cluster	Erythromycin-*S. erythraea*	0.676	0.593	0.539	0.444	1278-1305	72.29-75.12	M63676.1:3276-4577, 7695-8999. M63677.1:1132-2427, 5503-6789, 11827-13104.
	Erythromycin-*A. erythreum*	0.751	0.614	0.658	0.528	1290-1353	72.88-76.36	AY623658.2:27638-26286, 32060-30729, 38431-37142, 42799-41498, 47597-46290 51992-50700.
	Borrelidin	0	0	0.096	0.064	1260-1260	67.86-68.02	AJ580915.1: 29881-31140, 35310-36569, 40241-41500.
	Epithilone	0.202	0.161	0.626	0.561	1293-1296	70.52-72.84	AF210843.1: 17298-18593, 38072-39364, 44637-45932.
	Laidlomycin	0.273	0.189	0.472	0.261	1293-1317	69.35-74.03	JQ793783.1: 20204-21502, 24536-25840, 36534-37835, 43099-44400, 54806-56107, 82618-81326, 78013-76697, 87560-86265.
	Monensin	0.285	0.118	0.4	0.278	1293-1311	70.22-73.03	AF440781.1: 18381-19685, 24324-25619, 36925-38217, 55733-57025, 87589-86297, 82939-81629, 92613-91315.
	Amb/Jer cluster	0.306	0.213	0.741	0.496	1308-1312	74.30-76.36	Jerangolid DQ897668.1: 16882-18207, 31196-32524, 36533-37870 Ambruticin DQ897667: 15676-17001, 18951-20276,
								30020-31342, 46662-47978.
	Nyastatin	0.501	0.335	0.295	0.227	1311-1329	73.32-75.82	AF263912.1: 17018-18346, 67939-69261, 72565-73875.