A Thesis Submitted for the Degree of PhD at the University of Warwick

Permanent WRAP URL:

<http://wrap.warwick.ac.uk/151039>

Copyright and reuse:

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

The structure and function of microbial communities at the soil-atmosphere interface

Christopher James O'Grady

A thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences, University of Warwick, Gibbet Hill
Road, Coventry, Warwickshire, CV4 7AL, UK

September 2019

TABLE OF CONTENTS

LIST OF TABLES	VIII
LIST OF FIGURES	IX
ACKNOWLEDGEMENTS.....	XII
DECLARATION.....	XIII
LIST OF ABBREVIATIONS.....	XIV
CHAPTER 1 – GENERAL INTRODUCTION.....	1
1.1. Microbial ecology	1
1.1.1. Advances in microbial ecology	2
1.1.2. Amplicon DNA sequencing	3
1.1.3. Shotgun metagenomics	7
1.1.4. Advances in ecological theory	16
1.2. Soil ecosystems	16
1.2.1. The importance of soil	16
1.2.2. Soil health.....	17
1.2.3. Soil ecosystem functions	18
1.2.3.1 Biodiversity in soil	19
1.2.3.2 Soil biogeochemical cycles	20
1.2.4. Soil physics	21
1.2.4.1. Techniques in soil physics.....	22
1.2.4.2. Soil pores	24
1.2.4.3. Soil pore metrics.....	25
1.2.4.3.a. Soil porosity (ϕ).....	25
1.2.4.3.b. Pore categorization and size distribution	26
1.2.3.3.c. Pore connectivity	27
1.2.3.3.d. Pore circularity	28
1.2.5 Soil compartmentalization	28
1.2.5.1 The rhizosphere	29
1.2.5.2. Bulk soil	30
1.2.5.3. The soil surface	31
1.3. The Biological Soil Crust (BSC)	32

1.3.1. Soil crust functions.....	33
1.3.2. The soil crust microbiome.....	36
1.3.2.1. Cyanobacteria.....	36
1.3.2.2. Eukaryotic algae.....	38
1.3.2.3. Other primary producers and carbon fixers.....	39
1.3.2.4. Heterotrophs.....	40
1.4. Research gaps.....	40
1.5. Project aims.....	45
 CHAPTER 2 – THE TEMPORAL AND LOCAL-SCALE SPATIAL DYNAMICS OF SOIL CRUST AND BULK SOIL MICROBIAL COMMUNITIES.....	 47
2.1. Introduction.....	47
2.1.1. Spatial variation.....	47
2.1.2. Temporal variation.....	48
2.1.3. Perturbations of soil crusts.....	49
2.1.4. Assembly and microbial interactions of soil crust communities.....	50
2.1.5. Chapter 2 aims.....	51
2.2. Methods.....	53
2.2.1. Location, experimental design, sampling methods and core processing.....	53
2.2.2. Chlorophyll α content.....	53
2.2.3. pH.....	54
2.2.4. Soil moisture.....	55
2.2.5. Microbial carbon and nitrogen content.....	55
2.2.6. Bulk density.....	56
2.2.7. DNA extraction.....	56
2.2.8. Illumina MiSeq library preparation.....	56
2.2.9. Bioinformatics analyses.....	58
2.2.10. Statistical analysis.....	59
2.2.11. Network analysis.....	60
2.2.12. Community assembly processes.....	60
2.3. Results.....	62
2.3.1. Sample site temporal characteristics.....	62
2.3.1.1. Bulk density.....	63
2.3.1.2. Weather data.....	63
2.3.1.3. Soil moisture content.....	63
2.3.1.4. pH analysis.....	63
2.3.2. Microbial biomass.....	64

2.3.3. Chlorophyll α concentration and morphology	65
2.3.4. Rarefaction	66
2.3.5. Compartmental comparisons of soil crust and bulk soil microbial communities	67
2.3.5.1. Alpha diversity	67
2.3.5.2. Microbial community composition	67
2.3.5.3. Taxonomy	70
2.3.5.4. Soil crust-enriched OTUs	74
2.3.5.5. Co-occurrence microbial networks	79
2.3.6. Temporal dynamics of soil crust microbial communities	84
2.3.6.1. Alpha diversity	84
2.3.6.2. Community composition	86
2.3.6.3. Temporal taxonomic shifts	89
2.3.6.3.a. Bacteria	89
2.3.6.3.b. Fungi	90
2.3.6.3.c. Protists	91
2.3.6.3.d. Phototrophs	91
2.3.6.4. Temporal co-occurrence microbial networks	97
2.3.7. Community assembly processes	100
2.3.7.1. Phylogenetic signal.....	100
2.3.7.2. β -NTI.....	100
2.3.8. The relative impacts of time, distance and edaphic factors on soil crust microbial communities	104
2.4. Discussion	107
2.4.1. Alpha diversity	108
2.4.2. Community composition	109
2.4.3. Taxonomic comparisons	110
2.4.3.1. Bacterial communities	110
2.4.3.2. Fungal communities	113
2.4.3.3. Protist communities	114
2.4.3.4. Phototroph communities.....	115
2.4.4. Microbial co-occurrence networks.....	117
2.4.5. Community assembly processes	120
2.4.6. Soil physiochemical properties	121
2.4.6.1. Soil moisture.....	121
2.4.6.2. Soil pH.....	121
2.4.6.3. Physical perturbations to the soil surface	123
2.4.7. The relative contributions of edaphic factors, time and distance in shaping microbial community composition.....	124
2.4.8. Conclusions	125

CHAPTER 3 – THE INFLUENCE OF BIOLOGICAL SOIL CRUST DEVELOPMENT ON MICROBIAL COMMUNITIES AND SOIL STRUCTURE	128
3.1. Introduction	128
3.1.1. Soil physics at the soil surface	128
3.1.2. Chapter 3 aims.....	130
3.2. Methods.....	132
3.2.1. Location, experimental design, sampling methods and core processing	132
3.2.2. DNA extraction, Illumina MiSeq library preparation, DNA amplicon sequencing, bioinformatics and statistical analysis.....	133
3.2.3. X-ray CT scanning	133
3.2.4. Image processing.....	133
3.2.5. TopCap	134
3.2.6. Particle analysis and shape descriptors	135
3.2.7. Soil seal thickness and compartmentalization.....	136
3.3. Results	137
3.3.1. Metadata	137
3.3.1.1. Bulk Density	138
3.3.1.2. pH	138
3.3.1.3. Soil moisture.....	138
3.3.1.4. Crust morphology and chlorophyll a content	138
3.3.1.5. Microbial biomass	139
3.3.2. Soil crust and bulk soil microbial communities	140
3.3.2.1. Rarefaction	140
3.3.2.2. Alpha diversity	140
3.3.2.3. Community composition	141
3.3.2.4. Temporal development of soil crust communities	143
3.3.3. CT scanning data.....	149
3.3.3.1 Soil seal thickness	149
3.3.3.2. Soil porosity	149
3.3.3.3. Soil pore number	151
3.3.3.4. Soil pore connectivity.....	152
3.3.3.5. Soil pore surface area	153
3.3.3.6. Circularity.....	154
3.4. Discussion	156
3.4.1. Biology of the soil surface	156
3.4.2. Soil physics at the soil surface	157
3.4.2.1. Soil porosity and pore number	158

3.4.2.2. Soil pore connectivity	160
3.4.2.3. Soil pore shape descriptors	162
3.3.4.3. Conclusions	163

CHAPTER 4 – THE FUNCTIONAL GENETIC POTENTIAL OF DEVELOPED BIOLOGICAL SOIL CRUSTS166

4.1. Introduction	166
4.1.1. Metagenomics and metatranscriptomics of the soil surface	166
4.1.2. Chapter 4 aims.....	167
4.2. Methods.....	169
4.2.1. Sample selection.....	169
4.2.2. DNA extraction and sequencing	169
4.2.3. Raw sequences	169
4.2.4. Co-assembly	170
4.2.5. Open Reading Frame prediction and functional annotation	170
4.2.6. Mapping and normalisation.....	171
4.2.7. Bioinformatics and statistical analysis	171
4.3. Results	172
4.3.1. Assembly statistics	172
4.3.2. Taxonomy	172
4.3.2.1. Comparison of bacterial amplicon MiSeq 16S rRNA taxonomy and bacterial shotgun metagenomics taxonomy	181
4.3.3. Functional annotations	182
4.3.3.1 Nitrogen cycling	183
4.3.3.2. Carbon cycling	185
4.3.3.3. Phosphorus cycling.....	194
4.3.3.4. Sulphur metabolism.....	196
4.4. Discussion	198
4.4.1. Taxonomic differences	199
4.4.2. Functional analyses	201
4.4.2.1 Nitrogen cycling	201
4.4.2.2. Carbon cycling	203
4.4.2.2.a. Carbon fixation	203
4.4.2.2.b. Carbohydrate-Active Enzymes (CAZy)	205
4.4.2.2.c. Methane cycling.....	207
4.4.2.3. Phosphorus cycling.....	207
4.4.2.4. Sulphur cycling.....	208
4.4.3. Comparison of soil crust and bulk soil amplicon and metagenomic profiles	209

4.4.4. Conclusions	210
CHAPTER 5 – GENERAL DISCUSSION	212
5.1. Overview	212
5.2. The temporal and local-scale spatial dynamics of bacterial, fungal, protist and phototroph soil crust and bulk soil community composition	212
5.3. The influence of biological soil crust development on the physical soil structure throughout the soil profile.....	217
5.4. Taxonomic and functional profiles of developed biological soil crust and bulk soil metagenomes.....	220
5.5. Implications of studies	223
5.6. Future work	224
APPENDIX	227
BIBLIOGRAPHY	241

LIST OF TABLES

Table 1.1. Performance of metagenomic assembly of Tara Oceans metagenome.	12
Table 1.2. Ecosystem services obtained from soil.....	19
Table 1.3. Classification of micropores, mesopores and macropores based on diameter.	27
Table 1.4. Overview of current literature examining community structure of biological soil crusts using next-generation sequencing methods.	43
Table 2.1. Chemical application at the sampling site.	54
Table 2.2. Primer sequences used for amplicon DNA sequencing.....	57
Table 2.3. PCR conditions for each amplicon	57
Table 2.4. Number of sequences assigned to each amplicon.....	67
Table 2.5. Analysis of similarity (ANOSIM) of soil crust and bulk soil OTUs.....	69
Table 2.6. Compartmental OTU SIMPER analysis.....	78
Table 2.7. Correlation network statistics.	80
Table 2.8. Temporal alpha diversity statistics of soil crust and bulk soil communities.	85
Table 2.9. Temporal alpha diversity regression statistics.....	85
Table 2.10 Statistics of community composition similarity in soil crust and bulk soil communities across time.	88
Table 2.11. Regression statistics of community composition similarity in soil crust and bulk soil communities across time.....	89
Table 2.12. Redundancy analysis of soil crust microbial communities.....	105
Table 2.13. Redundancy analysis of bulk soil microbial communities.	106
Table 3.1. Chemical application at the sampling site.	133
Table 3.2. Pore size classification.....	135
Table 3.3. Statistics of community composition similarity in soil crust and bulk soil communities across time.	145
Table 4.1. Number of base pairs obtained from DNA sequencing.....	170
Table 4.2. Statistics of contigs after co-assembly.....	172
Table 4.3. Number of contigs assigned to each kingdom.	173
Table 4.4. Mapped coverage of soil crust and bulk soil reads to each kingdom.	174
Supplementary table 2.1. Differentially abundant bacterial OTUs between combined soil crust and bulk soil communities.....	229
Supplementary table 2.2. Differentially abundant fungal OTUs between combined soil crust and bulk soil communities.	231
Supplementary table 2.3. Differentially abundant protist OTUs between combined soil crust and bulk soil communities.	233
Supplementary table 2.4. Differentially abundant phototroph OTUs between combined soil crust and bulk soil communities	234
Supplementary table 4.1. KEGG numbers for CAZy gene enzyme classes.....	236
Supplementary table 4.2. Phosphorus cycling gene KEGG numbers.....	238

LIST OF FIGURES

Figure 1.1. Taxonomic shifts in relative abundance between treatments.	8
Figure 1.2. Summary of a metagenomics workflow.....	9
Figure 1.3. De Bruijn graph assembly.	11
Figure 1.4. Soil health schematic.....	18
Figure 1.5. X-ray computed tomography schematic.....	23
Figure 1.6. X-ray computed tomography images of a soil core.	24
Figure 1.7. Schematics of soil pore structure.....	25
Figure 1.8. X-ray computed tomography reconstruction displaying soil pore connectivity.	28
Figure 1.9. Soil compartmentalization.....	29
Figure 1.10. Photographs of Biological Soil Crusts (BSCs).	32
Figure 1.11. Estimates of global contributions of biocrusts to the nitrogen cycle.	34
Figure 1.12. Optical microscope images of biocrust cyanobacteria.	37
Figure 1.13. Optical microscope images of biocrust algae.....	39
Figure 2.1. Sampling location and strategy.	54
Figure 2.2. Environmental conditions and soil properties.	62
Figure 2.3. Biomass carbon and nitrogen.	64
Figure 2.4. Chlorophyll α concentration and soil crust morphology.....	66
Figure 2.5. Alpha diversity of soil crust and bulk soil microbial communities.....	68
Figure 2.6. NMDS plots of soil crust and bulk soil microbial communities combined across all time points.....	69
Figure 2.7. Soil crust and bulk soil microbial community composition combined across all time points.....	70
Figure 2.8. Differential abundance of soil crust and bulk soil taxa combined across all time points. .	73
Figure 2.9. Volcano plots of OTUs enriched in soil crust and bulk soil communities.....	77
Figure 2.10. Correlation networks of soil crust and bulk soil microbial communities.....	82
Figure 2.11. Cartographic representation of microbial correlation networks.....	83
Figure 2.12. Alpha diversity of soil crust and bulk soil communities across time.	84
Figure 2.13. Community composition similarity in soil crust and bulk soil communities across time.	86
Figure 2.14. Temporal NMDS plots of soil crust and bulk soil microbial communities.....	87
Figure 2.15. Relative abundance of bacterial phyla and fungal classes across time in soil crust and bulk soil communities	93
Figure 2.16. Relative abundance of microbial eukaryotic phyla and phototrophic classes across time in soil crust and bulk soil communities.....	94
Figure 2.17. Relative abundance of temporally variable microbial taxa identified by SIMPER analysis.....	95
Figure 2.18. Relative abundance of temporally variable microbial OTUs identified by SIMPER analysis.....	96

Figure 2.19. Correlation networks of soil crust and bulk soil microbial communities at each time point.	98
Figure 2.20. Temporal network statistics.....	99
Figure 2.21. β -NTI and time in soil crust and bulk soil microbial communities.....	102
Figure 2.22. Community assembly processes in soil crust and bulk soil communities across time...	103
Figure 3.1. Soil seal thickness after increasing durations of rainfall in three soil types.....	128
Figure 3.2. Structure of a biological soil crust obtained from the Negev desert.	129
Figure 3.3. A 2D cross section of a cyanobacterial-lichen dominated soil crust.....	130
Figure 3.4. Sampling location and strategy.	132
Figure 3.5. The TopCap process.	135
Figure 3.6. Environmental conditions and soil properties.	137
Figure 3.7. Soil crust morphology at each time point.....	139
Figure 3.8. Alpha diversity of soil crust and bulk soil microbial communities.....	141
Figure 3.9. NMDS plots of soil crust and bulk soil microbial communities combined across all time points.....	142
Figure 3.10. Soil crust and bulk soil microbial community composition combined across all time points.....	143
Figure 3.11. Temporal NMDS plots of soil crust and bulk soil microbial communities.....	144
Figure 3.12. Community composition similarity in soil crust and bulk soil communities across time.	145
Figure 3.13. Relative abundance of bacterial, fungal, protist and phototroph taxa across time in soil crust and bulk soil communities.	148
Figure 3.14. Soil seal thickness.	149
Figure 3.15. Calculated soil porosity throughout the soil profile at each time point.....	150
Figure 3.16. Number of soil pores in different soil compartments at each sampling time point.....	152
Figure 3.17. Soil pore connectivity throughout the soil profile at each time point.	153
Figure 3.18. Soil pore surface area in different soil compartments at each sampling time point.....	154
Figure 3.19. Soil pore circularity in different soil compartments at each sampling time point.	155
Figure 4.1. NMDS plots of soil crust and bulk soil microbial communities from metagenomic sequencing.	173
Figure 4.2. Soil crust and bulk soil microbial community composition.	175
Figure 4.3. Differential abundance of soil crust and bulk soil taxa.	176
Figure 4.4. Volcano plots of phyla enriched in soil crust and bulk soil metagenomes.	177
Figure 4.5. Relative abundance of archaeal and bacterial phyla in soil crust and bulk soil communities.....	179
Figure 4.6. Relative abundance of eukaryotic and viral taxa in soil crust and bulk soil communities.	180
Figure 4.7. Bacterial community composition of soil crust and bulk soil phyla determined from amplicon and shotgun metagenomic sequencing.....	181

Figure 4.8. Abundance of genes involved in biogeochemical processes in the soil crust and bulk soil.	182
Figure 4.9. Abundance of nitrogen cycling genes in the soil crust and bulk soil.	183
Figure 4.10. Abundance of taxa assigned to nitrogen fixation genes in the soil crust and bulk soil.	184
Figure 4.11. Gene abundance of carbon fixing genes in the soil crust and bulk soil.	186
Figure 4.12. Abundance of genes involved in reductive-TCA, Wood-Ljungdahl and 3HP pathways in the soil crust and bulk soil.	187
Figure 4.13. Gene abundance of CAZy genes in the soil crust and bulk soil.	188
Figure 4.14. Gene abundance of CAZy enzyme categories in the soil crust and bulk soil.	190
Figure 4.15. Gene abundance of photosynthesis-related genes in the soil crust and bulk soil.	191
Figure 4.16. Gene abundance of methane and methanol oxidation genes in the soil crust and bulk soil.	193
Figure 4.17. Gene abundance phosphorus cycling genes in the soil crust and bulk soil.	195
Figure 4.18. Abundance of sulphur metabolism genes in the soil crust and bulk soil.	197
Supplementary figure 2.1. Mantel correlograms of bacterial and fungal soil crust and bulk soil communities.	227
Supplementary figure 2.2. Mantel correlograms of protist and phototroph soil crust and bulk soil communities.	228
Supplementary figure 3.1. Soil seal determination.	235

ACKNOWLEDGEMENTS

I would like to thank all three of my supervisors, Gary Bending, Hendrik Schäfer and Chris van der Gast, for their continued guidance, encouragement and passion throughout my project. In particular, I would like to thank Gary for allowing me to enjoy a large degree of freedom in undertaking my project and conducting my research in the direction I have wanted - it has been a pleasure to work with him for the past three years.

I would like to thank my collaborators at The University of Nottingham, Craig Sturrock and Sacha Mooney, for the great deal of support and help with both performing X-Ray CT scanning and interpreting the results.

I would like to thank all past and present members of the Bending and Schäfer lab groups for making my day-to-day working life extremely pleasant and thoroughly enjoyable. In particular, I would like to thank Bex Southwell for helping me set up and troubleshoot my experiments (even after she had left!), as well as providing guidance and help in a number of situations throughout my project; Sally Hilton for an enormous amount of help in the lab, in particular with MiSeq library preparation and analysis; Emma Picot for a vast amount of help with coding and for troubleshooting my horrendous error codes, which always seemed to have been generated when she was away on holiday (sorry!); Amy Newman for help with sampling and improving my talks, as well as saving me countless hours by giving me many lifts home. Helen Jones, Richard Fetherston, Emma Johnstone, Ellis Monaghan, Jess Palmer and Katy Faulkner for good tea breaks, help with understanding network analysis, assistance with sampling and providing extremely useful perspectives on my results; and most importantly Nello Gregori and Chris Hale for allowing me to rant about Liverpool FC and my poor performances in fantasy football whenever they would listen.

From outside the lab groups, I would also like to thank Seb Raguideau for an enormous amount of help with the metagenomics pipeline; Julia Lipecki for solving my week-long R problems in just 5 minutes of her time; and Maëva Vinet for her continued morale support, tremendous amounts of help sampling and guidance (especially with presentations!) throughout the last three years.

Lastly, I would like to thank the MIBTP for selecting me as part of their programme and the BBSRC for providing the funding to conduct my research.

DECLARATION

I declare that the work presented in this thesis was conducted by me under direct supervision of Professor Gary Bending, Doctor Hendrik Schäfer and Professor Chris van der Gast, with the exception of those instances where the contribution of others has been specifically acknowledged. The work in this thesis has not been previously submitted for any other degree.

Christopher James O’Grady

LIST OF ABBREVIATIONS

ANOSIM	– Analysis of similarity
ANOVA	– Analysis of variance
BAT	– Bin annotation tool
BC	– Bray-Curtis
β MNTD	- β Mean nearest taxon distance
β NTI	- β Nearest taxon index
BLAST	– Basic local alignment search tool
Bp	– Base pair
BSC	– Biological soil crust
C	– Carbon
CAT	– Contig annotation tool
CAZy	– Carbohydrate active enzymes
COG	– Clusters of Orthologous Groups
CT	– Computed tomography
DNA	– Deoxyribonucleic acid
°E	– Degrees East
ECD	– Equivalent Cylindrical Diameter
FAO	– Food and Agriculture Organisation of the United Nations
FIJI	– Fiji Is Just ImageJ
Ha	– Hectare
HSD	– Honestly significant difference
HMM	– Hidden Markov Model
ITS	– Internal transcribed spacer region
KEGG	– Kyoto Encyclopaedia of Genes and Genomes
L ₂ FC	– Logarithm ₂ fold change
L ₁₀ FC	– Logarithm ₁₀ fold change
Log	– Logarithm
NCBI	– National Centre for Biotechnology Information
NGS	– Next generation sequencing
N	– Nitrogen
°N	– Degrees North

Nr – Non-redundant
NMDS – Non metric multidimensional scaling
ORF – Open reading frame
P – Phosphorus
Pg - Picogram
PCNM – Principal coordinates of neighbour matrices
PCR – Polymerase chain reaction
Pfam – Protein families
PSI/PSII – Photosystem I/II
QIIME – Quantitative Insights into Microbial Ecology
OTU – Operational taxonomic unit
RDA – Redundancy analysis
RPKM – Reads per kilobase million
rRNA – Ribosomal ribonucleic acid
SIMPER – Similarity of percentages
SPARCC – Sparse Correlations for Compositional data
SSU – Small Subunit
Tg - Teragram
TIF – Tagged Image Format
UV - Ultraviolet

SUMMARY

The soil surface exhibits distinct physical and biotic characteristics in comparison to bulk soil due to the specific environmental conditions it is exposed to. The uppermost layer of soil can form a microbiotic layer, known as the Biological Soil Crust (BSC), that is composed of bryophytes, autotrophic algae, cyanobacteria, fungi and bacteria. In arid regions, BSCs are known to perform substantial roles in nutrient cycling, water infiltration and protection from soil erosion. However, our understanding of BSCs in temperate agricultural soils is currently limited and their functional importance in these ecosystems remains unknown. Here, using DNA amplicon sequencing it was revealed that temperate agricultural BSCs and bulk soils had substantial differences in microbial community composition, diversity and co-occurrence network topology, and were driven by different community assembly processes. Compositional differences between BSCs and bulk soils were mainly attributed to significant enrichments of Cyanobacteria, Bacteroidetes, Agaricomycetes, Dothideomycetes, Rhizaria, and Dinophyceae in the soil crust. Furthermore, BSC communities displayed significant temporal variation, marked by significant reductions in diversity and shifts in community composition. Substantial enrichments of moss sequences, which are keystone taxa for soil crust development in arid regions, were detected after just 7-8 months of development. X-ray CT image analysis revealed that a structural seal forms at the soil surface, that had substantially lower porosity, higher soil pore connectivity and more complex soil pore descriptors relative to lower depths. Additionally, soil pore differences across depths became exacerbated as BSCs developed over time. Shotgun metagenomic analysis of mature BSCs and bulk soils revealed that soil crust communities had a higher genetic potential for a number of key processes, including nitrogen fixation, nitrification, oxygenic photosynthesis, phosphonate degradation and breakdown of complex carbohydrates. This highlights that communities at the soil surface may have substantially different contributions to biogeochemical cycling than soils from lower depths. Together, results presented here all indicate that BSCs in a temperate agroecosystem form a distinct soil compartment that has a substantially different physical soil structure, biological composition and functional capacity relative to lower depth soil.

CHAPTER 1 – GENERAL INTRODUCTION

1.1. Microbial ecology

Microorganisms underpin an array of irreplaceable global processes that are essential for sustaining life on earth. They are the foundations for every biogeochemical cycle and represent significant stocks of organic matter. For example, it is estimated that global carbon, nitrogen and phosphorus stocks are up to 10-fold higher in prokaryotic organisms than in all plants (Whitman et al., 1998). Microorganisms are also ubiquitously distributed across the biosphere (with a predicted abundance between 9.2×10^{29} – 31.7×10^{29} microbial cells) and are capable of inhabiting almost every conceivable habitat (Kallmeyer et al., 2012). However, as many ecological theories and frameworks do not suitably align with microbial distribution, understanding the processes that determine spatial and temporal distribution of microbiota remains one of the most widely disputed and compelling challenging topics in ecology (Prosser et al., 2007).

One of the first mainstream theories in microbial ecology comes from the work of Beijerinck early in the twentieth century, articulated by Baas Becking (1931), in which he suggested a globally cosmopolitan distribution of microorganisms and is summarized by the famous statement that “everything is everywhere, but the environment selects”. This theory derives from the fact that microorganisms have large populations and display incredibly high rates of dispersal, suggesting that environmental conditions dictate species distribution, not geographic distance or the ability to disperse (Barberán et al., 2014, O’Malley, 2008). Since then, a vast body of work has tested this theory, which both supports and challenges this statement. For example, Gibbons et al. (2013) found in marine systems that deeper sequencing efforts uncovered greater phylogenetic diversity through the discovery of rare taxa; whereas Papke et al. (2003) suggested that certain thermophilic cyanobacterial populations were geographically restricted and endemic to particular locations. A review by Martiny et al. (2006) evaluated a wide array of such studies examining whether environmental factors (Baas Becking theory) or historical assemblages (such as dispersal limitation from previous populations),

or a combination of both drove microbial species distribution. They concluded that not one argument fits all and it is likely that both scenarios are true. This is highlighted by more recent studies such as Herbold et al. (2014) which found a cosmopolitan distribution of taxa in surface geothermal sediments, whereas endemic archaeal and candidate bacterial taxa were found in the immediate subsurface geothermal sediments.

1.1.1. Advances in microbial ecology

Advances in the use of molecular and genomic techniques have revolutionised the field of microbial ecology over the past three decades. Traditional methods in microbial ecology centred around microscopy and culturing individual populations isolated from environmental samples, but these approaches are often low-throughput and biased towards selecting certain groups of microbial taxa (Handelsman, 2004). Cultivation of microorganisms in artificial nutrient-rich media can result in the proliferation of microorganisms that are not normally dominant in environmental systems, which have been described as the ‘weeds’ of the microbial world (Hugenholtz, 2002). Therefore, only easily-culturable taxa have been defined using culturing methods, which is problematic as less than 1% of all microbial species are culturable by standard techniques – a phenomenon known as the “great plate-count anomaly” (Staley and Konopka, 1985). Furthermore, culture-based approaches are performed in artificial nutrient-rich systems, which can frequently result in the production of unnatural or toxic conditions, for example high levels of oxidative stress (Morris et al., 2011, Tanaka et al., 2014). Conversely, microscopy is an alternative traditional approach to culture-based methods, but is extremely low-throughput and relies heavily on observations, therefore making functional or taxonomic conclusions extremely difficult (Moreira and López-García, 2002). As a result, the use of traditional methods in microbial ecology provides little context on the diversity and complexity of microbial communities in environmentally-relevant systems.

To address these issues ecologists moved towards using molecular fingerprinting approaches to obtain community profiles of environmental samples (Hugerth and Andersson, 2017). Molecular phylogenetic analysis was initiated by the work from Carl Woese and colleagues in 1977, who inferred phylogenetic relationships between

prokaryotes based on the 16S rRNA subunit, which led to the discovery that prokaryotes were composed of separate lineages of bacteria and archaea (Woese and Fox, 1977). Further work by Woese applied the same phylogenetic analysis on the 18S rRNA small subunit to identify eukaryotes as a distinct microbial domain, and additionally identified 11 taxonomically-distinct bacterial lineages (phyla) from 16S rRNA DNA sequences collected from pure cultures (Woese et al., 1985, Woese, 1987). Characterisation of 16S/18S rRNA molecular phylogenetic markers along with other markers, such as the ribosomal internal transcribed spacer region (ITS) and 5S large rRNA subunit, were shortly after applied to analyse microbial taxonomy in natural populations obtained from environmental samples (Pace et al., 1986, Stahl et al., 1985). These progressions in molecular ecology led to the development of community-profiling techniques that could examine all the extracted genetic material of an environmental sample, such as denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE), which made it possible to quickly assess microbial diversity and community composition between samples based on nucleotide differences in phylogenetic markers (Muyzer et al., 1993). The application of molecular analysis of ribosomal subunits in microbial ecology remains highly appealing over traditional methods as rRNA genes are found in all life forms, they consist of both highly conserved and variable regions (which can be exploited by PCR) and are rarely subject to horizontal gene transfer (Hugerth and Andersson, 2017).

1.1.2. Amplicon DNA sequencing

Recent advances in sequencing technologies, such as the development of next-generation sequencing (NGS) platforms, have made it possible to perform rapid, sensitive and high-throughput analysis of microbial communities, which can generate extremely large data sets at relatively low costs (Sboner et al., 2011). The use of NGS platforms to sequence specific regions within phylogenetic marker genes (such as the 16S/18S rRNA genes) has become a routine practice in microbial ecology, and is known as amplicon (DNA) sequencing (Lundberg et al., 2013). In amplicon sequencing, DNA is extracted from an environmental sample and specific primers are designed that target conserved regions in a gene of interest. These regions are subject to PCR amplification and 150-300 bp amplicons are sequenced in parallel on an NGS

platform (such as an Illumina MiSeq) that produces millions of DNA reads. Sequencing data is processed, quality controlled and individual reads are clustered together based on sequence similarity to provide an insight on taxonomy. A popular approach for taxonomic clustering groups sequences that are >97% identical in sequence similarity, which is termed as an Operational Taxonomic Unit (OTU) and acts as the 'species' level in taxonomic classification. Taxonomic information of OTUs is obtained by querying OTU DNA sequences with one of a number of databases that provide reference taxonomic information on a range of sequences, such as the Ribosomal Database Project (RDP), the SILVA database or Greengenes (Cole et al., 2014, Quast et al., 2013, McDonald et al., 2012). After OTUs have taxonomic annotations, sequence data is summarized in a table that contains the number of read counts per sample per OTU, which can be used to create a community profiles of each sample at a desired taxonomic level. Amplicon sequencing has been successfully used in a plethora of studies to understand the diversity and composition of microbial communities, which range from analysing the global topsoil microbiome, to characterising the human gut microbiome (Bahram et al., 2018, The Human Microbiome Project et al., 2012).

A number of factors make amplicon sequencing a desirable choice for molecular microbial studies. Firstly, a selection of large taxonomic databases exists for commonly used amplicons (16S rRNA, 18S rRNA, 23S rRNA, ITS) and taxonomic annotations of DNA amplicons can often reach both the species or genus level, which can be used to identify rare taxa that are not detectable using traditional approaches (Huggerth and Andersson, 2017). Such databases include the SILVA database, which contains over 5 million sequences that encode small ribosomal subunits (16S rRNA and 18S rRNA) and 700,000 sequences that encode large ribosomal subunits (23S rRNA and 28S rRNA); the Greengenes database, which combines taxonomic annotations from the NCBI database with cyanoDB and has over 400,000 full-length 16S rRNA sequences for bacteria and archaea; and the RDP database, which includes over three million 16S rRNA gene sequences and >60,000 28S rRNA and ITS sequences (Cole et al., 2014, Quast et al., 2013, McDonald et al., 2012). However, whilst a number of large taxonomic databases exist, annotation success can vary drastically between different taxa and studies (Wen et al., 2017, Almeida et al., 2018).

Secondly, several open-source and user-friendly pipelines are available for amplicon sequencing, making bioinformatics and statistical analysis relatively fast and simple. For example, Quantitative Insights Into Microbial Ecology (QIIME) is an open-source and highly customizable workflow that can be used for every stage of analysis in amplicon sequencing, starting from filtering raw reads directly obtained from sequencing, to performing normalisation techniques (such as rarefaction), clustering sequences based on similarity (for example into OTUs), assigning taxonomy and visualising data to generate figures (Caporaso et al., 2010). These pipelines therefore make it possible to perform amplicon sequencing analysis with little prior knowledge of bioinformatics. Thirdly, by multiplexing barcodes of numerous samples during library preparation, amplicon DNA sequencing can be a cheap molecular method to obtain microbial community profiles (Kozich et al., 2013). Depending on the desired sequencing depth and the sequencing technology used, amplicon sequencing can work out as little as \$1/sample (\$1/10,000 reads), but standard costs range between \$5-\$30/sample (Glenn et al., 2019, Meek and Larson, 2019). As alternative modern molecular approaches, such as deep-sequencing shotgun metagenomics can cost ~\$50 per 0.5 gigabases of data, amplicon DNA sequencing is therefore a sensible approach to analyse microbial communities for large experimental data sets with strict budgets (Thomas et al., 2012).

Whilst presenting a large amount of benefits, amplicon sequencing has several drawbacks and limitations that should be considered before designing an experiment to use this approach. Firstly, different taxa often contain a variable copy number of marker genes within their genome. For example, a number of archaeal and bacterial taxa have several copies of the 16S rRNA gene, with some taxa reaching up to 15 copies of the gene; therefore amplicon DNA sequencing will lead to an overrepresentation of these taxa in community profiles (Louca et al., 2018, Lee et al., 2009). A study by Větrovský and Baldrian (2013) highlighted the variability in the copy number of the 16S rRNA gene between taxa, in which they deciphered that the relative abundance of Firmicutes was artificially overestimated in amplicon data, whereas the relative abundance of Acidobacteria was significantly underestimated. Secondly, amplicon sequencing relies heavily on the use of PCR for both initial amplification and library preparation, which is vulnerable to a number of problems

and biases. For example, PCR can preferentially amplify template sequences of specific taxa, such as those that exhibit low GC nucleotide base pair content, therefore resulting in their biased amplification and overrepresentation in taxonomic abundance (Krehenwinkel et al., 2017). Moreover, further issues of PCR include misincorporation of specific bases by Taq polymerase, the formation of secondary structures in template strands and the production of chimeric molecules, which can all result in amplification bias amplification of specific PCR templates that can impact taxonomic profiles (Becker et al., 2000, Wang and Wang, 1997, Eckert and Kunkel, 1991). Thirdly, the approach to normalise amplicon sequencing data remains a highly controversial topic, and no one universal method is used across different studies. More specifically, one highly debated topic is whether to normalise sequencing depth across samples, a process known as rarefaction. In this process, the same number of reads are randomly sampled from the total population of reads within each sample, and only selected sequences are used for downstream analysis (Gotelli and Colwell, 2001). A study by McMurdie and Holmes (2014) opened the debate of rarefaction on amplicon sequencing data, suggesting that rarefaction leads to an increased abundance of false positives (that relative abundance of specific taxa is significantly different between treatments) and encouraged avoiding any kind of rarefaction approach on amplicon data. However, Weiss et al. (2017) later suggested that increases in differential abundance false positives were not due to rarefaction, and rarefaction actually lowers the false discovery rates in data sets where large discrepancies in library size are seen between samples. There currently remains no agreement as to whether rarefaction is a suitable approach for normalising amplicon sequencing data, and consequently the literature remains split on the issue. Furthermore, the performance of sequencing clustering methods has additionally been a heavily debated area in amplicon sequencing analysis. Traditional clustering of sequences into OTUs (based on a 97% sequence identity threshold) is often not stringent enough to distinguish between microbial species. For example Fox et al. (1992) identified that several *Bacillus* species share a 99.5% sequence similarity, and OTU clustering would group these species together. However, recent studies have addressed these issues and have managed to separate individual species (and further at the strain level) using new sequencing clustering algorithms (such as DADA2), which can discriminate sequence differences of a single nucleotide (Callahan et al., 2016). Fourthly, as amplicon DNA

sequencing only profiles phylogenetic marker genes (such as the 16S rRNA), very little functional information can be inferred from this approach. Whilst some tools exist that can predict composite metagenomes based on 16S rRNA data, for example PiCRUST, such tools are heavily reliant on the availability of reference genome sequences and can only be used to infer information on bacterial and archaeal genomes (Langille et al., 2013). The last major limitation of amplicon sequencing arises from the inherent problem of the compositional data that is generated. Community composition obtained from amplicon sequencing data is normally presented as a relative abundance (the proportion of reads assigned to a given species compared to the total number of reads) of each species. But due to the nature of compositional data, it is difficult to infer from relative abundance data alone whether a taxonomic shift of a given species is due to a change in abundance of that species, or is due to shifts in the abundance of other species (Figure 1.1) (Mandal et al., 2015). However, rarefaction can help reduce artefacts in apparent differences in the relative abundance of a specific taxa between treatments, and additionally other complex algorithms have been produced, such as GNEISS and ANCOM, that address the problem of compositional data and claim to reduce false discoveries by considering abundance data across multiple samples (Morton et al., 2017, Mandal et al., 2015). Overall, despite having a number of drawbacks, amplicon sequencing represents a cheap, accessible and fast molecular approach to profile microbial communities.

1.1.3. Shotgun metagenomics

Shotgun metagenomics is a relatively new field in microbial ecology that aims to address the issues of analysing microbial communities using amplicon DNA sequencing. Instead of sequencing a portion of a ribosomal marker gene, shotgun metagenomics uses deep sequencing approaches to sequence the complete genetic material of all microbial genomes that are present in an environmental sample (Quince et al., 2017). The main benefit of shotgun metagenomics is that sequence data can be used to gain both a taxonomic profile, as well as a functional profile by determining the abundance of genes associated with specific pathways, which is not possible with amplicon sequencing data. Furthermore, by using shotgun metagenomics it is possible to obtain community profiles of bacteria, archaea, viruses and eukaryotes simultaneously from the same DNA sample. A comprehensive review by Quince et

al. (2017) summarised the workflow of a standard metagenomics study into five key steps (Figure 1.2): 1) experimental design and DNA extraction, 2) pre-processing/quality control of reads, 3) sequence analysis, 4) post-processing and normalisation and 5) validation of methods.

Figure 1.1. Taxonomic shifts in relative abundance between treatments. Community composition (relative abundance) of unrarefied mock data (numbers indicate read count). The increase in relative abundance of species D in between treatments 1 and 2 were due to increased read count of species D. However, increased relative abundance of species D between treatments 1 and 3 were due to decreased read count of species A-C.

Shotgun metagenomics begins with the same approach as amplicon sequencing, whereby the complete genetic material is extracted from an environmental sample. However, instead of amplifying specific target sequences during library preparation, the complete genetic material within the sample is subject to sequencing, and library preparation is normally performed using PCR-free based methods to reduce bias. After library preparation sequencing is performed on an NGS platform, which is frequently an Illumina system (such as the Illumina HiSeq 2500) as these platforms provide large data outputs (several terabytes of data per run) with low sequencing error rates (0.1-1%) (Pfeiffer et al., 2018). However, Illumina systems often result in read lengths of 150-300bp and recent studies are coupling Illumina results with real-time long-read sequencing technologies (such as the Oxford Nanopore MinION) to sequence at greater read lengths, which have recently been reported to reach > 1 million bp

(Nicholls et al., 2019, Shabardina et al., 2019). Selection of the correct sequencing platform can dictate the output of metagenomics results, as downstream analyses are heavily reliant on sequencing depth per sample.

Figure 1.2. Summary of a metagenomics workflow. The standard metagenomics workflow, as presented by Quince et al. (2017), can be split into 5 steps. 1) Experimental study design. 2) Computational pre-processing and quality control to remove sequencing bias. 3) Sequence analysis using assembly-free or co-assembly techniques. 4) Post-processing, normalisation and statistical approaches to determine differences between samples. 5) Validation of results. Figure obtained from Quince et al. (2017).

After DNA sequencing, filtering and quality control of reads is performed to remove low quality base pairs and adapter sequences, which is performed in a similar manner to amplicon DNA sequencing. Once high-quality reads have been obtained, there are two main approaches that are often performed: assembly-free metagenomic profiling, or metagenomic (co)assembly. The traditional approach, assembly-free metagenomic profiling, assigns reads to external sequence resources (with taxonomic and/or functional annotations) without performing any kind of assembly step (combining reads to generate longer contigs) (Segata et al., 2013). The abundance of mapped reads to reference sequences is then collated, normalised and information on gene abundance and taxonomy can then be inferred. A number of tools exist that can provide both a rapid and sensitive approach to annotate mapped reads, for example

Kraken or Centrifuge (Wood and Salzberg, 2014, Kim et al., 2016). The main appeal of this approach is that is not computationally intensive in comparison to assembly-based methods and can therefore speed up downstream bioinformatics whilst using a limited amount of resources. Furthermore, by using assembly-free based approaches it is possible to assign taxonomy to low-abundance taxa that cannot be assembled *de novo*, and can consequently provide an insight into rare taxa (Quince et al., 2017). However, despite the benefits of assembly-free metagenomic profiling, there are several limitations and drawbacks that should be considered. Assembly-free taxonomic profiling of short reads, which are usually 150-300 bp in length, provides a lower mapping resolution, where both highly-repetitive sequences and reads mapping to uncharacterised taxa are extremely difficult to assign (Ayling et al., 2019). Therefore, direct mapping of unassembled reads to reference databases can result in a low proportion of annotated reads within a sample (Petersen et al., 2017, Breitwieser et al., 2017). This can be exacerbated when examining complex microbial communities, for example those from soils, where there remains a lack of reference genome sequences in current databases (Quince et al., 2017). Another drawback of assembly-free profiling arises from the number of false positives of taxonomic assignments. For example, Anwar et al. (2019) concluded that assembly of metatranscriptomic (sequencing RNA instead of DNA) reads identified genes with only a 0.6% false-positive discovery rate, whereas assembly-free approaches resulted in a 15% false positive discovery. This clear disparity can result in overrepresentation of specific taxa that have better taxonomic annotations in current databases.

Conversely, metagenomic assembly is the process of combining short reads into larger contigs, which are used to map against reference databases that contain sequence annotations. The main advantages of metagenomic assembly is that longer DNA sequences are able to map to reference sequences at a higher accuracy, produce better mapping statistics and lead to a significantly lower number of false positives in comparison to assembly-free methods (Sangwan et al., 2016). Additionally, the larger DNA fragments (which are normally filtered to contain at least 1000 bp) that are generated from metagenomic assembly can encode for whole genes, or even gene clusters, meaning sequences can be examined with greater context. Furthermore, metagenomic assembly of reads can also correct sequencing errors as multiple reads

are used to align contigs, and can therefore improve the overall quality of sequences (Ayling et al., 2019). Consequently, if good genome coverage is obtained (for example >20X coverage), metagenomic assembly is often the preferred choice of metagenomic analysis.

(i) Make kmers

(ii) Build graph

(iii) Walk graph and output contigs

Figure 1.3. De Bruijn graph assembly. In de Bruijn graph assembly, (i) reads are broken down into different kmers using a sliding window size (k) across all reads. (ii) The kmers become vertices in the de Bruijn graph, and edges connect overlapping kmers. (iii) Contigs are produced by walking across different paths across the de Bruijn graph (made in step ii). Better coverage paths are normally taken, and low-coverage paths are avoided. Figure obtained from Ayling et al. (2019).

A number of tools exist to perform *de novo* metagenomic assembly, which are normally based on the de Bruijn graph approach. In this process, individual reads are broken up and aligned into overlapping sub-sequences of a fixed length, k , which are used to construct de Bruijn graphs (Quince et al., 2017, Pevzner et al., 2001). De Bruijn graphs are a multi-dimensional construct where each node represents a k mer (nucleotides of a specific length, k) and connections between nodes occur when the first nucleotide of one k mer is the same as the last nucleotide of another k mer (Figure 1.3) (Ayling et al., 2019). Metagenomic assemblers iteratively add each read one by

one to generate complex de Bruijn graphs that contain overlapping sequences of all reads in a metagenomic co-assembly. However, as metagenomic samples are inherently convoluted and contain sequences from an abundance of organisms, de Bruijn graphs can be incredibly complex where not one clear path or solution exists, and selecting the correct *k*mer length as well as the correct assembler is critical for a strong assembly (Quince et al., 2017). Furthermore, de Bruijn graph assemblers often struggle to assemble highly-repetitive sequences, but this is a common problem for other assembly approaches, such as overlap, layout, consensus assembly (Ayling et al., 2019). A number of different metagenomic assemblers exist and a comprehensive review by van der Walt et al. (2017) quantified the performance statistics of an array of assemblers that use different assembly techniques (Table 1.1).

Following metagenomic assembly and annotation of contigs, quality-controlled short-reads (generated from sequencing) from each sample are mapped onto the assembled contigs to provide context to the samples, which is usually presented as (mapped) coverage information. The extent of mapped coverage to each contig is then combined, normalised (for example to account for sequencing differences between samples) and gene abundance can be inferred from the output.

Table 1.1. Performance of metagenomic assembly of Tara Oceans metagenome. Statistics and table were generated and taken from van der Walt et al. (2017).

	Tara Ocean								
	CLC	IDBA-UD	MEGAHIT	metaSPAdes	MetaVelvet	Omega	Ray Meta	SPAdes	Velvet
Number of contigs (≥ 500 bp)	50,716	163,815	216,938	185,419	67,161	15,982	6128	220,178	57,816
Total length	46,069,409	179,686,756	210,621,485	202,770,058	55,972,515	34,861,819	7,277,214	275,920,632	45,425,460
No. of long contigs (≥ 1 kbp)	10,720	50,498	56,243	48,640	12,590	13,305	2179	70,711	8802
No. of ultra-long contigs (≥ 50 kbp)	0	2	1	37	0	9	0	54	0
Largest contig	39,748	101,400	62,649	141,519	30,177	102,255	41,443	197,381	21,980
<i>N50</i>	880	1166	982	1124	805	2691	1329	1415	749
<i>L50</i>	14,113	38,236	58,246	39,033	21,544	2737	1345	39,617	19,631
Mapping rate (%)	38.98	52.24	55.92	64.03	4117	13.64	8.25	64.46	48.19
Time (seconds)	3527	69,782	10,455	125,862	2527	168,213	16,419	80,039	2342
Time (minutes)	58.78	1163.03	174.25	2097.70	42.12	2803.55	273.65	1333.98	39.03
Time (hours)	0.98	19.38	2.90	34.96	0.70	46.73	4.56	22.23	0.65
Memory required (GB)	16.23	42.84	10.58	66.53	109.37	30.7	42	157.75	109.37

A number of studies have extended the metagenomic assembly pipelines beyond the assembly of short reads. Despite generating large contiguous segments of DNA, metagenomic assembly can still result in the production of highly fragmented sequences that are from a variety of different microorganisms (Quince et al., 2017). Metagenomic binning aims to group high-quality contigs into specific taxonomic

groups, ideally at the species level. In an ideal situation, each metagenomic ‘bin’ will contain a number of high-quality sequences which could be used to create a partial or complete assembled genome, termed a metagenomic assembled genome (MAG). A variety of methods and different tools exist to perform metagenomic binning, for example CONCOCT or MetaBAT, which can taxonomically cluster groups of sequences based on their sequence properties, such as GC content (Alneberg et al., 2014, Kang et al., 2015). By using metagenomic binning approaches to reconstruct nearly 8,000 MAGs, Parks et al. (2017) discovered 20 new archaeal and bacterial candidate phyla that were not previously detected using alternative approaches, expanding our current knowledge of the microbial tree of life. However, current tools that perform metagenomic binning are computationally intensive, and high-quality sequences with good coverage (>20X) are required to create good quality bins. Therefore, in data sets that have used a small number of samples for co-assembly, metagenomic binning will lead to the production of low-quality bins and may not be a good approach for analysis.

Functional and taxonomic annotation of contigs can be performed by querying a number of databases that have annotations of specific sequences. Taxonomic annotations can be curated from the NCBI taxonomy database, which provides standard nomenclature and classification of all organisms currently in the public sequence database (which is estimated as 10% of global species) (Federhen, 2012). Alternatively, the non-redundant BLAST protein database, which contains all non-duplicate sequences from Genbank, RefSeq, NCBI, Swissprot, PIR, PDF and PDB, is a comprehensive database that is commonly queried for taxonomy, which can be performed using a BLASTX or DIAMOND (an alternative aligner) search (Buchfink et al., 2014, Pruitt et al., 2007). For functional annotations, the Kyoto Encyclopaedia of Genes and Genomes (KEGG) database is popular to obtain functional annotations of predicted protein sequences (Kanehisa et al., 2015). The database consists of three components for metagenome annotation: 1) a collection of annotated gene catalogues for complete genomes, 2) a knowledge database of high-level functions and 3) a molecular-level functional knowledge database of known functional associations (Kanehisa et al., 2011). KEGG annotations can therefore be used to gain insights on complete functional pathways (for example nitrogen metabolism), specific reactions

(nitrogen fixation) or individual genes (*nifH*). The protein families (Pfam) database consists of a large collection of annotated functional protein domains that are represented using multiple sequence alignments and hidden Markov models (HMMs) (Finn et al., 2014). The Pfam database is often desirable as it is generated using HMM-profile analysis, whereby sets of known protein sequences are aligned and used to train models that can search for matching domains, allowing for annotations of previously unannotated taxa (Sonnhammer et al., 1998). Alternatively, HMM-profiles can be trained manually from a small set of known protein sequences, which can then be used to query other larger databases (such as the NCBI non-redundant protein database) for matches. Cluster of Orthologous Groups (COG) is also a popular functional annotation database, which consists of orthologs produced through comparisons of predicted and known protein sequences in fully-sequenced microbial genomes (Tatusov et al., 2000). Predicting orthologous groups is a desired approach as current gene products remain poorly characterised, for example one of the best studied microorganisms, *Escherichia coli*, only has experimental annotations on ~40% of its gene products (Koonin, 1997). However, it is predicted that ~70% of microbial proteins are highly conserved across different taxa, therefore predicting orthologous groups represents a good approach to bridge this gap (Tatusov et al., 2000, Koonin et al., 1997).

Whilst metagenomics has been revolutionary in the field of microbial ecology, for example in the discovery of new candidate phyla, there are a number of limitations and drawbacks that should be considered before using this approach. Firstly, metagenomics is substantially more expensive in comparison to amplicon DNA sequencing. For example, in this experiment sequencing costs were ~\$450 (USD) per sample (including library preparation) for a targeted sequencing depth of 30 gigabases per sample on an Illumina HiSeq 4000 platform. In comparison, by multiplexing samples amplicon DNA sequencing costs can range between \$5-20 per sample (depending on the desired sequencing depth) and therefore makes it more accessible in large-scale experiments (Glenn et al., 2019). However, cost estimates can vary dramatically between different sequencing companies, and sequencing costs are rapidly lowering with time (Sboner et al., 2011). Secondly, metagenomic taxonomic and functional databases are poorly annotated and large portions of the microbial tree of life are currently missing (Quince et al., 2017). Randle-Boggis et al. (2016)

additionally showed that a trade-off exists between metagenomic annotation accuracy and taxonomic resolution (using several pipelines), within which they presented that the number of false positive annotations substantially increased with a greater taxonomic resolution. Therefore, whilst analysis at lower taxonomic levels (such as the genus or species level) may provide deeper biological insights into metagenomic communities, only a small proportion of metagenomic reads are usually accurately annotated at these levels, and analysis at higher taxonomic levels is often desired due to a greater confidence in annotation accuracy. Although taxonomic and functional databases will become better annotated with time, which will be exacerbated by a reduction in sequencing costs, current metagenomic studies are limited by the annotation databases available. In comparison, for amplicon DNA sequencing a number of taxonomic sequence databases exist for an array of both small and large ribosomal subunit sequences (see section 4.1.2). However, much like amplicon sequencing pipelines, there is no one universal method to analyse metagenomic data, and large disparities exist between different workflows, such as the type of assembly tools that are used or the methods to normalise data, which can provide different outcomes from the same data set. For example, Randle-Boggis et al. (2016) found large differences in the performance of sequence annotation (such as the number of false positive annotations and proportion of annotated reads) using four different annotation tools (MEGAN, MG-RAST, One Codex and Megablast) on the same simulated metagenomic data set. Lastly, whilst a variety of pipelines and tools exist for metagenomic analysis, the majority of them require a substantial amount of prior knowledge (both computational and biological knowledge) and are not considered as 'entry level' tools. Furthermore, metagenomic tools that are currently available are incredibly computationally intensive and can take extremely long periods of time to run (several weeks or months) whilst requiring state-of-the-art computer hardware to do so. For example, for metagenomic assembly of short reads most assemblers require roughly a 1:1 ratio of file size to computational random access memory (RAM) in order to produce contiguous sequences (van der Walt et al., 2017). Therefore, to perform large metagenomic co-assemblies of deeply sequenced samples, it is not uncommon to require up to several terabytes of RAM. A study by van der Walt et al. (2017) highlighted the high computational requirements of metagenomic assembly, where metagenomic assemblies took up to 47 hours and required up to 157.75 GB

RAM to assemble <300,000,000 bp (0.3 gigabases) using state-of-the-art computers. Overall, the number and performance of tools available for metagenomics has dramatically improved over recent years, and with an increasing number of studies incorporating metagenomics, sequence databases and pipelines are likely to improve.

1.1.4. Advances in ecological theory

As well as advances in technological tools, ecological theory has progressed substantially in microbial ecology over the past decade. Recently, a conceptual framework was proposed by Vellend (2010), which condensed a range of ecological processes that determine the composition and diversity of communities into four categories: 1) Selection, 2) Ecological drift, 3) Speciation and 4) Dispersal. In short, selection is defined as an increased fitness due to adaptation due to biotic and abiotic pressures; drift is caused by stochastic differences in birth and death rates causing shifts in population structure; speciation is the genesis of new species caused by a lack of dispersal amongst populations; and dispersal is the movement of individuals between populations across time and space (Vellend, 2010). In microbial ecology, a call for use of such a universal framework has been invited by a number of studies and considerable effort has been made to align Vellend's framework with microbial communities (Hanson et al., 2012, Chase et al., 2011, Stegen et al., 2013). Using null models, Stegen et al. (2012) proposed a model that quantifies the extent that each of Vellend's stochastic (ecological drift and dispersal limitation) and deterministic (selection) community processes have in shaping microbial metacommunities (temporally or spatially separated communities) (Stegen et al., 2013, Leibold et al., 2004). This model has since been used in a vast number of studies, which has furthered our understanding of the ecological processes that drive microbial communities in a host of ecosystems (Danczak et al., 2018, Martínez et al., 2015, Chai et al., 2016).

1.2. Soil ecosystems

1.2.1. The importance of soil

Soil is a vastly important and unique natural resource that performs irreplaceable roles required for sustaining the human population. It is estimated that 95% of global food production either directly (consumable crops) or indirectly (secondary consumables, e.g. livestock feed) originates from soil, all of which is currently

being produced from a combined area that occupies just 11.06% of total terrestrial land (FAO, 2015, FAO, 2019). However, with conservative estimates suggesting that the human population will reach 9.1 billion by 2050, FAO reports have predicted that global food production will need to increase by ~70% (from 2005) to support the growing population, most of which will be sourced from soils (Bergeret, 2017). Declines in soil fertility and health, for example by soil erosion (which is estimated to cost £0.9 – £1.4 billion per annum in the United Kingdom alone), threatens the prospects of soils being able to function correctly in order to meet both current and future demands (Syers, 1997, Graves et al., 2015). As a result, there is rising pressure to increase agricultural efficiency to produce higher yields on a decreasing area of available arable land.

1.2.2. Soil health

Soil health is an important concept in soil science, and it refers to the general condition and quality of soil as a natural resource (Kibblewhite et al., 2008). Historically, soil health was viewed from an agricultural perspective and was determined based on soil fertility and crop yield, which is usually dictated by the organic carbon content or bioavailability of nutrients. However, more recent concepts of soil health are assessed on the capacity of soils to perform soil functions (see section 1.2.3) which are beneficial for humans, plants and animals. A recent schematic proposed by Ritz (2014) (Figure 1.4) identified five branches of soil health: 1) Organic matter, 2) Water and Air Space, 3) Soil structure, 4) Nutrients and 5) Biota; all of which co-interact and are reliant on one another to perform soil functions. As a result, in order to fully assess soil health, it is important for current studies to target as many branches of the soil health schematic.

Figure 1.4. Soil health schematic. Soil health is composed of 5 components all of which influence and interact with one another. Schematic proposed by Ritz (2014).

1.2.3. Soil ecosystem functions

Historically soil functions have been poorly defined, but a clear definition was first proposed by Blum (1990), in which he defined soil functions as the benefits that humans, animals and plants obtain from soils. Blum (1990) further proposed six major categories of soil functioning: 1) the production of biomass, 2) acting as an environmental filter and buffer, 3) acting as a gene bank, 4) technical infrastructure, 5) a source for raw materials and 6) a source of historical information. Since then, a number of frameworks have been proposed to assess soil functions further, but one popular framework was proposed by the Millennium Ecosystem Assessment (MEA) launched in 2000, which was an intergovernmental call by the United Nations to obtain an updated synthesis of knowledge on human impacts on the earth's ecosystems (Reid et al., 2005). The project categorized ecosystem services (the benefits humans obtain from the environment) into 4 groups: 1) supporting ecosystem - processes that are required for the production of other ecosystem services, 2) regulating services - the benefits that humans obtain from the regulation of ecosystem services, 3) provisioning services - products or goods obtained from ecosystems and 4) cultural services - non-material benefits obtained from ecosystems (Reid et al., 2005). Soils are one of the few ecosystems that have been shown to fit every category of the MEA

framework (Table 1.2); for example, they perform pivotal roles in biogeochemical cycling (supporting), climate regulation (regulating), production of biomass (provisioning) and have heritage value (cultural), which emphasizes their importance to the human population.

Table 1.2. Ecosystem services obtained from soil. Based on the Millennium Ecosystem Assessment framework, adapted from (Jónsson and Davidsdottir, 2016)

Supporting	Regulating	Provisioning	Cultural
<ul style="list-style-type: none"> • Maintenance of biodiversity • Nutrient cycling • Water cycling 	<ul style="list-style-type: none"> • Biological control of pests and diseases • Climate regulation • Carbon storage and greenhouse gas regulation • Hydrological control • Recycling of wastes and detoxification • Filtering of nutrients and contaminants 	<ul style="list-style-type: none"> • Production of biomass (for use as food, wood, fuel, fibre) • Provision of clean water • Production of raw materials (e.g. topsoil, turf, clay and peat) • Physical environment for infrastructure 	<ul style="list-style-type: none"> • Heritage (geological and archaeological archive) • Recreation (ecotourism, sports) • Cognitive (landscapes)

1.2.3.1 Biodiversity in soil

Soil is considered to be the most biodiverse ecosystem on the planet, and it is estimated to account for as much as 25% of global diversity (George et al., 2019, M. Hart et al., 2017). Soil biota range from broad groups of microorganisms, such as archaea, bacteria, fungi and protists, to higher-level organisms including nematodes, earthworms and arthropods (Wall et al., 2015, Fierer and Jackson, 2006). Several studies have highlighted this enormous biodiversity, estimating that 1 gram of terrestrial soil can contain up to 200 metres of fungal hyphae and 1 billion bacterial cells, which can encompass a range from 2000 to 8.3 million different bacterial species (Roesch et al., 2007, Bardgett, 2005, Wardle, 2006). However, whilst it is known that soil contains a rich microbial biodiversity, the vast majority of these organisms remain uncharacterized, and therefore the functions of individual strains and species is mostly unknown.

1.2.3.2 Soil biogeochemical cycles

Soil ecosystems are major contributors to global biogeochemical cycles, and act as both a significant source and sink for a variety of nutrients. Soil contains a large range of carbon-fixing organisms, which are estimated to fix between 0.6 to 4.9 Pg (10^{12} g) carbon globally per annum (Yuan et al., 2012). However, global carbon emissions from soils, primarily by microbial respiration, are estimated to be 20-35 times higher than global carbon fixation from soils, resulting in a net annual release of soil carbon (Miltner et al., 2005). For example, from carbon dioxide (CO_2) emissions alone, it is estimated that soils emit 68-75 Pg CO_2 per year, which is the equivalent to approximately 10% of atmospheric CO_2 being cycled through soils every year (Raich and Potter, 1995, Raich and Schlesinger, 1992). Additionally soils act as a large carbon store, and conservative estimates suggest that soils currently store 1400-1600 Pg of carbon, which equates to roughly three times more than the organic carbon held in phytomass, and twice as much carbon that is held in the atmosphere (Scharlemann et al., 2014, Raich and Potter, 1995). However, the stability of long-term soil carbon stocks, especially in wetland and peatland areas, is under threat from anthropogenic-associated elevated atmospheric CO_2 , which will potentially increase soil carbon emissions through accelerated respiration and decomposition rates associated with a rise in temperature (Davidson and Janssens, 2006). Conversely, it is also possible that with elevated atmospheric CO_2 , both soil and plant carbon fixation rates will dramatically increase, leading to increased storage in these systems.

Nitrogen compounds are essential for all organisms as they are required for the synthesis of proteins and nucleic acids. Soil ecosystems are major determinants of the global nitrogen cycle and it is estimated that global soil nitrogen stocks range between 90-140 Pg - although forms of soil N are dynamic and under constant change (Batjes, 2014, Zinke et al., 1984). In terrestrial ecosystems, N-availability is often a limiting factor for plant productivity, and plants rely on soil nitrogen-fixing microorganisms to convert inaccessible dinitrogen to bioavailable forms (Chapin, 1980). It is estimated that between 40-290 Tg N is fixed in soil ecosystems annually, which equates to 10-60% of all annual global fixation (Fowler et al., 2013). The conversion of ammonium to nitrate (nitrification), which is the main N source for the majority of land plants, is

also mainly performed by soil microorganisms in terrestrial systems (Tischner, 2000). However, as nitrates are much more soluble than ammonia, the conversion to nitrates can often lead to leaching and eutrophication and an overall loss of nitrogen in the system (Van Der Heijden et al., 2008). Soil microorganisms also mediate the conversion of nitrates back to gaseous nitrogen compounds (dinitrogen or nitrous oxide) by nitrate reduction (assimilatory and dissimilatory) and denitrification, and it is estimated that agricultural soils are responsible for as much as 60% of global nitrous oxide emissions (IPCC, 2007).

Soils also contribute to major roles in a host of other biogeochemical cycles, which include the phosphorus, sulphur, potassium, magnesium and calcium cycles. On top of this, soils also host a range of micronutrients, which along with the previously listed biogeochemical cycles, are all essential for plant growth.

1.2.4. Soil physics

Soil science has traditionally consisted of three main domains: soil biology, soil chemistry and soil physics. Soil physics is a longstanding discipline within soil science that examines both the physical processes and properties of solid, liquid and gaseous particles within soil, which can have large impacts on both soil biology and soil chemistry (Lin et al., 2005, Miedema, 1997). The foundations of soil physics stems from the works of Schlübler, Schumacher and Wollny in the nineteenth century, who in separate studies examined soil metrics such as bulk density, water-holding capacity and electrical conductivity of agricultural soils, initiated the characterization of the movement of water and air through soil, and examined the influences of plant growth and management on soil properties (Lal, 2009). Since then, soil physics has progressed from a discipline focusing on small-scale and simple laboratory studies, to one that examines more complex systems and dynamics using integrative approaches that link to soil health (Lin et al., 2005). For example, soil physics has recently been applied to investigate the effects of soil management practices on the distribution of organic matter, examine the extent of soil compaction and erosion under various treatments, and model the distribution and transport of water and chemicals through the soil profile on the field scale (Kim et al., 2015, Rensburg, 2010, Teixeira et al., 2014,

Nawaz et al., 2013). New integrative approaches that utilize both soil physics and soil biology are providing new insights on biological functioning of soils.

1.2.4.1. Techniques in soil physics

Traditional techniques employed in soil physics centred around using morphological descriptions and manual measurements (for example soil moisture and bulk density) of thin sections or subsamples of soil (Lin et al., 2005). Information from these studies would then be incorporated into various mathematical models, for example ones to predict the movement of water through the soil profile (Schaap and Lebron, 2001). However, whilst an array of core principles in soil physics have derived from using such techniques, traditional methods are low-throughput, invasive and require destructive sampling in order to be performed, which can often result in the context of the samples being lost (Grevers et al., 1989). In recent years, technological advances have made it possible to use non-invasive techniques to measure a plethora of internal features and soil properties *in situ*, providing a more relevant context on the natural environment (Ketcham and Carlson, 2001). Examples of such techniques include X-ray computed tomography, soft-X-ray, nuclear magnetic resonance and gamma-ray tomography. X-ray computed tomography (CT) has emerged as one of the favoured non-invasive techniques to measure physical properties of soil, as the method excels in working on highly porous materials, such as soil. In this process, X-ray beams are passed through a sample at different angles over 360° and charged coupled device (CCD) detectors measure the attenuation of the X-ray signals from multiple angles through the sample (Figure 1.5).

Figure 1.5. X-ray computed tomography schematic. An X-ray source is pointed at the sample (green) which rotates. Attenuation information is then picked up by a detector. Image obtained from Mangalassery (2013).

Once analysed from a number of angles, high speed computers are capable of reconstructing detection signals to generate a three-dimensional projection - which can either be analysed in three dimensions, or separately in 2D horizontal or vertical cross sections, known as slices (Figure 1.6) (Pierret et al., 2002). Using this method, it is possible to extract information on the size, shape, distribution and arrangement of different solid materials, such as soil particles or mineral grains, and air spaces (known as pores) on an extremely fine resolution (reaching $<1\ \mu\text{m}$) on intact samples. This technique has been transformational in soil science and has been applied to a wide range of studies, which have included examining the effects of soil compaction, tracking the structural development of the rhizosphere and determining the structural impact of tillage on the soil profile (Périard et al., 2016, Helliwell et al., 2017, Abdollahi et al., 2013). Furthermore, X-ray CT scanning has been used to obtain insights on biological processes in soil, including predicting the distribution of microbial habitats and examining the distribution of macrobiota such as earthworm burrow systems or insect distribution (Taina et al., 2008).

Figure 1.6. *X-ray computed tomography images of a soil core. A: XY-oriented slice. B: ZX-oriented slice. C: A reconstructed 3D image from all slices. X-ray CT images were taken from a soil core obtained in July 2017 from a field growing a winter wheat crop.*

1.2.4.2. Soil pores

One of the major components in soil physics concerns the abundance and distribution of the voids between solid soil particles, known as soil pores (Gupta et al., 2008). Soil pores consist of complex heterogeneous mixtures of nutrient-rich liquid (such as water, chemicals, plant exudates) and gaseous (including carbon dioxide, oxygen, nitrogen and methane) substrates, which can result in the production of drastically different environmental conditions at the micrometre scale in the soil pore profile (Figure 1.7) (Rieu and Sposito, 1991, Nortcliff et al., 2006, Ruamps et al., 2011). Soil pores have major impacts on physical processes as well as biological functioning by influencing the distribution of macro and microbiota within soil. For example, the presence of soil pores can influence the activity and directionality of plant root growth, which can in turn determine the distribution nutrient-rich root exudates in the soil environment, shaping the spread of microbiota (Alaoui et al., 2011). Furthermore, in well aerated soils, the presence of soil pores can help distribute oxygen throughout the soil profile, resulting in larger aerobic and saturated zones that can promote microbial soil respiration (Fan et al., 2014). Conversely, in poorly aerated and waterlogged soils (such as peatlands), anaerobic environments are generated that can promote different microbial processes such as methanogenesis or denitrification (Munch et al., 2007, Rezanezhad et al., 2016, Mustamo et al., 2016). A study by Ruamps et al. (2011) highlighted the large impacts of soil pores on microbial communities, indicating that different types of soil pores contain significantly different microbial community compositions, suggesting that microbial biogeography can occur at the pore scale (micrometre scale).

Figure 1.7. Schematics of soil pore structure. Illustrations of different components that make up soil pores under varying moisture contents. Images obtained from (Lajos, 2008).

1.2.4.3. Soil pore metrics

1.2.4.3.a. Soil porosity (ϕ)

Several metrics have been developed to examine and classify pore size, structure, shape, distribution and abundance within soil, all of which provide different functional insights into soil pores. Soil porosity is the most common soil metric, and it is calculated as either a percentage or proportion (0-1) of the total measured volume that is occupied by soil pores (which is the inverse of bulk density, a measure of soil compaction) (Lawrence and Jiang, 2017, Saini, 1966). Soil porosity is generally considered a strong indicator of soil health, where more porous soils have better infiltration and drainage of water and nutrients, which leads to a greater and a more homogenous distribution of organic matter (Pagliai and Vignozzi, 2006, Arriaga et al., 2017). Soil porosity can heavily dictate the distribution and activity of microbial communities; for example, in poorly aerated soils microscale habitats can be produced that vary drastically in physical conditions (such as geochemistry and water content), which results in heterogenous niches in the soil profile that are occupied by different microorganisms (Wilpiseski et al., 2019). Soil porosity is often measured in sections, for example in selected regions of interest or whole soil compartments such as the rhizosphere, as soil is considered as considered a continuum of solid, liquid and gaseous materials that readily interchange (Eringen, 1994, Nimmo, 2004). The

porosity of different compartments has been shown to vary substantially across different soil compartments (Armenise et al., 2018, Helliwell et al., 2017). Furthermore, porosity of soils can vary drastically depending on a plethora of other factors, such as soil type, organic matter content and management practices, but typical values range between 0.3 – 0.7 in agricultural soils (Nimmo, 2004). Values of soil porosity can additionally provide insight into the upper limit of volumetric water content (saturation) and gaseous content of soils, and can act as a proxy for soil complexity, where higher values of soil porosity often are indicative of more complex pore systems with a larger capacity for interaction with soil aggregates (Nimmo, 2004).

1.2.4.3.b. Pore categorization and size distribution

There are a number of soil pore classification methods which categorize soil pores based on their diameter. Whilst there is no universal agreement regarding the exact diameter of each pore size category, soil pores are normally classified into three main groups: micropores, mesopores and macropores (Table 1.3.) (Beven and Germann, 2013). Micropores are the smallest pore size, within which little movement of gases and liquids occur and they are generally considered as the optimal habitat in the pore environment for microorganisms (Luxmoore, 1981, Smucker et al., 2007). Conversely, macropores are the largest pore size, reaching up to several millimetres in length, and are capable of forming channels that connect different soil depths. They can dictate a number of physical processes, for example they are involved with rapid movement of water and solutes through the soil profile (but have low water retention), and can additionally provide an environment in which roots can readily proliferate (Luxmoore, 1981, van den Akker and Soane, 2005, Alaoui et al., 2011, Yang et al., 2018). The intermediate pore size category, mesopores, are considered to perform a mixture of both micro and macropore functions – acting as either water-retaining or transporting pores (depending on their shape), as well as microbial habitats (Beven and Germann, 2013, Schoonover and Crim, 2015). Analysing the number of pores that fall into each size category therefore provides a proxy for determining the functionality of a soil sample.

Table 1.3. Classification of micropores, mesopores and macropores based on diameter.

Reference	Micropore (μm)	Mesopore (μm)	Macropore (μm)
Jongierius (1957)	< 30	30-100	> 100
Marshall (1959)	< 30	N/A	> 30
Brewer (1964)	5-30	30-75	> 75
McIntyre (1974)	< 0.3	0.3 – 30	30 – 300
Luxmoore (1981)	< 10	10 - 1000	> 1000

1.2.3.3.c. Pore connectivity

Soil pores can combine to create complex networks and channels that are distributed across soil horizons and throughout the soil profile (Figure 1.8). The connectivity of soil pores is highly sensitive to physical processes, for example several studies showed that greater soil compaction can lead to greater micropore, but lower macropore soil pore connectivity (Lipiec et al., 2009, Richard et al., 2001). Pore networks have complex shape characteristics, such as tortuosity, as they can consist of individual circular pores connected by cylindrical channels, which makes connectivity a difficult measurement to investigate. A number of methods exist to calculate soil pore connectivity, but the most popular method was proposed by Vogel (1997) in which soil pore connectivity was quantified as a function of the minimum pore diameter within a pore space, determined using the Euler-Poincaré characteristic (connections in 3D space) of each pore. In other words, it is a measurement of the number of connections that each pore (of a defined diameter) has to other pores within a specified volume. Higher soil pore connectivity is associated with increased transport of water and nutrients throughout the soil profile occurs, which can lead to better distribution of organic matter (Fei et al., 2008). Furthermore, soil pore connectivity has also been suggested to have a significant role in shaping microbial community structure, for example Carson et al. (2010) suggested that low soil pore connectivity is associated with higher bacterial diversity in soils due to lower amounts of dominance. Low soil pore connectivity can additionally lead to isolated pore structures in the soil profile, which can have different environmental niches (Wilpiseski et al., 2019). Therefore, soil pore connectivity can have large impacts on both biotic and abiotic functions of soil.

Figure 1.8. X-ray computed tomography reconstruction displaying soil pore connectivity. A) Connectivity of soil pores (grey particles indicate soil pores). B) Connectivity of soil particles (white particles indicate soil particles). X-ray CT images were taken from a soil core obtained from a wheat field in Wellesbourne (UK) in July 2016. The 3D projection was reconstructed using ImageJ. Dimensions of the region of interest: 2 cm x 2 cm x 3 cm.

1.2.3.3.d. Pore circularity

Soil pore circularity is a measurement of the roundness of individual soil pores in three-dimensional space, where values range between 0 – 1. A value of 1 indicates that a soil pore is perfectly spherical, whereas values towards 0 represent irregular non-spherical three-dimensional structures. Morphology and circularity of soil pores can provide an insight both physical and biological soil functions, for example circular soil pores have a low ability to retain water and gases and are more effective in the transport of water through the soil profile in comparison to irregular-shaped pores (Yang et al., 2018). Furthermore, higher soil pore circularity (0.8-1.0) can be indicative of gaseous vesicle formation from microbial respiration or metabolism (Yonovitz and Drohan, 2009). Therefore, a higher proportion of spherical pores can indicate a greater extent of microbial activity.

1.2.5 Soil compartmentalization

The soil horizon is comprised of numerous compartments, each of which exhibit distinct structural and chemical properties, and are often referred to as the ‘spheres’ of soil. Prior studies have categorized soil spheres into (at least) five main functional groups: the drilosphere (soil influenced by earth worms), the porosphere (pore spaces in the soil profile), the detritosphere (soil influenced by plant residues), the aggregatusphere (soil influenced by solid aggregates) and the rhizosphere (soil influenced by plant roots) (Figure 1.9) (Berg, 2012). Additional frameworks also exist that include a higher number of soil compartments, such as bulk soil (soil not

associated with a specific biota or plant roots), the myrmecosphere (soil associated with ants) or the termitosphere (soil associated by termites) (Brown et al., 2000).

Figure 1.9. Soil compartmentalization. A schematic proposed by Berg (2012) of different compartments within soil.

1.2.5.1 The rhizosphere

By far the most studied compartments in soil science are the rhizosphere, which is defined as soil that is closely adhering and interacting with plant roots, and the bulk soil, which is soil that is not associated with any plant roots (Daly et al., 2015, Raaijmakers et al., 2009, Mendes et al., 2013, Hiltner, 1904). Plant roots are responsible for the exudation of a variety of nutrient-rich compounds into the rhizosphere, which include amino acids, carbohydrates, hormones, proteins and secondary metabolites (Huang et al., 2014b, Walker et al., 2003). A study by Jones et al. (2004) highlighted the vast nutrient availability of the rhizosphere, in which it estimated that as much as 5-10% of plant net C fixed is lost to root exudates. As a result, rhizosphere biota are adapted to survive in environments with high nutrient input, which results in a distinct community composition in comparison to bulk soil (Mendes et al., 2013). For example, a number of studies have shown that rhizosphere

community composition is often less diverse than surrounding bulk soil, which may be attributed to a dominance of fewer species thriving off rich nutrient availability (Uroz et al., 2010). Microbiota of the rhizosphere have been shown to be beneficial for plant health in several ways. For example, rhizosphere microorganisms are known to perform significant roles in plant nutrient acquisition, such as nitrogen-fixing rhizobia in legumes, and suppression against a number of soil-borne pathogens (Hawkins et al., 2000, Pozo and Azcón-Aguilar, 2007). As well as displaying differences in microbial structure, it has also been reported that the physical structure of the rhizosphere is vastly different to bulk soil, where the rhizosphere has a greater proportion of large pore spaces, which are important for gaseous exchange, due to root growth, and the pH is often lower than in bulk soil due to exudation (Whalley et al., 2005, Shi et al., 2011).

1.2.5.2. Bulk soil

Bulk soil is often seen as a region that is as less nutrient-rich than the rhizosphere but is more biodiverse. However, it is known that bulk soil properties and community structure display large variation depending on a plethora of edaphic factors including soil type, pH and climate - which has been shown by a number of large-scale studies. A recent national-scale study by George et al. (2019) examined the composition of animal and microbial (bacteria, archaea, fungi and protist) communities from 436 soils from 7 different types of ecosystem located across the United Kingdom. They found that microbial community composition, but mainly species richness, was highly influenced by environmental properties and different land uses. This has been shown to extend to the global scale, a study by Bahram et al. (2018) examined bacterial community composition of 7,560 topsoil samples obtained from across the globe, in which they found that microbial gene composition varied dramatically with environmental parameters. As well as varying on a geographic scale, bulk soil varies considerably, both physically and biotically, with depth. With increasing soil depth, a number have studies have shown that the diversity of microbial species decreases, and that community structure shifts dramatically towards archaea and Gram-positive bacteria (Fierer et al., 2003, Tripathi et al., 2018). Additionally, physical properties of soil change with increasing depth, such as a reduction in soil porosity due to compaction, and an increase in soil organic carbon, making it a vastly different

environment relative to topsoil for microorganisms (Lorenz and Lal, 2005, Mahmoudi et al., 2015).

1.2.5.3. The soil surface

One vital soil compartment that is not as well studied as the rhizosphere or bulk soil, is the soil surface. The soil surface is a critical interface that connects the above ground environment with the soil ecosystem and can additionally dictate the extent that water, chemicals (pesticides, fertilizers) and nutrients may enter or leave the soil ecosystem. The soil surface has been previously described as an extreme environment due to the both the harsh environmental conditions it is exposed to, as well as the variability of these conditions (Csotonyi et al., 2010). The soil surface is unprotected and is therefore exposed to extremely high levels of ultraviolet radiation, which is known to irreversibly damage protein structure and stop DNA synthesis and replication (Rastogi et al., 2010). For example, soil surfaces in the Moab (Utah, USA) are exposed to excessive levels of UV light for ~4400 hours annually, yet wet conditions required for C-fixation, DNA repair and maintenance only exist <500 hours annually (Belnap et al., 2008). The soil surface additionally experiences extreme fluxes in temperature on a diurnal basis and is regularly subject to freezing and thawing cycles (Sharma et al., 2006). Freezing and thawing cycles can destroy and lyse microbial cells releasing an array of nutrients to surviving microorganisms, leading to a rapid increases in their biomass, thus dramatically changing microbial community composition (Sharma et al., 2006, Koponen et al., 2006). For example, freezing and thawing in agricultural soils have been shown to increase denitrifying populations due to an increased nutrient load – suggesting both a structural and functional change (Koponen and Martikainen, 2004). Additionally, the soil surface is also subject to a vast amount of dispersal, both as acquisition of nutrients, as well as the loss of nutrients through erosion. Studies have shown that dispersal of dust and other particulate matter to the soil surface can provide nutrients required for algae, mosses and lichens to form biomass and stabilize the soil surface (Belnap and Gillette, 1997). The soil surface is also under mechanical pressure, which mainly derives from rain water droplets applying physical pressure which can both decrease porosity and disperse soil particles, but also includes pressure from the

movement of a range of biota (Armenise et al., 2018, Souty et al., 1992). Therefore, in order to become established in the soil surface environment, a superfluity of strict environmental conditions must be overcome.

1.3. The Biological Soil Crust (BSC)

However, complex communities have adapted to withstand the extreme pressures found at the soil surface. The top few millimetres of soil is capable of forming a mesh-like microbiotic layer, known as the biological soil crust (BSC) or biocrusts, that are comprised of communities of bryophytes, autotrophic algae, cyanobacteria, fungi, bacteria, moss and lichens (Figure 1.10) (Belnap, 2003c, Bowker, 2007). Components and secretions of these organisms, which include fungal hyphae, cyanobacterial filaments, rhizoids of mosses and mucilage of algae bind tightly and intertwine with soil particles, which creates a physical barrier that separates the soil surface from underlying bulk soil (Veluci et al., 2006, Bu et al., 2015). These components stabilize the soil surface, providing an environment for the development of microbial heterotrophic communities, as well as higher plants.

Figure 1.10. Photographs of Biological Soil Crusts (BSCs). A) Soil crusts on soils growing a winter wheat field crop, Wellesbourne, UK (2016). B) Soil crusts from Utah, United States (Couradeau, 2013). C) Lichen crusts from Brandenburg, Germany Gypser et al. (2015)

1.3.1. Soil crust functions

Biological soil crusts have been shown to perform several critical soil functions that can improve soil health. Soil crusts perform integral roles in stabilizing the soil surface, which have been shown to drastically reduce the extent of wind and water erosion (Belnap and Lange, 2003, Warren, 2003). Mucilage produced by soil crusts can bind together small soil particles to create larger soil aggregates, which are much harder to displace than loose soil (Belnap, 2003a). A study by Bullard et al. (2018) found that the presence of cyanobacterial soil crusts can reduce the amount of soil lost by rain water erosion by as much as 80% in comparison to bare soils. Additionally, studies investigating the disturbance of soil crusts have shown that trampling and destruction reduces soil surface aggregation and promotes agricultural run-off and loss of soil (Kuske et al., 2012). Soil crusts have also been shown to perform critical roles in the hydration of soils. It has been reported that mosses and lichens in developed soil crusts can increase the ability of soils to retain soil moisture and maintain hydrated, which is especially important in arid ecosystems (Colica et al., 2014). Moreover, it has also been reported that soil crusts are highly influential in the transportation of water on the landscape scale, for instance a study by Faist et al. (2017) showed that less-developed soil crusts can help supply downslope vegetative patches with water in arid regions. However, conflicting studies have also suggested that soil crusts can produce hydrophobic layer that reduces water infiltration and may have detrimental impacts on soil health (Wei et al., 2015).

Soil surface biocrusts have been estimated to cover ~12% of global terrestrial ecosystems, and recent studies have suggested that they perform significant contributions to global biogeochemical cycles (Maier et al., 2018, Rodriguez-Caballero et al., 2018). Most notably, biological soil crusts make significant contributions to carbon and nitrogen cycling, due to their abundance of nitrogen-fixing and photosynthetic microbial populations, such as Cyanobacteria and green algae. A study by Elbert et al. (2012) estimated that nitrogen fixation by biocrusts (termed cryptogamic ground covers, which included biological soil crusts, rock crusts and bryophyte and lichen carpets) equated to roughly 26 Tg a^{-1} (10^{10} g), which is equivalent to just over 40% of total biological nitrogen fixation performed in global terrestrial

ecosystems (as estimated by the IPCC) (Ciais et al., 2013). However, a study by Weber et al. (2015) highlighted that biocrusts also contribute significantly to reverse reactions in the nitrogen cycle, such as the emission of reactive nitrogen compounds, mainly nitric oxide (NO) and nitrous acid (HONO), into the atmosphere. They estimated that global emissions of reactive nitrogen by biocrusts equated to $\sim 1.7 \text{ Tg a}^{-1}$ (1.1 Tg a^{-1} NO-N and 0.6 Tg a^{-1} of HONO-N), which is the equivalent of $\sim 20\%$ of global nitrogen oxide emissions from soils under natural vegetation as estimated by the 2013 IPCC report (Weber et al., 2015, Ciais et al., 2013). Together, these studies suggest that biological soil crusts accelerate the nitrogen cycle (Figure 1.11.) due to their large contributions to both atmospheric nitrogen fixation and emission. However, both of these estimates were based on soil crusts obtained from arid desert ecosystems, and extrapolation to fit other ecosystems (such as temperate ecosystems or peatlands), which have vastly different nutrient cycling dynamics, on a global scale may not be accurate.

Figure 1.11. Estimates of global contributions of biocrusts to the nitrogen cycle. Inner arrows represent global annual fixation of molecular nitrogen by biocrusts and emission of reactive nitrogen by biocrusts on the ground surface of drylands estimated by Weber et al., (2015). Outer arrows indicate global annual fixation of molecular nitrogen and emission of reactive nitrogen by soils under natural vegetation indicated by the latest IPCC report (Ciais et al., 2013). Figure obtained from Weber et al., (2015).

Elbert et al. (2012) also estimated that biocrust carbon fixation equates to roughly 3.9 Pg a^{-1} , which is the equivalent to 7% of global net primary production by terrestrial vegetation as according to the 2013 IPCC report (Ciais et al., 2013). A study by Huang et al. (2014a) modelled carbon fixation of algal and moss soil crusts over a 30 month period from January 2010 to November 2012 in the Tengger Desert, China. In this study, they concluded that more developed moss-dominated soil crusts exhibited higher levels of carbon fixation ($64.9 \text{ g C/m}^2 \text{ a}^{-1}$) in comparison to less developed algal-dominated soil crusts ($38.6 \text{ g C/m}^2 \text{ a}^{-1}$), which extrapolated to global contributions of $\sim 0.96 \text{ Pg a}^{-1}$ and $\sim 0.57 \text{ Pg a}^{-1}$ (respectively). Similar results have been shown in further studies, for example Li et al. (2012) revealed that cyanobacteria and

algae-dominated crusts exhibited lower carbon fixation rates ($11.36 \text{ g C/m}^2 \text{ a}^{-1}$) in comparison to developed lichen and moss-dominated soil crusts ($26.75 \text{ g C/m}^2 \text{ a}^{-1}$) from soil crusts also obtained from the Tengger Desert. The large capacity of carbon fixation by soil crusts could therefore act as a sink for anthropogenic-associated elevated atmospheric CO_2 , which have been estimated to mitigate as much as much as 3.6 tons of atmospheric $\text{CO}_2 \text{ ha}^{-1} \text{ a}^{-1}$ (Sears and Prithiviraj, 2012). However, relatively little work has been performed to investigate CO_2 emissions of biological soil crusts. One study by Dettweiler-Robinson et al. (2018), which examined the net carbon fluxes (by considering both respiration and photosynthesis) of arid soil crusts from grassland, shrubland, savanna and woodland sites in the New Mexico Elevation Gradient (NMEG), concluded that carbon fluxes actually resulted in a net-emission of CO_2 during summer periods in woodland and savanna sites. Furthermore, Angel et al. (2011) detected a production of methane, coupled with abundances of methanogenic taxa, in mesocosm experiments incubating soil crusts extracted from the Negev Desert, Israel. In addition, they did not detect the presence of any methanotrophic taxa, suggesting that all methane produced from soil crusts is released into the atmosphere, which may contribute to global methane emissions. Detection of methanogenic taxa, such as *Methanosarcina* and *Methanocella* have since been shown to occur in soil crusts from non-arid environments, such as glacial regions, which have a suggested role in the development of soil crusts (Aschenbach et al., 2013). Further work is required, on a range of ecosystems, to understand the global potential for biocrusts to act as either a carbon sink or source.

Biological soil crusts have also been proposed to perform significant roles in other biogeochemical cycles, beyond nitrogen and carbon. A number of studies have examined the contributions of biological soil crusts to phosphorus cycling, and have found that soil crusts had enriched concentrations of total and bioavailable phosphorus (P) in comparison to lower (bulk) soil (Wu et al., 2013, Pointing and Belnap, 2012). A recent study by Baumann et al. (2019), which investigated the roles of biological soil crust communities (obtained from sandy soils in Northern Germany) to phosphorus cycling, found that soil crusts perform a significant role in transforming inaccessible stable P into bioavailable (labile) P. Furthermore, Crain et al. (2018) investigated bioavailable P in biocrusts from the Chihuahuan Desert, and suggested

that soil crusts may prevent the adsorption of P to mineral surfaces, and promote incorporation into biomass. On top of this, the role of soil crusts in P cycling has additionally recently been examined in temperate soils. Glaser et al. (2017) identified a significant correlation of species richness of algal populations with inorganic phosphorus availability in soil crusts obtained from forest plots in Northeast Germany and suggested a potential role in P cycling of BSCs in this environment. Together, these studies suggest a significant contribution of biological soil crusts in biogeochemical cycling of P. Due to their widespread global abundance, it is likely that BSCs contribute to other global biogeochemical cycles, for example the sulphur cycle, which has recently been suggested by Rippin et al. (2018), but there is currently a lack of literature that explores their roles beyond N, C and P cycling.

1.3.2. The soil crust microbiome

Soil surface biocrust communities are both phylogenetically diverse and extremely dynamic, but display a high degree of variability over temporal and spatial scales (Heindel et al., 2019, Weber et al., 2016). Depending on an array of biogeophysical factors, biocrust communities can be dominated by a variety of species, ranging from microbial phototrophic colonisers, such as filamentous cyanobacteria, to higher organisms such as bryophytes and mosses in less perturbed environments (Garcia-Pichel and Wojciechowski, 2009, Belnap et al., 2003a, Lalley and Viles, 2008). However, despite showing a large degree of spatiotemporal variability in community structure, a number of studies have identified that biocrust communities almost always contain a large heterotrophic component, as well as the essential groups of primary producers (Dettweiler-Robinson et al., 2018, Maier et al., 2018).

1.3.2.1. Cyanobacteria

Primary producers are integral to the community structure of biocrusts, and perform vast roles in carbon and nitrogen fixation, as well as stabilizing the soil surface to allow development of other biota. Biocrust microbial primary producers are dominated by two main groups - cyanobacteria and microeukaryotic algae (Belnap, 2006, Frey et al., 2013). Büdel et al. (2016) categorized the major biocrust cyanobacterial species into three main groups: 1) Structural filamentous cyanobacteria, 2) opportunistic Cyanobacteria, and 3) Cyanobacteria that dispersed to

crusts by chance. Filamentous biocrust cyanobacteria, have been extensively studied and are functionally important in stabilizing the soil surface. Important members of this group include *Nostoc* spp (Figure 1.12.A), *Microcoleus steenstrupii* and the keystone species *Microcoleus vaginatus* (Figure 1.12.B), which have been shown to excrete polysaccharides that form an extracellular matrix that ‘glues’ together soil aggregates and particles (Mazor et al., 1996, Kuske et al., 2012, Couradeau et al., 2019, Belnap and Gardner, 1993). This extracellular matrix has been proven to drastically reduce soil erosion, as well as increase soil stability (Bowker et al., 2018).

Figure 1.12. Optical microscope images of biocrust cyanobacteria. A) *Nostoc commune* isolated from the Negev Desert, Israel. B) *Microcoleus vaginatus* isolated from the Negev Desert, Israel. C) *Chroococcidiopsis* isolated from the Western Cape region, South Africa. D) *Phormidium ambiguum* isolated from African soil. Images A-C obtained from (Büdel et al., 2016), image D obtained from (Chamizo et al., 2018).

Members from the second biocrust cyanobacterial group include opportunists such as *Chroococcidiopsis* (Figure 1.12.C) and *Scytonema*, which are adapted to survive in harsh environments (such as low N and moisture), and overall enhance the functional role of biocrusts through primary production and nitrogen fixation (Becerra-Absalón et al., 2019, Büdel, 1999). The final group includes cyanobacterial species that have dispersed to soil surface crusts, and often originate from other environments (Büdel et

al., 2016). Members of this group includes *Gleocapsa*, *Cylindrospermum* and more commonly a variety of *Phormidium* species (Figure 1.12.D), which are widely distributed and have been found in abundance in arctic, desert and aquatic ecosystems (Comte et al., 2007, Chamizo et al., 2018, Komárek et al., 2003).

1.3.2.2. Eukaryotic algae

Phototrophic microbial eukaryotic algae are also a vital component of biocrusts but are often low in abundance in comparison to cyanobacteria. Budel et al., (2016) also categorized biocrust eukaryotic algae into four functional groups: 1) crust-forming algae, 2) soil particle-attaching algae 3) free-living and lichen symbiotic algae and 4) algae dispersed other environments, normally of aquatic origin. Crust-forming filamentous algae are normally high in biomass and include *Zygonium*, *Apatococcus* and *Klebsormidium* species, which can create dense algal mats through a combination of excreting a fibrous mucilage layer and by binding their filaments to lichens and mosses (Mikhailyuk et al., 2014, Hoppert et al., 2004, Büdel et al., 2016). This mucilage layer has also been shown to increase hydration of crusts, allowing survival in desiccated environments (Karsten and Holzinger, 2014). In particular, *Klebsormidium flaccidum* (Figure 1.13.A) has been well characterised in soil crusts, and has been shown to tolerate drought, freezing and high intensity light exposure (Morison and Sheath, 1985, Nagao et al., 2008). Additionally, analysis of its genome has additionally shown that it produces a variety of plant hormones which may interact with higher plants (Hori et al., 2014). Soil-particle attaching algae include a range of diatoms, *Spongiochloris* (Figure 1.13.B) and *Neochlorosarcina* species, but this group normally only contributes to a small fraction of the community and contains a number of rare taxa (Büdel et al., 2009). A study in 2016 showed that whilst not abundant, diatoms had a disproportionally high species richness in cyanobacterial and algal-dominated crusts (Schulz et al., 2016). Symbiotic algae often form associations with lichens and fungi, whereby the fungal symbiont provides structural protection, while the algal symbionts perform photosynthesis (Root and Dodson, 2016). This can occur in a range of lichen species, including *Cladonia*, *Acarospora* and *Psora* species, whereas algal species include *Asterochloris*, *Stichococcus*, *Trebouxia* (Figure 1.13.C) and *Myrmecia* (Williams et al., 2017b, Büdel et al., 2016, Ruprecht et al., 2014). Lastly, opportunistic algae, including unicellular green algae such as *Chlamydomonas* (Figure 1.13.D), *Acutodesmus* and *Scenedesmus* can also be found in biocrusts (which

are normally abundant in aquatic ecosystems), but this normally occurs in biocrusts with high soil moisture.

Figure 1.13. Optical microscope images of biocrust algae. A) *Klebsormidium flaccidum*, scale bar indicates 10 µm (Hori et al., 2014). B) *Spongiochloris minor*, (Bischoff and Bold, 1963). C) *Trebouxia* (Silverside, 2019). D) *Chlamydomonas* (Denhartog, 1985).

1.3.2.3. Other primary producers and carbon fixers

Despite the main primary producers consisting of phototrophs, a number of studies have also isolated chemolithotrophic bacteria and archaea from biocrusts (Johnson et al., 2005, Delgado-Baquerizo et al., 2016). Of these, ammonia-oxidizing bacteria (including *Nitrosospira* and *Nitrosomonas*) and archaea (mainly Thaumarchaeota) are thought to play significant roles in nitrification through ammonia oxidation in desert soil crusts (Marusenko et al., 2013, Marusenko et al., 2014). But only a select number of studies have examined the capacity of these functions in soil crusts. Additionally, methanogenic archaea (but not methanotrophs), including *Methanosarcina* and *Methanocella* have been isolated from biocrusts from the Negev desert, Israel, suggesting that biocrusts may also contribute to the global methane cycle (Angel et al., 2011). Methanogenesis is coupled tightly with the Wood-Ljungdahl carbon fixation pathway in Archaea, however there are currently no studies that have examined carbon fixation by Archaea utilising this pathway.

1.3.2.4. Heterotrophs

Whilst phototrophs are the key component of biocrust communities, their stabilization of the soil surface makes it habitable by an array of heterotrophic microorganisms. A key study in community profiling of biocrusts was performed by Gundlapally and Garcia-Pichel (2006), within which they surveyed three 500m transects in the Colorado Plateau and identified 48 non-phototrophic genera (Gundlapally and Garcia-Pichel, 2006). Of these, genera belonging to the actinobacteria phylum dominated (37.5%), whilst Proteobacteria, Bacteroidetes, Thermomicrobiales and Acidobacteria genera were all well represented (>5%). However, advances in next-generation sequencing technologies have increased our understanding of soil crust communities, and many more heterotrophic organisms have been identified in biocrusts. For example, more recent studies have identified a range eukaryotic heterotrophs, which include fungal taxa such as Ascomycota and Hyphomycetes, as well as Protists and archaea (which are thought to make up ~5% of the prokaryotic communities), in biocrusts (Rippin et al., 2018, Soule et al., 2009).

1.4. Research gaps

Whilst a multitude of studies exist that have investigated soil crust community structure and functions, only a select number of these studies have used next-generation sequencing methods (Table 1.4). The vast majority of these studies have been performed in arid desert ecosystems - for example in the Negev desert (Israel), The Great Basin (US, Utah), The Chihuahuan Desert (USA/Mexico) and The Colorado Plateau (USA). Dryland and arid desert ecosystem top soils are dominated by sandy soils (90-95%), which have the extremely low soil organic matter C:N ratios (~9.9), low moisture content, low amounts of nutrient leaching, high salt content (such as calcium carbonate/phosphate) and consist of a large dust component (up to 30% of top soils) that readily disperses (Batjes, 1996, Reynolds et al., 2001, Dwevedi et al., 2017). These properties are contrary to soils found in temperate regions, which as a generalization are often naturally-fertile, rich in organic matter, experience a vast amount of precipitation and have high levels of drainage (although gross variation in all these factors exists in these regions) (Verheye, 2009). Therefore, it is unknown if current studies on arid soil crusts are applicable to temperate ecosystems.

Only a diminutive proportion of studies have examined soil crust community structure in temperate regions and as a result community development of soil crust biota in these ecosystems remains mostly unknown (Langhans et al., 2009). Prior studies that have used next-generation sequencing to examine community profiles of soil crusts in temperate regions have been performed in either artificial mesocosm experiments, have made no comparisons to bare or bulk soils, or have not considered the impact of time (Davies et al., 2013a, Davies et al., 2013b, Peng and Bruns, 2019, Williams et al., 2016). Furthermore, there are currently no studies that have utilised shotgun metagenomics to explore the functional potential of soil crust communities in temperate regions, and only a small number of studies have used this technique on soil crusts from other climatic regions (see section 4.1) (Steven et al., 2018, Steven et al., 2013a, Swenson et al., 2018). Consequently, further research is required to understand the development and functional potential of soil crust microbial communities in temperate regions, and to determine whether they follow similar trends to arid ecosystems.

Very little is also known about the ecological processes and the soil and environmental factors that drive soil crust community development. Whilst it is known that several soil and environmental factors are influential in the development of biological soil crusts, such as light and water availability, little focus has been made on the factors that determine microbial development of biological soil crusts. Instead, previous studies have typically focused on soil crust development based on morphological observations, such as crust coverage by moss and lichen species, which is solely descriptive (Belnap et al., 2003a, Belnap and Lange, 2003, Li et al., 2002, Heindel et al., 2019, Knapen et al., 2007). Furthermore, there are currently no studies that have applied the contextual framework proposed by Vellend (2010) on soil crust communities, and we lack an understanding of the relative contributions that each ecological process (selection, drift, dispersal, speciation) has in driving microbial community assembly of soil crusts from any climatic region.

In addition, a vast number of studies on biocrusts have been performed in undisturbed ecosystems that do not experience any major perturbation events. In these ecosystems, gradual successional development of communities of cyanobacterial and algal-dominated communities to moss and lichen-dominated communities can occur over several years, or even decades (Li et al., 2002), (Kuske et al., 2012). However in agricultural systems, soils are vulnerable to a vast amount of perturbations that can destroy soil structure, such as tillage, the use of heavy machinery, livestock tramping, crop harvesting and the application of chemical pesticides or fertilizers (Durham et al., 2018, Belnap, 2003b). With this in mind, very little research has been conducted to assess the viability of soil crust development in perturbed agricultural systems, despite their reported widespread abundance in these systems (Veluci et al., 2006). A small number of studies have assessed the effects of soil crusts in agricultural systems, which include their impacts on seed viability, plant biomass yield and soil erosion, but there currently are no studies that have used next-generation sequencing to track the temporal development of microbial communities in agricultural soils (Song et al., 2017, Gilbert and Corbin, 2019, Awadhwai and Thierstein, 1985).

Table 1.4. Overview of current literature examining community structure of biological soil crusts using next-generation sequencing methods. Literature searches were performed using the 'Web of Science' portal, and only studies that used next-generation sequencing methods to assess taxonomy were included.

Ecosystem	Location	Dominant taxa	Reference	Agricultural context
Arid	Great Basin, Chihuahuan Desert (USA)	<i>Microcoleus vaginatus</i> : 30-90% (Cyanobacteria)	(Couradeau et al., 2019)	None
Arid	Mojave Desert, California (USA)	<i>Phormidium</i> : 30-64% (Cyanobacteria)	(Mogul et al., 2017)	None
Arid	Colorado Plateau (USA)	Cyanobacteria: 20-30%	(Lee et al., 2016)	None
Arid	Chihuahuan Desert (Mexico)	Cyanobacteria: 45%	(Becerra-Absalón et al., 2019)	None
Arid	Oman	Cyanobacteria: 21%	(Abed et al., 2019)	None
Arid	Great Basin, Chihuahuan Desert (USA)	<i>Lyngbya nigra</i> : ~30-40% (Cyanobacteria)	(Bethany et al., 2019)	None
Arid	Gurbantunggut Desert (China)	Cyanobacteria: ~25%	(Zhang et al., 2016)	None
Arid	Almeria (Spain)	Cyanobacteria: 22-48%	(Williams et al., 2016)	None
Arid	Shapotou Desert (China)	<i>Bacillus</i> : 25-30% (Firmicutes)	(Liu et al., 2017a)	None
Arid	New South Wales (Australia)	Proteobacteria: 25%	(Chilton et al., 2017)	None
Arid	Northern Cape Province (South Africa)	Cyanobacteria: 5-35% Bacteroidetes: 10-20%	(Maier et al., 2018)	None
Arid	Utah (USA)	Cyanobacteria: 55%	(Kuske et al., 2012)	None
Arid	Utah (USA)	Cyanobacteria: 20-90%	(Johnson et al., 2012)	None

Arid	Utah (USA)	Cyanobacteria: 70-90%	(Steven et al., 2015)	None
Arid	Nevada (USA)	Proteobacteria: 20-30%	(Mueller et al., 2015)	None
Arid	South Africa	Cyanobacteria	(Venter et al., 2018)	None
Semi-Arid	Spain	<i>Microcoleus vaginatus</i> : 1-85% (Cyanobacteria)	(Muñoz-Martín et al., 2018)	None
Semi-Arid	Negev Desert (Israel)	<i>Microcoleus vaginatus</i> : 10-65% (Cyanobacteria)	(Nejidat et al., 2016)	None
Temperate	Ruine Homburg, (Germany) Gynge Alvar, (Sweden)	<i>Leptolyngbya</i> : 20-55% (Cyanobacteria)	(Büdel et al., 2014) (Williams et al. 2016)	None
Temperate	Pennsylvania (USA)	<i>Cylindrospermum</i> : 90% (Cyanobacteria)	(Peng and Bruns, 2019)	Soils sampled from an agricultural site
Temperate	Les Barges (Switzerland)	Cyanobacteria: 63-83% (23S rRNA) (<i>Nostoc punctiforme</i>)	(Davies et al., 2013a)	Soils sampled from an agricultural site
Alpine	Hohe Tauern, (Austria)	<i>Leptolyngbya</i> : 23-45% (Cyanobacteria)	(Williams et al. 2016)	None
Alpine	Reynolds Creek Critical Zone Observatory (USA)	Actinobacteria: 35-55%	(Blay et al., 2017)	None
Polar	Svalbard, Norway	Archaeplastida: 50-90% (Eukaryotes) Cyanobacteria: 30-70% (Prokaryotes)	(Rippin et al., 2018)	None
Polar	Svalbard, Norway	Synechococcales: 45-60% (Cyanobacteria)	(Pushkareva et al., 2015)	None
Polar	Ward Hunt Island, Canada	Cyanobacteria: 10-30%, Acidobacteria: 10-30%	(Steven et al., 2013c)	None

1.5. Project aims

It is unknown whether current studies on biocrusts are applicable to agricultural ecosystems containing temperate soils, and we currently lack an understanding of the soil structure, community composition and functional capacity of biocrusts in these ecosystems. This project aims to address these gaps by unravelling the dynamics and functioning of BSCs in a real-world agricultural environment. This leads to the following aims:

1. To decipher the temporal and local-scale spatial dynamics of microbial community composition at the soil surface in a temperate agricultural soil

- a. To determine whether the soil crust and bulk soil microbiome are compositionally different*
- b. To determine whether there are significant temporal dynamics in diversity and composition of the soil crust microbiome across the growing season*
- c. To examine local scale variation in soil crust community composition*
- d. To determine the relative influences of time, geographical distance and environmental factors on determining soil crust microbial community composition on a local scale*
- e. To compare co-occurrence networks of microbial communities in the bulk and surface soil to identify the extent at which key taxa are conserved across compartments*
- f. To quantify the relative contributions that each fundamental ecological process has in driving soil crust and bulk soil microbial community assembly*

2. To determine changes in the physical soil structure at the soil surface across a growing season in an agricultural field containing soil crusts

- a. To determine the depth at which the soil crust influences soil structure in the soil profile*
- b. To determine changes in soil pore abundance, distribution, connectivity and structure throughout the soil profile*

- c. *To determine changes in soil pore abundance, distribution, connectivity and structure across a growing season*
- d. *Investigate how annual trends in crust community development and community composition determined in aim 1 occur at a different study site*

3. To analyse developed biological soil crust and bulk soil metagenomes obtained from a temperate agricultural ecosystem

- a. *To compare taxonomic profiles of soil crust and bulk soil metagenomes*
- b. *To compare functional profiles of biogeochemical cycling genes between soil crust and bulk soil metagenomes*

CHAPTER 2 – THE TEMPORAL AND LOCAL-SCALE SPATIAL DYNAMICS OF SOIL CRUST AND BULK SOIL MICROBIAL COMMUNITIES

2.1. Introduction

2.1.1. Spatial variation

Soil crusts are extensively distributed in terrestrial habitats across the biosphere and recent estimates have suggested that they may cover as much as 12% of total surfaces on global land masses (Maier et al., 2018, Rodriguez-Caballero et al., 2018). They are capable of forming on every geographical continent and on a variety of soil types, ranging from hyper arid deserts to glacial forelands (Bowker et al., 2006, Belnap, 2003b, Yoshitake et al., 2010). However, whilst being broadly distributed, the morphology, physical structure and community composition of soil crusts can differ drastically depending on their geographic location. These differences are thought to be mainly driven by changes in climatic and environmental conditions such as temperature, light exposure, water availability, nutrient availability, plant cover and level of disturbance (Weber et al., 2016, Belnap, 2003b). For example, soil crusts found in dryland and arid ecosystems are often dark in appearance, only consist of a thin microbiotic layer, have a smooth microtopography and microbial communities are largely dominated by filamentous cyanobacterial taxa (up to 40% relative abundance) such as *Microcoleus spp.* (Becerra-Absalón et al., 2019, Kuske et al., 2012, Büdel et al., 2016, Garcia-Pichel et al., 2013). Conversely, soil crusts found in cooler and wetter climates are thicker three-dimensional assemblages, reaching up to several centimetres in height, and are more often dominated by higher organisms such as mosses, lichens and liverworts (Belnap, 2003b). Furthermore, in these climates cyanobacterial species normally contribute to a lower proportion of overall relative abundance (<10%) (Pushkareva et al., 2015). However, despite there being clear differences in community assemblages in soil crusts found in different climatic regions, their ecosystem functions in different regions have so far been shown to be similar (Williams et al., 2017a, Pointing and Belnap, 2012).

Soil crusts also exhibit large spatial variation over smaller geographic scales. Prior work in arid ecosystems have shown that soil crusts can have major differences in the composition of cyanobacteria, lichens and mosses despite being found just meters apart (Steven et al., 2013b). Local scale heterogeneity is a characteristic property of soil and a multitude of factors can determine variation in the community composition and functional capability at a given site (Strickland et al., 2009). In most habitats, soil crusts show high levels of patchiness in their distribution, where preferential development occurs in regions that are unperturbed and have consistent exposure to light (Kuske et al., 2012, Garcia-Pichel et al., 2003, Pajares et al., 2016). However, the extent of their development on a local scale is governed mostly by soil chemistry and physical factors such rainwater run-off, shading, light availability, microclimate, nutrient availability, distance from plants and physical disturbance history (Ullmann and Büdel, 2003). Combined, the formation of soil crusts in favourable areas additionally creates hotspots of biodiversity, as soil crusts stabilize the local soil environment and increase soil nutrient availability, which can attract higher plants and heterotrophic populations (Kleefeld et al., 2018). However, current literature assessing the local-scale spatial dynamics of soil crust communities is restricted to arid desert ecosystems, and we know relatively little about the about local-scale spatial heterogeneity in temperate ecosystems.

2.1.2. Temporal variation

As well as varying over spatial scales, soil crusts display a considerable amount of temporal variation. In arid ecosystems, soil crusts have been shown to follow a successional mode of development. In this process, initial communities are dominated by a few taxa (which are normally phototrophic taxa such as Cyanobacteria or green algae) that enhance soil quality and stabilize the soil surface, making it more habitable for a host of other organisms over time (Belnap, 2003b). For example, a study by Li et al. (2002), which tracked soil crust development in the Tengger Desert, China, between 1956 and 1999, categorized soil crust development into three developmental stages. Firstly, development of biological soil crusts began with the compaction of the soil surface by rainwater droplets and colonization by pioneer taxa, in this case the filamentous cyanobacterial species *Microcoleus vaginatus*. Secondly, over the period of 8-10 years, crusts became thicker in morphology and were dominated by increasing

abundances of mosses and lichens. Finally, after 25 years of undisturbed development, soil crusts reached a climax community governed by complex assemblages dominated by a variety of mosses, lichens and liverworts. The same developmental stages have since been shown to form in a variety of arid ecosystems (Veluci et al., 2006, Becerra-Absalón et al., 2019, Liu et al., 2017b). However, the temporal development of soil crusts is extremely variable between different geographical areas, and the characteristics and extent and duration of crust development to reach a climax community also depends on a number of soil and physical factors (Ullmann and Büdel, 2003, Bowker, 2007).

2.1.3. Perturbations of soil crusts

As mentioned prior (section 2.1.2) if left unperturbed (for example in polar, alpine or desert ecosystems) soil crusts will progressively develop until reaching climax communities that are dominated by mosses, lichens and liverworts (Belnap, 2003c, Li et al., 2002, Brankatschk et al., 2013, Liu et al., 2017b, Zhang et al., 2016, Frindte et al., 2019). However, soil crusts are extremely vulnerable to physical disturbances to the soil surface, and major perturbation events can destroy soil crust structure and consequently alter their community composition and functional capacity (Pietrasiak et al., 2011, Belnap, 2006, Faist et al., 2017, Belnap and Eldridge, 2003). For example, Steven et al. (2018) showed that metatranscriptomic functional profiles of soil crusts were chronically altered to have lower expression of nitrogen fixation genes after disturbance (trampling) events in the Moab Desert, Utah. Furthermore, frequent exposure to disturbance events can restrict soil crust communities to an early developmental stage, as early colonizing communities (such as cyanobacteria) are less sensitive to disturbance in comparison to climax community members such as mosses and lichens (Belnap and Eldridge, 2003, Harper and Marble, 1988).

In agricultural ecosystems, the soil surface is subject to a vast number of major perturbations, such as tillage, crop harvesting, trampling by livestock and the application of chemical pesticides and fertilizers (Durham et al., 2018, Belnap et al., 2003b). Whilst the impacts of all these factors on soil crust communities have not been explored, the use of heavy machinery and tillage can bury soil surface biota, and the application of herbicides is thought to kill crustal organisms - therefore these processes

will likely transform species composition of soil crusts (Belnap and Eldridge, 2003, Durham et al., 2018). However, despite the large number of disturbances to the soil surface in agriculture, soil crusts have been reported to form in agricultural systems across the planet (Sears and Prithiviraj, 2012, Bethany et al., 2019, Knapen et al., 2007). But in these systems, the microbial community composition, temporal development and ecosystem functions of soil crusts remain poorly defined.

2.1.4. Assembly and microbial interactions of soil crust communities

Community assembly concerns the processes that are responsible for determining the species composition within a given community. A popular conceptual framework was proposed recently by Vellend (2010) which categorised community assembly processes into four categories: natural selection, genetic drift, speciation and random dispersal (see section 1.1.4). Furthermore, Stegen et al. (2012) generated a null model that could be applied to microbial communities to quantify the extent that each of Vellend's processes have in driving community assembly. This has been a popular approach and has been used in a plethora of studies to further our understanding of the ecological processes that drive microbial community assembly in a host of ecosystems (Danczak et al., 2018, Martínez et al., 2015, Chai et al., 2016). However, this model is yet to be used on soil crust microbial communities, and we currently lack an understanding of the dominant processes that drive community assembly in soil crusts from any climatic region.

Interactions between different species are also important for determining microbial community composition. Microbial network analysis is an emerging approach that is used to assess the interactions that different organisms within a microbial community may have with one another. The approach correlates shifts in the relative abundance of different taxa to predict whether there are possible positive (commensalism or symbioses) or negative (parasitism or predation) ecological interactions between organisms (Hugerth and Andersson, 2017). Soil crust networks have been previously suggested to exhibit a distinct network topology, for example Chilton et al. (2017) presented substantially higher connectivity (number of positive or negative interactions between species) in developed soil crusts relative to undeveloped (bare) soils from arid regions in New South Wales. However, only a small number of studies

have analysed soil crust microbial networks, which have been performed exclusively in unperturbed arid ecosystems (Chilton et al., 2017, Karaoz et al., 2018). Furthermore, soil crust microbial networks have only been previously generated from bacterial communities (16S rRNA) and do not consider the effects of eukaryotes (such as protists or fungi), which have been suggested to largely contribute to soil crust community composition (States et al., 2003).

2.1.5. Chapter 2 aims

There are no studies in current literature that consider the spatio-temporal development of soil crust communities in temperate agricultural ecosystems. Prior studies that have investigated soil crust community dynamics have almost exclusively been performed in perturbed and arid ecosystems (see sections 1.4 and 1.5), and it is unknown whether these results are relevant beyond these systems. This chapter investigated this research gap, and had the following aims:

To decipher the temporal and local-scale spatial dynamics of microbial community composition at the soil surface in a temperate agricultural soil

- 1. To determine whether the soil crust and bulk soil microbiome are compositionally different*
- 2. To determine whether there are significant temporal dynamics in diversity and composition of the soil crust microbiome across the growing season*
- 3. To examine local scale variation in soil crust community composition*
- 4. To determine the relative influences of time, geographical distance and environmental factors on determining soil crust microbial community composition on a local scale*
- 5. To compare co-occurrence networks of microbial communities in the bulk and surface soil to identify the extent at which key taxa are conserved across compartments*
- 6. To quantify the relative contributions that each fundamental ecological process has in driving soil crust and bulk soil microbial community assembly*

From these aims, the following hypotheses were made:

1. *Community composition of soil crust and bulk soil communities are significantly different*
2. *Substantial shifts in soil crust community composition and diversity occur across a growing season*
3. *Local scale variation significantly impacts soil crust community composition*
4. *Environmental factors will have the larger influences on soil crust microbial community composition than distance and time*
5. *Co-occurrence network analysis will identify different key taxa in soil crust and bulk soil communities*
6. *Deterministic processes will dominate community assembly in soil crust communities, whereas stochastic processes will dominate community assembly in bulk soil communities*

2.2. Methods

2.2.1. Location, experimental design, sampling methods and core processing

Soil samples were obtained from Bradley's field (Figure 2.1) at the Warwick Crop Centre, Wellesbourne, United Kingdom (52.213833°N, -1.605683°E), growing a winter wheat crop. Previous crops grown on the sampling site were oil seed rape (2015-2016) and winter wheat (2014-2015). The sampling site was subject to 6-inch conservational tillage 2 weeks prior to the initial sampling date, and the crop was drilled the following week. Sampling occurred once a month, roughly every 30 days, for ten months (November 2016 – August 2017), which encapsulated the whole growing season. At each sampling date, a 40 meter transect was laid out along the field, which was separated into five 4m x 8m boxes (Figure 2.1). In these boxes, 5 cylindrical soil cores were taken (5 cm x 10 cm – diameter by height) in a 'W' shape to capture microsite variation. In order to avoid edge effects, transects were laid out at least 30 m from the edge of the field. The location of the 40 m transect was shifted by roughly 10 meters at each sampling date to avoid sampling in previously disturbed areas. Soil cores were then sealed in sterile plastic bags and were stored at 4°C until further processing on the following day. Alongside each core, a separate soil core (same dimensions) was taken to record bulk density measurements at each sampling location. Historical information on the weather conditions at the time of sampling were also recorded (Wellesbourne weather station, UK), and additionally the agricultural management practices (tillage and chemical application) throughout the year were recorded following discussions with the land manager (table 2.1).

In the lab, soil crusts (top 2 mm) and bulk soil (up to 2 cm in depth) were removed from each core. Soil crust and bulk soil samples were pooled within each of the five 4 m x 8 m boxes, creating 5 soil crust and 5 bulk soil samples each month (Figure 2.1). Pooled samples were sieved twice to 2 mm and stored at 4°C until further use. Subsamples from each sample were additionally stored at -80°C as a backup.

2.2.2. Chlorophyll α content

From pooled soil crust and bulk soil samples, chlorophyll α was extracted for use as a proxy for phototrophic biomass (Kidron et al., 2003). This was performed using adaptations of both Ritchie (2006) and Hansson (1988) methods, in which 1 g of soil

was shaken with 5 ml of 100% ethanol in 50 ml falcon tubes for 5 hours at 200 rpm and 20°C in the dark (to prevent chlorophyll photolysis). From this, the absorbance of 1 ml of the solution at 665 nm and 750 nm was measured using an Ultrospec 1100 pro UV/Visible spectrophotometer (GE Healthcare, UK). The solution was then acidified through the addition of 3M HCl - converting chlorophyll α to pheophytin (Moed and Hallegraeff, 1978). The absorbance of the acidified solution was again measured at 665 nm and 750 nm and chlorophyll α concentrations were calculated using equations given in (Castle et al., 2011, Lorenzen, 1967).

Table 2.1. Chemical application at the sampling site.

Date	Application	Amount
24/10/2016 - 28/10/2016	Flufenacet, Pendimethalin (Crystal®)	4 L Ha ⁻¹
	Diflufenican	0.1 L Ha ⁻¹
	Glyphosate	1 L Ha ⁻¹
10/03/2017	Granular Urea	225 Kg Ha ⁻¹
27/03/2017	Ammonium Nitrate	225 Kg Ha ⁻¹
04/04/2017	Ammonium Nitrate	250 Kg Ha ⁻¹

Figure 2.1. Sampling location and strategy. Left: Location of sampling site (Google maps). Right: Sampling and pooling strategy. Sampling was performed once a month between November 2016 and August 2017. Cores were taken at each circle and samples were pooled within each 'W'. Additional bulk density cores were taken at each circle.

2.2.3. pH

An adapted version of Emmett et al. (2008) was used to determine soil pH. In this process, 5 ml of ultrapure water was added to 2 g of soil and was shaken at 200 rpm for 30 minutes. After this, the solution was left to stand to allow the soil to settle. The

pH was then measured with a Mettler Toledo SevenEasy pH meter with a Mettler Toledo LE438 pH electrode (Mettler Toledo, Leicester, UK).

2.2.4. Soil moisture

Soil moisture was determined on pooled BSC and bulk samples. This was performed by adding 10 g soil into a beaker, which was left to dry overnight at 105°C. The weight difference between wet and dry soil was then calculated to determine the percentage soil moisture.

2.2.5. Microbial carbon and nitrogen content

A chloroform fumigation was performed to determine microbial biomass carbon and nitrogen in pooled BSC and bulk samples. This was performed following a protocol adapted from (Joergensen, 1996). 10 g soil was added to glass beakers and placed in a glass desiccator containing one beaker with 30 mL ethanol-free chloroform. Air was evacuated from the desiccator using a N 035.3 AN.18 diaphragm pump (KNF Neuberger, UK) until the chloroform boiled, at which point the air was vented and the process was repeated. On the third evacuation, the air was not vented and the vacuumed desiccator was then stored at room temperature for 3 days in the dark (to prevent breakdown of chloroform) to allow full lysis of microbial cells. After 3 days, 50 mL 0.5M potassium sulphate was added to fumigated and unfumigated (10 g of the same soil samples not incubated for 3 days with chloroform) soils, which were shaken at 125 rpm for 1 hour. Samples were then passed through type 1 Whatman filter paper prewashed with 0.5M potassium sulphate and the extracts were collected in 50 mL falcon tubes and frozen until analysis.

Total Carbon and Nitrogen content of potassium sulphate extracts were determined using a TOC-L Shimadzu analyser. Potassium sulphate extracts were briefly centrifuged after defrosting and were diluted 1:10 with ultrapure water. A standard curve for Non-Purgeable Organic Carbon (NPOC) and Total Nitrogen (TN) was generated from standard solutions made up of potassium hydrogen phthalate and potassium nitrate, respectively. For each diluted potassium sulphate extract, five 100 µl subsamples, alongside 0.5M potassium sulphate and ultrapure water controls, were injected into the TOC-L analyser and NPOC and TN concentrations were determined

using a standard curve. Control values were subtracted from each sample, and microbial NPOC and TN concentrations per gram of soil were calculated by subtracting unfumigated concentrations from fumigated concentrations.

2.2.6. Bulk density

Soil obtained from the separate cores at each sampling site were used to determine bulk density. Soil collected from the cores (10 cm x 5 cm x 10 cm) was removed, weighed and subsequently dried in an oven for 3 days at 60 °C until no moisture was present, and re-weighed. Bulk density was calculated over dividing dry soil weight by the volume of the core.

2.2.7. DNA extraction

DNA was extracted from 0.5 g of BSC and 0.5 g of bulk soil using the MPBio FastDNA™ SPIN Kit for Soil, according to the manufacturer's instructions. To ensure successful DNA extraction had occurred, 2 µl of DNA was loaded onto a 1% agarose gel and was run at 100V for 20 minutes. Smears of DNA were visualized under UV light using a transilluminator to confirm successful extraction. DNA was quantified using an Invitrogen Qubit® 2.0 broad range fluorometric quantitation protocol and was diluted to 5 ng/µl with ultrapure DNase free water.

2.2.8. Illumina MiSeq library preparation

PCR amplification was performed using the 515F and 806R primers to amplify the 16S rRNA V4 region of bacteria, Euk-A and Euk-570 primers to amplify the 18S rRNA V1-3 region of protists, fITS7 and fITS4 primers to amplify the ITS2 region of fungi and lastly the rVF1 and rVR1 to amplify the 23S region of phototrophs (Table 2.2). Sequences for these primers can be seen in Table 2.2. For compatibility with Illumina sequencing, each primer was modified at the 5' end with adaptor sequences (TGG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG for the forward primer, and GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G for the reverse primer).

The reaction volumes for each PCR were 25 µl and consisted of 3 µl of template DNA (5 ng/µl), 1.25 µl of forward primer (10 µM), 1.25 µl of reverse primer (10 µM), 12.5

μ l of Q5 Hot Start High-Fidelity DNA polymerase 2x Master Mix (New England Biolabs, Hitchin, UK) and 7 μ l of ultrapure DNase free water. This resulted in the following final reaction concentrations for each PCR: < 1000 ng of template DNA, 0.5 μ M of forward and reverse primers and a 1X concentration of Q5 Hot Start High-Fidelity DNA polymerase 2x Master Mix.

Table 2.2. Primer sequences used for amplicon DNA sequencing.

Target	Region	Primer	Primer sequence	Fragment size (bp)
Bacteria	16S (V4 SSU rRNA)	515f	5' GTCCCAGCMGCCGCGGTAA	291
		806r	5' GGACTACHVGGGTWTCTAAT	
Fungi	ITS (ITS2)	fITS7	5' GTGARTCATCGAATCTTTG	349
		fITS4	5' TCCTCCGCTTATTGATATGC	
Protists	18S (V1-3 SSU rRNA)	Euk-A	5' AACCTGGTTGATCCTGCCAGT	599
		Euk-570	5' GCTATTGGAGCTGGAATTAC	
Phototrophs	23S (LSU rRNA)	rV f1	5' GGACAGAAAGACCCTATGAA	410
		rV r1	5' TCAGCCTGTTATCCCTAGAG	

All PCRs were performed using a GeneAmp PCR System 9700 (Applied Biosystems, United States). Each PCR consisted of an initial denaturation step at 98 °C for 30 seconds. Following this, an amplicon-specific step comprised an initial denaturation step at 98 °C for 10 seconds, an annealing step (Table 4 for temperatures) for 15 seconds and an elongation step at 72 °C for 20 seconds. This step was repeated a different number of times depending on the amplicon (Table 2.3.). Lastly, a final elongation step at 72 °C for 5 minutes was performed. PCR products were stored at 4 °C until further use.

Table 2.3. PCR conditions for each amplicon

Amplicon	Number of cycles	Annealing temperature (°C)
16S	25	50
ITS	30	57
18S	30	57
23S	30	55

After amplification, PCR products were purified using AgenCourt AMPure XP beads (Beckman Coulter, UK) according to the manufacturer's instructions in order to

remove excess primers, nucleic acids and enzymes from the PCR products. In order to determine whether the PCR and purification step was successful, 5 μ l of the product was loaded into a 1% agarose gel and run at 100V for 10 minutes and viewed under UV light using a transilluminator. For unsuccessful samples, the PCR and purification steps were repeated.

Illumina sequencing adapters and unique dual indices were added to purified PCR products by a PCR using the v1 and v2 Nextera XT Index kit (Illumina, USA) according to the manufacturer's instructions. Each PCR contained 4 μ l of the purified PCR product, 2.5 μ l of the forward index primer (5 μ M), 2.5 μ l of the reverse index primer (5 μ M), 13 μ l Q5 Hot Start High-Fidelity DNA polymerase 2x Master Mix (New England Biolabs, Hitchin, UK) and 4 μ l of ultrapure DNase free water. Each PCR consisted of an initial denaturation step of 98 °C for 3 minutes, followed by 8 cycles of: 98 °C for 20 seconds, 55 °C for 15 seconds, and 72 °C for 15 seconds. This was followed by a final elongation step at 72 °C for 5 minutes. PCR products were compared to purified products (template DNA obtained prior to the index PCR) on a 1% agarose gel to determine whether indexing was successful. Indexed PCR products were then normalized to the same DNA concentration using the SequalPrep Normalization Kit, pooled and diluted to 4 nM. Pooled samples were then sent for 300 bp paired-end sequencing on an Illumina MiSeq system at the University of Warwick Genomics Facility, Coventry, United Kingdom.

2.2.9. Bioinformatics analyses

DNA sequences were automatically de-multiplexed by the Illumina MiSeq. Trimmomatic v0.35 was used to remove any low-quality bases at the ends of each read (Bolger et al., 2014). Forward and reverse reads were aligned using USEARCH (Edgar, 2010), which created paired-end reads. Primers were trimmed from sequences, and quality filtering was performed (-fastq_maxee = 0.5) to remove low quality reads using the following settings: minimum overlap between forward and reverse reads = 20 (0 for 18S rRNA), maximum mismatch = 3, minimum fragment length = 240, maximum primer mismatch = 2. Unique sequences were filtered by abundance, and any singletons were removed. A minimum identity threshold of 97% was used to cluster similar reads into an operational taxonomic unit (OTU), and chimeras were

filtered (-cluster_otus command). Quantitative Insights Into Microbial Ecology (QIIME 1.8) was used to assign taxonomy using the reference databases Greengenes and Unite for 16S rRNA and ITS amplicons, respectively, whereas the Silva reference database was used to assign taxonomy for both 18S rRNA and 23S rRNA amplicons (Caporaso et al., 2010, McDonald et al., 2012, Kõljalg et al., 2005, Quast et al., 2013). The abundance of each OTU, along with a taxonomic classification annotation, was generated using the make_otu_table.py function. For 16S rRNA and 18S rRNA, OTUs assigned to mitochondria or chloroplast were removed. Further filtering was performed on 18S rRNA OTU tables to remove any “metazoan”, whereas for 23S rRNA any OTUs assigned to non-phototrophic taxa were removed (which will further be referred to as ‘protist’ and ‘phototroph’ communities, respectively). All OTU tables were rarefied using a single random subsampling algorithm to the lowest number of reads across all samples for each amplicon, which was 8700, 10600, 1360, 1050 for 16S rRNA, ITS, 18S rRNA and 23S rRNA, respectively.

2.2.10. Statistical analysis

Alpha diversity, using Fisher’s method, was calculated on RStudio using the Phyloseq package (McMurdie and Holmes, 2013). Tests for normal distribution were performed using a Shapro-Wilk test, and statistical comparisons between treatments were performed using either a student’s t-test or a Wilcoxon signed-rank test (depending on normality) for pair-wise comparisons, or using two-way ANOVA (Tukey’s HSD multiple correction) or Kruskal-Wallis test (Dunn’s test) for multiple comparisons. Bray-Curtis similarity matrices from OTU tables were generated using the vegan package, which were used to create non-metric multidimensional scaling plots (NMDS) and to calculate an analysis of similarity (ANOSIM), which determines statistical dissimilarity between samples on the community level (ranging between -1 to 1) (Oksanen et al., 2018). Similarity percentages (SIMPER) analysis was performed to determine which phyla, class or OTUs were contributing the most to differences. Statistical differences between individual phyla, classes or OTUs between compartments were determined using a Wilcoxon signed-rank test, a Kruskal-Wallis (using Dunn’s multiple correction) for temporal comparisons. The deseq2 R-package was used to determine differential abundance of OTUs between soil compartments

using a Wald's test, which were plotted using the enhanced volcano package (Love et al., 2014).

Redundancy analysis (RDA) was performed to correlate metadata with changes in soil crust and bulk soil microbial community compositions. Before analysis, GPS locations of sampling sites were converted to principle coordinates of neighbour matrices (PCNM) and added as an explanatory variable to infer distance in RDA (Borcard and Legendre, 2002, Borcard et al., 2004, Dray et al., 2006). RDA was performed using environmental variables, including PCNMs, with forward selection (999 Monte Carlo permutations; FDR, $p < 0.05$), using relative abundance OTU tables for each amplicon (Peres-Neto et al., 2006). Bi-plots were generated with both significant and non-significant environmental variables.

2.2.11. Network analysis

OTU tables were first filtered by combining the upper-quartile of OTUs (based on read abundance) for each amplicon in soil crust and bulk soil samples. This resulted in 1818 bacterial OTUs, 483 fungal OTUs, 372 microbial eukaryote OTUs and 102 phototroph OTUs in soil crust samples; and 1804 bacterial OTUs, 485 fungal OTUs, 388 microbial eukaryote OTUs and 102 phototroph OTUs in bulk soil samples. Filtered OTU tables were then used with the SparCC algorithm with 20 iterations (Friedman and Alm, 2012). 999 bootstrap replicates were produced to calculate Pseudo p -values (corrected using Benjamini Hochberg) to infer significance of correlations. Networks were generated on significant correlations $>|0.7|$ (for compartmental networks) and $>|0.9|$ (for temporal networks), which were generated using the iGraph package on R-Studio (Csardi and Nepusz, 2005).

2.2.12. Community assembly processes

Microbial community assembly processes were determined according to the Vellend (2010) framework following a modified approach proposed by Stegen et al. (2012) (Stegen et al., 2013, Stegen et al., 2015). Sequences were first aligned and filtered in QIIME. Filtered sequences were used to create a maximum-likelihood phylogenetic tree using muscle algorithm. To determine whether there was significant phylogenetic signal (that is if phylogenetic distance correlates with temporal distance), the sampling

time at which each OTU had its maximum relative abundance was determined - this is referred to as the 'temporal niche' of an OTU. A temporal Euclidean distance matrix (absolute difference between temporal niche of each OTU) was produced using the 'picante' package in R and normalized between 0-1. Pairwise phylogenetic distance between OTUs was determined using the 'ape v.5.2' package in R (Paradis et al., 2004). Pairwise (between each OTU) phylogenetic distances were binned in intervals of 0.1, and the corresponding Euclidean temporal distance (for the given pairwise comparison) was placed in the given phylogenetic bin. Mantel correlograms were produced using the 'vegan' package on R between phylogenetic and Euclidean temporal distance matrices with 999 permutations (Oksanen et al., 2018).

Phylogenetic community turnover was determined by calculating β -mean nearest taxon index (β -MNTD), which is a pairwise measurement of the average phylogenetic distance to the closest neighbour. OTUs were randomly shuffled across the tips of the phylogenetic tree, and β -MNTD was recalculated 999 times to create a null distribution of β -MNTD. The β -nearest taxon index (β -NTI), which is the standardized measure of how far observed β -MNTD deviates from the null distribution of β -MNTD, was then calculated. Values of $|\beta\text{-NTI}| > 2$ indicate that observed species turnover (between a pair of communities) is due to environmental selection processes, $\beta\text{-NTI} > 2$ indicates variable selection (environmental selection between two samples is significantly different), $\beta\text{-NTI} < -2$ indicates homogenous selection (environmental selection between two samples are the same). For instances where no selection processes dominated ($|\beta\text{-NTI}| < 2$), stochastic processes were determined using the Raup-Crick algorithm using pairwise Bray-Curtis matrices (RC_{Bray}). RC_{Bray} values > 0.95 indicate homogenising dispersal (individuals are freely able to move between populations), $RC_{\text{Bray}} < -0.95$ infer dispersal limitation (a temporal or spatial barrier preventing movement of individuals across populations), and $|RC_{\text{Bray}}| < 0.95$ indicate drift (random changes through time of species abundance due to stochastic birth and death events). Pairwise β -NTI comparisons were made across all possible spatial and temporal comparisons within each compartment (within soil crust, and within bulk soil).

2.3. Results

2.3.1. Sample site temporal characteristics

Figure 2.2. Environmental conditions and soil properties. Average bulk density measurements (A) averaged across each sampling point across the transect. Cumulative monthly rainfall (B) average daily temperature (C) and humidity (D) obtained from the Wellesbourne weather station, UK, across the sampling period (days since first sampling event on 31/10/16). Temperature and humidity were fit with a non-linear model, red lines indicate times of sampling. Average soil moisture content (E) and pH (F) in pooled soil crust (green) and bulk soil (brown) samples each month. Error bars (shaded) indicate ± 1 standard deviation. Asterisks indicate a significant compartmental difference based on Tukey's HSD multiple-comparison test (after a two-way ANOVA).

2.3.1.1. Bulk density

Significant temporal variation in bulk density (Figure 2.2.A) was seen across the sampling period ($p < 0.0001$, one-way ANOVA). Average bulk density ranged from 1.04 (May) to 1.23 (August). However, no significant correlation (positive or negative) was detected between bulk density and time ($R^2 = 0.0019$, $p = 0.4825$).

2.3.1.2. Weather data

Temporal variation was seen in cumulative monthly precipitation (Figure 2.2.B). Monthly precipitation ranged from 8 mm (July) to 87 mm (June). Average daily temperature at the sampling site (Figure 2.2.C) showed large seasonal trends. The lowest temperature was 2.3 °C (December), whereas the highest temperature was 19.2 °C (August). Average daily humidity was above 74% on every day of sampling, and ranged from 74% (May) to 92% (November) (Figure 2.2.D).

2.3.1.3. Soil moisture content

Two-way ANOVA revealed that soil moisture significantly varied over time ($p < 0.0001$) and between compartments ($p = 0.0345$) and there was a significant interaction between time and compartment (Figure 2.2.E). Soil moisture had a peak value of 15.3% in December bulk soil samples, but had the lowest value (4.2%) in November soil crust samples. Multiple comparison tests (Tukey's HSD) showed only minor differences between moisture of crust and bulk soil with only significant differences occurring in November ($p = 0.0059$). No clear temporal patterns were seen across the time course, but soil moisture did show a strong positive correlation with monthly precipitation ($R^2 = 0.7930$, $p < 0.0001$, Figure 2.2.B).

2.3.1.4. pH analysis

Large variation was seen in the pH of both soil crust and bulk soil samples across the time course (Figure 2.2.F). Two-way ANOVA revealed that pH significantly varied over time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$). Across the time course, average pH varied by 1.11 pH units in soil crust samples, ranging from 7.57 (January) to 6.46 (May), and by 0.64 pH units in bulk soil samples, ranging from 7.50 (January) to 6.86 (June).

Multiple comparison tests (Tukey's HSD) revealed significant compartmental differences in March ($p < 0.0001$), April ($p = 0.0007$) and May ($p = 0.0034$) time points. This coincided with the application of 225 kg granular (urea one day prior to March sampling time point) and the application of 475 kg of ammonium nitrate (between March and April time points) (Table 2.1.).

Figure 2.3. Biomass carbon and nitrogen. Biomass carbon (A) and biomass nitrogen (B) in soil crust (green) and bulk soil (brown) samples each month. Error bars (shaded) indicate ± 1 standard deviation. Asterisks indicate a significant compartmental difference based on Tukey's HSD multiple-comparison test (after a two-way ANOVA).

2.3.2. Microbial biomass

Temporal development of biomass Carbon (biomass C) was seen in the soil crust (Figure 2.3.A). Two-way ANOVA revealed that biomass C significantly varied over time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$). Regression analysis revealed that biomass C had a strong, positive-correlation with time ($R^2 = 0.7321$, $p < 0.0001$) in the soil crust – within which average monthly concentration increased from 112.2 $\mu\text{g/g}$ in November to 303.4 $\mu\text{g/g}$ in August. However, no correlation between biomass C and time was found in bulk soil samples ($R^2 = 0.00018$, $p = 0.9258$). In biomass C, multiple comparison tests (Tukey's HSD) revealed significant compartmental differences in biomass C in April, June, July and August (April: $p = 0.329$, June-August: $p < 0.0001$).

Large variation was seen in biomass nitrogen (biomass N) across the time course in the soil crust (Figure 2.3.B). Two-way ANOVA revealed that biomass N significantly varied over time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was

a significant interaction between time and compartment ($p < 0.0001$). Regression analysis showed that biomass N in the soil crust had a positive correlation with time, but this analysis had a poor goodness of fit ($R^2 = 0.1671$, $p = 0.0032$). Unlike biomass C, in soil crust samples biomass N did not show cumulative increase with time, and there was a large spike in biomass N in March ($78.2 \mu\text{g/g}$), which gradually decreased over time before returning to pre-March concentrations in July ($15.4 \mu\text{g/g}$). A further spike biomass N occurred in August ($56.3 \mu\text{g/g}$). Biomass N did not correlate with time in the bulk soil ($R^2 = 1.8 \times 10^{-5}$, $p = 0.9761$). Multiple comparison tests (Tukey's HSD) revealed significant compartmental differences in biomass N in March, April, May, June, and August ($p = 0.0003$ in June, $p < 0.0001$ in March, April, May and August).

2.3.3. Chlorophyll α concentration and morphology

Average chlorophyll α concentration, a proxy for phototrophic biomass, increased progressively across the time course in soil crust samples (Figure 2.4.A). Two-way ANOVA revealed that chlorophyll α concentration was significantly affected by time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$). Chlorophyll α concentration displayed a strong, positive-correlation with time in soil crust samples ($R^2 = 0.8125$, $p < 0.0001$), but not in bulk soil samples ($R^2 = 0.008621$, $p = 0.5213$). Chlorophyll α concentrations ranged from 1.74 mg/m^3 (December), to 27.5 mg/m^3 (August) in the soil crust, and 0.71 mg/m^3 to 1.69 mg/m^3 in bulk soil samples. Significant compartmental differences were seen from January-August (January: $p = 0.0108$, February-August: $p < 0.0001$).

Additionally, clear shifts in the morphology of soil crusts were seen across time (Figure 2.4.B). No clear establishment of soil crusts were present in November, but the development of green algal layers can be seen by March. By August, the soil surface was dominated by a green biotic layer that had an abundance of mosses.

Figure 2.4. Chlorophyll α concentration and soil crust morphology. A) Average chlorophyll α concentrations in soil crust (green) and bulk soil (brown) samples at each month. Error bars (shaded) indicate ± 1 standard deviation. Asterisks indicate a significant compartmental difference based on Tukey's HSD multiple-comparison test (after a two-way ANOVA). B) Photographs of a random soil core from each month, taken on the day of sampling.

2.3.4. Rarefaction

Before any microbial community analysis, raw sequences were filtered and rarefied for each amplicon (Table 2.4.). Bacterial samples were assigned a total number of 2,188,198 reads, which resulted in an average 13,261 sequences per sample. Bacterial samples were rarefied to 8,700 sequences, resulting in a total of 7,316 OTUs. Fungal samples were assigned a total number of 2,862,482 reads, which resulted in an average 16,357 sequences per sample. Fungal samples were rarefied to 10,800 sequences, which had a total of 1954 OTUs. Protists were assigned a total of 1,315,052 sequences, which resulted on average number 12,892 reads per sample. Protist samples had a large amount of filtering applied (due to a high number of metazoan and fungi in the samples), and were rarefied to 1360 sequences per sample, with a total of 1,559 OTUs across all samples. Phototroph samples were assigned a total number of 1,560,524 sequences, which resulted on average 8,345 sequences per sample. Like protists, a high amount of processing was performed (to remove non-phototrophic organisms), and the samples were rarefied to 1,050 sequences per sample. From this, there was a total of 332 OTUs.

Table 2.4. Number of sequences assigned to each amplicon. Rarefaction across all samples for each amplicon was performed to the sample with the lowest number of reads, after processing to remove non-target DNA (chloroplasts and mitochondria for 16S rRNA, metazoan and fungi for 18S rRNA and non-phototrophic taxa for 23S rRNA).

Amplicon	Pre-filtering number of sequences	Average sequences per sample	Rarefaction level (number of sequences per sample)	Number of OTUs across all samples after rarefaction and filtering
16S rRNA (Bacterial)	2188198	13261	8700	7316
ITS (Fungal)	2862482	16357	10800	1954
18S rRNA (Protists)	1315052	12892	1360	1559
23S rRNA (Phototrophs)	1560524	8345	1050	332

2.3.5. Compartmental comparisons of soil crust and bulk soil microbial communities

2.3.5.1. Alpha diversity

Alpha diversity of bacterial, fungal, protist and phototroph communities were examined for soil crust and bulk soil samples across combined time points using Fisher's diversity index (Figure 2.5). Wilcoxon signed-rank sum tests revealed significant differences between compartments for bacterial, protist and phototroph ($p < 0.0001$ for all), but not fungal ($p = 0.2551$), communities. Fisher's alpha diversity was 1.18 (bacteria), 1.37 (protist) and 1.46 (phototrophs) times higher on average in the bulk soil in comparison to the soil crust.

2.3.5.2. Microbial community composition

Microbial community composition was visualized on non-metric multidimensional scaling (NMDS) plots generated using Bray-Curtis similarity (Figure 2.6). For all amplicons, stress values were between 0.1 and 0.2, suggesting a strong representation of reduced dimensionality. NMDS plots revealed a level of compartmental separation in all amplicons. Bulk soil samples displayed a relatively low level of separation between time points and mostly clustered together. However, in soil crust samples there was a higher degree of separation between time points (see section 2.3.6.2), suggesting community composition was more variable. Additionally, in all amplicons a number of soil crust samples clustered together with bulk soil samples, suggesting there was a group of soil crust samples that was compositionally similar to bulk soil.

Figure 2.5. Alpha diversity of soil crust and bulk soil microbial communities. Alpha diversity was calculated using Fisher's diversity index on combined soil crust (green) and bulk soil (brown) samples across all time points for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities. Asterisks indicate a significant result from a Wilcoxon signed-rank test ($p < 0.05 = *$, $< 0.01 = **$, $p < 0.001 = ***$).

Figure 2.6. NMDS plots of soil crust and bulk soil microbial communities combined across all time points. Non-metric multi-dimensional scaling plots for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities were generated based on Bray-Curtis similarity comparisons between samples.

Table 2.5. Analysis of similarity (ANOSIM) of soil crust and bulk soil OTUs. R values above 0.2 indicate significantly different community compositions.

Amplicon	R-value
16S rRNA	0.3165
ITS	0.1966
18S rRNA	0.3576
23S rRNA	0.438

2.3.5.3. Taxonomy

Analysis of similarity (ANOSIM) calculated from the relative abundance of OTUs in each sample revealed there were significant differences in community composition between soil crust and bulk soil samples for bacterial ($R = 0.3165$), protist ($R = 0.3576$) and phototroph ($R = 0.438$) communities (Table 2.5). ANOSIM revealed no significant difference between soil crust and bulk soil fungal communities, but the test statistic was close to the significance threshold of 0.2 ($R = 0.1966$). All ANOSIM values were robust ($p < 0.001$).

Figure 2.7. Soil crust and bulk soil microbial community composition combined across all time points. The average relative abundance (combined across all time points) of soil crust and bulk soil taxa contributing to >1% total relative abundance. Bacterial (16S rRNA) communities were examined at the phylum level, protist (18S rRNA) communities at the ‘ta2’ taxonomic level (phylum equivalent) and fungal (ITS) and phototroph (23S rRNA) communities were assessed at the class level (higher taxonomic levels displayed).

The relative abundance of microbial taxa (bacteria – phylum; fungi, protist and phototrophs – kingdom, phylum, class) contributing to >1% relative abundance was determined for each amplicon (Figure 2.7). In bacterial communities, Actinobacteria, Proteobacteria and Acidobacteria were the most abundant phyla in both compartments, contributing to 29.91%, 29.51 and 7.23% relative abundance in the soil crust, and 28.75%, 26.54% and 10.07% relative abundance respectively in the bulk soil.

In fungal communities, Ascomycota was the dominant phylum contributing to 76.47% of the total relative abundance in the soil crust and 74.04% in the bulk soil (Figure 2.7). The main fungal classes that dominated were Soradiomycetes, Dothideomycetes and Eurotiomycetes, contributing to 28.45%, 21.48% and 11.47% relative abundance in the soil crust, and 35.27%, 13.76% and 11.02% relative abundance respectively in the bulk soil.

In protist communities, Rhizaria (SAR), Amoebozoa and Stramenopiles (SAR) were the most abundant taxa at the phylum-equivalent level (ta2), contributing to 42.47%, 17.82% and 11.12% relative abundance in the soil crust, and 36.35%, 15.03% and 15.55% relative abundance respectively in the bulk soil (Figure 2.7).

In phototroph communities, Stramenopiles (SAR) and Viridiplantae were the most abundant taxa at the phylum-equivalent level (ta2), contributing to 38.77% and 36.36% relative abundance in the soil crust (respectively), and 21.10% and 51.20% relative abundance respectively in the bulk soil (Figure 2.7). The PX clade (Stramenopiles), Streptophyta (Viridiplantae) and Chlorophyta (Viridiplantae) were the dominant classes, contributing to 33.07%, 26.74% and 12.03% relative abundance in the soil crust, and 18.44%, 32.09% and 19.12% relative abundance in the bulk soil (respectively).

Log₁₀ fold change (L₁₀FC – the relative abundance of soil crust taxon divided by relative abundance of bulk soil taxon, transformed to log₁₀) was determined for taxa displaying significant differences (determined by a Wilcoxon signed-rank test) in the relative abundance between compartments (Figure 2.8). In bacterial communities,

nine out of twelve phyla (contributing to >1% relative abundance) had significant differences in the relative abundance between soil crust and bulk soil samples (Figure 2.8). Soil crust samples had a large, significant enrichment of Cyanobacteria, which was on average 14.17 times higher in the soil crust ($L_{10}FC = 1.168$, $p < 0.001$). Additionally, Bacteroidetes ($L_{10}FC = 0.111$, $p = 0.034$), and Proteobacteria ($L_{10}FC = 0.041$, $p < 0.001$) were also significantly enriched in the soil crust. Conversely, Nitrospirae had the highest fold-change of enriched bulk soil phyla, and was on average 1.61 times higher in bulk soil samples ($L_{10}FC = -0.207$, $p < 0.001$). Other significantly enriched bulk soil bacterial phyla included Acidobacteria ($L_{10}FC = -0.139$, $p < 0.001$), Verrucomicrobia ($L_{10}FC = -0.087$, $p < 0.001$), Chloroflexi ($L_{10}FC = -0.072$, $p < 0.001$), Planctomycetes ($L_{10}FC = -0.066$, $p < 0.001$), and Firmicutes ($L_{10}FC = -0.048$, $p = 0.037$).

In fungal communities, five out of nine classes (contributing to >1% relative abundance) displayed significant compartmental differences (Figure 2.8). Agaricomycetes ($L_{10}FC = 0.222$, $p < 0.001$) and Dothideomycetes ($L_{10}FC = 0.189$, $p < 0.001$) were significantly enriched in the soil crust, and were on average 1.67 and 1.55 times higher in the soil crust relative to bulk soil. In bulk soil samples, Mortierellomycotina ($L_{10}FC = 0.255$, $p < 0.001$) was on average 1.80 times higher in the bulk soil. Sodiariomycetes ($L_{10}FC = -0.093$, $p < 0.001$) and Tremellomycetes ($L_{10}FC = -0.043$, $p < 0.001$) were also enriched in the bulk soil relative to the soil crust.

In protist communities, four out of seven phylum-equivalent taxa (ta2 level) displayed significant compartmental differences (Figure 2.8). Rhizaria was 1.18 times higher in the soil crust than in bulk soil ($L_{10}FC = -0.072$, $p = 0.00319$). However, Centrohelida ($L_{10}FC = -0.123$, $p < 0.001$) was on average 1.33 times higher in the bulk soil relative to the soil crust. Alveolata ($L_{10}FC = -0.069$, $p < 0.001$) and Opisthokonta ($L_{10}FC = -0.047$, $p = 0.0073$) were also significantly enriched in bulk soil relative to soil crust.

In phototroph communities, eight out of nine classes had significant differences in relative abundance between compartments (Figure 2.8). Cyanobacteria (no class assigned, $L_{10}FC = 0.591$, $p = 0.0035$) and Oscillatoriales (Cyanobacteria, $L_{10}FC =$

0.469, $p < 0.001$) had the highest fold change in the soil crust, and were on average 3.89 and 2.94 times higher in soil crust samples relative to bulk soil. Additional classes significantly enriched the soil crust were Dinophyceae ($L_{10}FC = 0.336$, $p < 0.001$) and the PX clade ($L_{10}FC = 0.275$, $p < 0.001$). However, Alphaproteobacteria ($L_{10}FC = -0.296$, $p < 0.001$) was on average 1.977 times higher in the bulk soil relative to the soil crust. Chlorophyta ($L_{10}FC = -0.182$, $p < 0.001$), Nostocales ($L_{10}FC = -0.138$, $p = 0.0267$) and Streptophyta ($L_{10}FC = -0.070$, $p = 0.0191$) were also significantly enriched in the bulk soil relative to soil crust.

Figure 2.8. Differential abundance of soil crust and bulk soil taxa combined across all time points. Bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototrophic (23S rRNA) taxa (bacteria: phylum; fungi, protists and phototrophs: class) exhibiting significantly different abundance (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$) between soil crust and bulk soil communities. Relative abundance was transformed to fold change (relative abundance of given taxa in soil crust/bulk soil) and subsequently $\log_{10}$ transformed.

2.3.5.4. Soil crust-enriched OTUs

OTUs that were significantly enriched in the soil crust were identified using differential abundance analysis (Figure 2.9 and supplementary Tables 2.1-2.4) and SIMPER analysis (Table 2.6.). Differential abundance analysis identified OTUs which had the highest (significant) fold-change (L_2FC) between treatments, whereas SIMPER analysis identified OTUs that contributed the most to differences between treatments. In bacterial communities, differential abundance analysis identified 88 OTUs that were significantly enriched in the soil crust relative to the bulk soil ($L_2FC > 2, p < 0.05$) (Figure 2.9). Of these 88 OTUs, 27 were from the Bacteroidetes phylum, 25 from Proteobacteria and 9 were from Cyanobacteria. Two OTUs assigned to *Leptolyngbya* (Cyanobacteria), and an OTU assigned to *Phormidium* (Cyanobacteria) were the most enriched OTUs in soil crust communities ($L_2FC = 5.37$, 5.16 for *Leptolyngbya* OTUs, and $\log_2$ fold change = 5.32 for *Phormidium*; $p < 0.0001$ for all comparisons).

In fungal communities, differential abundance analysis identified 71 OTUs that were significantly enriched in the soil crust relative to the bulk soil. Of these, 15 were assigned to the class Agaricomycetes (Basidiomycota), 14 to Dothideomycetes (Ascomycota) and 7 were assigned to Tremellomycetes (Basidiomycota). OTUs assigned to *Xylographa vermicularis* (Ascomycota), Endogonales (Mucoromycota) and *Rhizophagus intraradices* (Glomeromycota) were the most enriched taxa in soil crust communities ($L_2FC = 6.016$, 5.6495 and 5.5579 , respectively).

In protist communities, differential abundance analysis identified 53 OTUs that were enriched in the soil crust relative to the bulk soil. Of these, 17 were assigned to Rhizaria (SAR), 6 to Alveolata (SAR) and 5 to Stramenopiles (SAR). OTUs assigned to Bodomorpha (Rhizaria), Tubulinida (Amoebozoa) and Thaumatomonadida (Rhizaria) were the most enriched taxa in soil crust communities ($L_2FC = 5.1757$, 4.9631 and 4.3371 , respectively)

In phototroph communities, differential abundance analysis identified 10 OTUs that were significantly enriched in the soil crust relative to the bulk soil. Of these, 5 were assigned to Stramenopiles, 3 were Viridiplantae and 2 were Cyanobacteria. OTUs

assigned to *Leptolyngbya* (Cyanobacteria), *Vaucheria litorea* (Stramenopiles) and *Nodosilinea* (Cyanobacteria) were the most enriched taxa in soil crust communities ($L_2FC = 3.1459, 2.2853$ and 2.2575 , respectively).

Similarity of percentages (SIMPER) analysis (Table 2.6.), which identified specific OTUs that contributed largely to differences between soil crust and bulk soil communities, revealed that OTUs assigned to *Oxalobacteraceae* (Proteobacteria), *Phormidium* (Cyanobacteria), *Pseudomonas* (Proteobacteria) and *Micrococcaceae* (Actinobacteria) contributed the most to compartmental differences in bacterial communities (1.77%, 1.21%, 1.16% and 0.93% contribution to change, respectively), and were enriched in soil crust samples. In particular, *Oxalobacteraceae* and *Phormidium* were on average 1.98 and 9.13-fold higher in soil crust samples (respectively). SIMPER analysis identified a *Flavobacterium* (Bacteroidetes) OTU that was enriched in bulk soil bacterial communities, which had a 0.85% contribution to difference.

SIMPER analysis of fungal communities identified that OTUs assigned to *Plenodomus biglobosus* (Ascomycota), *Cladosporium* (Ascomycota) and *Exophiala equina* (Ascomycota) were enriched in the soil crust and contributed the most to differences (3.66%, 3.07% and 2.75% contribution to difference) in community composition with bulk soil. These OTUs were on average 7.31, 1.54 and 1.03-fold higher in soil crust samples, respectively. SIMPER analysis identified a *Mortierella minutissima* OTU and a *Soradiomycetes* (Ascomycota) OTU that had high contributions to difference (2.44% and 2.35%, respectively) but were enriched in bulk soil. These OTUs were on average 2.17 and 1.36-fold higher in bulk soil samples.

SIMPER analysis of protist communities identified OTUs assigned to *Thaumatomonadida* (Cercozoa), Eimeriidae (Amoebozoa) and *Leptomyxida* (Amoebozoa) that had high contributions to difference and were enriched in soil crust samples (14.97%, 4.91% and 2.75% contribution to difference). These OTUs were 1.51, 1.80 and 2.36-fold higher in soil crust samples, respectively. OTUs assigned to *Thecofilosea* (Cercozoa) and *Chromulinales* (Heterokonta) were responsible for the most difference between compartments, but enriched in bulk soil samples (2.72% and

2.67% contribution to difference). The relative abundance of these OTUs was on average 2.14 and 1.87-fold higher in bulk soil samples.

SIMPER analysis of phototroph communities identified OTUs assigned to *Vaucheria litorea* (Stramenopiles), *Klebsormidium flaccidum* (Viridiplantae), *Dicranum scoparium* (Viridiplantae) and *Microcoleus vaginatus* (Cyanobacteria) as OTUs that were enriched in the soil crust and contributed most to differences (13.67%, 5.89%, 5.72% and 5.35%, respectively) in community composition with bulk soil. The average relative abundance of these OTUs was 2.71, 2.48, 1.18 and 1.83-fold higher (respectively) in soil crust samples. Conversely, an OTU assigned to *Acetobacteraceae* was contributing largely to compartmental differences (6.93%) that was enriched in bulk soil samples – which was on average 2.05 times more abundant in bulk soil samples.

Table 2.6. Compartmental OTU SIMPER analysis. OTUs contributing to the highest percentage change between soil crust and bulk soil samples. OTU ID numbers were generated separately for each amplicon

Amplicon	OTU ID	Phylum/Superphylum	Highest taxonomic level information	Soil Crust mean relative abundance (%)	Bulk Soil mean relative abundance (%)	Average Dissimilarity (%)	Contribution to difference (%)
16S rRNA	2	Proteobacteria	Oxalobacteraceae	2.00	0.73	0.7	1.77
16S rRNA	15	Cyanobacteria	<i>Phormidium</i>	1.01	0.08	0.48	1.21
16S rRNA	31	Proteobacteria	<i>Pseudomonas</i>	0.81	0.5	0.46	1.16
16S rRNA	1	Actinobacteria	<i>Micrococcaceae</i>	1.85	1.4	0.37	0.93
16S rRNA	32	Bacteroidetes	<i>Flavobacterium</i>	0.57	0.74	0.34	0.85
ITS	16	Ascomycota	<i>Plenodomus biglobosus</i>	3.22	0.44	1.51	3.66
ITS	2	Ascomycota	<i>Cladosporium exasperatum</i>	4.79	3.1	1.26	3.07
ITS	1	Ascomycota	<i>Exophiala equina</i>	9.36	9.01	1.13	2.75
ITS	8	Mortierellomycota	<i>Mortierella minutissima</i>	1.49	3.24	1	2.44
ITS	3	Ascomycota	<i>Sordariomycetes</i>	2.97	4.05	0.97	2.35
18S rRNA	1569	Cercozoa	Thaumatomonadida	18.6	12.29	9.24	14.97
18S rRNA	1561	Amoebozoa	Eimeriidae	6.57	3.65	4.91	4.91
18S rRNA	2992	Amoebozoa	Leptomyxida	3.46	1.46	1.7	2.75
18S rRNA	1566	Cercozoa	Thecofilosea	2.65	5.66	1.68	2.72
18S rRNA	1562	Heterokonta	Chromulinales	3.34	6.26	1.65	2.67
23S rRNA	8	Stramenopiles	<i>Vaucheria litorea</i>	15.05	5.54	7.49	13.67
23S rRNA	3	Proteobacteria	<i>Acetobacteraceae</i>	6.49	13.33	3.8	6.93
23S rRNA	1	Viridiplantae	<i>Klebsormidium flaccidum</i>	6.95	2.8	3.22	5.89
23S rRNA	5	Viridiplantae	<i>Dicranum scoparium</i>	6.82	5.82	3.13	5.72
23S rRNA	1718	Cyanobacteria	<i>Microcoleus vaginatus</i>	6.19	3.38	2.93	5.35

2.3.5.5. Co-occurrence microbial networks

Multi-amplicon correlation networks were generated from OTU abundance by combining bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) amplicons across all time points using the SparCC algorithm (Figure 2.10) (Friedman and Alm, 2012). In correlation networks, points (nodes) indicate individual OTUs, and lines (edges) between nodes indicate either a positive correlation (gray) or negative correlation (red) in the relative abundance of the two OTUs. The soil crust had a higher number of nodes (136) in comparison to the bulk soil (77) (Table 2.7). But both compartments were represented by a similar proportion of nodes from each amplicon; ITS amplicons had the highest proportion of nodes (43.4% in the soil crust, and 42.9% in the bulk soil), followed by 16S rRNA (38.9% and 37.7%), 18S rRNA (7.4% and 9.1%) and then 23S rRNA (7.4% and 7%).

Connectivity statistics, which include the number of edges, degree (number of edges per node), density (proportion of real connections of possible connections), transitivity (proportion of real connections of possible connections between more than 3 nodes) and polarity (positive or negative correlations), highlighted that the soil crust was more connected than bulk soil. The soil crust had a higher number of total edges (687) and average degree (10.1) than in the bulk soil (183 and 4.8 respectively). Additionally, density and transitivity were both higher in the soil crust (7.5% and 4.4%, respectively) in comparison to the bulk soil (6.3% and 4.0%, respectively). However, both compartments were dominated by positive connections (soil crust: 61.4% positive and 38.6% negative; bulk soil: 60.6% positive, 39.4% negative).

Network clusters, which are regions of the network where nodes are more connected to each other than other regions of the network, were more abundant in the bulk soil (11 clusters) than in the soil crust (9 clusters). However, the range and average number of nodes in each cluster was higher in the soil crust (35, and 15.1, respectively) than in the bulk soil (16 and 7.0), respectively.

Table 2.7. Correlation network statistics. Network statistics of soil crust and bulk soil microbial correlation networks of combined amplicons (16S rRNA, ITS, 18S rRNA and 23S rRNA) across all time points, generated using the SparCC algorithm. Statistics for significant connections with a threshold >|0.7| are shown.

	Soil Crust	Bulk Soil
Total number of nodes (%)	136	77
16S rRNA nodes	53 (38.9%)	29 (37.7%)
ITS nodes	59 (43.4%)	33 (42.9%)
18S rRNA nodes	14 (10.3%)	8 (10.3%)
23S rRNA nodes	10 (7.4%)	7 (9.1%)
Number of edges	687	183
Positive connections	422 (61.4%)	111 (60.6%)
Negative connections	265 (38.6%)	72 (39.4%)
Mean degree	10.1	4.8
Density	7.54%	6.31%
Transitivity	4.43%	4.00%
Centrality	234.71	58.36
Total number of clusters	9	11
Mean cluster size	15.1	7.0
Max. cluster size	37	18
Min. cluster size	2	2

Node connectivity was visualized as according to Guimera and Nunes Amaral (2005) cartographic representation framework, which categorizes nodes based on their within and between module connectivity (Figure 2.11). In soil crust communities, one module hub was identified, which was a fungal OTU assigned to *Meliniomyces* (Fungi, Ascomycota). 20 OTUs were identified as connector nodes (16S rRNA - 3, ITS - 13, 18S rRNA - 2, 23S rRNA - 2), of which OTUs assigned to *Plenodomus* (Fungi, Ascomycota), *Sodariomycetes* (Fungi, Ascomycota) and *Vaucheria litorea* (Eukaryote, Stramenopiles) displayed the highest within-module degree connectivity. 50 peripheral nodes were identified (16S rRNA - 18, ITS - 24, 18S rRNA - 4, 23S rRNA - 4), of which *Glissomonadida* (Eukaryote, Rhizaria), *Endogonomycetes* (Fungi, Mucoromycota) and *Exophiala* (Fungi, Ascomycota) displayed the highest within-module connectivity. 64 ultra-peripheral nodes were identified (16S - 32, ITS - 22, 18S - 8, 23S - 4), of these *Basidiobolus ranarum* (Fungi, Entomophthoromycota), *Luteolibacter* (Bacteria, Verrucomicrobia) and *Phormidium* (Cyanobacteria) displayed the highest within-module degree.

In bulk soil networks, one module hub was identified, which was an OTU assigned to *Celivibrio* (Bacteria, Proteobacteria). Nine nodes were identified as connector nodes

(16S: 4, ITS: 5), of which OTUs assigned to *Arenimonas* (Bacteria, Proteobacteria), *Luteolibacter* (Bacteria, Verrucomicrobia) and *Leotiomycetes* (Ascomycota) displayed the highest within-module connectivity. 17 nodes were identified as peripheral nodes (16S - 6, ITS - 7, 18S - 1, 23S - 2), of which OTUs assigned to *Leotiomycetes* (Fungi, Ascomycota), *Sodariomycetes* (Fungi, Ascomycota) and *Microdochium* (Fungi, Ascomycota) displayed the most within-module connectivity. 51 nodes were identified as ultraperipheral nodes (16S - 19, ITS - 21, 18S - 7, 23S - 5), of these *Dicranum* (Eukaryota, Viridiplantae) and *Phaeosphaeriaceae* (Fungi, Ascomycota) displayed the highest within-module connectivity.

Figure 2.10. Correlation networks of soil crust and bulk soil microbial communities. Co-occurrence microbial networks of combined amplicons (16S rRNA, ITS, 18S rRNA and 23S rRNA) across all time points, generated using the SparCC algorithm displaying significant ($p < 0.05$) positive (grey lines) and negative (red lines) connections above a threshold of 10.71. Size of nodes indicates the number of connections.

2.3.6. Temporal dynamics of soil crust microbial communities

2.3.6.1. Alpha diversity

Two-way ANOVA revealed that Fisher's alpha diversity significantly varied over time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$) for bacterial, protist and phototroph communities (Figure 2.12, Table 2.8.). In fungal communities, there was no significant compartmental effect ($p = 0.2017$). In bacterial, fungal and phototroph communities, time explained the largest proportion of variation in alpha diversity (86.58%, 43.67% and 82.58%, respectively), whereas in protist communities compartment explained the most variation in alpha diversity (38.02%).

Figure 2.12. Alpha diversity of soil crust and bulk soil communities across time. Alpha diversity was calculated using Fisher's diversity index for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities. Linear regression models were fit to soil crust and bulk soil samples in each amplicon (green and brown lines). Red ticks and letters (top y-axis) denote the month of sampling.

Table 2.8. Temporal alpha diversity statistics of soil crust and bulk soil communities. Alpha diversity statistics (generated using a two-way ANOVA) of soil crust and bulk soil bacterial (16S rRNA), fungal (ITS), protist (18S) and phototroph (23S rRNA) communities.

	16S	ITS	18S	23S
Time				
% Variation	86.58%	43.67%	20.54%	82.58%
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Compartment				
% Variation	7.27%	0.39%	38.02%	10.13%
<i>p</i> -value	< 0.0001	0.2017	< 0.0001	< 0.0001
Interaction				
% Variation	2.96%	24.56%	25.09%	2.789%
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001

In soil crust communities, regression analysis revealed that alpha diversity had a strong, significant decline across time in all amplicons ($R^2 = 0.3301, 0.4564, 0.7709$ and 0.5475 for bacteria, fungi, protists and phototrophs, respectively, $p < 0.0001$ in all cases) (Table 2.9). Most notably, in protist alpha diversity reduced by 53.9% in the soil crust from November to August. However, in bulk soil regression analysis revealed that only fungal communities had a significant correlation with time, although the goodness of fit of this correlation was poor (bacteria: $p = 0.1456$, protists: $p = 0.4285$, phototrophs: $p = 0.7771$).

Table 2.9. Temporal alpha diversity regression statistics. Regression statistics of soil crust and bulk soil bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities. Statistics were generated using a linear regression model.

	16S		ITS		18S		23S	
	Crust	Bulk	Crust	Bulk	Crust	Bulk	Crust	Bulk
Slope	-0.4987	-0.0989	-0.0743	-0.0238	-0.1749	0.0123	0.0055	0.0063
F	23.65	2.18	40.30	6.65	158.20	0.63	56.08	0.08
R^2	0.3301	0.0438	0.4564	0.0121	0.7709	0.1339	0.5475	0.0016
<i>P</i>	< 0.0001	0.1456	< 0.0001	0.013	< 0.0001	0.4285	< 0.0001	0.7771

2.3.6.2. Community composition

NMDS plots, generated using Bray-Curtis similarity, revealed clear temporal separation in soil crust samples, but relatively little temporal separation in bulk soil samples – which was the case in all amplicons (Figure 2.14). In soil crust communities, early time points clustered together with bulk soil samples (November-March for bacterial and fungal communities, but November-December for protist and phototroph communities), suggesting high similarity. For all amplicons, across time soil crust communities became less similar to the initial time points – and June-August time points clustered separately from earlier time points. However, in bulk soil samples for all amplicons there was little temporal separation with groups of November-May and June-August timepoints.

Figure 2.13. Community composition similarity in soil crust and bulk soil communities across time. Bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) community similarity was calculated by comparing samples the first time point (within amplicon and compartment) using Bray-Curtis similarity. Linear regression models were fit to soil crust and bulk soil samples in each amplicon (green and brown lines). Red ticks and letters (top y-axis) denote the month of sampling.

Figure 2.14. Temporal NMDS plots of soil crust and bulk soil microbial communities. Non-metric multi-dimensional scaling plots for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities were generated based on Bray-Curtis similarity comparisons between samples. The axis scales for soil crust and bulk soil samples are identical within each amplicon.

Pairwise Bray-Curtis comparisons of each sample were made with samples obtained from the initial November timepoint within each amplicon and compartment (Figure 2.13). This was performed to determine the extent that community composition had changed from the initial community structure within each compartment. Two-way ANOVA revealed that time ($p < 0.0001$, all amplicons) and compartment ($p < 0.0001$, all amplicons) had significant effects on Bray-Curtis similarity, and there was a significant interaction of time and compartment ($p < 0.0001$, all comparisons) (Table 2.10). In bacterial and fungal communities, time explained the largest proportion of variation (53.25% and 37.32%, respectively), however in protist and phototroph communities, compartment was responsible for the largest proportion of variation (42.92% and 46.88%, respectively).

Table 2.10 Statistics of community composition similarity in soil crust and bulk soil communities across time. Bray-Curtis comparisons were made with the initial time November time point for soil crust and bulk soil bacterial (16S rRNA), fungal (ITS), protist (18S) and phototroph (23S rRNA) communities. Statistics were generated using a two-way ANOVA

	16S rRNA	ITS	18S rRNA	23S rRNA
Time				
% Variation	53.25%	37.32	16.60	17.68
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Compartment				
% Variation	3.51%	13.95	42.92	46.88
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Interaction				
% Variation	25.43%	13.04	5.82	7.838
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Further, regression analysis also revealed strong negative correlations between Bray-Curtis similarity and time for each amplicon ($p < 0.0001$ in all cases, except fungal bulk soil communities $p = 0.0014$) (Table 2.11). However, analysis of gradients revealed Bray-Curtis similarity decreased at a faster rate over time in soil crust communities relative to bulk soil communities for all amplicons ($p < 0.0001$ in all cases).

Table 2.11. Regression statistics of community composition similarity in soil crust and bulk soil communities across time. Bray-Curtis comparisons were made with the initial time November time point for soil crust and bulk soil bacterial (16S rRNA), fungal (ITS), protist (18S) and phototroph (23S rRNA) communities. Statistics were generated using a linear regression model.

	16S		ITS		18S		23S	
	Crust	Bulk	Crust	Bulk	Crust	Bulk	Crust	Bulk
Slope	-7.6(10 ⁻⁴)	-1.7(10 ⁻⁴)	-5.1(10 ⁻⁵)	-4.5(10 ⁻⁵)	-1.1(10 ⁻³)	-3.8(10 ⁻³)	-8.6(10 ⁻⁴)	-3.8(10 ⁻⁴)
F	303.6	97.07	335.4	38.16	190.6	52.22	123.9	43.16
R ²	0.5658	0.2941	0.59	0.1407	0.45	0.1831	0.3472	0.1563
<i>p</i>	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001	< 0.0001

2.3.6.3. Temporal taxonomic shifts

2.3.6.3.a. Bacteria

Soil crust bacterial phyla displayed considerable variation in abundance across time, whereas the bulk soil phyla remained relatively stable throughout (Figure. 2.15). SIMPER analysis identified Actinobacteria, Proteobacteria, Bacteroidetes, Acidobacteria and Cyanobacteria as the main phyla which changed over time in the soil crust (2.28%, 1.19%, 1.58%, 1.54% and 1.29% average dissimilarity with time, respectively) (Figure 2.17). In particular, the relative abundance of Cyanobacteria was 18.5-fold higher ($p < 0.0001$) in August in comparison to November (reaching up 29.5-fold higher in May) in soil crust communities, but displayed no significant change in bulk soil communities ($p = 0.5889$). Conversely, in soil crust communities Acidobacteria displayed a significant reduction in relative abundance between November to August, and was 1.64 times lower in August ($p < 0.0001$) than in November. However, in bulk soil communities there was a significant increase in the relative abundance of Acidobacteria during this time period (1.2-fold increase, $p = 0.0027$).

SIMPER analysis of bacterial communities at the OTU level identified OTUs assigned to Oxalobacteraceae (Proteobacteria), *Pseudomonas* (Proteobacteria), *Phormidium* (Cyanobacteria), Micrococcaceae (Actinobacteria) and a *Flavobacterium* (Bacteroidetes) with the most substantial change over time (Figure 2.18). In particular, the relative abundance of the Oxalobacteraceae and *Phormidium* OTUs displayed

large temporal shifts, which were 6.4-fold and 18.1-fold higher ($p < 0.0001$), respectively, in August in comparison to November in soil crust communities, but displayed no significant change in bulk soil communities.

2.3.6.3.b. Fungi

Large temporal shifts occurred within soil crust fungal classes across the time course, notably from April-August, however bulk soil communities remained relatively stable across time (Figure 2.15). SIMPER analysis identified Dothideomycetes (Ascomycota), Sodiariomycetes (Ascomycota), Eurotiomycetes (Ascomycota), Leotiomycetes (Ascomycota) and Basidiobolomycetes (Entomophthoromycota) as the main classes which changed with time in the soil crust (5.38%, 4.9%, 1.51%, 1.34% and 1.09% average dissimilarity with time, respectively) (Figure 2.17). In particular, Dothideomycetes and Sodiariomycetes displayed drastic shifts in relative abundance between April and May, when the relative abundance of Dothideomycetes increased from 14.64% to 32.46% (2.2-fold increase, $p < 0.0001$), and Sodiariomycetes decreased from 31.74% to 17.9% (1.77-fold decrease, $p < 0.0001$) – these changes did not revert at subsequent time points. During the same time period there were no significant shifts in the abundances of Dothideomycetes ($p = 0.7012$) or Sodiariomycetes ($p = 0.4815$) bulk soil communities.

SIMPER analysis at the OTU level identified OTUs assigned to *Plenodomus biglobosus* (Ascomycota; Dothideomycetes), *Exophiala equina* (Ascomycota; Chaetothyriomycetes), *Cladosporium exasperatum* (Ascomycota; Dothideomycetes), *Zymoseptoria* (Ascomycota; Dothideomycetes) and *Meliniomyces* (Ascomycota; Leotiomycetes) as the main OTUs which changed with time in soil crust communities (Figure 2.18). Of these, *Plenodomus biglobosus* displayed a large increase in relative abundance between April (0.46%) to May (6.75% - 14-fold increase, $p < 0.0001$) in soil crust communities, and remained abundant until the end of the time course. Bulk soil communities also saw an increase in the relative abundance of *Plenodomus biglobosus* during this time period (0.15% to 1.34% from May to July, $p = 0.0002$). Significant increases in relative abundance in later time points was also seen in *Cladosporium exasperatum* (increase from 2.63% to 11.12% between May to July)

and *Meliniomyces* (increase from 0.49% to 5.30% from April to June) in soil crust communities.

2.3.6.3.c. Protists

Large variation occurred in protist phyla (or superphylum) across the time course in both compartments, but was more pronounced in the soil crust (Figure 2.16). SIMPER analysis identified Amoebozoa, Rhizaria, Alveolata, Stramenopiles and Discoba as the main phyla which changed with time in the soil crust (6.67%, 6.64%, 3.63%, 1.78% and 1.57% average dissimilarity with time, respectively) (Figure 2.17). In particular, Amoebozoa significantly increased in relative abundance within the soil crust from 6.16% to 44.26% between May-August ($p < 0.0001$), but did not change in bulk soil communities ($p = 0.2791$). Conversely, relative abundance of Alveolata (SAR) displayed large temporal shifts in soil crust communities, decreased from 11.88% to 4.25% in the soil crust between November to August ($p < 0.0001$), but no significant differences were seen in bulk soil communities ($p = 0.3961$).

SIMPER analysis at the OTU level identified OTUs assigned to *Allantion* (SAR, Rhizaria), Eimeriidae (SAR, Rhizaria), Retaria (SAR, Rhizaria), *Preleptomonas faecicola* (SAR, Rhizaria) and Chromulinales (SAR, Rhizaria) as the main OTUs which changed with time in soil crust communities (Figure 2.18). Of these, Eimeriidae increased from 4.80% in November to 24.44% in August ($p < 0.0001$) in soil crust communities, but did not change in bulk soil communities ($p = 0.9414$). Conversely, Chromulinales significantly decreased from 5.38% to 1.63% between November and August in soil crust communities ($p < 0.0001$), but no significant differences were seen in bulk soil communities ($p = 0.5813$).

2.3.6.3.d. Phototrophs

Large variation occurred in phototroph communities across the time course in both compartments, but was more pronounced in the soil crust (Figure 2.16). SIMPER analysis revealed Xanthophyceae, Klebsormidiophyceae, Bryophyta, Microcoleus and Rhodospirillales as the main classes that changed with time in the soil crust (10.89%, 7.25%, 4.96%, 2.59% and 2.41% average dissimilarity with time, respectively) (Figure 2.17). In particular, Klebsormidiophyceae increased in relative

abundance from 1.6% in November to 31.6% in July ($p < 0.0001$) in soil crust communities, and from 1.7% to 10.35% in bulk soil communities ($p < 0.0001$). Additionally, *Microcoleus* (Cyanobacteria) significantly increased from 1.36% to 13.91% between November to August ($p < 0.0001$) in soil crust communities; but displayed no changes over time in bulk soil communities ($p = 0.1154$). The relative abundance of Bryophyta (which contained only moss and lichen sequences) significantly increased in the soil crust between November (15.33%) to August (25.95%, $p < 0.0001$), but did not significantly change in the bulk soil. Conversely, the relative abundance of Rhodospirillales decreased significantly from 16.00% to 6.98% between November to August ($p < 0.0001$) in soil crust communities; but no significant changes in abundance occurred over this time frame in the bulk soil ($p = 0.5776$).

SIMPER analysis at the OTU level identified OTUs assigned to *Vaucheria litorea* (Stramenopiles), *Klebsormidium flaccidum* (Viridiplantae, Streptophyta), *Dicranum scoparium* (Viridiplantae, moss), *Microcoleus Vaginatus* (Cyanobacteria) and *Acetobacteraceae* (Proteobacteria) as the main OTUs which changed in abundance over time (Figure 2.18). *Vaucheria litorea* displayed a large increase in relative abundance between April (0.46%) and May (6.75%, $p < 0.0001$) only in the soil crust, and remained abundant until the end of the time course ($p = 0.251$). Additionally, *Klebsormidium flaccidum* increased significantly from 0.12% to 2.51% between November and July in soil crust communities ($p < 0.0001$), and from 0.12% to 0.85% in bulk soil ($p < 0.0001$). *Dicranum scoparium* (moss) significantly increased in relative abundance between November (0.60%) to August (1.72%) in the soil crust ($p < 0.0001$) but not in the bulk soil.

Figure 2.15. Relative abundance of bacterial phyla and fungal classes across time in soil crust and bulk soil communities. Only taxa that contributed to >1% relative abundance are shown. Any taxa contributing to less than 1% are in the 'low abundance' category.

Figure 2.16. Relative abundance of microbial eukaryotic phyla and phototrophic classes across time in soil crust and bulk soil communities. Only taxa that contributed to >1% relative abundance are shown. Any taxa contributing to less than 1% are in the 'low abundance' category.

Figure 2.17. Relative abundance of temporally variable microbial taxa identified by SIMPER analysis. The relative abundance of the most variable soil crust (green) and bulk soil (brown) taxa across time identified by SIMPER analysis. Shaded regions indicate $\pm$ standard deviation.

Figure 2.18. Relative abundance of temporally variable microbial OTUs identified by SIMPER analysis. The relative abundance of the most variable soil crust (green) and bulk soil (brown) OTUs across time identified by SIMPER analysis. Shaded regions indicate $\pm$ standard deviation.

2.3.6.4. Temporal co-occurrence microbial networks

Multi-amplicon co-occurrence networks generated from OTU read count were performed for each time point and compartment (Figure 2.19). In soil crust networks, the number of nodes displayed temporal variation but there was no clear correlation with time ($R^2 = 0.08727$, $p = 0.4073$). The relative proportions of nodes from each amplicon was similar regardless of time points (~60% positive, ~40% negative) (Figure 2.0). In bulk soil networks, there was no clear correlation between the number of nodes and time ($R^2 = 0.00737$, $p = 0.8136$).

Regression analysis of connectivity statistics (density, transitivity and degree) revealed no clear correlation with time in either compartment ($p > 0.05$ in all comparisons). However, there was a higher proportion of positive than negative connections in both compartments at every time point ($> 51.1\%$ in the soil crust, $> 51.6\%$ in the bulk soil). Cluster number and size did not show any significant correlations with time. Betweenness centrality displayed large temporal variation, but no clear patterns were seen.

Figure 2.19. Correlation networks of soil crust and bulk soil microbial communities at each time point. Microbial co-occurrence networks of combined amplicons (16S rRNA, ITS, 18S rRNA and 23S rRNA) within each month generated using the SparCC algorithm displaying significant ($p < 0.05$) positive (grey lines) and negative (red lines) connections above a threshold of 10.91. Size of nodes indicates the number of connections.

Figure 2.20. Temporal network statistics. Statistics of microbial co-occurrence networks of combined amplicons (16S rRNA, ITS, 18S rRNA and 23S rRNA) within each month generated using the SparCC algorithm for soil crust (green) and bulk soil (brown) samples. Statistics were produced from only significant correlations.

2.3.7. Community assembly processes

2.3.7.1. Phylogenetic signal

Phylogenetic signals were determined using mantel correlograms and by calculating Pagel's λ , using time as a factor, for each amplicon. Mantel correlograms (supplementary Figure 2.1-2.2) showed a significant phylogenetic signal over short phylogenetic distances for both soil crust and bulk soil communities in all amplicons (bacteria: ~ 0.2 phylogenetic units, fungi: ~ 0.8 phylogenetic units, protist: ~ 0.8 phylogenetic units, phototrophs: ~ 0.4 phylogenetic units).

Pagel's λ also revealed significant phylogenetic signals in both compartments for bacterial (soil crust: $\lambda = 0.998$, $p < 0.0001$; bulk soil: $\lambda = 0.831$, $p = 0.0024$), fungal (soil crust: $\lambda = 0.738$, $p = 0.0081$; bulk soil: $\lambda = 0.701$, $p = 0.0024$), protist (soil crust: $\lambda = 0.791$, $p = 0.0078$; bulk soil: $\lambda = 0.712$, $p = 0.0089$) and phototroph (soil crust: $\lambda = 0.992$, $p = 0.0031$; bulk soil: $\lambda = 0.697$, $p = 0.0008$) communities. As a result, β -NTI (using β -MNTD) was used to determine community assembly processes – as this model assumes a phylogenetic relationship over short phylogenetic distance.

2.3.7.2. β -NTI

To determine phylogenetic turnover, β -NTI was calculated across time in soil crust and bulk soil bacterial communities (Figure 2.21). In soil crust communities, regression analysis revealed that β -NTI had a significant positive-correlation with time in bacterial ($R^2 = 0.2685$, $p < 0.0001$) and phototroph communities ($R^2 = 0.01480$, $p < 0.0001$), a significant negative-correlation in microbial eukaryote communities ($R^2 = 0.02094$, $p < 0.0001$), but no significant temporal correlation in fungal communities ($R^2 = 0.001753$, $p = 0.1382$).

In soil crust samples, the relative contributions of community assembly processes displayed large shifts across time (Figure 2.22). In bacterial communities, deterministic processes, mainly homogenous selection, were dominant across shorter time periods (81.6%, 66.2% and 61.0% in the first 3 time points, respectively). However, deterministic processes became less influential as time increased (8.0%, 24.2% and 12.0% in the last 3 time categories, respectively), and variable selection (not homogenising) became the dominant deterministic process. In contrast, stochastic

processes became more dominant with time in soil crust bacterial communities, which were initially dominated by homogenous dispersal (66.67%, 45.15% and 36% of stochastic processes in the first three time categories, respectively), but became dominated by dispersal limitation at later time points (84.21% and 91.10% of stochastic processes in the last two time points, respectively).

However, in fungal protist and phototroph soil crust communities were mainly dominated by stochastic processes across the course (Figure 2.22). These communities were all initially dominated by random genetic drift in the first 3 time categories, but became dominated by dispersal limitation at later time points. Deterministic processes had minimal contributions to selection processes for fungal communities in all time points (<6.0% in all time points), but variable selection was the main contributor towards this small fraction. In protist and phototroph communities, deterministic processes contributed mainly in early time points (microbial eukaryotes: 12.5% – 23.0%; phototrophs: 13.1% – 30.2% in the first 5 time categories), of which variable selection dominated in microbial eukaryotes, but homogenous selection dominated in phototroph communities.

In bulk soil communities, regression analysis revealed that β -NTI had a significant positive-correlation with time in bacterial ($R^2 = 0.07825$, $p < 0.0001$), fungal ($R^2 = 0.01480$, $p < 0.0001$), and phototroph communities ($R^2 = 0.01195$, $p < 0.0001$), but no significant temporal correlation in microbial eukaryote communities ($R^2 = 0.0001646$, $p = 0.5895$) (Figure 2.21).

In bulk soil samples, the relative contributions of community assembly processes were largely dominated by stochastic processes in fungal, microbial eukaryote and phototroph communities (Figure 2.22). In these communities, random genetic drift was the main contributor in all time points. In bacterial bulk soil communities, community assembly processes were initially dominated by homogenising selection (deterministic), but became dominated by genetic drift (stochastic) as time increased.

Figure 2.21. β -NTI and time in soil crust and bulk soil microbial communities. Pairwise β -NTI values and absolute difference in time (between samples) in soil crust and bulk soil samples. Red lines indicate linear regression models. Dashed grey lines are at β -NTI values of ± 2 , values that fall inside this category indicate stochastic processes dominate, whereas if values fall outside this category selection processes dominate.

Figure 2.22. Community assembly processes in soil crust and bulk soil communities across time. Community assembly processes were determined following Vellend *et al.* (2012, 2013, 2015) framework. HS (homogenous selection) - $\beta\text{-NTI} > 2$, VS (variable selection) $\beta\text{-NTI} < -2$, HD (homogenising dispersal) $\beta\text{-NTI} < |2|$ and $B_{RC} > 0.95$, DR (drift), and $B_{RC} < |0.95|$ and DL and $B_{RC} < -0.95$ (dispersal limitation). Temporal categories (x-axis) correspond to the sampling times.

2.3.8. The relative impacts of time, distance and edaphic factors on soil crust microbial communities

Multivariate redundancy analysis (RDA) was performed on OTU abundance tables (using integrated forward selection of explanatory variables) to determine the relative influence of measured variables on community composition (Table 2.12 – 2.13). In the soil crust the measured variables accounted for was 56.9% for bacteria, 54.1% for fungal, 59.6% for protists and 62.3% for phototrophs communities (Table 2.12).

In the soil crust, time was the main contributing factor to community variation in bacterial (15.7%, $p = 0.0045$) and protist communities (31.9%, $p = 0.045$), but did not significantly affect contributions of fungal ($p = 0.068$) or phototroph communities ($p = 0.17$). In fungal communities, chlorophyll *a* content explained the most variation in community composition (28.1%, $p = 0.0045$); whereas temperature explained the most variation in phototroph community composition (36.4%, $p = 0.0045$). Distance did not contribute significantly towards explained variation for any amplicon ($p > 0.05$ in all cases).

In bulk soil communities, the variation significantly accounted ($p < 0.05$) for by measured variables towards 38% in bacteria, 28.4% in fungi and 26.4% in phototrophs (Table 2.13). No measured variables were significantly contributing to explained variance in bulk soil microbial eukaryote communities ($p > 0.05$). Time was the main contributing factor to community variation for bacterial (15.9%, $p = 0.0045$) and fungal communities (15.3%, $p = 0.009$), but was not significant in microbial eukaryotes or phototrophs. In phototrophic communities, temperature was the main contributing factor to community composition (12.3%, $p = 0.045$).

Table 2.12. Redundancy analysis of soil crust microbial communities. Relative contributions of measured variables towards soil crust microbial community composition. Redundancy analysis was performed using integrative forward-selection

Amplicon	Rank	Factor	Variation explained (%)	Contribution to explained variance (%)	Adjusted F-statistic	P-value
16S	1	Time	15.7	25.5	8.9	0.0045
	2	Chlorophyll <i>a</i>	13.3	21.6	10.4	0.0036
	3	Nitrogen	12.4	20.1	8.1	0.0045
	4	Soil moisture	9.7	15.8	8.9	0.0026
	5	pH	3.1	5.0	2.9	0.0036
	6	Temperature	2.7	4.4	2.7	0.0060
	7	Carbon	2.2	3.6	2.2	0.021
	8	Bulk Density	1.3	2.2	1.4	0.14
	9	Distance	1.2	1.9	1.2	0.22
ITS	1	Chlorophyll <i>a</i>	28.1	49.4	18.8	0.0045
	2	Bulk Density	7.8	13.7	5.7	0.0036
	3	Nitrogen	7.0	12.4	5.7	0.0036
	4	pH	3.8	6.8	3.2	0.0036
	5	Carbon	3.0	5.3	2.6	0.0036
	6	Temperature	2.3	4.1	2.1	0.0090
	7	Soil moisture	2.1	3.7	1.9	0.010
	8	Time	1.6	2.9	1.5	0.068
	9	Distance	1.0	1.8	0.9	0.48
18S	1	Time	31.9	50.6	22.5	0.0045
	2	pH	10.0	15.8	8.1	0.0036
	3	Nitrogen	5.7	9.0	5.0	0.0036
	4	Temperature	6.1	9.7	5.9	0.0030
	5	Carbon	3.1	4.9	3.1	0.011
	6	Chlorophyll <i>a</i>	2.8	4.5	3.0	0.024
	7	Soil moisture	1.5	2.3	1.6	0.18
	8	Bulk Density	1.1	1.8	1.2	0.34
	9	Distance	0.9	1.4	1.0	0.37
23S	1	Temperature	36.4	55.0	27.5	0.0045
	2	Nitrogen	8.6	13.0	7.3	0.0090
	3	Bulk Density	7.1	10.7	6.8	0.0060
	4	pH	4.2	6.3	4.3	0.0090
	5	Carbon	3.6	5.5	4.0	0.014
	6	Chlorophyll <i>a</i>	2.4	3.6	2.7	0.050
	7	Time (days)	1.6	2.4	1.9	0.17
	8	Distance	1.3	2.0	1.5	0.19
	9	Soil moisture	1.1	1.7	1.3	0.22

Table 2.13. Redundancy analysis of bulk soil microbial communities. Relative contributions of measured variables towards bulk soil microbial community composition. Redundancy analysis was performed using integrative forward-selection

Amplicon	Rank	Factor	Variation explained (%)	Contribution to explained variance (%)	Adjusted F-statistic	P-value
16S	1	Time	15.9	37.9	9.1	0.0045
	2	Nitrogen	8.2	19.5	5	0.0045
	3	Temperature	4.7	11.2	3	0.0036
	4	pH	3.4	8.1	2.2	0.0045
	5	Soil moisture	2.7	6.5	1.8	0.014
	6	Carbon	3.1	7.4	2.1	0.011
	7	Bulk Density	1.6	3.9	1.1	0.32
	8	Distance	1.2	3.0	0.9	0.79
	9	Chlorophyll <i>a</i>	1.1	2.7	0.8	0.8
ITS	1	Time	15.3	40.7	8.7	0.0090
	2	Temperature	4.9	12.9	3.0	0.0090
	3	Nitrogen	4.8	12.6	2.8	0.012
	4	Soil moisture	3.4	8.9	2.1	0.013
	5	Carbon	2.5	6.6	1.6	0.12
	6	pH	2.6	6.9	1.7	0.068
	7	Chlorophyll <i>a</i>	2.0	5.4	1.3	0.18
	8	Distance	1.3	3.5	0.8	0.75
	9	Bulk Density	0.9	2.5	0.6	0.95
18S	1	Bulk Density	5.1	24.5	2.6	0.072
	2	Time	4.7	22.7	2.5	0.072
	3	pH	3.5	16.8	1.9	0.25
	4	Nitrogen	1.9	9.3	1.0	0.71
	5	Distance	1.9	9.1	1.0	0.67
	6	Temperature	1.1	5.3	0.6	0.77
	7	Soil moisture	1.0	4.9	0.5	0.92
	8	Chlorophyll <i>a</i>	0.9	4.4	0.5	0.87
	9	Carbon	0.6	3.0	0.3	0.94
23S	1	Temperature	12.3	32.0	6.7	0.0045
	2	pH	7.8	20.4	4.6	0.012
	3	Bulk Density	6.3	16.5	3.9	0.0090
	4	Nitrogen	3.1	8.1	2.0	0.185
	5	Soil moisture	2.6	6.9	1.7	0.180
	6	Carbon	3.2	8.3	2.1	0.094
	7	Distance	1.6	4.3	1.1	0.46
	8	Time	0.9	2.3	0.6	0.73
	9	Chlorophyll <i>a</i>	0.4	1.2	0.3	0.96

2.4. Discussion

This chapter presents the first study that uses next generation sequencing techniques to track the development of microbial communities of soil crusts in a temperate agricultural environment. Results here revealed significant disparities between the soil surface and bulk soil bacterial, protist and phototrophic community composition and diversity. Furthermore, these communities exhibited significant reductions in alpha diversity and became less similar to initial communities across time in the soil crust. Additionally, a substantial number of moss and lichen sequences were detected in soil crust phototrophic communities in later time points, suggesting considerable development of soil crusts had occurred. Soil crust communities displayed substantial enrichments of Cyanobacteria which were over 10-fold more abundant in comparison to the bulk soil, which is in accordance with a number of prior studies performed in arid regions. Furthermore, significant enrichments of Bacteroidetes, Agaricomycetes, Dothideomycetes, Rhizaria, Oscillatoriales and Dinophyceae were all presented in the soil crust relative to bulk soil, suggesting a currently underappreciated significance of these taxa in soil crust communities. At the OTU level, using both SIMPER and differential abundance analysis, findings here revealed that sequences assigned to *Phormidium* (Cyanobacteria), *Vaucheria litorea* (yellow-green algae) and *Klebsormidium flaccidum* (charophyte green algae) were significantly enriched in the soil crust, which were substantially more abundant at the end of the time course. Soil crust microbial co-occurrence networks were more complex than bulk soil networks, within which they had a higher number of nodes, edges and connectivity, but the proportion of nodes assigned to each amplicon and the polarity of connections was conserved across compartments. Cartographic representation of microbial networks revealed that fungal OTUs such as *Meliniomyces* and *Plenodomus* were highly connected in the soil crust, suggesting a significant role of fungal taxa in the soil crust, despite there being no overall differences in fungal community composition or diversity between compartments. Furthermore, the correlative importance of fungal OTUs in soil crust communities identified by network analysis, coupled with significantly higher levels of chlorophyll α in soil crust samples provides evidence that suggests a functional role in the degradation phototroph biomass by fungal saprotrophs. This is further supported by substantial shifts in fungal community composition in developed crusts between April and August, during which significant

increases in phototrophic biomass were observed, and therefore additionally provides evidence of ecological succession during this time frame. Community assembly processes of bacterial communities were initially governed by deterministic processes (homogenous selection), but stochastic processes became more prevalent across time. Fungal, protist and phototroph communities were all dominated by stochastic community assembly processes regardless of time, which consisted mainly of dispersal limitation in the soil crust but drift in the bulk soil. Lastly, redundancy analysis revealed that local-scale distance effects had a negligible influence on soil crust and bulk soil microbial community composition. Together, these results highlight that the soil surface is a taxonomically distinct soil compartment, that has enrichments of a number of taxa that are not abundant in lower soils. Moreover, results here show that substantial development of soil crusts can occur within a single growing season in a real-world agricultural ecosystem.

2.4.1. Alpha diversity

Alpha diversity was significantly lower in bacterial, protist and phototrophic soil crust communities in comparison to bulk soil communities. This finding not only complements prior research performed in arid regions, which has shown that bacterial alpha diversity is substantially lower in soil crusts relative to subsurface and bare soils (Mojave Desert and Colorado Plateau), but additionally extends this finding to protist and phototrophic communities (Mogul et al., 2017, Lee et al., 2016). The soil surface is subject to a harsher extent of environmental stresses relative to the bulk soil, which includes large temperature fluxes, freezing and thawing cycles, high UV light exposure, mechanical pressure and high levels of dispersal (see section 1.2.5.3), all of which will select for a tighter range of taxa that are adapted to tolerate these stresses and could reduce diversity (Csotonyi et al., 2010). A number of studies have also shown that bulk soil communities exhibit a higher species richness than other soil compartments, for example the rhizosphere, which is often associated with increased resistance and resilience, a greater multifunctionality and the ability to act as a seed bank for rare taxa to flourish in altered environmental conditions (Fuhrman, 2009, Shi et al., 2015). In the context of agriculture, increased diversity of the bulk soil could therefore act as a reservoir that recruits rare taxa to help re-establish soil surface communities after perturbation events that mix surface and bulk soils, for example due

to tillage or harvest - which has previously been implied in the recruitment of arid soil crust communities (Lee et al., 2016). This is supported by temporal analysis of alpha diversity, within which no significant compartmental differences were seen in November (2 weeks post-tillage) in alpha diversity, suggesting a shared microbial pool between compartments shortly after perturbation. Furthermore, temporal analysis also revealed that alpha diversity of bacterial, fungal, protist and phototroph soil crust communities significantly reduced with time. This result is actually contrary to findings present by Maier et al. (2018), who revealed that alpha diversity of 16S rRNA and 18S rRNA community profiles increased as soil crusts became more developed in arid soils obtained from South Africa. While results shown by Maier and colleagues suggest that increased soil crust development in arid regions encourage proliferation and recruitment of taxa to the soil surface, results shown here suggest that soil crust development instead leads to increased exclusivity of soil surface microbial communities.

2.4.2. Community composition

NMDS plots of bacterial, fungal, protist and phototrophic communities showed considerable compartmental clustering, within which soil crust communities were compositionally less similar to one another in comparison to bulk soil communities. Further analysis showed that these differences were due to greater temporal variation in soil crust communities, and additionally NMDS plots showed that soil crust communities at early time points were compositionally similar to bulk soil communities. This strengthens the proposal that agricultural disturbances that mix the soil surface and bulk soils, such as conservational tillage and crop drilling, result in a shared microbial pool between these two compartments. Community-level similarity comparisons revealed that both soil crust and bulk soil communities exhibited significant temporal changes in community composition across the growing season. Although significant reductions in community similarity were witnessed in communities from both soil compartments, a larger shift in composition was seen in soil crust communities in comparison to bulk soil communities. Temporal development of soil crusts has been widely reported before, but current analysis of soil crust development is limited to morphological descriptions or culturing individual isolates from soil crusts (Li et al., 2002, Veluci et al., 2006, Liu et al., 2017a, Faist et

al., 2017, Tamm et al., 2018, Belnap, 2003c). Results here for the first time use molecular methods to show significant shifts in microbial community structure across time and suggest that substantial development of soil crusts occurs within the space of a single growing season in a temperate region.

2.4.3. Taxonomic comparisons

2.4.3.1. Bacterial communities

Significant compartmental differences in taxonomic composition were seen between bacterial, protist and phototroph communities. At the whole community level, fungal communities did not exhibit significant differences between compartments. In bacterial communities, significant enrichments of Cyanobacteria, Bacteroidetes and Proteobacteria occurred in the soil crust. Of these, the average relative abundance (across all time points) of Cyanobacteria (detected by the 16S rRNA region, as opposed to 23S rRNA) was almost 15-fold higher in soil crust communities in comparison to bulk soil communities. A plethora of studies have highlighted the importance of Cyanobacterial populations in both the establishment and development of soil crusts in a number of arid ecosystems, and results here too display their taxonomic significance at the soil surface (Belnap and Harper, 1995, Mazor et al., 1996, Büdel, 1999, Belnap et al., 2003a). However, in arid ecosystems Cyanobacterial populations have been shown to constitute a much larger proportion of soil crust bacterial communities, normally ranging between 20-50% and are often the dominant bacterial phylum (Becerra-Absalón et al., 2019, Kuske et al., 2012, Büdel et al., 2016). Results here revealed that although soil crust Cyanobacterial abundance was significantly higher than bulk soil, their contribution to total relative abundance was much lower, only reaching a maximum of 6% in May, which is similar prior to studies performed in cooler climates (Pushkareva et al., 2015). However, temporal analysis revealed that the abundance Cyanobacteria substantially increased across the time course, and if soil crusts were left to develop for longer time periods comparable proportions of Cyanobacteria, which have been presented in unperturbed arid ecosystems, may have been seen. At the OTU level, *Phormidium* and *Leptolyngbya* displayed the highest fold-change in soil crust communities, and additionally SIMPER analysis identified *Phormidium* as the second highest OTU contributing to compartmental differences. Furthermore, temporal analysis revealed that *Phormidium*

was substantially increasing in relative abundance across time. A number of studies have shown that *Phormidium* species are widely distributed and can thrive in a number of habitats, such as arctic mats and aquatic biofilms; but more specifically they are also thought to be integral for the initial colonization and formation of soil crusts due to their filamentous nature, but are often replaced in more developed crusts (Pushkareva et al., 2015, Komárek et al., 2003, Dojani et al., 2014, Bowker et al., 2018). *Leptolyngbya* species are also filamentous, and prior studies have suggested that they are a dominant Cyanobacterial group in polar and desert soil crusts (Pushkareva et al., 2015, Dojani et al., 2014). The enrichment of these OTUs therefore suggests the functional capability in soil surface establishment and stabilization by Cyanobacteria. Furthermore, increasing enrichments of *Phormidium* over time suggest that soil crust communities were still under development and had not yet reached their climax community. Results here surprisingly detected no 16S rRNA signals of *Microcoleus vaginatus*, which has been prior suggested to be an abundant keystone species in early development of arid soil crust communities (Kuske et al., 2012, Rajeev et al., 2013).

The relative abundance of Bacteroidetes was on average 1.3-fold higher in soil crust communities in comparison to bulk soil communities. A number of studies have shown that members of Bacteroidetes make up a considerable amount of the heterotrophic component of soil crust communities, and perform significant roles in the production of exopolysaccharides that can contribute to soil stabilization, as well as the catabolism of different compounds due to their flexible metabolic repertoire (Abed et al., 2019, Gundlapally and Garcia-Pichel, 2006). Temporal analysis revealed that Bacteroidetes abundance was fairly stable across the time course, except for a large increase in abundance during March, which coincided with a series of fertilizer application events and a sharp increase in pH (see 2.4.6.2). At the OTU level, log-fold change analysis identified 28 Bacteroidetes OTUs that were significantly-enriched in soil crust communities, most of which belonged to the classes Cytophagia and Sphingobacteria. Previous studies have reported that Cytophagia were isolated from arid soil crusts obtained from the Northern Cape Province, South Africa, and additionally *Hymenobacter* (which 7/13 Cytophagia OTUs were assigned to) species have been successfully isolated from the Hopq Desert, China (Maier et al., 2018, Liang

et al., 2018). However, their functional significance of these taxa in soil crusts remains unstudied.

Proteobacteria were also significantly enriched in soil crust communities and were on average 1.2-fold higher than in bulk soil communities. Proteobacterial populations contributed to a large proportion of overall relative abundance (~30%) and were the second most abundant phylum in both soil compartments. Proteobacteria are abundant in a number of soils found globally, but they have also been reported to be dominant in soil crusts found in the Tengger Desert in China (Liu et al., 2017b). Prior studies have suggested that members of the Proteobacteria phylum can also excrete exopolysaccharides that support surface stabilization, and have shown enrichments of several Proteobacterial taxa in arid soil crusts (Gundlapally and Garcia-Pichel, 2006, Steven et al., 2013b). At the OTU level, SIMPER analysis identified Oxalobacteraceae as the most enriched soil crust OTU, and log-fold change analysis identified 26 Proteobacterial OTUs that were significantly enriched in soil crust communities. Prior studies in desert regions have identified a signature accumulation of Oxalobacteria in soil crusts, which has been linked with their ability to degrade and utilize oxalate, a compound excreted by desert plants, lichens and fungi (Nagy et al., 2005). A number of enriched Proteobacteria OTUs were additionally assigned to *Sphingomonas*, which have been previously isolated from desert crusts, but their overall function in soil crusts remains mostly unknown (Reddy and Garcia-Pichel, 2007). Overall, results here suggest that both phototrophic and heterotrophic bacterial taxa are critical members of developing soil crusts in a temperate agricultural soil.

In bulk soil communities, there were significant enrichments of Nitrospirae, Acidobacteria, Verrucomicrobia, Planctomycetes and Firmicutes. Of these, Nitrospirae was the most enriched bacterial phylum in bulk soil communities, which has also been shown in prior studies (Mogul et al., 2017). This further emphasizes the large differences in bacterial community structure between the soil surface and bulk soils.

2.4.3.2. Fungal communities

Whilst overall community composition did not differ between compartments in fungal communities, there were significant enrichment of Agaricomycetes and Dothideomycetes in soil crust communities. Agaricomycetes are normally a dominant class in soil microbial communities and belong to Basidiomycota, which contribute significantly to carbon cycling through the decomposition of plant litter and organic material, and their enrichment in the soil crust suggest a potential functional role in these processes (Lynch and Thorn, 2006). Prior studies have additionally found abundances of Agaricomycetes in soil crusts, but so far this has only been shown in crusts obtained from arid regions (Mueller et al., 2015). At the OTU level, 15 OTUs assigned to Agaricomycetes were significantly enriched in soil crust communities, however most of these had no annotation beyond the order Agaricales (Bates et al., 2012).

Dothideomycetes (a class within Ascomycota) were also significantly enriched in soil crust communities. Dothideomycetes are known to tolerate extreme environments, which include high UV radiation, drought and fluctuations in temperature, which may explain their abundance at the soil surface (Sun et al., 2017). A study by Steven et al. (2016) which examined fungal communities of soil crusts across a 15 km spatial gradient in the Colorado Plateau, identified that Dothideomycetes populations were always abundant despite other taxa exhibiting large variation, suggesting they are part of the core biocrust community. Furthermore, the abundance of Dothideomycetes increased ~2 fold in the final four time points, which suggests a larger taxonomic role in more developed soil crusts. Abundances of Dothideomycetes presented here represents the first study to show enrichment of this class in soil crusts outside an arid ecosystem. SIMPER analysis identified *Plenodomus biglobosus* (belonging to Dothideomycetes) as the main fungal OTU contributing to compartmental and temporal differences in fungal community composition, which had a substantial contribution to relative abundance in the final 4 months (6-10%). *Plenodomus biglobosus* has previously been described as an opportunist fungal pathogen that is capable of infecting a range of plants, that include common crops such as *Brassica* species (de Gruyter et al., 2013, Cai et al., 2014). However, it is capable of producing ascospores that are incredibly vulnerable to wind dispersal, therefore enrichments at

the soil surface are likely due to dispersion. Fold-change analysis identified *Xylographa vermicularis* as the most enriched OTU in soil crust communities. Members of *Xylographa* are amongst the most abundant lichen mycobionts in temperate regions, and their abundance at the soil surface is likely linked with the development of soil crusts through the association with lichens (which were enriched in soil crusts in later months) (Spribille et al., 2014). Additionally, Endogonales (Mucoromycotina) were also significantly enriched in soil crust communities. Although Endogonales are currently poorly defined, they form arbuscular mycorrhizal associations with higher plants, and mycorrhiza-like associations with non-spermatophyte plants and bryophytes including liverworts, ferns, lycophytes, hornworts and mosses (Chang et al., 2019, Desirò et al., 2017). Overall, although community-scale analysis did not identify large community-scale shifts in fungal taxa, changes in abundance of fungal species are likely to be critical for development into mature crusts through symbioses with autotrophs.

2.4.3.3. Protist communities

The roles of protist in soil microbial communities are rarely considered in the current literature. Results in this chapter presents one of the few studies that examine soil crust protist communities, and findings here revealed for the first time that the soil crust is enriched of a number of protist taxa in a temperate soil. Rhizaria (SAR) were the only taxa that were significantly enriched in soil crust communities at the ‘ta1’ taxonomic level. Rhizaria are an extremely diverse supergroup that are composed of several phyla, and consequently are functionally diverse. They are capable of forming long pseudopodia, which are cytoplasmic hair-like extensions that are capable of forming networks that can reach over a centimeter in diameter, and their enrichment in the soil surface may too contribute towards soil surface stability (Geisen et al., 2018, Burki and Keeling, 2014). Furthermore, Rhizaria are known to form photosymbiotic relationships with microalgae and some groups are actually considered algae themselves, which may suggest they are associated with the photosynthetic capability of the soil crust (Burki and Keeling, 2014). A large number of Rhizaria OTUs were also enriched in the soil crust, which mainly consisted of Cercozoa. Cercozoa are vastly abundant in soils and contain large groups of bacterial grazers, and due to an increased recorded microbial biomass at the soil surface, this may be a more desirable

habitat for these organisms. A recent study by Fiore-Donno et al. (2018) designed a primer set to assess Cercozoan community composition of soil crusts from arid soils in Chile and a deciduous mixed forest in Germany. Fiore-Donno and colleagues showed that biocrusts from varying climatic regions contained an abundance of diverse Cercozoa taxa, but no functional information was inferred from this experiment. Results here too reveal that Cercozoa make up a considerable proportion of soil crust protist communities, but in a temperate agricultural setting. Temporal analysis revealed that Amoebozoa were significantly enriched across time in soil crust communities, which were notably abundant in the final three time points. Amoebozoa are known to perform considerable roles in bacterial grazing in soils, and their abundance in the final time points (when biomass C and N were high in the soil surface) may explain their abundance. Prior studies have shown that Amoebozoa taxa make up considerable proportions of heterotrophic populations, but current literature on this topic is restricted to two studies that sampled from polar (Norway) and arid (South Africa) soils (Rippin et al., 2018, Dumack et al., 2016). Results presented here suggest that protist communities contain significant enrichments of numerous taxa in temperate soil crusts, including groups that have previously reported roles in bacterial grazing and photosynthesis. However, protists are vastly understudied in comparison to bacterial and fungal communities, and current taxonomic annotations are relatively poor.

2.4.3.4. Phototroph communities

Analysis of phototrophic communities (determined by the 23S rRNA subunit) revealed significant enrichments of Oscillatoriales, Dinophyceae and the PX clade were in the soil surface at the class level. Oscillatoriales are a group of filamentous Cyanobacteria, whereas Dinophyceae (Dinoflagellates) and the PX clade (Stramenopiles) are groups of algae. All of these taxa have been shown to be abundant components of early colonizing phototroph communities in both arid and arctic soil crusts (Pushkareva et al., 2015, Condon et al., 2019), and results here too suggest a significant taxonomic role in a temperate soil. At the OTU level, *Leptolyngbya* (also identified from 16S rRNA sequences, see 2.4.1.3.1), *Vaucheria litorea*, *Klebsormidium flaccidum* and *Nodosilinea* were identified as highly enriched soil crust taxa; and additionally *Vaucheria litorea*, *Klebsormidium flaccidum* and

Microcoleus vaginatus were identified as displaying the greatest temporal variation in the soil crust. Nodosilinea belong to the same family as *Leptolyngbya* (Leptolyngbyaceae) and are also filamentous in nature and may perform similar functions as *Leptolyngbya* in soil surface stabilization (Türk and Gärtner, 2003). *Vaucheria litorea* is a species of filamentous, tubular yellow-green algae which is associated with early soil surface colonization and formation of yellow-green algal mats (Türk and Gärtner, 2003). *Klebsormidium flaccidum* and *Microcoleus vaginatus* are both widely reported as keystone species of soil crust colonization in a number of arid soils due to their excretion of mucilage and exopolysaccharide that binds and stabilizes the soil surface (Veluci et al., 2006, Bu et al., 2015). Enrichments of these taxa in the soil crust, notably in the last four months, highlights the functional capability of phototrophic communities for stabilizing the soil surface.

Furthermore, significant detection of Bryophyta sequences (which were exclusively moss and lichen sequences) were detected in soil crust communities, which increased substantially in the final three months, reaching a peak relative abundance of 28%. This result correlates with visual assessments and morphological analysis of soil cores, within which after 8 months the soil surface was almost exclusively covered by a green biotic layer and contained an abundance of mosses. This result is one of the first studies that shows substantial development of mosses in soil crusts within a relatively short time span (<1 year). For example, a study by Li et al. (2002) revealed that progression towards moss-dominated soil crusts occurred after 8-10 years in the Tengger Desert, China (an arid ecosystem), but only if left unperturbed. A similar timescale to show the presence of mosses in soil crust development has additionally been described in a number of studies, all of which have been in non-agricultural arid ecosystems (Maestre et al., 2011, Belnap and Eldridge, 2003, Dojani et al., 2011). The quickest progression of a moss-dominated soil crusts was recorded by Wang et al. (2009) in the desert areas of Mongolia, in which the presence of mosses occurred after 2 years of soil crust formation. However, this study was performed in a semi-artificial system, whereby soils were inoculated with cyanobacteria in order to simulate soil crust growth and were protected from physical perturbations by using sand barriers and straw checkboards. The timescale for natural biological soil crust development in temperate ecosystems has been recorded to occur much quicker than in arid regions.

For example, Fischer et al. (2010) described the development of 2.5 mm thick soil crusts, which were dominated by cyanobacteria and algae, three years after a perturbation event in a natural system containing temperate soils from Germany, although no presence of mosses were recorded. A recent study by Crouzet et al. (2019) recorded the rapid progression of algal biocrusts after just 50 days in a mesocosm experiment using agricultural temperate soils in France. Although this study was performed in an experimental system with controlled environmental conditions, this highlights the capacity for the rapid development of soil crusts in temperate regions. Overall, results shown here highlight the quickest recorded development of soil crust communities to show the presence of mosses in any soil ecosystem.

2.4.4. Microbial co-occurrence networks

Only a small number of studies have explored soil crust communities using co-occurrence microbial networks and results here represent the first use of this method in temperate soil crusts. Microbial co-occurrence network analysis revealed that soil crust networks were more complex in comparison to bulk soil networks. Both soil compartments had a higher proportion of positive (~60%) than negative (~40%) connections, suggesting the majority of microbial correlations were potentially beneficial, such as symbioses or commensalism. However, soil crust communities displayed a higher level of connectivity and centrality but had a lower number of clusters in comparison to bulk soil communities. Similar findings were shown by Chilton et al., (2017), who examined bacterial networks from arid soil crusts in New South Wales and found that undeveloped and bare soil crusts had lower connectivity than highly developed crusts. Whilst a higher level of connectivity is normally considered a sign of a more robust community, higher connectivity combined with high centrality and low modularity are typical of microbial communities that are less stable under disturbance events, which have been shown in bacterial responses to drought (de Vries et al., 2018, Santolini and Barabási, 2018). Therefore, connectivity features shown in soil crust microbial networks here are typical of microbial communities that are less stable under disturbance events.

Soil crust networks had a higher number of nodes (136) in comparison to bulk soil networks (77), which suggests that bulk soil communities exhibit a lower number of

important interacting OTUs. The number of nodes in both compartments however were represented by a similar proportion from each amplicon, in which ITS amplicons had the highest proportion, followed by 16S rRNA, 18S rRNA and 23S rRNA (in that order) amplicons. Prior work has suggested that fungal co-occurrence networks are more stable in perturbation events, which may explain their abundance (de Vries et al., 2018). Surprisingly, 23S rRNA nodes (phototroph amplicons) represented the least proportion of nodes in soil crust networks, suggesting that a lower number of 23S rRNA OTUs are correlatively important. Cartographic representation of soil crust nodes revealed that *Meliniomyces* (ITS) was classified as a ‘module hub’, which are normally considered as keystone taxa due to their importance in network topology (Cong et al., 2019). *Meliniomyces* are groups of mycorrhizal fungi that can form symbioses with plants in a structural plant-fungal symbiotic complex, known as the *Rhizoscyphus ericae* aggregate, or in plant roots (Hambleton and Sigler, 2005). However, these associations have previously been shown to only occur with plants belonging to the Ericaceae family (such as ericaceous shrubs and ectomycorrhizal trees), which were not present in this system; and additionally there have been no prior studies that have reported abundances of *Meliniomyces* in soil crusts. Whilst results here suggest an important functional significance of *Meliniomyces* in the soil crust, it is not clear the exact role may be. Further highly-connected ITS OTUs in soil crust communities included the potential plant pathogen *Plenodomus* and plant symbiotic Endogonomycetes (section 2.4.1.3.2), which were both previously identified as significantly enriched OTUs in soil crust communities – suggesting plant-microbe interactions are important in shaping soil crust communities (Bates et al., 2010, Adhikari et al., 2014). For 16S rRNA, 18S rRNA and 23S rRNA amplicons, *Phormidium*, Rhizaria, *Vaucheria litorea* and *Klebsormidium flaccidum* were all identified as highly connected OTUs, which were all taxonomically enriched in soil crust communities (section 2.4.1.3). Taxonomic enrichment, coupled with a large number of correlative connections emphasizes the proposal that these taxa are important members for soil surface stabilization and recruitment of other taxa to soil crusts, which has been suggested by a number of prior studies in arid regions. Furthermore, these results highlight that whilst fungal communities did not display significant differences between compartments, a number of fungal taxa have important interactions in soil crust communities.

It was hypothesized that a decreased connectivity would occur in soil crust communities over time, as more dominant phototrophic OTUs would become more prevalent but no trends were seen in connectivity across time. Furthermore, there were no differences in the proportion of positive and negative connections in either compartment with time, but the ratio always remained roughly 60% positive, and 40% negative. This suggests there was a higher level of beneficial relationships (e.g. symbiosis) as opposed to negative relationships irrespective of time. No differences were seen in the number of nodes, or the relative proportion of nodes represented by each amplicon across time. Cartographic representation (results not presented) revealed no large changes in connected taxa between months in either compartment, and the same OTUs discussed earlier had the highest connectivity regardless of time. These results are contrary to results presented by Chilton et al., (2017), who showed that co-correlation networks of arid soil crusts displayed significant structural changes in network topology as soil crust development increased, which mainly consisted of increased connectivity. However, this study only examined bacterial co-occurrence networks, and did not consider the fungal, protist and non-bacterial phototroph communities. Overall, these results highlight that compartmental, not temporal, differences are more important in determining network topology.

However, microbial network analysis remains a young field in microbial ecology and there is yet to remain a universally accepted method to generate microbial correlation networks. Network analysis is extremely prone to thresholding techniques and the method of filtering OTU tables (for example, rarefaction, use of multi-amplicon networks, removal of rare taxa) can significantly impact network structure and topology (Ju et al., 2014, Berry and Widder, 2014). Furthermore, the community remains split over the optimal correlation algorithm to use, for example SparCC or SPIEC-EASI, and the use of different algorithms can additionally contribute to differences in network topology. As a result, results of microbial network analyses should be assessed with caution.

2.4.5. Community assembly processes

Only a small selection of studies have previously used null models to quantify community assembly processes of microbial communities over time, and this experiment represents the first study to use null models to quantify community assembly processes in soil crust communities from any environment. Results presented here revealed that fungal, protist and phototroph communities were mainly governed by stochastic community assembly processes, whereas bacterial community assembly processes were governed by both deterministic and stochastic processes. Furthermore, clear differences were seen between compartments in community assembly processes in every amplicon, within which bulk soil communities were all dominated by random genetic drift, or drift and selection (bacteria) – which suggests assembly in the bulk soil is largely random. Bacterial soil crust communities were initially dominated by homogenous selection, which occurs in environments within which there is a consistent geophysical selection pressure, resulting in a low species turnover (Stegen et al., 2015). Similar findings were presented by Tripathi et al. (2018), who quantified community assembly processes from soil samples obtained from various arctic glacial chronosequences in Svalbard (Norway). Tripath and colleagues showed that in early successional stage soils, bacterial communities were initially dominated by homogenous selection (30-100%) in 4 geographically separated chronosequences, but gradually became dominated by stochastic processes over time. Results presented here show similar trends to Tripath et al., (2018), and in the context of this study suggest that in undeveloped soil crusts there is a strict environmental pressure at the soil surface that is restricting propagation to select biota. However, over time bacterial, fungal, protist and phototroph soil crust communities all became dominated by dispersal limitation. Dispersal limitation occurs when there are weak selection pressures and physical barriers constrain organisms from different populations from mixing. In this situation, this result suggests that in developed soil crusts there is dispersal barrier that is preventing assembly at the soil surface, and that the ability to disperse to the soil surface by chance events is the main factor determining assembly. Together, these results support the proposal that the ability to assemble at the soil surface is initially restricted to well adapted taxa due to the strict environmental pressures, but in developed soil crusts these selective pressures are more negligible and the stochastic events determine assembly. As this is the first study

to use null model approaches to quantify community assembly processes of soil crust communities across time, further research is warranted to test whether results shown here hold true in other ecosystems.

2.4.6. Soil physiochemical properties

2.4.6.1. Soil moisture

The soil surface is often described as an extreme environment due to the large fluctuations in physical pressures it experiences (Csotonyi et al., 2010). Results here suggest that the immediate soil surface and bulk soil were under similar environmental stresses. Significant differences in recorded physical parameters between compartments were only seen in November (soil moisture), March (pH), April (pH) and June (pH). Large variation was seen in soil moisture content across the time course and the same trends were seen in both soil crust and bulk soil samples. Soil moisture is known to significantly impact important soil functions, for example an increase in soil moisture is often associated with higher levels of carbon and nitrogen fixation (Niederberger et al., 2015, Zielke et al., 2002). A number of studies have shown that soil moisture is a highly important driver in shaping microbial communities. For example, a recent study by Frindte et al. (2019) showed that soil moisture was a main long-term driver in shaping community structure in arctic-alpine soils. However, redundancy analysis here revealed that soil moisture was not the main contributor to explained variance of microbial communities in both compartments, and only contributed significantly in bacterial and fungal community composition.

2.4.6.2. Soil pH

Extremely large fluctuations were seen in pH across the time course, which ranged by over a pH unit in soil crust samples and by just over half a pH unit in bulk soil samples. Seasonal fluctuations in soil pH are widely reported, which can often be associated with microbial processes, for example soil acidification mediated by nitrification, or by non-biotic processes such as rainfall events and nutrient leaching (You et al., 2015, Miegroet and Cole, 1984, Wang et al., 2015). Soil pH is arguably the most important factor in shaping microbial communities, and a variety of studies have examined the impacts of soil pH on community structure over local and continental scales (Fierer and Jackson, 2006). For example, Bartram et al. (2014) identified sizable, scalable

shifts in the abundance of a number of soil bacterial phyla, including Acidobacteria, Alphaproteobacteria, Verrucomicrobia and Gammaproteobacteria, across a pH gradient ranging from 4.5 – 7.5. Results here revealed that variation in soil crust communities (all amplicons) was always significantly impacted by pH and additionally in bacterial and phototroph bulk soil communities.

More specifically, a significant decline in pH was seen in both compartments from November to May, but an abnormal spike in pH was seen only in the soil surface in March, which coincided with the application of nitrogen-based fertilizers. Application of nitrogen-based fertilizers can either acidify and alkalify soils depending on type of fertilizer that has been used (Barak et al., 1997). Urea fertilizers, which were applied one day prior to sampling in March, can temporarily increase local soil pH and can lead to high rates of volatilization due to ammonium conversion (up to 50% loss of nitrogen) (Jones, 2013). However, microbial oxidation of nitrogen-based fertilizers over greater time periods has been shown to drastically reduce soil pH, which was witnessed in subsequent months (Schroder et al., 2011). Therefore, it is likely that the short-term increase in pH at the soil surface in March samples was due to fertilizer application. Very little work has been performed on the impacts of fertilizer application and soil crust development, but it has been suggested that large chemical application may be detrimental to soil crust health (Belnap, 2003). In March, soil crusts appeared lighter in colour and experienced a sharp increase in chlorophyll *a* content in comparison to general trends seen in prior months. Furthermore, a large shift was seen in bacterial community composition and a reduction of bacterial alpha diversity occurred during this month, suggesting the large shift in pH altered bacterial community structure. However, these changes were only seen in bacterial soil crust communities, which suggests that fungal, protist and phototrophic soil crust communities were more resistant to pH fluctuations. A number of studies have shown that soil bacterial communities are less resistant, but more resilient, to environmental perturbations than fungal communities (Bapiri et al., 2010, de Vries et al., 2018). For example, Barnard et al. (2013) highlighted that community composition of bacterial communities were significantly different after drought treatments but recovered rapidly upon re-wetting, however fungal communities remained relatively unchanged throughout the whole process. However, the impacts on soil protist and phototroph

communities are less well known. Furthermore, the compositional shift that was seen in bacterial communities in March is likely attributed to the large increase (>2-fold) in the relative abundance of Bacteroidetes, which are known to increase substantially with a rise in soil pH (Lauber et al., 2009). A number of studies have shown that Bacteroidetes make up a considerable amount of the heterotrophic component of soil crust communities, and may perform significant roles in the production of exopolysaccharides that can contribute to soil stabilization, and the catabolism of different compounds due to their flexible metabolic repertoire (Abed et al., 2019, Gundlapally and Garcia-Pichel, 2006). Overall, this result highlights that temperate agricultural soils are prone to large changes in pH over time, which likely influence community structure.

2.4.6.3. Physical perturbations to the soil surface

In arid desert ecosystems, prior studies have shown that it can take up to 25 years for soil crusts to fully develop into communities containing mosses and lichens, which is only possible if they are left unperturbed (Li et al., 2002). However, in different climatic regions the development of soil crusts into mature late-successional crusts can occur in a much shorter time scale (Wang et al., 2009, Fischer et al., 2010, Crouzet et al., 2019). Results here showed that development of soil crusts to complex communities, including mosses, can occur within just a few months within an agricultural system. However, during this time frame soil crusts were exposed to a number of major physical disturbance events. Shortly prior to the initial time point, the soil surface was subject to herbicide application (glyphosate, pendimethalin, and Diflufenican) and 6-inch minimum tillage (the process of mechanically replacing surface soil with lower layers). Despite the benefits of tillage in increasing surface nutrients and preparing the soil for drilling, tillage is currently a widely-debated and controversial practice as it can result in the mineralization soil organic carbon, increase surface wind and water erosion, increase soil compaction and reduce water infiltration, all of which are associated with a reduction in soil health (Haddaway et al., 2017, Gao et al., 2016). On top of this, differences in tillage regimes have also been shown to significantly alter the composition and functionality of microbial communities (Smith et al., 2016, Degruene et al., 2017). Soil crusts are readily destroyed by tillage practices and a recent study by Durham et al. (2018) suggested that past tillage practices had

long-lasting detrimental impacts on the species richness of soil crust lichen and bryophyte functional groups from grassland soils in Western Montana (USA). Furthermore, the application of herbicides also has been reported to detrimentally impact soil crust communities (Belnap and Eldridge, 2003). For example, Zaady et al. (2004) concluded that application of simazine, a pre-emergence herbicide, disturbed soil crust populations, resulting in a reduction of microphytes and mosses, thus leading to increased soil erosion.

Results here revealed that there were no significant compartmental differences at the first sampling point in November, in microbial community composition (NMDS and Bray-Curtis comparisons), alpha diversity (all amplicons), biomass C/N and chlorophyll *a* content, and additionally no crusts were visualized morphologically shortly after tillage and pesticide application. This suggests that at the start of the growing season soil surface and bulk soil communities were the same, and that no pre-existing soil crust development persisted from prior seasons. Therefore, the physical agricultural disturbances of crop harvest, tillage and pesticide application resets the development of crusts after each growing season, creating a level-playing field where soil crust pioneers can re-establish.

2.4.7. The relative contributions of edaphic factors, time and distance in shaping microbial community composition

No single edaphic factor was determined to contribute the most towards community composition in soil crust and bulk soil samples. In bacterial and protist soil crust communities, time was the most important factor for determining community composition. However, time displayed strong, significant correlations with chlorophyll *a* content and temperature, which were the most important factors determining fungal and phototroph community composition. Time is often considered the most important factor for the development of soil crust communities, but only if communities can remain unperturbed (Belnap and Eldridge, 2003). In fungal communities, a large proportion of variation correlated with chlorophyll *a* content, which suggests phototrophic biomass is important for explaining variation in fungal soil crust communities. For phototrophic soil crust and bulk soil communities, temperature, which is known to significantly influence photosynthesis and nitrogen

fixation, explained a large proportion of explained community variance (Stal, 2017). Less variation was explained by measured variables in bulk soil communities, suggesting that bulk soil communities are more resistant to environmental variables. Time (bacteria and fungi) and temperature (phototroph) were the most important factors explaining compositional variance in bulk soil communities.

Redundancy and distance-decay analysis revealed that within-field geographical distance did not contribute significantly towards microbial community composition in soil crust or bulk soil microbial communities. A number of studies on soil microbial communities have shown that local-scale distance effects can significantly influence community composition. For example Deakin et al. (2018) identified significant shifts in community structure, notably the abundance of several fungal OTUs, across just a 20 meter scale. Furthermore, previous studies have shown that soil crusts are spatially heterogenous on a local scale, and community composition can differ greatly even in soil crusts that are several meters apart (Steven et al., 2013b). However, local-scale spatial studies of soil crusts have only been performed in unperturbed desert environments, where there is heterogenous distribution of soil nutrients that can result in accumulation of more-developed soil crusts in hotspots of higher nutrient availability (Goberna et al., 2007). Conversely, in agricultural systems the distribution of nutrients is often more homogenous (due to processes such as tillage), which may reduce variation in the development of soil crusts. Furthermore, the effects of pooling samplings (within each 'W' shape), which aimed to reduce microsite variation (variation under the meter scale) may have reduced distance-decay impacts. Overall, results here conclude that local-scale distance effects over a 40 meter transect did not significantly influence soil crust community composition.

2.4.8. Conclusions

Results presented in this chapter represent the first use of next-generation sequencing technologies to track the development of biological soil crust and bulk soil communities in a temperate agricultural ecosystem. The first aim of this chapter was to determine whether the soil crust and bulk soil microbiome are compositionally different in a temperate agricultural ecosystem. Amplicon DNA sequencing results revealed that there were significant compartmental differences in bacterial, protist and

phototrophic community composition and diversity. Soil crust communities had substantial enrichments of an array of taxa, which included Cyanobacteria, Agaricomycetes, Dothideomycetes, Rhizaria, Oscillatoriales and Dinophyceae. At the OTU level, enrichments of taxa associated with soil surface stabilization were seen in soil crusts relative to bulk soil, which included *Phormidium*, *Vaucheria litorea* and *Klebsormidium flaccidum*. These results suggest that the soil crust has a distinct microbiome relative to bulk soils in a temperate agricultural ecosystem.

The second aim of this chapter was to determine whether significant temporal dynamics in diversity and composition occurred in soil crust communities. Results here revealed that significant reductions in alpha diversity as well as large shifts in community composition were seen across time in bacterial, fungal, protist and phototrophic soil crust communities. Furthermore, results here showed the quickest progression of soil crust development to contain an abundance of mosses in the current literature, emphasizing that substantial development of soil crusts can occur within a highly perturbed agricultural ecosystem. However, several agricultural management practices, such as harvest, tillage and chemical application, were shown to impact community structure of soil crusts.

The third aim was to examine local scale variation in soil crust community composition. Redundancy analysis revealed that there were no significant contributions of local-scale distance effects (over a 40 meter transect) to explained variance in soil crust community composition in any amplicon.

The fourth aim was to examine the relative influences of time, geographical distance and environmental factors in determining soil crust community composition on a local scale. Redundancy analysis revealed that no one single factor explained the most variation in soil crust microbial community composition for all amplicons; but time (bacterial, protist), temperature (phototroph) and chlorophyll *a* (fungal) were the most important factors in different amplicons.

The fifth aim was to compare co-occurrence networks of microbial communities in soil crusts and bulk soils to identify which key taxa are conserved across

compartments. Soil crust microbial networks were more complex, within which they had a higher number of nodes, edges and connectivity, than bulk soil networks. Cartographic representation identified *Meliniomyces* and *Plenodomus* were highly connected and correlatively-important soil crust OTUs, suggesting their functional importance in soil crust communities. Cartographic representation and analysis of network topology did not reveal any clear trends in soil crust networks over time.

The last aim was to quantify the relative contributions that each fundamental ecological process has in driving soil crust and bulk soil microbial community assembly. Using null models to quantify community assembly processes, results showed that deterministic processes dominated soil crust bacterial community composition in early time points, but stochastic processes became dominant across time. For fungal protist and phototroph communities, stochastic processes (dispersal limitation) dominated regardless of time point. Bulk soil community assembly was dominantly driven by genetic drift.

Overall, the ecological importance of soil crusts is not currently considered in temperate agricultural regions, and due to the previous studies showing the ecological importance of soil crusts in arid regions, work here suggests that this distinct compartment should be considered in agroecosystems. However, this chapter represents a single case study, and further work is warranted to try and identify if results found here are reproducible and applicable to other temperate agricultural fields.

CHAPTER 3 – THE INFLUENCE OF BIOLOGICAL SOIL CRUST DEVELOPMENT ON MICROBIAL COMMUNITIES AND SOIL STRUCTURE

3.1. Introduction

3.1.1. Soil physics at the soil surface

A small number of studies have explored structure at the soil surface. As discussed in section 1.2.5.3, the soil surface is exposed to environmental conditions that are not experienced in bulk soil, such as high exposure to UV light, mechanical pressure from rain water droplets and trampling, and extreme fluxes in temperature, which have been suggested to impact on soil structure, as well as community structure (Csotonyi et al., 2010). A recent study by Armenise et al. (2018) used X-ray CT to investigate the structural changes in artificial physical soil crusts generated after exposure to increasing durations of rain water. They found that in three different soil types, the soil surface experienced a zone of reduced porosity, termed the ‘soil seal’, which increased in thickness after exposure to longer rainfall events (Figure 3.1) (Armenise et al., 2018). However, this study exclusively examined physical soil crusts on sieved repacked soil cores, and therefore provides little context in reference to naturally occurring biological soil crusts.

Figure 3.1. Soil seal thickness after increasing durations of rainfall in three soil types. Seal thickness was calculated as the depth at which porosity returns to within the 95% confidence interval of reference (bulk) soil from X-ray CT images. ZCL: Silty clay loam, SZL: Sandy silt loam, SL: Sandy loam. D2-D14 indicates duration of rainfall (minutes). Figure obtained from Armenise et al. (2018).

Raanan et al. (2016) used X-ray microtomography to examine the soil structure of biological soil crusts obtained from hyper-arid soils obtained from Egypt and Israel. This study revealed that in the presence of soil crusts, two morphologically distinct layers were produced – which they termed the top crust, and the sub crust (Figure 3.2). The top crust was classified as a seal found at the immediate soil surface that exhibited exceptionally low porosity, whereas the sub crust was directly below the top crust and contained a large number of vesicular pores, suggesting higher biological activity in this region (Raanan et al., 2016).

Figure 3.2. Structure of a biological soil crust obtained from the Negev desert. Structural analysis was performed using X-ray microtomography on Cyanobacterial-dominated crusts. Separation of a top crust (~2mm) and subcrusts (~1-2mm) were identified. Image taken from Raanan et al. (2016).

Couradeau et al. (2018) used X-ray microtomography to examine the *in-situ* response of desiccation on the soil structure of developed soil crusts sampled from the Moab, Utah. Using imaging techniques, Couradeau and colleagues examined the proportion of air, water and mineral particles in an exopolysaccharide Cyanobacterial bundle sheath isolated from the soil surface of an arid soil crust. They showed that the Cyanobacterial bundle sheath was capable of remaining hydrated whilst surrounding soil pores outside the sheath were rapidly drying, suggesting a significant functional role of soil crusts in retaining soil moisture in arid soils.

Lastly, Menon et al. (2011) used X-ray microtomography to quantify water flow in Cyanobacterial and Cyanobacterial-lichen dominated soil crusts isolated from Botswana. They concluded that less developed (Cyanobacteria-dominated) crusts had a higher porosity and were more prone to mechanical disturbance in comparison to more developed crusts (Cyanobacterial-lichen) (Menon et al., 2011). Further, the soil surface also identified as a region of lower porosity and increased density in comparison to bulk soil (Figure 3.3), which agrees with conclusions made by Armenise et al. (2018) and Raanan et al. (2016) (Menon et al., 2011).

Figure 3.3. A 2D cross section of a cyanobacterial-lichen dominated soil crust. Structural analysis was performed using X-ray microtomography on a cyanobacterial-lichen dominated crust obtained from Botswana. Scans had an image resolution of $3\ \mu\text{m}$ ($5805 \times 2700\ \mu\text{m}$). Image obtained from Menon et al. (2011).

3.1.2. Chapter 3 aims

Prior studies using non-invasive techniques to examine structure of soil crusts have all highlighted that the soil surface exhibits a distinct region of reduced porosity (in comparison to lower soils) in the presence of developed soil crusts (Menon et al., 2011, Armenise et al., 2018). Further, the clear structural differences displayed at the soil surface have been suggested to have important implications for physical processes, such as water retention or distribution (Menon et al., 2011, Armenise et al., 2018). However, current literature studying soil structure of naturally developed biological

soil crusts have been performed exclusively in arid desert ecosystems, which have different chemical, biological and structural compositions in comparison to temperate soils (see section 1.4). Furthermore, current literature investigating soil structure of biological soil crusts have sampled from unperturbed deserts, have used repacked soil cores or have made no biological links, particularly by examining shifts in the community structure (Armenise et al., 2018, Raanan et al., 2016, Menon et al., 2011). These studies therefore provide little insights into real world agricultural environments, and it remains unknown how the structure of biological soil crusts develop outside these systems. This chapter investigated these current research gaps in understanding how soil structure of biological soil crusts develop in a real-world temperate agricultural system, and how whether this differs to bulk soil. The aims of this chapter were:

To determine changes in the physical soil structure at the soil surface across a growing season in an agricultural field containing soil crusts

1. *To determine the depth at which the soil crust influences structure in the soil profile*
2. *To determine changes in soil pore abundance, distribution, connectivity and structure throughout the soil profile*
3. *To determine changes in soil pore abundance, distribution, connectivity and structure across a growing season*
4. *Investigate how annual trends in crust community development and community composition determined in chapter 2 occur at a different study site*

From these aims, the following hypotheses were made:

1. *Soil crusts have a larger influence on soil structure than the ~2 mm depth in which they occupy*
2. *Soil porosity decreases with increasing depth, and soil pore characteristics are more complex at the soil surface*
3. *Soil pore characteristics significantly change across time in the soil crust*
4. *Soil crust microbial communities display similar trends (to those identified in chapter 2) in composition across time and between compartments on a different sampling site*

3.2. Methods

3.2.1. Location, experimental design, sampling methods and core processing

Soil samples were obtained from Big Cherry field (Figure 3.4.A) at the Warwick Crop Centre, Wellesbourne, United Kingdom (52.205086°N, -1.603735°E), under a winter barley crop. The sampling site was subject to 6-inch conservational tillage in the second week of October 2017, and the crop was drilled the following week. Sampling occurred three times in the 2017-2018 growing season (November, March and July), which were chosen based on first year metadata in order to select three developmental stages of soil crusts. At each sampling date, a 40 metre transect was laid out along the field, which was separated into five 4 m x 8 m boxes (Figure 3.4.B). Sampling was conducted according to the methodology described in section 2.2.1. A plastic soil core (10 cm x 5 cm x 10 cm) was taken at the centre of each ‘W’ (letter ‘C’) for X-Ray Computed Tomography (CT) scanning. X-Ray CT cores were sampled carefully without forceful impact (no use of a mallet) to avoid compaction of the soil, which would impact soil pore metrics. Soil cores were kept at 4 °C overnight and X-Ray CT scanning was performed the following day for all time points. Processing of soil cores and measurements of chlorophyll *a*, biomass C/N, soil moisture, soil pH and bulk density were performed using methods described in sections 2.2.1 – 2.2.6.

Figure 3.4. Sampling location and strategy. A) Location of sampling site for this chapter (1) and chapter two (2) (Google maps, satellite image). B) Sampling and pooling strategy. Sampling was performed in November 2017, March 2018 and July 2018. Cores were taken at each circle and samples were pooled within each ‘W’. Additional bulk density cores were taken at each circle.

Historical information on the weather conditions at the time of sampling were also recorded (Wellesbourne weather station, UK), and additionally the agricultural management practices (tillage and chemical application) used throughout the year were recorded following discussions with the land manager (Table 3.1).

Table 3.1. Chemical application at the sampling site.

Date	Application	Amount
22/10/2017 –	Flufenacet, Pendimethalin (Crystal®)	4 L Ha ⁻¹
26/10/2017	Diflufenican	0.1 L Ha ⁻¹
	Glyphosate	1 L Ha ⁻¹
21/03/2018	Granular Urea	225 Kg Ha ⁻¹
07/04/2018	Ammonium Nitrate	200 Kg Ha ⁻¹

3.2.2. DNA extraction, Illumina MiSeq library preparation, DNA amplicon sequencing, bioinformatics and statistical analysis

DNA extraction, Illumina MiSeq library preparation, DNA amplicon sequencing, bioinformatics and statistical analysis were all performed using methods described in sections 2.2.7 – 2.2.10. Following DNA sequencing and filtering of raw reads, OTU tables were rarefied using a single random subsampling algorithm to the lowest number of reads across all samples for each amplicon, which was 6,300, 16,000, 1,600 and 2,900 for 16S rRNA, ITS, 18S rRNA and 23S rRNA, respectively.

3.2.3. X-ray CT scanning

Soil cores were scanned using a Phoenix vltomelx M scanner (GE Sensing and Inspection Technologies GmbH, Wunstorf, Germany) at 140 kV at a current of 180 μ A. Each run lasted 40 minutes, and a total of 2400 projection images were taken (with an average 4 images, skipping 1 for each projection) with 200 ms detector timing at a resolution of 30 μ m.

3.2.4. Image processing

Scanned images were reconstructed using Phoenix datoslx2 rec reconstruction software (version 2.2.3; GE Sensing and Inspection Technologies GmbH, Wunstorf, Germany). Radiographs were visually assessed for any movement during the scan by comparing the first and last projection image using the autoscantoptimizer module,

and were manually adjusted for any movement or shrinkage if necessary. Any cores displaying a substantial amount of movement during the scan were re-scanned. Reconstructed X-Ray CT images were loaded into VG StudioMAX and converted to 16-bit format (version 2.2.0; Volume Graphics GmbH, Heidelberg, Germany). The image intensity was normalized and scaled by mapping the intensity of pore space to a value of 5,000, and material (column wall) to 15,000. Images were aligned and stacked to generate 3D volumes. 3D volumes were levelled to remove any tilts, and XY-orientated cross sections from a region of interest (which encapsulated the top 3 cm of each core) were exported as TIF images. TIF image stacks were loaded into ImageJ (FIJI), and any low-quality images occurring at the start or end of the stacks were removed (Schindelin et al., 2012). A uniform region of interest was selected in each sample from the middle of the core (24 x 24 x 30 mm) to avoid edge effects from sampling. Image stacks from the region of interest were then resliced to produce a ZX orientated image stack. Contrast was enhanced across the image stack by 0.4%, using the stacked histogram setting. Images were converted to binary images (air converted to black, and matter converted to white) using the Otsu thresholding algorithm (Otsu, 1978). A separate threshold was not made for each image, but was instead made from the whole stack. Visual assessment was performed on all slices to ensure there were no thresholding errors, and any errors were manually amended. The scale of the image stack was then defined according to the scanning resolution (1 voxel set to $0.3 \mu\text{m}^3$).

3.2.5. TopCap

Processed binary image stacks (ZX-orientation) were loaded into TopCap for further analysis. TopCap is an ImageJ/FIJI plugin that captures the shape of soil surface, extracts a binary image mask of this shape, and measures various pore metrics in the shape of the mask iteratively through the soil profile at a specified thickness (Garbout et al., 2018) (Figure 3.5). Rolling curve parameters were tested on each core, and a value of 15 was selected to use on all cores, as this had the highest resolution whilst not producing artefacts. Each iteration was performed in groups of 5 slices ($150 \mu\text{m}$), starting from the soil surface to 3 cm in depth. The first groups at the immediate soil surface (up to 0.9 mm) were removed from all further analyses as this was shown to overestimate porosity by producing a mask above the soil surface.

Figure 3.5. The TopCap process. X-ray computed tomography images of an example soil core displaying the TopCap process. A) The soil surface is captured using a specific rolling curve parameter. B) A mask is generated (of specified thickness) that captures the shape of the soil surface. C) Soil pore metrics are calculated within the mask. D-F) Steps B and C iteratively repeat until the bottom of the soil core is reached, capturing soil pores at each depth. The TopCap macro was obtained from Garbout *et al.*, (2018).

3.2.6. Particle analysis and shape descriptors

Pore metrics were calculated using Minkowski functionals, according to Vogel *et al.* (2010). Porosity, pore count, diameter, thickness, circularity, surface area and pore shape descriptors were all calculated from binary image stacks generated from each TopCap output using the ‘particle analyzer’ tool within BoneJ (Doube *et al.*, 2010). Particle analysis of individual slices was performed using the ‘analyze particles’ function within the FIJI extension of ImageJ (Schindelin *et al.*, 2012). Binary image stacks for each depth were purified and connectivity was calculated using the ‘connectivity’ tool within BoneJ (Doube *et al.*, 2010). Pore size classification was performed using the method proposed by Jongerius (1957) (Table 3.2) (Perret *et al.*, 1999).

Table 3.2. Pore size classification. Classification of soil pores based on their Equivalent Cylindrical Diameter (ECD). Pore size category terminology was determined following the framework proposed by Jongerius (1957).

Terminology	ECD (μm)
Micropore	< 30
Mesopore	30 - 100
Macropore	> 100

Total porosity was calculated as the proportion of total air volume that occupied a segmented area. Pore count was calculated as the number of separate pores ($>30\ \mu\text{m}$ in diameter) that occupy a segmented area. Pore diameter was calculated using Feret's diameter algorithm, which calculates the largest possible diameter of an object within a boundary across two parallel axes (in a given orientation). Pore thickness was calculated as the number of stacks that occupy a given pore. Connectivity density was calculated from the Euler characteristic (a metric of connectivity, where negative values indicate highly connected structures, whereas positive values indicate poorly connected structures), normalized by the total volume of the segmented area. Circularity was calculated for pores only greater than 5 voxels in size.

3.2.7. Soil seal thickness and compartmentalization

Soil seal thickness was determined using a modified version of the approach described in Armenise et al. (2018). The porosity from a region of soil (2.5 – 3.0 mm in depth), where the effects of the soil surface will be negligible, was used as a reference zone of standard porosity. The depth at which the porosity returned to within the median porosity ($\pm$ the median confidence interval) of the reference zone was determined as the 'soil seal'. The depth immediately after the soil seal (with the same thickness as the soil seal) was defined as the 'transitional zone'. The reference zone (the bottom 10 mm of the region of interest) was termed bulk soil.

3.3. Results

3.3.1. Metadata

Figure 3.6. Environmental conditions and soil properties. Average values of measured soil properties at each time point in the soil crust (green) and bulk soil (brown). A) Bulk density measurements averaged across the whole transect. B) Average soil pH. C) Average measured soil moisture content. D) Average chlorophyll α concentration. E) Average biomass carbon. F) Average biomass nitrogen. Error bars (shaded) indicate ± 1 standard deviation. Asterisks indicate a significant compartmental difference based on Tukey's HSD multiple-comparison test (after a two-way ANOVA).

3.3.1.1. Bulk Density

One-way ANOVA revealed that bulk density did not significantly vary across time (Figure 3.6A). Bulk density only displayed minor shifts across time, and from ranged from 1.214 g/cm³ (July) to 1.239 g/cm³ (March).

3.3.1.2. pH

Two-way ANOVA revealed that soil pH significantly varied over time ($p = 0.0405$) but not compartment ($p = 0.8473$) and there was no significant interaction between time and compartment ($p = 0.0625$) (Figure 3.6.B.). Average soil pH ranged from 6.83 in March soil crust samples, to 7.04 in July soil crust samples. Multiple comparison tests (Tukey's HSD) did not reveal any significant compartmental differences in soil pH at any time point.

3.3.1.3. Soil moisture

Two-way ANOVA revealed that soil moisture significantly varied over time ($p = 0.0005$) and compartment ($p = 0.0073$) but there was not a significant interaction between time and compartment ($p = 0.1914$) (Figure 3.6.C). Soil moisture had a peak value of 16.01% in March bulk soil samples, and had the lowest value of 13.10% in July soil crust samples. Multiple comparison tests (Tukey's HSD) did not reveal any significant compartmental differences in soil moisture at any time point.

3.3.1.4. Crust morphology and chlorophyll a content

Clear shifts in the morphology of soil crusts were seen from November to July (Figure 3.7). In November, the soil surface was highly perturbed and did not display any evidence of soil crust formation, and mainly consisted of large soil aggregates. In March a thin green biotic film was present in certain patches on the soil surface, which is indicative of progression to intermediate successional crusts. In July, the soil surface was covered by a green biotic layer, and visual analysis of this identified them to include mosses, which is a signature of developed crusts (Li et al., 2002).

Figure 3.7. Soil crust morphology at each time point. Photographs were taken of a representative soil core from each time point. Left to right: November 2017, March 2018 and July 2018.

Two-way ANOVA revealed that chlorophyll *a* concentration was significantly affected by time ($p < 0.0001$) and compartment ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$) (Figure 3.6.D). Chlorophyll *a* concentration in the soil crust ranged from 2.60 mg/m³ in November, to 19.92 mg/m³ in July. In the bulk soil, chlorophyll *a* concentration ranged from 2.46 mg/m³ in July to 2.88 mg/m³ in March. Multiple comparison tests (Tukey's HSD) revealed significant compartmental differences in chlorophyll *a* concentration in March ($p < 0.0001$) and July ($p < 0.0001$).

3.3.1.5. Microbial biomass

Two-way ANOVA revealed that biomass carbon significantly changed across time ($p < 0.0001$) and between compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$) (Figure 3.6.E). Biomass carbon ranged from 102.02 µg/g in November, to 280.74 µg/g in July in the soil crust, but only 118.64 µg/g in March to 131.99 µg/g in July in the bulk soil. Multiple comparison tests revealed significant compartmental differences in biomass carbon in March ($p = 0.0016$) and July ($p < 0.0001$).

Two-way ANOVA revealed that biomass nitrogen was significantly affected by time ($p < 0.0001$) and compartment ($p < 0.0001$) and there was a significant interaction between time and compartment ($p < 0.0001$) (Figure 3.6.F). Biomass nitrogen ranged from 10.93 µg/g in November to 76.15 µg/g in July in the soil crust, and from 17.69 µg/g in November to 22.73 µg/g in March in the bulk soil. Multiple comparison tests revealed significant compartmental differences in biomass nitrogen in March ($p < 0.0001$) and July ($p < 0.0001$).

3.3.2. Soil crust and bulk soil microbial communities

3.3.2.1. Rarefaction

Before any microbial community analysis, raw sequences were filtered and rarefied for each amplicon. Bacterial samples were assigned a total number of 290,832 reads, which resulted in an average 9,640 sequences per sample. Bacterial samples were rarefied to 6,300 sequences, resulting in 3706 OTUs. Fungal samples were assigned a total number of 714,287 reads, which resulted in an average 23,809 sequences per sample. Fungal samples were rarefied to 16,000 sequences, which had a total of 1166 OTUs. Protists were assigned a total of 315,860 sequences, which resulted in an average of 10,528 reads per sample. Protist samples had a large amount of filtering applied (due to a high number of metazoan and fungi in the samples), and were rarefied to 1,600 sequences per sample, with a total of 927 OTUs across all samples. Phototroph samples were assigned a total of 464,416 sequences, which resulted in an average of 15,840 sequences per sample. Like protists, a high amount of processing was performed (to remove non-phototrophic organisms), and the samples were rarefied to 2,900 sequences per sample. From this, there was a total of 288 OTUs.

3.3.2.2. Alpha diversity

Alpha diversity of bacterial, fungal, protist and phototroph communities were examined in soil crust and bulk soil samples across combined time points using Fisher's diversity index (Figure 3.8.). Wilcoxon signed-rank sum tests revealed that alpha diversity (calculated using Fisher's diversity index) was significantly different between compartments in bacterial, protist and phototroph ($p < 0.0001$ for all cases), but not fungal ($p = 0.8842$), communities. Alpha diversity was 1.18 (bacteria), 1.41 (protist) and 1.42 (phototrophs) times higher on average in the bulk soil in comparison to the soil crust. One-way ANOVA revealed that alpha diversity significantly reduced across time in soil crust bacterial, fungal, protist and phototroph communities ($p < 0.05$ for all cases). No significant reductions in alpha diversity were seen across time in bulk soil communities.

Figure 3.8. Alpha diversity of soil crust and bulk soil microbial communities. Alpha diversity was calculated using Fisher's diversity index on combined soil crust (green) and bulk soil (brown) samples across all time points for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities. Asterisks indicate a significant result from a Wilcoxon signed-rank test ($p < 0.05$ = *, < 0.01 = **, $p < 0.001$ = ***).

3.3.2.3. Community composition

Microbial community composition was visualized on non-metric multidimensional scaling (NMDS) plots generated using Bray-Curtis similarity (Figure 3.9). For all amplicons, stress values were below 0.2, suggesting a strong representation of reduced dimensionality. NMDS plots revealed compartmental separation in all amplicons. Bulk soil samples displayed a relatively low level of separation between time points and clustered together. In soil crust samples, a higher degree of separation occurred between time points, indicating community composition was more variable.

Analysis of similarity (ANOSIM) calculated from the relative abundance of OTUs (for soil crust and bulk soil samples combined across all time points) revealed that bacterial ($R = 0.400$), fungal ($R = 0.416$), protist ($R = 0.586$) and phototroph ($R = 0.449$) communities were all significantly different between the soil crust and bulk soil. All ANOSIM comparisons were robust ($p < 0.001$ for all comparisons).

Figure 3.9. NMDS plots of soil crust and bulk soil microbial communities combined across all time points. Non-metric multi-dimensional scaling plots for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities were generated from Bray-Curtis similarity comparisons between samples.

The relative abundance of microbial taxa (bacteria – phylum; fungi, protist and phototrophs - class) contributing to >1% relative abundance (averaged across all time points) was determined for each amplicon (Figure 3.10). In bacterial communities, Wilcoxon ranked-sum tests revealed that Cyanobacteria were the only significantly enriched bacterial phyla in the soil crust, and were on average 47.2 times higher in the soil crust ($p = 0.0017$). However, Chloroflexi, Acidobacteria, Verrucomicrobia and Nitrospirae were all significantly enriched in the bulk soil ($p < 0.05$ for all cases). In fungal communities, there were significant enrichments of Chytridiomycetes and Rozellomycota in the soil crust, which were on average 7.48 and 3.06 times higher (respectively) in the soil crust in comparison to bulk soil. In the bulk soil, there was significant enrichments of Soderiomycetes, Eurotiomycetes, Tremellomycetes and Mortierellomycotina. In protist communities, Chloroplastida was significantly enriched in the soil crust and was on average 1.46 times more abundant in the soil

crust than in the bulk soil. In the bulk soil, there were significant enrichments of Discosea, Schizoplasmodiida and Tubulinea. In phototroph communities, Oscillatoriales was significantly enriched in the soil crust, and was on average 5.99 times more abundant in the soil crust than in the bulk soil. In the bulk soil, there were significant enrichments of Nostocales, Alphaproteobacteria and Chlorophyta.

Figure 3.10. Soil crust and bulk soil microbial community composition combined across all time points. The average relative abundance (combined across all time points) of soil crust and bulk soil taxa contributing to >1% total relative abundance. Bacterial (16S rRNA) communities were examined at the phylum level, protist (18S rRNA) communities at the 'ta3' taxonomic level and fungal (ITS) and phototroph (23S rRNA) communities were assessed at the class level (higher taxonomic levels displayed).

3.3.2.4. Temporal development of soil crust communities

NMDS plots, generated using Bray-Curtis similarity, revealed a higher degree of temporal separation in the soil crust relative to the bulk soil for all amplicons (Figure 3.11). Furthermore, to assess the extent that community structure had changed from the initial communities, pairwise Bray-Curtis comparisons of each sample were made with the initial November time point within each amplicon and compartment (Figure 3.12). Two-way ANOVA revealed that time ($p < 0.0001$, all amplicons) and compartment ($p < 0.0001$, all amplicons) had significant effects on Bray-Curtis

similarity, and there was a significant interaction of time and compartment ($p < 0.0001$ for all comparisons) (Table 3.3).

Figure 3.11. Temporal NMDS plots of soil crust and bulk soil microbial communities. Non-metric multi-dimensional scaling plots for bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities were generated based on Bray-Curtis similarity comparisons between samples.

Multiple comparison testing (Tukey's HSD) revealed that there were significant differences in Bray-Curtis similarity (comparisons with November) in bacterial, fungal, protist and phototroph communities between each time point in the soil crust ($p < 0.0001$ for all cases) (Figure 3.12). In the bulk soil, multiple comparison testing only revealed significant differences between November and July Bray-Curtis comparisons (comparisons with November) for all amplicons ($p < 0.05$ for all cases).

Table 3.3. Statistics of community composition similarity in soil crust and bulk soil communities across time. Bray-Curtis comparisons were made with the initial time November time point for soil crust and bulk soil bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) communities. Statistics were generated using a two-way ANOVA.

	16S	ITS	18S	23S
Time				
% Variation	49.07%	57.12%	68.46%	40.94%
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Compartment				
% Variation	11.11%	7.19%	3.39%	20.59%
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001
Interaction				
% Variation	19.59%	6.34%	6.11%	7.47%
<i>p</i> -value	< 0.0001	< 0.0001	< 0.0001	< 0.0001

Figure 3.12. Community composition similarity in soil crust and bulk soil communities across time. Bacterial (16S rRNA), fungal (ITS), protist (18S rRNA) and phototroph (23S rRNA) community similarity was calculated by comparing samples the first time point (within amplicon and compartment) using Bray-Curtis similarity. Time indicates days after the first sampling time point.

Significant taxonomic shifts were seen in bacterial, fungal, protist and phototroph soil crust communities across time (Figure 3.13). In bacterial communities, one-way ANOVA revealed that the relative abundance of Cyanobacteria (0.24% - 10.9%, $p <$

0.0001) and Actinobacteria (29.06% - 40.13%, $p = 0.0031$) significantly increased from November to July in the soil crust. However, the relative abundance of Acidobacteria (9.85% - 2.46%, $p < 0.0001$) and Proteobacteria (25.04% - 18.97%, $p = 0.0410$) significantly decreased between November to July in soil crust communities. No significant temporal differences in relative abundance were seen in bacterial bulk soil communities. ANOSIM revealed that significant differences in bacterial community composition between compartments occurred in March ($R = 0.3014$) and July ($R = 0.7610$). Both ANOSIM values were robust ($p < 0.001$). Wilcoxon signed-rank sum tests revealed significant enrichments of Actinobacteria (soil crust: 40.12%, bulk soil: 33.17%, $p = 0.0005$) in the soil crust relative to the bulk soil in July; whereas significant enrichments of Cyanobacteria were seen in the soil crust relative to the bulk soil in March (soil crust: 1.61%, bulk soil: 0%, $p = 0.0048$) and July (soil crust: 10.89%, bulk soil: 0%, $p < 0.0001$).

In soil crust fungal communities, one-way ANOVA revealed that the relative abundance of Chytridiomycota (3.30% - 15.01% $p < 0.0001$) significantly increased from November to July, whereas Rozellomycota significantly (4.35% - 1.06%, $p = 0.0058$) decreased in abundance during this time frame. In bulk soil communities, the relative abundance of Basidiomycota (6.69% - 3.88%, $p = 0.0319$) significantly decreased from November to July. ANOSIM revealed that significant differences in fungal community composition between compartments occurred in March ($R = 0.2101$) and July ($R = 0.3048$). Both ANOSIM comparisons were robust ($p < 0.001$). Wilcoxon signed-rank sum tests revealed that there were significant enrichments in the relative abundance of Chytridiomycota in the soil crust relative to the bulk soil in March (soil crust: 9.28%, bulk soil: 2.88%, $p < 0.0001$) and July (soil crust: 15.02%, bulk soil: 3.34%, $p < 0.0001$); Basidiomycota in March (soil crust: 6.55%, bulk soil: 3.93%, $p = 0.0139$) and July (soil crust: 6.91%, bulk soil: 3.88%, $p = 0.0141$) and Rozellomycota in March (soil crust: 3.91%, bulk soil: 0.53%, $p = 0.0031$) and July (soil crust: 1.06%, bulk soil: 0.18%, $p = 0.0081$). Significant enrichments of Mortierellomycota were seen in the bulk soil relative to the soil crust in July (soil crust: 1.06%, bulk soil: 6.91%, $p < 0.0001$).

In protist communities, one-way ANOVA revealed that in both soil compartments, the relative abundance of Archaeplastida (soil crust: 30.93% - 56.96%, $p < 0.0001$, bulk soil: 21.98% - 36.64%, $p < 0.0001$) significantly increased between November to July, whereas SAR significant decreased between November to July (soil crust: 53.66% - 30.86%, $p < 0.0001$, bulk soil: 53.35% - 41.09%, $p = 0.0014$). ANOSIM revealed that significant differences in protist community composition between compartments occurred in March ($R = 0.3571$) and July ($R = 0.5918$) Both ANOSIM comparisons were robust ($p < 0.001$). Significant enrichments of Archaeplastida occurred in the soil crust relative to the bulk soil in November (soil crust: 30.93%, bulk soil: 21.97, $p < 0.0001$), March (soil crust: 20.11%, bulk soil: 16.40, $p = 0.0391$) and July (soil crust: 56.96%, bulk soil: 35.63, $p < 0.0001$). Significant enrichments of SAR were seen in the bulk soil relative to the soil crust in July (soil crust: 30.86%, bulk soil: 41.08%, $p < 0.0001$).

In phototroph communities, one-way ANOVA revealed that in both communities the relative abundance of Streptophyta significantly increased between November to July (soil crust: 11.95% - 67.29%, $p < 0.0001$, bulk soil: 13.20% - 35.26%, $p < 0.0001$), whereas the relative abundance of the PX clade (soil crust: 35.13% - 4.48%, $p < 0.0001$, bulk soil: 19.09% - 9.98%, $p < 0.0001$) significant decreased between during this time period. In soil crust communities, the relative abundance of Alphaproteobacteria (11.09% - 1.12%, $p < 0.0001$) and Bacillariophyta (14.97% - 0.58%, $p < 0.0001$) also significantly decreased from November to July. ANOSIM revealed that significant differences in phototroph community composition between compartments occurred in November ($R = 0.7014$), March ($R = 0.7581$) and July ($R = 0.8112$). All ANOSIM comparisons were robust ($p < 0.001$) Wilcoxon signed-rank sum tests revealed significant enrichments of Oscillatoriales were seen in the soil crust relative to bulk soil in November (soil crust: 3.18%, bulk soil: 0.75%, $p = 0.0096$), March (soil crust: 17.20%, bulk soil: 1.71%, $p < 0.0001$), and July (soil crust: 20.26%, bulk soil: 4.31%, $p < 0.0001$); Bacillariophyta in November (soil crust: 14.97%, bulk soil: 7.84%, $p = 0.0096$) and Streptophyta in July (soil crust: 67.29%, bulk soil: 35.26%, $p < 0.0001$). Significant enrichments of Alphaproteobacteria were seen in the bulk soil relative to the soil crust in November (soil crust: 11.09%, bulk soil: 26.95%, $p = 0.0058$), March (soil crust: 9.55%, bulk soil: 27.17%, $p = 0.0015$)

and July (soil crust: 1.12%, bulk soil: 20.93%, $p = 0.0003$); Chlorophyta in March (soil crust: 7.05%, bulk soil: 21.04%, $p < 0.0001$) and July (soil crust: 4.47%, bulk soil: 18.09%, $p < 0.0001$) and PX clade in July (soil crust: 4.48%, bulk soil: 9.97%, $p = 0.0306$).

Overall, analysis of microbial communities revealed significant biological development of soil crust communities, which were taxonomically distinct from bulk soil.

Figure 3.13. Relative abundance of bacterial, fungal, protist and phototroph taxa across time in soil crust and bulk soil communities. Bacterial (16S rRNA), fungal (ITS) and protist (18S rRNA) were shown at the phylum-equivalent level. Phototrophs (23S rRNA) were shown at the class level. Only taxa that contributed to >1% relative abundance in any treatment were shown.

3.3.3. CT scanning data

3.3.3.1 Soil seal thickness

Soil seal thickness was calculated as the depth at which the soil porosity returned to the median reference porosity (bulk soil) $\pm$ the 95% confidence interval, a method proposed by Armenise et al. (2018) (Supplementary figure 3.1). Soil seals had the lowest average thickness (3.39 mm) in November, but had the highest average thickness (8.70 mm) in July (Figure 3.14). Pairwise Wilcoxon signed-rank sum tests revealed significant differences in seal thickness between November and March ($p = 0.0044$), November and July ($p < 0.0001$) but not March and July ($p = 0.1344$).

Figure 3.14. Soil seal thickness. The thickness of soil seals was determined as the depth at which soil porosity returned to median bulk soil porosity ($\pm$ 95% confidence interval), as proposed by Armenise et al (2018). Asterisks indicate a significant test statistic from a Kruskal-Wallis comparison (* = $p < 0.05$, ** = $p < 0.01$, *** = $p < 0.001$).

3.3.3.2. Soil porosity

Two-way ANOVA revealed that soil porosity significantly varied over time ($p < 0.0001$) and depth ($p < 0.0001$), and there was a significant interaction between time and depth ($p < 0.0001$) (Figure 3.15.A). Average soil porosity (combined across all depths) had a peak value in July (24.50%), followed by November (19.99%) and March (14.86%). However, each time point displayed dissimilar trends in soil porosity and depth.

Figure 3.15. Calculated soil porosity throughout the soil profile at each time point. Porosity was calculated from X-ray computed tomography images obtained after running the TopCap macro at intervals of 150 μm . A) Calculated soil porosity with increasing depth (intervals of 150 μm). B) Soil porosity of soil seals (the depth at which porosity returns to the median porosity, $\pm$ 95% confidence interval, of bulk soil), transitional layer (the depth immediately below the soil seal of the same thickness) and bulk soil layers (bottom 10 mm). Mesopores (30-100 μm) and macropores (>100 μm) were defined as according to Jongerius (1957).

Soil porosity was also calculated for the soil seal (the depth at which soil porosity returns to $\pm$ the confidence interval of bulk soil porosity), the transitional (trans) layer (the adjacent layer immediately below the soil seal, of the same thickness as the seal) and for bulk soil (bottom 10 mm of soil core) (Figure 3.15.B). Two-way ANOVA revealed that soil porosity significantly varied over time ($p < 0.0001$) and compartment ($p < 0.0001$) and there was a significant interaction between time and compartment ($p = 0.0086$). Compartmental multiple comparison tests (Tukey's HSD) revealed significant differences in soil porosity between the soil seal and transitional layer (November: $p = 0.0156$, July: $p = 0.0265$) and between the soil seal and bulk soil (July: $p < 0.0001$). Significant temporal differences were seen across every compartmental comparison ($p < 0.05$).

In all compartments and time points, macropores (> 100 μm in diameter) were responsible for the majority of soil porosity (>80.1% in all cases) (Figure 3.15.B). March had the lowest proportion of total porosity by mesopores (30 – 100 μm in diameter - 0.56%, 0.30% and 0.37% in the soil seal, transitional layer and bulk soil, respectively), whereas July samples had the highest proportion of porosity by mesopores (2.91%, 4.19% and 5.41% in the soil seal, transitional layer and bulk soil, respectively). Wilcoxon-signed rank sum tests revealed that mesopore soil porosity

was significantly lower in the soil seal in comparison to the transitional layer ($p < 0.05$, $p < 0.05$) and bulk soil ($p < 0.05$, $p < 0.05$) in November and July.

3.3.3.3. Soil pore number

Two-way ANOVA revealed that the number of soil pores significantly varied over time ($p < 0.0001$) and compartment ($p < 0.0001$) and there was a significant interaction between time and compartment ($p = 0.0039$) (Figure 3.16). March had the lowest average number of soil pores across all compartments (1.44 mm^3), whereas July had the highest average number of pores across all compartments (4.81 mm^3). Multiple comparison tests across soil compartments (Tukey's HSD) revealed that there were significant differences in the number of soil pores between the soil seal and the transitional layer (November: $p = 0.0108$), between the soil seal and bulk soil (March: $p = 0.0276$, July: $p < 0.0001$) and between the transitional layer and bulk soil (July: $p < 0.0001$). Significant temporal differences in soil pore count ($p < 0.05$) within compartments occurred in all comparisons except between the transitional layers in November and March ($p = 0.3160$), and between bulk soil layers in November and July ($p = 0.5713$).

Mesopores were the most abundant pore type across all time points and compartments, and contributed to $> 63.84\%$ of the total pores in all cases. July had the highest proportion of mesopores in each compartment (74.69%, 79.43% and 80.52% in the soil seal, transitional layer and bulk soil, respectively), whereas March had the lowest proportion of mesopores (70.79%, 63.89% and 66.71% in the soil seal, transitional layer and bulk soil, respectively). Wilcoxon signed-rank sum tests revealed that the proportion of mesopores was significantly lower in the soil seal in comparison to bulk soil in November ($p = 0.023$) and July ($p = 0.0086$), but not in March ($p = 0.222$).

Figure 3.16. Number of soil pores in different soil compartments at each sampling time point. Number of soil pores in the soil seals (the depth at which porosity returns to the median porosity, $\pm$ 95% confidence interval, of bulk soil), transitional layer (the depth immediately below the soil seal of the same thickness) and bulk soil (bottom 10 mm) for pores $>30 \mu\text{m}$ in diameter. The number of pores was calculated from X-ray computed tomography images obtained after running the TopCap macro at intervals of $150 \mu\text{m}$. Mesopores ($30\text{--}100 \mu\text{m}$) and macropores ($>100 \mu\text{m}$) were defined according to Jongerius (1957).

3.3.3.4. Soil pore connectivity

Two-way ANOVA revealed that soil pore connectivity significantly varied over time ($p < 0.0001$) and depth ($p < 0.0001$), and there was a significant interaction between time and depth ($p < 0.0001$) (Figure 3.17.A). Average soil pore connectivity across all depths was lowest in March (0.041 mm^{-3}), followed by November (0.065 mm^{-3}) and July (0.104 mm^{-3}).

In relation to soil compartments, two-way ANOVA revealed that that soil pore connectivity significantly varied across time ($p < 0.0001$) and compartments ($p < 0.0001$) and there was a significant interaction between time and compartment ($p = 0.0039$) (Figure 3.17.B). Multiple comparison tests across compartments (Tukey's HSD) revealed that there were significant differences between the soil seal and the transitional layer (November: $p = 0.0108$) and between the soil seal and bulk soil (November: $p = 0.0015$, July: $p < 0.0001$). Multiple comparison tests further revealed significant temporal differences in soil pore connectivity ($p < 0.05$) within compartments for all comparisons, except between the transitional layers in November and March ($p = 0.9461$), and in bulk soil between November and July ($p = 0.5713$).

Figure 3.17. Soil pore connectivity throughout the soil profile at each time point. Soil pore connectivity density (Euler number/volume) was calculated from X-ray computed tomography images obtained after running TopCap at intervals of 250 μm . A: Soil pore connectivity density with increasing depth (intervals of 250 μm). B: The connectivity density of soil pores within the soil seals (the depth at which porosity returns to the median porosity, $\pm$ 95% confidence interval, of bulk soil), transitional layer (the depth immediately below the soil seal of the same thickness) and bulk soil layers (bottom 10 mm).

3.3.3.5. Soil pore surface area

Two-way ANOVA revealed that soil pore surface area was significantly influenced by time ($p < 0.0001$) and compartment ($p < 0.0001$), but there was no significant interaction between time and compartment ($p = 0.1065$) (Figure 3.18). March had the lowest average soil pore surface area across all compartments (2.76 mm²/mm³), whereas July had the highest average soil pore surface area (11.92 mm²/mm³). Multiple comparison tests (Tukey's HSD) revealed that there were only significant compartmental differences in the soil pore surface area between the soil seal and bulk soil in July ($p < 0.0001$). Temporal multiple comparison tests within compartment revealed significant differences in the soil crust between November and July ($p = 0.0337$) and March and July ($p < 0.0001$); between all comparisons in the transitional layer ($p < 0.05$); but between no comparisons in the bulk soil.

Macropores contributed the most towards total soil pore surface area in all time points and compartments (>80.34% of total pore surface area in all cases). March had the highest proportion of total surface area by macropores (93.90%, 95.15% and 95.29% in the soil seal, transitional layer and bulk soil, respectively), whereas July had the lowest proportion of total surface area by macropores (85.00%, 83.26% and 80.34% in the soil seal, transitional layer and bulk soil, respectively). Wilcoxon-signed rank sum tests of the relative abundance of macropores between the soil seal and the

transitional layer/bulk soil revealed no significant differences at any time point ($p > 0.05$ for all cases).

Figure 3.18. Soil pore surface area in different soil compartments at each sampling time point. The soil pore surface area was calculated within the soil seals (the depth at which porosity returns to the median porosity, $\pm$ 95% confidence interval, of bulk soil), transitional layer (the depth immediately below the soil seal of the same thickness) and bulk soil layers (bottom 10 mm). The surface area was calculated from X-ray computed tomography images obtained after running the TopCap macro at intervals of 150 μm .

3.3.3.6. Circularity

Two-way ANOVA revealed that soil pore circularity was significantly influenced by time ($p < 0.0001$) and compartment ($p < 0.0001$), and there was a significant interaction between time and compartment ($p = 0.0227$) (Figure 3.19). Average pore circularity across all compartments increased with time, rising from 0.734 in November to 0.815 in July. Multiple comparison tests (Tukey's HSD) revealed that there were significant compartmental differences in soil pore circularity between the soil seal and the transitional layer, and between the soil seal and bulk soil in March and July ($p < 0.05$). Temporal multiple comparison tests within compartment revealed significant differences in soil seal pore circularity between all comparisons of soil seals ($p < 0.05$), and between November and July ($p = 0.0007$, $p = 0.0410$) and March and July ($p = 0.0429$, $p = 0.0167$) for both transitional layers and bulk soil layers.

Figure 3.19. Soil pore circularity in different soil compartments at each sampling time point. Soil pore circularity was calculated within soil seals (the depth at which porosity returns to the median porosity, $\pm$ 95% confidence interval, of bulk soil), transitional layer (the depth immediately below the soil seal of the same thickness) and bulk soil layers (bottom 10 mm). Soil pore circularity was calculated from X-ray computed tomography images obtained after running the TopCap macro at intervals of 150 μm .

3.4. Discussion

This study represents the first application of X-ray CT scanning to investigate soil structure of biological soil crusts obtained from a real-world temperate agricultural ecosystem. Amplicon DNA sequencing results confirmed trends presented in chapter 2, within which significant differences in bacterial, protist and phototroph community composition and diversity were seen between the soil crust and the bulk soil. Additionally, morphology, biomass and amplicon sequencing data also confirmed that soil crust communities exhibited significant biological development within a single growing season. X-ray CT image analysis further revealed that a region of reduced porosity at the soil surface, termed the soil seal, had developed after 4 months in an agricultural temperate soil, which had increased in thickness after 8 months. The soil seal had a higher number of pores, which were more complex in terms of connectivity, surface area and circularity and had a higher abundance of mesopores, in comparison to lower depths, but this was only seen when substantial biological development of soil crusts had occurred (March and July). Furthermore, a region directly below the soil seal, termed the transitional zone, had higher porosity in comparison to the soil seal, but had a higher pore number and more complex shape descriptors (surface area, circularity and size) in comparison to bulk soil. The structural discrepancies of the transitional layer with other soil layers also became more prominent with increasing soil crust development over time. Together, these results demonstrate that the soil surface is a distinct biological and structural compartment in comparison to bulk soil, which undergoes significant development (abiotic and biotic) in a temperate agricultural environment. Furthermore, work here suggests that soil crusts have larger structural impacts on the soil profile than previously anticipated, and influence the soil structure beyond the 2-5 mm in which they occupy (Belnap, 2003b).

3.4.1. Biology of the soil surface

Visual analysis of morphology showed that the soil surface had no significant biological development at the initial time point (November), and instead the soil surface consisted of large, loose debris and soil particles shortly after tillage. However, visual assessment of soil cores showed that patches of green biotic layers, which contained mosses, had established after just 4 months (March) of the initial time point (November). After 8 months (July) the soil surface was almost exclusively covered by

a green biotic layer, which contained an abundance of mosses, suggesting that substantial development of biological soil crusts had occurred in this time frame. Amplicon DNA sequencing results revealed that alpha diversity was significantly lower in bacterial, protist and phototrophic soil crust communities in comparison to bulk soil communities. Temporal analysis of microbial communities revealed that soil crust community composition significantly changed over time, where significant enrichments of Cyanobacteria, Actinobacteria, Chytridiomycota, Archaeplastida and Streptophyta were all seen as soil crusts developed. These results all confirm trends seen in chapter 2 (see section 2.4), where substantial development of soil crust morphology, biomass and microbial communities were seen within a single growing season, with large enrichments of filamentous taxa or early colonisers that are involved in soil surface stabilisation.

3.4.2. Soil physics at the soil surface

Using non-invasive X-ray CT scanning, results here revealed that significant structural differences in pore abundance, surface area, shape, connectivity and porosity occur at the soil surface in comparison to lower layers in natural temperate soils containing biological soil crusts. Using a methodology proposed by Armenise et al. (2018), a region of reduced porosity was found at the soil surface, termed the soil seal, which increased in thickness as soil crusts became more developed. Furthermore, results here identified a separate zone in the soil profile immediately below the soil seal that had altered porosity characteristics in comparison to both the soil surface and bulk soil, which was termed the “transitional zone”. Several studies have found a reduction in porosity at the soil surface, for example Armenise et al. (2018) reported the production of soil seals (of reduced porosity) in three soil types after exposure to artificial rainfall events in repacked soil cores – which reached a peak thickness of up to ~5.5 mm after just 14 minutes of rainfall. Additionally, Raanan et al. (2016) witnessed that in developed soil crusts from arid regions in Egypt and Israel, two distinct soil compartments formed – a ‘topcrust’ (a region of reduced porosity, roughly ~2 mm thick) and the ‘subcrust’ (a vesicle-rich porous layer directly below the topcrust, roughly 1-2 mm in thickness). These findings complement results shown here, whereby the topcrust is equivalent to the soil seal, and the subcrust is equivalent to the more porous transitional layer. However, results here revealed that soil seals in more

developed soil crusts (July) were over 5-fold higher than in developed crusts reported in Raanan et al. (2016), which reached a peak average thickness of 8.7 mm (or 17.4 mm if the effects of the transitional layer are also considered). Together, these results suggest that the presence of soil crusts may have deeper structural impacts on the soil profile, beyond the upper ~2-5 mm reported (that used morphological analysis) in prior studies, and are additionally thicker than soil crusts reported in arid regions (Belnap, 2003b).

3.4.2.1. Soil porosity and pore number

Significant differences were reported in a number of soil pore metrics between compartments at each sampling time point. In November, there were significant differences in soil porosity and the number of pores between the soil seal and transitional layer. Furthermore, there was a significantly lower proportion of macropores in the soil surface in comparison to the bulk soil. Two weeks prior to the November sampling time point, the sample site was subject to 6-inch conservational tillage, which has been shown in an array of studies to increase the breakdown of macroaggregates, increase porosity, reduce bulk density and destroy complex soil structures (such as fungal hyphal networks) over short timescales (Six et al., 1998, Ambert-Sanchez et al., 2016). Therefore, the soil structure of the upper ~6 inches of the soil profile was essentially reset as a result of tillage, and results in November indicate the extent of tillage on soil structure. Significant differences in soil porosity and pore number between the soil seal and transitional layer in November suggests that over short recovery periods structural differences in the soil surface were beginning to establish. Biological data presented here (and in chapter 2) showed that microbial community composition (bacterial, fungal and protist) and diversity (all amplicons) were not significantly different between the soil surface and bulk soil at this time point. Therefore, results here suggest that physical factors had larger impacts than microbial community composition on soil porosity and pore count in November.

No significant differences were seen in soil porosity between compartments in March. However, significant reductions in porosity, coupled with increased proportions of macropore porosity (which are associated with non-microbial processes), across time were found in all compartments between November and March samples (Beven and

Germann, 2013). Whilst samples taken in November act as the “post tillage” samples, samples obtained in March have experienced the effect of “wintering”, which included exposure to heavy amounts of snow and sub-zero temperatures. Both the effects of snow and freeze-thawing cycles can have large short-term impacts on soil structure, which include altered porosity (due to rearrangement and breakdown of soil particles due to freezing), increase in macropores, increased short-term soil compaction (due to pressure of snow), and water-logging of soil pores after snow melting (LaMalfa and Ryle, 2008, Decharme et al., 2016). Biological data showed that by March, moderate development of soil crust biological communities had occurred, within which there were significant changes in community similarity (relative to initial communities) in all amplicons. Additionally, the soil surface contained significant enrichments of Cyanobacteria (Oscillatoriales), Chytridiomycota and Archaeplastida (dominated by green-algae) relative to the bulk soil, all of which have previously been reported as performing roles in early colonisation or stabilisation of the soil surface (Abed et al., 2019, Pushkareva et al., 2015). Together, shifts in community structure, but increases in macroporosity (which are generated through non-microbial processes) suggest that the major reductions in soil porosity in March are likely due to a combination of both environmental conditions and soil crust maturity.

Significant differences in soil porosity were revealed between the soil seal and both the transitional layer and bulk soil in July. Between the March and July sampling time points, there were no major perturbation events that were experienced at the soil surface, and visual assessments of soil crust morphology, biomass data and amplicon sequencing data all suggested that substantial biological development of soil crusts had occurred. Therefore, in the presence of a considerably developed biological soil crust (July), the soil seal experienced significantly lower soil porosity in comparison to the transitional layer and the bulk soil, which is in agreement with results found in Raanan et al. (2016) and Menon et al. (2010). Amplicon DNA sequence results showed large enrichments of Cyanobacteria (11%), Archaeplastida (56%, dominated by green-algae), Streptophyta (67%, dominated by moss and green algae sequences) and Actinobacteria (40%) in the soil crust relative to bulk soil. These taxa have all been previously shown to increase soil surface stability, either by the excretion of mucilage or exopolysaccharides (green algae and Actinobacteria), or due to their

filamentous nature that binds together soil particles (Cyanobacteria, Streptophyta) – all of which could reduce soil porosity (Mikhailyuk et al., 2014, Hoppert et al., 2004, Büdel et al., 2016, Belnap, 2003b, Rego et al., 2019). Therefore, the abundance of these taxa in developed soil crust communities (which showed similar composition to crusts studied in chapter 2) shown here suggests a substantial biological influence on soil structure in surface seals through reduction of soil surface porosity. Soil porosity is an indicator of the upper water holding capacity of soils, and a reduced porosity in the soil seal suggests that the soil surface may act as a filter that limits the rate of water infiltration (which is limited by soil porosity) into the soil profile (Nimmo, 2004). However, in order to confirm this, the hydraulic conductivity of soil crusts would have to be measured experimentally. Furthermore, structural features described at the soil surface, in which a thick structural seal of reduced porosity and tightly packed soil aggregates is produced, are features that are indicative of a functional role in the protection from soil erosion, which is a widely reported function of soil crusts in arid regions (Belnap, 2006, Belnap, 2003a, Jimenez Aguilar et al., 2009).

3.4.2.2. Soil pore connectivity

Soil pore connectivity has been shown to be a vital factor that influences microbial activity, decomposition rates, movement of nutrients and gases through the soil profile and determining soil organic matter complexity (Smith et al., 2017, Negassa et al., 2015). In November, there were no significant differences in soil pore connectivity between compartments. This is likely due to the impacts of tillage, which would have destroyed previously established soil pore networks and redistributed soil particles in the upper 6 inches, creating new soil pore channels in the process. However, in March and July the soil seal exhibited significantly higher levels of soil pore connectivity and soil pore count in comparison to the bulk soil. Therefore, results here suggest that pore networks in more developed biological soil crusts are highly connected, yet experience an overall level of reduced porosity in comparison to lower soils. Whilst methods here did not measure hydraulic properties of soil crusts, increased soil pore connectivity coupled with a lower soil porosity in the soil seal could suggest both a greater ability for soil crusts to retain soil moisture at the soil surface to maintain hydrated, which has been reported in a number prior studies on soil crust functioning, and additionally limited infiltration to lower soil depths (Belnap and Lange, 2002, Belnap, 2003b,

Colica et al., 2014). Furthermore, these results are likely explained by the abundance of several surface stabilising and filamentous microbial taxa (as mentioned above), such as Cyanobacteria and green-algae, in developed soil crust communities, which have prior been shown to heavily influence soil structure by creating mesh-like networks that bind together soil aggregates (Belnap, 2003b, Belnap and Gardner, 1993, Zhang, 2005). The action of these biota binding together soil aggregates could account for the production of a higher number of smaller pores between large soil aggregates, which increases the probability of pore channel formation and pore connection.

Furthermore, significant increases in soil pore connectivity were seen across time in the soil seal between November and July. A number of studies have shown that tillage can lead to lower soil pore connectivity, and zero-till systems can actually lead to long-term increases in soil macroporosity and connectivity (Galdos et al., 2019). Therefore, the effects of tillage prior to the November sampling time point may explain the low soil pore connectivity in the soil seal during November (2 weeks post-till) when biological impacts were deemed minimal (as no significant biological development of soil crusts had occurred). Furthermore, the reduced amount of disturbance to the soil surface after November may have contributed to increased soil pore connectivity over time (which is also seen in zero-till systems), which was likely exacerbated by substantial enrichments of surface-stabilizing biota that could reshape pore channels at the soil surface.

Moreover, prior findings presented by Carson et al. (2010) have suggested that soil pore connectivity can drastically impact microbial community structure, in which low soil pore connectivity results in an increased bacterial community diversity. Results here also show similar trends, with increases of soil pore connectivity but decreases in alpha diversity were seen across time in soil crust bacterial communities, but additionally we show that this extends to protist and phototroph communities. Low soil pore connectivity has been suggested to prevent exchange of oxygen, water and nutrients in the soil environment and can lead to the production of different environmental niches on the pore scale, which can in turn increase diversity (Wilpiseski et al., 2019). However, there are no studies that have investigated soil

pore connectivity of biological soil crusts, and results here for the first time identify a region of low porosity, but high connectivity which forms under developed biological soil crusts.

3.4.2.3. Soil pore shape descriptors

Soil pore shape descriptors, which included soil pore surface area, circularity and size categorisation, all exhibited significant differences between the soil seal and lower layers under biologically developed soil crusts (July). Soil pore surface area was significantly higher in the soil seal relative to the bulk soil in July, but not in November or March. Soil pore surface area is an important metric in soil physics as it indicates the size of the interface between the soil and the pore environment, within which higher values are indicative of increased sorption of water and nutrients into the soil ecosystem (Hajnos et al., 2000). Results presented here suggest that in biologically developed soil crusts, the soil pore/soil particulate interface is more complex relative to lower layers, highlighting a greater capacity for exchange and acquisition of water and nutrients, which are previously reported functions of biocrusts found in arid regions (Rossi et al., 2012, Wei et al., 2015, Belnap, 2003b). Large temporal variation was seen in soil pore surface area within compartments, but no clear trends were deciphered with time and instead soil pore surface area strongly correlated with the number of pores ($R^2 = 0.9031$, $p < 0.001$).

Soil pore circularity (where higher values indicate rounder pores) was significantly higher in the soil seal relative to lower layers in March and July, and significant increases in soil pore circularity were seen within each compartment across time. Higher values of soil pore circularity are indicative of biological processes, as they are predominantly composed of gaseous vesicles that are released due to microbial metabolism (Yonovitz and Drohan, 2009). This therefore suggests that in a temperate agricultural setting, soils become more microbially active across time, and this is exacerbated closer to the surface. Furthermore, results here also revealed that under developed soil crusts (July), there was a significantly lower proportion of macropores (relative to mesopores) in the soil seal in comparison to lower layers. Macropores ($>100 \mu\text{m}$) are generated as a result of large disturbances to soils such as desiccation, soil cracking, freeze-thawing cycles, plant root growth, animal burrows (mainly

worms) and fungal mycelia growth (Beven and Germann, 2013). However, micro (<30 μm) and mesopores (30-100 μm) are considered the preferred habitats for microbial activity, and it has been shown that the majority of microbial biomass in the soil profile originates from pores of these sizes (Amato and Ladd, 1992, Hassink et al., 1993, Kirchmann and Gerzabek, 1999). Together, increased soil pore circularity and enrichment of mesopores in the soil seal all suggest a substantial influence of biological processes in shaping pore structure in soil crusts from July; and amplicon sequencing data suggests that these biological processes are likely attributed to significant enrichments of filamentous taxa at the soil surface.

3.3.4.3. Conclusions

This is the first study to couple the use of non-invasive X-ray CT scanning and next-generation sequencing to track development of biological soil crusts in a temperate agricultural ecosystem. The first aim of this chapter was to determine the depth at which the soil crust influences soil structure in a temperate agricultural soil. Results here showed that a distinct structural seal forms at the soil surface (termed the ‘soil seal’) that exhibits a substantially lower porosity relative to the bulk soil. The structural influence of biological soil crusts was shown to extend beyond the soil seal, and a zone of altered porosity relative to both the soil seal and bulk soils was identified directly beneath the soil seal, which we termed the ‘transitional zone’. Both the structural seal and the transitional zone reached a peak combined thickness of 17.4 mm, suggesting that biological soil crusts have an extended structural significance on the soil profile than the 2-5 mm depth in which they span.

The second aim was to determine changes in soil pore abundance, distribution, connectivity and structure throughout the soil profile in a temperate agricultural soil. Results revealed that soil surface structural seals exhibited substantially lower soil porosity, but higher pore connectivity relative to lower soil layers. Furthermore, the pore structures of the soil seal were shown to be more complex relative to lower layers, within which soil seal pores were more abundant, had a higher surface area, higher circularity and lower proportion of macropores, but only under developed biological soil crusts. Additionally, the transitional zone had significantly lower porosity, higher circularity and a higher number of pores than bulk soils, but only when substantial

biological development of soil crusts had occurred (July). These results highlight the structural significance of the soil seal in temperate agricultural soils.

The third aim of this chapter was to determine changes in soil pore abundance, distribution, connectivity and structure across the growing season in a temperate agricultural soil. Results here showed that whilst there was substantial temporal variation in soil pore metrics across all soil compartments, no clear universal temporal trends were shared across compartments. However, the clear structural differences in pore architecture, porosity and connectivity between the soil seal and lower layers were exacerbated with time, and samples in July displayed the largest structural discrepancies between the soil seal and lower layers. This data may suggest that structural features of the soil surface are exacerbated with significant biological development of soil crusts.

The last aim was to confirm trends in biological development of soil crusts obtained from results presented in chapter 2. Sequencing data presented here revealed that composition and diversity of bacterial, protist and phototrophic communities was significantly different at the soil surface relative to bulk soils. Furthermore, significant shifts in microbial community composition (all amplicons), coupled with reductions in alpha diversity (bacteria, protist, phototrophs) were seen in the soil crust across time. Soil surface communities also showed significant enrichments of Cyanobacteria, Actinobacteria, Chytridiomycota, Archaeplastida and Streptophyta relative to bulk soils. Overall, these results not only confirm trends seen in chapter 2, but also extend them to show significant enrichments of Chytridiomycota at the soil surface. Moreover, enrichments of filamentous taxa, such as Cyanobacteria, green-algae and mosses at the soil surface, when there was substantial structural development of soil, suggests that these taxa likely influenced characteristics in pore structure in the soil seal.

Overall, results here emphasise that the soil surface is a structurally and biologically distinct soil compartment, which is capable of forming in a perturbed temperate agricultural ecosystem, and may have functional roles in soil erosion and water infiltration. However, results here only represent a single case study and further studies

are required in order to assess whether these results are translatable to other agricultural temperate ecosystems.

CHAPTER 4 – THE FUNCTIONAL GENETIC POTENTIAL OF DEVELOPED BIOLOGICAL SOIL CRUSTS

4.1. Introduction

4.1.1. Metagenomics and metatranscriptomics of the soil surface

A small number of studies have applied shotgun metagenomic sequencing to examine functional profiles of biological soil crusts. A study by Steven et al. (2013a) coupled amplicon DNA sequencing with shotgun metagenomics to examine taxonomic and functional profiles of soil crust and shrubland root zone (rhizosphere) samples obtained from the Mojave Desert, containing arid soils. In this study, assembly-free metagenomic sequences were aligned to the KEGG and COG databases, and gene abundances between soil crust and root zone samples were determined at different SEED levels. Their results revealed that soil crusts, which were largely dominated by Cyanobacterial reads (~40%), had significant enrichments of genes with KEGG annotations assigned to ‘potassium metabolism’, ‘photosynthesis’ and ‘cell division and cell cycle’. Further breakdown revealed that the soil crust had significant enrichments of genes that encoded light harvesting complexes, electron transport chain proteins and circadian clock genes relative to the root zone. They also revealed that metagenomic and taxonomic profiles of soil crusts did not change after 10 years of exposure to elevated atmospheric carbon dioxide, suggesting that soil crust microbiomes show a degree of resistance to climate change. Whilst this study emphasises the different metabolic capabilities of soil crusts in comparison to the rhizosphere, this was performed in an arid desert ecosystem which lacks relevance to temperate ecosystems.

A more recent study by Swenson et al. (2018) utilised exometabolomics, shotgun metagenomics and metatranscriptomics to investigate functional responses of wetting on microbial communities sampled from the Moab Desert, Utah. In this experiment, within which metagenomics was mainly used for taxonomic profiling, Cyanobacterial communities dramatically reduced in response to wetting (from 17-28% to 1-3% after 50 hours of wetting), whereas Firmicutes displayed the opposite response (4-5% to 19-39% after 50 hours of wetting). However, metatranscriptomic reads (extracted from active mRNA instead of DNA) assigned to KEGG pathways revealed

‘biosynthesis of amino acids’, ‘phenylalanine, tyrosine and tryptophan biosynthesis’ and ‘valine, leucine and isoleucine biosynthesis’ significantly increased as a direct response to wetting. Whilst this study represents an impressive showcase using an integrative approach to determine microbial functioning of soil crusts, it provides little context to functioning of soil crusts, especially in context to bare or bulk soils, in temperate systems.

A comprehensive study by Steven et al. (2018) used metatranscriptomics to investigate the effects of wetting and disturbance (trampling) on soil crust microbial communities from arid soils near the Moab Desert, Utah. Firstly, they concluded that functional profiles changed relatively little between soil crusts (0-1 cm) and subsurface soils (1-2 cm). They additionally found that rapid enrichments in the abundance of transcripts associated with photosynthesis and carbon fixation occurred in biological soil crusts shortly after wetting. However, they could not single out an individual gene that was responsible for this functional response. Lastly, they additionally highlighted the vulnerability of soil crusts to disturbance events, in finding that chronic disturbance of soil crusts significantly altered their functional profile, leading to significant reductions in expression of transcripts associated with nitrogen fixation.

4.1.2. Chapter 4 aims

Only a small number of studies have used metagenomic approaches to gain both a taxonomic and functional insight of biological soil crusts. Furthermore, studies that have used these approaches (or metatranscriptomic approaches) have been performed exclusively in arid desert ecosystems, and provide little context for temperate (agricultural) environments (see section 1.4) (Steven et al., 2018, Swenson et al., 2018, Steven et al., 2013a). Additionally, there is a lack of comparative metagenomic approaches, in which metagenomic profiles of soil crust and bulk (or bare) soil communities have been compared. Consequently, there is currently no study that presents the functional roles of soil crusts temperate agroecosystems, and their roles in these ecosystems remains speculative. This chapter investigates the current research gaps in understanding the taxonomic and functional metagenomes of developed soil crusts in a real-world temperate agricultural ecosystem, and how they differ to bulk soil, using a genetic approach. The aims of this chapter were:

To analyse developed biological soil crust and bulk soil metagenomes obtained from a temperate agricultural ecosystem

1. *To compare taxonomic profiles of soil crust and bulk soil metagenomes*
2. *To compare functional profiles of biogeochemical cycling genes between soil crust and bulk soil metagenomes*

From these aims, the following hypotheses were made:

1. *Taxonomic composition of soil crust and bulk soil metagenomes are significantly different, and soil crust metagenomes have enrichments of similar taxa identified by DNA amplicon sequencing (chapter 2)*
2. *Soil crust metagenomes are enriched with a number of genes associated with biogeochemical cycles, mainly carbon and nitrogen fixation, relative to bulk soil metagenomes*

4.2. Methods

4.2.1. Sample selection

DNA isolated from soil crust and bulk soil samples collected in August 2017 (sampled in chapter 3) were selected for metagenomic analysis. Samples were chosen using metadata presented in chapter 3, which highlighted both significant compartmental differences between the soil crust and the bulk soil, as well as significant development of soil crusts. A full description of sampling methodologies is described in section 3.2.1.

4.2.2. DNA extraction and sequencing

DNA was extracted from pooled soil crust and bulk soil samples, which was described in section 3.2. After extraction, DNA was loaded onto a 1% agarose gel and run at 100V for 20 minutes and smears of DNA were visualised under UV light using a transilluminator. DNA concentrations were determined using an Invitrogen Qubit® 2.0 broad range fluorometric quantitation protocol and DNA quality (A_{260}/A_{280}) was assessed using a NanoPhotometer® N60. DNA (at concentrations of at least 30 ng/ μ L) was sent to Novogene Ltd. (Hong Kong) for library preparation and 150 base pair paired-end shotgun sequencing on an Illumina HiSeq 4000 with a target depth of 30 gigabases per sample. Sequencing resulted in an average number of 9,028,805,826 bp per sample in forward reads, and an average number of 10,318,355,835 bp per sample in reverse reads (Table 4.1).

4.2.3. Raw sequences

Trimmomatic (version 0.39) was used to remove 3'- and 5'- adaptor sequences, low-quality base pairs or low quality reads and to align forward and reverse reads from raw sequences (Bolger et al., 2014). Removal of adapters (supplied in the TruSeq. 3 adapter file) was performed using a maximum number of two mismatches in the initial seed, a clipping threshold of 30 and a simple clip threshold of 10. Quality filtering of bases was performed using a leading and trailing threshold of 12, a sliding window of 4 bases with a minimum Phred score of 15 and removal of any reads shorter than 36 nucleotides. Cleaned paired reads were quality assessed (per base quality) using FASTQC to ensure suitable trimming had occurred (Andrew, 2010).

Table 4.1. Number of base pairs obtained from DNA sequencing. DNA sequencing was performed on an Illumina HiSeq 4000 (150 bp paired-end sequencing).

Sample	Forward read	Reverse read
Soil crust 1	10,184,323,118	11,555,091,035
Soil crust 2	9,146,933,486	10,302,701,505
Soil crust 3	9,209,381,122	10,208,837,283
Soil crust 4	8,718,734,603	9,836,302,344
Soil crust 5	8,674,453,142	9,709,621,639
Bulk soil 1	8,488,390,014	9,856,814,128
Bulk soil 2	8,797,308,936	10,161,410,268
Bulk soil 3	8,719,933,403	10,035,460,588
Bulk soil 4	9,906,765,627	11,747,132,512
Bulk soil 5	8,441,834,804	9,770,187,049

4.2.4. Co-assembly

Quality-controlled sequences from all samples (5 soil crust and 5 bulk soil) were co-assembled using MEGAHIT (version 2.0.), a de novo de Bruijn graph based metagenomic assembler (Li et al., 2015). Reads were assembled using a k -mer range between 27 to 127, with step intervals of 10. Contigs below 1000 base pairs in length were discarded and removed from any further analysis. Contigs generated from the metagenomic co-assembly were quality assessed using MetaQUAST with default parameters (Mikheenko et al., 2015).

4.2.5. Open Reading Frame prediction and functional annotation

Coding regions within contigs, known as open reading frames (ORFs), were predicted using Prodigal (version 2.6.3.) with the ‘meta’ pre-set function (Hyatt et al., 2010). ORFs were translated into amino acid sequences and functional annotations were obtained by performing a DIAMOND (version 0.9.9.110) search which queried the Kyoto Encyclopaedia of Genes and Genomes (KEGG), Clusters of Orthologous Groups (COG), the protein families (Pfam) and the carbohydrate-active enzymes (CAZy) databases using an expected value (e-value) threshold of $1e-15$, a minimum amino acid length of 50 and a minimum identity of 90% (Galperin et al., 2014, Finn et al., 2014, Buchfink et al., 2014, Kanehisa et al., 2015, Lombard et al., 2013). Taxonomic annotations of assembled contigs were performed using ‘CAT and BAT’, which queried contigs against The National Centre for Biotechnology Information

(NCBI) non-redundant protein sequences database using an e-value threshold of 1e-15 (Bastiaan von Meijenfeldt et al., 2019, Kanehisa et al., 2015).

4.2.6. Mapping and normalisation

Quality controlled reads (obtained after step 4.2.3) from each sample were mapped to annotated contigs and annotated ORFs obtained after co-assembly using Bowtie2 were sorted, indexed and filtered using SAMtools (Langmead and Salzberg, 2012, Li et al., 2009). Mapping information from each sample was collated across all samples to obtain coverage information (read hits per contig/ORF) using the ‘coverage’ function within the BEDTools suite (Quinlan and Hall, 2010). Coverage information was normalised to account for gene length and sequencing depth per sample (equation 4.1) and gene abundance was expressed as reads per kilobase million (RPKM).

Equation 4.1. Reads per kilobase million calculation:

$$RPKM = \frac{\text{Number of reads}}{\left(\frac{\text{gene length}}{1000}\right) \left(\frac{\text{total reads}}{1,000,000}\right)}$$

4.2.7. Bioinformatics and statistical analysis

Non-metric multidimensional scaling plots (NMDS) were generated from Bray-Curtis similarity using sample coverage information using the vegan package on R (version 3.5.1) (Oksanen et al., 2018). Tests for a normal distribution were performed using a Shaprio-Wilk test, and statistical comparisons between treatments were performed using either a student’s t-test or a Wilcoxon signed-rank sum test (depending on normality) for pair-wise comparisons. The DESeq2 R-package was used to determine differential abundance between soil compartments using a Wald’s test, which were plotted using the enhanced volcano package on R (Love et al., 2014). Stacked bar plots were generated using the Phyloseq package on R, and were plotted using the ggplot2 R package (McMurdie and Holmes, 2013).

4.3. Results

4.3.1. Assembly statistics

Metagenomic co-assembly resulted in the generation of 5,823,845 contigs that were over 1000 bp in length (Table 4.2). Contigs consisted of a combined total of 11,278,116,646 base pairs, which equated to 14.57% of the total base pairs input for metagenomic co-assembly. N50 and N75 values (the minimum contig lengths required to cover 50% and 75% of all contigs) were calculated as 1954 and 1333 (respectively), which indicates a strong assembly performance based on similar sized data sets (van der Walt et al., 2017, Vollmers et al., 2017). The vast majority of contigs were between 1000-5000 bp (95.57%). A small number of large contigs (>50,000 bp) were produced (441, 0.007% of total contigs). A total number of 14,973,970 ORFs (minimum length of 50 amino acids) were predicted from the 5,823,845 contigs, which was the equivalent of 2.57 ORFs per contig.

Table 4.2. Statistics of contigs after co-assembly. Statistics were generated using MetaQUAST on contigs produced from a metagenomic co-assembly of combined soil crust and bulk soil reads.

Statistic	Value
Total bp	11,278,116,646
Number of contigs	5,823,845
> 1000 bp	5,823,845
> 5000 bp	213,005
> 10,000 bp	40,847
> 25,000 bp	3462
> 50,000 bp	441
N50	1954
N75	1333

4.3.2. Taxonomy

Microbial community composition was visualised on non-metric multidimensional scaling (NMDS) plots generated using Bray-Curtis similarity from mapped coverage of reads to co-assembled contigs (Figure 4.1). A stress value of 0.101 was calculated, suggesting a strong representation of reduced dimensionality. NMDS plots revealed clear separation between the soil crust and the bulk soil in mapped coverage to contigs.

Figure 4.1. NMDS plots of soil crust and bulk soil microbial communities from metagenomic sequencing. Non-metric multi-dimensional scaling plots of mapped coverage (quality-controlled reads to co-assembled contigs) were generated using Bray-Curtis similarity.

After quality-control, a total number of 5,852,165 contigs (99.51% of all contigs from metagenomic co-assembly) were used for taxonomic annotation (Table 4.3). The vast majority of contigs were assigned to bacterial sequences (5,480,673 – 93.65%), followed by unassigned contigs (274,602 – 4.72%), eukaryotes (35,574 – 0.61%), archaea (30,658 – 0.53%) and viruses (2338 – 0.04%).

Table 4.3. Number of contigs assigned to each kingdom. Taxonomic annotations were assigned based on the NCBI non-redundant protein database using a DIAMOND search with a threshold of 10^{-15} .

Kingdom	Number of contigs	Proportion (%)
Archaea	30658	0.53
Bacteria	5480673	94.11
Eukaryotes	35574	0.61
Unassigned	274602	4.72
Viruses	2338	0.04

Coverage of mapped reads to co-assembled contigs showed similar trends to the number of contigs assigned to each kingdom (Table 4.4). In the soil crust, bacterial contigs had the highest proportion of mapped coverage (93.96%) followed by

unassigned contigs (4.11%), eukaryotes (1.04%), archaea (0.85%) and viruses (0.04%). In the bulk soil, bacterial contigs had the highest proportion of mapped coverage (94.77%), followed by unassigned contigs (3.70%), archaea (1.34%), eukaryotes (0.17%) and viruses (0.02%). Wilcoxon signed-rank sum tests revealed significant differences in mapped coverage between compartments for archaeal ($p = 0.0062$) and eukaryotic ($p = 0.0360$) sequences.

Table 4.4. Mapped coverage of soil crust and bulk soil reads to each kingdom. The relative abundance of coverage information obtained from quality-controlled reads to co-assembled contigs. Contigs were annotated against the NCBI non-redundant protein database using a DIAMOND search with an e -value threshold of 10^{-15} .

Kingdom	Soil Crust (%)	Bulk Soil (%)	p -value
Archaea	0.85	1.34	0.0062
Bacteria	93.96	94.77	0.1250
Eukaryotes	1.04	0.17	0.0360
Unassigned	4.11	3.70	0.1013
Viruses	0.04	0.02	0.0625

The relative abundance of phyla contributing to >1% of total mapped coverage was determined for individual soil crust and bulk soil samples (Figure 4.2). ANOSIM (calculated from the relative abundance of total mapped coverage) revealed significant differences between compartments ($R = 0.8013$). Calculated ANOSIM values were robust ($p < 0.001$). In both compartments Actinobacteria, Proteobacteria and Acidobacteria were the most abundant phyla, contributing to an average relative abundance of 35.84%, 32.84% and 9.32% in the soil crust, and 32.83%, 31.41% and 17.61% in the bulk soil, respectively. Thaumarchaeota were the only abundant archaeal phylum, and contributed 0.97% in the soil crust and 1.61% in the bulk soil. Chlorophyta was the most abundant eukaryotic phyla, and contributed to an average relative abundance of 0.62% in the soil crust, and 0.03% in the bulk soil.

Figure 4.2. Soil crust and bulk soil microbial community composition. The relative abundance was calculated based on the mapped coverage of each sample to co-assembled contigs. Taxonomic annotations were obtained using a DIAMOND search against the NCBI non-redundant protein database using an e -value threshold of 10^{-15} . (BSC = biological soil crust samples, BULK = bulk soil samples). Only taxa that contribute to $>1\%$ of relative abundance in any sample are shown.

Wilcoxon signed-rank sum tests revealed that eight out of eleven phyla (contributing to $>1\%$ relative abundance) had significant differences in relative abundance between compartments. $\log_{10}$ fold change ($L_{10}FC$ – the relative abundance of soil crust taxon divided by relative abundance of bulk soil taxon, transformed to $\log_{10}$) was determined for taxa displaying significant differences in the relative abundance between compartments (Figure 4.3). The soil crust had significant enrichments of Chlorophyta ($L_{10}FC = 1.34$, $p = 0.0002$), Cyanobacteria ($L_{10}FC = 1.01$, $p = 0.0003$), Bacteroidetes ($L_{10}FC = 0.23$, $p = 0.0041$), Actinobacteria ($L_{10}FC = 0.04$, $p = 0.0158$) and Proteobacteria ($L_{10}FC = 0.02$, $p = 0.0491$). However, Chloroflexi ($L_{10}FC = -0.21$, $p = 0.0071$), Thaumarchaeota ($L_{10}FC = -0.22$, $p = 0.0003$) and Acidobacteria ($L_{10}FC = -0.28$, $p = 0.0001$) were significantly enriched in the bulk soil relative to the soil crust.

Figure 4.3. Differential abundance of soil crust and bulk soil taxa. Taxa exhibiting significantly different relative abundance in mapped coverage (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$) between soil crust and bulk soil communities. Relative abundance of mapped coverage was transformed to fold change (relative abundance of a given taxa in soil crust/bulk soil) and subsequently $\log_{10}$ transformed.

Differential abundance analysis of mapped coverage to annotated contigs was performed at the phylum level between compartments (Figure 4.4). This was performed by plotting the $\log_2$ fold change (soil crust/bulk soil, L_2FC) against the $\log_{10}$ p -value (determined by a Wald's test). Differential abundance analysis identified 13 phyla that were significantly enriched in the soil crust ($L_2FC > 2$, $p < 0.05$), and 2 phyla that were significantly enriched in the bulk soil ($L_2FC < -2$, $p < 0.05$). Of these phyla, Cryptomycota, Chlorophyta and Blastocladiomycetes displayed the highest fold-change and were enriched in the soil crust ($L_2FC = 4.72, 4.67$ and 4.25 ; $-L_{10}P = -4.75, -14.48$ and -14.48 , respectively). Streptophyta, Bacillariophyta and Cyanobacteria had the highest significance statistic ($-L_{10}P$) of significantly enriched soil-crust phyla ($L_2FC = 4.21, 4.25$ and 3.64 , $-L_{10}P = 87.14, 59.23$ and 29.29 , respectively). Two phyla were significantly enriched in the bulk soil by differential abundance analysis, which were assigned to Apicomplexa and the candidate phyla WPS2 ($L_2FC = -3.51, -2.16$; $-L_{10}P = -3.75, -3.71$, respectively).

Figure 4.4. Volcano plots of phyla enriched in soil crust and bulk soil metagenomes. Log_2 fold change was calculated from mapped coverage to annotated contigs between soil crust and bulk soil samples. $\text{Log}_{10}(p)$ was generated from a Wald's test. Dashed horizontal line: $p = 0.05$, dashed vertical lines: log_2 fold change = ± 2 threshold. Plots generated using 'enhanced volcano' R package

ANOSIM revealed significant differences in the relative abundance of mapped coverage to archaeal sequences between compartments ($R = 0.908$) (Figure 4.5). The calculated ANOSIM value was robust ($p < 0.001$). The majority of mapped reads to archaeal contigs were assigned to Thaumarchaeota in both compartments, contributing to an average relative abundance of 96.13% in the soil crust and 96.67% in the bulk soil. Wilcoxon signed-rank sum tests revealed that the abundance of Thaumarchaeota was significantly enriched in the bulk soil relative to the soil crust ($p = 0.0079$). The relative abundance of Crenarchaeota was also significantly higher in the bulk soil relative to the soil crust (soil crust: 1.26%, bulk soil: 1.61%, $p = 0.0073$), whereas Euryarchaeota was significantly enriched in the soil crust relative to the bulk soil (soil crust: 2.11%, bulk soil: 1.61%, $p = 0.0151$).

ANOSIM revealed significant differences in the relative abundance of mapped coverage to bacterial sequences between compartments ($R = 0.9115$) (Figure 4.5). The

calculated ANOSIM value was robust ($p < 0.001$). Bacterial contigs with the highest abundance of mapped reads were assigned to Actinobacteria, Proteobacteria and Acidobacteria in both compartments (soil crust: 36.63%, 33.56% and 9.52%, bulk soil: 33.45% 31.99% and 17.94%, respectively). Wilcoxon signed-rank sum tests revealed that Actinobacteria ($p = 0.0158$), Proteobacteria ($p = 0.0079$), Bacteroidetes ($p = 0.0073$) and Cyanobacteria ($p = 0.0001$) were all significantly enriched in the soil crust relative to the bulk soil. Conversely, Acidobacteria ($p = 0.0019$), Chloroflexi ($p = 0.0071$) and Rokubacteria ($p = 0.0079$) were significantly enriched in the bulk soil relative to the soil crust.

ANOSIM revealed significant differences in the relative abundance of mapped coverage to eukaryotic reads between compartments ($R = 0.9684$) (Figure 4.6). The calculated ANOSIM value was robust ($p < 0.001$). Eukaryotic contigs with the highest abundance of mapped reads were assigned Chlorophyta (soil crust: 53.84%, bulk soil: 16.63%), Ascomycota (soil crust: 25.06%, bulk soil: 60.7%) and Streptophyta (soil crust: 14.07%, bulk soil: 5.94%) in both compartments. Wilcoxon signed-rank sum tests revealed significant enrichments of Bacillariophyta ($p < 0.0001$), Chlorophyta ($p < 0.0001$) and Streptophyta ($p < 0.0001$) in the soil crust relative to the bulk soil. However, significant enrichments of Chordata ($p = 0.0103$), Ascomycota ($p < 0.0001$), Basidiomycota ($p = 0.0002$) and Mucoromycota ($p < 0.0001$) were revealed in the bulk soil relative to the soil crust.

ANOSIM revealed significant differences in the relative abundance of mapped coverage to viral reads between compartments ($R = 0.7014$) (Figure 4.6). The calculated ANOSIM value was robust ($p < 0.001$). Viral contigs with the highest abundance of mapped reads were assigned to Siphoviridae (77.22%), Podoviridae (7.57%) and Myoviridae (5.83%) in the soil crust, and Siphoviridae (77.49%), Mimiviridae (7.66%) and Podoviridae (7.57%) in the bulk soil. Wilcoxon signed-rank sum tests revealed Podoviridae ($p = 0.0079$) and Phycodnaviridae ($p = 0.0098$) were significantly enriched in the soil crust relative to the bulk soil. However, Siphoviridae ($p = 0.0017$) was significantly enriched in the bulk soil relative to the soil crust.

Figure 4.5. Relative abundance of archaeal and bacterial phyla in soil crust and bulk soil communities. Only phyla that contributed to >1% relative abundance of mapped coverage (reads to contigs) are shown. Any taxa contributing less than 1% are in the 'low abundance' category. * Indicates candidate phyla.

Figure 4.6. Relative abundance of eukaryotic and viral taxa in soil crust and bulk soil communities. Only phyla (eukaryotes) and families (viruses) that contributed to >1% relative abundance of mapped coverage (reads to contigs) are shown. Any taxa contributing less than 1% are in the 'low abundance' category.

4.3.2.1. Comparison of bacterial amplicon MiSeq 16S rRNA taxonomy and bacterial shotgun metagenomics taxonomy

Comparisons of bacterial community profiles were made between 16S rRNA amplicon DNA sequencing (from section 2.3.5.3) and shotgun metagenomic sequencing (Figure 4.7). ANOSIM of bacterial composition between the two methodologies revealed significant differences in both the soil crust ($R = 0.9011$) and the bulk soil ($R = 0.9119$). Both calculated ANOSIM values were robust ($p < 0.001$). Both sequencing methods identified Actinobacteria, Proteobacteria and Acidobacteria as the most abundant bacterial phyla in both compartments. Wilcoxon signed-rank sum tests revealed significant differences in the relative abundance of Actinobacteria, Acidobacteria, Bacteroidetes, Chloroflexi, Firmicutes, Nitrospirae, Planctomycetes and Verrucomicrobia ($p < 0.05$ for all comparisons) between the two sequencing methods (both compartments). Firmicutes, Nitrospirae and Planctomycetes were underrepresented in metagenomic profiles, contributing to an average relative abundance of 0.18%, 0.64% and 0.57% (respectively) across both compartments. However, in amplicon sequencing Firmicutes, Nitrospirae and Planctomycetes contributed to an average relative abundance of 1.81%, 0.76% and 5.88% (respectively), which were on average 9.83, 1.14 and 10.34-fold more abundant than in metagenomic profiles.

Figure 4.7. Bacterial community composition of soil crust and bulk soil phyla determined from amplicon and shotgun metagenomic sequencing. Only phyla representing >1% relative abundance are shown. Metagenomic relative abundance was determined from the percentage of mapped reads to annotated contigs. Amplicon relative abundance was determined as the relative abundance of annotated read count (after rarefaction to 8700 sequences).

4.3.3. Functional annotations

Functional profiles of soil crust and bulk soil sequences were created by querying predicted ORF protein sequences against protein sequences of annotated genes in the KEGG database. ORFs were additionally annotated against the COG and Pfam databases to confirm correct annotations had occurred. The abundance of genes (reads per kilobase million – RPKM) involved in nitrogen metabolism, photosynthesis, carbon fixation, phosphorus cycling, sulphur metabolism and methane metabolism according to KEGG annotations (pathway KEGG level) was determined in soil crust and bulk soil samples (Figure 4.8). Significant enrichments of genes involved in nitrogen metabolism (soil crust: 5461.1 RPKM, bulk soil: 5326.7, $p = 0.0081$), photosynthesis (soil crust: 542.9 RPKM, bulk soil: 111.9, $p = 0.0001$) and carbon fixation (soil crust: 5084.5 RPKM, bulk soil: 4971.5 RPKM, $p = 0.0003$) were seen in the soil crust relative to the bulk soil. However, there were significant enrichments of genes involved in phosphorus cycling (soil crust: 5369.7 RPKM, bulk soil: 5562.3, $p < 0.0001$) and methane metabolism (soil crust: 32.4 RPKM, bulk soil: 45.2, $p = 0.0041$) in the bulk soil relative to the soil crust. No significant compartmental differences were seen in sulphur metabolism ($p = 0.1041$).

Figure 4.8. Abundance of genes involved in biogeochemical processes in the soil crust and bulk soil. The combined abundance (reads per kilobase million) of genes involved in KEGG pathways for nitrogen metabolism (map00910), oxygenic photosynthesis (map00195), carbon fixation (map00720), sulphur metabolism (map00920) and methane metabolism (map00920). KEGG genes involved in phosphorus cycling were obtained from Liu et al. (2018). Error bars indicate $\pm$ the standard deviation. Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

4.3.3.1 Nitrogen cycling

Abundance of genes involved in nitrogen cycling were determined in soil crust and bulk soil samples (Figure 4.9). The abundance of genes involved in nitrogen fixation were significantly higher in the soil crust ($p = 0.0035$), and were on average 3.55-fold more abundant in the soil crust (266.24 RPKM) relative to the bulk soil (74.46 RPKM).

Figure 4.9. Abundance of nitrogen cycling genes in the soil crust and bulk soil. The combined average abundance (reads per kilobase million) of genes involved in nitrogen fixation (KEGG module M00175 - *nifHDK*), nitrification (M00528 - *amoABC*, *hao*, *nirAB*), dissimilatory nitrate reduction (M00530 - *narIV*, *napAB*, *nirBD*, *nrFAH*), assimilatory nitrate reduction (M00531 - *nasAB*, *narB*, *nirA*, *NIT-6*, *NR*) and denitrification (M00529 - *narGHI*, *napAB*, *nirKS*, *norBC*, *CYP55*, *nosZ*). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Taxonomic annotations of sequences mapped to nitrogen-fixation genes were assigned to the bacterial phyla Cyanobacteria and Proteobacteria. Wilcoxon signed-rank sum tests revealed that the gene abundance of Cyanobacterial nitrogen-fixation genes was significantly higher in the soil crust (146.37 RPKM) relative to the bulk soil (6.35 RPKM) ($p < 0.0001$). Further taxonomic annotations (Figure 4.10) revealed Cyanobacterial nitrogen fixation contigs consisted of sequences assigned Nostocales

(soil crust: 114.75 RPKM and bulk soil: 4.87 RPKM), or no further taxonomic annotations (soil crust: 31.62 RPKM, bulk soil: 1.49 RPKM), whereas Proteobacterial nitrogen fixation contigs were assigned to Rhodospirillales (soil crust: 44.73 RPKM, bulk soil: 17.66 RPKM), Rhizobiales (soil crust: 0 RPKM, bulk soil: 7.46 RPKM) and Myxococcales (soil crust: 7.07 RPKM, bulk soil: 1.10 RPKM). All comparisons at the order level were significantly different between compartments ($p < 0.05$ for all cases).

Figure 4.10. Abundance of taxa assigned to nitrogen fixation genes in the soil crust and bulk soil. The combined average abundance (reads per kilobase million) of genes involved in nitrogen fixation (KEGG module M00175 – *nifHDK*). Taxonomic annotations were performed at the ‘order’ level. Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Wilcoxon signed-rank sum tests revealed that the abundance of genes involved in nitrification were significantly higher in the soil crust relative to the bulk soil ($p < 0.0001$) (Figure 4.9). Bacterial taxonomic annotations of nitrification genes were assigned to Actinobacteria (soil crust: 209.27 RPKM, bulk soil: 148.04 RPKM), Proteobacteria (soil crust: 42.11 RPKM, bulk soil: 46.75 RPKM) and Nitrospirae (soil crust: 6.55 RPKM, bulk soil: 13.39 RPKM). Archaeal taxonomic annotations of nitrification genes were assigned to Thaumarchaeota (soil crust: 4.87 RPKM, bulk soil: 8.56 RPKM), or were unassigned (soil crust: 6.97 RPKM, bulk soil: 10.39 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene

abundance of every nitrification phyla between compartments ($p < 0.05$ for all cases), except Proteobacteria ($p = 0.0784$).

However, Wilcoxon signed-rank sum tests revealed that the abundance of genes involved in denitrification were significantly higher in the bulk soil relative to the soil crust ($p = 0.0078$) (Figure 4.9). Bacterial taxonomic annotations of sequences mapped to denitrification genes were assigned to Proteobacteria (soil crust: 149.44 RPKM, bulk soil: 151.23 RPKM), Actinobacteria (soil crust: 84.79 RPKM, bulk soil: 67.68 RPKM), Acidobacteria (soil crust: 16.67 RPKM, bulk soil: 25.15 RPKM), Bacteroidetes (soil crust: 15.97 RPKM, bulk soil: 16.92 RPKM), Chloroflexi (soil crust: 0 RPKM, bulk soil: 5.71 RPKM) Gemmatimonadates (soil crust: 12.86 RPKM, bulk soil: 7.46 RPKM) and Verrucomicrobia (soil crust: 4.87 RPKM, bulk soil: 5.91 RPKM). Archaeal sequences were only present in the bulk soil, which were assigned to Crenarchaeota (6.24 RPKM) and Thaumarchaeota (5.47 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of every denitrification phyla between compartments ($p < 0.05$ for all cases). No significant differences in gene abundance were seen between the soil crust and bulk soil in assimilatory and dissimilatory nitrate reduction, but ANOSIM of the taxonomic profiles of genes associated with these processes revealed that composition was significantly different between compartments (assimilatory: $R = 0.4011$, dissimilatory: $R = 0.2110$). Both calculated ANOSIM values were robust ($p < 0.001$).

4.3.3.2. Carbon cycling

The abundance of genes involved in carbon dioxide fixation, production of photosynthetic apparatus and carbohydrate-active enzymes (CAZy) were determined in soil crust and bulk soil samples (Figure 4.11). Wilcoxon signed-rank sum tests revealed a significant enrichment of genes involved in the reductive TCA cycle (soil crust: 108.81 RPKM, bulk soil: 145.00 RPKM) and the Wood-Ljungdahl cycle (soil crust: 3.62 RPKM, bulk soil: 7.77 RPKM) in the bulk soil relative to the soil crust. Taxonomic annotations (phyla) of sequences mapped to reductive TCA genes were assigned to Actinobacteria (soil crust: 22.49 RPKM, bulk soil: 16.21 RPKM), Proteobacteria (soil crust: 15.90 RPKM, bulk soil: 21.97 RPKM) Nitrospirae (soil crust: 8.13 RPKM, bulk soil: 14.15 RPKM) and Rokubacteria (soil crust: 1.77 RPKM,

bulk soil: 6.50 RPKM) (Figure 4.12). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the gene abundance of all phyla for detected reductive-TCA genes. No significant compartmental differences were seen in the abundance of genes associated in the 3-HP bicycle.

Figure 4.11. Gene abundance of carbon fixing genes in the soil crust and bulk soil. The abundance of genes involved in carbon fixation, CAZy enzymes and photosynthetic apparatus. A) KEGG modules: Reductive TCA cycle (M00173), 3HP bi-cycle (M00376), Wood-Ljungdahl (M00377). B) CAZy enzymes (see supplementary information for full list). C) KEGG modules: PSII (M00161), PSI (M00136), Cytochrome b6f (M00162) and rubisco (K01601, K01602). Error bars indicate +/- standard deviation. Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Taxonomic annotations of sequences mapped to Wood-Ljungdahl cycle genes (Figure 4.12) were assigned to Gemmatimonadetes (soil crust: 0.47 RPKM, bulk soil: 0.56 RPKM), Acidobacteria (soil crust: 0.29 RPKM, bulk soil: 0.33 RPKM), Actinobacteria (soil crust: 0.266 RPKM, bulk soil: 0.66 RPKM), Proteobacteria (soil crust: 0.24 RPKM, bulk soil: 0.45 RPKM), Chloroflexi (soil crust: 0.23 RPKM, bulk

soil: 0.48 RPKM) and Verrucomicrobia (soil crust: 0.09 RPKM, bulk soil: 0.03 RPKM). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the relative abundance of Gemmatimonadetes ($p = 0.0131$), Acidobacteria ($p = 0.0198$), Proteobacteria ($p = 0.0012$) and Chloroflexi ($p = 0.0081$).

Figure 4.12. Abundance of genes involved in reductive-TCA, Wood-Ljungdahl and 3HP pathways in the soil crust and bulk soil. KEGG modules were used to obtain functional annotations of sequences. Reductive TCA cycle (M00173), 3HP bi-cycle (M00376), Wood-Ljungdahl (M00377). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Wilcoxon signed-rank sum tests revealed a significant enrichment of sequences assigned to CAZy enzymes genes in the soil crust (11620.80 RPKM) relative to the bulk soil (11083.11 RPKM, $p = 0.0036$), which were 1.05-fold more abundant in the soil crust (Figure 4.13). Bacterial taxonomic annotations of sequences assigned to CAZy enzyme genes were assigned to Actinobacteria (soil crust: 3353.73 RPKM, bulk soil: 2919.05 RPKM), Proteobacteria (soil crust: 2536.79 RPKM, bulk soil: 2221.88 RPKM), Acidobacteria (soil crust: 833.64 RPKM, bulk soil: 1491.81 RPKM), Bacteroidetes (soil crust: 823.57 RPKM, bulk soil: 521.91 RPKM), Chloroflexi (soil

crust: 151.39 RPKM, bulk soil: 241.83 RPKM), Cyanobacteria (soil crust: 303.533 RPKM, bulk soil: 0 RPKM), Gemmatimonadetes (soil crust: 267.96 RPKM, bulk soil: 226.07 RPKM) and Verrucomicrobia (soil crust: 248.42 RPKM, bulk soil: 305.15 RPKM). Chlorophyta (eukaryotes) were additionally detected in the soil crust (205.93 RPKM) but not in the bulk soil (0 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of all phyla assigned to CAZy enzyme genes between compartments ($p < 0.05$ for all cases).

Figure 4.13. Gene abundance of CAZy genes in the soil crust and bulk soil. KEGG gene numbers were used to obtain functional annotations of sequences. See the supplementary information for the full list of KEGG annotations. Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Taxonomic profiles as well as the abundance of genes associated with each CAZy class was further broken down in the soil crust and the bulk soil (Figure 4.14). Wilcoxon signed-rank sum tests revealed a significant compartmental difference of CAZy carbohydrate-binding module gene abundance, which were 1.07-fold more abundant in the soil crust (1076.59 RPKM) relative to the bulk soil (1000.44 RPKM, $p = 0.0079$). Taxonomic annotations of CAZy carbohydrate-binding module genes were assigned to Acidobacteria (soil crust: 54.07 RPKM, bulk soil: 93.88 RPKM), Actinobacteria (soil crust: 459.45 RPKM, bulk soil: 360.37 RPKM), Bacteroidetes (soil crust: 32.62 RPKM, bulk soil: 14.89 RPKM), Cyanobacteria (soil crust: 31.30 RPKM, bulk soil: 0 RPKM), Gemmatimonadetes (soil crust: 22.23 RPKM, bulk soil: 12.42 RPKM), Proteobacteria (soil crust: 163.03 RPKM, bulk soil: 152.19 RPKM), Verrucomicrobia (soil crust: 33.01 RPKM, bulk soil: 41.66 RPKM) and Chlorophyta

(soil crust: 12.49 RPKM, bulk soil: 0 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of all phyla assigned to CAZy carbohydrate-binding module genes between compartments ($p < 0.05$ for all cases), except for Proteobacteria ($p = 0.0713$).

Wilcoxon signed-rank sum tests revealed a significant compartmental difference of CAZy glycoside hydrolase gene abundance, which were 1.09-fold more abundant in the soil crust (5910.13 RPKM) relative to the bulk soil (5412.62 RPKM, $p = 0.00319$) (Figure 4.14). Taxonomic annotations of CAZy glycoside hydrolase genes were assigned to Acidobacteria (soil crust: 366.72 RPKM, bulk soil: 633.26 RPKM), Actinobacteria (soil crust: 2047.22 RPKM, bulk soil: 1626.23 RPKM), Bacteroidetes (soil crust: 566.66 RPKM, bulk soil: 377.29 RPKM), Chloroflexi (soil crust: 66.62 RPKM, bulk soil: 99.03 RPKM), Cyanobacteria (soil crust: 135.25 RPKM, bulk soil: 0 RPKM), Gemmatimonadetes (soil crust: 134.51 RPKM, bulk soil: 108.66 RPKM), Proteobacteria (soil crust: 1048.06 RPKM, bulk soil: 877.02 RPKM), Verrucomicrobia (soil crust: 111.91 RPKM, bulk soil: 135.39 RPKM) and Chlorophyta (soil crust: 86.23 RPKM, bulk soil: 0 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of all phyla assigned to CAZy glycoside hydrolase genes between compartments ($p < 0.05$ for all cases).

Wilcoxon signed-rank sum tests revealed a significant compartmental difference of CAZy polysaccharide lyase gene abundance, which were 1.39-fold more abundant in the soil crust (32.26 RPKM) than in the bulk soil (23.23 RPKM, $p = 0.0061$) (Figure 4.14). Taxonomic annotations of CAZy polysaccharide lyase genes were assigned to Acidobacteria, (soil crust: 2.61 RPKM, bulk soil: 4.45 RPKM), Actinobacteria (soil crust: 2.40 RPKM, bulk soil: 2.95 RPKM), Bacteroidetes (soil crust: 12.36 RPKM, bulk soil: 5.99 RPKM), Gemmatimonadetes (soil crust: 0.58 RPKM, bulk soil: 0.63 RPKM), Planctomycetes (soil crust: 0 RPKM, bulk soil: 0.29 RPKM), Proteobacteria (soil crust: 4.21 RPKM, bulk soil: 0.618 RPKM) and Verrucomicrobia (soil crust: 2.77 RPKM, bulk soil: 2.25 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of Bacteroidetes ($p < 0.0001$), Planctomycetes ($p = 0.0013$) and Proteobacteria ($p < 0.0001$) CAZy polysaccharide lyase genes between compartments.

Figure 4.14. Gene abundance of CAZy enzyme categories in the soil crust and bulk soil. KEGG gene numbers were used to obtain functional annotations of sequences. See the supplementary information for the full list of KEGG annotations. CAZy categories were split according to the CAZy database (Lombard et al., 2013). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Wilcoxon signed-rank sum tests revealed a significant enrichment in the abundance of sequences assigned to photosystem I (PSI - soil crust: 133.58 RPKM, bulk soil: 8.26 RPKM), photosystem II (PSII – soil crust: 254.89 RPKM, bulk soil: 17.00 RPKM), cytochrome B6F (soil crust: 154.37 RPKM, bulk soil: 122.9 RPKM) and rubisco (soil crust: 122.39 RPKM, bulk soil: 66.84 RPKM) in the soil crust relative to the bulk soil (Figure 4.15). The abundance of genes associated with these processes were on average 16.18, 14.99, 1.78 and 1.83-fold higher in the soil crust relative to the bulk soil.

Figure 4.15. Gene abundance of photosynthesis-related genes in the soil crust and bulk soil. KEGG module numbers were used to obtain functional annotations of sequences. Photosystem I (M00163), photosystem II (M00163), cytochrome B6F (M00162) and rubisco (K01601, K01602 - which encoded *rbcl* and *rbcS* subunits). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Taxonomic annotations (at the phylum level) of sequences mapped to PSI genes were assigned to Cyanobacteria (soil crust: 32.01 RPKM, bulk soil: 1.78 RPKM), Streptophyta (soil crust: 6.29 RPKM, bulk soil: 0 RPKM), Chlorophyta (soil crust: 2.32 RPKM, bulk soil: 0 RPKM) and Bacillariophyta (soil crust: 1.20 RPKM, bulk soil: 0 RPKM). Wilcoxon signed rank-sum tests revealed significant differences in the relative abundances of all phyla assigned to PSI genes between compartments ($p < 0.05$ for all cases). Taxonomic annotations (at the phylum level) of PSII sequences were assigned to Cyanobacteria (soil crust: 64.54 RPKM, bulk soil: 3.43 RPKM), Chlorophyta (soil crust: 19.04 RPKM, bulk soil: 1.88 RPKM), Streptophyta (soil crust: 17.83 RPKM, bulk soil: 0.92 RPKM) and Bacillariophyta (soil crust: 0.57 RPKM, bulk soil: 0.11 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance of all PSII phyla between compartments ($p < 0.05$ for all comparisons).

Taxonomic annotations (at the phylum level) of sequences mapped to cytochrome B6F were assigned to Cyanobacteria (soil crust: 26.64 RPKM, bulk soil: 1.43 RPKM), Streptophyta (soil crust: 14.00 RPKM, bulk soil: 0 RPKM), Acidobacteria (soil crust: 11.08 RPKM, bulk soil: 24.83 RPKM), Actinobacteria (soil crust: 9.51 RPKM, bulk soil: 13.70 RPKM), Chlorophyta (soil crust: 9.45 RPKM, bulk soil: 1.09 RPKM), Chloroflexi (soil crust: 9.17 RPKM, bulk soil: 16.09 RPKM), Proteobacteria (soil crust: 2.58 RPKM, bulk soil: 2.77 RPKM) and Verrucomicrobia (soil crust: 1.78 RPKM, bulk soil: 0 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the gene abundance in all phyla assigned to cytochrome B6F genes between compartments.

Taxonomic annotations (at the phylum level) of sequences mapped to rubisco genes were assigned to Proteobacteria (soil crust: 29.85 RPKM, bulk soil: 26.64 RPKM), Streptophyta (soil crust: 12.79 RPKM, bulk soil: 0.79 RPKM), Cyanobacteria (soil crust: 9.95 RPKM, bulk soil: 0.68 RPKM), Actinobacteria (soil crust: 7.70 RPKM, bulk soil: 7.58 RPKM), Bacteroidetes (soil crust: 0 RPKM, bulk soil: 1.44 RPKM) and Chlorophyta (soil crust: 3.19 RPKM, bulk soil: 0 RPKM) (Figure 4.15). Wilcoxon signed-rank sum tests revealed significant differences ($p < 0.05$ for all cases) in the

gene abundance of all phyla, except Actinobacteria ($p = 0.0810$), assigned to rubisco genes between compartments.

The abundance of genes involved in methane (*pmoA*) and methanol (*mxoF*, *mdh*) oxidation were examined in soil crust and bulk soil communities (Figure 4.16). Wilcoxon signed-rank sum tests revealed an enrichment of genes involved in methane oxidation (soil crust: 31.17 RPKM, bulk soil: 43.88 RPKM, $p = 0.0013$), but not methanol oxidation ($p = 0.1034$) in the bulk soil relative to the soil crust (Figure 4.16). Taxonomic annotations of sequences mapped to methane oxidation were assigned to Proteobacteria (soil crust: 8.64 RPKM, bulk soil: 8.68 RPKM), Thaumarchaeota (soil crust: 5.38 RPKM, bulk soil: 9.12 RPKM), Nitrospirae (soil crust: 1.24 RPKM, bulk soil: 3.36 RPKM), Crenarchaeota (soil crust: 0.36 RPKM, bulk soil: 0.93 RPKM) and Acidobacteria (soil crust: 0.19 RPKM, bulk soil: 0.31 RPKM). Wilcoxon signed-rank sum tests revealed significant differences in the abundance of genes mapped to methane oxidation in every phylum, ($p < 0.05$ for all comparisons), except Proteobacteria ($p = 0.3119$), between compartments.

No significant compartmental differences in gene abundance were detected in methanol oxidation ($p = 0.1013$). Additionally, there were no detected signals for genes encoding any subunit of methyl coenzyme-M reductase (*mcrA*, *mcrB*, *mcrG*), which are functional markers for methanogens, in either soil compartment.

Figure 4.16. Gene abundance of methane and methanol oxidation genes in the soil crust and bulk soil. KEGG gene numbers were used to obtain functional annotations of sequences for methane oxidation (*pmoABC*: K10944-46, *mmoXYZ*: K16157-59) and methanol oxidation (*mxoIF*: K14028-29, *mdh*: K00093). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

4.3.3.3. Phosphorus cycling

Abundance of genes associated with phosphorus cycling (phosphatases, phytases, phosphate starvation response, phosphonate degradation and phosphorus transporters) were determined in soil crust and bulk soil samples (Figure 4.17). Genes associated with phosphorus cycling were chosen based on those provided in Liu et al. (2018). Wilcoxon signed-rank sum tests revealed a significant enrichment of genes involved in phosphonate degradation in the soil crust (soil crust: 243.43 RPKM, bulk soil: 234.01 RPKM, $p = 0.0132$), which were on average 1.04-fold more abundant in the soil crust relative to the bulk soil. However, genes associated with phosphate starvation regulation (soil crust: 744.99 RPKM, bulk soil: 786.85 RPKM, $p = 0.0005$) and phosphorus transporters (soil crust: 3431.24, bulk soil: 3597.96, $p < 0.0001$) were significantly more abundant in the bulk soil relative to the soil crust, and were on average 1.06 and 1.04-fold more abundant in the bulk soil (respectively). No significant compartmental differences were found in gene abundance between compartments for genes encoding either phosphatases or phytases.

Taxonomic annotations (phylum level) of sequences assigned to genes associated with phosphonate degradation were assigned to Proteobacteria (soil crust: 149.06 RPKM, bulk soil: 142.90 RPKM), Bacteroidetes (soil crust: 18.04 RPKM, bulk soil: 7.74 RPKM), Acidobacteria (soil crust: 17.81 RPKM, bulk soil: 29.07 RPKM), Gemmatimonadetes (soil crust: 9.68 RPKM, bulk soil: 8.19 RPKM), Verrucomicrobia (soil crust: 5.75 RPKM, bulk soil: 5.19 RPKM) and Actinobacteria (soil crust: 3.06 RPKM, bulk soil: 3.42 RPKM). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the gene abundance of phyla assigned to phosphonate genes for Acidobacteria ($p < 0.0001$) and Bacteroidetes ($p = 0.0048$).

Taxonomic annotations (phylum level) of sequences assigned to genes associated with the regulation of phosphate starvation response were assigned to Proteobacteria (soil crust: 243.29 RPKM, bulk soil: 212.24 RPKM), Actinobacteria (soil crust: 132.33 RPKM, bulk soil: 133.55 RPKM), Acidobacteria (soil crust 69.32 RPKM, bulk soil: 129.99 RPKM), Bacteroidetes (soil crust: 57.09 RPKM, bulk soil: 33.58 RPKM), Gemmatimonadetes (soil crust: 26.011 RPKM, bulk soil: 23.75 RPKM), Verrucomicrobia (soil crust: 16.93 RPKM, bulk soil: 21.14 RPKM), Chloroflexi (soil

crust: 16.66 RPKM, bulk soil: 28.81 RPKM) and Cyanobacteria (soil crust: 12.93 RPKM, bulk soil: 0 RPKM). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the gene abundance of phyla assigned to genes associated with the phosphate starvation response in Proteobacteria ($p = 0.0291$), Acidobacteria ($p < 0.0001$), Bacteroidetes ($p = 0.0038$), Chloroflexi ($p = 0.0030$) and Cyanobacteria ($p < 0.0001$).

Figure 4.17. Gene abundance phosphorus cycling genes in the soil crust and bulk soil. KEGG gene numbers were used to obtain functional annotations of sequences associated with phosphorus cycling as categorised in Liu et al (2018). For a full list of KEGG numbers see the supplementary information. Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

Taxonomic annotations (phylum level) of sequences assigned to genes involved in phosphorus transport were assigned to Proteobacteria (soil crust: 1055.40 RPKM, bulk soil: 914.47 RPKM), Actinobacteria (soil crust: 900.33 RPKM, bulk soil: 790.01 RPKM), Acidobacteria (soil crust: 324.26 RPKM, bulk soil: 642.24 RPKM), Bacteroidetes (soil crust: 135.04 RPKM, bulk soil: 92.83 RPKM), Cyanobacteria (soil

crust: 95.55 RPKM, bulk soil: 0 RPKM), Gemmatimonadetes (soil crust: 78.09 RPKM, bulk soil: 68.92 RPKM), Verrucomicrobia (soil crust: 58.63 RPKM, bulk soil: 68.44 RPKM) and Chloroflexi (soil crust: 45.76 RPKM, bulk soil: 69.88 RPKM). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the gene abundance of phyla assigned to genes associated with phosphorus transport in Proteobacteria ($p < 0.0001$), Actinobacteria ($p = 0.0011$), Acidobacteria ($p < 0.0001$), Bacteroidetes ($p = 0.0479$) and Cyanobacteria ($p < 0.0001$).

4.3.3.4. Sulphur metabolism

Abundance of genes involved in sulphur metabolism were determined in soil crust and bulk soil samples (Figure 4.18). The abundance of genes involved in thiosulphate oxidation were significantly enriched in the bulk soil ($p < 0.0001$), and were on average 1.25-fold more abundant in the bulk soil. Taxonomic annotations of sequences mapped to thiosulphate oxidation were assigned to Proteobacteria (soil crust: 299.55 RPKM, bulk soil: 338.66 RPKM), Acidobacteria (soil crust: 26.24 RPKM, bulk soil: 57.64 RPKM) and Gemmatimonadetes (soil crust: 5.97 RPKM, bulk soil: 6.99 RPKM). Wilcoxon signed-rank sum tests revealed significant compartmental differences in the gene abundance of phyla assigned to thiosulphate oxidation for Proteobacteria and Acidobacteria ($p < 0.0001$ for both cases).

No significant differences in the abundance of genes assigned to sulphate reduction (assimilatory or dissimilatory) were found between the soil crust and the bulk soil ($p > 0.05$).

Figure 4.18. Abundance of sulphur metabolism genes in the soil crust and bulk soil. The combined average abundance (reads per kilobase million) of genes involved in sulphur metabolism (KEGG module map00920), which included assimilatory sulphate reduction (M00176), thiosulphate oxidation (M00595) and dissimilatory sulphate reduction (M00596). Asterisks indicate a significant compartmental difference (Wilcoxon signed-rank test, $p < 0.05 = *$, $p < 0.01 = **$, $p < 0.001 = ***$).

4.4. Discussion

This study is the first application of shotgun metagenomics to examine both taxonomic and functional profiles of soil crust microbiomes obtained from a temperate environment. Results here revealed significant differences in archaeal, bacterial, eukaryotic and viral taxonomic metagenomic profiles between soil crust and bulk soil communities. Soil crust metagenomes exhibited large enrichments of Chlorophyta, Cyanobacteria and Bacteroidetes whereas the bulk soil had enrichments of Acidobacteria, Thaumarchaeota and Chloroflexi, which both strengthens and add to taxonomic profiles presented in prior chapters (section 2.4). Functional analysis demonstrated that there were significant enrichments of genes involved in nitrogen metabolism, oxygenic photosynthesis and carbon fixation in the soil crust relative to bulk soil. More specifically, significant enrichments of genes associated with nitrogen fixation, nitrification, photosynthetic apparatus (PSI, PSII, cytochrome B6F), rubisco and CAZy enzymes were shown in the soil crust. However, in the bulk soil there were significant enrichments of genes involved in methane metabolism and phosphorus cycling. Further analysis revealed significant enrichments of genes associated with denitrification, the reductive TCA, the Wood-Ljungdahl pathway, methane oxidation, the phosphorus starvation response, phosphorus transport and thiosulphate oxidation in the bulk soil relative to the soil crust. Together, results presented here show for the first time that soil crusts from temperate agricultural ecosystems have a substantial capability in performing high levels of carbon and nitrogen fixation relative to bulk soil, and additionally populations that likely perform these functions are not exclusively composed of Cyanobacteria, which has been suggested in arid regions. As temperate regions occupy an incredibly large proportion (~39%) of global terrestrial environments, temperate soil crusts therefore have a colossal potential to contribute to carbon and nitrogen cycles on a global scale. With this in mind, results here suggest that these ecosystems have the hypothetical possibility to mitigate large amounts of atmospheric carbon dioxide annually (if scaled), and further research into the functional dynamics of terrestrial soil crusts in different sites is warranted to explore whether this can be seen on larger scales.

4.4.1. Taxonomic differences

The application of shotgun metagenomics to investigate taxonomic profiles of environmental samples permits the analysis of archaeal, bacterial, eukaryotic and viral community compositions simultaneously, without introducing major PCR biases. Using this approach, results here demonstrated that bacterial sequences constitute the vast majority of total reads (~94%) in soil microbiomes; archaeal and eukaryote reads contributed to small proportions (~1%), whereas viral reads were sparse in abundance (0.02-0.04%). Bacterial reads have previously been shown to dominate a large array of soil microbiomes using shotgun metagenomics. For example Imchen et al. (2017) showed bacterial sequences made up a large proportion (~90%) of shotgun metagenomic reads from soil and sediment samples obtained from tropical mangroves and rainforests. Significant enrichments of bacterial metagenomic sequences assigned to Cyanobacteria and Bacteroidetes were seen in the soil crust relative to the bulk soil, which corresponds with results seen in earlier chapters using amplicon DNA sequencing (chapters 2 and 3), as well as a number of prior studies investigating soil crust microbiomes in a range of ecosystems (Steven et al., 2013c, Rippin et al., 2018, Abed et al., 2019). Cyanobacteria have been shown to perform key roles in the establishment and development of soil crusts in arid, temperate and polar regions, and results here too display their significance in an agricultural soil surface context (Belnap and Harper, 1995, Mazor et al., 1996, Büdel, 1999, Belnap, 2003b). Whilst Bacteroidetes have been shown in several studies to be abundant members of soil crust communities, most notably in arid or high pH soils, their functional significance in soil crusts remains poorly understood (Maier et al., 2014, Mogul et al., 2017, Kuske et al., 2012, Zhang et al., 2016). However, some Bacteroidetes species are known to produce exopolysaccharides, which promote stabilisation of the soil surface and development of biological soil crusts, therefore their abundance at the soil surface may suggest a contribution to these functions (Gundlapally and Garcia-Pichel, 2006).

Log-fold change analysis revealed that sequences assigned to eukaryotes, which included Chlorophyta, Cryptomycota and Bacillariophyta were largely enriched in the soil crust relative to bulk soil. The importance of eukaryotic taxa in soil crusts has been reported in a number of prior studies, for example in arid regions a number of green algal species (belonging to Chlorophyta) have been shown to perform key roles

in providing carbon and organic matter, as well as stabilising the soil surface for colonisation of mosses and lichens (Zhao et al., 2008). Furthermore, prior studies have found an abundance of diatoms (Bacillariophyta) in biological soil crusts obtained from North America, Europe, Asia, Africa and South America (Büdel et al., 2016). In arid soils, enrichments of Bacillariophyta have proposed roles in enhancing soil crust succession by providing the soil surface with nutrients (Hawkes and Flechtner, 2002, Schulz et al., 2016). Whilst there are no previously reported enrichments of Cryptomycota in soil crusts, Cryptomycota have been shown to be abundant in young developing soils, suggesting a role in early colonisation (Tian et al., 2017). Therefore, log-fold analysis identified enrichments of eukaryotes at the soil surface that are previously associated with soil surface stabilisation, early colonisation and driving soil crust development.

Significant differences were seen between metagenomic taxonomic profiles of soil crust and bulk soil archaeal communities, where an overall higher proportion of reads were assigned to contigs with archaeal annotations in the bulk soil in comparison to the soil crust. Archaeal abundance is often associated with reduced oxygen content or anoxic conditions, which are found at lower depths in the soil profile. For example, Lee et al. (2015) showed an increased abundance of methanogenic archaea, such as *Methanoregula*, *Methanosaeta* and *Methanosphaerula* with increasing depth in bulk and rhizosphere soils obtained from rice paddies. Archaeal sequences were largely dominated by Thaumarchaeota, which were 1.65-fold more abundant in the bulk soil relative to the soil crust. Previous studies have shown that Thaumarchaeota are sensitive to depth in agricultural soils, where increased depth results in higher diversity (Lu et al., 2016). More specifically, a study by Marusenko et al. (2013) showed that *Thaumarchaeota* were abundant ammonia-oxidising archaea in soil crusts obtained from arid desert regions, however their relative abundance in comparison to lower layers of the soil profile was not assessed. Overall, results here suggest the abundance, diversity and functional significance of archaea is lower in the soil crust relative to bulk soil.

Significant differences were seen between metagenomic taxonomic profiles of soil crust and bulk soil viral communities, where soil crust communities had a two-fold

higher proportion of viral reads in comparison to bulk soil communities. A higher viral load is often associated with a greater impact on microbial populations, and soils often exhibit a high abundance of viruses which can reach as much as 10^9 g^{-1} (Wilpiseski et al., 2019, Ashelford et al., 2003). An enrichment of reads assigned to Phycodnaviridae was seen at the soil surface in comparison to the bulk soil. Phycodnaviridae are a family of dsDNA viruses that infect a host of algae (including green-algae), therefore their abundance at the soil crust is likely attributed to an increase abundance of green algae (Wilson et al., 2009). Furthermore, Phycodnaviridae abundance at the soil surface may contribute to control of algal populations and development of soil crusts towards moss and lichen dominated communities; however supplementary work would be required to explore this further.

4.4.2. Functional analyses

4.4.2.1 Nitrogen cycling

Functional analysis of soil crust and bulk soil metagenomes revealed that the soil crust had a significant enrichment in genes associated with nitrogen metabolism, photosynthesis and carbon fixation according to KEGG pathway annotations. Metagenomic analysis of nitrogen cycling pathways demonstrated that the soil crust had a significantly higher abundance of genes involved in nitrogen fixation and nitrification in comparison to the bulk soil. A number of studies have shown that biological soil crusts perform substantial amounts of nitrogen fixation and have been estimated to contribute as much as ~40% of global biological nitrogen fixation performed in terrestrial environments (Elbert et al., 2012). Results presented here show that the genetic potential for nitrogen fixation in a developed temperate agricultural soil crust is ~3.5-fold higher in comparison to the immediate lower soil layer. Previous studies that have examined the genetic potential for nitrogen fixation by soil crusts have been exclusively limited to unperturbed arid and semi-arid ecosystems. For example, previous studies have used qPCR to show that the gene abundance of *nifH* (a gene encoding nitrogenase iron protein, a functional marker for nitrogen fixation) in soil crusts were roughly 100-fold higher relative to bare and bulk soils in the Chihuahuan Desert (arid), the Great Basin Desert (arid) and the Anhui Province (semi-arid, China) (Couradeau et al., 2019, Huang et al., 2011). Data here for the first time suggests that similar enrichment of genes associated with nitrogen

fixation rapidly occurs in a highly perturbed temperate agricultural soil. Taxonomic annotations of nitrogen fixation genes were exclusively assigned to Cyanobacteria and Proteobacteria at the phylum level. Further breakdown of Cyanobacterial nitrogen fixation sequences revealed that the majority of reads were assigned to Nostocales, and remaining reads contained no further taxonomic annotation beyond 'Cyanobacteria'. Nostocales taxa have been previously shown to be abundant Cyanobacterial members of soil crusts in arid and temperate regions, with suggested roles in nitrogen fixation, soil surface stabilisation and additionally acting as photobionts to support lichen colonisation (Li et al., 2014, Belnap, 2006, Belnap, 2003c, Davies et al., 2013a, Yeager et al., 2012). Furthermore, Proteobacterial nitrogen fixation sequences were assigned to Myxococcales, Rhizobiales and Rhodospirillales. Whilst these taxa have been reported to be present in soil crust microbiomes in several prior studies, the functional significance has not been explored (Maier et al., 2018, Angel and Conrad, 2013). Overall, results here suggest that soil crusts in temperate agricultural soils have a high genetic capability to perform nitrogen fixation in comparison to bulk soil. Additionally, taxonomic annotations to nitrogen fixation genes were not only assigned to Cyanobacterial taxa, but they additionally consisted of significant proportions of Proteobacteria, suggesting diverse communities are involved in nitrogen fixation in soil crusts.

Significant enrichments in the abundance of nitrification genes (oxidation of ammonia to nitrate, and oxidation of nitrite to nitrate) were seen in the soil crust relative to the bulk soil, which were on average 1.14-fold more abundant in the soil crust. These results are in agreement with previous findings suggesting that soil crusts perform significant amounts of nitrification (Weber et al., 2015). For example, Brankatschk et al. (2013) found that in developed soil crusts sampled from temperate inland sand dunes (Germany), the abundance of ammonia monooxygenase (a functional gene for nitrification - *amoA*) was ~10-fold higher in comparison to bare soil surfaces. However, results presented here extend this finding to an agricultural temperate environment. Taxonomic annotations of nitrification genes were predominantly composed of Actinobacterial sequences in both soil compartments. Actinobacteria are a dominant phyla in a range of soil microbial communities, and whilst they are often detected in soil crust microbiomes, their functional significance in the soil crust has

mostly been ignored (Mogul et al., 2017, Steven et al., 2013b, Liu et al., 2017b, Abed et al., 2019). However, both compartments displayed low abundances of nitrification genes assigned to Nitrospirae, within which significant enrichments were seen in Nitrospirae abundance in bulk soil relative to the soil crust. Nitrospirae are widely reported to perform central roles in nitrification in a range of soils, and data here suggest that the majority of soil crust and bulk soil nitrification is not performed this phylum (Lücker et al., 2010, Koch et al., 2019). Overall, results here suggest that temperate agricultural soil crusts have an increased genetic capability to perform nitrification in comparison to bulk soil, which is predominantly due to increased contribution of Actinobacteria taxa.

Significant enrichments in gene abundance of denitrification genes (reduction of nitrate to molecular nitrogen) were seen in the bulk soil relative to the soil crust. Previous studies have suggested that nitrogen loss via biological denitrification is a major process that occurs in developed biological soil crusts, but this has only been shown in arid environments where the extent at which this happens varies considerably (Weber et al., 2015, Johnson et al., 2005, Strauss et al., 2012). For example, Abed et al. (2013) showed significant nitrogen loss via biological denitrification by tracking radiolabelled nitrates and using qPCR to detect denitrification genes (*nirS*, *narG*) in arid biological soil crusts from the Oman desert. Results here suggest that the genetic potential for denitrification was higher in bulk soil relative to the soil crust, therefore indicating that temperate agricultural soil crust contributions to denitrification are no higher than bulk soil. Denitrification genes were mostly assigned to Proteobacteria and Actinobacteria, and both of these groups are commonly associated with several steps in soil denitrification (Abed et al., 2013, Palmer and Horn, 2012).

Overall, results here show that metagenomes of soil crusts in a temperate agricultural setting have a significantly higher genetic capability to perform nitrogen fixation and nitrification, but a lower capability to perform denitrification, relative to bulk soils.

4.4.2.2. Carbon cycling

4.4.2.2.a. Carbon fixation

Functional analysis of carbon fixation processes revealed significant enrichments in the abundance of genes encoding PSI, PSII, cytochrome B6F and Rubisco in the soil crust relative to the bulk soil. It is well documented in the literature that soil crusts perform substantial amounts of carbon fixation in a host of climates, and may account for as much as 7% of terrestrial NPP by vegetation (Elbert et al., 2012, Ciais et al., 2013, Huang et al., 2014a, Li et al., 2012, Niederberger et al., 2015). The enrichments of genes encoding PSI and PSII, which along with cytochrome B6F are vital machinery in oxygenic photosynthesis, were ~15-fold higher in the soil crust relative to the bulk soil, representing a substantially higher genetic capability of soil crusts to perform photosynthesis. A study performed by Steven et al. (2013a) also presented significant enrichments of photosynthetic KEGG-annotated genes, including light harvesting complexes and electron transport chain genes, in soil crusts relative to root-zone soils – but this was performed in arid soils from the Mojave Desert. Furthermore, Rubisco was 1.83-fold more abundant in the soil crust relative to the bulk soil. Rubisco is the most abundant enzyme in the biosphere and catalyses the first major step in carbon fixation through the Calvin-Benson cycle by combining carbon dioxide with ribulose 1,5-bisphosphate to produce 3-phosphoglycerate (Badger and Bek, 2008). Similar enrichments of Rubisco genes have been exhibited in metatranscriptomes of arid soil crusts from the Moab Desert (Utah), which were shown to increase rapidly in response to wetting events (Steven et al., 2018). Together, results presented here show for the first time that soil crusts from a temperate agricultural setting have a substantially higher genetic capability to perform carbon fixation through oxygenic photosynthesis in comparison to bulk soils. Taxonomic annotations of carbon fixation genes were dominated by Cyanobacteria and Eukaryotes (with no further annotation) in PSI, PSII and cytochrome B6F genes, but Proteobacteria in Rubisco genes. The same phyla (Cyanobacteria and Proteobacteria) were also revealed as dominant taxa in nitrogen fixation (section 2.4.2.1) associated genes. This suggests that the abundance of both Cyanobacteria and Proteobacteria in the soil crust are central for both nitrogen and carbon fixation in a temperate agricultural ecosystem.

However, significant enrichments of genes associated with carbon fixation through the reductive TCA cycle and the Wood-Ljungdahl pathway were revealed in the bulk soil in comparison to the soil crust. Autotrophic organisms that perform the reductive

TCA cycle, also known as the Reverse Krebs Cycle, utilise carbon dioxide (and water) to generate cellular carbon through a series of biochemical reactions (Berg, 2011). This pathway was originally identified in the anaerobic green sulphur bacterium *Chlorobi* (*Chlorobium limicola*), and is now known to be performed mostly by obligate anaerobes or microaerobic taxa (Evans et al., 1966, Berg, 2011). Therefore, it is likely that the abundance of genes associated with the reductive TCA at a lower soil depth is due to a lower oxygen availability in comparison to the soil surface. This is further supported by taxonomic annotations of reductive-TCA genes, where significant enrichments of sequences in bulk soil were assigned to Proteobacteria and Nitrospirae, which contain a number of anaerobes that utilise this pathway (Berg, 2011). However, taxonomic annotations of sequences assigned to reductive-TCA genes were fairly poor, where just over 50% of the sequences did not have a taxonomic annotation beyond 'Bacteria', and more thorough annotations are required to fully scope the taxonomic profiles of these sequences. Enrichments of sequences assigned to the Wood-Ljungdahl pathway, a non-cyclic pathway that utilises carbon dioxide to produce carbon monoxide and formic acid, were additionally seen in the bulk soil (Berg, 2011). Like the reductive-TCA, taxonomic sequences assigned to the Wood-Ljungdahl pathway were poor, where just under 65% of sequences were not annotated beyond 'Bacteria'. Therefore, it is difficult to conclude which taxa contributed the most to compartmental differences. Furthermore, the abundance of sequences assigned to the Wood-Ljungdahl pathway was incredibly low in both compartments (<10 RPKM), which suggests the contribution of this pathway to overall carbon fixation is negligible. Overall, whilst the soil crust exhibited an abundance of genes associated with carbon fixation through oxygenic photosynthesis, these results reveal that lower layers may also contribute to carbon fixation through alternative pathways.

4.4.2.2.b. Carbohydrate-Active Enzymes (CAZy)

The abundance of combined CAZy enzyme genes were significantly higher in the soil crust relative to the bulk soil. CAZy enzymes consist of a large group of proteins that are directly involved in the biological breakdown, modification or assembly of complex oligosaccharides and polysaccharides (Cantarel et al., 2009). The current CAZY database consists of over 300 protein families, which have been classified into 6 categories: 1) auxiliary activities, 2) glycoside hydrolases, 3) glycosyltransferases,

4) oligosaccharide lyases, 5) carbohydrate esterases and 6) carbohydrate-binding modules (Lombard et al., 2013). Analysis here revealed that the glycoside hydrolases, polysaccharide lyase and carbohydrate-binding modules CAZy enzyme classes were significantly enriched in the soil crust relative to the bulk soil. Glycoside hydrolases are responsible cleaving glycosidic bonds (covalent bonds between carbohydrates) by hydrolysis or transglycosylation to produce small metabolizable products; whereas polysaccharide lyases use the β -elimination mechanism to cleave glycosidic bonds in hexose-1-4- α - or β -uronic acid sequences (Henrissat and Davies, 2000, Linhardt et al., 1987, Sutherland, 1995). In particular, glycoside hydrolases can act on a range of common substrates found in soils, for example cellulose, starch, fructan and xylan from plant matter, or chitin from fungi or microarthropods (Berlemont and Martiny, 2016). It is widely reported that soil crust biota excrete complex mixtures of high-molecular weight exopolysaccharides, which make up a considerable component of the biological matrix that forms in soil crusts (Colica et al., 2014, Rossi et al., 2012, Chamizo et al., 2018, Mazor et al., 1996, Belnap, 2003b, Rossi and De Philippis, 2015). Therefore, it is highly possible that enrichments in gene abundance of both glycoside hydrolases and polysaccharide lyases at the soil surface could be explained by an increased substrate availability and utilisation of exopolysaccharides excreted by soil crust inhabitants. Further, results here revealed that the majority of glycoside hydrolase sequences were dominated by Actinobacteria (34.6% relative abundance), whereas polysaccharide lyase sequences were dominated by Bacteroidetes, suggesting important roles of these phyla in the utilisation of exopolysaccharides in soil crusts. This result aligns with findings presented in Yeager et al. (2017), who suggested large roles in the degradation of hemicellulose, cellulose and chitin in arid crusts by a range of Actinobacteria species, which included *Pseudonocardiaceae*, *Streptomyces* and *Micromonosporaceae*.

Significant enrichments were also found in the gene abundance of carbohydrate-binding modules in the soil crust relative to the bulk soil. Whilst this protein family does not have any direct enzymatic activity, carbohydrate-binding modules facilitate activity of other CAZy classes by promoting folding, targeting and prolonging interaction of enzyme-substrate complexes (Cantarel et al., 2009). Most notably, carbohydrate-binding modules are known to work closely in conjunction with

glycoside hydrolases (Boraston et al., 2004). Sequences assigned to carbohydrate-binding modules were also dominated by Actinobacteria (44.5% relative abundance), which also dominated glycoside hydrolase profiles.

Overall, the increased abundance of polysaccharide lyases, glycoside hydrolases and carbohydrate binding modules at the soil surface suggests a functional role of temperate agricultural soil crusts in the deconstruction of complex carbohydrates and polysaccharides at the soil surface.

4.4.2.2.c. Methane cycling

Prior studies have suggested that soil crusts in arid and glacial ecosystems results in a net-release of methane due to development of increased abundances of methanogenic taxa within soil crusts (Angel et al., 2011, Aschenbach et al., 2013). However, results shown here did not reveal any detection of methanogenic functional genes (methyl coenzyme-M reductase) in either soil compartment. Conversely, detection of methanotrophic (utilisation of methane) taxa involved in both methane (particulate and soluble methane monooxygenases) and methanol (methanol dehydrogenase) oxidation were detected in both compartments, albeit in small abundances (< 50 RPKM). Methane oxidation gene abundance was significantly higher in the bulk soil relative to the soil crust, suggesting that lower layers in the soil profile could have a greater significance for utilising methane, which has been shown to occur optimally at depths of around 10-20 cm in other soil systems (Saari et al., 1998, Roslev et al., 1997). Furthermore, the detection of methanotrophs here is contrary to results found by Angel et al. (2011), who found considerable numbers of methanogens but no detectable signals of methanotrophs in desert soil crusts. Due to the low abundance of the detected methanotroph genes, the effects of methane cycling in a temperate agricultural ecosystem are likely negligible to overall carbon cycling.

4.4.2.3. Phosphorus cycling

Several studies have suggested substantial roles of soil crusts in arid regions in phosphorus cycling, particularly in the release of inaccessible organic-P (Crain et al., 2018, Glaser et al., 2017, Baumann et al., 2019). Significant enrichments in the abundance of genes assigned to phosphonate degradation were seen in the soil crust

relative to the bulk soil, which were on average 1.04-fold more abundant in the soil crust. Phosphonates are a class of carbon-bound organophosphorus compounds that exhibit high chemical stability, and are common structural compounds used by bacteria, archaea and eukaryotes (Kononova and Nesmeyanova, 2002, Yu et al., 2013). Phosphonate degradation usually occurs through the action of C-P lyases, which result in the production of orthophosphate (Pi – a more bioavailable form of P) (Villarreal-Chiu et al., 2012). Results here suggest increased ability to utilise phosphonates at the soil surface relative to the bulk soil. However, phosphonates are minor components of soil P pools, and the main process that releases bioavailable phosphorus in soils is usually through the action of acid and alkaline phosphatases, which hydrolyse anhydrides and esters of phosphoric acids (Margalef et al., 2017, Eivazi and Tabatabai, 1977, Huang et al., 2017). No significant differences were seen in the abundance of genes encoding phosphatases between compartments. Furthermore, significant enrichments of phosphorus transporters (which catalyse Pi uptake into the cytoplasm) and phosphate starvation response genes (which are responsible for regulating P-cycling genes under P-limiting conditions) were seen in the bulk soil relative to the soil surface (Biber et al., 2013, Furtwängler et al., 2010). Both these processes displayed higher compartmental fold-changes and gene abundance in comparison to phosphonate degradation. Therefore, it is likely that developed temperate agricultural soil crusts do not have a larger role in the net release and utilisation of Pi than bulk soils.

4.4.2.4. Sulphur cycling

Significant involvements of soil crusts in sulphur cycling, which is tightly linked with the carbon cycle, have been proposed in prior studies (Rippin et al., 2018, Canfora et al., 2016). For example, results presented by Canfora et al. (2016) revealed that bacteria associated with the sulphur cycle (with potential roles in sulphide, sulphur, sulphite or thiosulphate oxidation) co-occurred with cyanobacterial populations in soil crusts sampled from semi-arid Mediterranean environments in Italy. However, results here exhibited no significant enrichments of genes associated with sulphate reduction or thiosulphate oxidation in the soil crust relative to the bulk soil. Instead, enrichments of genes associated with thiosulphate oxidation were seen in the bulk soil relative to the soil crust. Thiosulphate represents a pool of naturally-occurring sulphur in soils,

and can be utilised by a number of chemotrophic and phototrophic taxa as an energy source (Hensen et al., 2006). But despite a range of phototrophic taxa showing enrichment in the soil crust, enrichments of thiosulphate oxidation sequences were not seen at the soil surface. Therefore, results shown here indicate that lower layers in temperate crust-dominated soil profiles may have larger contributions to sulphur cycling in comparison to the soil surface.

4.4.3. Comparison of soil crust and bulk soil amplicon and metagenomic profiles

Soil crust and bulk soil samples used in this chapter were used for both 16S rRNA DNA amplicon sequencing (chapter 2) and shotgun metagenomic sequencing (chapter 4). Comparison of results revealed that in both compartments community composition was significantly different between sequencing methods. Amplicon DNA sequencing displayed significant enrichments of Firmicutes, Nitrospirae and Planctomycetes relative to shotgun metagenomic sequencing. DNA amplicon sequencing has been heavily criticized as an approach to profile microbial communities, as the method can introduce sequence biases through PCR. Additionally there are irregular copy numbers of the 16S rRNA gene across different taxa and taxonomy is assigned based on only a small 150-300 bp region in the bacterial genome (Krehsenwinkel et al., 2017, Becker et al., 2000). For example, the relative abundance of Firmicutes, which is known to contain several copies of the 16S rRNA gene, has been shown to be overrepresented in amplicon DNA sequencing (Větrovský and Baldrian, 2013). Conversely, metagenomics aims to reduce these biases through avoiding PCR methods in library preparation and assembling short reads to larger contigs that provide a higher taxonomic resolution. Only a small number of studies have made direct comparisons of community profiles generated from 16S rRNA amplicon sequencing and shotgun metagenomics, for example, Laudadio et al. (2018) who compared the two approaches to profile the human gut microbiome, showed that shotgun metagenomics resulted in a substantially deeper characterization of composition through identifying a higher number of species per sample. Results presented here suggested that substantial variation in community composition can occur between the two methods, and amplicon DNA communities are likely inflated with taxa assigned to Firmicutes, Nitrospirae and Planctomycetes. However, until the price of metagenomic sequencing

reduces, amplicon DNA sequencing will be the most cost-effective approach to use next-generation sequencing to profile microbial communities.

4.4.4. Conclusions

Prior studies that have investigated metagenomes of soil crusts have been performed exclusively in arid ecosystems, and results presented in this chapter signify the first study to use metagenomics to assess the functional capability of temperate soil crusts. The first aim of this chapter was to compare metagenomic taxonomic profiles of soil crust and bulk soil communities. Analysis of unbiased taxonomic profiles here revealed that bacterial, eukaryote, archaeal and viral communities all displayed significantly different composition in the soil crust relative to the bulk soil. Furthermore, soil crust communities were enriched with a plethora of taxa relative to bulk soil communities, and differences were dominated by sequences assigned to Cyanobacteria, Chlorophyta, Bacteroidetes, Actinobacteria and Proteobacteria – all of which have previously-reported roles in soil surface stabilisation. This finding both supports and adds to prior work examining taxonomic profiles of soil crusts performed in earlier chapters (chapters 2 and 3). Additionally, for the first time in the current literature, this experiment studied the taxonomic and functional roles of the soil crust viral metagenome. Results presented here showed that soil crust viral communities were significantly enriched with taxa that are capable of infecting green algae, a key member of soil crust communities, which could suggest a significant ecological role of viruses in the soil crust.

The second aim of this chapter was to compare functional profiles of biogeochemical cycling genes between soil crust and bulk soil metagenomes. Soil crust and bulk soil metagenomes were functionally annotated using the KEGG database, and compartmental differences in gene abundance were assessed for nitrogen, carbon, phosphorus and sulphur biogeochemical pathways. Significant enrichments in the abundance of genes associated with nitrogen metabolism, oxygenic photosynthesis and non-photosynthetic carbon fixation were seen in the soil crust relative to bulk soils; whereas significant enrichments of genes associated with phosphorus cycling and methane metabolism were seen in bulk soils relative to soil crusts. Further analysis revealed that in the soil surface there were significant enrichments in the abundance

of genes associated with nitrogen fixation, nitrification, oxygenic photosynthesis apparatus (PSI, PSII, cytochrome B6F, rubisco), phosphonate degradation and CAZy genes (carbohydrate-binding module, glycoside hydrolases and polysaccharide lyases). Of these, photosynthetic apparatus and nitrogen fixation genes displayed the highest enrichment in the soil crust, and were dominated by Cyanobacteria (both processes) and Proteobacteria (nitrogen fixation only) sequences. Conversely, in the bulk soil there were significant enrichments of genes associated with denitrification, the reductive TCA, the Wood-Ljungdahl pathway, methane oxidation, the phosphorus starvation response, phosphorus transport and thiosulphate oxidation relative to the soil crust – suggesting different contributions of lower soil layers to a number of biogeochemical processes. These results for the first time highlight the substantial genetic potential that temperate agricultural crusts have in the contribution to global nitrogen and carbon cycling, and further provide insights into the biota that perform these processes.

The use of metagenomics may prove a revolutionary tool in microbial ecology as it permits for unbiased taxonomic and functional analysis of microbial communities. However, metagenomic analysis remains conceptually challenging and there are numerous limitations that govern the current performance of this approach. Current metagenomic experiments are regulated by the quality of annotation databases, which to date poorly represent all aspects of the tree of life, and as a result a large number of sequences obtained from metagenomic sequencing cannot be fully utilised. Furthermore, cost and resource restrictions often constrain the size of experiments, but with reductions of sequencing costs these boundaries should lower. The use of metagenomics in this experiment was confined to a single time point, and metagenomic analysis of additional time points would uncover functional changes that occur during soil crust development. Further use of metagenomics to compare taxonomic and functional profiles of bare soils (or undeveloped crusts) to well-developed crusts could also uncover further capabilities of the soil crust.

CHAPTER 5 – GENERAL DISCUSSION

5.1. Overview

Prior research that has investigated either the structure or function of biological soil crusts have almost exclusively been performed in non-agricultural arid desert ecosystems. For example, by querying the Web of Science database against all current online scientific literature to find studies on “soil crusts” or “biocrusts” only 2.3% of all soil crust literature contained the keyword “temperate” in the main body of text, and 0.2% contained both “temperate” and “agriculture”. Research presented in this study aimed to address a current research gap in understanding the structural and biological development of soil crusts in a temperate agricultural ecosystem. Three separate studies were presented, which utilised (a) amplicon DNA sequencing, (b) X-Ray computed tomography scanning and (c) shotgun metagenomic sequencing to explore microbial community composition, soil structure and potential functioning of temperate agricultural soil crusts. Overall, results presented here demonstrated that the soil crust is a distinct soil compartment that has a substantially different structure, biological composition and functional capacity relative to lower bulk soils in a temperate agroecosystem.

5.2. The temporal and local-scale spatial dynamics of bacterial, fungal, protist and phototroph soil crust and bulk soil community composition

Chapter 2 aimed to determine the temporal and local-scale spatial dynamics of soil crust and bulk soil microbial communities in a temperate agricultural soil. Only a small number of studies that have investigated soil crust microbial community structure have made direct comparisons to lower soil layers in the soil profile, and instead comparisons are normally made to ‘bare’ soils which contain no visible soil crusts (Davies et al., 2013b, Mueller et al., 2015, Maier et al., 2018). As ‘bare soils’ are often determined based on morphological analysis (and may be in the early stages of soil crust development and contain considerable Cyanobacterial populations) comparisons to lower soils provides a better comparison to understand distinct features in soil crust communities. Results here revealed that bacterial, protist and phototrophic community composition and diversity was significantly different in soil crusts relative to bulk soils

lower in the soil profile. Soil crust communities were enriched with a number of taxa such as Cyanobacteria, Agaricomycetes, Dothideomycetes, Rhizaria, Oscillatoriales and Dinophyceae. At the OTU level, these differences were mainly attributed to *Phormidium*, *Vaucheria litorea* and *Klebsormidium flaccidum*, all of which are groups of filamentous phototrophs that have prior been shown to have roles in soil surface stabilisation, as well as carbon and nitrogen fixation (Dojani et al., 2014, Pushkareva et al., 2015, Türk and Gärtner, 2003, Mikhailyuk et al., 2014, Bu et al., 2015). Therefore, results here concluded that the soil surface and bulk soil had substantial differences in microbiome structure and diversity; and furthermore suggested an important role of filamentous phototrophs in soil crusts from temperate agricultural soils, which have previously only been explored in arid, polar or non-agricultural temperate ecosystems.

Whilst there were enrichments of a number of fungal taxa between the soil surface and bulk soil, results here revealed that there were no significant differences in fungal community composition between compartments. Fungal communities in this study were examined by amplifying the internal transcribed spacer (ITS) region between the small 5.8S rRNA and large 28S rRNA subunits in eukaryotic DNA. The ITS region has previously been criticized for its use as a phylogenetic marker to target fungal populations, as it experiences insertion and deletion induced alignment problems, and it often exhibits poor phylogenetic diversity (Nagy et al., 2012). However, the poor phylogenetic resolution of ITS regions is usually only applicable above the species taxonomic level, and compositional fungal data here was analysed at the class level. Therefore ‘poor’ phylogenetic resolution of ITS regions would have had negligible effects on results presented here, and findings here suggest that fungal populations are less sensitive to physiochemical changes experienced at the soil surface relative to other microbial communities.

Temporal analysis revealed substantial differences in bacterial, fungal, protist and phototroph community composition and diversity over time. At early time points, visual analysis revealed that there were no soil crust structures present at the soil surface, and additionally there were no differences in alpha diversity, community composition or biomass data between compartments. This suggested that soil crust

and bulk soil communities were initially the same, and that any development of soil crust communities from prior seasons was reset – which is likely caused by the physical perturbations of tillage, harvest and chemical application to the soil surface. However, significant increases in biomass C, biomass N and chlorophyll *a*, as well as reductions in alpha diversity and shifts in community composition, were seen in the soil crust over time. These changes were coupled with increased abundances of OTUs assigned to Cyanobacteria (*Phormidium*, Oscillatoriales), *Vaucheria litorea*, *Klebsormidium flaccidum* and *Dicranum scoparium* in the soil crust, all of which are important for soil crust stabilisation in either unperturbed desert or polar ecosystems, suggesting that there is global taxonomic continuity in soil crusts irrespective of climate and habitat (Pushkareva et al., 2015, Dojani et al., 2014). In particular, both morphological and DNA sequencing analysis revealed that soil crust communities had substantial enrichments of moss sequences, which prior studies have used as a proxy to show substantial soil crust development, after just 7-8 months of soil crust development (Li et al., 2002). In arid ecosystems, the quickest progression to show the presence of mosses was recorded by Wang et al. (2009), which occurred after 2 years of development in desert areas of Mongolia, but development of moss-dominated soil crusts is less well-documented in temperate regions. In addition to this, shifts in fungal community composition during periods of increased biomass C, N and chlorophyll *a* provides evidence of ecological succession of soil crust communities to support fungal saprotrophs. Together, these data all suggested that substantial development of soil crusts occurs in a temperate, agricultural ecosystem, and results here further showed the quickest progression of soil development to show an abundance of mosses in any prior reports.

Prior studies have shown that soil crusts found just metres apart can have substantially different morphology and community composition, which is thought to be mainly driven by differences in physiochemical properties, such as water and soil nutrient availability (Steven et al., 2013b). Redundancy analysis and distance decay plots here revealed that within-field local-scale variation (across 40 metres) was not an important factor in determining soil crust (or bulk soil) community composition. However, the effects of pooling samples (5 samples over an 8 x 4 metre box), which aimed to reduce sampling noise on the microsite scale (<10 metres), may have minimised the impacts

of within-field spatial variation. Results from redundancy analysis also showed that no one single factor determined composition of soil crust microbial communities, but time, temperature and chlorophyll *a* were important factors.

Co-occurrence microbial network analysis revealed that both soil crust and bulk soil microbial communities were dominated mainly by positive (~60%) as opposed to negative (~40%) correlations, suggesting a higher number of beneficial interactions between microbes such as symbioses or commensalism. However, soil crust microbial network topology was vastly more complex relative to bulk soil networks, within which soil crust networks had a higher node number, number of edges and connectivity. Together, soil crust network features were typical of networks that are less stable (less resistant) under disturbance events, which is a widely reported feature of soil crusts (Kuske et al., 2012, Santolini and Barabási, 2018). Cartographic representation of soil crust networks revealed that fungal taxa contained the greatest number of connections and were therefore the most important for soil crust network structure. This suggests an importance of individual fungal OTUs in soil crust community structure, despite prior results showing that there were no differences in fungal community structure between compartments. Furthermore, network analysis identified a *Meliniomyces* OTU as the most connected OTU in soil crust networks, which has not been reported as an important soil crust taxon in any prior study. Network analysis also identified *Phormidium*, *Rhizaria*, *Vaucheria litorea* and *Klebsormidium flaccidum*, as correlatively important OTUs, which were all identified as important OTUs in the soil crust based on relative abundance. However, temporal network analysis did not identify any major shifts in network topology over time in either the soil crust or bulk soil, suggesting that compartmental differences in co-occurrence networks had a greater effect on topology than time.

However, microbial network analysis remains a young field in microbial ecology and there is yet to remain a universally accepted method to generate microbial correlation networks. Network analysis is extremely prone to thresholding techniques and the method of filtering OTU tables, for example, rarefaction, use of multi-amplicon networks, removal of rare taxa, can all significantly impact network structure and topology (Ju et al., 2014, Berry and Widder, 2014). Furthermore, the community

remains split over the optimal correlation algorithm to use, for example SparCC or SPIEC-EASI, and the use of different algorithms can additionally contribute to differences in network topology. As a result, it is not feasible to compare co-occurrence networks across studies, and results of microbial network analyses should always be interpreted with consideration of the methods used.

This study represents the first experiment to quantify the relative contributions that stochastic and deterministic processes have on soil crust microbial community assembly (using the framework proposed by Vellend, 2010). Quantification of community assembly processes revealed that fungal, protist and phototroph soil crust communities were mainly governed by stochastic community assembly processes regardless of time. However, soil crust bacterial communities were initially dominated by deterministic processes, which was mainly homogenous selection, but became dominated by stochastic processes across time. Conversely, bulk soil communities were mainly dominated stochastic community assembly processes regardless of amplicon and time point. These results provide the first insight into the processes that drive assembly of soil crust microbial communities.

Microbial community assembly is influenced by a host of interacting factors, and categorisation of these factors to fit Vellend's framework represents a crude method to simplify these complex processes (Stegen et al., 2012, Vellend, 2010). As a result, β NTI makes a number of assumptions in order to quantify microbial communities under this framework (Stegen et al., 2012). As β NTI is performed on OTU tables obtained from amplicon data, there is no way to measure the effects of speciation – which is one of Vellend's major processes. However, the effects of speciation are proposed to have minimal impacts on microbial communities, and therefore β NTI assumes that the effects of speciation on microbial community assembly are negligible (Leibold et al., 2004, Stegen et al., 2013). The β NTI framework also assumes that there is a considerable phylogenetic signal between each OTUs optimal habitat conditions. In this case, phylogenetic signal refers to a scenario where closely related OTUs have preferred habitats relative to less closely-related OTUs. Whilst this is tested using Mantel tests and Pagel's λ , this should be tested experimentally to confirm whether this holds true. In summary, the β NTI framework is a young methodology in

microbial ecology, and further work is warranted to determine whether these assumptions are applicable to a host of other microbial systems.

Overall, results presented in chapter 2 emphasise that the soil surface is a biologically distinct soil compartment relative to the bulk soil. Additionally, soil crust communities experience complex successional shifts in community structure over time, emphasised by transitions from early-establishing phototrophic and soil stabilising taxa, to the propagation of heterotrophs and fungal saprotrophs that can utilise the increased nutrients and biomass supplied at the soil surface by pioneer populations. Network analysis results further support this notion, within which several correctively-important fungal saprotroph OTUs were identified in soil crust, but not bulk soil, communities. Fundamental ecological processes governing soil crust communities were shown to be different to those influencing bulk soil communities. Initially, ecological processes dictating soil crust community assembly were deterministic, and selection pressures restricted establishment to fast-growing pioneer phototroph communities. However, over time stabilisation of the soil surface by early colonisers enabled establishment of other taxa, such as fungal saprotrophs, and the ability to disperse to the soil surface governed community assembly. Together, these results all highlight that complex soil crust communities indeed form in temperate agroecosystems, and furthermore undergo convoluted successional changes within the space of a single growing season.

5.3. The influence of biological soil crust development on the physical soil structure throughout the soil profile

Chapter 3 aimed to determine the structural impacts that biological soil crust development had on the soil profile across a growing season in a temperate agricultural field (a different field studied in chapter 2). In this chapter, non-invasive X-ray CT scanning was used to determine structural features throughout the soil profile, whereas amplicon DNA sequencing was used to assess biological development of soil crusts. Amplicon DNA sequencing confirmed trends seen in chapter 2, where significant differences in bacterial, protist and phototroph community composition and diversity were seen between the soil crust and the bulk soil, within which soil crust communities had significant enrichments of Cyanobacteria, Actinobacteria, Chytridiomycota,

Archaeplastida and Streptophyta; and temporal analysis revealed that soil crust and bulk soil community composition became less similar across time. This emphasised that significant biological development of soil crusts occurred in this experiment, and additionally emphasised that findings from chapter 2 were replicated on a different temperate agricultural field supporting a different crop.

X-ray CT analysis of soil structures revealed that a structural seal forms at the soil surface, termed the 'soil seal', that had substantially lower porosity but higher connectivity in comparison to lower soil depths. Furthermore, pore structures within the soil seal were more complex, within which there was a higher number of pores, they had a higher surface area, higher circularity and a higher proportion of macropores relative to lower layers, but these features were only present under substantially developed biological soil crusts. The structural region of influence of biological soil crusts on the soil profile was shown to extend beyond the soil seal, and a region of altered porosity relative to the soil seal and reference bulk soil was identified immediately beneath the soil seal, which was termed the 'transitional zone'. The transitional zone had lower porosity, higher circularity and a higher pore number relative to bulk soils, but only when substantial development of biological soil crusts had occurred. Together, these results suggest that biological soil crusts have a larger structural influence on the soil profile than the 2-5 mm depth in which they span, and these structural influences are likely associated with shifts in community composition towards filamentous colonisers.

X-ray CT analysis has been revolutionary in the field of soil science as it allows *in situ* analysis of solid, liquid and gaseous material within soil samples in real time, which allowed the progression away from the analysis of thin slices or subsamples that lack environmental context. However, there are several limitations of X-ray CT scanning that can restrict the effectiveness of this technique in experiments. Firstly, X-ray CT image analysis is limited by the scanning resolution, which is the size of individual slices (or distance between voxels), and the typical resolution for industrial scanners ranges between 5 – 150 μm for biological studies (du Plessis et al., 2017). In this study, a scanning resolution of 30 μm was used, which meant that it was only possible to analyse soil structures larger than 30 μm , making it impossible to study the structural

impacts of micropores ($< 30 \mu\text{m}$), which are generally accepted as the main habitats for microorganisms in the soil profile (Luxmoore, 1981, Smucker et al., 2007). Secondly, agricultural soils are composed of a number of different materials, which include large or small soil aggregates, clay, rocks, stones, liquid material and air spaces, all of which make soil pore data incredibly noisy. For example, the presence of a single large stone (which would not be assigned to a soil pore in image analysis) in the soil profile would shift soil porosity in that region to a value of 0%, which would drastically skew data on soil porosity and soil pore structure (circularity, pore number, connectivity, surface area). However, to combat these effects, soil physicists normally measure soil structures in a representative subsample, known as a region of interest, that is not skewed by any large artefacts; but the process of selecting a region of interest to avoid particular features can introduce bias into the experiment. In this experiment, a uniform region of interest (2.4 cm x 2.4 cm x 3.0 cm) was taken from the same location in each core to minimize artefacts produced from sampling whilst minimizing bias. Thirdly, thresholding methods are required in order to differentiate solid from liquid or gaseous material in the soil profile, which is either performed on a global (all slices) or individual (manually thresholding each slice) scale. Global thresholding algorithms can often result in the incorrect allocation of solid to liquid or gaseous material (or the reverse), whereas individual thresholding is subject to human error and interpretation – both of which can severely impact downstream analyses (Iassonov et al., 2009). In order to reduce thresholding artefacts, in this experiment a global thresholding algorithm (Otsu's algorithm) was used to process all slices, and individual manual adjustment of each slice (where necessary) was performed to ensure correct thresholding was performed. Lastly, X-ray CT requires expensive state-of-the-art machinery (with high maintenance costs), and consequently is an incredibly expensive process to perform. In this experiment, it cost ~£200/sample in order to scan each core once to a ~5cm depth, which considerably restricted the sampling size of this study. Overall, X-ray CT image analysis has a number of limitations, and in this experiment these limitations were addressed to try and reduce the impacts they would have on downstream analysis.

The TopCap ImageJ algorithm was used to capture the shape of the soil surface by creating a mask that can be used to measure pore metrics iteratively throughout the

soil profile (Garbout et al., 2018). The algorithm works by using a ‘rolling curve’ diameter, within which a circular object of a specified diameter ‘rolls’ over the soil surface and records the path in which it travels. However, this algorithm was designed to capture the surface of physical soil crusts, which have a smooth microtopography and are morphologically less complex than biological soil crusts. The use of a rolling curve diameter that was too high resulted in an overestimation of the soil surface, whereas the use of a rolling curve diameter that was too low resulted in complete miscalculation of the soil surface (as the circular object was small enough to fit in crevices in the soil profile). In this experiment, a conservative rolling curve diameter was used that marginally over-estimated the soil surface. However, this resulted in soil pores being assigned above the soil surface, which significantly overestimated soil porosity in the uppermost slices. Despite manual intervention these artefacts could not be corrected, and as a result the upper slices (0.9 mm) had to be removed from downstream analysis – which contains perhaps the most important feature of the biological soil crust structure. Optimization of the rolling curve diameter algorithm, coupled with a greater scanning resolution, would reduce artefacts at the immediate soil surface, making it possible to analyse this section in future studies.

Overall, results from chapter 3 emphasised that the soil surface forms a distinct structural compartment, and biological soil crust development has a large structural impact on the soil profile than the 2-5 mm in which they span. Furthermore, biological data from this chapter confirmed that results presented in chapter 2 could be replicated on a different temperate agricultural site.

5.4. Taxonomic and functional profiles of developed biological soil crust and bulk soil metagenomes

Chapter 4 aimed to determine differences in taxonomic and functional profiles of metagenomes obtained from developed biological soil crusts and bulk soil from a temperate agricultural field. Analysis of unbiased taxonomic profiles presented here revealed that bacterial, eukaryote, archaeal and viral communities were all significantly different between the soil crust and the bulk soil. Soil crust communities displayed significant enrichments in a number of taxa that are associated with roles in soil surface stabilisation, which included Cyanobacteria, Chlorophyta, Bacteroidetes,

Actinobacteria and Proteobacteria. Furthermore, as shotgun metagenomics permits the analysis of bacterial, eukaryote, archaeal and viral communities simultaneously, results shown here for the first time show soil crust and bulk soil viral communities had a significantly different composition. Functional analysis of shotgun metagenomes was performed by assigning assembled sequences to the KEGG, COG and Pfam databases, and results were presented based on KEGG annotations. Soil crust communities were significantly enriched with genes associated with nitrogen fixation, nitrification, oxygenic photosynthesis, phosphonate degradation and CAZy genes (carbohydrate-binding modules, glycoside hydrolases and polysaccharide lyases). Of these, oxygenic photosynthesis and nitrogen fixation associated genes displayed the highest fold-change between compartments and were dominated by Cyanobacteria and Proteobacteria – suggesting a functional significance of these taxa in soil crust communities. Bulk soil communities displayed significant enrichments in genes associated with carbon fixation (through the reductive TCA and the Wood-Ljungdahl pathway), methane oxidation, phosphorus starvation, phosphorus transport and thiosulphate oxidation, which suggested differential contributions of bulk soil and soil surface communities to several biogeochemical cycles. This chapter represents the first study to analyse shotgun metagenomes of temperate agricultural soil crusts and the results suggest that they perform enriched roles in a number of biogeochemical cycles, specifically nitrogen and carbon fixation. This work remains the only study to explore the functional capabilities of soil crusts using a genetic approach, and further work is warranted to determine whether findings present here extend to different ecosystems within temperate regions.

Taxonomic analysis revealed that Eukaryotes were only assigned to a very low proportion of contigs (0.61%, by count), and mapped coverage eukaryotic contigs were ~5-fold higher in the soil crust (1.04%) than the bulk soil (0.17%). This highlights the dominance of bacteria in soil microbial communities (~90% coverage), whereas fungi and protists (which belong to the eukaryotic kingdom) only contribute to minute abundances (<1% in both compartments), which were more comparable to the proportion archaea (~1%). However, Chlorophyta (a diverse group of green algae belonging to Eukaryotes) were identified as one of the most abundant phyla across all kingdoms and were revealed as the most enriched phyla (highest fold change which

contributed >1% relative abundance) in the soil crust. Therefore, despite their low contributions to overall metagenomic reads, these results suggest that eukaryotes have an important functional role in developed temperate agricultural soil crusts, which has only been suggested previously in polar and arid ecosystems (Dumack et al., 2016, Maier et al., 2018, Abed et al., 2019). Moreover, the importance of eukaryotes in soil crust communities is supported by the multi-amplicon microbial co-occurrence networks presented in chapter 2, within which fungal OTUs displayed the highest number of connections of any amplicon.

Whilst metagenomics has been revolutionary in microbial ecology, for example in the discovery of new candidate phyla, it is not a perfect methodology and still contains a number of limitations (Parks et al., 2017). The largest limitation of metagenomics arises from the databases that are available, and large portions of the microbial tree of life are currently unannotated across all databases (Quince et al., 2017). In this study, 26.2% of contigs were not assigned beyond the phylum taxonomic level, which limited the conclusions that could be made. For example, genes annotated to assimilatory and dissimilatory nitrate reduction, denitrification, the reductive TCA and the Wood-Ljungdahl pathway contained 20-65% of sequences with no taxonomic annotation beyond 'bacteria', providing little insight into the taxa that perform these functions. However, as the price of metagenomics becomes cheaper to perform, taxonomic and functional databases will become better annotated and consequently the effectiveness of metagenomics will improve. Another problem in metagenomics arises from the approach to normalize and present gene coverage data. Whilst the future of metagenomic analysis is likely to be directed towards taxonomic binning techniques, metagenomic bins produced in this study were low quality and only a small number were suitable for further analysis, therefore functional analyses of metagenomic bins was not presented. Instead, gene abundance was converted to reads per kilobase million (RPKM), which was calculated as the mapped coverage of a given contig with the consideration of both contig length and sequencing depth of that given sample – both of which are considered as major factors that can impact coverage data across samples. Converting metagenomic gene abundance to RPKM is a popular approach that has been used in a number of studies as it accounts for sequencing depths across samples, which is also widely debated and problematic topic in amplicon sequencing,

therefore making cross-sample comparisons more applicable relative to alternative normalisation approaches such as relative gene abundance (that do not consider differences in sequencing depth) (Lawson et al., 2017, Annavajhala et al., 2018, Zhang et al., 2018). An alternative approach is to express gene abundance as a proportion of the coverage of housekeeping genes, for example RecA, an essential gene encoding a subunit of the bacterial DNA recombination protein. However, as metagenomics amplifies DNA obtained from all kingdoms simultaneously, normalisation of housekeeping genes that are conserved across all kingdoms would have to be performed – most of which can display substantially variability in gene abundance. A common approach is to divide gene abundance data by the median coverage of 32 single copy core housekeeping genes to account for the abundance of different individuals. However, use of this approach in this study resulted in extremely large variation across samples – where median coverage ranged by >100-fold difference between samples, and therefore normalization with this approach would introduce larger amounts of variation than it would have accounted for. Overall, methods used in this chapter aimed to maximize the effectiveness whilst reducing the number of artefacts in metagenomic analysis. This was performed through the use of metagenomic co-assembly, by querying sequences with three functional annotation databases and by using normalization methods to reduce variability across samples. Overall, results from chapter 3 emphasised that developed biological soil crusts had a distinct taxonomic and functional metagenomic profile relative to bulk soil. These results suggested that soil crusts had a greater capacity to perform nitrogen and carbon fixation, relative to lower depth soil, in a temperate agricultural ecosystem.

5.5. Implications of studies

Soil crusts have been estimated to cover ~12% of global terrestrial ecosystems and have predicted contributions of 40% of total biological nitrogen fixation and 7% of NPP (under vegetation) by terrestrial ecosystems (Maier et al., 2018, Rodriguez-Caballero et al., 2018, Weber et al., 2015, Elbert et al., 2012). However, these estimates have been based on extrapolations from experiments performed in unperturbed arid ecosystems. Results presented in the three experimental chapters indicate that biological soil crusts undergo substantial physical and biotic development in a highly perturbed temperate agroecosystem. Physical analysis showed that a

structural seal forms at the soil surface, and pore features of the structural seal are indicative of a functional role in protection from soil erosion, which is a widely reported role of biological soil crusts from arid regions (Belnap and Lange, 2003). Decline in soil fertility through soil erosion is a large factor contributing to losses in soil health, and it is estimated in the UK that the costs of soil erosion are between £0.9 – £1.4 billion pounds per annum (Graves et al., 2015). Encouraging or protecting development of soil crusts in agroecosystems could therefore help negate the impacts of soil erosion, but further work is required to explore the functional role of soil crusts in soil erosion in temperate ecosystems.

Results here also showed that taxonomic profiles of soil crusts contained large enrichments of filamentous taxa that are associated with soil surface stabilisation, as well as carbon and nitrogen fixing taxa. Furthermore, metagenomic analysis revealed that soil surface communities had substantially higher genetic capabilities to perform nitrogen fixation, nitrification, carbon fixation and breakdown complex carbohydrates (CAZy) relative to bulk soils. These results all suggest that soil crusts may perform considerable roles in nutrient cycling in agricultural environments, which may have implications for plant development. As bioavailable nitrogen is often a limiting factor for growth of plants, increased nitrogen fixation and nitrification by soil crusts may help increase soil fertility and therefore plant production, which has been suggested in arid regions (Song et al., 2017, Belnap and Lange, 2002, Bowker et al., 2010, Li et al., 2012). Furthermore, increased amounts of carbon fixation, coupled with the utilization of complex polysaccharides, suggests that soil crusts may promote formation of soil organic carbon, which is considered a vital factor contributing to soil health. Overall, results here suggest that soil crusts have the capacity to perform substantial roles in nutrient cycling relative to lower depth soils in a temperate agroecosystem. These roles could help increase soil health and plant productivity, but further research is warranted to explore this claim.

5.6. Future work

Together, data presented from experimental chapters here indicates for the first time that biological soil crusts from temperate agroecosystems form a separate physical soil compartment that exhibits a distinct microbial composition and has a substantially

higher capability to contribute to nitrogen and carbon cycling relative to bulk soils. However, this study was limited to just two separate temperate agroecosystems, both of which contained the same soil type. Additionally, these two fields were growing a winter wheat crop (chapters 2 and 4) and a winter barley crop (chapter 3), which are closely related, exhibit similar resource requirements and have similar morphology (and therefore similar shading effects on the soil surface) (Li et al., 2010). However, different crops would permit a varied amount of light to reach the soil surface, and could consequently impact the ability for soil crusts to develop in these systems, as light is an important factor that drives soil crust development (Belnap and Lange, 2003). Therefore, a future direction could be to investigate whether biological and structural soil crust development occurs under a number of crops that exhibit varied amounts of light coverage. Furthermore, as temperate regions have extremely variable climatic conditions, soil crust development could fluctuate dramatically across large spatial scales in temperate regions. To address this, a future study could examine long-distance spatial variation of soil crust community development in different temperate regions, which has previously been performed in arid regions, to help uncover whether substantial variation in community structure occurs in these climates (Couradeau et al., 2019, Pajares et al., 2016, Pietrasiak et al., 2011).

Moreover, temporal analysis in this experiment was performed over a single growing season (on two separate sites), and development (or destruction) of soil crusts post-harvest were not assessed. A future direction could be to assess the recovery of soil crust communities and soil structure after major perturbations across several years in a temperate agroecosystem. This could provide an insight as to whether agricultural practices can lead to a chronic disturbance on physical structure, composition and functioning of soil crust communities. Furthermore, tracking soil crust development in agroecosystems within which the soil surface is not actively destroyed each year (harvest), for example in reduced or no-till systems, would provide an insight as to whether soil crusts can reach climax communities in these environments.

In this study, shotgun metagenomics was performed using a short-read sequencing platform (150 bp, paired-end). Whilst metagenomic co-assembly was performed to help combat the problems with the analysis of short-read sequences, the use of long-

read sequencing technologies (for example the Oxford Nanopore) could permit analysis of metagenomes with a greater resolution whilst providing a greater confidence (relative to assembled contigs) in sequences (Nicholls et al., 2019).

Metagenomic data only provides an insight on the genetic capability to perform a specific function, and provides no insight into whether a given gene is actively transcribed. While we identified a potential capacity for biological soil crusts to contribute to a number of biogeochemical cycles, there is need to investigate gaseous emissions and soil nutrient analysis in biological soil crusts to provide insight into the products released from metabolic processes as well as the availability of substrates. A future study could be to use other 'omics approaches to determine whether enriched gene abundance found in metagenomes correlates with activity. Analysis of soil crust and bulk soil metatranscriptomes, all of the actively transcribed RNA, would uncover the abundance and identity of genes that are actively expressed and transcribed. However, metatranscriptomics can be difficult to perform and has a number of limitations as mRNA only represents a small pool of total RNA (~1%), has an incredibly short half-life, is easily contaminated and is vastly more expensive to study than DNA (Bashiardes et al., 2016). Furthermore, analysis of the soil crust metaproteome, all of the expressed proteins in a given sample, would provide an insight into all the proteins that are actively translated, providing a greater functional context. However, metaproteomics also has a number of limitations, the most important being the extraction efficiency, which is exacerbated in complex systems such as soil, as well as the relatively poor annotation of proteins, with significant proportions remaining uncharacterised (Petriz and Franco, 2017). Combined, the use of all three of these 'omics approaches would generate a strong dataset that could be used to understand both the functional capacity and activity of biological soil crusts.

APPENDIX

Supplementary figure 2.1. Mantel correlograms of bacterial and fungal soil crust and bulk soil communities. Significant correlations according to a mantel test are indicated by filled symbols ($p < 0.05$). Mantel correlograms were calculated using the vegan package on R.

Supplementary figure 2.2. Mantel correlograms of protist and phototroph soil crust and bulk soil communities. Significant correlations according to a mantel test are indicated by filled symbols ($p < 0.05$). Mantel correlograms were calculated using the *vegan* package on R.

Supplementary table 2.1. Differentially abundant bacterial OTUs between combined soil crust and bulk soil communities. Differentially abundant OTUs were calculated if they exhibited a $\log_2$ Fold Change $>|2|$ and a corrected p -value of <0.05 (Wald's test).

Phylum	Class	Order	Family	Genus	Species	log ₂ FoldChange	stat	p-value
[Thermi]	Deinococci	Deinococcales	Deinococcaceae	<i>Deinococcus</i>		3.16702205	4.78388508	2.50E-05
Acidobacteria	[Chloracidobacteria]	RB41	Ellin6075			2.35236084	4.78921586	2.46E-05
Actinobacteria	Actinobacteria	Actinomycetales	Nocardioidaceae	<i>Nocardioides</i>	<i>plantarum</i>	2.38645633	6.86450972	3.14E-10
Actinobacteria	Actinobacteria	Actinomycetales				3.54813896	6.55086153	2.40E-09
Armatimonadetes	[Fimbriimonadia]	[Fimbriimonadales]	[Fimbriimonadaceae]	<i>Fimbriimonas</i>		2.54856086	4.61572349	5.12E-05
Armatimonadetes	Armatimonadia	Armatimonadales	Armatimonadaceae			3.29363696	5.08922133	6.74E-06
Armatimonadetes	Armatimonadia	Armatimonadales	Armatimonadaceae			2.0342853	3.62193143	0.00225493
Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	<i>Segetibacter</i>		2.57208978	6.76040117	6.11E-10
Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	<i>Segetibacter</i>		2.48932292	4.94363745	1.25E-05
Bacteroidetes	[Saprospirae]	[Saprospirales]	Chitinophagaceae	<i>Segetibacter</i>		2.47244367	3.23370301	0.00748453
Bacteroidetes	[Saprospirae]	[Saprospirales]	Saprospiraceae			2.38730541	4.92596677	1.35E-05
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		4.59406346	8.29501651	1.42E-14
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		3.9011819	8.19794032	3.03E-14
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		3.82780289	6.97274753	1.62E-10
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		3.33430528	4.83640554	2.01E-05
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		2.90378991	3.16743376	0.00897558
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		2.61966907	3.41021527	0.00441249
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Hymenobacter</i>		2.27536615	5.81733142	1.76E-07
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Rudanella</i>		5.00710598	8.52011559	2.34E-15
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Spirosoma</i>		3.11235491	3.70709036	0.00174243
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Spirosoma</i>		3.08707563	2.94269075	0.01616288
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae	<i>Spirosoma</i>		2.85745366	3.40117937	0.0045114
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae			3.23306683	4.37745541	0.00013981
Bacteroidetes	Cytophagia	Cytophagales	Cytophagaceae			2.29462174	4.16615287	0.00033422
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Flavobacterium</i>		2.31203992	3.45747482	0.00381684
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Flavobacterium</i>		2.30099312	4.58247303	5.84E-05
Bacteroidetes	Flavobacteriia	Flavobacteriales	Flavobacteriaceae	<i>Flavobacterium</i>		2.26045267	4.60776881	5.26E-05
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	<i>Pedobacter</i>		3.08458866	5.02885863	8.60E-06
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae	<i>Pedobacter</i>		2.89596177	8.63119062	1.02E-15
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			3.73333964	4.96055163	1.17E-05
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			3.67510144	4.34800652	0.00015762
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			2.72704352	2.94169736	0.01616288
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			2.34751612	2.88396049	0.01865958
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			2.25219303	7.0910778	7.45E-11
Bacteroidetes	Sphingobacteriia	Sphingobacteriales	Sphingobacteriaceae			2.21013478	9.92171505	2.63E-20
Chloroflexi	Chloroflexi	Herpetosiphonales				2.74147155	8.77289352	3.41E-16
Chloroflexi	Chloroflexi	Herpetosiphonales				2.49390371	5.97379672	7.47E-08

Chloroflexi	Thermomicrobia					2.10256295	6.44503452	4.53E-09
Chloroflexi	Thermomicrobia					2.06126063	4.638741	4.68E-05
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Phormidium</i>		5.32225307	10.1357103	4.51E-21
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Phormidium</i>		4.097968	5.02047115	8.91E-06
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Phormidium</i>		3.92606645	7.96386866	1.57E-13
Cyanobacteria	Oscillatoriophycideae	Oscillatoriales	Phormidiaceae	<i>Phormidium</i>		3.85489847	11.0886037	3.35E-25
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	<i>Leptolyngbya</i>		5.37751252	8.03952654	1.01E-13
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	<i>Leptolyngbya</i>		5.16601543	9.31395072	4.14E-18
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae	<i>Leptolyngbya</i>		3.04370353	4.43042778	0.00011115
Cyanobacteria	Synechococcophycideae	Pseudanabaenales	Pseudanabaenaceae			4.41690882	8.74434735	4.06E-16
Cyanobacteria						3.87191847	7.48668112	5.03E-12
FBP						2.32398071	3.78986531	0.0013179
FBP						2.1231012	4.02218275	0.00058338
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	A1-B1			3.42118656	7.16431679	4.60E-11
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales	Ellin5301			2.32075271	3.31815054	0.00585516
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales				2.71985571	2.54715606	0.04222126
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales				2.26636265	4.00455927	0.00062422
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales				2.25788372	3.96410771	0.00071519
Gemmatimonadetes	Gemmatimonadetes	Gemmatimonadales				2.14833559	3.77234857	0.00138818
Planctomycetes	Phycisphaerae	WD2101				3.52622077	3.6542084	0.00204297
Planctomycetes	Phycisphaerae	WD2101				2.31452627	3.91889483	0.0008369
Planctomycetes	Phycisphaerae	WD2101				2.10092276	3.95231318	0.00074608
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae	<i>Mycoplasma</i>		2.80034748	5.30200992	2.49E-06
Proteobacteria	Alphaproteobacteria	Caulobacterales	Caulobacteraceae			2.37136067	6.51501184	3.00E-09
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	<i>Methylobacterium</i>	<i>adhaesivum</i>	2.06253982	5.46564074	1.08E-06
Proteobacteria	Alphaproteobacteria	Rhizobiales	Methylobacteriaceae	<i>Methylobacterium</i>		2.5354778	6.66514674	1.13E-09
Proteobacteria	Alphaproteobacteria	Rhodobacterales	Rhodobacteraceae	<i>Rhodobacter</i>		2.55697177	8.99837354	5.39E-17
Proteobacteria	Alphaproteobacteria	Rhodospirillales	Acetobacteraceae			2.32963621	3.69824394	0.00178535
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	<i>Sphingomonas</i>	<i>echinoides</i>	4.03409997	6.47099891	3.95E-09
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	<i>Sphingomonas</i>		2.92246712	4.15101506	0.00035227
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	<i>Sphingomonas</i>		2.67482034	5.0744907	7.03E-06
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae	<i>Sphingomonas</i>		2.19809883	4.33870634	0.00016285
Proteobacteria	Alphaproteobacteria	Sphingomonadales	Sphingomonadaceae			3.96725723	5.85492892	1.44E-07
Proteobacteria	Alphaproteobacteria	Sphingomonadales				2.68019	5.27824961	2.79E-06
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	<i>Hydrogenophaga</i>		3.9179488	6.93068269	2.10E-10
Proteobacteria	Betaproteobacteria	Burkholderiales	Comamonadaceae	<i>Limnobacter</i>		2.91998873	3.89028068	0.00092717
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae	<i>Janthinobacterium</i>		2.41723487	4.40457001	0.00012403
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae			2.40718917	8.95511026	7.26E-17
Proteobacteria	Betaproteobacteria	Burkholderiales	Oxalobacteraceae			2.08101487	5.23038241	3.49E-06
Proteobacteria	Betaproteobacteria	Burkholderiales				2.00930407	6.17686557	2.30E-08
Proteobacteria	Deltaproteobacteria	Bdellovibrionales	Bacteriovoracaceae			2.13431394	3.36122435	0.00514113
Proteobacteria	Deltaproteobacteria	Myxococcales	Myxococcaceae	<i>Myxococcus</i>		2.71841417	3.8612496	0.00102048
Proteobacteria	Deltaproteobacteria	Myxococcales	Polyangiaceae			4.26192681	6.00875439	6.19E-08
Proteobacteria	Deltaproteobacteria	Spirobacillales				2.05571167	6.24575237	1.55E-08

Proteobacteria	Gammaproteobacteria	Alteromonadales	Alteromonadaceae	<i>Cellvibrio</i>		2.12881537	5.35155575	1.95E-06
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	<i>Pseudomonas</i>	<i>viridiflava</i>	4.38866944	7.11091713	6.61E-11
Proteobacteria	Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	<i>Pseudomonas</i>		2.73312915	2.65867229	0.03239773
Proteobacteria	Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae			4.48791123	5.62559265	4.78E-07
TM7	TM7-1					2.06968815	3.72040571	0.00165894
Verrucomicrobia	[Spartobacteria]	[Chthoniobacterales]	[Chthoniobacteraceae]			2.59465096	4.64225211	4.63E-05
Verrucomicrobia	Verrucomicrobiae	Verrucomicrobiales	Verrucomicrobiaceae			2.44968162	2.8162864	0.02224066

Supplementary table 2.2. Differentially abundant fungal OTUs between combined soil crust and bulk soil communities. Differentially abundant OTUs were calculated if they exhibited a $\log_2\text{Fold Change} > |2|$ and a corrected p -value of < 0.05 (Wald's test).

Phylum	Class	Order	Family	Genus	Species	log2FoldChange	stat	p-value
Ascomycota	Dothideomycetes	Capnodiales	Teratosphaeriaceae	<i>Devriesia</i>	<i>fraseriae</i>	3.8863	5.4406	7.40E-07
Ascomycota	Dothideomycetes	Dothideales	Dothioraceae	<i>Aureobasidium</i>	<i>pullulans</i>	2.9244	4.1977	0.00026836
Ascomycota	Dothideomycetes	Myriangiales	Incertae sedis	<i>Endosporium</i>	<i>aviarium</i>	2.4581	2.7508	0.03419855
Ascomycota	Dothideomycetes	Incertae sedis	Dothideomycetes	<i>Septoriella</i>	<i>hirta</i>	3.8273	7.5155	3.23E-12
Ascomycota	Dothideomycetes	Incertae sedis	Dothideomycetes	<i>Zymoseptoria</i>	<i>brevis</i>	3.2059	5.2179	2.32E-06
Ascomycota	Dothideomycetes	Pleosporales	Didymellaceae	<i>Neosascochyta</i>	<i>exitialis</i>	3.5542	3.4597	0.00423784
Ascomycota	Dothideomycetes	Pleosporales	Didymellaceae	<i>Neosascochyta</i>	<i>graminicola</i>	4.9899	6.3173	8.34E-09
Ascomycota	Dothideomycetes	Pleosporales	Leptosphaeriaceae	<i>Leptosphaeria</i>		3.3579	6.9799	1.32E-10
Ascomycota	Dothideomycetes	Pleosporales	Leptosphaeriaceae	<i>Plenodomus</i>	<i>biglobosus</i>	3.9135	10.8252	1.64E-24
Ascomycota	Dothideomycetes	Pleosporales	Lophiostomataceae	<i>Trichometasphaeria</i>		2.4812	5.6460	2.86E-07
Ascomycota	Dothideomycetes	Pleosporales	Phaeosphaeriaceae	<i>Phaeosphaeria</i>	<i>juncicola</i>	2.0859	2.6311	0.04410719
Ascomycota	Dothideomycetes	Pleosporales	Pleosporaceae	<i>Alternaria</i>	<i>hordeicola</i>	3.5335	9.7033	6.10E-20
Ascomycota	Dothideomycetes	Pleosporales				3.1579	4.4968	7.86E-05
Ascomycota	Dothideomycetes					3.5918	3.1126	0.01291882
Ascomycota	Lecanoromycetes	Agyriales	Agyriaceae	<i>Xylographa</i>	<i>vermicularis</i>	6.0616	4.2986	0.00018265
Ascomycota	Lecanoromycetes	Teloschistales	Teloschistaceae	<i>Xanthoria</i>	<i>coomae</i>	2.3780	2.6706	0.04058791
Ascomycota	Leotiomycetes	Helotiales	Helotiaceae	<i>Articulospora</i>		3.1441	6.7156	6.91E-10
Ascomycota	Leotiomycetes	Incertae sedis	Leotiomycetes	<i>Meliniomyces</i>		3.0263	10.1910	6.82E-22
Ascomycota	Leotiomycetes	Incertae sedis	Leotiomycetes	<i>Meliniomyces</i>		2.6175	8.7346	3.07E-16
Ascomycota	Leotiomycetes	Incertae sedis	Leotiomycetes	<i>Meliniomyces</i>		2.5645	5.7015	2.26E-07
Ascomycota	Leotiomycetes					3.4334	3.8796	0.00096486
Ascomycota	Sordariomycetes	Chaetosphaeriales	Chaetosphaeriaceae			3.7587	6.3022	8.76E-09
Ascomycota	Sordariomycetes	Xylariales	Incertae sedis	<i>Microdochium</i>	<i>colombiense</i>	4.0488	5.5456	4.71E-07
Ascomycota						4.2193	5.3616	1.12E-06
Ascomycota						2.8630	8.4311	3.07E-15
Ascomycota						2.4961	3.5215	0.00354074
Basidiomycota	Agaricomycetes	Agaricales				4.0066	5.7425	1.83E-07
Basidiomycota	Agaricomycetes	Agaricales				3.9902	4.3723	0.00013289
Basidiomycota	Agaricomycetes	Agaricales				3.8184	6.0892	2.92E-08

Basidiomycota	Agaricomycetes	Agaricales				3.6250	5.6514	2.85E-07
Basidiomycota	Agaricomycetes	Agaricales				3.4700	5.5189	5.35E-07
Basidiomycota	Agaricomycetes	Agaricales				3.3558	5.4638	6.96E-07
Basidiomycota	Agaricomycetes	Agaricales				3.2854	5.9513	6.04E-08
Basidiomycota	Agaricomycetes	Agaricales				2.9156	2.6757	0.04058791
Basidiomycota	Agaricomycetes	Agaricales				2.7911	5.0732	4.91E-06
Basidiomycota	Agaricomycetes	Agaricales				2.6236	3.8682	0.00099615
Basidiomycota	Agaricomycetes	Agaricales				2.3951	5.6853	2.41E-07
Basidiomycota	Agaricomycetes	Agaricales				2.3630	7.4787	3.92E-12
Basidiomycota	Agaricomycetes	Agaricales				2.3450	5.7498	1.81E-07
Basidiomycota	Agaricomycetes	Agaricales				2.2495	4.3866	0.00012666
Basidiomycota	Agaricomycetes	Russulales	Bondarzewiaceae	<i>Heterobasidion</i>	<i>australe</i>	2.1450	3.9859	0.00063872
Basidiomycota	Agaricomycetes					2.5286	2.9470	0.02012189
Basidiomycota	Exobasidiomycetes					2.3266	3.1702	0.01085233
Basidiomycota	Microbotryomycetes	Leucosporidiales	Leucosporidiaceae	<i>Leucosporidium</i>	<i>drummii</i>	3.4729	6.4796	3.03E-09
Basidiomycota	Microbotryomycetes	Sporidiobolales	Incertae sedis	<i>Sporobolomyces</i>	<i>roseus</i>	3.9488	5.4509	7.20E-07
Basidiomycota	Tremellomycetes	Filobasidiales	Filobasidiaceae	<i>Filobasidium</i>	<i>stepposum</i>	3.0213	3.2112	0.00985072
Basidiomycota	Tremellomycetes	Tremellales	Bulleraceae	<i>Bullera</i>	<i>crocea</i>	2.5679	3.6830	0.00200743
Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	<i>Cryptococcus</i>	<i>frias</i>	4.0708	5.3524	1.16E-06
Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	<i>Cryptococcus</i>		4.8342	7.5561	2.60E-12
Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	<i>Dioszegia</i>	<i>fristingsensis</i>	2.1339	2.7578	0.03378425
Basidiomycota	Tremellomycetes	Tremellales	Tremellaceae	<i>Dioszegia</i>	<i>hungarica</i>	3.0059	3.7593	0.00152601
Basidiomycota	Tremellomycetes	Tremellales				2.7174	2.6705	0.04058791
Chytridiomycota	Chytridiomycetes					4.3271	6.5974	1.46E-09
Chytridiomycota	Spizellomycetes	Spizellomycetales	Spizellomycetaceae	<i>Powellomyces</i>		4.4963	8.1468	2.60E-14
Chytridiomycota	Spizellomycetes	Spizellomycetales	Spizellomycetaceae	<i>Powellomyces</i>		2.0792	3.4969	0.00383289
Chytridiomycota						3.3016	6.0945	2.92E-08
Chytridiomycota						2.5367	3.0676	0.01470807
Chytridiomycota						2.2899	2.6924	0.0390474
Entomophthoromycota	Basidiobolomycetes	Basidiobolales	Basidiobolaceae	<i>Basidiobolus</i>	<i>ranarum</i>	3.3386	6.7749	4.88E-10
Entomophthoromycota	Basidiobolomycetes	Basidiobolales	Basidiobolaceae	<i>Basidiobolus</i>	<i>ranarum</i>	3.1767	7.1606	3.87E-11
Glomeromycota	Glomeromycetes	Glomerales	Glomeraceae	<i>Rhizophagus</i>	<i>intraradices</i>	5.5578	8.3332	6.17E-15
Mortierellomycota	Mortierellomycotina	Mortierellales				2.6087	2.9155	0.02161648
Mucoromycota	Endogonomycetes	Endogonales				5.6495	4.9680	8.32E-06
Unidentified						3.8263	9.1804	6.73E-18
Unidentified						3.7890	6.0851	2.92E-08
Unidentified						3.0794	3.4337	0.00460877
Unidentified						3.0257	4.5253	7.00E-05
Unidentified						2.7474	2.9421	0.02023797
Unidentified						2.6531	3.0258	0.01636665
Unidentified						2.5754	4.2043	0.00026487
Unidentified						2.4790	4.1142	0.00038067

Supplementary table 2.3. Differentially abundant protist OTUs between combined soil crust and bulk soil communities. Differentially abundant OTUs were calculated if they exhibited a $\log_2$ Fold Change $>|2|$ and a corrected p -value of <0.05 (Wald's test).

<i>ta2</i>	<i>ta3</i>	<i>ta4</i>	<i>ta5</i>	<i>ta6</i>	<i>ta7</i>	log2FoldChange	stat	p-value
Amoebozoa	Fractovitelliida	Soliformovum	Amoebozoa			2.7169	3.9705	0.00033427
Amoebozoa	Myxogastria	Cavostelium	apophysatum			3.1856	5.2256	1.32E-06
Amoebozoa	Myxogastria	Protophysarum	Lamproderma	scintillans		2.5256	3.1798	0.00489548
Amoebozoa	Tubulinea	Euamoebida	BOLA868	Hartmannellidae	Hartmannellidae	3.3119	7.5127	1.40E-12
Amoebozoa	Tubulinea	Euamoebida	Glaeseria	Copromyxa protea	Copromyxa protea	3.4303	6.3242	3.50E-09
Amoebozoa	Tubulinea	Euamoebida	Glaeseria	Hartmannella	cantabrigiensis	2.4985	6.4388	1.74E-09
Amoebozoa	Tubulinea	Euamoebida	Glaeseria	Tubulinida		4.9631	7.1364	1.98E-11
Amoebozoa	Tubulinea	Leptomyxida	Eimeriidae	Eimeriidae	Eimeriidae	2.3012	5.6722	1.41E-07
Archaeplastida	Chloroplastida	Charophyta	Charales	Tolypella	Tolypella porteri	2.3882	8.6625	1.67E-16
Archaeplastida	Chloroplastida	Chlorophyta	Chlorophyceae	Kirchneriella	Neocystis brevis	3.6256	5.2168	1.35E-06
Archaeplastida	Chloroplastida	Chlorophyta	Trebouxiphyceae	Banisveld	Banisveld	3.5640	4.6711	1.97E-05
Archaeplastida	Chloroplastida	Desmidiaceae				2.1625	2.3848	0.04185442
Archaeplastida	Glaucochyta	Cyanoptyche	gloeocystis			3.0897	4.0221	0.00027323
Centrohelida	Lobosea					2.6756	5.2740	1.04E-06
Excavata	Discoba	Discicristata	Diplonemea	Diplonema	marine diplomemid	2.6440	3.7779	0.00069253
Excavata	Discoba	Discicristata	Heteronematina	Petalomonas		2.4254	7.0977	2.44E-11
Excavata	Discoba	Discicristata	Prokinetoplastina	Ichthyobodo		2.4583	5.6260	1.78E-07
Incertae Sedis	Apusomonadidae	Apusomonas	Proboscidea			2.6863	3.5826	0.00136535
Opisthokonta	Holozoa	Choanomonada	Stephanoecidae			2.9081	5.4190	5.25E-07
Opisthokonta	Holozoa	Corallochytreia	Corallochytridae	Corallochytrium	limacisporum	2.2952	2.5556	0.02810409
Opisthokonta	Nucleomycea					2.3206	2.5662	0.0275171
RT5iin25	Eimeriidae					4.8964	7.3084	6.01E-12
RT5iin25						3.2367	3.1304	0.00560483
RT5iin25						2.5897	6.5962	6.42E-10
SAR	Alveolata	Apicomplexa	Eugregarinorida	Gregarina	Heterocapsaceae	2.8260	8.5294	4.73E-16
SAR	Alveolata	Ciliophora	Colpodida	Notoxoma	Colpodidae	2.8083	6.1075	1.33E-08
SAR	Alveolata	Ciliophora	Heterotrichia	Blepharisma		3.0272	3.5978	0.00130608
SAR	Alveolata	Ciliophora	Hypotrichia			2.7369	4.2182	0.00012937
SAR	Alveolata	Protalveolata	Perkinsidae	A31	freshwater	3.1048	5.8265	6.82E-08
SAR	Alveolata	Protalveolata	Syndiniales	Amoebophrya	clone OLI11261	2.8237	5.7216	1.17E-07
SAR	Rhizaria	Cercozoa	Cercomonadida	Cercomonadida		3.0929	5.7403	1.09E-07
SAR	Rhizaria	Cercozoa	Cercomonadidae	Cercomonas		2.3332	2.6951	0.01944572
SAR	Rhizaria	Cercozoa	Cercomonadidae	Eocercomonas		3.2046	10.0910	4.38E-22
SAR	Rhizaria	Cercozoa	Glissomonadida	Allantion	Thaumatomonadida	4.2224	12.0341	3.40E-31
SAR	Rhizaria	Cercozoa	Glissomonadida	Bodomorpha	cerczoan	2.7025	3.2606	0.0038244
SAR	Rhizaria	Cercozoa	Glissomonadida	Heteromita	Bodomorpha	5.1757	4.4004	6.12E-05
SAR	Rhizaria	Cercozoa	Glissomonadida	Proleptomonas	faecicola	3.7497	10.3709	3.24E-23
SAR	Rhizaria	Cercozoa	Glissomonadida	Thaumatomonadida		4.3371	4.7438	1.41E-05
SAR	Rhizaria	Cercozoa	Glissomonadida			2.5175	2.6938	0.01944572
SAR	Rhizaria	Cercozoa	Incertae Sedis	Trachelocorythion	pulchellum	3.4846	12.9898	4.04E-36

SAR	Rhizaria	Cercozoa	Novel Clade Gran-6	cercozoan	cercozoan	2.9804	5.3942	5.85E-07
SAR	Rhizaria	Cercozoa	Peregriniidae	Gyromitus		3.6953	5.9811	2.79E-08
SAR	Rhizaria	Cercozoa	Peregriniidae	Gyromitus		2.6924	4.5863	2.90E-05
SAR	Rhizaria	Cercozoa	RT5iin19	Eimeriidae		3.4594	4.3108	8.87E-05
SAR	Rhizaria	Cercozoa	RT5iin19	Eimeriidae		2.8009	4.0438	0.00025455
SAR	Rhizaria	Cercozoa	RT5iin19	Eimeriidae		2.6571	3.4031	0.00253372
SAR	Rhizaria	Cercozoa	Thaumatomonadidae	Thaumatomastix	cercozoan	4.1082	5.3648	6.69E-07
SAR	Rhizaria	Cercozoa	Thecofilosea			2.1101	8.9442	1.80E-17
SAR	Stramenopiles	Bicosoecida	P34.6			2.3108	7.6259	6.37E-13
SAR	Stramenopiles	Ochrophyta	Ochromonadales	Paraphysomonas	butcheri	2.1683	5.3310	7.84E-07
SAR	Stramenopiles	Ochrophyta	Tribonematales	Bumilleriopsis		2.7728	8.2968	3.09E-15
SAR	Stramenopiles	Peronosporomycetes	Apodachlya	brachynema		2.7213	5.0786	2.68E-06
SAR	Stramenopiles	Peronosporomycetes	Saprolegnia	Saprolegnia parasitica		2.1049	9.2911	8.83E-19

Supplementary table 2.4. Differentially abundant phototroph OTUs between combined soil crust and bulk soil communities. Differentially abundant OTUs were calculated if they exhibited a $\log_2$ Fold Change $>|2|$ and a corrected p -value of <0.05 (Wald's test).

<i>ta1</i>	<i>ta2</i>	<i>ta3</i>	<i>ta4</i>	<i>ta5</i>	<i>ta7</i>	<i>ta8</i>	log₂FoldChange	stat	<i>p</i>-value
Bacteria	Cyanobacteria	Oscillatoriales	Leptolyngbya				3.1459	5.3684	6.14E-07
Bacteria	Cyanobacteria	Oscillatoriales	Nodosilinea				2.2575	4.6258	1.81E-05
Eukaryota	Stramenopiles	Bacillariophyta	Bacillariophyceae	Bacillariophycidae	Phaeodactylaceae	Phaeodactylum	2.2444	4.5980	1.91E-05
Eukaryota	Stramenopiles	Bacillariophyta	Coscinodiscophyceae	Thalassiosirophycidae	Thalassiosiraceae	Thalassiosira	2.0178	4.0624	0.00017565
Eukaryota	Stramenopiles	PXclade	Xanthophyceae	Vaucheriales	Vaucheria	litorea	2.1409	4.7074	1.28E-05
Eukaryota	Stramenopiles	PXclade	Xanthophyceae	Vaucheriales	Vaucheria	litorea	2.0763	3.0358	0.00599708
Eukaryota	Stramenopiles	PXclade	Xanthophyceae	Vaucheriales	Vaucheria	litorea	2.2853	5.1905	1.37E-06
Eukaryota	Viridiplantae	Chlorophyta	Trebouxiophyceae	Microthamniales	Stichococcusbacillaris		2.1771	3.3597	0.00232678
Eukaryota	Viridiplantae	Streptophyta	Klebsormidiophyceae	Klebsormidiales	Klebsormidium	flaccidum	2.1547	6.3876	1.79E-09
Eukaryota	Viridiplantae	Streptophyta	Klebsormidiophyceae	Klebsormidiales	Klebsormidium	flaccidum	2.2403	3.5261	0.00132785

Supplementary figure 3.1. Soil seal determination. Porosity was calculated from X-ray computed tomography images obtained after running the TopCap macro at intervals of 150 μm . The top 0.9 mm were removed from calculations as the TopCap algorithm overestimated porosity at the immediate soil surface. Horizontal lines indicate $\pm$ median confidence interval calculated from the porosity of the bottom 10 mm of the core. Soil seals were calculated as the depth at which porosity returned to within the median confidence interval of reference porosity.

Supplementary table 4.1. KEGG numbers for CAZy gene enzyme classes. KEGG numbers for different CAZy gene enzyme classes were obtained from the KEGG database (Kanehisha et al. 2015).

CAZy class	KO number
Auxiliary activities	K03933 K21712 K21713 K22033 K00516 K19356 K22032
Carbohydrate binding modules	K18578 K13381 K20547 K20692 K20697 K23550 K20844 K19891 K19892 K19893 K22832 K22833 K13268 K21606 K19223 K19224 K22409 K15921 K01070 K03930 K03932 K01044 K03927 K03929 K15743
Glycoside Hydrolases	K01188 K12695 K13031 K13032 K14756 K15748 K20834 K20835 K20852 K21362 K22279 K22404 K22425 K17624 K08304 K08305 K15532 K17108 K18576 K18206 K18205 K00700 K00701 K01200 K01214 K01236 K02438 K22253 K00689 K00690 K01176 K01182 K01196 K01208 K01226 K05341 K05343 K05992 K06044 K16147 K21350 K21575 K22597 K23079 K20885 K01177 K01178 K22308 K22934 K16559 K20830 K20832 K20846 K21850 K01199 K01183 K17523 K17524 K17525 K17526 K03791 K18577 K12373 K14459 K20730 K08306 K08308 K08309 K21687 K21690 K21691 K01218 K20845 K20847 K01213 K18580 K18650 K01206 K05349 K15920 K20836 K01201 K22276 K01187 K01203 K01811 K12047 K12316 K12317 K15922 K15925 K18820 K22307 K01193 K01212 K18775 K20848 K20849 K21351 K21352 K01186 K12357 K23591 K19392 K06617 K01194 K01191 K12311 K12312 K15524 K21000 K01232 K07406 K22933 K01198 K06113 K23741 K01230 K20829 K01179 K05991 K15924 K18579 K22277 K01209 K22268 K07405 K16149 K22451 K01202 K19668 K03931 K23086 K00691 K05342 K10231 K18783 K18784 K21355 K22078 K21353 K05988 K00692 K20811 K01225 K19357 K22834 K22835 K22836 K03796 K13714 K18651 K08257 K00705 K07964 K20833 K15531 K20542 K20850 K19247 K15719 K18581 K20831 K13688 K18675 K21298 K15923 K20851 K21574 K15538 K21132
Glycosyl Transferases	K00699 K00711 K00724 K01223 K04628 K05841 K07432 K07441 K08236 K08237 K10757 K12338 K12711 K12717 K12726 K12930 K12933 K12937 K12938 K12939 K13030 K13068 K13078 K13080 K13227 K13235 K13263 K13269 K13492 K13493 K13494 K13495 K13496 K13691 K13692 K14367 K14368 K14375 K14595 K14596 K15774 K15775 K15776 K15787 K15929 K15931 K15934 K15935 K15936 K15937 K15938 K15939 K15946 K15948 K15951 K15960 K15961 K15964 K15965 K15989 K15993 K15994 K16004 K16009 K16010 K16388 K16440 K16441 K16442 K16443 K16444 K18100 K18101 K19566 K19888 K20079 K21187 K21194 K21229 K21230 K21259 K21260 K21262 K21263 K21354 K21370 K21371 K21372 K21373 K21374 K22426 K22706 K22707 K22771 K22772 K22794 K22845 K23154 K23260 K23261 K23452 K23520 K23521 K23522 K23523 K23524 K23525 K23725 K00716 K00753 K03663 K07632 K07633 K07634 K07635 K09669 K14412 K14464 K20151 K23249 K23424 K00718 K20883 K21367 K00725 K09655 K00726 K09666 K00727 K00742 K00771 K09662 K09663 K20891 K03854 K05537 K10967 K00736 K00737 K00744 K09661 K00748 K20444 K00694 K00698 K00721 K00729 K00745 K00752 K00770 K02359 K03669 K03820 K07011 K07027 K10012 K10999 K11936 K12983 K12984 K12987 K12988 K12990 K12991 K12992 K12997 K12998 K12999 K13002 K13500 K13670 K13680 K13683 K14666 K16555 K16556 K16557 K16562 K16564 K16649 K16650 K16698 K16700 K16702 K17250 K18788 K19003 K19004 K19290 K19354 K19423 K19425 K19427 K20327 K20438 K20534 K20887 K20923 K20924 K20926 K21216 K21304 K21305 K21363 K21366 K22708 K22709 K22844 K22907 K23103 K23105 K23130 K23131 K23190 K23229 K23230 K23251 K23683 K23684 K23685 K23686 K00697 K03692 K16055 K00720 K03846 K03847 K05286 K08098 K00717 K11718 K07270 K11703 K23067 K02852 K05946 K13660 K21364 K00710 K02563 K03429 K03715 K00778 K00779 K00780 K00781 K03368 K03369 K03370 K03371 K03373 K03374 K03375 K03376 K03479 K03494 K03792 K06612 K06613 K06614 K06615 K06616 K16153 K00693 K02527 K00715 K00719 K00731 K00734 K00739 K00741 K03766 K03877 K05948 K07819 K07820 K07970 K07971 K09653

	K09654 K09664 K09665 K13675 K14413 K18705 K20843 K20854 K20855 K20856 K01988 K05528 K05534 K08251 K22721 K03842 K23082 K05531 K05532 K08238 K00688 K00702 K13681 K00728 K03280 K00695 K00696 K00712 K00754 K01222 K02840 K02844 K03208 K03843 K03844 K03857 K03867 K06119 K08256 K09480 K12583 K12989 K12993 K12994 K12995 K12996 K13000 K13001 K13003 K13004 K13057 K13657 K13668 K13677 K13678 K13684 K14335 K15914 K16148 K16150 K16697 K16701 K16703 K17248 K17249 K19002 K19422 K19424 K20578 K20582 K20589 K20921 K20922 K20999 K21001 K21002 K21011 K21365 K21369 K21586 K22248 K22710 K22908 K22916 K23073 K23101 K23102 K23104 K23132 K23160 K23189 K23199 K23200 K23201 K23202 K23250 K13672 K09667 K00735 K10157 K10158 K10812 K20869 K20871 K02366 K02367 K02368 K02370 K18789 K20870 K20888 K20889 K00706 K11000 K21032 K00703 K00749 K13679 K20430 K20812 K05284 K03693 K03814 K04478 K05365 K05366 K05367 K12551 K12555 K18770 K21464 K00785 K11386 K11387 K00738 K13748 K12582 K03848 K03849 K03845 K03850 K00707 K00709 K00722 K00743 K20736 K20886 K21368 K13666 K03714 K18134 K18207 K05529 K05530 K05533 K02369 K03691 K07151 K21306 K13673 K13690 K00747 K13499 K00708 K00723 K00733 K00746 K07553 K07966 K07967 K07968 K07969 K09656 K09657 K09905 K13659 K05535 K05536 K05538 K10968 K10969 K12981 K13379 K07542 K11714 K20783 K20784 K20892 K18818 K13674 K09668 K00750 K03275 K03276 K03278 K03279 K12916 K12985 K12986 K13648 K18819 K20867 K20890 K20893 K22809 K12248 K13693 K21349 K13689 K23084 K23085 K13686 K13669 K13671 K13687 K02841 K02843 K02849 K12982 K13667 K13682 K14434 K20782
Polysaccharide	K18197 K18198 K19052 K19050 K19051 K18195 K19053 K01727 K19049
Lyases	

Supplementary table 4.2. Phosphorus cycling gene KEGG numbers. KEGG gene numbers were used to obtain functional annotations of sequences associated with phosphorus cycling as categorised in Liu et al (2018).

	Gene category	Gene	KO number	KEGG description
Phosphatase	Acid phosphatase	PHO	K01078	PHO; acid phosphatase [EC:3.1.3.2]
	Acid phosphatase	phoN	K09474	phoN; acid phosphatase (class A) [EC:3.1.3.2]
	Acid phosphatase	aphA	K03788	aphA; acid phosphatase (class B) [EC:3.1.3.2]
	Alkaline phosphatase	phoA	K01077	E3.1.3.1, phoA, phoB; alkaline phosphatase [EC:3.1.3.1]
	Alkaline phosphatase	phoD	K01113	phoD; alkaline phosphatase D [EC:3.1.3.1]
	GP phosphodiesterase	ugpQ	K01126	E3.1.4.46, glpQ, ugpQ; glycerophosphoryl diester phosphodiesterase [EC:3.1.4.46]
	Phosphotriesterase	php	K07048	PTER, php; phosphotriesterase-related protein
Phytase	4-Phytase	appA	K01093	appA; 4-phytase / acid phosphatase [EC:3.1.3.26 3.1.3.2]
	3-Phytase	3-phytase	K01083	E3.1.3.8; 3-phytase [EC:3.1.3.8]
Phosphonate degradation	C-P lyase multi-enzyme complex	phnF	K02043	phnF; GntR family transcriptional regulator, phosphonate transport system regulatory protein
	C-P lyase multi-enzyme complex	phnG	K06166	phnG; alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnG [EC:2.7.8.37]
	C-P lyase multi-enzyme complex	phnH	K06165	phnH; alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnH [EC:2.7.8.37]
	C-P lyase multi-enzyme complex	phnI	K06164	phnI; alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnI [EC:2.7.8.37]
	C-P lyase multi-enzyme complex	phnJ	K06163	phnJ; alpha-D-ribose 1-methylphosphonate 5-phosphate C-P lyase [EC:4.7.1.1]
	C-P lyase multi-enzyme complex	phnK	K05781	phnK; putative phosphonate transport system ATP-binding prot
	C-P lyase multi-enzyme complex	phnL	K05780	phnL; alpha-D-ribose 1-methylphosphonate 5-triphosphate synthase subunit PhnL [EC:2.7.8.37]
	C-P lyase multi-enzyme complex	phnM	K06162	phnM; alpha-D-ribose 1-methylphosphonate 5-triphosphate diphosphatase [EC:3.6.1.63]

	C-P lyase multi-enzyme complex	phnN	K05774	phnN; ribose 1,5-bisphosphokinase [EC:2.7.4.23]
	C-P lyase multi-enzyme complex	phnO	K09994	phnO; (aminoalkyl)phosphonate N-acetyltransferase [EC:2.3.1.280]
	C-P lyase multi-enzyme complex	phnP	K06167	phnP; phosphoribosyl 1,2-cyclic phosphate phosphodiesterase [EC:3.1.4.55]
	Phosphonatase	phnX	K05306	phnX; phosphonoacetaldehyde hydrolase [EC:3.11.1.1]
	Phosphonacetate hydrolase	phnA	K06193	phnA; protein PhnA
Phosphorus transporter	Inorganic pyrophosphatase	ppa	K01507	ppa; inorganic pyrophosphatase [EC:3.6.1.1]
	Exopolyphosphatase	ppx	K01514	PRUNE, PPX1; exopolyphosphatase [EC:3.6.1.11]
	Polyphosphate kinase	ppk	K00937	ppk; polyphosphate kinase [EC:2.7.4.1]
	PQQGDH	gcd	K00117	gcd; quinoprotein glucose dehydrogenase [EC:1.1.5.2]
	Phosphate inorganic transporter	pit	K03306	TC.PIT; inorganic phosphate transporter, PiT family
	Low-affinity inorganic phosphate transporter	pitA	K16322	pit; low-affinity inorganic phosphate transporter
	Phosphate-specific transport system subunit	pstA	K02038	pstA; phosphate transport system permease protein
	Phosphate-specific transport system subunit	pstB	K02036	pstB; phosphate transport system ATP-binding protein [EC:7.3.2.1]
	Phosphate-specific transport system subunit	pstC	K02037	pstC; phosphate transport system permease protein
	Phosphate-specific transport system subunit	pstS	K02040	pstS; phosphate transport system substrate-binding protein
	Phosphonate transporter subunit	phnC	K02041	phnC; phosphonate transport system ATP-binding protein [EC:7.3.2.2]
	Phosphonate transporter subunit	phnD	K02044	phnD; phosphonate transport system substrate-binding protein
	Phosphonate transporter subunit	phnE	K02042	phnE; phosphonate transport system permease protein
	Glycerol-3-phosphate transporter subunit	ugpA	K05814	ugpA; sn-glycerol 3-phosphate transport system permease protein

	Glycerol-3-phosphate transporter subunit	ugpB	K05813	ugpB; sn-glycerol 3-phosphate transport system substrate-binding protein
	Glycerol-3-phosphate transporter subunit	ugpC	K05816	ugpC; sn-glycerol 3-phosphate transport system ATP-binding protein [EC:7.6.2.10]
	Glycerol-3-phosphate transporter subunit	ugpE	K05815	ugpE; sn-glycerol 3-phosphate transport system permease protein
Regulation of phosphate starvation genes	Phosphate regulon response regulator	phoB	K07657	phoB; two-component system, OmpR family, phosphate regulon response regulator PhoB
	Phosphate regulon sensor histidine kinase	phoR	K07636	phoR; two-component system, OmpR family, phosphate regulon sensor histidine kinase PhoR [EC:2.7.13.3]
	PhoR/PhoB inhibitor protein	phoU	K02039	phoU; phosphate transport system protein

BIBLIOGRAPHY

- ABDOLLAHI, L., MUNKHOLM, L. & GARBOUT, A. 2013. Tillage System and Cover Crop Effects on Soil Quality: II. Pore Characteristics. *Soil Sci. Soc. Am. J.*
- ABED, R. M. M., LAM, P., DE BEER, D. & STIEF, P. 2013. High rates of denitrification and nitrous oxide emission in arid biological soil crusts from the Sultanate of Oman. *The Isme Journal*, 7, 1862.
- ABED, R. M. M., TAMM, A., HASSENRÜCK, C., AL-RAWAHI, A. N., RODRÍGUEZ-CABALLERO, E., FIEDLER, S., MAIER, S. & WEBER, B. 2019. Habitat-dependent composition of bacterial and fungal communities in biological soil crusts from Oman. *Scientific reports*, 9, 6468-6468.
- ADHIKARI, M., KIM, S., YADAV, D., BABU, G., KIM, C., LEE, H. & SU LEE, Y. 2014. A New Report on *Oidiodendron flavum* Isolated from Field Soil in Korea. *The Korean journal of Mycology*.
- ALAOUI, A., LIPIEC, J. & GERKE, H. H. 2011. A review of the changes in the soil pore system due to soil deformation: A hydrodynamic perspective. *Soil and Tillage Research*, 115-116, 1-15.
- ALMEIDA, A., MITCHELL, A. L., TARKOWSKA, A. & FINN, R. D. 2018. Benchmarking taxonomic assignments based on 16S rRNA gene profiling of the microbiota from commonly sampled environments. *GigaScience*, 7, giy054.
- ALNEBERG, J., BJARNASON, B. S., DE BRUIJN, I., SCHIRMER, M., QUICK, J., IJAZ, U. Z., LAHTI, L., LOMAN, N. J., ANDERSSON, A. F. & QUINCE, C. 2014. Binning metagenomic contigs by coverage and composition. *Nature Methods*, 11, 1144.
- AMATO, M. & LADD, J. N. 1992. Decomposition of ¹⁴C-labelled glucose and legume material in soils: Properties influencing the accumulation of organic residue C and microbial biomass C. *Soil Biology and Biochemistry*, 24, 455-464.
- AMBERT-SANCHEZ, M., MICKELSON, S., AHMED, S., GRAY, J. N. & WEBBER, D. 2016. Evaluating Soil Tillage Practices Using X-Ray Computed Tomography and Conventional Laboratory Methods. *Transactions of the ASABE*, 59, 455-463.
- ANDREW, S. 2010. *FASTQC. A quality control tool for high throughput sequence data* [Online]. [Accessed].
- ANGEL, R. & CONRAD, R. 2013. Elucidating the microbial resuscitation cascade in biological soil crusts following a simulated rain event. *Environmental Microbiology*, 15, 2799-2815.
- ANGEL, R., MATTHIES, D. & CONRAD, R. 2011. Activation of Methanogenesis in Arid Biological Soil Crusts Despite the Presence of Oxygen. *PLOS ONE*, 6, e20453.
- ANNAVAJHALA, M. K., KAPOOR, V., SANTO-DOMINGO, J. & CHANDRAN, K. 2018. Structural and Functional Interrogation of Selected Biological Nitrogen

Removal Systems in the United States, Denmark, and Singapore Using Shotgun Metagenomics. *Frontiers in Microbiology*, 9, 2544.

- ANWAR, M. Z., LANZEN, A., BANG-ANDREASEN, T. & JACOBSEN, C. S. 2019. To assemble or not to resemble-A validated Comparative Metatranscriptomics Workflow (CoMW). *GigaScience*, 8, giz096.
- ARMENISE, E., SIMMONS, R. W., AHN, S., GARBOUT, A., DOERR, S. H., MOONEY, S. J., STURROCK, C. J. & RITZ, K. 2018. Soil seal development under simulated rainfall: Structural, physical and hydrological dynamics. *Journal of hydrology*, 556, 211-219.
- ARRIAGA, F. J., GUZMAN, J. & LOWERY, B. 2017. Chapter 5 - Conventional Agricultural Production Systems and Soil Functions. In: AL-KAISI, M. M. & LOWERY, B. (eds.) *Soil Health and Intensification of Agroecosystems*. Academic Press.
- ASCHENBACH, K., CONRAD, R., ŘEHÁKOVÁ, K., DOLEŽAL, J., JANATKOVÁ, K. & ANGEL, R. 2013. Methanogens at the top of the world: occurrence and potential activity of methanogens in newly deglaciated soils in high-altitude cold deserts in the Western Himalayas. *Frontiers in Microbiology*, 4, 359.
- ASHELFORD, K. E., DAY, M. J. & FRY, J. C. 2003. Elevated Abundance of Bacteriophage Infecting Bacteria in Soil. *Applied and Environmental Microbiology*, 69, 285.
- AWADHWAL, N. K. & THIERSTEIN, G. E. 1985. Soil crust and its impact on crop establishment: A review. *Soil and Tillage Research*, 5, 289-302.
- AYLING, M., CLARK, M. D. & LEGGETT, R. M. 2019. New approaches for metagenome assembly with short reads. *Briefings in Bioinformatics*.
- BADGER, M. R. & BEK, E. J. 2008. Multiple Rubisco forms in proteobacteria: their functional significance in relation to CO₂ acquisition by the CBB cycle. *Journal of Experimental Botany*, 59, 1525-1541.
- BAHRAM, M., HILDEBRAND, F., FORSLUND, S. K., ANDERSON, J. L., SOUDZILOVSKAIA, N. A., BODEGOM, P. M., BENGTSSON-PALME, J., ANSLAN, S., COELHO, L. P., HAREND, H., HUERTA-CEPAS, J., MEDEMA, M. H., MALTZ, M. R., MUNDRA, S., OLSSON, P. A., PENT, M., PÖLME, S., SUNAGAWA, S., RYBERG, M., TEDERSOO, L. & BORK, P. 2018. Structure and function of the global topsoil microbiome. *Nature*, 560, 233-237.
- BAPIRI, A., BÅÅTH, E. & ROUSK, J. 2010. Drying–Rewetting Cycles Affect Fungal and Bacterial Growth Differently in an Arable Soil. *Microbial Ecology*, 60, 419-428.
- BARAK, P., JOBE, B. O., KRUEGER, A. R., PETERSON, L. A. & LAIRD, D. A. 1997. Effects of long-term soil acidification due to nitrogen fertilizer inputs in Wisconsin. *Plant and Soil*, 197, 61-69.
- BARBERÁN, A., CASAMAYOR, E. O. & FIERER, N. 2014. The microbial contribution to macroecology. *Frontiers in Microbiology*, 5, 203.

- BARDGETT, R. D. 2005. *The biology of soil : a community and ecosystem approach*, New York, Oxford University Press.
- BARNARD, R. L., OSBORNE, C. A. & FIRESTONE, M. K. 2013. Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *The Isme Journal*, 7, 2229.
- BARTRAM, A. K., JIANG, X., LYNCH, M. D. J., MASELLA, A. P., NICOL, G. W., DUSHOFF, J. & NEUFELD, J. D. 2014. Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. *FEMS Microbiology Ecology*, 87, 403-415.
- BASHIARDES, S., ZILBERMAN-SCHAPIRA, G. & ELINAV, E. 2016. Use of Metatranscriptomics in Microbiome Research. *Bioinformatics and biology insights*, 10, 19-25.
- BASTIAAN VON MEIJENFELDT, F., ARKHIPOVA, K., D. CAMBUY, D., HERNANDES COUTINHO, F. & E. DUTILH, B. 2019. *Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT: Supplementary table 1*.
- BATES, S. T., NASH, T. H. & GARCIA-PICHEL, F. 2012. Patterns of diversity for fungal assemblages of biological soil crusts from the southwestern United States. *Mycologia*, 104, 353-361.
- BATES, S. T., NASH, T. H., SWEAT, K. G. & GARCIA-PICHEL, F. 2010. Fungal communities of lichen-dominated biological soil crusts: Diversity, relative microbial biomass, and their relationship to disturbance and crust cover. *Journal of Arid Environments*, 74, 1192-1199.
- BATJES, N. H. 1996. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 47, 151-163.
- BATJES, N. H. 2014. Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, 65, 10-21.
- BAUMANN, K., SIEBERS, M., KRUSE, J., ECKHARDT, K.-U., HU, Y., MICHALIK, D., SIEBERS, N., KAR, G., KARSTEN, U. & LEINWEBER, P. 2019. Biological soil crusts as key player in biogeochemical P cycling during pedogenesis of sandy substrate. *Geoderma*, 338, 145-158.
- BECERRA-ABSALÓN, I., MUÑOZ-MARTÍN, M. Á., MONTEJANO, G. & MATEO, P. 2019. Differences in the Cyanobacterial Community Composition of Biocrusts From the Drylands of Central Mexico. Are There Endemic Species? *Frontiers in Microbiology*, 10, 937.
- BECKER, S., BÖGER, P., OEHLMANN, R. & ERNST, A. 2000. PCR Bias in Ecological Analysis: a Case Study for Quantitative Taq Nuclease Assays in Analyses of Microbial Communities. *Applied and Environmental Microbiology*, 66, 4945.
- BECKING, L. G. M. B. 1931. *Gaia of leven en aarde*, Martinus Nijhoff.

- BELNAP, J. 2003a. Biological Soil Crusts and Wind Erosion. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BELNAP, J. 2003b. Comparative Structure of Physical and Biological Soil Crusts. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BELNAP, J. 2003c. The world at your feet: desert biological soil crusts. *Frontiers in Ecology and the Environment*, 1, 181-189.
- BELNAP, J. 2006. The potential roles of biological soil crusts in dryland hydrologic cycles. *Hydrological Processes*, 20, 3159-3178.
- BELNAP, J., BÜDEL, B. & LANGE, O. L. 2003a. Biological Soil Crusts: Characteristics and Distribution. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BELNAP, J. & ELDRIDGE, D. 2003. Disturbance and Recovery of Biological Soil Crusts. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BELNAP, J. & GARDNER, J. S. 1993. SOIL MICROSTRUCTURE IN SOILS OF THE COLORADO PLATEAU: THE ROLE OF THE CYANOBACTERIUM MICROCOLEUS VAGINATUS. *The Great Basin Naturalist*, 53, 40-47.
- BELNAP, J. & GILLETTE, D. A. 1997. Disturbance of biological soil crusts: impacts on potential wind erodibility of sandy desert soils in southeastern Utah. *Land Degradation & Development*, 8, 355-362.
- BELNAP, J. & HARPER, K. T. 1995. Influence of cryptobiotic soil crusts on elemental content of tissue of 2 desert seed plants. *Arid Soil Research and Rehabilitation*, 9, 107-115.
- BELNAP, J. & LANGE, O. L. 2002. *Biological Soil Crusts: Structure, Function, and Management*, Springer Berlin Heidelberg.
- BELNAP, J. & LANGE, O. L. 2003. Structure and Functioning of Biological Soil Crusts: a Synthesis. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.
- BELNAP, J., PHILLIPS, S. L., FLINT, S., MONEY, J. & CALDWELL, M. 2008. Global change and biological soil crusts: effects of ultraviolet augmentation under altered precipitation regimes and nitrogen additions. *Global Change Biology*, 14, 670-686.
- BELNAP, J., PRASSE, R. & HARPER, K. T. 2003b. Influence of Biological Soil Crusts on Soil Environments and Vascular Plants. *In: BELNAP, J. & LANGE, O. L. (eds.) Biological Soil Crusts: Structure, Function, and Management.* Berlin, Heidelberg: Springer Berlin Heidelberg.

- BERG, I. A. 2011. Ecological Aspects of the Distribution of Different Autotrophic CO₂ Fixation Pathways. *Applied and Environmental Microbiology*, 77, 1925.
- BERG, M. 2012. Patterns of Biodiversity at Fine and Small Spatial Scales.
- BERGERET, P. 2017. The future of food and agriculture: trends and challenges [Note de lecture]. *Futuribles*.
- BERLEMONT, R. & MARTINY, A. C. 2016. Glycoside Hydrolases across Environmental Microbial Communities. *PLoS computational biology*, 12, e1005300-e1005300.
- BERRY, D. & WIDDER, S. 2014. Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Frontiers in Microbiology*, 5, 219.
- BETHANY, J., GIRALDO-SILVA, A., NELSON, C., BARGER, N. N. & GARCIA-PICHEL, F. 2019. Optimizing the Production of Nursery-Based Biological Soil Crusts for Restoration of Arid Land Soils. *Applied and Environmental Microbiology*, 85, e00735-19.
- BEVEN, K. & GERMANN, P. 2013. Macropores and water flow in soils revisited. *Water Resources Research*, 49, 3071-3092.
- BIBER, J., HERNANDO, N. & FORSTER, I. 2013. Phosphate Transporters and Their Function. *Annual Review of Physiology*, 75, 535-550.
- BISCHOFF, H. W. & BOLD, H. C. 1963. *Some soil algae from Enchanted Rock and related algal species*, Austin, Tex., University of Texas.
- BLAY, E. S., SCHWABEDISSEN, S. G., MAGNUSON, T. S., AHO, K. A., SHERIDAN, P. P. & LOHSE, K. A. 2017. Variation in Biological Soil Crust Bacterial Abundance and Diversity as a Function of Climate in Cold Steppe Ecosystems in the Intermountain West, USA. *Microbial Ecology*, 1-10.
- BLUM, W. E. H. 1990. The Challenge of Soil Protection in Europe. *Environmental Conservation*, 17, 72-74.
- BOLGER, A. M., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30, 2114-2120.
- BORASTON, ALISDAIR B., BOLAM, DAVID N., GILBERT, HARRY J. & DAVIES, GIDEON J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochemical Journal*, 382, 769.
- BORCARD, D. & LEGENDRE, P. 2002. All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecological Modelling*, 153, 51-68.
- BORCARD, D., LEGENDRE, P., AVOIS-JACQUET, C. & TUOMISTO, H. 2004. DISSECTING THE SPATIAL STRUCTURE OF ECOLOGICAL DATA AT MULTIPLE SCALES. *Ecology*, 85, 1826-1832.
- BOWKER, M. A. 2007. Biological Soil Crust Rehabilitation in Theory and Practice: An Underexploited Opportunity. *Restoration Ecology*, 15, 13-23.

- BOWKER, M. A., BELNAP, J. & MILLER, M. E. 2006. Spatial modeling of biological soil crusts to support rangeland assessment and monitoring. *Rangeland Ecology and Management*, 59, 519-529.
- BOWKER, M. A., MAESTRE, F. T. & ESCOLAR, C. 2010. Biological crusts as a model system for examining the biodiversity–ecosystem function relationship in soils. *Soil Biology and Biochemistry*, 42, 405-417.
- BOWKER, M. A., REED, S. C., MAESTRE, F. T. & ELDRIDGE, D. J. 2018. Biocrusts: the living skin of the earth. *Plant and Soil*, 429, 1-7.
- BRANKATSCHK, R., FISCHER, T., VESTE, M. & ZEYER, J. 2013. Succession of N cycling processes in biological soil crusts on a Central European inland dune. *FEMS Microbiology Ecology*, 83, 149-160.
- BREITWIESER, F. P., LU, J. & SALZBERG, S. L. 2017. A review of methods and databases for metagenomic classification and assembly. *Briefings in Bioinformatics*.
- BROWN, G. G., BAROIS, I. & LAVELLE, P. 2000. Regulation of soil organic matter dynamics and microbial activity in the drilosphere and the role of interactions with other edaphic functional domains. Paper presented at the 16th World Congress of Soil Science, 20–26 August 1998, Montpellier, France. *European Journal of Soil Biology*, 36, 177-198.
- BU, C., WU, S., HAN, F., YANG, Y. & MENG, J. 2015. The Combined Effects of Moss-Dominated Biocrusts and Vegetation on Erosion and Soil Moisture and Implications for Disturbance on the Loess Plateau, China. *PLOS ONE*, 10, e0127394.
- BUCHFINK, B., XIE, C. & HUSON, D. H. 2014. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 12, 59.
- BÜDEL, B. 1999. Ecology and diversity of rock-inhabiting cyanobacteria in tropical regions. *European Journal of Phycology*, 34, 361-370.
- BÜDEL, B., COLESIE, C., GREEN, T. G. A., GRUBE, M., LÁZARO SUAU, R., LOEWEN-SCHNEIDER, K., MAIER, S., PEER, T., PINTADO, A., RAGGIO, J., RUPRECHT, U., SANCHO, L. G., SCHROETER, B., TÜRK, R., WEBER, B., WEDIN, M., WESTBERG, M., WILLIAMS, L. & ZHENG, L. 2014. Improved appreciation of the functioning and importance of biological soil crusts in Europe: the Soil Crust International Project (SCIN). *Biodiversity and conservation*, 23, 1639-1658.
- BÜDEL, B., DARIENKO, T., DEUTSCHEWITZ, K., DOJANI, S., FRIEDL, T., MOHR, K. I., SALISCH, M., REISSER, W. & WEBER, B. 2009. Southern African Biological Soil Crusts are Ubiquitous and Highly Diverse in Drylands, Being Restricted by Rainfall Frequency. *Microbial Ecology*, 57, 229-247.
- BÜDEL, B., DULIĆ, T., DARIENKO, T., RYBALK, N. & FRIEDL, T. 2016. Cyanobacteria and Algae of Biological Soil Crusts. In: WEBER, B., BÜDEL, B. & BELNAP, J. (eds.) *Biological Soil Crusts: An Organizing Principle in Drylands*. Cham: Springer International Publishing.

- BULLARD, J. E., OCKELFORD, A., STRONG, C. & AUBAULT, H. 2018. Effects of Cyanobacterial Soil Crusts on Surface Roughness and Splash Erosion. *Journal of Geophysical Research: Biogeosciences*, 123, 3697-3712.
- BURKI, F. & KEELING, P. J. 2014. Rhizaria. *Current Biology*, 24, R103-R107.
- CAI, X., YANG, L., ZHANG, J. & LI, G. Q. 2014. First Report of *Leptosphaeria biglobosa* Causing Black Leg on *Raphanus sativus* in Central China. *Plant Disease*, 98, 993-993.
- CALLAHAN, B. J., MCMURDIE, P. J., ROSEN, M. J., HAN, A. W., JOHNSON, A. J. A. & HOLMES, S. P. 2016. DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, 13, 581.
- CANFORA, L., VENDRAMIN, E., VITTORI ANTISARI, L., LO PAPA, G., DAZZI, C., BENEDETTI, A., IAVAZZO, P., ADAMO, P., JUNGBLUT, A. D. & PINZARI, F. 2016. Compartmentalization of gypsum and halite associated with cyanobacteria in saline soil crusts. *FEMS Microbiology Ecology*, 92.
- CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic acids research*, 37, D233-D238.
- CAPORASO, J. G., KUCZYNSKI, J., STOMBAUGH, J., BITTINGER, K., BUSHMAN, F. D., COSTELLO, E. K., FIERER, N., PEÑA, A. G., GOODRICH, J. K., GORDON, J. I., HUTTLEY, G. A., KELLEY, S. T., KNIGHTS, D., KOENIG, J. E., LEY, R. E., LOZUPONE, C. A., MCDONALD, D., MUEGGE, B. D., PIRRUNG, M., REEDER, J., SEVINSKY, J. R., TURNBAUGH, P. J., WALTERS, W. A., WIDMANN, J., YATSUNENKO, T., ZANEVELD, J. & KNIGHT, R. 2010. QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, 7, 335.
- CARSON, J. K., GONZALEZ-QUINONES, V., MURPHY, D. V., HINZ, C., SHAW, J. A. & GLEESON, D. B. 2010. Low Pore Connectivity Increases Bacterial Diversity in Soil. *Applied and Environmental Microbiology*, 76, 3936.
- CASTLE, S. C., MORRISON, C. D. & BARGER, N. N. 2011. Extraction of chlorophyll a from biological soil crusts: A comparison of solvents for spectrophotometric determination. *Soil Biology and Biochemistry*, 43, 853-856.
- CHAI, Y., YUE, M., LIU, X., GUO, Y., WANG, M., XU, J., ZHANG, C., CHEN, Y., ZHANG, L. & ZHANG, R. 2016. Patterns of taxonomic, phylogenetic diversity during a long-term succession of forest on the Loess Plateau, China: insights into assembly process. *Scientific Reports*, 6, 27087.
- CHAMIZO, S., ADESSI, A., MUGNAI, G., SIMIANI, A. & DE PHILIPPIS, R. 2018. Soil Type and Cyanobacteria Species Influence the Macromolecular and Chemical Characteristics of the Polysaccharidic Matrix in Induced Biocrusts. *Microbial Ecology*.
- CHANG, Y., DESIRÒ, A., NA, H., SANDOR, L., LIPZEN, A., CLUM, A., BARRY, K., GRIGORIEV, I. V., MARTIN, F. M., STAJICH, J. E., SMITH, M. E., BONITO, G. & SPATAFORA, J. W. 2019. Phylogenomics of Endogonaceae and evolution of mycorrhizas within Mucoromycota. *New Phytologist*, 222, 511-525.

- CHAPIN, S. F. 1980. *The mineral nutrition of wild plants*.
- CHASE, J. M., KRAFT, N. J. B., SMITH, K. G., VELLEND, M. & INOUE, B. D. 2011. Using null models to disentangle variation in community dissimilarity from variation in α -diversity. *Ecosphere*, 2, art24.
- CHILTON, A., NEILAN, B. & ELDRIDGE, D. 2017. Biocrust morphology is linked to marked differences in microbial community composition. *Plant and Soil*.
- CIAIS, P., SABINE, C., BALA, G., BOPP, L., BROVKIN, V., CANADELL, J., CHHABRA, A., DEFRIES, R., GALLOWAY, J., HEIMANN, M., JONES, C., LE QUÉRE, C., MYNENI, R. B., PIAO, S. & THORNTON, P. 2013. Carbon and Other Biogeochemical Cycles. In: STOCKER, T. F., QIN, D., PLATTNER, G.-K., TIGNOR, M., ALLEN, S. K., BOSCHUNG, J., NAUELS, A., XIA, Y., BEX, V. & MIDGLEY, P. M. (eds.) *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press.
- COLE, J. R., WANG, Q., FISH, J. A., CHAI, B., MCGARRELL, D. M., SUN, Y., BROWN, C. T., PORRAS-ALFARO, A., KUSKE, C. R. & TIEDJE, J. M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic acids research*, 42, D633-D642.
- COLICA, G., LI, H., ROSSI, F., LI, D., LIU, Y. & DE PHILIPPIS, R. 2014. Microbial secreted exopolysaccharides affect the hydrological behavior of induced biological soil crusts in desert sandy soils. *Soil Biology and Biochemistry*, 68, 62-70.
- COMTE, K., ŠABACKÁ, M., CARRÉ-MLOUKA, A., ELSTER, J. & KOMÁREK, J. 2007. Relationships between the Arctic and the Antarctic cyanobacteria; three Phormidium-like strains evaluated by a polyphasic approach. *FEMS Microbiology Ecology*, 59, 366-376.
- CONDON, L. A., PIETRASIAK, N., ROSENTERTER, R. & PYKE, D. A. 2019. Passive restoration of vegetation and biological soil crusts following 80 years of exclusion from grazing across the Great Basin. *Restoration Ecology*, 0.
- CONG, J., ZHU, J., ZHANG, C., LI, T., LIU, K., LIU, D., ZHOU, N., JIANG, M., HOU, H. & ZHANG, X. 2019. Chemotherapy Alters the Phylogenetic Molecular Ecological Networks of Intestinal Microbial Communities. *Frontiers in Microbiology*, 10, 1008.
- COURADEAU. 2013. *Researchers discover global warming may affect microbe survival* [Online]. [Accessed 27/08/2019 2019].
- COURADEAU, E., FELDE, V. J. M. N. L., PARKINSON, D., UTEAU, D., ROCHET, A., CUELLAR, C., WINEGAR, G., PETH, S., NORTEN, T. R. & GARCIA-PICHEL, F. 2018. In Situ X-Ray Tomography Imaging of Soil Water and Cyanobacteria From Biological Soil Crusts Undergoing Desiccation. *Frontiers in Environmental Science*, 6, 65.
- COURADEAU, E., GIRALDO-SILVA, A., DE MARTINI, F. & GARCIA-PICHEL, F. 2019. Spatial segregation of the biological soil crust microbiome around its

- foundational cyanobacterium, *Microcoleus vaginatus*, and the formation of a nitrogen-fixing cyanosphere. *Microbiome*, 7, 55.
- CRAIN, G., MCLAREN, J., BRUNNER, B. & DARROUZET-NARDI, A. 2018. Biologically Available Phosphorus in Biocrust-Dominated Soils of the Chihuahuan Desert. *Soil Systems*, 2, 56.
- CROUZET, O., CONSENTINO, L., PÉTRAUD, J.-P., MARRAUD, C., AGUER, J.-P., BUREAU, S., LE BOURVELLEC, C., TOULOUMET, L. & BÉRARD, A. 2019. Soil Photosynthetic Microbial Communities Mediate Aggregate Stability: Influence of Cropping Systems and Herbicide Use in an Agricultural Soil. *Frontiers in Microbiology*, 10, 1319.
- CSARDI, G. & NEPUSZ, T. 2005. *The Igraph Software Package for Complex Network Research*.
- CSOTONYI, J. T., SWIDERSKI, J., STACKEBRANDT, E. & YURKOV, V. A New Extreme Environment for Aerobic Anoxygenic Phototrophs: Biological Soil Crusts. In: HALLENBECK, P. C., ed. *Recent Advances in Phototrophic Prokaryotes*, 2010// 2010 New York, NY. Springer New York, 3-14.
- DALY, K. R., MOONEY, S. J., BENNETT, M. J., CROUT, N. M. J., ROOSE, T. & TRACY, S. R. 2015. Assessing the influence of the rhizosphere on soil hydraulic properties using X-ray computed tomography and numerical modelling. *Journal of experimental botany*, 66, 2305-2314.
- DANCZAK, R. E., JOHNSTON, M. D., KENAH, C., SLATTERY, M. & WILKINS, M. J. 2018. Microbial Community Cohesion Mediates Community Turnover in Unperturbed Aquifers. *mSystems*, 3, e00066-18.
- DAVIDSON, E. A. & JANSSENS, I. A. 2006. Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, 440, 165-173.
- DAVIES, L. O., BRAMKE, I., FRANCE, E., MARSHALL, S., OLIVER, R., NICHOLS, C., SCHÄFER, H. & BENDING, G. D. 2013a. Non-UV Light Influences the Degradation Rate of Crop Protection Products. *Environmental Science & Technology*, 47, 8229-8237.
- DAVIES, L. O., SCHÄFER, H., MARSHALL, S., BRAMKE, I., OLIVER, R. G. & BENDING, G. D. 2013b. Light Structures Phototroph, Bacterial and Fungal Communities at the Soil Surface. *PLOS ONE*, 8, e69048.
- DE GRUYTER, J., WOUTENBERG, J. H. C., AVESKAMP, M. M., VERKLEY, G. J. M., GROENEWALD, J. Z. & CROUS, P. W. 2013. Redisposition of phoma-like anamorphs in Pleosporales. *Studies in mycology*, 75, 1-36.
- DE VRIES, F. T., GRIFFITHS, R. I., BAILEY, M., CRAIG, H., GIRLANDA, M., GWEON, H. S., HALLIN, S., KAISERMANN, A., KEITH, A. M., KRETZSCHMAR, M., LEMANCEAU, P., LUMINI, E., MASON, K. E., OLIVER, A., OSTLE, N., PROSSER, J. I., THION, C., THOMSON, B. & BARDGETT, R. D. 2018. Soil bacterial networks are less stable under drought than fungal networks. *Nature Communications*, 9, 3033.

- DEAKIN, G., TILSTON, E. L., BENNETT, J., PASSEY, T., HARRISON, N., FERNÁNDEZ-FERNÁNDEZ, F. & XU, X. 2018. Spatial structuring of soil microbial communities in commercial apple orchards. *Applied Soil Ecology*, 130, 1-12.
- DECHARME, B., BRUN, E., BOONE, A., DELIRE, C., LE MOIGNE, P. & MORIN, S. 2016. Impacts of snow and organic soils parameterization on northern Eurasian soil temperature profiles simulated by the ISBA land surface model. *The Cryosphere*, 10, 853-877.
- DEGRUNE, F., THEODORAKOPOULOS, N., COLINET, G., HIEL, M.-P., BODSON, B., TAMINIAU, B., DAUBE, G., VANDENBOL, M. & HARTMANN, M. 2017. Temporal Dynamics of Soil Microbial Communities below the Seedbed under Two Contrasting Tillage Regimes. *Frontiers in Microbiology*, 8, 1127.
- DELGADO-BAQUERIZO, M., MAESTRE, F. T., ELDRIDGE, D. J. & SINGH, B. K. 2016. Microsite Differentiation Drives the Abundance of Soil Ammonia Oxidizing Bacteria along Aridity Gradients. *Frontiers in Microbiology*, 7, 505.
- DENHARTOG, C. 1985. Süßwasserflora von mitteleuropa 9 chlorophyta I phytomonadinaH. Ettl, Gustav Fischer Verlag, Stuttgart, New York, 1983, I-XIV + 807 pp., price DM 180. ISBN 3-347-30408-9. *Aquatic Botany*.
- DESIRÒ, A., RIMINGTON, W., JACOB, A., VANDE POL, N., E. SMITH, M., TRAPPE, J., BIDARTONDO, M. & BONITO, G. 2017. *Multigene phylogeny of Endogonales, an early diverging lineage of fungi associated with plants*.
- DETTWEILER-ROBINSON, E., NUANEZ, M. & LITVAK, M. E. 2018. Biocrust contribution to ecosystem carbon fluxes varies along an elevational gradient. *Ecosphere*, 9, e02315.
- DOJANI, S., BÜDEL, B., DEUTSCHEWITZ, K. & WEBER, B. 2011. Rapid succession of Biological Soil Crusts after experimental disturbance in the Succulent Karoo, South Africa. *Applied Soil Ecology*, 48, 263-269.
- DOJANI, S., KAUFF, F., WEBER, B. & BÜDEL, B. 2014. Genotypic and Phenotypic Diversity of Cyanobacteria in Biological Soil Crusts of the Succulent Karoo and Nama Karoo of Southern Africa. *Microbial Ecology*, 67, 286-301.
- DOUBE, M., KŁOSOWSKI, M. M., ARGANDA-CARRERAS, I., CORDELIÈRES, F. P., DOUGHERTY, R. P., JACKSON, J. S., SCHMID, B., HUTCHINSON, J. R. & SHEFELBINE, S. J. 2010. BoneJ: Free and extensible bone image analysis in ImageJ. *Bone*, 47, 1076-1079.
- DRAY, S., LEGENDRE, P. & PERES-NETO, P. R. 2006. Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). *Ecological Modelling*, 196, 483-493.
- DU PLESSIS, A., BROECKHOVEN, C., GUELPA, A. & LE ROUX, S. G. 2017. Laboratory x-ray micro-computed tomography: a user guideline for biological samples. *GigaScience*, 6, 1-11.

- DUMACK, K., KOLLER, R., WEBER, B. & BONKOWSKI, M. 2016. Estimated abundance and diversity of heterotrophic protists in South African biocrusts. *South African journal of science*, 112, 1-5.
- DURHAM, R. A., DOHERTY, K. D., ANTONINKA, A. J., RAMSEY, P. W. & BOWKER, M. A. 2018. Insolation and disturbance history drive biocrust biodiversity in Western Montana rangelands. *Plant and Soil*, 430, 151-169.
- DWEVEDI, A., KUMAR, P., KUMAR, P., KUMAR, Y., SHARMA, Y. K. & KAYASTHA, A. M. 2017. 15 - Soil sensors: detailed insight into research updates, significance, and future prospects. In: GRUMEZESCU, A. M. (ed.) *New Pesticides and Soil Sensors*. Academic Press.
- ECKERT, K. A. & KUNKEL, T. A. 1991. DNA polymerase fidelity and the polymerase chain reaction. *Genome Research*, 1, 17-24.
- EDGAR, R. C. 2010. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*, 26, 2460-2461.
- EIVAZI, F. & TABATABAI, M. A. 1977. Phosphatases in soils. *Soil Biology and Biochemistry*, 9, 167-172.
- ELBERT, W., WEBER, B., BURROWS, S., STEINKAMP, J., BUDEL, B., ANDREAE, M. O. & POSCHL, U. 2012. Contribution of cryptogamic covers to the global cycles of carbon and nitrogen. *Nature Geosci*, 5, 459-462.
- EMMETT, B. A., FROGBROOK, Z., CHAMBERLAIN, P. M., GRIFFITHS, R. I., PICKUP, R., POSKITT, J., REYNOLDS, B., ROWE, E., ROWLAND, P., SPURGEON, D., WILSON, J. & WOOD, C. M. 2008. Countryside Survey Technical Report No.03/07. .
- ERINGEN, A. C. 1994. A continuum theory of swelling porous elastic soils. *International Journal of Engineering Science*, 32, 1337-1349.
- EVANS, M. C., BUCHANAN, B. B. & ARNON, D. I. 1966. A new ferredoxin-dependent carbon reduction cycle in a photosynthetic bacterium. *Proceedings of the National Academy of Sciences of the United States of America*, 55, 928-934.
- FAIST, A. M., HERRICK, J. E., BELNAP, J., VAN ZEE, J. W. & BARGER, N. N. 2017. Biological soil crust and disturbance controls on surface hydrology in a semi-arid ecosystem. *Ecosphere*, 8, e01691.
- FAN, Z., NEFF, J. C., WALDROP, M. P., BALLANTYNE, A. P. & TURETSKY, M. R. 2014. Transport of oxygen in soil pore-water systems: implications for modeling emissions of carbon dioxide and methane from peatlands. *Biogeochemistry*, 121, 455-470.
- FAO 2015. Healthy soils are the basis for healthy food production. *The importance of soil organic matter: key to drought-resistant soil and sustained food production*. FAO.
- FAO. 2019. *Arable land (% of land area)* [Online]. The world bank. [Accessed 12 June 2019 2019].
- FEDERHEN, S. 2012. The NCBI Taxonomy database. *Nucleic acids research*, 40, D136-D143.

- FEI, L., JIANLI, L. & JUAN, H. 2008. Prediction of near saturated soil hydraulic properties by using CT images and network model. *Transactions of the Chinese Society of Agricultural Engineering*.
- FIERER, N. & JACKSON, R. B. 2006. The diversity and biogeography of soil bacterial communities. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 626.
- FIERER, N., SCHIMEL, J. P. & HOLDEN, P. A. 2003. Variations in microbial community composition through two soil depth profiles. *Soil Biology and Biochemistry*, 35, 167-176.
- FINN, R. D., BATEMAN, A., CLEMENTS, J., COGGILL, P., EBERHARDT, R. Y., EDDY, S. R., HEGER, A., HETHERINGTON, K., HOLM, L., MISTRY, J., SONNHAMMER, E. L. L., TATE, J. & PUNTA, M. 2014. Pfam: the protein families database. *Nucleic acids research*, 42, D222-D230.
- IORE-DONNO, A. M., RIXEN, C., RIPPIN, M., GLASER, K., SAMOLOV, E., KARSTEN, U., BECKER, B. & BONKOWSKI, M. 2018. New barcoded primers for efficient retrieval of cercozoan sequences in high-throughput environmental diversity surveys, with emphasis on worldwide biological soil crusts. *Molecular Ecology Resources*, 18, 229-239.
- FISCHER, T., VESTE, M., SCHAAF, W., DÜMIG, A., KÖGEL-KNABNER, I., WIEHE, W., BENS, O. & HÜTTL, R. F. 2010. Initial pedogenesis in a topsoil crust 3 years after construction of an artificial catchment in Brandenburg, NE Germany. *Biogeochemistry*, 101, 165-176.
- FOWLER, D., COYLE, M., SKIBA, U., SUTTON, M. A., CAPE, J. N., REIS, S., SHEPPARD, L. J., JENKINS, A., GRIZZETTI, B., GALLOWAY, J. N., VITOUSEK, P., LEACH, A., BOUWMAN, A. F., BUTTERBACH-BAHL, K., DENTENER, F., STEVENSON, D., AMANN, M. & VOSS, M. 2013. The global nitrogen cycle in the twenty-first century. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368.
- FOX, G., WISOTZKEY, J. D. & JURTSCHUK, J. P. 1992. How Close Is Close: 16S rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *International journal of systematic bacteriology*, 42, 166-70.
- FREY, B., BÜHLER, L., SCHMUTZ, S., ZUMSTEG, A. & FURRER, G. 2013. Molecular characterization of phototrophic microorganisms in the forefield of a receding glacier in the Swiss Alps. *Environmental Research Letters*, 8, 015033.
- FRIEDMAN, J. & ALM, E. J. 2012. Inferring Correlation Networks from Genomic Survey Data. *PLOS Computational Biology*, 8, e1002687.
- FRINDTE, K., PAPE, R., WERNER, K., LÖFFLER, J. & KNIEF, C. 2019. Temperature and soil moisture control microbial community composition in an arctic-alpine ecosystem along elevational and micro-topographic gradients. *The ISME Journal*.
- FUHRMAN, J. A. 2009. Microbial community structure and its functional implications. *Nature*, 459, 193.

- FURTWÄNGLER, K., TARASOV, V., WENDE, A., SCHWARZ, C. & OESTERHELT, D. 2010. Regulation of phosphate uptake via Pst transporters in *Halobacterium salinarum* R1. *Molecular Microbiology*, 76, 378-392.
- GALDOS, M. V., PIRES, L. F., COOPER, H. V., CALONEGO, J. C., ROSOLEM, C. A. & MOONEY, S. J. 2019. Assessing the long-term effects of zero-tillage on the macroporosity of Brazilian soils using X-ray Computed Tomography. *Geoderma*, 337, 1126-1135.
- GALPERIN, M. Y., MAKAROVA, K. S., WOLF, Y. I. & KOONIN, E. V. 2014. Expanded microbial genome coverage and improved protein family annotation in the COG database. *Nucleic Acids Research*, 43, D261-D269.
- GAO, Y., DANG, X., YU, Y., LI, Y., LIU, Y. & WANG, J. 2016. Effects of Tillage Methods on Soil Carbon and Wind Erosion. *Land Degradation & Development*, 27, 583-591.
- GARBOUT, A., STURROCK, C. J., ARMENISE, E., AHN, S., SIMMONS, R. W., DOERR, S., RITZ, K. & MOONEY, S. J. 2018. TopCap: A Tool to Quantify Soil Surface Topology and Subsurface Structure. *Vadose Zone Journal*, 17.
- GARCIA-PICHEL, F., BELNAP, J., NEUER, S. & SCHANZ, F. 2003. Estimates of global cyanobacterial biomass and its distribution. *Algological Studies*, 109, 213-227.
- GARCIA-PICHEL, F., LOZA, V., MARUSENKO, Y., MATEO, P. & POTRAFKA, R. M. 2013. Temperature Drives the Continental-Scale Distribution of Key Microbes in Topsoil Communities. *Science*, 340, 1574.
- GARCIA-PICHEL, F. & WOJCIECHOWSKI, M. F. 2009. The Evolution of a Capacity to Build Supra-Cellular Ropes Enabled Filamentous Cyanobacteria to Colonize Highly Erodible Substrates. *PLOS ONE*, 4, e7801.
- GEISEN, S., MITCHELL, E. A. D., ADL, S., BONKOWSKI, M., DUNTHORN, M., EKELUND, F., FERNÁNDEZ, L. D., JOUSSET, A., KRASHEVSKA, V., SINGER, D., SPIEGEL, F. W., WALOCHNIK, J. & LARA, E. 2018. Soil protists: a fertile frontier in soil biology research. *FEMS Microbiology Reviews*, 42, 293-323.
- GEORGE, P. B. L., LALLIAS, D., CREER, S., SEATON, F. M., KENNY, J. G., ECCLES, R. M., GRIFFITHS, R. I., LEBRON, I., EMMETT, B. A., ROBINSON, D. A. & JONES, D. L. 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, 10, 1107.
- GIBBONS, S. M., CAPORASO, J. G., PIRRUNG, M., FIELD, D., KNIGHT, R. & GILBERT, J. A. 2013. Evidence for a persistent microbial seed bank throughout the global ocean. *Proceedings of the National Academy of Sciences*, 110, 4651.
- GILBERT, J. A. & CORBIN, J. D. 2019. Biological soil crusts inhibit seed germination in a temperate pine barren ecosystem. *PLOS ONE*, 14, e0212466.
- GLASER, K., BAUMANN, K., LEINWEBER, P., MIKHAILYUK, T. & KARSTEN, U. 2017. Algal diversity of temperate biological soil crusts depends on land use intensity and affects phosphorus biogeochemical cycling. *Biogeosciences Discussions*, 1-24.

- GLENN, T. C., PIERSON, T. W., BAYONA-VÁSQUEZ, N. J., KIERAN, T. J., HOFFBERG, S. L., THOMAS, J. C., LEFEVER, D. E., FINGER, J. W., GAO, B., BIAN, X., LOUHA, S., KOLLI, R. T., BENTLEY, K., RUSHMORE, J., WONG, K., SHAW, T. I., ROTHROCK, M. J., MCKEE, A. M., GUO, T. L., MAURICIO, R., MOLINA, M., CUMMINGS, B. S., LASH, L. H., LU, K., GILBERT, G. S., HUBBELL, S. P. & FAIRCLOTH, B. C. 2019. Adapterama II: Universal amplicon sequencing on Illumina platforms (TaggiMatrix). *bioRxiv*, 619544.
- GOBERNA, M., PASCUAL, J. A., GARCÍA, C. & SÁNCHEZ, J. 2007. Do plant clumps constitute microbial hotspots in semiarid Mediterranean patchy landscapes? *Soil Biology and Biochemistry*, 39, 1047-1054.
- GOTELLI, N. & COLWELL, R. 2001. Quantifying biodiversity: Procedures and pitfalls in the measurement and comparison of species richness. *Ecology Letters*, 4, 379-391.
- GRAVES, A. R., MORRIS, J., DEEKS, L. K., RICKSON, R. J., KIBBLEWHITE, M. G., HARRIS, J. A., FAREWELL, T. S. & TRUCKLE, I. 2015. The total costs of soil degradation in England and Wales. *Ecological Economics*, 119, 399-413.
- GREVERS, M. C. J., JONG, E. D. & ST. ARNAUD, R. J. 1989. THE CHARACTERIZATION OF SOIL MACROPOROSITY WITH CT SCANNING. *Canadian Journal of Soil Science*, 69, 629-637.
- GUNDLAPALLY, S. R. & GARCIA-PICHEL, F. 2006. The Community and Phylogenetic Diversity of Biological Soil Crusts in the Colorado Plateau Studied by Molecular Fingerprinting and Intensive Cultivation. *Microbial Ecology*, 52, 345-357.
- GUPTA, R. K., ABROL, I. P., FINKL, C. W., KIRKHAM, M. B., ARBESTAIN, M. C., MACÍAS, F., CHESWORTH, W., GERMIDA, J. J., LOEPPERT, R. H., COOK, M. G., SCHWAB, G. O., KONSTANKIEWICZ, K., PYTKA, J., OERTLI, J. J., SINGER, A., EDMONDS, W. J., FENG, Y., FELDMAN, S. B., SHANG, C., ZELAZNY, L. W., FORD, P. W. & CLOTHIER, B. E. 2008. Soil Pores. In: CHESWORTH, W. (ed.) *Encyclopedia of Soil Science*. Dordrecht: Springer Netherlands.
- GYPSER, S., VESTE, M., FISCHER, T. & LANGE, P. 2015. Formation of soil lichen crusts at reclaimed post-mining sites, Lower Lusatia, North-east Germany. *Graphis Scripta*, 27, 3-14.
- HADDAWAY, N. R., HEDLUND, K., JACKSON, L. E., KÄTTERER, T., LUGATO, E., THOMSEN, I. K., JØRGENSEN, H. B. & ISBERG, P.-E. 2017. How does tillage intensity affect soil organic carbon? A systematic review. *Environmental Evidence*, 6, 30.
- HAJNOS, M., KORSUNSKAIA, L. & PACHEPSKY, Y. 2000. Soil pore surface properties in managed grasslands. *Soil and Tillage Research*, 55, 63-70.
- HAMBLETON, S. & SIGLER, L. 2005. Meliniomyces, a new anamorph genus for root-associated fungi with phylogenetic affinities to Rhizoscyphus ericae (≡ Hymenoscyphus ericae), Leotiomycetes. *Studies in Mycology*, 53, 1-27.
- HANDELSMAN, J. 2004. Metagenomics: Application of Genomics to Uncultured Microorganisms. *Microbiology and Molecular Biology Reviews*, 68, 669.

- HANSON, C. A., FUHRMAN, J. A., HORNER-DEVINE, M. C. & MARTINY, J. B. H. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. *Nature Reviews Microbiology*, 10, 497.
- HANSSON, L.-A. 1988. Chlorophyll a determination of periphyton on sediments: identification of problems and recommendation of method. *Freshwater Biology*, 20, 347-352.
- HARPER, K. T. & MARBLE, J. R. 1988. A role for nonvascular plants in management of arid and semiarid rangelands. In: TUELLER, P. T. (ed.) *Vegetation science applications for rangeland analysis and management*. Dordrecht: Springer Netherlands.
- HASSINK, J., BOUWMAN, L. A., ZWART, K. B., BLOEM, J. & BRUSSAARD, L. 1993. Relationships between soil texture, physical protection of organic matter, soil biota, and c and n mineralization in grassland soils. *Geoderma*, 57, 105-128.
- HAWKES, C. V. & FLECHTNER, V. R. 2002. Biological Soil Crusts in a Xeric Florida Shrubland: Composition, Abundance, and Spatial Heterogeneity of Crusts with Different Disturbance Histories. *Microbial Ecology*, 43, 1-12.
- HAWKINS, H.-J., JOHANSEN, A. & GEORGE, E. 2000. Uptake and transport of organic and inorganic nitrogen by arbuscular mycorrhizal fungi. *Plant and Soil*, 226, 275-285.
- HEINDEL, R. C., GOVERNALI, F. C., SPICKARD, A. M. & VIRGINIA, R. A. 2019. The Role of Biological Soil Crusts in Nitrogen Cycling and Soil Stabilization in Kangerlussuaq, West Greenland. *Ecosystems*, 22, 243-256.
- HELLIWELL, J. R., STURROCK, C. J., MAIRHOFER, S., CRAIGON, J., ASHTON, R. W., MILLER, A. J., WHALLEY, W. R. & MOONEY, S. J. 2017. The emergent rhizosphere: imaging the development of the porous architecture at the root-soil interface. *Scientific Reports*, 7, 14875.
- HENRISSAT, B. & DAVIES, G. J. 2000. Glycoside Hydrolases and Glycosyltransferases. Families, Modules, and Implications for Genomics. *Plant Physiology*, 124, 1515.
- HENSEN, D., SPERLING, D., TRÜPER, H. G., BRUNE, D. C. & DAHL, C. 2006. Thiosulphate oxidation in the phototrophic sulphur bacterium *Allochrocatium vinosum*. *Molecular Microbiology*, 62, 794-810.
- HERBOLD, C. W., LEE, C. K., MCDONALD, I. R. & CARY, S. C. 2014. Evidence of global-scale aeolian dispersal and endemism in isolated geothermal microbial communities of Antarctica. *Nature Communications*, 5, 3875.
- HILTNER, L. 1904. Über neuere Erfahrungen und Probleme auf dem Gebiet der Boden Bakteriologie und unter besonderer Berücksichtigung der Grundung und Broche. *Arbeit. Deut. Landw. Ges. Berlin*, 98, 59-78.
- HOPPERT, M., REIMER, R., KEMMLING, A., SCHRÖDER, A., GÜNZL, B. & HEINKEN, T. 2004. Structure and Reactivity of a Biological Soil Crust from a Xeric Sandy Soil in Central Europe. *Geomicrobiology Journal*, 21, 183-191.

- HORI, K., MARUYAMA, F., FUJISAWA, T., TOGASHI, T., YAMAMOTO, N., SEO, M., SATO, S., YAMADA, T., MORI, H., TAJIMA, N., MORIYAMA, T., IKEUCHI, M., WATANABE, M., WADA, H., KOBAYASHI, K., SAITO, M., MASUDA, T., SASAKI-SEKIMOTO, Y., MASHIGUCHI, K., AWAI, K., SHIMOJIMA, M., MASUDA, S., IWAI, M., NOBUSAWA, T., NARISE, T., KONDO, S., SAITO, H., SATO, R., MURAKAWA, M., IHARA, Y., OSHIMA-YAMADA, Y., OHTAKA, K., SATOH, M., SONOBE, K., ISHII, M., OHTANI, R., KANAMORI-SATO, M., HONOKI, R., MIYAZAKI, D., MOCHIZUKI, H., UMETSU, J., HIGASHI, K., SHIBATA, D., KAMIYA, Y., SATO, N., NAKAMURA, Y., TABATA, S., IDA, S., KUROKAWA, K. & OHTA, H. 2014. Klebsormidium flaccidum genome reveals primary factors for plant terrestrial adaptation. *Nature Communications*, 5, 3978.
- HUANG, L., ZHANG, Z.-S. & LI, X. 2014a. Carbon fixation and its influence factors of biological soil crusts in a revegetated area of the Tengger Desert, northern China. *Journal of Arid Land*, 6.
- HUANG, L.-M., JIA, X., SHAO, M.-A. & ZHANG, G.-L. 2017. Soil organic phosphorus transformation during ecosystem development: A review. *Plant and Soil*.
- HUANG, L.-N., TANG, F.-Z., SONG, Y.-S., WAN, C.-Y., WANG, S.-L., LIU, W.-Q. & SHU, W.-S. 2011. Biodiversity, abundance, and activity of nitrogen-fixing bacteria during primary succession on a copper mine tailings. *FEMS Microbiology Ecology*, 78, 439-450.
- HUANG, X.-F., CHAPARRO, J. M., REARDON, K. F., ZHANG, R., SHEN, Q. & VIVANCO, J. M. 2014b. Rhizosphere interactions: root exudates, microbes, and microbial communities. *Botany*, 92, 267-275.
- HUGENHOLTZ, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biology*, 3, reviews0003.1.
- HUGERTH, L. W. & ANDERSSON, A. F. 2017. Analysing Microbial Community Composition through Amplicon Sequencing: From Sampling to Hypothesis Testing. *Frontiers in Microbiology*, 8, 1561.
- HYATT, D., CHEN, G.-L., LOCASCIO, P. F., LAND, M. L., LARIMER, F. W. & HAUSER, L. J. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC bioinformatics*, 11, 119-119.
- IASSONOV, P., GEBRENEGUS, T. & TULLER, M. 2009. Segmentation of X-ray computed tomography images of porous materials: A crucial step for characterization and quantitative analysis of pore structures. *Water Resources Research*, 45.
- IMCHEN, M., KUMAVATH, R., BARH, D., AZEVEDO, V., GHOSH, P., VIANA, M. & WATTAM, A. R. 2017. Searching for signatures across microbial communities: Metagenomic analysis of soil samples from mangrove and other ecosystems. *Scientific Reports*, 7, 8859.
- JIMENEZ AGUILAR, A., HUBER-SANNWALD, E., BELNAP, J., SMART, D. R. & ARREDONDO MORENO, J. T. 2009. Biological soil crusts exhibit a dynamic response to seasonal rain and release from grazing with implications for soil stability. *Journal of Arid Environments*, 73, 1158-1169.

- JOERGENSEN, R. G. 1996. The fumigation-extraction method to estimate soil microbial biomass: Calibration of the kEC value. *Soil Biology and Biochemistry*, 28, 25-31.
- JOHNSON, S., R KUSKE, C., CARNEY, T., C HOUSMAN, D., GALLEGOS-GRAVES, L. & BELNAP, J. 2012. Increased temperature and altered summer precipitation have differential effects on biological soil crusts in a dryland ecosystem. *Global Change Biology*, 18, 2583-2593.
- JOHNSON, S. L., BUDINOFF, C. R., BELNAP, J. & GARCIA-PICHEL, F. 2005. Relevance of ammonium oxidation within biological soil crust communities. *Environmental Microbiology*, 7, 1-12.
- JONES, C. B., B.; ENGEL, R.; HORNECK, D.; OLSON-RUTZ, KATHRIN. 2013. Factors Affecting Nitrogen Fertilizer Volatilization. *Management to Minimize Nitrogen Fertilizer Volatilization*. Land resources Montana: Montana State University.
- JONES, D. L., HODGE, A. & KUZYAKOV, Y. 2004. Plant and mycorrhizal regulation of rhizodeposition. *New Phytologist*, 163, 459-480.
- JONGERIUS, A. 1957. *Morfologische onderzoekingen over de bodemstructuur*. Staatsdrukkerij.
- JÓNSSON, J. & DAVIDSDOTTIR, B. 2016. Classification and valuation of soil ecosystem services. *Agricultural Systems*, 145, 24-38.
- JU, F., XIA, Y., GUO, F., WANG, Z. & ZHANG, T. 2014. Taxonomic relatedness shapes bacterial assembly in activated sludge of globally distributed wastewater treatment plants. *Environmental Microbiology*, 16, 2421-2432.
- KALLMEYER, J., POCKALNY, R., ADHIKARI, R. R., SMITH, D. C. & D'HONDT, S. 2012. Global distribution of microbial abundance and biomass in subseafloor sediment. *Proceedings of the National Academy of Sciences*, 109, 16213.
- KANEHISA, M., GOTO, S., SATO, Y., FURUMICHI, M. & TANABE, M. 2011. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Research*, 40, D109-D114.
- KANEHISA, M., SATO, Y., KAWASHIMA, M., FURUMICHI, M. & TANABE, M. 2015. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research*, 44, D457-D462.
- KANG, D. D., FROULA, J., EGAN, R. & WANG, Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. *PeerJ*, 3, e1165.
- KARAOZ, U., COURADEAU, E., DA ROCHA, U. N., LIM, H.-C., NORTHEN, T., GARCIA-PICHEL, F. & BRODIE, E. L. 2018. Large Blooms of Bacillales (Firmicutes) Underlie the Response to Wetting of Cyanobacterial Biocrusts at Various Stages of Maturity. *mBio*, 9, e01366-16.
- KARSTEN, U. & HOLZINGER, A. 2014. Green algae in alpine biological soil crust communities: acclimation strategies against ultraviolet radiation and dehydration. *Biodiversity and conservation*, 23, 1845-1858.

- KETCHAM, R. A. & CARLSON, W. D. 2001. Acquisition, optimization and interpretation of X-ray computed tomographic imagery: applications to the geosciences. *Computers & Geosciences*, 27, 381-400.
- KIBBLEWHITE, M. G., RITZ, K. & SWIFT, M. J. 2008. Soil health in agricultural systems. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363, 685-701.
- KIDRON, G. J., YAIR, A., VONSHAK, A. & ABELIOVICH, A. 2003. Microbiotic crust control of runoff generation on sand dunes in the Negev Desert. *Water Resources Research*, 39, n/a-n/a.
- KIM, D., SONG, L., BREITWIESER, F. P. & SALZBERG, S. L. 2016. Centrifuge: rapid and sensitive classification of metagenomic sequences. *Genome research*, 26, 1721-1729.
- KIM, Y., LEE, G., AN, H. & YANG, J. E. 2015. Uncertainty assessment of soil erosion model using particle filtering. *Journal of Mountain Science*, 12, 828-840.
- KIRCHMANN, H. & GERZABEK, M. H. 1999. Relationship between soil organic matter and micropores in a long-term experiment at Ultuna, Sweden. *Journal of Plant Nutrition and Soil Science*, 162, 493-498.
- KLEEFELD, A., GYPSER, S., HERPPICH, W. B., BADER, G. & VESTE, M. 2018. Identification of spatial pattern of photosynthesis hotspots in moss- and lichen-dominated biological soil crusts by combining chlorophyll fluorescence imaging and multispectral BNDVI images. *Pedobiologia*, 68, 1-11.
- KNAPEN, A., POESEN, J., GALINDO-MORALES, P., BAETS, S. D. & PALS, A. 2007. Effects of microbiotic crusts under cropland in temperate environments on soil erodibility during concentrated flow. *Earth Surface Processes and Landforms*, 32, 1884-1901.
- KOCH, H., VAN KESSEL, M. A. H. J. & LÜCKER, S. 2019. Complete nitrification: insights into the ecophysiology of comammox Nitrospira. *Applied microbiology and biotechnology*, 103, 177-189.
- KÖLJALG, U., LARSSON, K.-H., ABARENKOV, K., NILSSON, R. H., ALEXANDER, I. J., EBERHARDT, U., ERLAND, S., HØILAND, K., KJØLLER, R., LARSSON, E., PENNANEN, T., SEN, R., TAYLOR, A. F. S., TEDERSOO, L. & VRÅLSTAD, T. 2005. UNITE: a database providing web-based methods for the molecular identification of ectomycorrhizal fungi. *New Phytologist*, 166, 1063-1068.
- KOMÁREK, J., KLING, H. & KOMÁRKOVÁ, J. 2003. 4 - FILAMENTOUS CYANOBACTERIA. In: WEHR, J. D. & SHEATH, R. G. (eds.) *Freshwater Algae of North America*. Burlington: Academic Press.
- KONONOVA, S. & NESMEYANOVA, M. A. 2002. Phosphonates and Their Degradation by Microorganisms. *Biochemistry. Biokhimiia*, 67, 184-95.
- KOONIN, E. V. 1997. Genome sequences: Genome sequence of a model prokaryote. *Current Biology*, 7, R656-R659.

- KOONIN, E. V., MUSHEGIAN, A. R., GALPERIN, M. Y. & WALKER, D. R. 1997. Comparison of archaeal and bacterial genomes: computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Molecular Microbiology*, 25, 619-637.
- KOPONEN, H. T., JAAKKOLA, T., KEINÄNEN-TOIVOLA, M. M., KAIPAINEN, S., TUOMAINEN, J., SERVOMAA, K. & MARTIKAINEN, P. J. 2006. Microbial communities, biomass, and activities in soils as affected by freeze thaw cycles. *Soil Biology and Biochemistry*, 38, 1861-1871.
- KOPONEN, H. T. & MARTIKAINEN, P. J. 2004. Soil water content and freezing temperature affect freeze–thaw related N₂O production in organic soil. *Nutrient Cycling in Agroecosystems*, 69, 213-219.
- KOZICH, J. J., WESTCOTT, S. L., BAXTER, N. T., HIGHLANDER, S. K. & SCHLOSS, P. D. 2013. Development of a Dual-Index Sequencing Strategy and Curation Pipeline for Analyzing Amplicon Sequence Data on the MiSeq Illumina Sequencing Platform. *Applied and Environmental Microbiology*, 79, 5112.
- KREHENWINKEL, H., WOLF, M., LIM, J. Y., ROMINGER, A. J., SIMISON, W. B. & GILLESPIE, R. G. 2017. Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7, 17668.
- KUSKE, C. R., YEAGER, C. M., JOHNSON, S., TICKNOR, L. O. & BELNAP, J. 2012. Response and resilience of soil biocrust bacterial communities to chronic physical disturbance in arid shrublands. *The ISME Journal*, 6, 886-897.
- LAJOS, B. 2008. Soil Science. In: BELÜL, D. E. A. T. P. K. (ed.).
- LAL, R. 2009. *Agricultural Sciences - Volume I*, EOLSS Publishers Company Limited.
- LALLEY, J. S. & VILES, H. A. 2008. Recovery of lichen-dominated soil crusts in a hyper-arid desert. *Biodiversity and Conservation*, 17, 1-20.
- LAMALFA, E. M. & RYLE, R. 2008. Differential Snowpack Accumulation and Water Dynamics in Aspen and Conifer Communities: Implications for Water Yield and Ecosystem Function. *Ecosystems*, 11, 569-581.
- LANGHANS, T. M., STORM, C. & SCHWABE, A. 2009. Community Assembly of Biological Soil Crusts of Different Successional Stages in a Temperate Sand Ecosystem, as Assessed by Direct Determination and Enrichment Techniques. *Microbial Ecology*, 58, 394-407.
- LANGILLE, M. G. I., ZANEVELD, J., CAPORASO, J. G., MCDONALD, D., KNIGHTS, D., REYES, J. A., CLEMENTE, J. C., BURKEPILE, D. E., VEGA THURBER, R. L., KNIGHT, R., BEIKO, R. G. & HUTTENHOWER, C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nature biotechnology*, 31, 814-821.
- LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9, 357-359.

- LAUBER, C. L., 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. *Applied and Environmental Microbiology*, 75, 5111.
- LAWRENCE, M. & JIANG, Y. 2017. Porosity, Pore Size Distribution, Micro-structure. In: AMZIANE, S. & COLLET, F. (eds.) *Bio-aggregates Based Building Materials : State-of-the-Art Report of the RILEM Technical Committee 236-BBM*. Dordrecht: Springer Netherlands.
- LAWSON, C. E., WU, S., BHATTACHARJEE, A. S., HAMILTON, J. J., MCMAHON, K. D., GOEL, R. & NOGUERA, D. R. 2017. Metabolic network analysis reveals microbial community interactions in anammox granules. *Nature Communications*, 8, 15416.
- LEE, H. J., JEONG, S. E., KIM, P. J., MADSEN, E. & JEON, C. O. 2015. High resolution depth distribution of Bacteria, Archaea, methanotrophs, and methanogens in the bulk and rhizosphere soils of a flooded rice paddy. *Frontiers in Microbiology*, 6, 639.
- LEE, K. C., ARCHER, S. D. J., BOYLE, R. H., LACAP-BUGLER, D. C., BELNAP, J. & POINTING, S. B. 2016. Niche Filtering of Bacteria in Soil and Rock Habitats of the Colorado Plateau Desert, Utah, USA. *Frontiers in Microbiology*, 7, 1489.
- LEE, Z. M.-P., BUSSEMA, C., 3RD & SCHMIDT, T. M. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic acids research*, 37, D489-D493.
- LEIBOLD, M. A., HOLYOAK, M., MOUQUET, N., AMARASEKARE, P., CHASE, J. M., HOOPEs, M. F., HOLT, R. D., SHURIN, J. B., LAW, R., TILMAN, D., LOREAU, M. & GONZALEZ, A. 2004. The metacommunity concept: a framework for multi-scale community ecology. *Ecology Letters*, 7, 601-613.
- LI, D., LIU, C.-M., LUO, R., SADAKANE, K. & LAM, T.-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31, 1674-1676.
- LI, H., HANDSAKER, B., WYSOKER, A., FENNELL, T., RUAN, J., HOMER, N., MARTH, G., ABECASIS, G., DURBIN, R. & GENOME PROJECT DATA PROCESSING, S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics (Oxford, England)*, 25, 2078-2079.
- LI, H., JIANG, D., WOLLENWEBER, B., DAI, T. & CAO, W. 2010. Effects of shading on morphology, physiology and grain yield of winter wheat. *European Journal of Agronomy*, 33, 267-275.
- LI, H., RAO, B., WANG, G., SHEN, S., LI, D., HU, C. & LIU, Y. 2014. Spatial heterogeneity of cyanobacteria-inoculated sand dunes significantly influences artificial biological soil crusts in the Hopq Desert (China). *Environmental Earth Sciences*, 71, 245-253.
- LI, X.-R., WANG, X.-P., LI, T. & ZHANG, J.-G. 2002. Microbiotic soil crust and its effect on vegetation and habitat on artificially stabilized desert dunes in Tengger Desert, North China. *Biology and Fertility of Soils*, 35, 147-154.

- LI, X. R., ZHANG, P., SU, Y. G. & JIA, R. L. 2012. Carbon fixation by biological soil crusts following revegetation of sand dunes in arid desert regions of China: A four-year field study. *CATENA*, 97, 119-126.
- LIANG, Y., TANG, K., WANG, Y., YUAN, B., TAN, F., FENG, F. & LIU, H. 2018. *Hymenobacter crusticola* sp. nov., isolated from biological soil crust.
- LIN, H., BOUMA, J., WILDING, L. P., RICHARDSON, J. L., KUTÍLEK, M. & NIELSEN, D. R. 2005. Advances in Hydropedology. *Advances in Agronomy*. Academic Press.
- LINHARDT, R. J., GALLIHER, P. M. & COONEY, C. L. 1987. Polysaccharide lyases. *Applied Biochemistry and Biotechnology*, 12, 135-176.
- LIPIEC, J., WÓJCIGA, A. & HORN, R. 2009. Hydraulic properties of soil aggregates as influenced by compaction. *Soil and Tillage Research*, 103, 170-177.
- LIU, J., CADE-MENUN, B. J., YANG, J., HU, Y., LIU, C. W., TREMBLAY, J., LAFORGE, K., SCHELLENBERG, M., HAMEL, C. & BAINARD, L. D. 2018. Long-Term Land Use Affects Phosphorus Speciation and the Composition of Phosphorus Cycling Genes in Agricultural Soils. *Frontiers in Microbiology*, 9, 1643.
- LIU, L., LIU, Y., HUI, R. & XIE, M. 2017a. Recovery of microbial community structure of biological soil crusts in successional stages of Shapotou desert revegetation, northwest China. *Soil Biology and Biochemistry*, 107, 125-128.
- LIU, L., LIU, Y., ZHANG, P., GUANG, S., HUI, R. & WANG, J. 2017b. *Development of bacterial communities in biological soil crusts along a revegetation chronosequence in the Tengger Desert, northwest China*.
- LOMBARD, V., GOLACONDA RAMULU, H., DRULA, E., COUTINHO, P. M. & HENRISSAT, B. 2013. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research*, 42, D490-D495.
- LORENZ, K. & LAL, R. 2005. The Depth Distribution of Soil Organic Carbon in Relation to Land Use and Management and the Potential of Carbon Sequestration in Subsoil Horizons. *Advances in Agronomy*. Academic Press.
- LORENZEN, C. J. 1967. DETERMINATION OF CHLOROPHYLL AND PHEOPIGMENTS: SPECTROPHOTOMETRIC EQUATIONS¹. *Limnology and Oceanography*, 12, 343-346.
- LOUCA, S., DOEBELI, M. & PARFREY, L. W. 2018. Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. *Microbiome*, 6, 41-41.
- LOVE, M. I., HUBER, W. & ANDERS, S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology*, 15, 550-550.
- LU, X., SEURADGE, B. J. & NEUFELD, J. D. 2016. Biogeography of soil Thaumarchaeota in relation to soil depth and land usage. *FEMS Microbiology Ecology*, 93.
- LÜCKER, S., WAGNER, M., MAIXNER, F., PELLETIER, E., KOCH, H., VACHERIE, B., RATTEI, T., DAMSTÉ, J. S. S., SPIECK, E., LE PASLIER, D. & DAIMS, H.

2010. A *Nitrospira* metagenome illuminates the physiology and evolution of globally important nitrite-oxidizing bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 13479-13484.
- LUNDBERG, D. S., YOURSTONE, S., MIECZKOWSKI, P., JONES, C. D. & DANGL, J. L. 2013. Practical innovations for high-throughput amplicon sequencing. *Nature Methods*, 10, 999.
- LUXMOORE 1981. Micro-, meso-, and macroporosity of soil. v. 45.
- LYNCH, M. D. J. & THORN, R. G. 2006. Diversity of Basidiomycetes in Michigan Agricultural Soils. *Applied and Environmental Microbiology*, 72, 7050.
- M. HART, M., ANTUNES, P. & ABBOTT, L. 2017. *Unknown risks to soil biodiversity from commercial fungal inoculants*.
- MAESTRE, F. T., BOWKER, M. A., CANTÓN, Y., CASTILLO-MONROY, A. P., CORTINA, J., ESCOLAR, C., ESCUDERO, A., LÁZARO, R. & MARTÍNEZ, I. 2011. Ecology and functional roles of biological soil crusts in semi-arid ecosystems of Spain. *Journal of arid environments*, 75, 1282-1291.
- MAHMOUDI, M., JALALI, A., VALIZADEH, M. & SKANDRI, I. 2015. *The effects of forward speed and depth of conservation tillage on soil bulk density*.
- MAIER, S., SCHMIDT, T. S. B., ZHENG, L., PEER, T., WAGNER, V. & GRUBE, M. 2014. Analyses of dryland biological soil crusts highlight lichens as an important regulator of microbial communities. *Biodiversity and Conservation*, 23, 1735-1755.
- MAIER, S., TAMM, A., WU, D., CAESAR, J., GRUBE, M. & WEBER, B. 2018. Photoautotrophic organisms control microbial abundance, diversity, and physiology in different types of biological soil crusts. *The ISME Journal*, 12, 1032-1046.
- MANDAL, S., VAN TREUREN, W., WHITE, R. A., EGGESBØ, M., KNIGHT, R. & PEDDADA, S. D. 2015. Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microbial ecology in health and disease*, 26, 27663-27663.
- MARGALEF, O., SARDANS, J., FERNÁNDEZ-MARTÍNEZ, M., MOLOWNY-HORAS, R., JANSSENS, I. A., CIAIS, P., GOLL, D., RICHTER, A., OBERSTEINER, M., ASENSIO, D. & PEÑUELAS, J. 2017. Global patterns of phosphatase activity in natural soils. *Scientific reports*, 7, 1337-1337.
- MARTÍNEZ, I., STEGEN, JAMES C., MALDONADO-GÓMEZ, MARIA X., EREN, A. M., SIBA, PETER M., GREENHILL, ANDREW R. & WALTER, J. 2015. The Gut Microbiota of Rural Papua New Guineans: Composition, Diversity Patterns, and Ecological Processes. *Cell Reports*, 11, 527-538.
- 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., MORIN, P. J., NAEEM, S., ØVREÅS, L., REYSENBACH, A.-L., SMITH, V. H. & STALEY, J. T. 2006. Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology*, 4, 102-112.

- MARUSENKO, Y., BATES, S. T., ANDERSON, I., JOHNSON, S. L., SOULE, T. & GARCIA-PICHEL, F. 2013. Ammonia-oxidizing archaea and bacteria are structured by geography in biological soil crusts across North American arid lands. *Ecological Processes*, 2, 9.
- MARUSENKO, Y., GARCIA-PICHEL, F. & HALL, S. J. 2014. Ammonia-oxidizing archaea respond positively to inorganic nitrogen addition in desert soils. *FEMS Microbiology Ecology*, 91, 1-11.
- MAZOR, G., KIDRON, G. J., VONSHAK, A. & ABELIOVICH, A. 1996. The role of cyanobacterial exopolysaccharides in structuring desert microbial crusts. *FEMS Microbiology Ecology*, 21, 121-130.
- MCDONALD, D., PRICE, M. N., GOODRICH, J., NAWROCKI, E. P., DESANTIS, T. Z., PROBST, A., ANDERSEN, G. L., KNIGHT, R. & HUGENHOLTZ, P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *The ISME journal*, 6, 610-618.
- MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLOS ONE*, 8, e61217.
- MCMURDIE, P. J. & HOLMES, S. 2014. Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLOS Computational Biology*, 10, e1003531.
- MEEK, M. H. & LARSON, W. A. 2019. The future is now: Amplicon sequencing and sequence capture usher in the conservation genomics era. *Molecular Ecology Resources*, 19, 795-803.
- MENDES, R., GARBEVA, P. & RAAIJMAKERS, J. M. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiology Reviews*, 37, 634-663.
- MENON, M., YUAN, Q., JIA, X., DOUGILL, A. J., HOON, S. R., THOMAS, A. D. & WILLIAMS, R. A. 2011. Assessment of physical and hydrological properties of biological soil crusts using X-ray microtomography and modeling. *Journal of Hydrology*, 397, 47-54.
- MIEDEMA, R. 1997. Applications of Micromorphology of Relevance to Agronomy. In: SPARKS, D. L. (ed.) *Advances in Agronomy*. Academic Press.
- MIEGROET, H. & COLE, D. 1984. *The Impact of Nitrification on Soil Acidification and Cation Leaching in a Red Alder Ecosystem*.
- MIKHAILYUK, T., HOLZINGER, A., MASSALSKI, A. & KARSTEN, U. 2014. Morphology and ultrastructure of Interfilum and Klebsormidium (Klebsormidiales, Streptophyta) with special reference to cell division and thallus formation. *European journal of phycology*, 49, 395-412.
- MIKHEENKO, A., SAVELIEV, V. & GUREVICH, A. 2015. MetaQUAST: evaluation of metagenome assemblies. *Bioinformatics*, 32, 1088-1090.
- MILTNER, A., KOPINKE, F.-D., KINDLER, R., SELESI, D., HARTMANN, A. & KAESTNER, M. 2005. *Non-phototrophic CO₂ fixation by soil microorganisms*.

- MOED, J. R. & HALLEGRAEFF, G. M. 1978. Some Problems in the Estimation of Chlorophyll-a and Phaeopigments from Pre- and Post-acidification Spectrophotometric Measurements. *Internationale Revue der gesamten Hydrobiologie und Hydrographie*, 63, 787-800.
- MOGUL, R., VAISHAMPAYAN, P., BASHIR, M., MCKAY, C. P., SCHUBERT, K., BORNACCORSI, R., GOMEZ, E., THARAYIL, S., PAYTON, G., CAPRA, J., ANDAYA, J., BACON, L., BARGOMA, E., BLACK, D., BOOS, K., BRANT, M., CHABOT, M., CHAU, D., CISNEROS, J., CHU, G., CURNUTT, J., DIMIZIO, J., ENGELBRECHT, C., GOTT, C., HARNOTO, R., HOVANESIAN, R., JOHNSON, S., LAVERGNE, B., MARTINEZ, G., MANS, P., MORALES, E., OEI, A., PELOW, G., PIAGET, R., PONCE, N., RENTERIA, E., RODRIGUEZ, V., RODRIGUEZ, J., SANTANDER, M., SARMIENTO, K., SCHEPPELMANN, A., SCHROTER, G., SEXTON, D., STEPHENSON, J., SYMER, K., RUSSO-TAIT, T., WEIGEL, B. & WILHELM, M. B. 2017. Microbial Community and Biochemical Dynamics of Biological Soil Crusts across a Gradient of Surface Coverage in the Central Mojave Desert. *Frontiers in Microbiology*, 8, 1974.
- MOREIRA, D. & LÓPEZ-GARCÍA, P. 2002. The molecular ecology of microbial eukaryotes unveils a hidden world. *Trends in Microbiology*, 10, 31-38.
- MORISON, M. O. & SHEATH, R. G. 1985. Responses to desiccation stress by *Klebsormidium rivulare* (Ulotrichales, Chlorophyta) from a Rhode Island stream. *Phycologia*, 24, 129-145.
- MORRIS, J. J., JOHNSON, Z. I., SZUL, M. J., KELLER, M. & ZINSER, E. R. 2011. Dependence of the Cyanobacterium *Prochlorococcus* on Hydrogen Peroxide Scavenging Microbes for Growth at the Ocean's Surface. *PLOS ONE*, 6, e16805.
- MORTON, J. T., SANDERS, J., QUINN, R. A., MCDONALD, D., GONZALEZ, A., VÁZQUEZ-BAEZA, Y., NAVAS-MOLINA, J. A., SONG, S. J., METCALF, J. L., HYDE, E. R., LLADSER, M., DORRESTEIN, P. C. & KNIGHT, R. 2017. Balance Trees Reveal Microbial Niche Differentiation. *mSystems*, 2, e00162-16.
- MUELLER, R., BELNAP, J. & KUSKE, C. 2015. Soil bacterial and fungal community responses to nitrogen addition across soil depth and microhabitat in an arid shrubland. *Frontiers in Microbiology*, 6, 891.
- MUNCH, VELTHOF & GERARD 2007. Chapter 21 - Denitrification and Agriculture. In: BOTHE, H., FERGUSON, S. J. & NEWTON, W. E. (eds.) *Biology of the Nitrogen Cycle*. Amsterdam: Elsevier.
- MUÑOZ-MARTÍN, M., BECERRA-ABSALÓN, I., PERONA, E., FERNÁNDEZ-VALBUENA, L., GARCIA-PICHEL, F. & MATEO, P. 2018. Cyanobacterial biocrust diversity in Mediterranean ecosystems along a latitudinal and climatic gradient. *New Phytologist*, 221.
- MUSTAMO, P., HYVÄRINEN, M., RONKANEN, A. K. & KLØVE, B. 2016. Physical properties of peat soils under different land use options. *Soil Use and Management*, 32, 400-410.
- MUYZER, G., DE WAAL, E. C. & UITTERLINDEN, A. G. 1993. Profiling of complex microbial populations by denaturing gradient gel electrophoresis analysis of

- polymerase chain reaction-amplified genes coding for 16S rRNA. *Applied and Environmental Microbiology*, 59, 695.
- NAGAO, M., MATSUI, K. & UEMURA, M. 2008. Klebsormidium flaccidum, a charophycean green alga, exhibits cold acclimation that is closely associated with compatible solute accumulation and ultrastructural changes. *Plant, Cell & Environment*, 31, 872-885.
- NAGY, L. G., KOCSUBÉ, S., CSANÁDI, Z., KOVÁCS, G. M., PETKOVITS, T., VÁGVÖLGYI, C. & PAPP, T. 2012. Re-Mind the Gap! Insertion – Deletion Data Reveal Neglected Phylogenetic Potential of the Nuclear Ribosomal Internal Transcribed Spacer (ITS) of Fungi. *PLOS ONE*, 7, e49794.
- NAGY, M. L., PÉREZ, A. & GARCIA-PICHEL, F. 2005. The prokaryotic diversity of biological soil crusts in the Sonoran Desert (Organ Pipe Cactus National Monument, AZ). *FEMS Microbiology Ecology*, 54, 233-245.
- NAWAZ, M. F., BOURRIÉ, G. & TROLARD, F. 2013. Soil compaction impact and modelling. A review. *Agronomy for Sustainable Development*, 33, 291-309.
- NEGASSA, W. C., GUBER, A. K., KRAVCHENKO, A. N., MARSH, T. L., HILDEBRANDT, B. & RIVERS, M. L. 2015. Properties of Soil Pore Space Regulate Pathways of Plant Residue Decomposition and Community Structure of Associated Bacteria. *PLOS ONE*, 10, e0123999.
- NEJIDAT, A., POTRAFKA, R. M. & ZAADY, E. 2016. Successional biocrust stages on dead shrub soil mounds after severe drought: Effect of micro-geomorphology on microbial community structure and ecosystem recovery. *Soil Biology and Biochemistry*, 103, 213-220.
- NICHOLLS, S. M., QUICK, J. C., TANG, S. & LOMAN, N. J. 2019. Ultra-deep, long-read nanopore sequencing of mock microbial community standards. *GigaScience*, 8.
- NIEDERBERGER, T. D., SOHM, J. A., GUNDERSON, T., TIRINDELLI, J., CAPONE, D. G., CARPENTER, E. J. & CARY, S. C. 2015. Carbon-Fixation Rates and Associated Microbial Communities Residing in Arid and Ephemeral Wet Antarctic Dry Valley Soils. *Frontiers in microbiology*, 6, 1347-1347.
- NIMMO, J. R. 2004. Porosity and pore-size distribution. *Encyclopedia of soils in the environment*, 3, 295-303.
- NORTCLIFF, S., HULPKE, H., BANNICK, C., TERYTZE, K., KNOOP, G., BREDEMEIER, M., SCHULTE-BISPING, H., AUERSWALD, K., LITZ, N., MAYER, R., STOY, A., ALEF, K., KERNDORFF, H., CRÖSSMANN, G., EIKMANN, T., FRANZIUS, V., GRIMSKY, D., MÖHLENBRUCH, N., STORM, P. C. & DWORSHAK, U. 2006. Soil, Definition, Function, and Utilization of Soil.
- O'MALLEY, M. A. 2008. 'Everything is everywhere: but the environment selects': ubiquitous distribution and ecological determinism in microbial biogeography. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 39, 314-325.
- OKSANEN, J., GUILLAUME, B., FRIENDLY, M., KINDT, R., LEGENDRE, P., MCGLINN, D., MINCHIN, P., O'HARA, R., SIMPSON, G., SOLYMOS, P.,

- STEVENS, M., SZOECS, E. & WAGNER, H. 2018. *Vegan: Community Ecology Package. R package version 2.5-3*. [Online]. Available: <https://CRAN.R-project.org/package=vegan> [Accessed].
- OTSU, N. A. 1978. *A Threshold Selection Method From Gray-Level Histograms*.
- PACE, N. R., STAHL, D. A., LANE, D. J. & OLSEN, G. J. 1986. The Analysis of Natural Microbial Populations by Ribosomal RNA Sequences. *In*: MARSHALL, K. C. (ed.) *Advances in Microbial Ecology*. Boston, MA: Springer US.
- PAGLIAI, M. & VIGNOZZI, N. 2006. Soil porosity as an indicator of soil health. *Annals of Arid Zone*, 45, 259-286.
- PAJARES, S., ESCALANTE, A. E., NOGUEZ, A. M., GARCÍA-OLIVA, F., MARTÍNEZ-PIEDRAGIL, C., CRAM, S. S., EGUIARTE, L. E. & SOUZA, V. 2016. Spatial heterogeneity of physicochemical properties explains differences in microbial composition in arid soils from Cuatro Ciénegas, Mexico. *PeerJ*, 4, e2459-e2459.
- PALMER, K. & HORN, M. A. 2012. Actinobacterial Nitrate Reducers and Proteobacterial Denitrifiers Are Abundant in N₂-O-Metabolizing Palsa Peat. *Applied and Environmental Microbiology*, 78, 5584.
- PAPKE, R. T., RAMSING, N. B., BATESON, M. M. & WARD, D. M. 2003. Geographical isolation in hot spring cyanobacteria. *Environmental Microbiology*, 5, 650-659.
- PARADIS, E., CLAUDE, J. & STRIMMER, K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20, 289-290.
- PARKS, D. H., RINKE, C., CHUVOCHINA, M., CHAUMEIL, P.-A., WOODCROFT, B. J., EVANS, P. N., HUGENHOLTZ, P. & TYSON, G. W. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nature Microbiology*, 2, 1533-1542.
- PENG, X. & BRUNS, M. A. 2019. Cyanobacterial Soil Surface Consortia Mediate N Cycle Processes in Agroecosystems. *Frontiers in Environmental Science*, 6, 156.
- PERES-NETO, P. R., LEGENDRE, P., DRAY, S. & BORCARD, D. 2006. VARIATION PARTITIONING OF SPECIES DATA MATRICES: ESTIMATION AND COMPARISON OF FRACTIONS. *Ecology*, 87, 2614-2625.
- PÉRIARD, Y., GUMIERE, S. J., LONG, B., ROUSSEAU, A. N. & CARON, J. 2016. Use of X-ray CT scan to characterize the evolution of the hydraulic properties of a soil under drainage conditions. *Geoderma*, 279, 22-30.
- PERRET, J., PRASHER, S., KANTZAS, A. & LANGFORD, C. 1999. *Three-Dimensional Quantification of Macropore Networks in Undisturbed Soil Cores*.
- PETERSEN, T. N., LUKJANCENKO, O., THOMSEN, M. C. F., MADDALENA SPEROTTO, M., LUND, O., MØLLER AARESTRUP, F. & SICHERITZ-PONTÉN, T. 2017. MGmapper: Reference based mapping and taxonomy annotation of metagenomics sequence reads. *PloS one*, 12, e0176469-e0176469.
- PETRIZ, B. A. & FRANCO, O. L. 2017. Metaproteomics as a Complementary Approach to Gut Microbiota in Health and Disease. *Frontiers in chemistry*, 5, 4-4.

- PEVZNER, P. A., TANG, H. & WATERMAN, M. S. 2001. An Eulerian path approach to DNA fragment assembly. *Proceedings of the National Academy of Sciences*, 98, 9748.
- PFEIFFER, F., GRÖBER, C., BLANK, M., HÄNDLER, K., BEYER, M., SCHULTZE, J. L. & MAYER, G. 2018. Systematic evaluation of error rates and causes in short samples in next-generation sequencing. *Scientific Reports*, 8, 10950.
- PIERRET, A., CAPOWIEZ, Y., BELZUNCES, L. & MORAN, C. J. 2002. 3D reconstruction and quantification of macropores using X-ray computed tomography and image analysis. *Geoderma*, 106, 247-271.
- PIETRASIAK, N., JOHANSEN, J. R., LADOUX, T. & GRAHAM, R. C. 2011. Comparison of Disturbance Impacts to and Spatial Distribution of Biological Soil Crusts in the Little San Bernardino Mountains of Joshua Tree National Park, California. *Western North American Naturalist*, 71, 539-552.
- POINTING, S. B. & BELNAP, J. 2012. Microbial colonization and controls in dryland systems. *Nat Rev Micro*, 10, 551-562.
- POZO, M. J. & AZCÓN-AGUILAR, C. 2007. Unraveling mycorrhiza-induced resistance. *Current Opinion in Plant Biology*, 10, 393-398.
- PROSSER, J. I., BOHANNAN, B. J. M., CURTIS, T. P., ELLIS, R. J., FIRESTONE, M. K., FRECKLETON, R. P., GREEN, J. L., GREEN, L. E., KILLHAM, K., LENNON, J. J., OSBORN, A. M., SOLAN, M., VAN DER GAST, C. J. & YOUNG, J. P. W. 2007. The role of ecological theory in microbial ecology. *Nature Reviews Microbiology*, 5, 384.
- PRUITT, K. D., TATUSOVA, T. & MAGLOTT, D. R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research*, 35, D61-D65.
- PUSHKAREVA, E., PESSI, I. S., WILMOTTE, A. & ELSTER, J. 2015. Cyanobacterial community composition in Arctic soil crusts at different stages of development. *FEMS Microbiology Ecology*, 91.
- QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLÖCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic acids research*, 41, D590-D596.
- QUINCE, C., WALKER, A. W., SIMPSON, J. T., LOMAN, N. J. & SEGATA, N. 2017. Shotgun metagenomics, from sampling to analysis. *Nature Biotechnology*, 35, 833.
- QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26, 841-842.
- RAAIJMAKERS, J. M., PAULITZ, T. C., STEINBERG, C., ALABOUVETTE, C. & MOËNNE-LOCCOZ, Y. 2009. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil*, 321, 341-361.
- RAANAN, H., FELDE, V. J. M. N. L., PETH, S., DRAHORAD, S., IONESCU, D., ESHKOL, G., TREVES, H., FELIX-HENNINGSSEN, P., BERKOWICZ, S. M.,

- KEREN, N., HORN, R., HAGEMANN, M. & KAPLAN, A. 2016. Three-dimensional structure and cyanobacterial activity within a desert biological soil crust. *Environmental Microbiology*, 18, 372-383.
- RAICH, J. W. & POTTER, C. S. 1995. Global patterns of carbon dioxide emissions from soils. *Global Biogeochemical Cycles*, 9, 23-36.
- RAICH, J. W. & SCHLESINGER, W. H. 1992. The global carbon dioxide flux in soil respiration and its relationship to vegetation and climate. *Tellus B*, 44, 81-99.
- RAJEEV, L., DA ROCHA, U. N., KLITGORD, N., LUNING, E. G., FORTNEY, J., AXEN, S. D., SHIH, P. M., BOUSKILL, N. J., BOWEN, B. P., KERFELD, C. A., GARCIA-PICHEL, F., BRODIE, E. L., NORTHERN, T. R. & MUKHOPADHYAY, A. 2013. Dynamic cyanobacterial response to hydration and dehydration in a desert biological soil crust. *The Isme Journal*, 7, 2178.
- RANDLE-BOGGIS, R. J., HELGASON, T., SAPP, M. & ASHTON, P. D. 2016. Evaluating techniques for metagenome annotation using simulated sequence data. *FEMS microbiology ecology*, 92, fiw095.
- RASTOGI, R. P., RICHA, KUMAR, A., TYAGI, M. B. & SINHA, R. P. 2010. Molecular mechanisms of ultraviolet radiation-induced DNA damage and repair. *Journal of nucleic acids*, 2010, 592980-592980.
- REDDY, G. S. N. & GARCIA-PICHEL, F. 2007. *Sphingomonas mucosissima* sp. nov. and *Sphingomonas desiccabilis* sp. nov., from biological soil crusts in the Colorado Plateau, USA. *International Journal of Systematic and Evolutionary Microbiology*, 57, 1028-1034.
- REGO, A., RAIIO, F., MARTINS, T. P., RIBEIRO, H., SOUSA, A. G. G., SÉNECA, J., BAPTISTA, M. S., LEE, C. K., CARY, S. C., RAMOS, V., CARVALHO, M. F., LEÃO, P. N. & MAGALHÃES, C. 2019. Actinobacteria and Cyanobacteria Diversity in Terrestrial Antarctic Microenvironments Evaluated by Culture-Dependent and Independent Methods. *Frontiers in Microbiology*, 10, 1018.
- REID, W., MOONEY, H., CROPPER, A., CAPISTRANO, D., CARPENTER, S. & CHOPRA, K. 2005. *Millennium Ecosystem Assessment. Ecosystems and human well-being: synthesis*.
- RENSBURG, L. D. V. 2010. Advances in soil physics: Application in irrigation and dryland crop production. *South African Journal of Plant and Soil*, 27, 9-18.
- REYNOLDS, R., BELNAP, J., REHEIS, M., LAMOTHE, P. & LUISZER, F. 2001. Aeolian dust in Colorado Plateau soils: Nutrient inputs and recent change in source. *Proceedings of the National Academy of Sciences*, 98, 7123.
- REZANEZHAD, F., PRICE, J. S., QUINTON, W. L., LENNARTZ, B., MILOJEVIC, T. & VAN CAPPELLEN, P. 2016. Structure of peat soils and implications for water storage, flow and solute transport: A review update for geochemists. *Chemical Geology*, 429, 75-84.
- RICHARD, G., COUSIN, I., SILLON, J. F., BRUAND, A. & GUÉRIF, J. 2001. Effect of compaction on the porosity of a silty soil: influence on unsaturated hydraulic properties. *European Journal of Soil Science*, 52, 49-58.

- RIEU, M. & SPOSITO, G. 1991. Fractal Fragmentation, Soil Porosity, and Soil Water Properties: I. Theory. *Soil Science Society of America Journal*, 55, 1231-1238.
- RIPPIN, M., LANGE, S., SAUSEN, N. & BECKER, B. 2018. Biodiversity of biological soil crusts from the Polar Regions revealed by metabarcoding. *FEMS Microbiology Ecology*, 94.
- RITCHIE, R. J. 2006. Consistent Sets of Spectrophotometric Chlorophyll Equations for Acetone, Methanol and Ethanol Solvents. *Photosynthesis Research*, 89, 27-41.
- RITZ, K. 2014. *Soil Biota & Soil Health: A Primer* [Online].
<https://soilandwater.org.uk/Resources/Documents/Karl%20Ritz%20-%20Soil%20biota%20and%20soil%20health%20-%20primer.pdf>: Soil and water.
 [Accessed 27.10.2019 2019].
- RODRIGUEZ-CABALLERO, E., BELNAP, J., BÜDEL, B., CRUTZEN, P. J., ANDREAE, M. O., PÖSCHL, U. & WEBER, B. 2018. Dryland photoautotrophic soil surface communities endangered by global change. *Nature Geoscience*, 11, 185-189.
- ROESCH, L. F. W., FULTHORPE, R. R., RIVA, A., CASELLA, G., HADWIN, A. K. M., KENT, A. D., DAROUB, S. H., CAMARGO, F. A. O., FARMERIE, W. G. & TRIPLETT, E. W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *The ISME journal*, 1, 283-290.
- ROOT, H. T. & DODSON, E. K. 2016. Pssst...pass the algae: succession in lichen soil crusts. *Frontiers in Ecology and the Environment*, 14, 451-452.
- ROSLEV, P., IVERSEN, N. & HENRIKSEN, K. 1997. Oxidation and Assimilation of Atmospheric Methane by Soil Methane Oxidizers. *Applied and environmental microbiology*, 63, 874-80.
- ROSSI, F. & DE PHILIPPIS, R. 2015. Role of cyanobacterial exopolysaccharides in phototrophic biofilms and in complex microbial mats. *Life (Basel, Switzerland)*, 5, 1218-1238.
- ROSSI, F., POTRAFKA, R. M., PICHEL, F. G. & DE PHILIPPIS, R. 2012. The role of the exopolysaccharides in enhancing hydraulic conductivity of biological soil crusts. *Soil Biology and Biochemistry*, 46, 33-40.
- RUAMPS, L. S., NUNAN, N. & CHENU, C. 2011. Microbial biogeography at the soil pore scale. *Soil Biology and Biochemistry*, 43, 280-286.
- RUPRECHT, U., BRUNAUER, G. & TÜRK, R. 2014. High photobiont diversity in the common European soil crust lichen *Psora decipiens*. *Biodiversity and Conservation*, 23, 1771-1785.
- SAARI, A., HEISKANEN, J. & MARTIKAINEN, P. J. 1998. Effect of the organic horizon on methane oxidation and uptake in soil of a boreal Scots pine forest. *FEMS Microbiology Ecology*, 26, 245-255.
- SAINI, G. R. 1966. Organic Matter as a Measure of Bulk Density of Soil. *Nature*, 210, 1295-1296.
- SANGWAN, N., XIA, F. & GILBERT, J. A. 2016. Recovering complete and draft population genomes from metagenome datasets. *Microbiome*, 4, 8.

- SANTOLINI, M. & BARABÁSI, A.-L. 2018. Predicting perturbation patterns from the topology of biological networks. *Proceedings of the National Academy of Sciences*, 115, E6375.
- SBONER, A., MU, X. J., GREENBAUM, D., AUERBACH, R. K. & GERSTEIN, M. B. 2011. The real cost of sequencing: higher than you think! *Genome Biology*, 12, 125.
- SCHAAP, M. G. & LEBRON, I. 2001. Using microscope observations of thin sections to estimate soil permeability with the Kozeny–Carman equation. *Journal of Hydrology*, 251, 186-201.
- SCHARLEMANN, J. P. W., TANNER, E. V. J., HIEDERER, R. & KAPOV, V. 2014. Global soil carbon: understanding and managing the largest terrestrial carbon pool. *Carbon Management*, 5, 81-91.
- SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J.-Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9, 676.
- SCHOONOVER, J. E. & CRIM, J. F. 2015. An Introduction to Soil Concepts and the Role of Soils in Watershed Management. *Journal of Contemporary Water Research & Education*, 154, 21-47.
- SCHRODER, J. L., ZHANG, H., GIRMA, K., RAUN, W. R., PENN, C. J. & PAYTON, M. E. 2011. Soil Acidification from Long-Term Use of Nitrogen Fertilizers on Winter Wheat. *Soil Science Society of America Journal*, 75, 957-964.
- SCHULZ, K., MIKHAILYUK, T., DRESSLE, M., LEINWEBER, P. & KARSTEN, U. 2016. Biological Soil Crusts from Coastal Dunes at the Baltic Sea: Cyanobacterial and Algal Biodiversity and Related Soil Properties. *Microbial Ecology*, 71, 178-193.
- SEARS, J. T. & PRITHIVIRAJ, B. Seeding of Large Areas with Biological Soil Crust Starter Culture Formulations: Using an Aircraft Disbursable Granulate to Increase Stability, Fertility and CO₂ Sequestration on a Landscape Scale. 2012 IEEE Green Technologies Conference, 19-20 April 2012. 1-3.
- SEGATA, N., BOERNIGEN, D., TICKLE, T. L., MORGAN, X. C., GARRETT, W. S. & HUTTENHOWER, C. 2013. Computational meta'omics for microbial community studies. *Molecular Systems Biology*, 9, 666.
- SHABARDINA, V., KISCHKA, T., MANSKE, F., GRUNDMANN, N., FRITH, M. C., SUZUKI, Y. & MAKALOWSKI, W. 2019. NanoPipe—a web server for nanopore MinION sequencing data analysis. *GigaScience*, 8.
- SHARMA, S., SZELE, Z., SCHILLING, R., MUNCH, J. C. & SCHLOTTER, M. 2006. Influence of Freeze-Thaw Stress on the Structure and Function of Microbial Communities and Denitrifying Populations in Soil. *Applied and Environmental Microbiology*, 72, 2148.
- SHI, J.-Y., YUAN, X.-F., LIN, H.-R., YANG, Y.-Q. & LI, Z.-Y. 2011. Differences in soil properties and bacterial communities between the rhizosphere and bulk soil and

among different production areas of the medicinal plant *Fritillaria thunbergii*. *International journal of molecular sciences*, 12, 3770-3785.

- SHI, S., NUCCIO, E., HERMAN, D. J., RIJKERS, R., ESTERA, K., LI, J., DA ROCHA, U. N., HE, Z., PETT-RIDGE, J., BRODIE, E. L., ZHOU, J. & FIRESTONE, M. 2015. Successional Trajectories of Rhizosphere Bacterial Communities over Consecutive Seasons. *mBio*, 6, e00746-15.
- SILVERSIDE, A. 2019. *Trebouxia, puymaly* [Online]. [Accessed 2019].
- SIX, J., ELLIOTT, E. T., PAUSTIAN, K. & DORAN, J. W. 1998. Aggregation and Soil Organic Matter Accumulation in Cultivated and Native Grassland Soils. *Soil Science Society of America Journal*, 62.
- SMITH, A. P., BOND-LAMBERTY, B., BENSCOTER, B. W., TFAILY, M. M., HINKLE, C. R., LIU, C. & BAILEY, V. L. 2017. Shifts in pore connectivity from precipitation versus groundwater rewetting increases soil carbon loss after drought. *Nature Communications*, 8, 1335.
- SMITH, C. R., BLAIR, P. L., BOYD, C., CODY, B., HAZEL, A., HEDRICK, A., KATHURIA, H., KHURANA, P., KRAMER, B., MUTERSPAW, K., PECK, C., SELLS, E., SKINNER, J., TEGELER, C. & WOLFE, Z. 2016. Microbial community responses to soil tillage and crop rotation in a corn/soybean agroecosystem. *Ecology and evolution*, 6, 8075-8084.
- SMUCKER, A., J. PARK, E., DÖRNER, J. & HORN, R. 2007. Soil Micropore Development and Contributions to Soluble Carbon Transport within Macroaggregates. *Vadose Zone Journal*, 6, 282-290.
- SONG, G., LI, X. & HUI, R. 2017. Biological soil crusts determine the germination and growth of two exotic plants. *Ecology and Evolution*, 7, 9441-9450.
- SONNHAMMER, E. L., EDDY, S. R., BIRNEY, E., BATEMAN, A. & DURBIN, R. 1998. Pfam: multiple sequence alignments and HMM-profiles of protein domains. *Nucleic acids research*, 26, 320-322.
- SOULE, T., ANDERSON, I. J., JOHNSON, S. L., BATES, S. T. & GARCIA-PICHEL, F. 2009. Archaeal populations in biological soil crusts from arid lands in North America. *Soil Biology and Biochemistry*, 41, 2069-2074.
- SOUTY, N., STENGEL, P., RODE, C. & TUTTOBENE, R. 1992. A mechanistic study of maize emergence through superficial crusts. *Soil and Tillage Research*, 23, 125-140.
- SPRIBILLE, T., RESL, P., AHTI, T., PÉREZ-ORTEGA, S., TØNSBERG, T., MAYRHOFER, H. & LUMBSCH, H. T. 2014. Molecular systematics of the wood-inhabiting, lichen-forming genus *Xylographa* (Baeomycetales, Ostropomycetidae) with eight new species. *Acta Universitatis Upsaliensis. Symbolae botanicae Upsalienses : Arbeten fran Botaniska institutionerna i Uppsala*, 37, 1-87.
- STAHL, D. A., LANE, D. J., OLSEN, G. J. & PACE, N. R. 1985. Characterization of a Yellowstone hot spring microbial community by 5S rRNA sequences. *Applied and Environmental Microbiology*, 49, 1379.

- STAL, L. J. 2017. The effect of oxygen concentration and temperature on nitrogenase activity in the heterocystous cyanobacterium *Fischerella* sp. *Scientific Reports*, 7, 5402.
- STALEY, J. T. & KONOPKA, A. 1985. MEASUREMENT OF IN SITU ACTIVITIES OF NONPHOTOSYNTHETIC MICROORGANISMS IN AQUATIC AND TERRESTRIAL HABITATS. *Annual Review of Microbiology*, 39, 321-346.
- STATES, J. S., CHRISTENSEN, M. & KINTER, C. L. 2003. Soil Fungi as Components of Biological Soil Crusts. In: BELNAP, J. & LANGE, O. L. (eds.) *Biological Soil Crusts: Structure, Function, and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- STEGEN, J. C., LIN, X., FREDRICKSON, J. K., CHEN, X., KENNEDY, D. W., MURRAY, C. J., ROCKHOLD, M. L. & KONOPKA, A. 2013. Quantifying community assembly processes and identifying features that impose them. *The ISME journal*, 7, 2069-2079.
- STEGEN, J. C., LIN, X., FREDRICKSON, J. K. & KONOPKA, A. E. 2015. Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology*, 6, 370.
- STEGEN, J. C., LIN, X., KONOPKA, A. E. & FREDRICKSON, J. K. 2012. Stochastic and deterministic assembly processes in subsurface microbial communities. *The ISME Journal*, 6, 1653.
- STEVEN, B., BELNAP, J. & KUSKE, C. R. 2018. Chronic Physical Disturbance Substantially Alters the Response of Biological Soil Crusts to a Wetting Pulse, as Characterized by Metatranscriptomic Sequencing. *Frontiers in Microbiology*, 9, 2382.
- STEVEN, B., GALLEGOS-GRAVES, L., YEAGER, C., BELNAP, J. & R KUSKE, C. 2013a. Common and distinguishing features of the bacterial and fungal communities in biological soil crusts and shrub root zone soils. *Soil Biology and Biochemistry*, 69, 302-312.
- STEVEN, B., GALLEGOS-GRAVES, L. V., BELNAP, J. & KUSKE, C. R. 2013b. Dryland soil microbial communities display spatial biogeographic patterns associated with soil depth and soil parent material. *FEMS Microbiology Ecology*, 86, 101-113.
- STEVEN, B., HESSE, C., GALLEGOS-GRAVES, L., BELNAP, J. & R KUSKE, C. 2016. Fungal Diversity in Biological Soil Crusts of the Colorado Plateau.
- STEVEN, B., KUSKE, C. R., GALLEGOS-GRAVES, L. V., REED, S. C. & BELNAP, J. 2015. Climate Change and Physical Disturbance Manipulations Result in Distinct Biological Soil Crust Communities. *Applied and Environmental Microbiology*, 81, 7448.
- STEVEN, B., LIONARD, M., KUSKE, C. R. & VINCENT, W. F. 2013c. High bacterial diversity of biological soil crusts in water tracks over permafrost in the high arctic polar desert. *PloS one*, 8, e71489-e71489.

- STRAUSS, S. L., DAY, T. A. & GARCIA-PICHEL, F. 2012. Nitrogen cycling in desert biological soil crusts across biogeographic regions in the Southwestern United States. *Biogeochemistry*, 108, 171-182.
- STRICKLAND, M. S., LAUBER, C., FIERER, N. & BRADFORD, M. A. 2009. Testing the functional significance of microbial community composition. *Ecology*, 90, 441-451.
- SUN, S., LI, S., AVERA, B. N., STRAHM, B. D. & BADGLEY, B. D. 2017. Soil Bacterial and Fungal Communities Show Distinct Recovery Patterns during Forest Ecosystem Restoration. *Applied and Environmental Microbiology*, 83, e00966-17.
- SUTHERLAND, I. W. 1995. Polysaccharide lyases. *FEMS Microbiology Reviews*, 16, 323-347.
- SWENSON, T. L., KARAOZ, U., SWENSON, J. M., BOWEN, B. P. & NORTHERN, T. R. 2018. Linking soil biology and chemistry in biological soil crust using isolate exometabolomics. *Nature Communications*, 9, 19.
- SYERS, J. K. 1997. Managing soils for long-term productivity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352, 1011-1021.
- TAJNA, I. A., HECK, R. J. & ELLIOT, T. R. 2008. Application of X-ray computed tomography to soil science: A literature review. *Canadian Journal of Soil Science*, 88, 1-19.
- TAMM, A., CAESAR, J., KUNZ, N., COLESIE, C., REICHENBERGER, H. & WEBER, B. 2018. Ecophysiological properties of three biological soil crust types and their photoautotrophs from the Succulent Karoo, South Africa. *Plant and Soil*, 429, 127-146.
- TANAKA, T., KAWASAKI, K., DAIMON, S., KITAGAWA, W., YAMAMOTO, K., TAMAKI, H., TANAKA, M., NAKATSU, C. H. & KAMAGATA, Y. 2014. A Hidden Pitfall in the Preparation of Agar Media Undermines Microorganism Cultivability. *Applied and Environmental Microbiology*, 80, 7659.
- TATUSOV, R. L., GALPERIN, M. Y., NATALE, D. A. & KOONIN, E. V. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic acids research*, 28, 33-36.
- TEIXEIRA, W., CEDDIA, M., OTTONI, M. & KANGUSSU DONNAGEM, G. 2014. *Application of Soil Physics in Environmental Analyses: Measuring, Modelling and Data Integration*.
- THE HUMAN MICROBIOME PROJECT, C., HUTTENHOWER, C., GEVERS, D., KNIGHT, R., ABUBUCKER, S., BADGER, J. H., CHINWALLA, A. T., CREASY, H. H., EARL, A. M., FITZGERALD, M. G., FULTON, R. S., GIGLIO, M. G., HALLSWORTH-PEPIN, K., LOBOS, E. A., MADUPU, R., MAGRINI, V., MARTIN, J. C., MITREVA, M., MUZNY, D. M., SODERGREN, E. J., VERSALOVIC, J., WOLLAM, A. M., WORLEY, K. C., WORTMAN, J. R., YOUNG, S. K., ZENG, Q., AAGAARD, K. M., ABOLUDE, O. O., ALLEN-VERCOE, E., ALM, E. J., ALVARADO, L., ANDERSEN, G. L., ANDERSON, S., APPELBAUM, E., ARACHCHI, H. M., ARMITAGE, G., ARZE, C. A., AYVAZ, T., BAKER, C. C., BEGG, L., BELACHEW, T., BHONAGIRI, V., BIHAN, M., BLASER, M. J., BLOOM, T., BONAZZI, V., PAUL BROOKS, J., BUCK, G. A.,

- BUHAY, C. J., BUSAM, D. A., CAMPBELL, J. L., CANON, S. R., CANTAREL, B. L., CHAIN, P. S. G., CHEN, I. M. A., CHEN, L., CHHIBBA, S., CHU, K., CIULLA, D. M., CLEMENTE, J. C., CLIFTON, S. W., CONLAN, S., CRABTREE, J., CUTTING, M. A., DAVIDOVICS, N. J., DAVIS, C. C., DESANTIS, T. Z., DEAL, C., DELEHAUNTY, K. D., DEWHIRST, F. E., DEYCH, E., DING, Y., DOOLING, D. J., DUGAN, S. P., MICHAEL DUNNE, W., SCOTT DURKIN, A., EDGAR, R. C., ERLICH, R. L., FARMER, C. N., FARRELL, R. M., FAUST, K., FELDGARDEN, M., FELIX, V. M., FISHER, S., FODOR, A. A., FORNEY, L. J., FOSTER, L., DI FRANCESCO, V., FRIEDMAN, J., FRIEDRICH, D. C., FRONICK, C. C., FULTON, L. L., GAO, H., GARCIA, N., GIANNOUKOS, G., GIBLIN, C., GIOVANNI, M. Y., GOLDBERG, J. M., GOLL, J., GONZALEZ, A., et al. 2012. Structure, function and diversity of the healthy human microbiome. *Nature*, 486, 207.
- THOMAS, T., GILBERT, J. & MEYER, F. 2012. Metagenomics - a guide from sampling to data analysis. *Microbial informatics and experimentation*, 2, 3-3.
- TIAN, J., QIAO, Y., WU, B., CHEN, H., LI, W., JIANG, N., ZHANG, X. & LIU, X. 2017. Ecological Succession Pattern of Fungal Community in Soil along a Retreating Glacier. *Frontiers in Microbiology*, 8, 1028.
- TISCHNER, R. 2000. Nitrate uptake and reduction in higher and lower plants. *Plant, Cell & Environment*, 23, 1005-1024.
- TRIPATHI, B. M., KIM, M., KIM, Y., BYUN, E., YANG, J.-W., AHN, J. & LEE, Y. K. 2018. Variations in bacterial and archaeal communities along depth profiles of Alaskan soil cores. *Scientific Reports*, 8, 504.
- TÜRK, R. & GÄRTNER, G. 2003. Biological Soil Crusts of the Subalpine, Alpine, and Nival Areas in the Alps. In: BELNAP, J. & LANGE, O. L. (eds.) *Biological Soil Crusts: Structure, Function, and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- ULLMANN, I. & BÜDEL, B. 2003. Ecological Determinants of Species Composition of Biological Soil Crusts on a Landscape Scale. In: BELNAP, J. & LANGE, O. L. (eds.) *Biological Soil Crusts: Structure, Function, and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- UROZ, S., BUÉE, M., MURAT, C., FREY-KLETT, P. & MARTIN, F. 2010. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental Microbiology Reports*, 2, 281-288.
- VAN DEN AKKER, J. J. H. & SOANE, B. 2005. COMPACTION. In: HILLEL, D. (ed.) *Encyclopedia of Soils in the Environment*. Oxford: Elsevier.
- VAN DER HEIJDEN, M. G. A., BARDGETT, R. D. & VAN STRAALLEN, N. M. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters*, 11, 296-310.
- VAN DER WALT, A. J., VAN GOETHEM, M. W., RAMOND, J.-B., MAKHALANYANE, T. P., REVA, O. & COWAN, D. A. 2017. Assembling metagenomes, one community at a time. *BMC Genomics*, 18, 521.
- VELLEND, M. 2010. *Conceptual Synthesis in Community Ecology*.

- VELUCI, R. M., NEHER, D. A. & WEICHT, T. R. 2006. Nitrogen Fixation and Leaching of Biological Soil Crust Communities in Mesic Temperate Soils. *Microbial Ecology*, 51, 189-196.
- VENTER, A., SIEBERT, S., RAJAKARUNA, N., BARNARD, S., LEVANETS, A., ISMAIL, A., ALLAM, M., PETERSON, B. & SANKO, T. 2018. Biological crusts of serpentine and non-serpentine soils from the Barberton Greenstone Belt of South Africa. *Ecological Research*, 33, 629-640.
- VERHEYE, W. H. 2009. *Land Use, Land Cover and Soil Sciences - Volume VII: Soils and Soil Sciences - 2*, EOLSS Publishers Company Limited.
- VĚTROVSKÝ, T. & BALDRIAN, P. 2013. The Variability of the 16S rRNA Gene in Bacterial Genomes and Its Consequences for Bacterial Community Analyses. *PLOS ONE*, 8, e57923.
- VILLARREAL-CHIU, J. F., QUINN, J. P. & MCGRATH, J. W. 2012. The genes and enzymes of phosphonate metabolism by bacteria, and their distribution in the marine environment. *Frontiers in microbiology*, 3, 19-19.
- VOGEL, H. J. 1997. Morphological determination of pore connectivity as a function of pore size using serial sections. *European Journal of Soil Science*, 48, 365-377.
- VOGEL, H. J., WELLER, U. & SCHLÜTER, S. 2010. Quantification of soil structure based on Minkowski functions. *Computers & Geosciences*, 36, 1236-1245.
- VOLLMERS, J., WIEGAND, S. & KASTER, A.-K. 2017. Comparing and Evaluating Metagenome Assembly Tools from a Microbiologist's Perspective - Not Only Size Matters! *PLOS ONE*, 12, e0169662.
- WALKER, T. S., BAIS, H. P., GROTEWOLD, E. & VIVANCO, J. M. 2003. Root Exudation and Rhizosphere Biology. *Plant Physiology*, 132, 44.
- WALL, D. H., NIELSEN, U. N. & SIX, J. 2015. Soil biodiversity and human health. *Nature*, 528, 69.
- WANG, C., LI, W., YANG, Z., CHEN, Y., SHAO, W. & JI, J. 2015. An invisible soil acidification: Critical role of soil carbonate and its impact on heavy metal bioavailability. *Scientific Reports*, 5, 12735.
- WANG, G. C. & WANG, Y. 1997. Frequency of formation of chimeric molecules as a consequence of PCR coamplification of 16S rRNA genes from mixed bacterial genomes. *Applied and Environmental Microbiology*, 63, 4645.
- WANG, W., LIU, Y., LI, D., HU, C. & RAO, B. 2009. Feasibility of cyanobacterial inoculation for biological soil crusts formation in desert area. *Soil Biology and Biochemistry*, 41, 926-929.
- WARDLE, D. A. 2006. The influence of biotic interactions on soil biodiversity. *Ecology Letters*, 9, 870-886.
- WARREN, S. D. 2003. Synopsis: Influence of Biological Soil Crusts on Arid Land Hydrology and Soil Stability. In: BELNAP, J. & LANGE, O. L. (eds.) *Biological Soil Crusts: Structure, Function, and Management*. Berlin, Heidelberg: Springer Berlin Heidelberg.

- WEBER, B., BÜDEL, B. & BELNAP, J. 2016. *Biological Soil Crusts: An Organizing Principle in Drylands*.
- WEBER, B., WU, D., TAMM, A., RUCKTESCHLER, N., RODRÍGUEZ-CABALLERO, E., STEINKAMP, J., MEUSEL, H., ELBERT, W., BEHRENDT, T., SÖRGEL, M., CHENG, Y., CRUTZEN, P. J., SU, H. & PÖSCHL, U. 2015. Biological soil crusts accelerate the nitrogen cycle through large NO and HONO emissions in drylands. *Proceedings of the National Academy of Sciences*, 112, 15384.
- WEI, W., YU, Y. & CHEN, L. 2015. Response of Surface Soil Hydrology to the Micro-Pattern of Bio-Crust in a Dry-Land Loess Environment, China. *PLOS ONE*, 10, e0133565.
- WEISS, S., XU, Z. Z., PEDDADA, S., AMIR, A., BITTINGER, K., GONZALEZ, A., LOZUPONE, C., ZANEVELD, J. R., VÁZQUEZ-BAEZA, Y., BIRMINGHAM, A., HYDE, E. R. & KNIGHT, R. 2017. Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, 5, 27-27.
- WEN, C., WU, L., QIN, Y., VAN NOSTRAND, J. D., NING, D., SUN, B., XUE, K., LIU, F., DENG, Y., LIANG, Y. & ZHOU, J. 2017. Evaluation of the reproducibility of amplicon sequencing with Illumina MiSeq platform. *PLOS ONE*, 12, e0176716.
- WHALLEY, W. R., RISELEY, B., LEEDS-HARRISON, P. B., BIRD, N. R. A., LEECH, P. K. & ADDERLEY, W. P. 2005. Structural differences between bulk and rhizosphere soil. *European Journal of Soil Science*, 56, 353-360.
- WHITMAN, W. B., COLEMAN, D. C. & WIEBE, W. J. 1998. Prokaryotes: The unseen majority. *Proceedings of the National Academy of Sciences*, 95, 6578.
- WILLIAMS, L., BORCHHARDT, N., COLESIE, C., BAUM, C., KOMSIC-BUCHMANN, K., RIPPIN, M., BECKER, B., KARSTEN, U. & BÜDEL, B. 2017a. Biological soil crusts of Arctic Svalbard and of Livingston Island, Antarctica. *Polar Biology*, 40, 399-411.
- WILLIAMS, L., COLESIE, C., ULLMANN, A., WESTBERG, M., WEDIN, M. & BÜDEL, B. 2017b. Lichen acclimation to changing environments: Photobiont switching vs. climate-specific uniqueness in *Psora decipiens*. *Ecology and evolution*, 7, 2560-2574.
- WILLIAMS, L., LOEWEN-SCHNEIDER, K., MAIER, S. & BÜDEL, B. 2016. Cyanobacterial diversity of western European biological soil crusts along a latitudinal gradient. *FEMS microbiology ecology*, 92, fiw157.
- WILPISZESKI, R. L., AUFRECHT, J. A., RETTERER, S. T., SULLIVAN, M. B., GRAHAM, D. E., PIERCE, E. M., ZABLOCKI, O. D., PALUMBO, A. V. & ELIAS, D. A. 2019. Soil Aggregate Microbial Communities: Towards Understanding Microbiome Interactions at Biologically Relevant Scales. *Applied and Environmental Microbiology*, 85, e00324-19.
- WILSON, W. H., VAN ETTEN, J. L. & ALLEN, M. J. 2009. The Phycodnaviridae: the story of how tiny giants rule the world. *Current topics in microbiology and immunology*, 328, 1-42.
- WOESE, C. R. 1987. Bacterial evolution. *Microbiological Reviews*, 51, 221.

- WOESE, C. R. & FOX, G. E. 1977. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proceedings of the National Academy of Sciences*, 74, 5088.
- WOESE, C. R., STACKEBRANDT, E., MACKE, T. J. & FOX, G. E. 1985. A Phylogenetic Definition of the Major Eubacterial Taxa. *Systematic and Applied Microbiology*, 6, 143-151.
- WOOD, D. E. & SALZBERG, S. L. 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15, R46.
- WU, Y., RAO, B., WU, P., LIU, Y., LI, G. & LI, D. 2013. Development of artificially induced biological soil crusts in fields and their effects on top soil. *Plant and Soil*, 370, 115-124.
- YANG, Y., WU, J., ZHAO, S., HAN, Q., PAN, X., HE, F. & CHEN, C. 2018. Assessment of the responses of soil pore properties to combined soil structure amendments using X-ray computed tomography. *Scientific Reports*, 8, 695.
- YEAGER, C., KUSKE, C., CARNEY, T., JOHNSON, S., TICKNOR, L. & BELNAP, J. 2012. Response of Biological Soil Crust Diazotrophs to Season, Altered Summer Precipitation, and Year-Round Increased Temperature in an Arid Grassland of the Colorado Plateau, USA. *Frontiers in Microbiology*, 3, 358.
- YEAGER, C. M., GALLEGOS-GRAVES, L. V., DUNBAR, J., HESSE, C. N., DALIGAULT, H. & KUSKE, C. R. 2017. Polysaccharide Degradation Capability of *Actinomycetales* Soil Isolates from a Semiarid Grassland of the Colorado Plateau. *Applied and Environmental Microbiology*, 83, e03020-16.
- YONOVITZ, M. & DROHAN, P. J. 2009. Pore morphology characteristics of vesicular horizons in undisturbed and disturbed arid soils; implications for arid land management. *Soil Use and Management*, 25, 293-302.
- YOSHITAKE, S., UCHIDA, M., KOIZUMI, H., KANDA, H. & NAKATSUBO, T. 2010. Production of biological soil crusts in the early stage of primary succession on a High Arctic glacier foreland. *New Phytologist*, 186, 451-460.
- YOU, J., LIU, X., ZHANG, B., XIE, Z., HOU, Z. & YANG, Z. 2015. Seasonal changes in soil acidity and related properties in ginseng artificial bed soils under a plastic shade. *Journal of Ginseng Research*, 39, 81-88.
- YU, X., DOROGHAZI, J. R., JANGA, S. C., ZHANG, J. K., CIRCELLO, B., GRIFFIN, B. M., LABEDA, D. P. & METCALF, W. W. 2013. Diversity and abundance of phosphonate biosynthetic genes in nature. *Proceedings of the National Academy of Sciences of the United States of America*, 110, 20759-20764.
- YUAN, H., GE, T., CHEN, C., DONNELL, A. G. & WU, J. 2012. Significant Role for Microbial Autotrophy in the Sequestration of Soil Carbon. *Applied and Environmental Microbiology*, 78, 2328.
- ZAADY, E. L. I., LEVACOV, R. & SHACHAK, M. 2004. Application of the Herbicide, Simazine, and its Effect on Soil Surface Parameters and Vegetation in a Patchy Desert Landscape. *Arid Land Research and Management*, 18, 397-410.

- ZHANG, B., KONG, W., WU, N. & ZHANG, Y. 2016. Bacterial diversity and community along the succession of biological soil crusts in the Gurbantunggut Desert, Northern China. *Journal of Basic Microbiology*, 56, 670-679.
- ZHANG, X., XU, W., LIU, Y., CAI, M., LUO, Z. & LI, M. 2018. Metagenomics Reveals Microbial Diversity and Metabolic Potentials of Seawater and Surface Sediment From a Hadal Biosphere at the Yap Trench. *Frontiers in microbiology*, 9, 2402-2402.
- ZHANG, Y. 2005. The microstructure and formation of biological soil crusts in their early developmental stage. *Chinese Science Bulletin*, 50, 117-121.
- ZHAO, J., ZHANG, B. & ZHANG, Y. 2008. Chlorophytes of biological soil crusts in Gurbantunggut Desert, Xinjiang Autonomous Region, China. *Frontiers of Biology in China*, 3, 40-44.
- ZIELKE, M., EKKER, A. S., OLSEN, R. A., SPJELKAVIK, S. & SOLHEIM, B. 2002. The Influence of Abiotic Factors on Biological Nitrogen Fixation in Different Types of Vegetation in the High Arctic, Svalbard. *Arctic, Antarctic, and Alpine Research*, 34, 293-299.
- ZINKE, P. J., STANGENBERGER, A. G., POST, W. M., EMANUEL, W. R. & OLSON, J. S. 1984. *Worldwide organic soil carbon and nitrogen data*.