Bacterial biogeography of the *rare* Charitable Research Reserve

by

Brent Seuradge

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2015

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Soil microbial communities play a dominant role in global biogeochemical cycles, with profound effects on agriculture, ecosystem stability, human health, and global climate. As a result, assessing their biogeographic patterns can help to further reveal mechanisms shaping their diversity and function in the environment. Furthermore, due to extensive spatial heterogeneity and environmental gradients, there is potential for overlooking key biogeographical patterns, critical metabolic processes, and novel bacterial taxa existing within deeper soil horizons that can be highly dependent on changes in land-usage. Additionally, an active area of research in soil microbial biogeography is assessing the extent to which current environmental or past historical factors constrain microbial community assemblages.

The objectives of this study were to examine and characterize depth-dependent bacterial community characteristics across multiple land-use types to explore subsurface biogeographical patterns. I collected soil samples across seven distinct land-use types to depths of 45 cm, including old-growth and mature forests, decommissioned, and active agricultural fields from the ***rare*** Charitable Research Reserve (Cambridge, Ontario). Bacterial communities were characterized by sequencing of bacterial 16S rRNA gene amplicons coupled with multivariate statistical analyses from 376 soil samples. In addition, to explore functional and metabolic characteristics of collected soils, the PICRUSt algorithm was used to predict metagenomes of uncharacterized taxa.

Soil bacterial communities across all sites were strongly influenced by depth. Upper soils (0–15 cm) and open field sites maintained higher bacterial alpha-diversity than deeper soils and forested sites. The magnitude of soil depth effects appeared to differ across environment types highlighting that land-use type also plays a significant role in shaping communities; bacterial communities across the field sites (i.e., grasslands and agricultural sites) were shown to be more strongly affected than forested sites. Soil pH, which exhibited a large gradient across samples, appeared to be largely responsible for differential shifts in communities with depth across land-use types especially considering that C, NH_4^+, NO_3^-, moisture, and texture showed generally consistent trends with depth across all sites. This observation was further corroborated by NPMANOVA and CCA, which highlighted that pH was among the top explanatory variable explaining >15% of the variation in the dataset. This finding further emphasizes that pH is a strong predictor of bacterial community composition, not only across surface soils, but also within the soil subsurface. Overall, the impact of pH on soil bacterial community composition exceeded that of depth.

The effect of land-use type on subsurface bacterial communities was found to be largely attributed to differences in dominant plant communities. Field sites were characterized by tall grasses

whereas forest sites were characterized by woody tree species. Considering that plant inputs (i.e., root exudates, litter) are translocated through soils over time and affect the physicochemical environment, these findings further enforce that plants play important roles in structuring soil bacterial communities across environment types. In addition, contrary to evidence from the aboveground plant communities and site histories, there was no direct evidence of bacterial community succession throughout soils across the field sites sampled in this investigation. Instead, edaphic factors including soil texture, particularly sand, silt, clay, and moisture, appeared to govern changes in overall community composition across the field sites, highlighting the importance of the immediate physicochemical environment in shaping soil bacterial communities.

Soils across all sites and depths were dominated by members of the *Proteobacteria* (33.2%), *Actinobacteria* (27.8%), *Acidobacteria* (14.9%), *Chloroflexi* (6.6%), *Gemmatimonadetes* (4.7%), *Bacteroidetes* (3.0%), *Nitrospirae* (2.1%), *Firmicutes* (2.3%), *Verrucomicrobia* (1.7%), and *Latescibacteria* (formerly WS3; 1.2%). In addition to observing trends in specific phyla with depth (e.g., *Proteobacteria* and *Bacteroidetes*), data also highlighted consistent depth-specific changes in OTU relative abundances. Although the majority of significant correlations were negative (indicating a decrease in abundance with increasing depth), Spearman's correlation analysis found evidence for consistent positively correlated OTUs with depth. Notably, all positively depth-correlated OTUs were affiliated with uncultivated bacteria, further highlighting that subsurface environments are poorly studied. Correlation analyses were also conducted for pH. *Nitrospirae* and *Chloroflexi* members were among the top strongly and positively correlated taxa with pH, consistent with previous studies. *Acidomicrobiia* and *Solibacteres* classes, members of the *Acidobacteria* phylum, were found to be strongly and negatively correlated with pH, which is also consistent with previous research. These results further demonstrate the importance of pH in shaping soil bacterial communities considering that many taxa are adapted to narrow and specific growth and pH ranges.

The PICRUSt results reflected observations noted in the taxonomy-based analysis. "Transporter" associated genes appeared to show differential abundances across land-use type. Forest sites, in particular site CA, a mature forest environment, had the lowest abundance of "transporter" associated genes. This result may further highlight pH effects on soil bacterial communities, considering that site CA had samples with the lowest pH and, consequently, the lowest species diversity.

Overall, this research has set up baseline observations of bacterial community dynamics at the ***rare*** Charitable Research Reserve expanding on the few studies that have included soil depth as an environmental gradient and paving the way for future investigations. In addition, this study exemplifies important global environmental gradients including depth, land-usage, and soil

biogeochemistry operating at smaller geographical scales across consistent underlying geology. Furthermore, this work has added insight concerning the interplay of the immediate physicochemical environment and past historical legacies in shaping soil microbial communities. Future research with the dataset generated will further explore bacterial taxa that vary in relation to pH and depth, in addition to phylogenetically novel taxa existing at low relative abundance, providing additional insight into the unexplored biodiversity of soil microbial communities.

Acknowledgements

I would like to take this opportunity to express my thanks to the individuals who have helped, guided, and supported me throughout the course of this degree. First and foremost, I would like to express my deepest thanks to Josh D. Neufeld for his mentorship, support, enthusiasm, and dedication throughout Master's program. All of his insightful input and helpful critiques have truly inspired and helped build my confidence as a student and researcher. In addition, I would also like to thank the members of supervisory committee, Barbara J. Butler and Maren M. Oelbermann, both whom have provided significant insight, advice, and motivation during the course of my project.

I would also like to thank Katja Engel, who has been a mentor and teacher during my Master's program providing me with theoretical framework and necessary laboratory skills, particularly with Illumina library preparation and sequencing, which have been invaluable to the completion of this project. Her patience and knowledge has truly been an integral part during this project. Furthermore, I would like to thank Michael W. Hall and Michael D. Lynch, for their computational, statistical, and bioinformatic expertise during this project, particularly with AXIOME. Their patience and easy explanations were extremely helpful. I would also like to thank Colin Elliot for his help with AXIOME.

I also wish to thank Briallen Lobb for her tremendous help with soil sampling. Her enthusiasm and input during field sampling was extremely helpful. I would also like to thank members from the ***rare*** Charitable Research Reserve, including Jenna Quinn, Katherine McLeod, and John MacDonald for their help in organization, soil sampling, plant species identification, and navigation across the ***rare*** reserve.

Furthermore, I thank all the past and current members of the Neufeld lab, for all their knowledge, technical advice, and guidance, which have been invaluable throughout the course of my research program.

Finally, I would like to thank all my closest family and friends; Mom and Dad, Jason, Andrew, Marc, and Jenn. Without your support and encouragement, completing this degree would have been substantially more difficult.

Table of Contents

AUTHOR'S DECLARATION ii
Abstract iii
Table of Contents vii
List of Figures x
List of Tables xii
List of Abbreviations xiii
Chapter 1 Introduction 1
1.1 Soil microbial communities 1
1.1.1 The soil environment 1
1.1.2 Microbial diversity and function 3
1.1.3 Factors affecting soil microbial communities 4
1.1.3.1 Abiotic factors 5
1.1.3.2 Biotic factors 10
1.2 Microbial biogeography 10
1.2.1 Microbial biogeographical theory 10
1.2.1.1 Past considerations 10
1.2.1.2 Modern aspects of microbial biogeography 11
1.2.1.3 Processes shaping microbial biogeography 14
1.2.2 Patterns of soil microbial biogeography 15
1.2.2.1 Distance decay 15
1.2.2.2 Latitude 15
1.2.2.3 Depth 15
1.2.3 Importance of microbial biogeography 17
1.3 Characterizing microbial communities 17
1.3.1 Some prerequisites 17
1.3.2 Culture-dependent methods 18
1.3.3 Culture-independent methods 18
1.3.3.1 Metagenomics and the 16S rRNA marker gene 19
1.3.3.2 HTP sequencing and bioinformatics 20
1.3.3.2.1 Classifying microorganisms from marker genes 22
1.3.3.2.2 Methods in sequence clustering 23
1.4 Research description 23

1.4.1 Research overview 23
1.4.2 Objectives and hypotheses 24
Chapter 2 Materials and Methods 25
2.1 Site characterization 25
2.2 Sample collection 30
2.3 Soil physicochemical characterization 31
2.4 Microbial community analysis 32
2.4.1 DNA extraction & quality assessment 32
2.4.2 HTP sequencing 33
2.4.2.1 PCR amplification of the16S rRNA gene 33
2.4.2.2 Illumina library preparation 34
2.4.2.3 Illumina sequencing 35
2.4.3 Bioinformatic and statistical analyses 35
2.4.3.1 Data subsets 35
2.4.3.2 Sequence clustering, OTU table generation, and downstream analyses 36
2.4.3.3 Comparison of sequence clustering strategies 39
2.4.3.4 Community function predictions 40
Chapter 3 Results 42
3.1 Soil characteristics 42
3.2 HTP sequencing of soil samples from the *rare* Charitable Research Reserve 44
3.2.1 Sequence clustering analysis 45
3.3 Identifying patterns in soil bacterial communities 49
3.3.1 OTU richness 49
3.3.2 Depth 51
3.3.3 pH effects 55
3.3.4 Land-use responses 57
3.4 Additional edaphic factors 57
3.5 Exploring taxonomic distributions across the *rare* Charitable Research Reserve 62
3.5.1 Taxa associations 62
3.5.2 Indicator species 72
3.6 Bacterial community function predictions 74
3.6.1 Predicted metagenome shifts with depth and across land-usage 74
3.6.2 Assessing differences in predicted genes with depth and across land-use type 76
Chapter 4 Discussion 79

4.1 16S rRNA gene clustering analysis 79
4.2 Bacterial community dynamics across the *rare* Charitable Research Reserve 80
4.2.1 Depth influences 80
4.2.2 Land-use influences 83
4.2.3 Successional responses 84
4.3 Taxa dynamics 84
4.3.1 Phylum-level trends 84
4.3.2 Depth and pH specific taxa correlations 86
4.3.3 Indicator species 87
4.4 PICRUSt metagenome predictions 88
4.5 Microbiogeographical implications 90
Chapter 5 Conclusions & Future Directions 92
5.1 Contributions and perspectives 92
5.2 Future directions 94
Bibliography 97
Appendix A 118

List of Figures

Figure 1: General trend of bacterial diversity and pH across North American soils 7
Figure 2: Summary of the four alternative hypotheses addressing microbial biogeography 12
Figure 3*:* Map of the ***rare*** Charitable Research Reserve including the seven sampling sites 27
Figure 4: Overview of the sampling design 30
Figure 5: Flowchart showing the overall bioinformatic and sequence processing steps 37
Figure 6: pH with depth from all sites across soils from the ***rare*** Charitable Research Reserve 43
Figure 7: Inorganic and organic C content with depth from all sites across soils from the ***rare*** Charitable Research Reserve 43
Figure 8: NH_4^+ and NO_3^- content with depth from all sites across soils from the ***rare*** Charitable Research Reserve. 44
Figure 9: Rarefaction curves showing OTU counts from clustering schemes 48
Figure 10: Number of observed OTUs for sampled sites and depth 49
Figure 11: Number of observed OTUs in relation to pH and soil depth 50
Figure 12: Number of observed OTUs in relation to pH and site (land-use type) 50
Figure 13: PCoA ordinations based on Bray-Curtis dissimilarities showing the effect of depth (A and C) and land-use type (B and D) on soil bacterial communities from the ***rare*** Charitable Research Reserve at coarse (15 cm; C and D) and fine (5 cm; B and D) scale depth increments 52
Figure 14: PCoA ordinations based on Bray-Curtis dissimilarities and MRPP results for depth-constrained ordinations 53
Figure 15: PCoA ordinations based on Bray-Curtis dissimilarities for forest (A and B) and field (C and D) datasets by depth (A and C) and site (B and D) 54
Figure 16: PCoA ordination based off Bray-Curtis dissimilarities of samples in relation to pH, for all samples analyzed. 56
Figure 17: Environmental factors explaining variation in community dissimilarity (Bray-Curtis) from samples taken across the ***rare*** Charitable Research Reserve for the field, forest, and 15 cm increment dataset ($p \leq 0.005$) 59
Figure 18: CCA ordination biplot showing the 15 cm data subset of soil samples taken from the ***rare*** Charitable Research Reserve and continuous environmental factors 60
Figure 19: CCA ordination biplot showing A) forest data subset and B) field data subset of soil samples taken from the ***rare*** Charitable Research Reserve and continuous environmental factors 61
Figure 20: Shifts in relative abundance of the top ten phyla with depth from soils taken from the ***rare*** Charitable Research Reserve 63

Figure 21: CCA biplot showing OTU distributions among the most abundant classes of bacteria 64
Figure 22: Indicator species analysis for individual sites and depth based on an indicator value cut-off of ≥ 0.7, sequence abundance ≥ 100, $p \leq 0.05$). 72
Figure 23: Indicator species analysis for pH based on an indicator value cut-off of ≥ 0.7, sequence abundance ≥ 100, $p \leq 0.05$). 73
Figure 24: PCA analysis of predicted metagenomes with depth for the A) Full dataset, B) Field sites only, and C) Forest sites only 75
Figure 25: PCA analysis of predicted metagenomes for forests and field sites 76
Figure 26: Heat map showing filtered ($q < 0.05$; Benjamini-Hochberg corrected; effect size > 0.35) predicted metagenomes per sample 77
Figure 27: Post-hoc analysis (Kruskal-Wallis H-test, Tukey-Kramer) of predicted genes associated with the KEGG category "transporter" based on land-use type (p values are Benjamini-Hochberg corrected) 78

Appendix A

Figure 28: DCA for forest sites showing the reduction of the “arch effect” coloured by A) depth and B) pH. 125

List of Tables

Table 1: Common examples of soil prokaryotes and their major biogeochemical roles 3
Table 2: Summary of major factors affecting soil microorganisms 5
Table 3: Summary of studies that have found evidence for microbial biogeographical patterns 13
Table 4: Collection of literature examining soil microbial diversity with an emphasis on the impact of depth on soil microbial communities. 16
Table 5: Characteristics of example HTP sequencing platforms 21
Table 6: Summary of site information. 28
Table 7: Dominant plant species of sampled sites across the ***rare*** Charitable Research Reserve. 29
Table 8: Summary of the methods used by the Agriculture & Food Laboratory (University of Guelph) for metadata characterization. 31
Table 9: Summary of 16S rRNA gene data subsets 45
Table 10: Comparison of clustering algorithms and PANDAseq quality thresholds (PQT) on sequence and OTU count dynamics 47
Table 11: ANOSIM results for depth, land-use type, and site ID for various data subsets. 55
Table 12: Spearman's rank correlation analysis showing correlated (based on Spearman's ρ) taxa with depth for both the 15 and 5 cm increment datasets 65
Table 13: Spearman's rank correlation analysis showing correlated (based on Spearman's ρ) taxa with pH for both the 15 and 5 cm increment datasets 68

Appendix A

Table 14: Summary of physicochemical data 118
Table 15: Soil horizon characterization for each soil sample across ***rare*** Charitable Research Reserve 120

List of Abbreviations

%	percent
°C	degrees Celsius
µg	microgram
µL	microliter
µm	micrometer
µM	micromolar
ANOSIM	analysis of similarity
ANOVA	analysis of variance
ATP	adenosine triphosphate
AXIOME	automation, extension, and integration of microbial ecology
bp	base pair
C	carbon
Ca	calcium
CCA	constrained (canonical) correspondence analysis
CD-HIT	cluster database at high identity with tolerance
$C_{(i)}$	inorganic carbon
Cl	chloride
cm	centimeter
$C_{(o)}$	organic carbon
CO_2	carbon dioxide
COG	clusters of orthologous groups
$C_{(t)}$	total carbon
DCA	detrended correspondence analysis
DGGE	denaturing gradient gel electropheresis
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
EDTA	ethylenediaminetetraacetic acid
Fe	iron
g	gram
g	gravity
GSC	genomics standard consortium
ha	hectare
h	hour
H^+	proton
H_2O	water
HS^-	hydrogen sulfide
HTP	high-throughput
K	potassium
KCl	potassium chloride
KEGG	Kyoto encyclopedia of genes and genomes
kg	kilogram
km	kilometer
lbs	pounds
m	meter
M	molar
Mb	megabyte
Mg	magnesium
mg	milligram
MIMARKS	minimum information about a marker gene sequence
min	minutes
mL	milliliter
mM	millimolar
mm	millimeter
Mn	manganese
MRPP	multiple-response permutation procedure
N	nitrogen
n	sample size
N	population size
NA	not applicable
NaOH	sodium hydroxide

ng	nanogram
NH_3	ammonia
NH_4^+	ammonium
nM	nanomolar
nm	nanometer
NO_2^-	nitrite
NO_3^-	nitrate
NPMANOVA	non-parametric multivariate analysis of variance
NSTI	nearest sequenced taxon index
O_2	oxygen
OTU	operational taxonomic units
P	phosphorus
PANDASeq	paired-end assembler for DNA sequences
PCA	principal components analysis
PCoA	principal coordinates analysis
PCR	polymerase chain reaction
pH	power of hydrogen
PICRUSt	phylogenetic investigation of communities by reconstruction of unobserved states
PLFA	phospholipid-derived fatty acid
pM	picomolar
PO_4^{3-}	phosphate
PyNAST	python nearest alignment space termination tool
QIIME	quantitative insights into microbial ecology
qPCR	quantitative polymerase chain reaction
rRNA	ribosomal ribonucleic acid
RAM	random access memory
RDP	ribosomal data project
s	second
S	sulfur
SDS	sodium dodecyl sulfate
Se	selenium
SIC	soil inorganic carbon
SIP	stable isotope probing
SO_4^{2-}	sulfate
SOC	soil organic carbon
STAMP	Statistical analysis of metagenomic profiles
TAE	tris-acetate EDTA
Taq	*Thermus aquaticus*
UPARSE (+1)	UPARSE algorithm with singletons included
UPARSE(+2)	UPARSE algorithm without singletons included
UPGMA	Unweighted Pair Group Method with Arithmetic Mean
V3	16S rRNA gene variable region 3
V4	16S rRNA gene variable region 4
ρ	Spearman's rank correlation coefficient

Chapter 1
Introduction

1.1 Soil microbial communities

1.1.1 The soil environment

Soils are complex, heterogeneous mixtures of minerals, gases, H_2O, organic matter, and biological entities occupying the interface between the atmosphere and lithosphere (1). Whether providing a reservoir of nutrients for growing economically important foods or protecting the natural environment from contaminants, soils are critical for maintaining Earth's natural systems (2, 3). Due to extensive spatial heterogeneity, soils are presumed to support the majority of Earth's microbial diversity, with most species existing at low relative abundance (4–6). The spatial heterogeneity affecting the distribution and diversity of these organisms is greatly influenced by pedogenetic processes, which can lead to vastly distinct physical and chemical characteristics within soils (2). In the context of soil microbiology, characterizing the physical, chemical, and overall biological nature of soils is critical for assessing how microbial communities are distributed and thrive in these systems.

Soils are the result of the ongoing weathering of parent material, stimulated over time by both biotic and abiotic factors such as vegetation type, climate, parent material type, and topography (7). The development of soils is a result of the accumulation of organic material (i.e., living and decaying organisms as well as humic substances) in surface layers (7, 8). The downward movement of clays, soluble ions, and the development of soil structure eventually results in characteristic vertical gradients with increasing depth (7, 9). Maturation of soils via gains, losses, translocations, and transformations defines the physicochemical environment, ultimately generating characteristic layers, or horizons (7, 9). Horizons are key aspects of soil profiles and they are representative of the contemporary and historical biological, geological, and climatic environment at a given location (7, 9). Generally, soil profiles are characterized by an organic-rich litter layer (O-horizon) followed by A- (organic-rich mineral soil which generally hosts the majority of microbial activity), B- (chemically and physically altered parent material), and C- (parent material) horizons (9, 10).

The soil environment is a chemically rich and reactive body (9, 11). In the context of microbial ecology (i.e., the interactions of microorganisms with each other and their environment (1)), much of the availability of nutrients, activity, and distributions of microorganisms in soils are often dependent on the composition of the colloidal fraction of these systems (11, 12). The soil colloidal fraction is comprised of extremely small (<1 μm in diameter) clay and organic particles (9,

11). In temperate soils, clay particles maintain an overall net negative charge as a result of their silicon- and aluminum-oxide lattice structure, which preferentially allows substitutions with multivalent cations (e.g., Mg^{2+}, Al^{3+}) as well as frequent ionization of terminal hydroxyl groups (11). Changes in the concentration of common cations present in the soil solution such as Ca^{2+}, Mg^{2+}, K^{+}, Na^{+}, H^{+}, in addition to their characteristic adsorption affinities to the charged surfaces of colloids, can shift the equilibria in these systems, thus allowing displacement of other more weakly attached ions from clay surfaces (11). In a similar way, humic substances, which are complex polymers formed by the degradation of biomass, often maintain a net negative charge as a result of hydrophilic and negatively charged functional groups on the exterior of humic colloids (11). This dynamic movement of cations from soil particles to the soil solution in temperate environments, known as the cation exchange capacity, is critical to soil microorganisms because it directly affects microbial activity by sorbing cells to colloidal particles as well as changing the availability of important solutes present in the soil solution (12–14).

Together with macro- and micro-fauna, soils are home to immense biological diversity (15, 16). The soil ecosystem is inhabited by bacteria, archaea, fungi, and viruses as well as plants and animals (e.g., earthworms, mammals); all of these organisms collectively play important roles in soil development and architecture (17, 18). Beyond providing constant sources of organic matter in the form of crop residues and litter material, plants play important roles in modifying soils physically (19, 20). Plant roots enlarge pore spaces, help stabilize organic-mineral bonds, and increase soil stability through aggregation (19). In addition, because roots can extend to great depths (e.g., greater than 5 m for some woody plants), they often supply deeper soil horizons with sources of organic matter (9). Plant roots also alter the immediate adjacent environment, giving rise to chemically and biologically distinct zones collectively known as the rhizosphere (21). Shifts in rhizosphere properties as a result of root exudates (e.g., low-molecular-mass compounds, polymerized sugar, secondary metabolites, and dead root cells) and plant respiration play major roles in shaping soil microbial communities (21, 22). Additionally, mycorrhizal fungi participate in mutualistic associations with vascular plant roots playing important roles in defining the soil environment by affecting soil fertility and aggregation (23). Other organisms, such as burrowing animals, earthworms, arthropods, gastropods, nematodes, algae, and protists, play roles in defining soil structure via actions such as physical disruption, aeration, organic matter breakdown, and production of important organic substances and nutrients (15).

1.1.2 Microbial diversity and function

With billions of microorganisms in a single gram of soil, representing potentially tens of thousands of species, terrestrial environments are among the most diverse and heterogeneous habitats for microbes, with most species existing at low relative abundance (5, 24). Microorganisms are critical for maintaining soil fertility and catalyzing biogeochemical processes on Earth (25, 26). Organic matter decomposition by soil prokaryotic and eukaryotic activity is one of the most important functional roles of soil fauna (27, 28). Decomposition of labile organic matter, including dead leaves, roots, and other plant material, allows for the assimilation of important elements (e.g., C, N, P, S) by organisms such as bacteria, archaea, and fungi (27). In this way, the cycling of important biogeochemical elements is maintained by a rich diversity of soil microorganisms (27, 29). Inorganic transformations such as oxidation/reduction reactions of compounds that contain elements such as nitrogen (e.g., NO_2^-, NO_3^-, NH_4^+) (30), sulfur (e.g., SO_4^{2-}, HS^-) (31), and phosphorus (e.g., PO_4^{3-}) (32), are carried out by a host of microbial populations (Table 1).

Table 1: Common examples of soil prokaryotes and their major biogeochemical roles (adapted from Pepper *et al*. (1)).

	Organism	Function/roles	Reference(s)
[a]Autotrophs	*Nitrosomonas, Nitrososphaera*	Nitrification; $NH_4^+ \rightarrow NO_2^-$	(33, 34)
	Nitrobacter	Nitrification; $NO_2^- \rightarrow NO_3^-$	
	Thiobacillus	Sulfur oxidation; $S \rightarrow SO_4^{2-}$	(35, 36)
	Thiobacillus denitrificans	Sulfur oxidation/denitrification; $S \rightarrow SO_4^{2-}$	
	Thiobacillus ferrooxidans	Iron oxidation; $Fe^{2+} \rightarrow Fe^{3+}$	
[b]Heterotrophs	*Rhizobium, Frankia*	Nitrogen fixation; $N_2 \rightarrow NH_3$	(37)
	Actinomycetes Streptomycetes	Antibiotic production; organic matter degradation	(38, 39)
	Bacillus	Carbon cycling	(40, 41)
	Clostridium	Carbon cycling (fermentation)	(42)
	Methanotrophs (e.g., *Methylosinus*)	Carbon cycling (methane oxidation)/TCE* breakdown; $CH_4 \rightarrow CO_2$	(43, 44)
	Bacillus, Caulobacter, Hyphomicrobium, Pseudomonas	Manganese oxidation; $Mn^{2+} \rightarrow Mn^{3+}$	(45–47)

[a]Organisms that can fix carbon (e.g., synthesize complex molecules); [b]organisms that cannot fix carbon (e.g., relies on fixed carbon from other sources); *trichloroethylene

In addition, microorganisms in soils have important roles in heavy metal transformations (e.g., Fe, Mn, Se, and Cr), often influencing the overall quality of natural systems (46, 48, 49). These can include conversions of toxic metal forms to less toxic states (46, 48, 49). Soil microorganisms also play critical roles in the decomposition of toxic anthropogenically derived xenobiotics (i.e.,

foreign, non-naturally produced substances) (50–52). Although most detoxifying bacteria and archaea are concentrated in surface soils where organic matter is highest, many microbial taxa extend to subsurface horizons and are capable of anaerobic degradation of recalcitrant toxicants (28, 50). With such broad functional roles of soil microorganisms (Table 1), and with wide-reaching implications for Earth's natural systems, detailed studies of soil microbial ecology are important for a robust understanding of global environmental processes.

1.1.3 Factors affecting soil microbial communities

As discussed in section 1.1.1, soils are complex environments that have a myriad of dynamic physical, chemical, and biological processes operating within them (Table 2). Both biotic and abiotic factors play critical roles in affecting soil microbial diversity (Table 2) (24). As a result, the metabolic diversity and niche adaptations of microorganisms reflect complex physicochemical gradients within soil profiles (4, 25).

Table 2: Summary of major factors affecting soil microorganisms (adapted from Pepper *et al*. (1)).

Factor		Known effect(s)	Reference(s)
Abiotic	pH	Maximum diversity observed at near-neutral pH (6-8)	(53, 54)
	Organic matter content	Higher concentrations support relatively more diversity	(55)
	Water content	Moderate moisture supports higher diversity (e.g., at field capacity)	(56–58)
	Oxygen concentration	Oxic environments support more diversity	(1, 59)
	Temperature	Temperature extremes reduce diversity	(57, 60)
	Land management practices (e.g., soil tillage)	Gradual soil homogenization reduces micro-environment diversity leading to diversity reduction	(61)
	Soil structure & texture	More favourable environments with mixture of sand, silt, and clay particles. Well-defined aggregate structure supports more favourable environment types for microorganisms	(13)
	Available nutrients	Carbon and nitrogen forms are necessary for growth and activity; when in limiting amounts, will decrease diversity	(27)
	Salinity	Generally decreases diversity due to osmotic pressures on cells	(62)
Biotic	Vegetation	Can help stimulate or inhibit certain microbial populations. Rhizosphere implications	(63)
	Microbe-microbe interactions	Drives community dynamics through, for example, mutualism, competition, or predation	(55, 64)

1.1.3.1 Abiotic factors

The distribution and organization of soil microbial communities is largely dependent on soil structure (65, 66). Soils exhibit specific structural arrangements as a result of the composition of sand (0.05–2 mm), silt (0.002–0.05), clay (<0.002 mm), and organic particles. These structural arrangements, or aggregates, are characterized by natural zones of weakness when a soil mass is physically disrupted (13). The size of aggregates can play a role in defining overall microbial community structure. Smaller microaggregates (2–250 μm), which temporarily coalesce to form large and less stable macroaggregates (0.25–5 mm), are often considered to harbour immense microbial and enzyme diversity (13, 67). Relative to macroaggregates, the interior of microaggregates likely represent secluded and stable habitats characterized by low predation, consistent moisture availability, as well as low nutrient and O_2 levels (67). When assessed individually, organisms

inhabiting microaggregates potentially use distinct metabolic strategies and carry out important roles in ecosystem functions at local scales (13, 67).

Pore networks within and between soil macroaggregates and microaggregates affect the transfer of fluids (air, soil solution), breakdown of organic matter, and microbial activity and diversity (13, 65). Colonizing microorganisms are physically restricted in soils by pore space size (68). Pore spaces can act like microbial caves (i.e., open pockets within soils) or mazes (i.e., irregularly shaped networks) and the diameters of these structures have direct effects on important physicochemical constraints on microorganisms such as diffusion and availability of nutrients (68–70). Recent estimates suggest that, in some instances, greater than 80% of bacteria can be found in pore spaces located in soil microaggregates (24, 67). The influence of pore spaces on soil microbes has been demonstrated by Kilbertus (68), who determined that a consistent 3:1 relationship exists between pore space diameter and the diameter of cells or colony sizes within them. Furthermore, past work (67, 71, 72) has also revealed differences in microbial communities associated with differences in pore or aggregate sizes, but such studies were limited due to methodological approaches (13). However, using a photo-oxidation approach to selectively sterilize the surfaces of microaggregrates, Mummey *et al.* (13) showed specific shifts in bacterial communities on and within these structures in relation to changes in land-use type. These observations further confirmed that soil architecture can be an important determinant of the distribution and diversity of soil microorganisms.

An important global driver of soil microbial diversity and composition is pH (53, 73, 74). Surface soils tend to be more acidic as a result of more direct interactions with acidifying processes, such as fertilization and organic matter (particularly organic acids) inputs from plants (75). Common fertilizers such as urea ($(NH_2)_2CO$) or anhydrous NH_3 can decrease soil pH via oxidative processes as shown in equations 1 and 2 below (76, 77):

$$\mathbf{(NH_2)_2CO} + 4O_2 \rightarrow 2NO_3^- + \mathbf{2H^+} + CO_2 + H_2O \quad (1)$$

$$\mathbf{NH_3} + 2O_2 \rightarrow NO_3^- + \mathbf{H^+} + H_2O \quad (2)$$

Furthermore, organic matter inputs from plants have been considered an acidifying process in soils as a result of H^+ release associated with organic anions or nitrification processes (78). However, the degree of this impact on soil pH varies because it is the organic anion input that governs H^+ release, which often varies between soil types (78). As a result, the release of organic acids from roots are likely more important in acidifying top soil (78, 79).

Deeper soil horizons tend to have a higher pH in relation to surface soils, often as a result of buffering by bedrock material (80). Non-tilled soils (e.g., forest systems, croplands, and unplowed

grasslands) have well-stratified vertical pH gradients, with the majority of the stratification existing within the top few centimeters (80). Moreover, microbial community surveys demonstrate that increasing pH correlates with increasing bacterial diversity, with maximum taxonomic diversity observed at soil pH values between approximately 6–8 (Figure 1) (81).

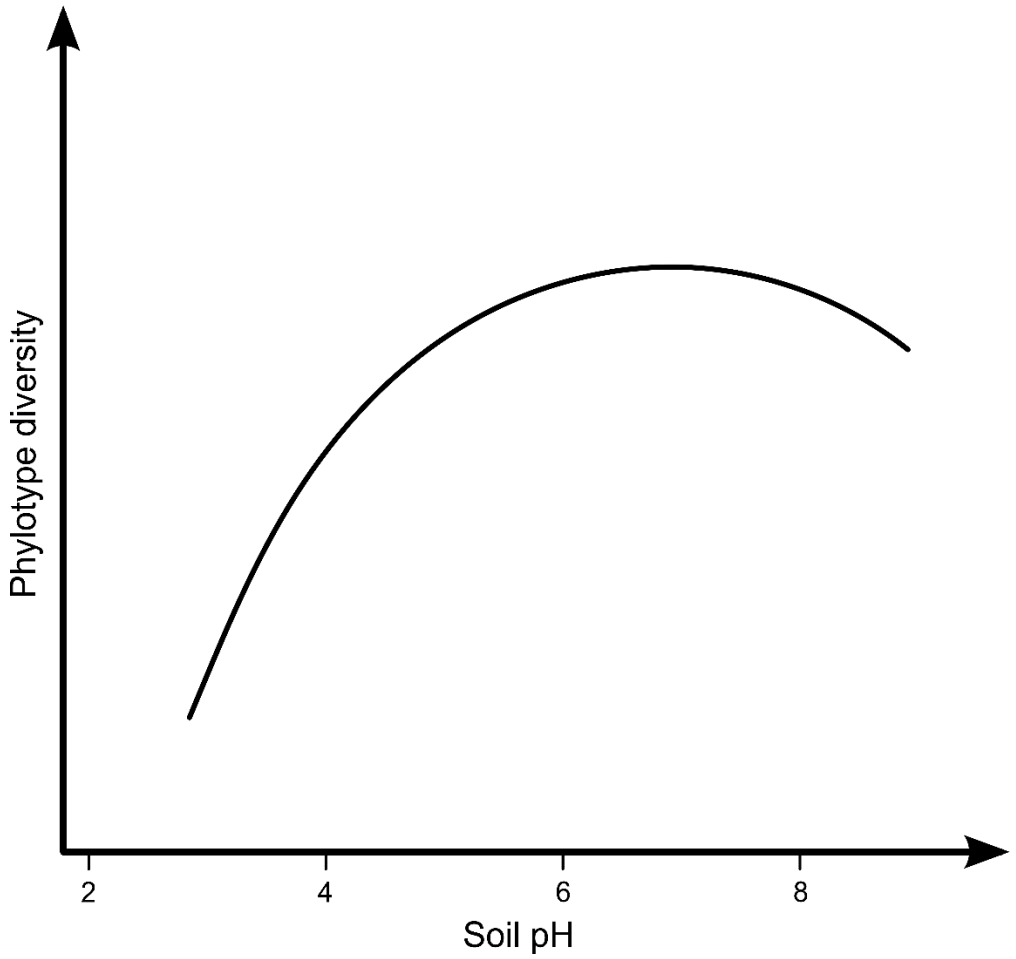

Figure 1: General trend of bacterial diversity and pH across North American soils. Bacterial diversity, measured using the Shannon index, increases with soil pH to a maximum diversity at neutral to near-neutral pHs (adapted from Fierer & Jackson (81)).

In addition to influencing overall bacterial diversity, soil pH influences specific microbial groups differentially (53, 82, 83). Prevalence of subgroups of the *Acidobacteria* phylum demonstrate specific trends with increasing or decreasing pH; relative abundances of subgroups 1, 2, and 3 reportedly decrease with increasing pH, whereas subgroups 5, 6, and 7 were shown to increase in abundance with increasing pH (53, 82). Members of the *Alphaproteobacteria* phylum were reported to increase in abundance with increasing pH (82). Furthermore, some evidence suggests characteristic changes in

the relative abundances of *Nitrospirae*, *Firmicutes*, *Actinobacteria*, and *Betaproteobacteria* with pH (53, 82). Data have also suggested that archaeal abundance decreases between pHs of ~4.5–5.1 before sharply and consistently increasing after this range to a pH of ~8 (84). This non-uniform pH influence on soil archaea is hypothesized to be attributed to the consistent observation of the contrasting abundances of archaea, bacteria and fungi at extreme pHs in soils resulting from competitive interactions (84). Although reported to be predominant in lower pH forest soils (85, 86), recent reports indicate that soil fungi are overall weakly influenced by pH due to their wider growth ranges (82, 87).

Soil organic carbon (SOC), a component of soil organic matter (SOM), is an important factor for maintaining soil quality (i.e., the measure of the capacity of soils to perform a particular ecological function such as nutrient cycling or temperature moderation (9)) and a diverse assemblage of microbial communities (27). Generally, surface soils are relatively rich in available C substrates as a result of detritus and root exudates (88). In some soil systems such as forestry and agro-ecosystems, the quantity and quality (i.e., "bioaccessibility") of organic matter decreases with depth as a result of lower inputs and decreased molecule lability (88, 89). As a result, there is an overall proportional relationship between microbial abundances and diversity in relation to SOC content with depth (88). Fierer *et al.* (88) reported that microbial diversity, as measured by PLFA profiles, declined with depth in response to SOC abundance and quality therefore leading to the hypothesis that the state of SOC in soils can be a predictor of microbial community variations throughout soil profiles.

N is a critical requirement for all microbial life and represents an important component of soil biogeochemistry (90). Because N quantity closely mimics SOC patterns within soils, typical decreases in N with depth generally correlate with proportional decreases in microbial diversity (91). Furthermore, many microbial N transformations are dependent on the O_2 mixing ratio of the system (e.g., oxic vs. anoxic environments) (90). In surface soils, where O_2 is usually elevated and inputs of available N compounds are high, N-cycling by nitrifying bacteria and archaea and N_2-fixing bacteria, in low-O_2 root nodule microenvironments, are dominant (92). Matejek *et al.* (92) illustrated that O-horizons within investigated soils showed the greatest net nitrification rates, which corresponded with greater microbial species diversity than mineral soil horizons. This was hypothesized to be attributed to faster metabolic processing, which is generally favourable in oxic surface soil environments (92). In some instances, the dominance of anaerobic reactions, such as denitrification and dissimilatory NO_3^- reduction to NH_3 (DNRA), can become dominant in subsurface horizons with implications for greenhouse gas generation (i.e., N_2O release) (92–94). Overall, N is an important element with considerable implications in shaping microbial community diversity and assemblages in soil profiles (90).

Moisture availability is an important parameter affecting soil microorganisms. Soil H_2O content affects the movement of materials, soil atmosphere, and temperature, therefore applying an important constraint on the diversity and activity of soil microorganisms (Table 2) (57, 95). Saturated soil environments often have lower microbial activity as a result of the low solubility of O_2 in H_2O and the initial rapid depletion of O_2 by aerobic microorganisms (57, 95). Furthermore, a variety of studies have reported moisture content as a principal factor affecting soil microbial communities (Table 2). Zhang *et al.* (58) noted that moisture content, which was hypothesized to affect soil proteobacteria, was the main driver of soil microbial community structure in Arctic permafrost soils. Furthermore, O_2 availability, a factor dependent on soil moisture, structures soil microbial communities by defining oxic and anoxic zones in soil profiles (96). In a typical soil profile, deeper soil horizons are generally depleted in O_2 relative to surface horizons (96). Although micro-scale regions of anoxia or O_2 enrichment may be found in surface and subsurface soils, oxic top soils generally support higher diversity due to the energetically favourable aspects of aerobic respiration (Table 2) (1, 59, 96).

Other factors that have been shown to affect soil microorganisms include temperature and salinity (57, 60, 62). Globally, temperature is known to indirectly impact soil microbial communities resulting from influences on evapotranspiration and decomposition rates, with optimal diversity and activity maintained at intermediate temperatures (promoting mesophilic activity; 20–45°C) (1, 73). Elevated soil salinity (i.e., resulting from land-use changes) has been associated with decreased enzyme activity and microbial biomass (97). Although soil microbial communities can adapt to increased salt stress, particularly from regular exposure, many microbial populations are subject to salinity-associated osmotic stress and desiccation (97, 98). Wichern *et al.* (97) observed a reduction in soil respiration, biomass, and decomposition and mineralization processes in soils as a result of increased salinity. Interestingly, this study noted that changes in these parameters were not necessarily associated with abundance changes in particular bacteria but did provide evidence that long-term exposure to high salt concentration can lead to increased energy rich metabolism mechanisms (i.e., less efficient substrate utilization) in affected microbial communities. Furthermore, in a global investigation of microbial communities from a range of environments, Lozupone and Knight (62) noted that salinity was the dominant overall factor influencing microbial diversity and community composition. Considering the contrast of important physicochemical gradients reported to affect microbial communities, this further exemplifies the importance of spatial scales when considering overall constraints on microbial diversity (81, 83).

1.1.3.2 Biotic factors

Soil microorganisms are also affected by interactions with other soil biota (1, 61). In particular, plant communities have been shown to influence belowground microbial communities (63, 74, 99). Recent work by Prober *et al.* (100) showed that although plant diversity did not appear to influence microbial alpha-diversity (variation in diversity within a sample or site), it did have an impact on measured beta-diversity (variation in diversity between sites). Furthermore, aside from obvious effects on the rhizosphere (21), dominant plant communities can affect soil microbial communities by affecting the type and quality of organic matter input into soils (63).

The close proximity of microorganisms to each other also may affect communities via microbe-microbe interactions or symbioses (101, 102). In some instances, microorganisms in soils are dependent on the products of another organism's metabolism (101, 102). This synergism is exemplified, for example, when fermentative bacteria supply methanogenic archaea with reduced C sources for energy (103). Furthermore, microbial predation, such as protozoa grazing on bacteria, which has implications for nutrient regeneration, also represents an important factor affecting soil microorganisms (104). Finally, competitive interactions, such as those involving antibiotic production by *Streptomyces* sp., may have broad implications for the overall dynamics of microbial communities (101). As a result, such fine-scale microbial interactions represent important additional factors shaping overall community structure (101).

1.2 Microbial biogeography

1.2.1 Microbial biogeographical theory

1.2.1.1 Past considerations

The study of microbial biogeography, distributions of microorganisms through space and time, has been transformed since its conception in the nineteenth century (105, 106). With the advent and recognition of bacteriology as a scientific discipline in the late 1800s, observations by microbiologists who noticed consistencies in organism types whenever specific environmental and nutritional conditions were mimicked, led to generalizations of microbial distributions (105, 107). The notion that microorganisms are selected based on the characteristics of their immediate environment, articulated by Dutch microbiologist Martinus Willem Beijerinck (1851–1931), became an accepted concept (105). He further proposed an "ecologically deterministic" concept to explain the biogeography of microorganisms based on some of his elective culturing (i.e., enrichment culturing) work where he designed selective media to isolate microorganisms with specific metabolic strategies

(105, 108). This was exemplified in an accompaniment to his 1877 thesis where he quoted Charles Darwin: “If it were possible to expose all the individuals of a species during many generations to absolutely uniform conditions of life, there would be no variability” (pp 242, *107*; *108*). This “geographic constancy theory”, derived from Beijerinck, was popularized at the Delft School of Microbiology with the well-known statement by Dutch microbiologist, Lourens G. M. Baas Becking (1895–1963), “everything is everywhere; but, the environment selects” and further reinforced by Dutch-American microbiologist Cornelis van Niel (1897–1985) (111, 112). van Niel proposed that Beijerinck and Bass Becking had developed a “unifying theory of general microbiology” (105, 113, 114). The popularization of microbial ecology in the 1960s and the growing use and impact of molecular methods in microbiology led to eventual rejection of the Beijerinck and Baas Becking theory as an unchallenged principle in microbial ecology, leading to reformation and modern revaluation of microbial biogeographical theory (105, 115–117).

1.2.1.2 Modern aspects of microbial biogeography

The debate as to whether microbial communities are shaped by current environmental factors or past historical constraints is an active area of investigation in microbiogeography (118). A useful framework, for assessing microbial biogeography, as described by Martiny *et al.* (119), employs two concepts derived from work by Swiss botanist Augustin P. de Candolle (1778-1841): biotic province and habitat-type (120). A biotic province represents a region where historical features shape the overall biological composition; habitat type refers to a region where the contemporary abiotic and biotic features shape the environment (118, 120). For example, a biotic province may represent the Australian continent where many macroorganisms are found as a result of past historical constraints, such as geographical isolation, whereas a habitat type includes the dry scrubland environments located throughout the continent (118, 120).

Defining the aforementioned concepts allows the identification of four different hypotheses (Figure 2) that address microbial biogeography (118). One hypothesis suggests that microorganisms are randomly distributed through space and time, that is, microorganisms do not have defined biogeography (e.g., one habitat-type and one province). This hypothesis is invalidated by many studies suggesting otherwise (Table 3), but remains useful as a null hypothesis when assessing microbial biogeographical principles (118, 121). A second hypothesis suggests that the immediate contemporary environment shapes biogeography (e.g., multiple habitat-types in one province) (118). This effectively represents the Baas-Becking hypothesis (see previous section), where high dispersal capabilities of microorganisms erase the effect of historical constraints on biogeography (118). In contrast, a third hypothesis underlines the importance of historical over contemporary constraints

(e.g., multiple provinces with one habitat-type) in shaping microbial distributions (118). The fourth hypothesis suggests that the interplay between historical and contemporary factors ultimately shape the distributions of microorganisms (e.g., multiple habitat-types and provinces) (118). It is important to note that the Baas-Becking hypothesis is, to some extent, useful for explaining biogeographical patterns amongst microorganisms but there is growing evidence suggesting otherwise; past historical constraints also play a role in shaping microbial distributions (118, 122). As suggested by Martiny *et al.* (118) and others, elucidating the interplay of contemporary and historical environmental constraints helps distinguish between factors influencing microbial biogeography, which is a necessary and fundamental aspect of microbial ecology (118, 123).

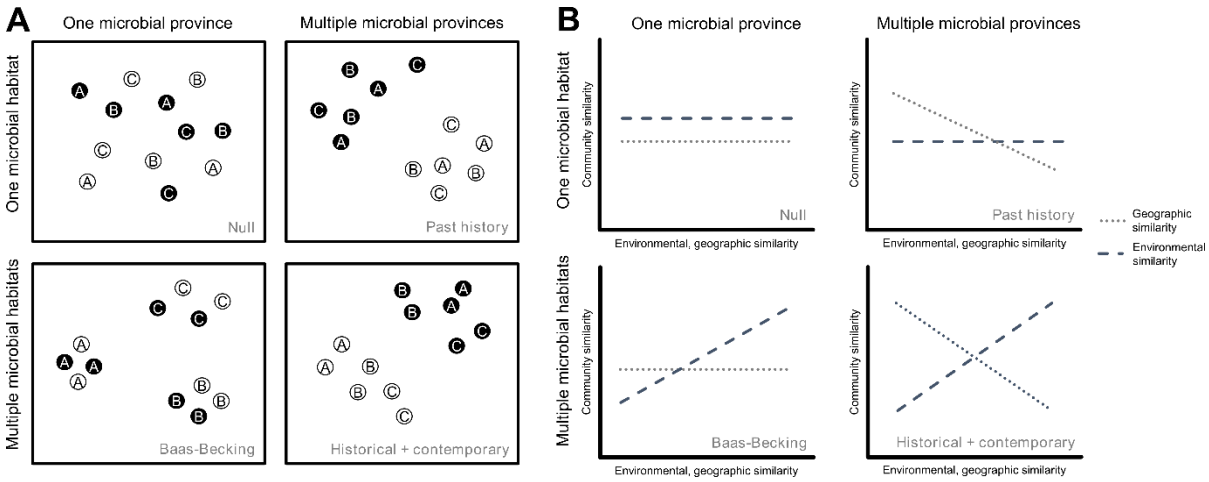

Figure 2: Summary of the four alternative hypotheses addressing microbial biogeography. A) Circles represent independent samples (filled and unfilled samples represent different locations), letters (A, B, C) represents predefined habitat types. Axes are arbitrary but samples that appear closer together are interpreted to have similar microbial assemblages. B) Dashed lines represent environmental similarity, dotted line represents geographic similarity (see section 1.2.1.2 for explanation; adapted from Martiny *et al*. (118)).

Table 3: Summary of studies that have found evidence for microbial biogeographical patterns (adapted from Martiny *et al.* (118)).

Organisms	Scale (km)	Habitat	Method of analysis	Correlated with
Pseudomonads	20000	Soil	BOX-PCR[a] isolation	Linear distance
3-CBD[b] bacteria	20000	Soil	ARDRA[c] isolation	
Aerobic, anoxygenic phototrophs	20000	Marine	Dissociation curves	Latitude
SAR11 bacteria and archaea	13000	Marine	16S rRNA[d]/ITS[e] sequencing	Depth
Green sulfur bacteria	8000	Lakes	16S rRNA gene sequencing	Continental divide
N_2-fixing bacteria	700	Desert crusts	Sequencing and T-RFLP[f] of *nifH*[g] and 16S rRNA genes	Mature versus poorly developed crusts
Crenarchaeota	200	Soil	PCR-SSCP[h] of 16S rRNA genes	At small scales, distance
Crenarchaeota	200	Soil	PCR-SSCP of 16S rRNA genes	Rhizosphere versus bulk soil
Bacteria	50	Marine	DGGE[i] 16S rRNA genes	Ocean front
Bacteria	35	Marine	DGGE 16S rRNA genes	Depth and ocean front
Bacteria	15	River plume	DGGE 16S rRNA genes	River-marine transition
Bacteria	5	River plume	DGGE 16S rRNA genes	Salinity
Bacteria, archaea, and eukaryotes	3	Salterns	DGGE, T-FRLP, RISA[j]	Salinity
Pseudomonas cepacia	3	Soil	Isolate allozymes	Vegetation
Bacteria and eukaryotes	1	Soil	RNA hybridization	Cultivation history
Gram-negative bacteria	0.8	Soil	sole carbon source	Latitude
Microorganisms	0.2	Groundwater	RAPD[k]	Oxygen zonation
Microorganisms	0.1	Agricultural soil	AFLP[l]	
Bacteria and archaea	0.02	Lakes	DGGE of 16S rRNA genes	Depth
Bacteria	0.01	Drinking warer	T-RFLP of 16S rRNA genes	Bulkwater versus pipe biofilm
Purple non-sulphur bacteria	0.01	Fresh marsh	BOX-PCR isolation	Linear distance
Bacteria	0.01	Soil	RFLP[f] of 16S rRNA genes	
Microorganisms	0.002	Salt marsh	RAPD	Marsh elevation

[a]BOX primer sets used for repetitive extragenic palindromic PCR genomic fingerprinting; [b]3-chlorobenzoate degrading; [c]amplified ribosomal DNA restriction analysis; [d]component of the 30S small subunit ribosome in prokaryotes; [e]intergenic transcribed spacer; [f]restriction fragment length polymorphisms; T-RFLP, terminal RFLP; [g]bacterial gene that encodes for nitrogenase; [h]polymerase chain reaction-single strand conformational polymorphism; [i]denaturing gradient gel electrophoresis; [j]ribosomal intergenic spacer analysis; [k]random amplified polymorphic DNA; [l]amplified fragment length polymorphism

1.2.1.3 Processes shaping microbial biogeography

Fundamentally, microbial distributions are shaped by the interplay between dispersal capabilities, which are also known as colonization events, and subsequent diversification or extinction (118). Dispersal rates of microorganisms affect biogeography as a result of changes in gene flow through horizontal gene transfer or even sexual reproduction (118, 123). For microorganisms, colonization of new environments is primarily the result of passive rather than active transport, especially in soil habitats where microorganisms are limited by pore space size and the suite of cell- and particle-surface interactions (e.g., electrostatic interactions of charged particles, van der Waals forces, and hydrophobic interactions) (124). Although motile microorganisms may, over time, travel large distances, increasing soil heterogeneity would likely increase selective pressures, therefore increasing the chance of speciation (118). Passive diffusion of soil microorganisms also likely does not play a large role in colonization due to limited overall mobility rates in soils (1, 125, 126). As a result, advective transport of microorganisms in soils represents the primary way that microorganisms move, likely driving the colonization of deeper soil environments. Airborne microorganisms may colonize new surface soils through passive transport from the atmosphere and subsequent deposition in new environments (118, 125, 127). Recent modelling work by Wilkinson *et al.* (127) found that dispersal capabilities of microorganisms over a one year period was vast for smaller organisms as a result of greater abundances and longer atmospheric residence times.

Contrasting the homogenizing potential of colonization events, diversification processes (e.g., mutation, genetic drift, various selective pressures) contribute to and maintain microbial biogeographic patterns (118, 123). Evidence suggests that the short generation times of many microorganisms ultimately allows rapid genetic divergence and therefore increased potential for defined biogeographic patterns (118, 128, 129). Disturbance of the physical environment may also alter biogeography of organisms by changing selective pressures (130). Evidence supporting the intermediate disturbance hypothesis, which suggests that diversity of microorganisms is highest when moderate ecological disturbance of a system is achieved, has been observed for bacterial communities (130). A study showed clear evidence of differential selection occurring when the physical environment was disrupted, thus generating and altering the biogeographical patterns of bacterial communities (130). Overall, it is likely the complex interplay between colonization (gene flow) and diversification processes that is a key factor underlying patterns of global microbial biogeography.

1.2.2 Patterns of soil microbial biogeography

1.2.2.1 Distance decay

Distinct patterns of soil microorganisms across environment types has been observed (118, 123). In many instances, biogeographical patterns of microorganisms can be considered as a distance-decay relationship, where community similarity decreases with increasing geographic distance (118, 123, 131). This ultimately implies that community differences are spatially auto-correlated. An example of a distance-decay relationship operating to influence biogeographical patterns across latitudinal gradients is exemplified in a study conducted by Cho and Tiedje (132), where *Pseudomonas* strains showed evidence of endemism and non-random distributions across a large (>100 km) transect of soil. An arguably more prominent example is the taxa-area relationship, a well-known extension of the distance-decay relationship observed for macroorganisms, demonstrated for bacteria across cm to m distances (121). Interestingly, environmental heterogeneity was attributed to be the main driver for this relationship rather than geographic distance, which further stimulates questions as to whether contemporary or historical constraints are more important in shaping microbial biogeography and whether these trends are consistent across different environment types (118, 121, 123).

1.2.2.2 Latitude

Although evidence for microbial biogeographical patterns have been observed across latitudinal gradients in soils (Table 3; *125*, *126*), a growing number of studies suggest that latitude does not play a significant role (74, 86, 135). For example, Fierer and Jackson (81) reported that across North and South America, bacterial diversity was affected by soil pH rather than latitude. In addition, Neufeld and Mohn (86) observed that latitude likely did not play a role in shaping bacterial community composition across a gradient spanning boreal and Arctic biomes. Yergeau *et al.* (135) found that the effect of geographic distance and vegetation were more important factors affecting terrestrial soil bacterial communities than latitudinal changes in a southern polar transect spanning greater than 3200 km. Although these studies highlight that latitude does not appear to greatly affect microbial biogeography, considerably more research is necessary to confirm these observations (136).

1.2.2.3 Depth

Microbial biomass and diversity generally decrease with depth, reflecting multiple changes in soil physicochemistry (25, 88, 137, 138). Microorganisms within deeper soil horizons are considered to play important roles in soil development and biogeochemical cycling (e.g., such as C

sequestration), especially due to their proximity to parent material (139), storage potential, and lower turn-over rates than surface soil microbiota (28, 89, 140). However, the biogeography of microorganisms throughout soil profiles is largely under-characterized with relatively few studies explicitly examining the effect of depth (25). A recent important contribution to depth-specific biogeography was conducted by Eilers *et al.* (25) where high throughput (HTP) sequencing of 16S rRNA genes from soils to depths of up to 180 cm revealed distinct trends in major groups of bacteria and archaea paving the way for future studies investigating the consistencies in these trends. Table 4 lists studies examining depth specific shifts in microbial communities.

Table 4: Collection of literature examining soil microbial diversity with an emphasis on the impact of depth on soil microbial communities.

Depth (cm)[a]	Method of analysis	Microbial group(s)	Environment type(s)	Reference
300	16S[b,‡] rRNA and functional genes	Bacteria, archaeal	Not specified[*]	(4)
200	PLFA[c]	Bacteria, fungi, protozoa	Not specified[*]	(88)
180	16S rRNA genes[‡]	Bacteria, archaeal	Forest/forest riparian zone, meadow	(25)
170	PLFA	Bacteria	Agricultural	(137)
149+	PLFA, DGGE[d]	Bacteria, fungi	Forest	(2)
118	16S rRNA genes	Bacteria, Archaea	Tundra permafrost	(141)
100+	16S rRNA genes[‡]	Bacteria, Archaea, Fungi	Arctic tundra	(142)
80	PLFA	Bacteria, Fungi, other[g]	Grassland	(143)
70	Cultures	Fungi	Agricultural	(144)
50	Cultures/microscopy	Fungi	Forest	(145)
40	16S rRNA genes	Bacteria, Archaea	Arctic tundra	(146)
20	16S rRNA genes	Bacteria	Subarctic tundra	(147)
20	DGGE	Bacteria	Grassland	(148)
18	T-RFLP[e]	Archaea	Forest	(149)
N/A[f]	16S rRNA genes[‡]	Bacteria	Grassland	(150)

[a]Maximum depth analyzed; [b]component of the 30S small subunit ribosome in prokaryotes; [c]phospholipid-derived fatty acid; [d]denaturing gradient gel electrophoresis; [e]terminal restriction fragment length polymorphism; [f]conducted analyses based on specific soil horizons; [g]general assessment of other non-specified microorganisms; [‡]analysis used HTP sequencing; [*]Environment type not explicitly specified

Furthermore, although there are reports of specific microbial populations associated with deeper soil horizons, an understanding of their ecology remains largely unknown (25). Bacteria, including organisms from the phyla *Acidobacteria*, *Actinobacteria*, *Bacteroidetes*, and *Verrucomicrobia*, as well as archaeal taxa, particularly from the *Thaumarchaeota* (formerly *Crenarchaeota*) phylum, are known to be distributed differentially throughout soil profiles to >1 m in

depth (4, 25). The presence of such microorganisms provides insight into the biogeochemical processes that are likely occurring throughout soil profiles. For example, several members of the phylum *Acidobacteria* associate with low-pH microenvironments of soil aggregates, located closer to surface acidifying processes (see section 1.1.3.1) (25). *Verrucomicrobia* are ubiquitous within soils, yet have been shown to increase in abundance at mid-depths in soil profiles, likely due to their putative oligotrophic nature and the known decrease of C with increasing depth (25). Furthermore, taxa from the phylum *Bacteroidetes* are hypothesized to be copiotrophic due to their relative abundance in surface soils (e.g., rhizosphere), where high concentrations of labile organic C are usually present (25). Depth-specific distributions of NH_4^+-oxidizing thaumarchaeotes suggests their important role in the biogeochemical cycling of N in multiple horizons of soil; distinct archaeal 16S rRNA genes were found to be associated with individual soil horizons (4). These archaea are likely important contributors to chemolithoautotrophic nitrification and are also hypothesized to be potentially involved in C metabolism and amino acid assimilation (4, 149, 151). Although several soil microorganisms demonstrate distinct abundance distributions throughout soil profiles, as previously mentioned, details of depth-specific biogeography and biogeochemical significance are still largely unknown (25).

1.2.3 Importance of microbial biogeography

Studying soil microbial distributions and the factors affecting these distributions, benefits a variety of disciplines (152). Aside from broadening our understanding of the diversity and breadth of microorganisms on Earth, an understanding of microbial biogeography offers additional benefits to soil fertility, productivity (i.e., efficiency in nutrient cycling), epidemiology, biocontrol, bioremediation, forensics, and bioprospecting (152). For example, biogeography is used for predicting the response and localization of human, animal, and plant pathogens (152, 153). Furthermore, by understanding biogeographical principles, the fate of biocontrol- or bioremediation-organisms can be better predicted (152). Finally, studying microbial biogeography can provide useful insight into where valuable and beneficial (e.g., antibiotic-producing organisms) microorganisms are most likely to be found in the environment (152).

1.3 Characterizing microbial communities

1.3.1 Some prerequisites

Complex interactions between populations, which are collections of individuals of the same species that live in a specific geographic area, are necessary for the majority microorganisms in soil

environments (154). As result, it is important to study the microbial community, which is the collection of microbial populations that exist within a given area, to understand the overall roles of microorganisms in soils. Unlike eukaryotic organisms that are more compatible with the traditional species concept, the prokaryotic species definition is much less well defined (154, 155). Prokaryotic species are usually defined as operational taxonomic units (OTUs) based on phylogenetic similarity (154, 155). In many cases, microbial communities can be studied based on their taxonomic diversity, which takes into account richness (e.g., number of species) and evenness (the relative abundance of species) (154, 155). The diversity of microbial species found in a community provides important information regarding the overall community function in soil systems.

1.3.2 Culture-dependent methods

Characterizing microbial communities from environmental samples has traditionally employed culturing techniques on a variety of media types designed to grow and isolate specific groups of microorganisms (156). Comparisons of direct cell counts from soil samples with the observed colony forming units suggests that less than 1% of soil microorganisms are readily cultivated, leading to a reliance on culture-independent techniques for microbial community analysis (5, 24). Furthermore, a variety of soil-related microbial community analyses examine whole soil characteristics, such as overall microbial biomass and activity (e.g., respiration rates and enzyme activity) (156). These process-level measurements have been used for assessing community stability as well as soil function and quality (see section 1.1.3.1) (156). Another culture-dependent method is community-physiological profiling (156). Although this technique is useful for assessing overall soil function and quality, it lacks the ability to constrain which specific community members are responsible for certain functions (156).

1.3.3 Culture-independent methods

The limitations of culture-dependent methods represents an important motivation for using culture-independent methods for characterizing microbial communities. Although other methods exist, two principle molecule types are useful in culture-independent analyses: PLFAs and nucleic acids (156). Because different types of microorganisms have different profiles of fatty acids present in the cell membrane, analyzing and characterizing these components from the environment can be useful for identifying and assessing the soil microbial community (156). This technique, known as PLFA analysis, has allowed assessment of community structural changes resulting from shifts in soil quality (156). However, because some microorganisms share similar PLFA profiles, taxonomic resolution is difficult to detect when using PLFAs as the main technique for analysis. This, in

combination with the low-throughput and lengthy analysis process, often makes studies utilizing nucleic acids, DNA and RNA, much more practical (156).

1.3.3.1 Metagenomics and the 16S rRNA marker gene

Within the last decade, advances in DNA sequencing technology have enabled HTP analysis of microbial diversity in environmental and host-associated samples. In particular, metagenomics—analysis of genomes from environmental samples—have revealed significant insight into the functional potential of microbial communities, allowing for the discovery of novel enzymes and species (157). For example, metagenomic analyses of bacterioplankton and archaea from marine environments led to the discovery of proteorhodopsin-based photoheterotrophy and ammonia-oxidizing archaea, respectively (158, 159). Although metagenomic research has seen rapid cost reduction, the overall price of deep metagenomic sequencing for multiple samples remains prohibitively expensive (160, 161). As a result, the vast majority of microbial ecological studies utilize marker genes for the detection and phylogenetic characterization of environmental microorganisms from larger numbers of samples in field-scale community comparisons (161).

The majority of studies analyzing bacterial and archaeal communities in environmental samples utilize a component of the 30S small subunit rRNA gene (16S rRNA in bacteria and archaea). However, other universal marker genes, such as those encoding RNA polymerase subunits, sigma factors, DNA gyrases, and heat-shock proteins, can also be used (161–163). Besides the overall cost reduction in analysis, the 16S rRNA gene remains the current "gold-standard" in microbial ecological research because the gene itself is universal, maintains structural conservations as well as distinct regions of genetic variability, and is resistant to horizontal gene transfer (HGT) (161). In addition, there are considerably more well-annotated databases (e.g., SILVA and GreenGenes) for 16S rRNA genes than other marker genes allowing comprehensive analysis of taxa in different environments (162, 164). The obvious limitation of 16S rRNA approaches is the lack of functional and metabolic information provided from these datasets because only a single or a few universal genes are used (161). Furthermore, because DNA is stable for relatively long periods of time (165), it is not necessarily known whether the observed taxa are active in the environment, which may limit the interpretation of the results. This can, to some degree, be circumvented using RNA (166) to analyze active microorganisms or discriminative agents (e.g., propidium monoazide) (167) to prevent detection of dead cells. However, there are other inherent issues with these methods, including the instability of RNA outside of cells and the efficacy of various chemicals (166, 167).

1.3.3.2 HTP sequencing and bioinformatics

HTP sequencing of 16S rRNA genes has been impactful for microbial ecology, allowing in-depth analysis and comparison of hundreds of samples from various environment types in a single sequencing run (168). There are a variety of HTP sequencing technologies commercially available, each utilizing different variations of three principle steps: nucleic acid template preparation, sequencing and imaging, and bioinformatic analysis (168). The variation and unique strategies taken to address these three steps ultimately defines the sequencing technology. Some of the well-known sequencing platforms include Illumina, 454 pyrosequencing, the SOLiD platform (Applied Biosystems), and Pacific Biosciences (PacBio), each with specific advantages and disadvantages (summarized in Table 5) (168).

Table 5: Characteristics of example HTP sequencing platforms (169, 170).

Platform	Template preparation	Chemistry	Max read length (bp)	Error type	Overall error rate (%)	Performance notes
454 (all models)	emPCR	Pyrosequencing	1000	Indel	1	Long reads High error in homopolymer[a] repeats
Illumina (all models)	Solid-phase	Reversible terminator	300	Substitution	~0.1	Widely used capabilities Library preparation technically challenging and requires control of template concentration
Ion-torrent all chips	emPCR	Ion semiconductor	400	Indel	~1	Lower-cost instrument High error in homopolymer repeats
SOLiD[b] – 5500xl	emPCR	Cleavable probe; sequencing by ligation	75	A-T bias	~5	Inherent error correction (via two-base encoding) Long run times
PacBio RS	Single molecule	Real-time	20000	CG deletions	~15	Greatest potential for long read analysis Very high error rates

[a]Regions of identical bases; [b]Sequencing by oligonucleotide ligation and detection

Furthermore, technological advances have allowed access to a "third" generation of sequencing platforms, particularly the Oxford Nanopore system, which uses nanopore and single-molecule detection to sequence DNA (168, 170, 171). Ultimately, improvements in DNA sequencing strategies will allow for increasingly detailed analyses of microbial communities from complex environments, enabling robust analyses of the distributions of microorganisms on Earth (168, 170).

The immense amount of data generated from HTP sequencing studies requires bioinformatic techniques for processing and analysis. Within 16S rRNA gene studies, paired-end sequencing is often used to generate high-quality reads for subsequent downstream analysis (168, 172). Paired-end sequences are generated by sequencing both the forward and reverse strands of a DNA fragment

(173). Merging the two sequences not only extends the length of the overall assembled sequence but also allows for base error correction thereby producing higher quality data (173). Read-merging pipelines are often an important step in data processing in HTP 16S rRNA gene studies (172, 173). Although there are many methods for merging paired-reads including COPE (174) and PEAR (173), the PANDAseq algorithm is particularly useful because it includes a quality filtering component (172). Briefly, PANDAseq uses a three step process to merge paired-reads which include: 1) determination of the location of primers, 2) finding optimal overlap between paired-reads, and 3) correcting for length and quality based on Q-scores (Phred quality scores; represents the probability of an incorrect base call during sequencing) (172).

An important component of 16S rRNA gene studies is sequence clustering at a predefined threshold and OTU table generation using predefined clustering algorithms (e.g., CD-HIT, UCLUST, UPARSE; see section 1.3.3.2.2). In microbiology, a typical sequence similarity cut-off for the 16S rRNA gene is at 97%, which is approximately equivalent to species-level taxonomic identification (154, 175). This cut-off is based on DNA-DNA hybridization experiments; 70% genome-genome similarity of different strains corresponds to approximately 97% sequence identity across the entire 16S rRNA gene (154, 175). Although this cut-off is somewhat arbitrary for short amplified fragments of the 16S rRNA gene, it is still useful for reducing the complexity of datasets when considering alpha- and beta- diversity (154, 175). With increasing development and cost-reduction of metagenome sequencing, full genomes of organisms collected from environmental samples likely represents the next-step in microbial ecological approaches (157).

1.3.3.2.1 Classifying microorganisms from marker genes

In microbial ecology, two general approaches are used to characterize microbial communities based on 16S rRNA gene sequences: taxonomy-based and OTU-based methods (175, 176). The OTU-based methods, which rely on sequences clustered at a specified similarity threshold, are commonly used to group 16S rRNA genes and maintain significant advantages over taxonomy-dependent methods, which rely on databases of annotated sequences from previously described representative organisms (175). The main advantage of OTU-based methods is the lack of data loss because sequences are assembled into phylotypes, groups based on observed sequence similarity, independently of a reference taxonomy (175). However, limitations of OTU-based approaches include the potential inflation of observed phylotypes due to sequencing errors, differential evolution of 16S rRNA genes that may alter the identity thresholds for OTUs, and the appropriate biological interpretation of OTUs (175). Taxonomy-based methods help circumvent these issues (175). Choosing appropriate sequence clustering methods is an important step prior to data exploration.

1.3.3.2.2 Methods in sequence clustering

There are a variety of methods for clustering 16S rRNA genes including CD-HIT, UCLUST, and UPARSE algorithms (175). The CD-HIT and UCLUST approaches are relatively similar, differing primarily in the way sequences are sorted and mapped to representative sequence clusters (175). CD-HIT sorts sequences by decreasing length (177), whereas UCLUST typically sorts reads via decreasing abundance (178). Furthermore, UCLUST implements USEARCH (software that contains multiple algorithms for a variety of sequence processing tasks) as a subroutine for sequence assignment to clusters (178). The USEARCH software contrasts with the CD-HIT assignment algorithm in that it uses word counts (e.g., k-mers; short nucleotide fragments of length, k) to prioritize dataset searches rather than using them to exclusively estimate sequence identity from k-mer matched counts (177, 178). In either method, initial "seed" sequences are selected and each subsequent query sequence in the database of 16S rRNA gene reads is compared to the initial sequence. Sequences matched to the seed sequence within the predetermined quality threshold, which is typically 97% sequence similarity, become members of the cluster. Otherwise, they become new seed sequences (177, 178). Similar to UCLUST, UPARSE defines clusters by sorting sequences, by decreasing abundance, after first removing duplicated sequences. In contrast to CD-HIT and UCLUST, UPARSE assesses the addition of sequences that do not initially fall into the pre-existing clusters by invoking the UPARSE-REF greedy algorithm, which looks for the most parsimonious model of a query sequence to infer a novel cluster or chimeric characteristics (179). The accuracy of the UPARSE method was shown to be markedly better than other clustering methods (179).

1.4 Research description

1.4.1 Research overview

Because microbial communities play dominant roles in the biogeochemical cycling of Earth's elements, studying microbial biogeography has important implications for ecosystem stability, human health, and global climate (5, 24–26). Soils contain the majority of Earth's microbial diversity, with most species existing at low relative abundance (i.e., the "rare biosphere") (5). Studying soil microbial communities, with particular emphasis on the physicochemical state of the system, land-use histories, and the relatively understudied role of depth on their diversity, is critical for understanding soil function (25).

Previous soil microbial ecological research has focused on the analysis of surface soils and the latitudinal changes across particular sites of interest (25, 88, 121). However, it is well known that many important biogeochemical processes, such as C and N cycling, occur differentially at distinct

soil depths. Furthermore, the problem remains that depth-dependent responses of microbial communities are an understudied aspect of soil microbial diversity and function (25). Succession and changes in microbial community dynamics with depth, particularly following anthropogenic land-use alterations, is relatively uncharacterized, with most previous studies examining surface soils and seasonal environmental changes (61, 148). Comparing uncharacterized depth-dependent responses to changing land-use regimes represents a largely unexplored aspect of soil microbial ecology. Furthermore, relating these trends and attempting to address the underlying factors affecting these distributions remains a critical and common goal amongst microbial ecologists (123).

Using a dataset of 16S rRNA genes generated via HTP sequencing of soils from the ***rare*** Charitable Research Reserve (Cambridge, Ontario), this research assessed the effects of depth and land-use type in shaping soil microbial community distributions. The ***rare*** Charitable Research Reserve was ideal for this study because of the unique land variations (e.g., old/mature growth forests and active/decommissioned agricultural fields) as well as a consistent geologic background that allowed more consistent comparison of soils across depths throughout the property. Furthermore, the exploration of functional differences throughout soil profiles and across land-use types was investigated using a dataset of predicted metagenomes to further assess how bacterial communities are shaped by depth and land-usage characteristics.

1.4.2 Objectives and hypotheses

The objectives of this study were to investigate and explore the effects of 1) depth and 2) land-usage regimes on soil bacterial communities, as well as 3) assess the functional characteristics of collected soils in relation to soil depth and land-use type. The hypotheses were that microbial species diversity will decrease with depth and that subsurface soil microbial communities will show patterns of depth variation similar in magnitude to variations observed for important global gradients such as those influenced by pH (25, 53). Furthermore, considering the consistent geology across the ***rare*** Charitable Research reserve, surface soils are likely governed largely by contemporary physicochemical factors as a result of stronger environmental gradients, such as surface inputs and anthropogenic land-use alterations, with deeper soils affected by historical influences (e.g., parent material) and depth-specific soil properties. Legacies left over from agricultural practices are also hypothesized to affect bacterial communities undergoing succession. This research also tested the hypothesis that consistent indicator species will associate with depth and land-usage. Furthermore, it was hypothesized that changes in the overall functional characteristics would show distinctive patterns mimicking bacterial community shifts observed with depth and across land-use types further highlighting distinct community structures throughout subsoil environments.

Chapter 2
Materials and Methods

2.1 Site characterization

Soil sampling was conducted at the ***rare*** Charitable Research Reserve (Cambridge, Ontario, Canada; Figure 3). Founded in 2001, the ***rare*** Charitable Research Reserve is a 900 acre region of diverse land usage, ranging from old-growth forests to active agricultural fields, located within the Grand River Watershed (Figure 3). The Grand River borders the northern edge of the property. The majority of the property exists on fluvio-glacial materials deposited from the last ice age (~13,000 years ago) as well as upper middle Silurian dolostone bedrock. Outcrops of dolostone (alvars) can be observed on the north-eastern corner of the property. Soils across the property are mostly Luvisolic (i.e., well-developed eluviated clay-rich horizons), with the alvar areas particularly on the north-east side of the property (alvars) mostly classified as Brunisolic (i.e., less clay accumulation in B-horizon) (10). Seven distinct sampling sites were chosen across the property: one old-growth forest (Indian Woods), two mature forests (Hogsback, Cliffs and Alvars), three decommissioned agricultural fields (decommissioned in 2003, 2007, and 2010), and one active agricultural field (Preston Flats) (Table 6). In this study, the term "field" collectively refers to grasslands (i.e., decommissioned agricultural fields, alvars) and the agricultural site. The majority of forested sites across the ***rare*** Charitable Research Reserve date back to over 200 years. Old-growth forests are here defined as largely undisturbed areas for more than 240 years. Mature forests are classified as having been forests for 100–200 years, with the Cliffs and Alvars site being the youngest and most subjected to disturbance (e.g., cattle grazing and selective cutting in some areas).

All forested sites were dominated mainly by a mix of northern hardwood and Carolinian tree species (i.e., *Quercus rubra*, *Quercus alba*, *Quercus velutina*, *Acer saccharum*, *Acer nigrum*, *Fagus grandifolia*, *Pinus strobus*, *Fraxinus americana*, *Tilia* sp., and *Carya* sp.) with a diverse ground cover including ferns, shrubs, and forbs (Table 7). Decommissioned fields were dominated primarily by grasses and forbs, including *Daucus carota*, *Apocynum* sp., *Solidago* sp., *Dipsacus fullonum*, thistle, with some larger tree species including *Acer saccharum*, *Populus* sp., and *Juglans nigra* also present (Table 7). The Preston Flats agricultural site is a non-tilled field that has been subjected to pesticide and fertilizer use since 2002 (180). Urea is used as a N fertilizer and is applied annually across the Preston flats at a rate of ~291 kg ha^{-1} (180). Since 2011, the Preston Flats field has been used solely to grow *Zea mays* (corn) but had also been subjected to crop rotation with *Glycine max* (soy) between 2002 and 2011. (180). Agricultural practices across the ***rare*** Charitable Research Reserve date back

to the mid-1800s and have included areas for pasture lands, field crops, and horticultural crops. Furthermore, the general history of previous estate owners have suggested that lands were treated with conservative agricultural practices (e.g., using sloped lands for pasture fields and farming only to edge of drip lines) (180).

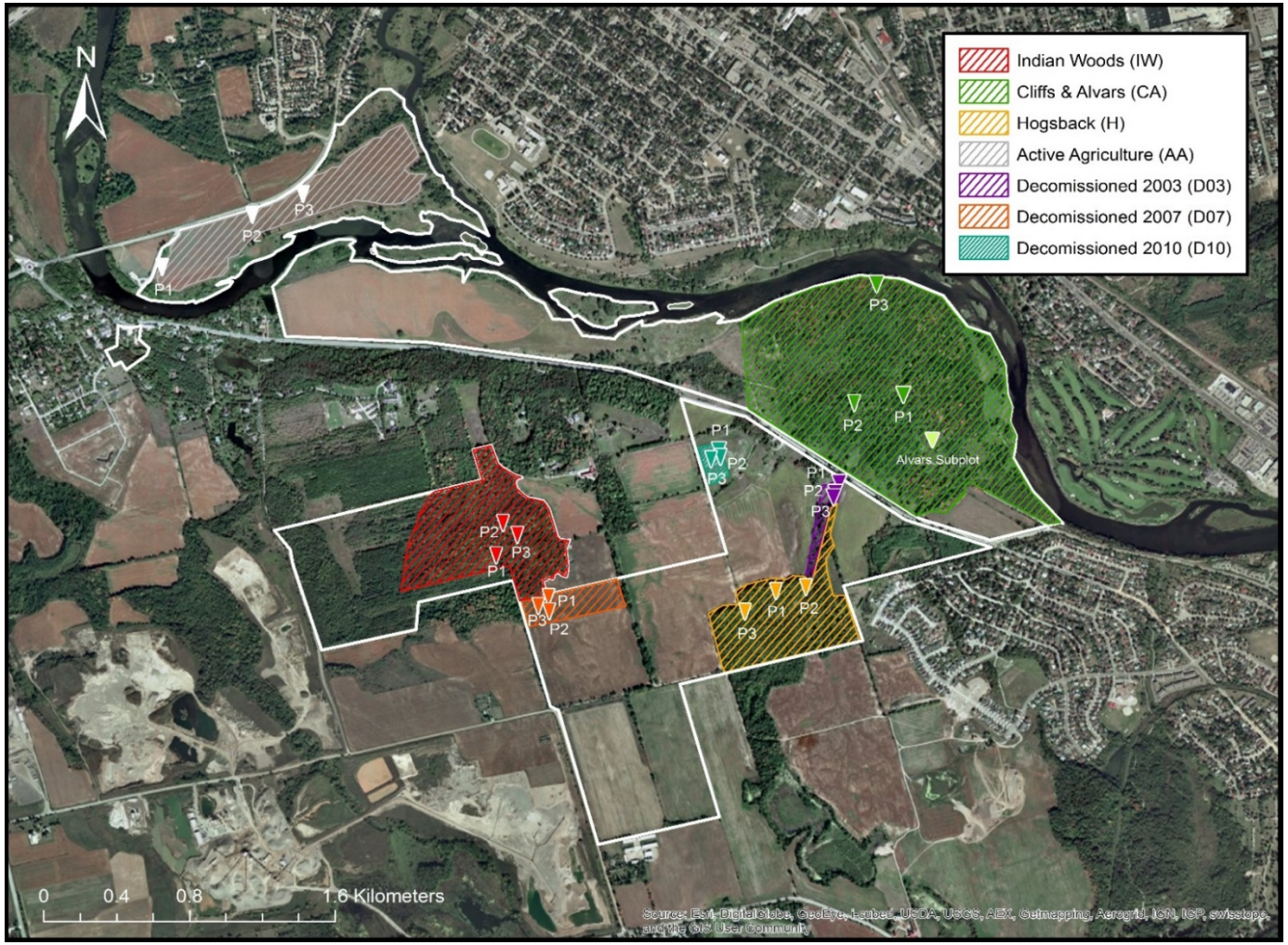

Figure 3*:* Map of the ***rare*** Charitable Research Reserve including the seven sampling sites. Sampling was conducted on one active agricultural field, three decommissioned agricultural sites (2003, 2007, and 2010), one old-growth forest (Indian Woods), and two mature forests (Hogsback and Cliffs & Alvars). Subplots ("P") for each site are indicated by vertical arrows. At the Cliffs & Alvars site, a fourth subplot was designated in order to sample alvar-type soils exclusively (designated by the light green vertical arrow).

Table 6: Summary of site information.

Site name	ID	Site type	Plot	Sampling date	Coordinates	Soil type[a]	Plant cover[b]
Active Agriculture	AA	Agriculture	1	09/10/13	43°23'08.8980" N, 80°23'00.1824" W	L	C
			2	09/10/13	43°23'17.0340" N, 80°22'43.7088" W	L	C
			3	09/10/13	43°23'20.3028" N, 80°22'34.8276" W	L	C
Alvars	Al	Alvars	1	08/16/13	43°22'46.3476" N, 80°20'43.0224" W	L	F/G
			2	08/16/13	43°22'46.5054" N, 80°20'42.7770" W	L	F/G
			3	08/16/13	43°22'46.4268" N, 80°20'42.7452" W	L	F/G
Cliffs and Alvars	CA	Mature Forest	1	08/16/13	43°22'52.5108" N, 80°20'47.6016" W	Br	W
			2	08/16/13	43°22'52.4208" N, 80°20'54.2544" W	Br	W
			3	08/19/13	43°23'07.5948" N, 80°20'54.9456" W	Br	W
Decommissioned Agriculture 2003	D03	Field	1	08/19/13	43°22'40.2492" N, 80°20'59.3916" W	Br	F/G
			2	08/20/13	43°22'39.4242" N, 80°21'00.4716" W	Br	F/G
			3	08/20/13	43°22'38.7876" N, 80°21'00.4860" W	Br	F/G
Decommissioned Agriculture 2007	D07	Field	1	08/15/13	43°22'24.5712" N, 80°21'51.7212" W	L	F/G
			2	08/15/13	43°22'23.0448" N, 80°21'51.4476" W	L	F/G
			3	08/16/13	43°22'23.6208" N, 80°21'53.4060" W	L	F/G
Decommissioned Agriculture 2010	D10	Field	1	08/22/13	43°22'44.0076" N, 80°21'20.6676" W	L	F/G
			2	08/23/13	43°22'43.6296" N, 80°21'20.8260" W	L	F/G
			3	08/23/13	43°22'43.2372" N, 80°21'22.6476" W	L	F/G
Hogsback	H	Mature forest	1	08/13/13	43°22'24.7656" N, 80°21'11.3112" W	L	W
			2	08/14/13	43°22'25.8780" N, 80°21'05.6664" W	L	W
			3	08/14/13	43°22'22.7424" N, 80°21'16.0560" W	L	W
Indian Woods	IW	Old-growth forest	1	08/09/13	43°22'30.0864" N, 80°22'00.0552" W	L	W
			2	08/12/13	43°22'33.7404" N, 80°21'58.6548" W	L	W
			3	08/13/13	43°22'31.6452" N, 80°21'56.3256" W	L	W

[a]L= luvisol, Br = brunisol; [b]Main plant cover; C = corn, F/G = forbs/grasses, W = woody

Table 7: Dominant plant species of sampled sites across the ***rare*** Charitable Research Reserve.

Site name	ID	Plot	Major plant species
Active Agriculture	AA	1	*Zea mays* (Corn)
		2	
		3	
Alvars	Al	1	*Daucus carota* (Queen Anne's lace), *Solidago* sp. (Goldenrod), *Rhamnus* sp. (Common Buckthorn), *Fraxinus* sp. (Ash tree), *Berberis thunbergii* (Japanese barberry), *Juniperus* sp. (Juniper), *Malus* spp. (Crabapple), grasses
		2	
		3	
Cliffs and Alvars	CA	1	*Fagus grandifolia* (American Beech), *Acer saccharum* (Sugar Maple), *Toxicodendron radicans* (Poison Ivy), *Prunus serotina* (Wild black cherry), *Ostrya virginiana* (American hophornbeam)
		2	*Fagus grandifolia* (American Beech), *Acer saccharum* (Sugar Maple), *Toxicodendron radicans* (Poison Ivy), *Allium tricoccum* (Wild leek), *Podophyllum peltatum* (Mayapple), *Fraxinus* sp. (Ash tree)
		3	*Fagus grandifolia* (American Beech), *Acer saccharum* (Sugar Maple), *Quercus rubra* (Red Oak), *Quercus alba* (white Oak)
Decommissioned Agriculture 2003	D03	1	*Apocynum* sp. (Dogbane), *Rubus* subsp. *Idaeobatus* (Wild raspberry), *Populus* sp. (Poplar), *Juglans nigra* (Black walnut), *Vitis* sp. (Wild grape), *Rhus* sp. (Sumac)
		2	*Daucus carota* (Queen Anne's lace), *Cirsium* sp. (Thistle), *Eutrochium* sp. (Joe-Pye weed), *Apocynum* sp. (Dogbane), *Anethum graveolens* (Dill), *Dipsacus fullonum* (Wild teasel)
		3	*Daucus carota* (Queen Anne's lace), *Eutrochium* sp. (Joe-Pye weed), *Vitis* sp. (Wild grape), *Apocynum* sp. (Dogbane), *Anethum graveolens* (Dill), *Acer saccharum* (Sugar maple)
Decommissioned Agriculture 2007	D07	1	*Rhus* sp. (Sumac), *Acer saccharum* (Sugar Maple), *Rubus subsp. Idaeobatus* (Wild raspberry), *Daucus carota* (Queen Anne's lace), *Dipsacus fullonum* (Wild teasel), *Asclepias* sp. (Milkweed), *Solidago* sp. (Goldenrod), grasses
		2	
		3	
Decommissioned Agriculture 2010	D10	1	*Daucus carota* (Queen Anne's lace), *Solidago* sp. (Goldenrod), *Dipsacus fullonum* (Wild teasel), *Heracleum maximum* (Cow parsnip), *Asclepias* sp. (Milkweed), *Taraxacum* (Dandelion), grasses
		2	
		3	
Hogsback	H	1	*Fagus grandifolia* (American Beech), *Quercus rubra* (Red Oak), mosses, grasses
		2	
		3	
Indian Woods	IW	1	*Fagus grandifolia* (American Beech), *Acer saccharum* (Sugar Maple), *Quercus rubra* (Red Oak), *Arisaema triphyllum* (Jack-in-the-pulpit), *Maianthemum racemosum* (False Solomon's seal)
		2	
		3	

2.2 Sample collection

Samples were collected from August to September, 2013 (Table 6). For each of the seven sites, three 5 m x 5 m sampling subplots were randomly placed (Figure 4A). Within each of the subplots, three randomly placed pits were excavated with a clean stainless steel shovel and spade (Figure 4A).

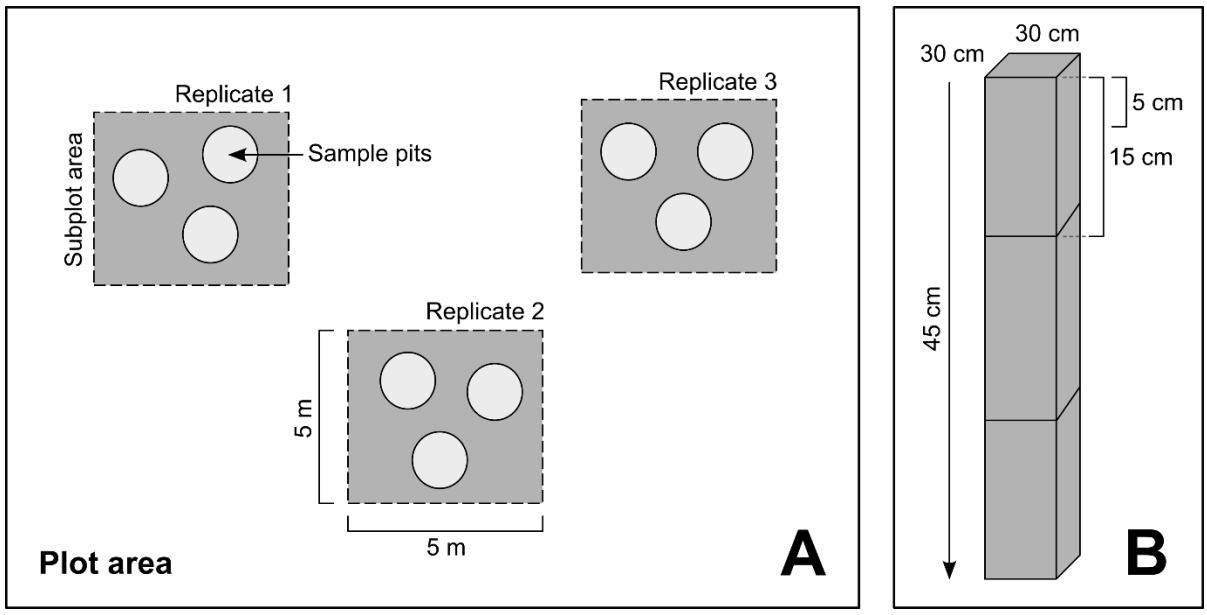

Figure 4*:* Overview of the sampling design. A) Overview of the sampling subplots and pits (not drawn to scale) and B) overview of the sampling scheme for each pit.

Pits were designated using a random number generator to assemble coordinates across each subplot. Each pit was sampled three times (i.e., every 15 cm) to a total depth of 45 cm (Figure 4B). In order to ensure that small-scale depth variations were not overlooked, samples were obtained at 5-cm intervals for one pit from each subplot (Figure 4B). Because soils at the Alvars site were shallow, pits were excavated to a total depth of 20 cm and sampled in 10 (coarse) and 5 (fine) cm increments. In addition, assessment of soil horizons for each depth increment was made using a combination of visual analysis of the soil profile, soil physicochemical results (Table 14, Appendix A), information from a previous survey of soils of ***rare*** Charitable Research Reserve (181), as well as guidelines from the Canadian system of soil classification (10) (Table 15, Appendix A).

Three sets of samples were taken for analyses. For soil physicochemistry, approximately 500 g of soil was removed from each 15 cm sampling increment using a clean stainless steel shovel, placed in a plastic bag on ice until transfer to a -80°C freezer until analysis. For bulk density analysis, a metal core ring (79.5 cm^3) was inserted into the side of each 15 cm depth increment profile (in

duplicate), carefully removed, placed into a plastic bag, and stored in the dark at room temperature. For molecular analyses, approximately 15 g of soil was removed from the appropriate depth interval using a clean stainless steel spatula and placed in a 15 mL falcon tube. Soils for molecular analyses were stored immediately on ice in the field and then at -80°C until analysis. For all soil collections, exposed soil surfaces were scraped and samples were collected as deep as possible within the soil profile to minimize cross-contamination between depth levels.

2.3 Soil physicochemical characterization

Soil physiochemical parameters measured in this study were chosen based on the recommended minimum sample information as determined by the MIMARKS guidelines (182) of the Genomic Standard Consortium (GSC). Approximately 100 g of soil from each 15 cm depth increment from each pit was pooled (3 soils per subplot, 63 samples total), thoroughly mixed, sealed in clean plastic bags, and sent on ice to the Agriculture & Food Laboratory (University of Guelph) for physicochemical analysis, with the exception of soils used for bulk density. For all samples sent for physicochemical characterization, prior to analysis, soils were first sieved to 2 mm. Because bulk density only requires dry soil weight expressed per unit volume, storage at room temperature was considered to have negligible effect on the final results. A summary of the methods used for physicochemical characterization is shown in Table 8.

Table 8: Summary of the methods used by the Agriculture & Food Laboratory (University of Guelph) for metadata characterization.

Analysis	Method	Reference(s)
pH	H_2O saturated paste method (1:2 ratio of soil to H_2O)	(183)
Ammonium	KCl-extractable NH_4^+-N; Spectrophotometric identification at 650-660 nm (USEPA 600/4-79-020: Method 350.1)	(184, 185)
Nitrate	KCl-extractable NO_3^--N; Spectrophotometric identification at 520 nm (USEPA 600/R93/100: Method 353.2)	(184, 186)
Nitrite	KCl-extractable NO_2^--N; Spectrophotometric identification at 520 nm (USEPA 600/4-79-020: Method 354.1)	(184, 187)
Organic Carbon	Combustion method using a high-temperature induction furnace	(188)
Inorganic Carbon		
Total Carbon		
Texture	Pipette method (sand fraction and textural classification).	(189)
Moisture (%)	Gravimetric moisture	(184, 190)

Briefly, soil pH was determined using the saturated paste method in a 1:2 ratio of soil to H_2O (i.e., 10 g of air dried soil to 20 mL of deionized/distilled H_2O) (Table 8) (183). To obtain organic C

measurements, subtraction of the total inorganic C portion of the soil (determined by ashing soils at 475°C for three hours) from the total organic C content was made (188). The total organic C data was collected using the LECO SC444 system which combusts and oxidizes soil C to CO_2 at 1350°C thereby making a calculation of the % C in the sample possible (188).

To obtain concentrations of inorganic N species, soils were first extracted with 2 M KCl and then analyzed and quantified using the Seal AQ2 discreet analyzer according to the manufacturer's standard operating procedures (Table 8) (184–187). The Seal AQ2 measures NH_4^+ by spectrophotometric analysis of indophenol blue (between 650–660 nm) created by reacting alkaline phenol and hypocholorite with NH_4^+ present in the sample (184, 185). The Seal AQ2 also measures NO_3^- by spectrophotometric analysis of a reddish-purple azo-dye (at 520 nm) derived from the reduction of NO_3^- by copperized cadmium to NO_2^-, which is subsequently reacted with sulphanilamide in dilute phosphoric acid (184, 186). This purple-reddish dye is also measured spectrophotometrically at 520 nm by the Seal AQ2 to quantify NO_2^- but uses an acetate buffer rather than the dilute phosphoric acid (184, 187).

Detailed soil textural analysis was conducted by passing soil samples through various sieve sizes as well as measuring the rates of particle settling in H_2O in relation to Stoke's Law (Table 8) (189). Bulk density calculations were made by weighing and oven drying soils at 105°C for 48 h (9). After drying, soils were weighed again and bulk density was calculated according to equation 3, below:

$$\rho_b = M_d \div V_R = M_d \div 79.5\ cm^3 \quad (3)$$

For equation 3, ρ_b refers to the bulk density, M_d refers to the dry weight of the soil, and V_R refers to the volume of the bulk density core rings, which is approximately 79.5 cm^3 (9). Gravimetric moisture content (mass of H_2O per mass of dry soil determined from drying soils at 105°C for 24 h) was calculated during preparation for NH_4^+ and NO_3^- determination (9, 184, 190).

2.4 Microbial community analysis

2.4.1 DNA extraction & quality assessment

Total DNA was extracted from approximately 0.5 g of thoroughly homogenized soil using the PowerSoil-htp 96 Well Soil DNA Isolation Kit (MO BIO) according to a modified version of the manufacturer's instructions. Soil extraction mass was adjusted as described in the manufacturer's guidelines according to the approximate soil type to ensure optimized DNA yield. The PowerSoil-htp 96 Well Soil DNA Isolation Kit uses a combination of both mechanical and chemical techniques for

cell lysis. In addition, the kit removes common soil contaminants such as humic acids, proteins, and other organic and inorganic materials through successive precipitation and washing steps. A modification of the protocol was required to circumvent the requirement of a mechanical plate shaker (for cell lysis). Instead, soils were extracted using individual garnet bead tubes, following a similar protocol as described in the manufacturer's instruction for the PowerSoil DNA Isolation Kit (MO BIO, steps 1–6). Mechanical and chemical (addition of SDS, kit supplied) lysis for individual tubes was carried out for 45 s at 5 m s^{-1} in a bead mill (FastPrep 24, MP Biomedicals). After mechanical and chemical lysis, samples were centrifuged at 10,000 x *g* for 30 s. The supernatant was transferred to a 96-well extraction plate and stored at 4°C overnight (<18 h) before completion of the extraction protocol the following day as described in the PowerSoil-htp protocol (steps 10–34). Extracted DNA was stored at -20°C in an EDTA-free 10 mM Tris solution (100 µL).

The quality of the extracted nucleic acids were assessed on a 1% (w/v) agarose gel (BioBasic; prepared with 1X TAE buffer and stained with GelRed; diluted 1:10,000; Biotium) and visualized using the AlphaImager HP (Alpha Innotech). A reference ladder (1 kb Plus ladder; Invitrogen) was used as a size marker. In addition, sample concentration and quality was assessed using the NanoDrop 2000 spectrophotometer (Thermo Scientific).

2.4.2 HTP sequencing

2.4.2.1 PCR amplification of the16S rRNA gene

Bacterial communities were characterized by amplification of a segment of the 16S rRNA gene spanning the V3-V4 hypervariable region (~ 465 bp) using a modified protocol as described elsewhere (191). Modifications of the protocol include addition of the V4 region (creating longer fragments with higher taxonomic resolution), increased PCR cycle number (necessary to ensure most samples generate PCR product), and the use of *Taq* over Phusion DNA polymerase to generate more product. Modified Illumina-specific primers, 341–F (5'–CCTACGGGAGGCAGCAG–3') and 806–R (5'–GGACTACHVGGGTATCTAAT–3'), were used to amplify the bacterial 16S rRNA fragments, adapter sequences (for flow-cell binding), complimentary forward and reverse priming regions required for Illumina-specific primers, and a 6-base barcode for sample multiplexing. Each PCR was carried out in 25 µL reactions containing 2.5 µL of ThermoPol *Taq* buffer (10X; New England BioLabs), 0.05 µL of the forward primer (100 µM; Integrated DNA Technologies), 0.5 µL of the reverse-indexed primer (10 µM; Integrated DNA Technologies), 0.05 µL of dNTPs (100 mM; New England BioLabs), 0.125 µL of *Taq* DNA polymerase (5000 U mL^{-1}; New England BioLabs), 1.5 µL of bovine serum albumin (10 mg mL^{-1}), 1 µL of normalized template DNA (1-20 ng μL^{-1}), and 19.3

μL of nuclease-free PCR-grade H_2O (Thermo Scientific). All PCRs were run using the T100 Thermal Cycler (Bio-Rad) and the following reaction conditions: initial denaturation step at 95°C for 30 s, 30 cycles of denaturation (95°C for 15 s), annealing (50°C for 30 s), and extension (68°C for 30 s) followed by a final extension step at 68°C for 5 min. Electrophoresis of PCR products on a 1% (w/v) agarose gel (BioBasic) prepared with 1X TAE buffer and stained with GelRed (diluted 1:10,000; Biotium) or ethidium bromide (1 μg mL^{-1}; Calbiochem) was used to confirm amplification as well as check for specificity and size of each PCR replicate. Samples that showed little to no amplification were noted and attempts were made to optimize PCR conditions (increasing template amount) to obtain PCR products. These PCR products were pooled and sequenced in a separate sequencing run.

2.4.2.2 Illumina library preparation

After the quality of the PCR products were assessed, triplicate PCR products were pooled and gel quantified with pre-prepared V3-V4 standards using the band analysis tool from the AlphaView Software (Alpha Innotech). Standards for gel quantification were prepared by pooling previously prepared Illumina-based PCR products (spanning the V3-V4 region), gel purifying the band corresponding to the 16S rRNA gene fragment using the Wizard SV Gel and PCR Clean-Up System (Promega), and quantifying the products using the NanoDrop 2000 spectrophotometer (Thermo Scientific). In some cases, quantification of pooled PCR products was carried out using relative concentrations so that an appropriate volume of each PCR product could be added to make a sufficient library volume. Pooling was carried out such that an equal amount of each PCR product was added to the final library. After pooling all PCR products into a single sample, approximately half of the library was electrophoresed on a 1% (w/v) agarose gel (BioBasic) prepared with 1X TAE buffer and stained with ethidium bromide (1 μg mL^{-1}; Calbiochem). The band corresponding to the 16S rRNA gene was excised and gel purified using the Wizard SV Gel and PCR Clean-Up System (Promega) and stored in EDTA and nuclease-free H_2O at -20°C until library quantification.

Because cluster generation is highly dependent on the concentration of the library loaded onto the flow cell, quantification was carried out using three different methods to ensure precision. Prior to quantification, libraries were diluted to 6 and 9 nM using the initial library concentration determined via spectrophotometry. Validation of these concentrations were carried out using the following methods: the Qubit 2.0 Fluorometer and the Qubit dsDNA HS Assay Kit (Invitrogen), gel quantification using previously prepared Illumina-based standards (see above), and qPCR. Library quantification via qPCR was done using the PerfeCta NGS Library Quantification Kit (Quanta Biosciences) designed specifically for the quantification of Illumina libraries. Briefly, the qPCRs were carried out in 20 μL reactions using the Illumina forward (P5) and reverse (P7) primers (5'–

AATGATACGGCGACCACCGA–3' and 5'–CAAGCAGAAGACGGCATACGA–3', respectively) as well as a SYBR Green SuperMix (Quanta Biosciences) on the C1000 Thermal Cycler coupled with a CFX96 optical module (Bio-Rad). The program used had the following conditions*:* initial denaturation step at 95°C for 3 min, 35 cycles of denaturation (95°C for 10 s), annealing (60°C for 20 s), and extension (72°C for 45 s) followed by a melt curve analysis from 65–95°C in 0.5°C increments held for 2 s. Comparison of the concentrations of the diluted libraries was made to ensure high degree of confidence in the final concentration of the library. Quantified libraries were stored at -20°C until sequencing. Libraries prepared greater than approximately one week before sequencing were also quantified again using fluorometry to ensure no changes in concentration during storage.

2.4.2.3 Illumina sequencing

Libraries were prepared for sequencing using the MiSeq Reagent v2 kit (500 cycles, Illumina) according to the manufacturer's instructions. Briefly, sample and PhiX control (for increasing sample diversity) libraries were denatured using a 0.2 N NaOH solution and diluted (8-10 pM for sample libraries and 10 pM for the PhiX control library) before being pooled in a 19:1 ratio and loaded onto the MiSeq Reagent v2 cartridge (Illumina) for sequencing. Paired-end sequencing of monoclonal clusters of V3-V4 sequences was done on the MiSeq platform (Illumina) using 250-base reads. In addition, an extra index read was included to allow sample mapping and sorting based on the 6-base sample barcode. Post-sequencing analysis, including image analysis, base calling, Phred quality score calculations, and demultiplexing (barcode sorting) were performed using the MiSeq Control Software (version 2.3.0.3).

2.4.3 Bioinformatic and statistical analyses

2.4.3.1 Data subsets

Because soils were sampled in two different depth resolutions (i.e., in 5 and 15 cm increments, see section 2.2) across various land-use types, the entire dataset was split into four main data subsets in addition to the master dataset (e.g., all samples) to increase resolution in trends. Datasets were divided into two depth resolution subsets, 15 and 5 cm, as well as two land-use type based datasets which discriminated forested (IW, H, CA) and field (AA, D03, D07, D10, Al) sites. The latter two datasets included soil samples taken exclusively from the 15 cm increments. This was done to ensure a high degree of sample replication considering that soils were taken in 15 cm increments from all pits whereas soils sampled in 5 cm increments were taken from only one pit in each subplot.

2.4.3.2 Sequence clustering, OTU table generation, and downstream analyses

After initial post-sequencing analysis (see section 2.4.2.3), resulting paired-end reads were assembled using the PAired-eND Assembler for DNA sequences (PANDAseq) algorithm using default parameters including a quality threshold (PQT) of 0.9 and a minimum sequence overlap threshold of 1 base (172). Briefly, PANDAseq assembles paired-end sequences by first sorting reads by barcode and then looking for overlapping regions in the sequence sets (191). Sequences with both mismatches and ambiguous base calls, based off Phred scores, were discarded allowing for high quality datasets to be acquired (191). Because greater than 98% of the samples had more than 10,000 sequences per sample, any sample below this threshold was considered an outlier and removed. As a result, three samples from the 15 cm increment dataset (a sample from the 30–45 cm increment in the active agricultural field, 736 sequences per sample; 2 samples from the 30–45 cm increment in the decommissioned 2010 field, 1016 and 2136 sequences per sample) and one sample from the 5 cm increment dataset (a sample from the 40–45 cm active agricultural field; 954 sequences per sample) were removed. After removing these samples, each depth increment was still represented by >60 samples. For downstream analyses requiring equal sequence contribution per sample, rarefaction (e.g., random subsampling of sequences from each sample) was carried out at a specified depth using the number of sequences per sample dictated by the sample with lowest sequence count.

Sequence processing and dataset exploration was largely managed using the AXIOME (Automation, eXtension, and Integration Of Microbial Ecology) pipeline, a tool allowing streamlined data analysis and visualization of small subunit rRNA datasets through implementation of a variety of standard computational tools used in microbial ecological studies (191, 192). In particular, *de novo* clustering of assembled reads into OTUs and removal of chimeric sequences and singletons (i.e., single sequences) were done using the UPARSE (179) pipeline at a 97% sequence identity threshold. Classification of OTUs was carried out using the RDP classifier (Ribosomal Database Project, version 2.2) (193) at a bootstrap confidence threshold of 0.8 against the Greengenes database (version 13.8) (194). A summary of the overall bioinformatic and sequence processing steps is shown in Figure 5.

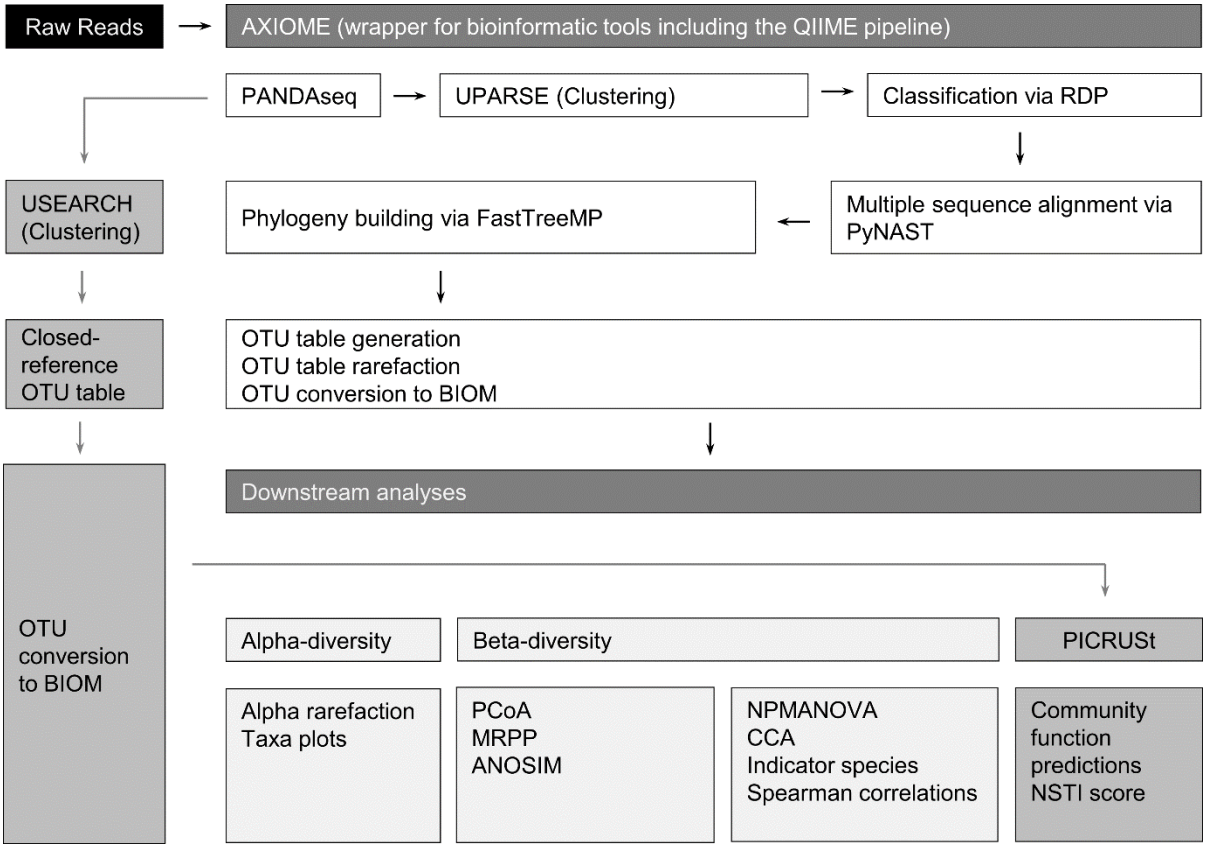

Figure 5: Flowchart showing the overall bioinformatic and sequence processing steps. Raw reads were assembled using PANDAseq and then clustered into OTUs using the UPARSE pipeline. Representative sequences were then classified using the RDP classifier against the Greengenes database. Phylogeny building was carried out using FastTreeMP from a multiple sequence alignment computed using the PyNAST algorithm (data not shown). OTU table generation, rarefaction, and conversion to BIOM format were carried in QIIME prior to further downstream analyses including alpha- and beta-diversity analysis (Principal Coordinates Analysis, PCoA, Multiple-Response Permutation Procedure, MRPP; ANalysis Of SIMilarity, ANOSIM, Non-Parametric Multivariate Analysis of Variance using distance matrices, NPMANOVA; Constrained Correspondence Analysis, CCA). For community functional predictions, closed-reference OTU picking was used to generate a dataset for downstream analysis using the PICRUSt algorithm.

Multivariate visualizations using Principal Coordinate Analysis (PCoA) (195) were generated in AXIOME (192). Principal coordinate analysis attempts to maximize the variance (often expressed as a percentage) in the dataset through synthetic variables expressed as principal axes (195). Bray-Curtis dissimilarities (196) were used to assess the compositional dissimilarity between sites. The Bray-Curtis method calculates the pairwise sample dissimilarities by first summing the number of common taxa present between two samples and comparing this to the sum of the total number of taxa in both samples (196). Non-parametric methods, Multiple-Response Permutation Procedure (MRPP; uses pairwise distances to calculate and compare expected and observed deltas) (197) and ANalysis

Of SIMilarity (ANOSIM; tests the average mean of the ranked dissimilarities between-groups and within-groups) (198), were carried out in AXIOME and used to test whether *a priori* groupings of samples were significantly different. For consistency in MRPP interpretations, thresholds for the chance corrected within-group statistic, *A*, were arbitrarily defined as weak, <0.1, moderate, 0.1–0.3, and strong, >0.3, based on the notion that values >0.3 are considered high in community ecology (199). The MRPP test-statistic, *T*, which represents between-group separation, was also given arbitrary threshold values where values >-10, represented weak, -10– -25, moderate, and <-25, strong.

To assess the influence of explanatory variables in affecting the overall 16S rRNA gene datasets, NPMANOVA (non-parametric multivariate analysis of variance using distance matrices) (200) was carried out using the adonis function in the vegan package (version 2.2-1) (201) in R (202) (managed in QIIME). Adonis provides an evaluation of the strength and significance of explanatory variables in explaining the differences in a dissimilarity matrix by partitioning the sums of squares in a multivariate dataset (200). NPMANOVA was carried out on each environmental variable individually to wholly assess the overall influence. Constrained (Canonical) Correspondence Analysis (CCA) (203) based on NPMANOVA results (i.e., R^2) was used to further evaluate the relationship between edaphic factors and soil bacterial community structure. The CCA analysis, a version of correspondence analysis which attempts to preserve χ^2 distances between samples, constrains ordinations based on linear combinations of the prescribed explanatory variables (203, 204). It is considered a common and robust approach for microbial ecological exploratory analyses even considering its ideal use for unimodal models of species response to the environmental variation (203, 204). Based on the explained variations and their interpreted ecological significance from NPMANOVA, the following parameters were used for CCA: depth, pH, NO_2^-, NO_3^-, NH_4^+, SOC (i.e., $C_{(o)}$), SIC (i.e., $C_{(i)}$), $C_{(t)}$, % H_2O, soil texture, site ID, bulk density and land-use type. To test the significance of the constrained ordination, the anova.cca function in the vegan package (version 2.2-1) in R was used (201, 202). All CCA models reported in this study were significant ($p < 0.01$). For CCA analyses, depth was considered as a continuous variable considering that it represents an environmental gradient. In all analyses, physiochemical results that were below the analytical detection limit were adjusted to half of the detection limit value (205). Detrended Correspondence Analysis (DCA) was carried out to help reduce the strong "arch effect" observed in the forest data subset using the phyloseq (version 1.10.0) package (206) in R (202).

Species richness shifts with depth and across land-use types were analyzed using observed OTU counts. Non-parametric two sample t-tests were used to test for significant differences in richness estimates among depth and land-use type categories using the "compare_alpha_diversity.py" script in QIIME (25, 207). In addition, indicator species analysis was used which looks for species

closely associated with a particular category based on fidelity and specificity (208). This analysis, managed by AXIOME, was conducted for the categories of depth, land-use type, and pH using a 0.7 indicator value threshold, sequence abundance threshold of 100, and p value cut-off of < 0.05.

To assess overall shifts in taxa with depth, bar plots based off relative abundances of the top ten most abundant phyla at both coarse- and fine-scale depth intervals were created in the phyloseq package (version 1.10.0) (206) in R. To test the significance of shifts of each phylum between depth increments, non-parametric pairwise Wilcoxon rank-sum tests were conducted for the top ten most abundant phyla using the pairwise.wilcox.test function in the "stats" package (version 3.1.2) in R (202). The p-values generated from the Wilcoxon rank-sum tests were Benjamini-Hochberg corrected to reduce false discovery rate (209).

For partial assessments of taxa distributions and dynamics, CCA biplots showing the top 30 OTUs classified to the class-level were generated using the phyloseq (version 1.10.0) and vegan packages (version 2.3-0) (206) in R. Depth, pH, and site ID were constrained in the CCA plots considering their importance in this study. In addition, vectors highlighting depth, pH, and land-use type were fitted onto ordination space using the "envit" function in the vegan package in R and significance of each correlation was assessed based on 999 permutations.

Spearman's rank correlation analysis was used to test for significant correlations ($p < 0.001$) between taxa and the metadata categories depth and pH using a correlation threshold ≥ 0.5. ("cor.test" function in the "stats" package in R). Correlations were calculated for taxa from both the 15 and 5 cm increment datasets to compare both large-scale (e.g., 15 cm increment) and fine-scale (e.g., 5 cm increment) dynamics.

2.4.3.3 Comparison of sequence clustering strategies

To assess the effects of different clustering and assembly algorithms on OTU table generation, a data subset of 21 soil samples taken from one arbitrarily chosen subplot from each of the seven sites was generated. Three common *de novo* heuristic OTU clustering schemes, including CD-HIT, UCLUST, and UPARSE were applied to the data subset using a 97% sequence identity threshold and PANDAseq quality thresholds of 0.6 and 0.9, respectively. The UPARSE clustering algorithm was tested using default parameters (e.g., removal of singletons) as well as with modified parameters (e.g., including singletons; UPARSE+1). For CD-HIT and UCLUST, chimeric sequences were removed using UCHIME (210). Similar to the full dataset analysis, classification of OTUs was carried out using the RDP classifier (version 2.2) (193) against the GreenGenes database. To visualize and assess changes in OTU counts from different clustering and read-assembly strategies

and thresholds, rarefaction curves were used and adjusted to an even sequencing depth of 17,550 sequences per sample.

2.4.3.4 Community function predictions

Community functional predictions was carried using the PICRUSt algorithm (161). Briefly, PICRUSt works by first precomputing the gene content in a reference 16S rRNA phylogenetic tree before predicting genomes of terminal nodes (unknown genome content) via ancestral state reconstruction and the assumption that gene content can be weighted by the reciprocal of the phylogenetic distance to the closest full genomes (211). Once computed and normalized by 16S rRNA gene copy number, gene predictions can be multiplied by the relative abundances of 16S rRNA genes in the associated samples, thereby providing a prediction of the entire community metagenome (161). The PICRUSt approach uses two sets of workflows: 1) gene content inference and 2) metagenome prediction inference (161). These workflows were carried out on the 15 cm depth increment subset. The 15 cm sample dataset was comprised of 187 samples (12,515,168 total input sequences).

PICRUSt is dependent on a reference database of known 16S rRNA genes therefore sequences were clustered into OTUs using the USEARCH (version 6.1.544) (178, 179) algorithm managed through QIIME (207) using a 97% sequence identity threshold against the GreenGenes database (version 5) (162). This allowed the generation of a closed-reference PICRUSt-compatible OTU table. Metagenome predictions were carried out as described previously (161) by using the predicted reference metagenomes from the output of the gene content inference step and the query OTU table generated from closed-reference OTU picking step.

Because no unique changes to the reference dataset were considered necessary, precalculated files obtained from the PICRUSt installation (version 1.0.0) were used. These files were prepared as described previously (161). Specifically, annotated 16S rRNA genes were obtained from the GreenGenes database (version 13.5, 408,135 annotated 16S rRNA gene sequences; matching the same version used in OTU table generation step) and the 16S rRNA gene copy number and reference genomes using KEGG (Kyoto Encyclopedia of Genes and Genomes) Ortholog annotations (KOs; 6,909 annotations) were derived from the IMG database (version 4.0) (162, 212, 213). KOs were chosen specifically for this analysis because this offered more annotations than other databases, such the Clusters of Orthologous Groups of proteins database (COG), thereby potentially further increasing the resolution in the predictions (161). KOs are categorized into KEGG pathways which summarize experimental knowledge on the molecular interaction of genes (213). The KEGG pathways are grouped in a hierarchy with the main parent levels (i.e., general pathway groups) as

follows: “metabolism”, “genetic information processing”, “environmental information processing”, “cellular processes and signalling”, “organismal systems”, and “human diseases”.

The reference phylogenetic tree for the 16S rRNA gene dataset was prepared using the “tax2tree” version from the GreenGenes database (194). Ancestral state reconstruction was carried out using the default “ace_pic” (Phylogenetically Independent Contrasts, PIC method) function (a least squares method) in the APE package (version 2.8) (214) in R.

Quality control on the resulting dataset was conducted using NSTI scores and confidence intervals for each predicted trait in the metagenome prediction step. NSTI scores assess the extent to which microorganisms in a sample are related to a reference genome (161). They are the average branch length that an OTU is to a reference genome weighted by the OTU’s abundance; the higher the NSTI value the more phylogenetically diverse the sample (i.e., not many OTUs are closely represented by a reference genome) (161). Confidence intervals are based off the variance of the predicted trait resulting from ancestral state reconstruction as well as uncertainty resulting from the evolutionary distance separating the query taxa (i.e., taxa for which traits are being predicted from the reference genome) and/or the reconstructed ancestors that are enabling the prediction. The confidence intervals were reported only when specific KEGG annotations were looked at specifically (e.g., for Post-hoc analyses; see below).

Downstream analysis of the predicted genomes, specifically ordination analysis via Principal Components Analysis (PCA) and heat maps, were generated by STAMP (version 2.0.0) (215). A heat map was specifically created using only the “Cellular processes and signalling”, “Environmental information processing”, “Genetic information processing”, and “Metabolism” KEGG pathways to help reduce the amount of non-applicable genes based on the presumption that these categories are most relevant to environmental systems. The dendrogram associated with the heat map was generated using the UPGMA (average neighbour) clustering method at a clustering threshold of 0.85. To further reduce the amount of seemingly uninformative gene predictions in the heat map, a *q*-value (Benjamini-Hochberg, false discovery rate assessment) threshold was set to 0.05 (i.e., those predictions with a value greater than 0.05 were removed). Furthermore, an effect size threshold was also set to 0.35 (i.e., those predictions that have an effect size less than 0.35 were removed). Effect size expresses the magnitude of the difference in gene counts between two groups by the calculating the standardized mean difference. In this analysis, the effect size threshold was set arbitrarily to primarily filter the number of uninformative predictions (i.e., those gene predictions that were proportionally low across all sites) (215). Post-hoc analyses were conducted using the Tukey-Kramer test (216) (pairwise-comparison; at 95% confidence intervals) based off Benjamini-Hochberg corrected *p*-values generated via non-parametric Kruskal-Wallis H-tests.

Chapter 3
Results

3.1 Soil characteristics

Soil physicochemistry varied with soil depth, individual plots, and across site locations (Figure 6–8, Table 14, Appendix A). Soil pH was consistently higher in deeper soils than in surface soils from all sites (Figure 6, Table 14, Appendix A). The forest sites (IW, H, CA) were more variable in pH than the field sites (D03, D07, D10, AA, Al), which were all consistently near–neutral to slightly alkaline (Figure 3, 6, Table 14, Appendix A). Some forest soils were acidic (CA, plot 3) whereas others were near–neutral to alkaline (IW, plot 1; Figure 6). The CA site had the widest range of pHs than any other site, with the pH of plots ranging from 4.5 to >7.5. Overall, SOC decreased with soil depth (Figure 7). In surface soils (i.e., 0–15 cm), SOC content ranged from <1% (D07, plot 1) to >8% dry weight (CA, plot 1). In contrast, SIC was generally higher in deeper soils (Figure 7). Soil NH_4^+ and NO_3^- was generally lower in deeper soils, except for site AA, plot 2, which appeared to have higher NH_4^+ in deeper samples than in surface samples (Figure 8). Surface NH_4^+ was generally higher in the forest than field sites whereas surface NO_3^- was generally higher in field than in forest sites. Soil moisture was lower in deeper soil, with the exception of samples from site D03, which were consistent throughout the soil profile (Table 14, Appendix A). In addition, soil samples were classified as loamy, or some variation thereof, with the exception of soils from site D07, plot 1, which were sandier (i.e., texturally classified as “fine sands”). A summary of the soil physicochemical results with the exception of bulk density (data not shown) are shown in Table 14, Appendix A.

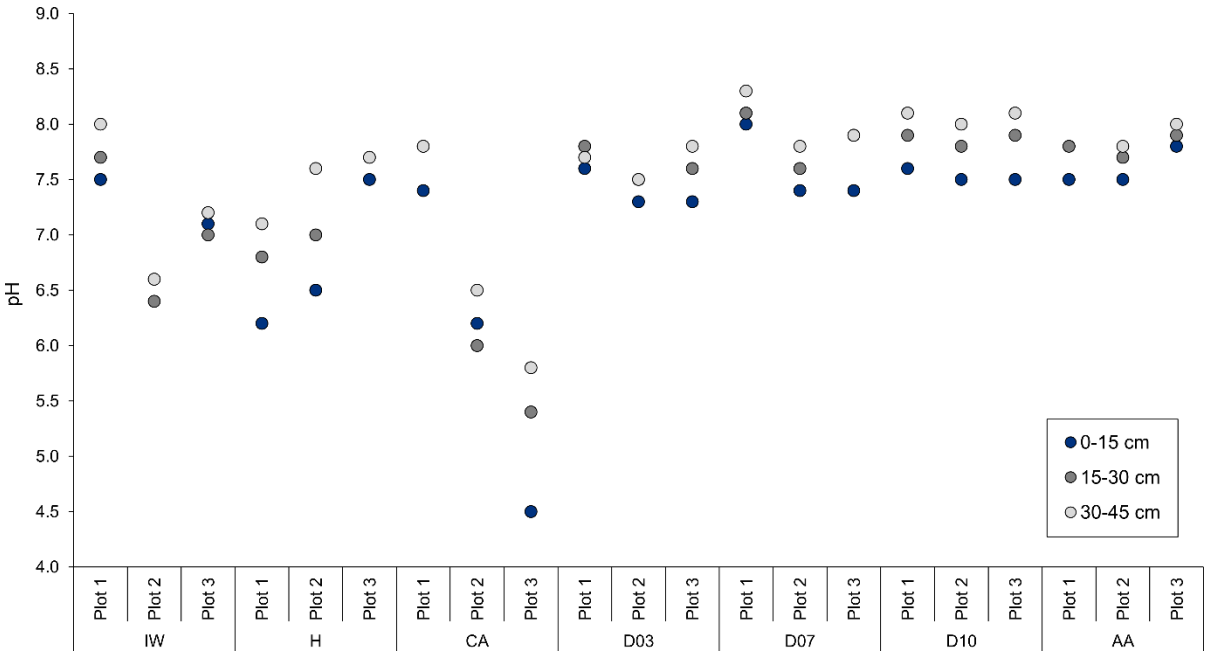

Figure 6: pH with depth from all sites across soils from the ***rare*** Charitable Research Reserve. Points represent pH for composite soils obtained by pooling soils from triplicate pits at each plot for each specific depth increment.

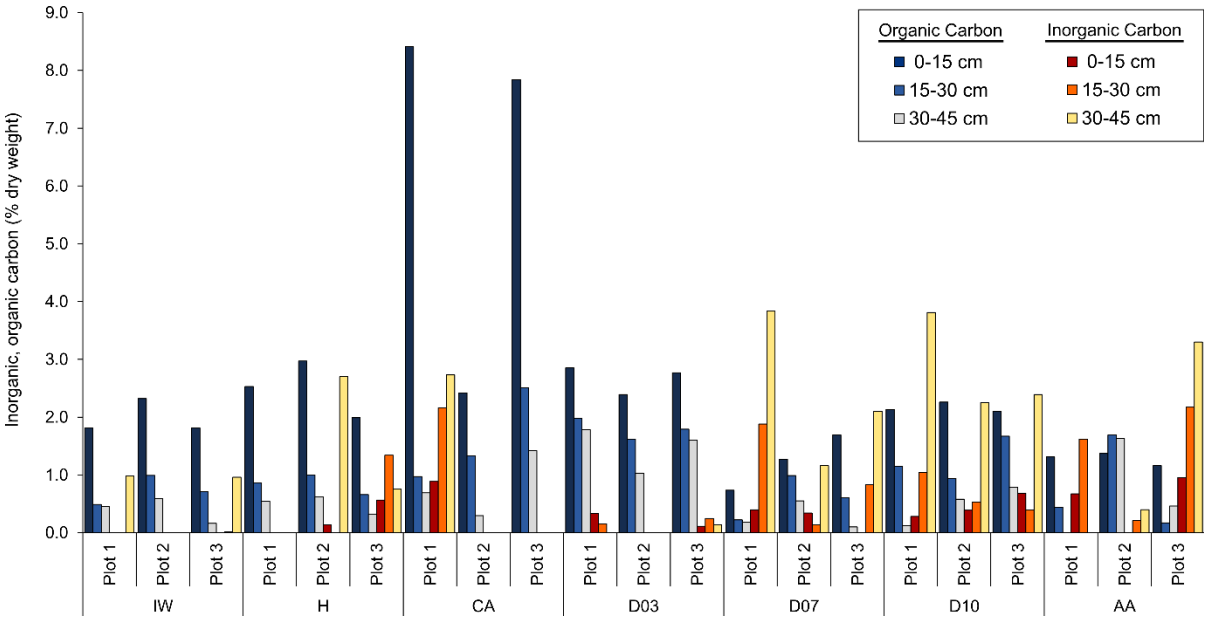

Figure 7: Inorganic and organic C content with depth from all sites across soils from the ***rare*** Charitable Research Reserve. Each bar represents values for composite soils obtained by pooling soils from triplicate pits at each plot for each specific depth increment.

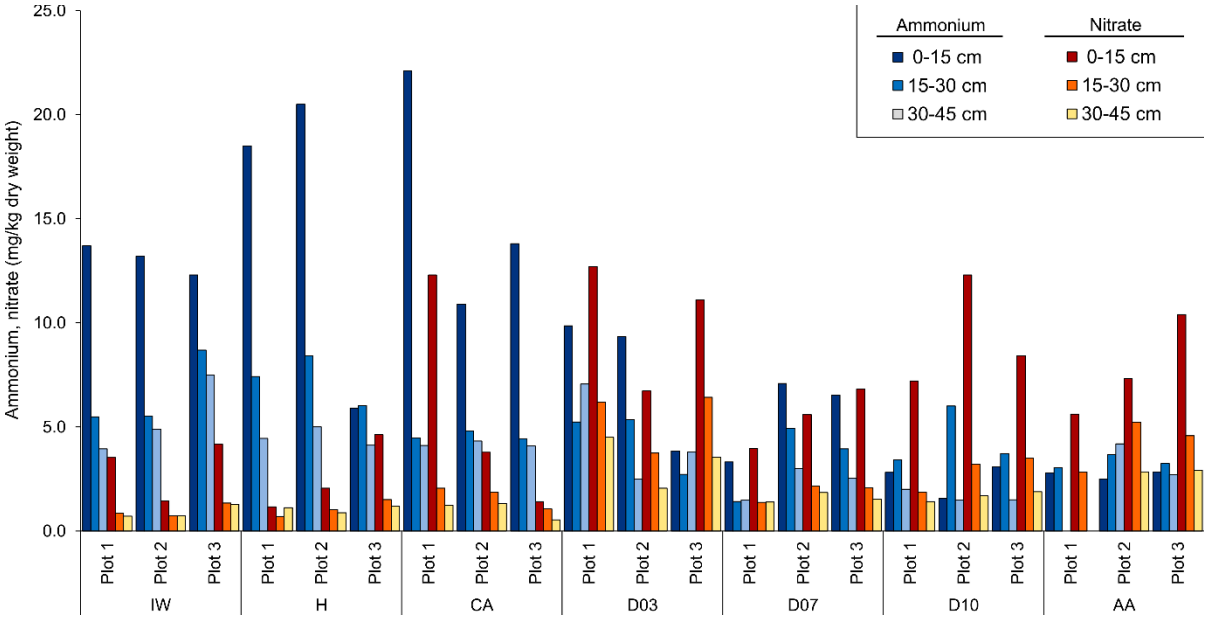

Figure 8: NH_4^+ and NO_3^- content with depth from all sites across soils from the ***rare*** Charitable Research Reserve. Each bar represents values for composite soils obtained by pooling soils from triplicate pits at each plot for each specific depth increment.

3.2 HTP sequencing of soil samples from the *rare* Charitable Research Reserve

Sequencing of 16S rRNA gene amplicons from 376 soil samples from 7 selected sites across the ***rare*** Charitable Research Reserve generated an overall dataset of 34,232,172 assembled paired-end reads. After sequence clustering and chimera checking via the UPARSE pipeline, 28,307,344 sequences remained in the overall dataset. Depending on the particular data subset used in downstream analyses, sequenced samples maintained a minimum, maximum, and mean sequence count of greater than 10,000, 100,000, and 60,000, respectively (Table 9). As a result, rarefaction depth for all data subsets was >10,000 sequences. Data subsets were clustered into OTUs with full-site subsets (i.e., full, 15 cm, and 5 cm increment subsets; see section 2.4.3.1) and land usage subsets maintaining greater than 17,000 OTUs (Table 9). Land usage subsets (e.g., forests and fields only) had OTU counts of greater than 11,000 (Table 9).

Table 9: Summary of 16S rRNA gene data subsets. The overall "Full" dataset was divided into smaller data subsets to enable data exploration and interpretation.

Dataset	Number of samples	Assembled sequences[a]	Unique sequences	Counts per sample			OTUs	Total sequences
				Min	Max	Mean		
Full	376	34,232,172	1,433,931	10,348	359,138	75,285	24,809	28,307,344
15 cm	187	14,912,183	658,720	16,680	358,286	66,926	17,549	12,515,168
5 cm	189	19,319,989	804,863	10,413	330,840	83,118	19,640	15,709,212
Forests[b]	81	6,985,667	313,394	16,646	108,560	75,552	12,577	5,919,450
Fields[c]	106	7,926,516	349,770	18,402	358,926	61,917	11,382	6,563,180

[a]via PANDAseq; [b]sites IW (Indian Woods), CA (Cliffs & Alvars), and H (Hogsback); [c]sites AA (Active agricultural), D03, D07, D10 (Decommissioned 2003, 2007, and 2010), and Al (alvars)

3.2.1 Sequence clustering analysis

To reinforce the validity of the sequence clustering strategy used in this study, a small subset of samples (n = 21) were clustered into OTUs under four clustering treatments (Table 10). The four clustering strategies were CD-HIT, UCLUST, UPARSE+2 (UPARSE algorithm with singletons removed), and UPARSE+1 (UPARSE algorithm with singletons included). With respect to the paired-end assembly thresholds, using a more stringent cut-off (i.e., 0.9) reduced assembled sequence counts by ~9% compared to a threshold of 0.6 (Table 10). Relaxing the assembly quality threshold increased OTU counts regardless of the clustering scheme (Table 10). Clustering algorithms that included singletons (e.g., CD-HIT, UCLUST, UPARSE+1) showed marked increases in OTUs (~30%) when the PANDAseq quality threshold was relaxed to 0.6 (Table 10). Although the UPARSE+2 clustering scheme also showed an increase in OTUs when the threshold was relaxed, the increase was much less (~3%; Table 10). As shown in Figure 9 and Table 10, the OTU counts generated from CD-HIT, UCLUST, and UPARSE+1 were orders of magnitude greater than counts generated by UPARSE+2. The rarefaction curves (Figure 9) showing the observed species richness generated from the UPARSE+2 clustering strategy showed evidence of an asymptote (global maximum), which contrasted with clustering subsets that included singletons (CD-HIT, UCLUST, and UPARSE+1) where a lack of an asymptotic trend was observed at a comparable sequencing depth (Figure 9).

Regardless of the number of OTUs generated, there were similar numbers of sequences retained in the final datasets after other processing steps (e.g., chimera checking; Table 10). However, in comparison to CD-HIT and UCLUST, the UPARSE subsets both showed a similar magnitude of total sequences lost after clustering (Table 10). Chimeric sequence counts in the UPARSE+2 subset represented ~8% of the total non-singleton input sequences and ~9–10% of the total input sequences in the CD-HIT and UCLUST datasets (Table 10). Chimeric sequences were highest in the UPARSE+1 dataset (~25–28% of the total input sequences; Table 10).

Furthermore, an important consideration when assessing the different clustering approaches is time and memory requirements, which govern the overall practicality of each approach (217). Although this analysis was conducted on a small dataset, there were time and memory characteristics associated with each method which can be extrapolated and interpreted for larger datasets. In particular, the UPARSE+2 method was the quickest and least memory intensive clustering method (Table 10). In contrast, CD-HIT required the most time and had comparable memory requirements to the UPARSE+1 clustering strategy (Table 10). UCLUST, which generated the most OTUs, was the second fastest clustering algorithm but had similar memory requirements to CD-HIT and UPARSE+1 (Table 10).

Table 10: Comparison of clustering algorithms and PANDAseq quality thresholds (PQT) on sequence and OTU count dynamics. Sequence clustering conducted on 21 samples from the 7 sites (1 subplot from each site).

Clustering method	CD-HIT		UCLUST		UPARSE+2[a]		UPARSE+1[b]	
	PQT 0.6	*PQT 0.9*	*PQT 0.6*	*PQT 0.9*	*PQT 0.6*	*PQT 0.9*	*PQT 0.6*	*PQT 0.9*
Assembled sequences	1,759,736	1,605,840	1,759,736	1,605,840	1,759,736	1,605,840	1,759,736	1,605,840
Unique sequences	N/A[c]				83,126	77,859	946,659	855,303
OTUs before chimera checking	181,939	149,892	217,006	181,258	N/A[d]			
Chimeric OTUs	87,606	86,344	98,027	95,096				
Chimeric sequences / % abundance	170,302 / 9.6	157,671 / 9.8	159,365 / 9.1	151,196 / 9.4	6,955[e] / 8.3	6,345[e] / 8.1	267,344 / 28.2	219,611 / 25.7
OTUs after chimera checking	91,831	63,548	118,979	86,162	6,326	6,131	59,796	39,692
Total sequences included in analyses	1,589,434	1,448,169	1,600,371	1,454,644	1,410,047	1,307,105	1,438,903	1,340,089
Total sequences lost (%)	9.6	9.8	9.1	9.4	19.8	18.6	18.2	16.5
Approximate clustering time / max memory usage[f]	232 min / 1303 Mb	189 min / 1232 Mb	27 min / 1027 Mb	22 min / 923 Mb	2 min / 70 Mb	1 min / 67 Mb	92 min / 1461 Mb	59 min / 1036 Mb

[a]Singletons removed from dataset; [b]Singletons included in dataset; [c]no dereplication step for CD-HIT and UCLUST; [d]UPARSE picks OTUs and checks chimeras at the same time; [e]refers to "chimeric non-singletons" because the singletons were discarded before the chimera check step; [f]Memory (max memory stored in RAM during clustering) and time approximations are based exclusively on the "usearch cluster_otus" command for UPARSE analyses, "pick_otus.py –m uclust" command for UCLUST, and "cdhit-est" command for CD-HIT

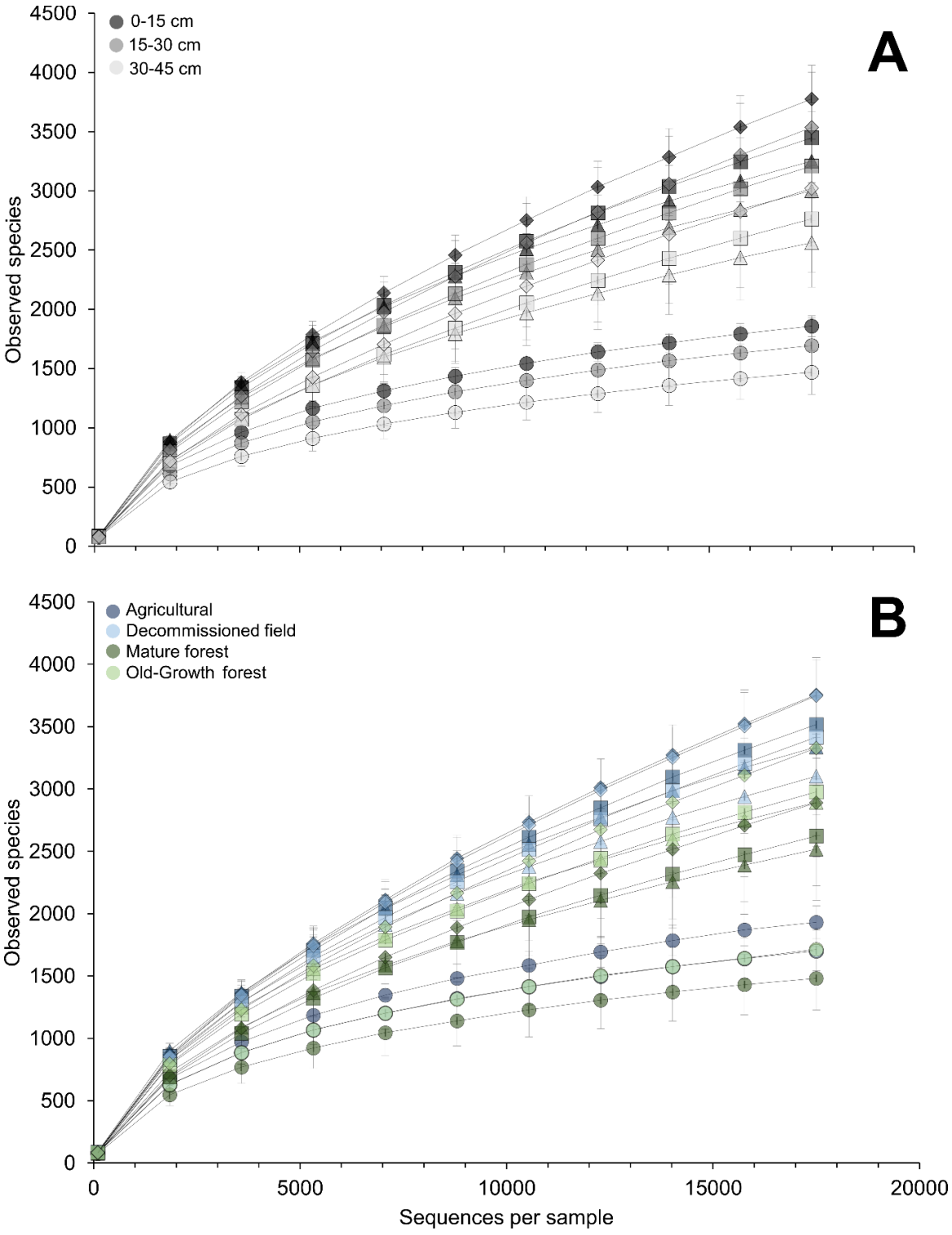

Figure 9: Rarefaction curves showing OTU counts from clustering schemes. Symbols represent different clustering schemes: □, CD-HIT; ◇, UCLUST; ○; UPARSE(+2); △ UPARSE (+1). Rarefaction depth was set to 17,550 sequences per sample. Curves are shown for each depth increment (A) and land-usage type (B).

3.3 Identifying patterns in soil bacterial communities

3.3.1 OTU richness

In general, richness estimates showed a consistent trend with depth across all sites, with surface soil samples associated with higher OTU richness than deeper samples (t-test, $p < 0.001$, between 0–15 and 0–30, 0–15 and 15–45 cm) (Figure 10). Although this trend was less apparent when assessed by overall depth increment across all sites (Figure 11), richness results from the clustering trial subset (section 3.2; Figure 9) showed a consistent pattern of richness shifts with depth, as shown in Figure 10. In relation to pH, greater richness was generally observed for samples maintaining near–neutral to slightly alkaline pHs (Figure 11). Field sites were consistently found to have near-neutral pHs corresponding to a generally higher OTU richness (Figure 12). Forest site samples fell across the pH range corresponding to a differential range of OTU richness (Figure 12). Site CA had an overall smaller OTU richness (t-test, $p < 0.005$) than all sites, except sites Al and H (Figure 10). This is further shown in Figure 12 where site CA samples tended to have lower richness (in some instances < 500 OTUs). Notably, samples obtained from site CA were in the lowest pH range (Figure 6 and 12).

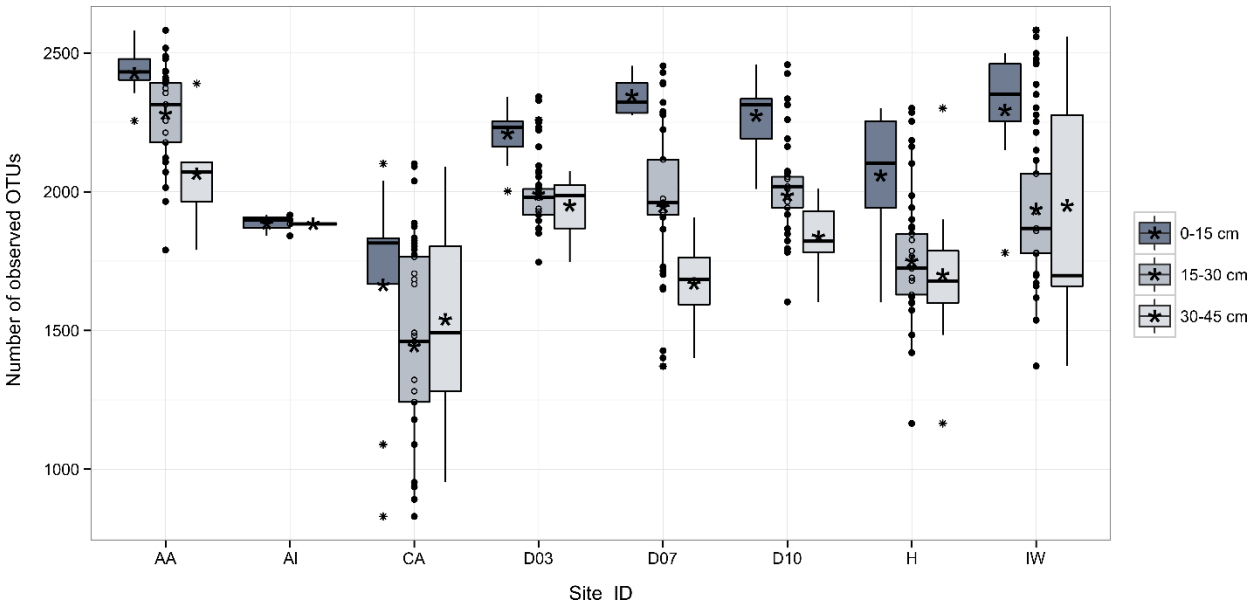

Figure 10: Number of observed OTUs for sampled sites and depth. Asterisks represent the average observed OTUs for each depth increment. Circles represent individual samples included for each site and stars represent outliers. Observed OTU richness is based on a rarefied OTU table (16,680 sequences/sample).

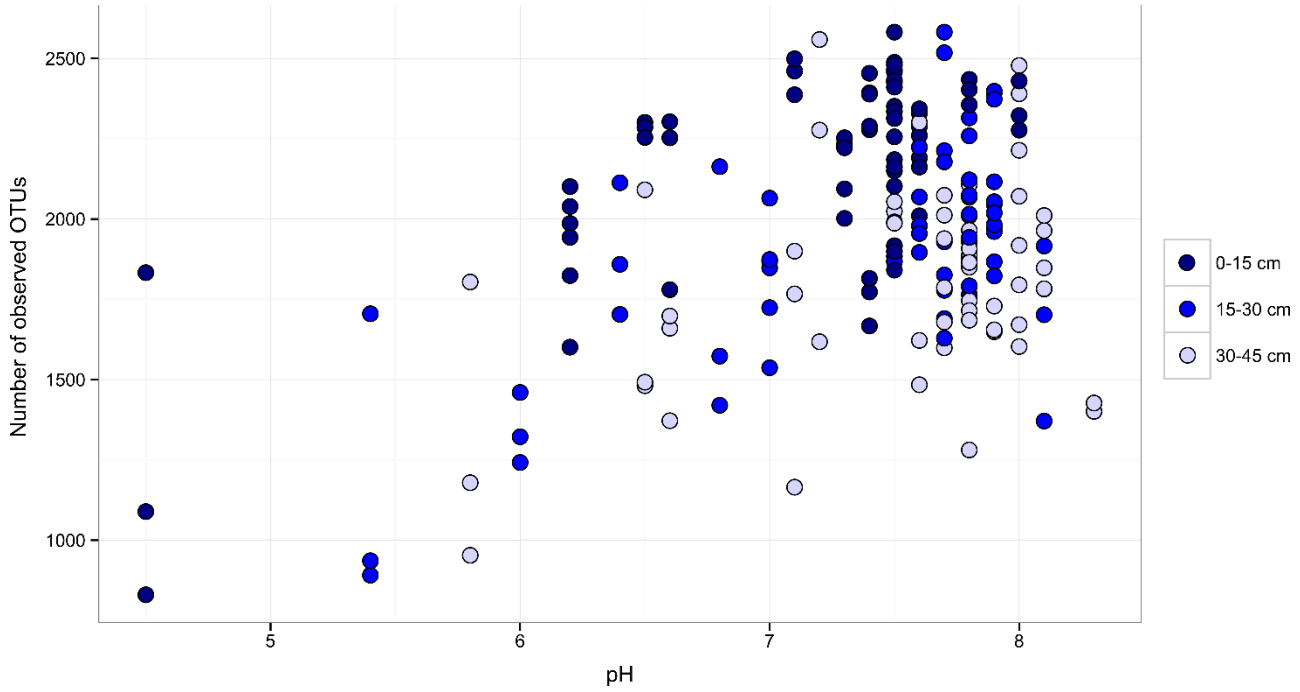

Figure 11: Number of observed OTUs in relation to pH and soil depth. Observed OTU richness based on a rarefied OTU table (16,680 sequences/sample).

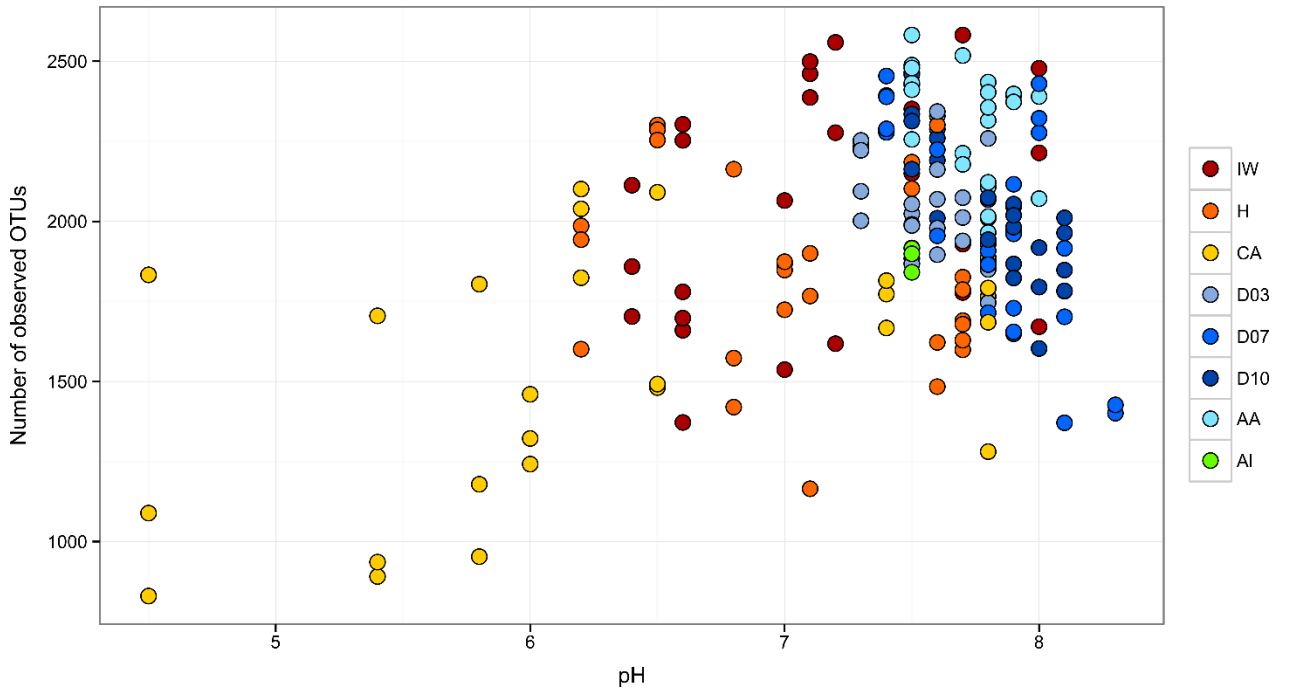

Figure 12: Number of observed OTUs in relation to pH and site (land-use type). Observed OTU richness based on a rarefied OTU table (16,680 sequences/sample).

3.3.2 Depth

The primary objective of this research was to characterize subsurface soil bacterial community composition across different land usages. To address this goal, soil depth was evaluated to 45 cm across 7 distinct sites at both fine and coarse spatial scales (i.e., 5 cm and 15 cm increments, respectively). Ordinations of amplified 16S rRNA gene datasets that were generated from soils sampled across the ***rare*** Charitable Research Reserve provided evidence of consistent differences in bacterial community composition with increasing soil depth (Figure 13A and B).

In general, top soil (i.e., 0–15 cm) bacterial communities grouped more closely than communities from greater depths within soil profiles (i.e., < 15 cm; Figure 13A). The influence of depth was significant at both spatial resolution scales ($p < 0.001$), albeit with relatively low within-group associations ($A = 0.052$ for the 15 cm data subsets; $A = 0.032$ for the 5 cm data subsets; Figure 13A and B). Although there was evidence for notable between-group separation at both spatial resolutions, the coarse spatial scale suggested stronger and more distinct depth-based groupings ($T = -20.1$; Figure 13A and C). This was further corroborated by ANOSIM results, which indicated that bacterial communities associated with each depth group were significantly, albeit weakly, different ($R = 0.166$ for the 15 cm data subset; $R = 0.105$ for the 5 cm data subset; Table 11). Top soil (0–15 cm) bacterial communities grouped strongly by site (Figure 14A), whereas samples located deeper within soil profiles were more variable (i.e., less distinct site-specific groupings; Figure 14B and C). The MRPP results for these constrained depth ordinations highlight shifts in the degree of sample groupings with depth where both the within-group associations and between-group separation both become less apparent (i.e., decreasing A, increasing T; Figure 14D).

To further assess the effect of depth across land-use types, data were ordinated based on overall land-use types (i.e., forest and field environments) (Figure 15A and C). Field sites showed much stronger depth and site groupings ($A = 0.117, 0.131$, respectively; $p < 0.001$) than forested sites ($A = 0.030, 0.044$, respectively; $p < 0.001$) (Figure 15A and C). Results from ANOSIM (Table 11) and NPMANOVA further reinforced these observations (Figure 17). Reduced separation of deeper samples was apparent even after consideration of the strong “arch effect” (an ordination artifact caused by the unimodal distribution of taxa along environmental gradients) that was observed for the forested sites, where there was more overlap with depth groupings (DCA, Figure 28, Appendix A). In particular, NPMANOVA highlighted that depth explained a larger majority of the variation in the field dataset (i.e., field samples only; 16.7%, $p = 0.001$) than in the forest dataset (i.e., forest samples only; 5.1%, $p = 0.001$) (Figure 17).

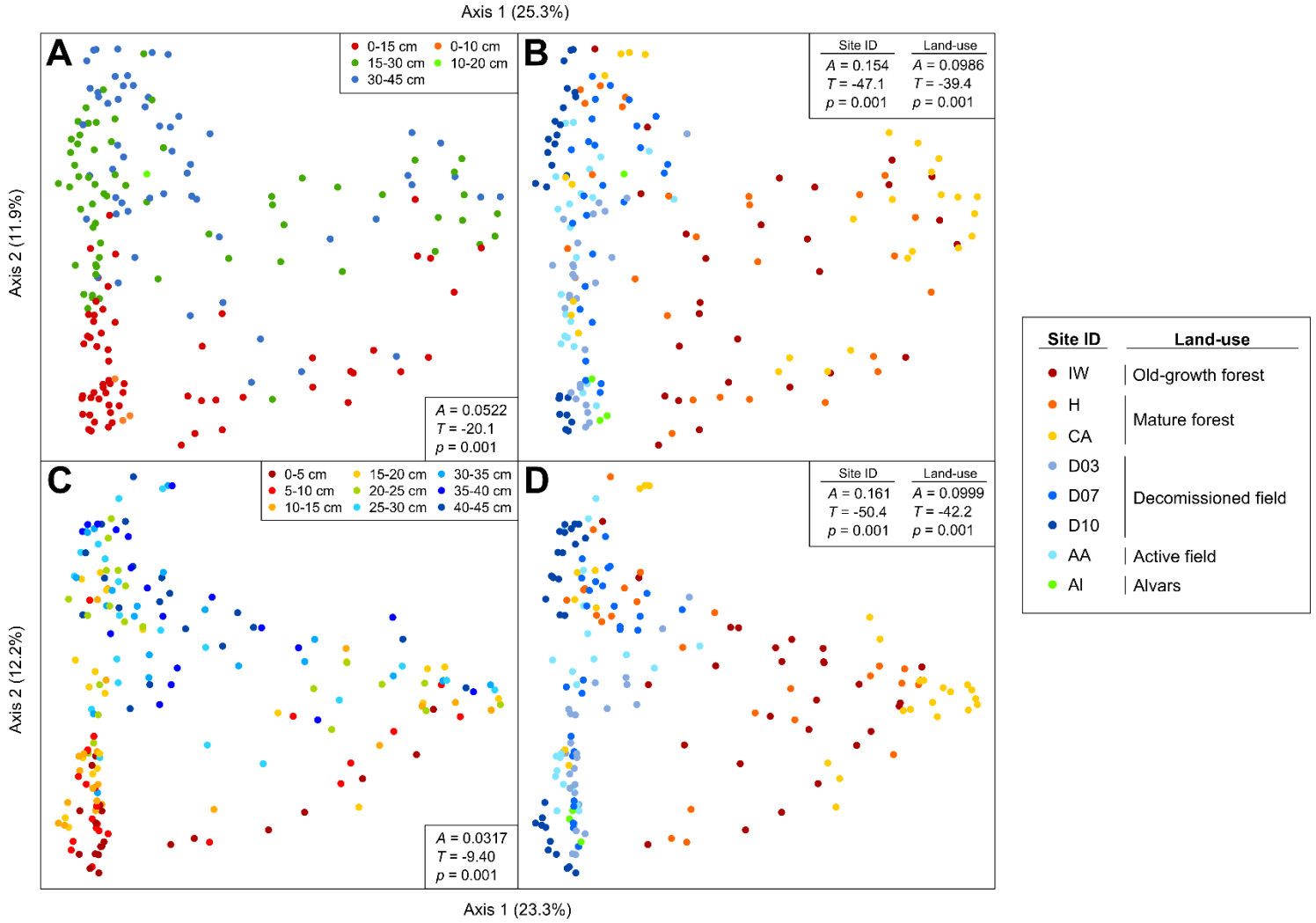

Figure 13*:* PCoA ordinations based on Bray-Curtis dissimilarities showing the effect of depth (A and C) and land-use type (B and D) on soil bacterial communities from the ***rare*** Charitable Research Reserve at coarse (15 cm; C and D) and fine (5 cm; B and D) scale depth increments. Depths at 0–10 and 10–20 cm correspond to alvars sites which were sampled at different depth increments. For panels B) and D), individual sites and the associated land-use type are shown in the legend (right). Insets show MRPP results.

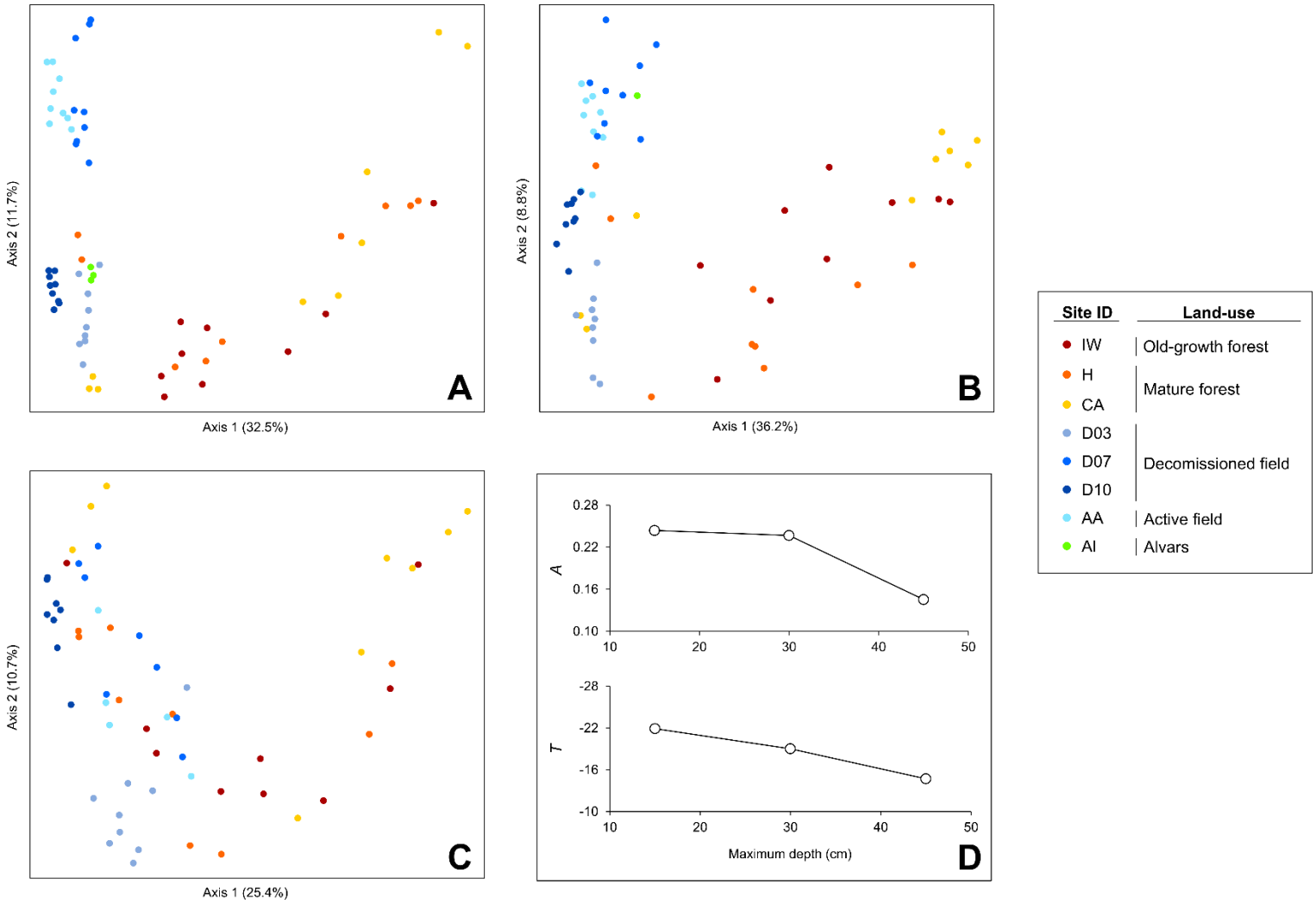

Figure 14: PCoA ordinations based on Bray-Curtis dissimilarities and MRPP results for depth-constrained ordinations. A) 0–15 cm, B) 15–30 cm, C) 30–45 cm, D) MRPP results ($p = 0.001$) for each depth-constrained ordination showing the *A* and *T* statistics.

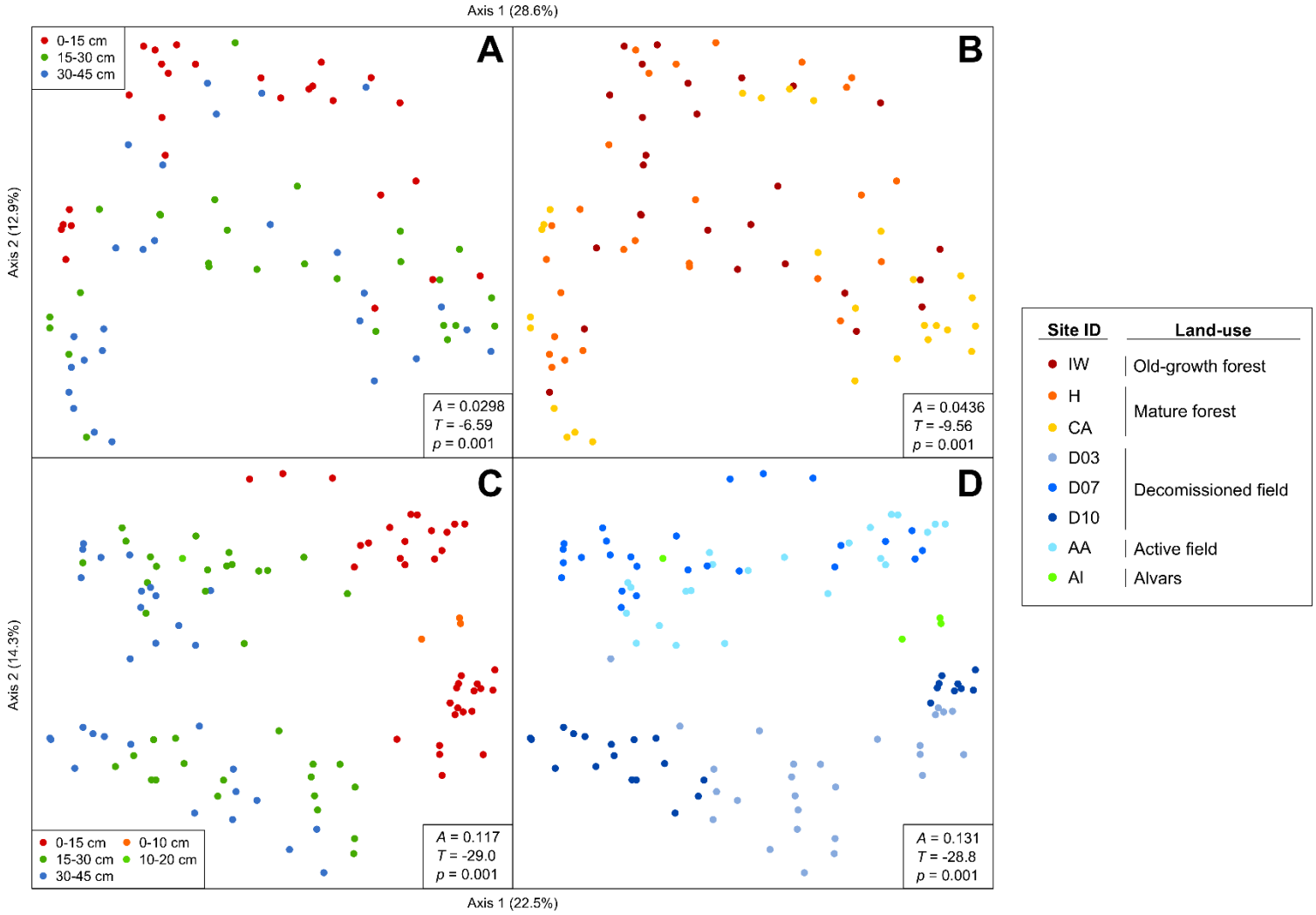

Figure 15: PCoA ordinations based on Bray-Curtis dissimilarities for forest (A and B) and field (C and D) datasets by depth (A and C) and site (B and D). Depths at 0–10 and 10–20 cm correspond to alvars sites which were sampled at different depth increments. For panels B) and D), individual sites and the associated land-use type are shown in the legend (right). Insets show MRPP results.

Table 11: ANOSIM results for depth, land-use type, and site ID for various data subsets.

	Depth		Land-use type		Site ID	
Data subset	*R* statistic	*p*	*R* statistic	*p*	*R* statistic	*p*
15 cm	0.17	0.001	0.36	0.001	0.39	0.001
5 cm	0.11	0.001	0.38	0.001	0.43	0.001
Forests	0.12	0.001	-0.06	0.991	0.16	0.001
Fields	0.49	0.001	0.09	0.045	0.45	0.001

In this work, soil horizon characterization was also conducted (Table 15, Appendix A). However, observed trends (i.e., PCoA; data not shown) were much less apparent, compared to depth specifically, likely resulting from inconsistencies in qualitative assessments of soil horizons. For soils across the ***rare*** Charitable Research Reserve characterizing horizons was particularly challenging due to transitional zones between horizons. Importantly, all of the 0–15 cm soil samples were consistently and qualitatively assigned to A-horizons, whereas there was more variability in defining the boundary between A and B or B and C horizons. Considering the nature of study, predefined depth increments were deliberately chosen for sampling to obtain a comparable assessment of bacterial communities with depth across a range of sites, As a result, the soil horizon data was used primarily for initial exploration of the datasets.

3.3.3 pH effects

A prominent trend observed across all sites was in relation to soil pH, as shown by the overall shift in bacterial community composition across ordination space (Figure 16). NPMANOVA further highlighted that soil pH was a strong and significant ($p = 0.001$) edaphic factor shaping bacterial communities explaining as much as ~20% of the total variation in the datasets (based on PCoA ordinations; Figure 17). When the effect of pH was assessed at each depth level (i.e., 0–15 cm only, 15–30 cm only, 30–45 cm only), pH was found to influence bacterial communities more strongly in top soils (24.5%, in the 0–15 cm increment, 28.7% 15–30 cm increment, $p = 0.001$) than in deeper soils (20.3% in the 30–45 cm increment, $p = 0.001$) (full data not shown). In addition, pH explained more of the variation in the forest (~19%, $p = 0.001$) than in the field datasets (~14%, $p = 0.001$) (Figure 17).

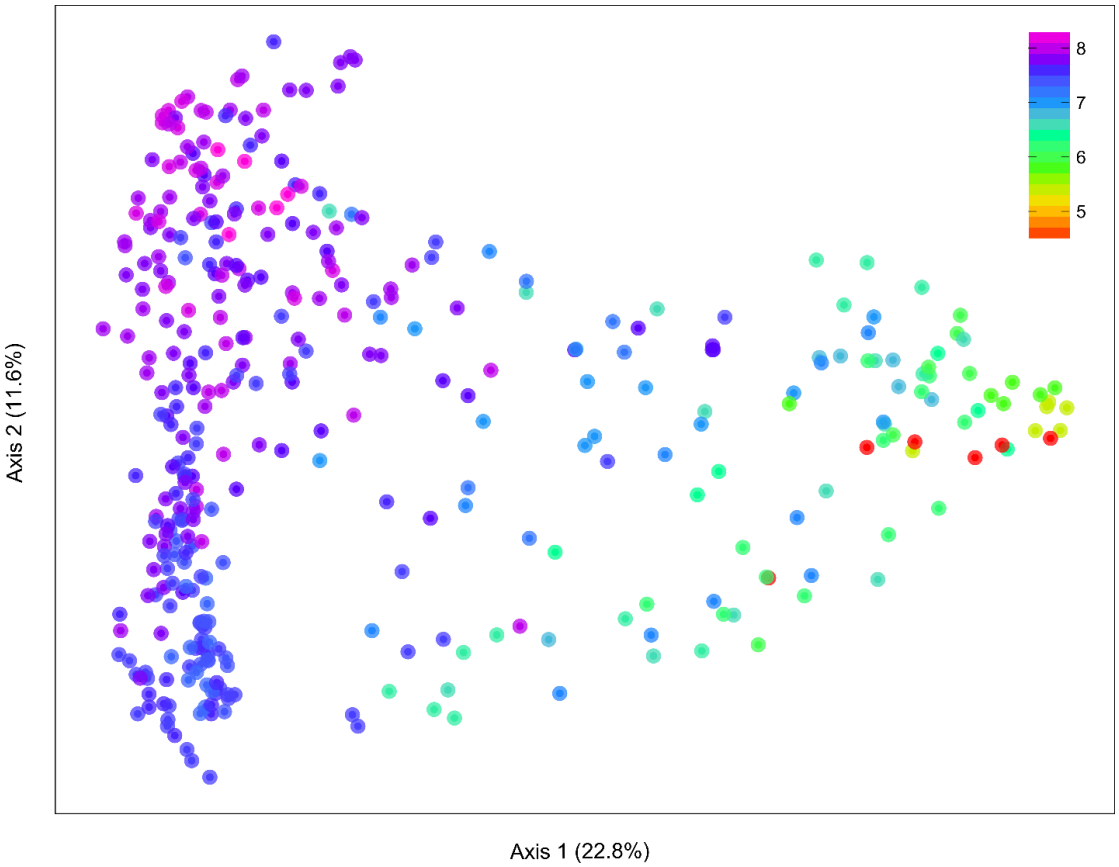

Figure 16: PCoA ordination based off Bray-Curtis dissimilarities of samples in relation to pH, for all samples analyzed.

3.3.4 Land-use responses

Visual assessment of ordinations indicated that bacterial communities were also strongly influenced by site ID (Figure 13B and D**)**. When grouped by individual sites, MRPP results suggested a relatively strong and significant ($p < 0.001$) difference in bacterial communities at both spatial scales (e.g., 15 cm and 5 cm increment datasets; $A = 0.154$, 0.161, $T = -41.1$, -50.4 for the 15 cm and 5 cm datasets, respectively) (Figure 13B and D). This was further highlighted by NPMANOVA results (Figure 17), which showed that land-use type explained the highest % of the variation (18.8%, $p = 0.001$) in the "Full dataset" (i.e., all samples ordinated). Moreover, ordinations showed that samples from field sites (D03, D07, D10, AA), as well as the alvars sites (Al), were more similar in bacterial community composition than the forest sites (IW, H, CA) (Figure 13B and D). Further visual assessment of the land-use type ordinations revealed that, in general, samples could be primarily grouped as forest or field communities (Figure 13B and D). When samples were ordinated by this classification, field sites showed much stronger site groupings ($A = 0.117$, 0.131, respectively; $p < 0.001$) than forested sites ($A = 0.030$, 0.044, respectively; $p < 0.001$) (Figure 15B and D). Results from ANOSIM (Table 11) and NPMANOVA (Figure 17) further highlighted this trend. In particular, NPMANOVA highlighted that site ID explained a larger majority of the variation in the field dataset (i.e., field samples only; ~25%, $p = 0.001$) than in the forest dataset (i.e., forest samples only; ~9%, $p = 0.001$) (Figure 17).

3.4 Additional edaphic factors

Aside from depth, land-use type, site ID, and pH, additional soil characteristics influenced bacterial communities across the ***rare*** Charitable Research Reserve. Across the full dataset, soil texture, plant richness, SIC, soil parent material, and NO_3^- were found to explain between 7–13% of the variation in the datasets (Figure 17). When dissimilarity matrices, representing data only from the fields and forested environments, were assessed via NPMANOVA, there were considerable differences in the dominant variables identified (Figure 17). For the field sites, the top three explanatory variables were soil texture (25.5%, $p = 0.001$), site location (25.3%, $p = 0.001$), and depth (16.7%, $p = 0.001$). In contrast, the top three variables influencing the forested sites were pH (19.0%, $p = 0.001$), texture (16.2%, $p = 0.001$), and SIC (14.5%, $p = 0.001$).

CCA plots constructed for the "full" dataset (i.e., 15 cm increment), as well as the field and forest dataset subsets, were consistent with NPMANOVA results and demonstrated that a variety of physicochemical factors operated in the same direction as depth and land-use type (Figure 18 and Figure 19). In the full dataset, NH_4^+, NO_3^-, NO_2^-, % H_2O, % sand (including fine and very fine sand), silt, clay, and $C_{(t)}$ generally shifted similarly in the direction of depth, as indicated by the directions

and length of the environmental vectors (Figure 18). In contrast, pH appeared to change most strongly along the same axis as the direction of change for land-use type, indicated both by the vector length and direction as well as sample points (Figure 18). In addition, coarse textural fractions (i.e., % gravel and coarse sand) were also shown to change along the same axis as pH and land-use type (Figure 18). In the forest-only ordination (Figure 19A), results were consistent with the NPMANOVA but demonstrated a strong "arch effect"; depth and pH gradients were generally found to operate in a similar direction relative to both the "full" and "forest" datasets (Figure 18, Figure 19A and B) (204). In the fields only ordination, changes in % sand (including fine sand), silt, and H_2O appeared to shift most strongly along the same axis as the direction of change for land-use type (i.e., sample points for individual sites; Figure 19B) (168).

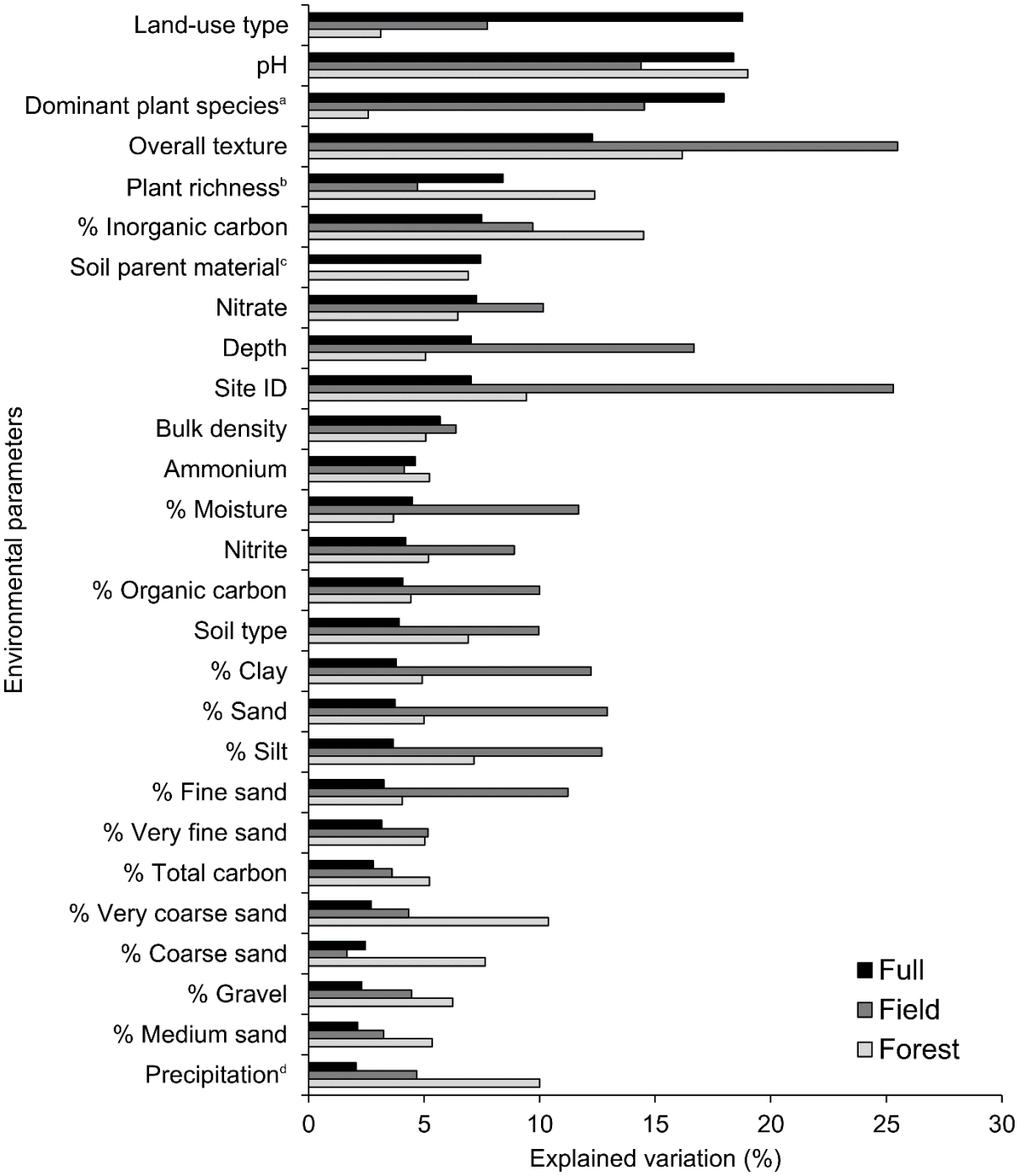

Figure 17: Environmental factors explaining variation in community dissimilarity (Bray-Curtis) from samples taken across the ***rare*** Charitable Research Reserve for the field, forest, and 15 cm increment dataset ($p \leq 0.005$). The 15 cm increment data were used to rank this plot (top to bottom) by decreasing explained variation (based on NPMANOVA). The values for each dataset do not sum to 100% because NPMANOVA analysis was conducted on each metadata category individually. [a,b]based on visual approximation; [c]all field sites classified as the same; [d]precipitation accumulation (i.e., accumulation over 24 h).

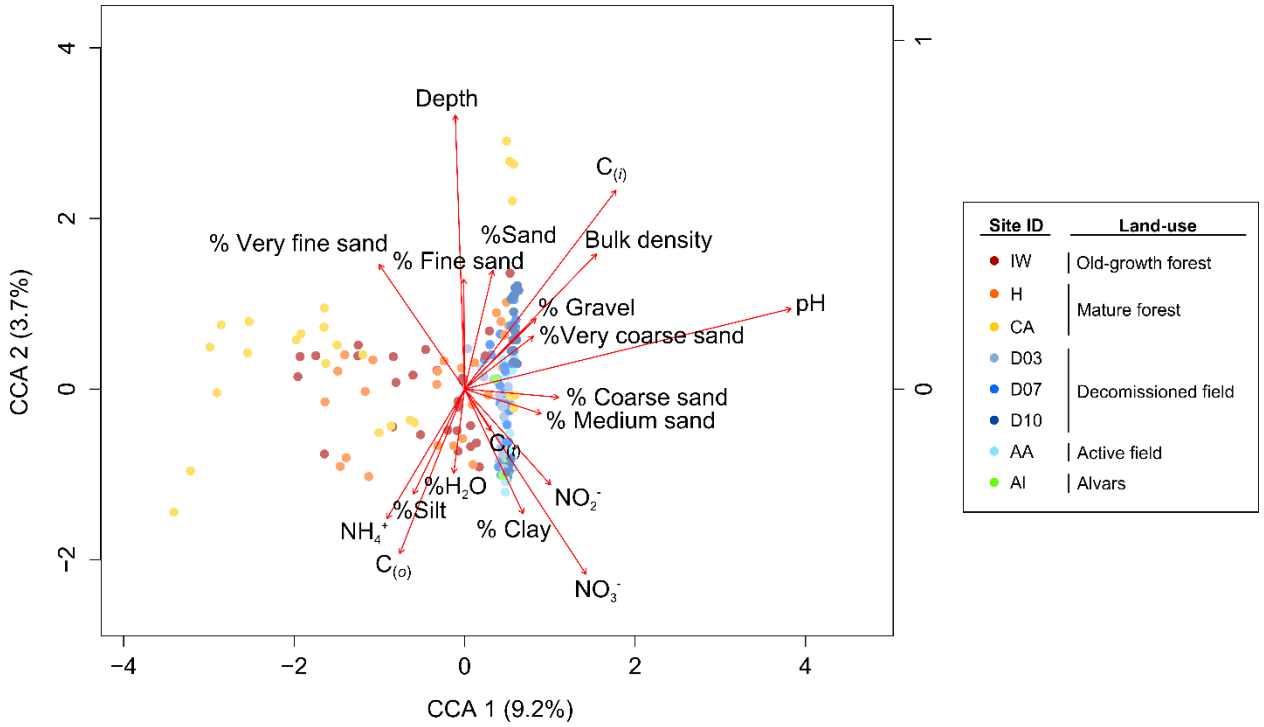

Figure 18: CCA ordination biplot showing the 15 cm data subset of soil samples taken from the ***rare*** Charitable Research Reserve and continuous environmental factors. Percentages on the CCA axes represent the total inertia (variance in species dispersion) explained by constrained environmental variables. Significant correlations ($p < 0.001$) of environmental variables and soil samples are shown as red arrows.

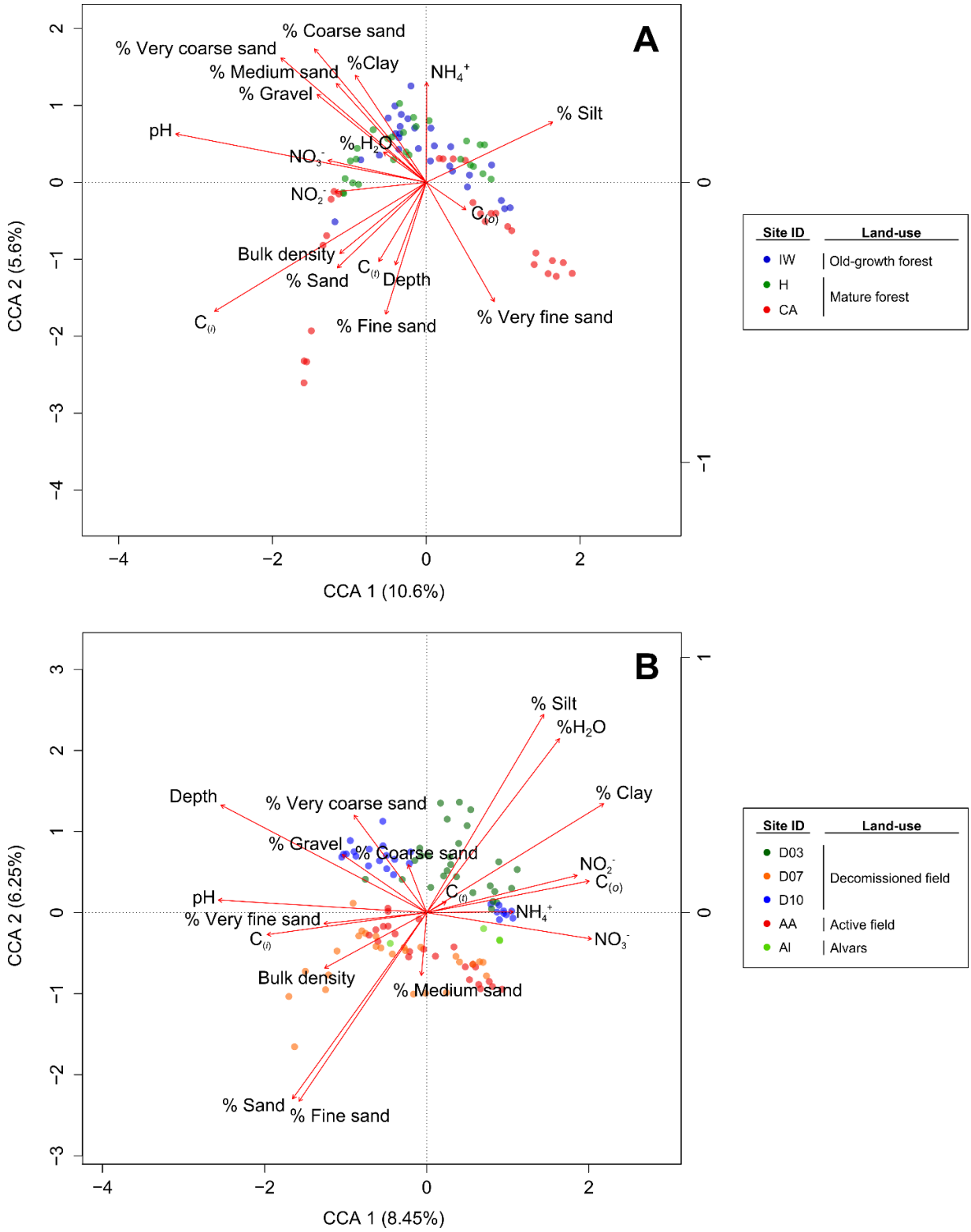

Figure 19: CCA ordination biplot showing A) forest data subset and B) field data subset of soil samples taken from the ***rare*** Charitable Research Reserve and continuous environmental factors. Percentages on the CCA axes represent the total inertia (variance in species dispersion) explained by constrained environmental variables. Significant correlations ($p < 0.001$) of environmental variables and soil samples are shown as red arrows.

3.5 Exploring taxonomic distributions across the *rare* Charitable Research Reserve

3.5.1 Taxa associations

Across the ***rare*** Charitable Research Reserve, 16S rRNA gene sequences were affiliated with members of the *Proteobacteria* (33.2%), *Actinobacteria* (27.8%), *Acidobacteria* (14.9%), *Chloroflexi* (6.6%), *Gemmatimonadetes* (4.7%), *Bacteroidetes* (3.0%), *Nitrospirae* (2.1%), *Firmicutes* (2.3%), *Verrucomicrobia* (1.7%), and *Latescibacteria* (formerly WS3; 1.2%) phyla. Based on a pairwise Wilcoxon rank sum test, consistent and significant shifts in the relative abundance of specific taxa with depth were observed at coarse-scale (i.e., 15 cm increments; Figure 20A) as well as fine-scale increments (i.e., 5 cm increments; Figure 20B). To further explore patterns of dominant taxa with depth, as well as across the pH and land usage gradient, CCA biplots were constructed (Figure 21). OTUs associated with the *Proteobacteria* phylum dominated ordinations and were observed across the entire pH and depth gradients. Although many OTUs were observed along the same axis as depth, robust affiliations of specific bacterial groups were more apparent with pH (Figure 21). Many members of the *Acidobacteria* phylum appeared to be strongly influenced by pH. In particular, members from the *Acidobacteriia* and the DA052 candidate class appeared to have many more OTUs associated with acidic environments (opposite direction of the pH vector). Similar to *Proteobacteria*, members from the *Actinobacteria* phylum were found across all soils, with the exception of the *Rubrobacteria* class, which appeared to have OTUs distributed across the depth gradient.

To assess which OTUs were most strongly correlated with depth and pH, Spearman's rank correlation analysis was conducted for all taxa in the 15 and 5 cm increment datasets (Table 12 and Table 13). There were more negatively than positively correlated taxa with depth (49 compared to 13, respectively) with most taxa from the *Actinobacteria*, *Proteobacteria*, and *Verrucomicrobia* phyla (Table 12). All taxa that were positively correlated with depth were unclassified representatives (Table 12). There were more OTUs that were found to be correlated with pH than depth. In addition, more OTUs were observed to be positively than negatively correlated with pH (54 compared to 45, respectively) (Table 13). Among the top positively and consistently correlated OTUs with pH were members from *Nitrospiraceae*, S0208, *Gemmatimonadetes*, and BPC102 from the *Nitrospirae*, *Chloroflexi*, *Gemmatimonadetes*, and *Acidobacteria* phyla, respectively (Table 13). Members from the groups *Solibacterales*, *Candidatus* Solibacter, *Acidobacteriaceae*, *Rhodoplanes*, and Ellin329 from the *Acidobacteria* and *Proteobacteria* phyla, respectively, were found to be consistently negatively correlated with pH (Table 13).

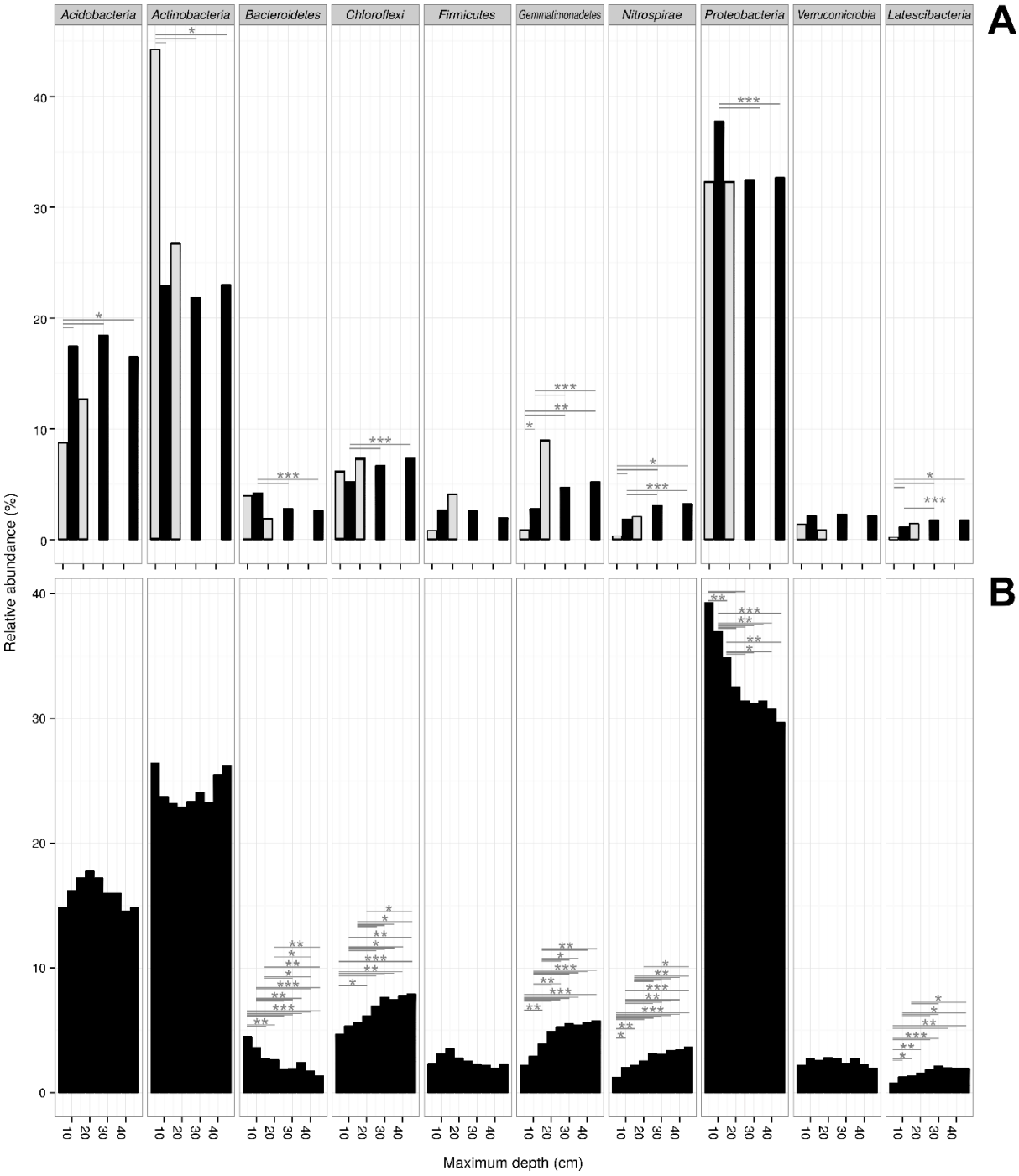

Figure 20: Shifts in relative abundance of the top ten phyla with depth from soils taken from the ***rare*** Charitable Research Reserve. Panel A) shows broad-scale (15 cm increment) shifts in taxa relative abundance (alvars sites are shown in grey and were sampled in "broad" increments of 10 and 20 cm) and panel B) shows fine-scale (5 cm increment) shifts associated with each depth increment. Bacterial phyla with a statistically significant differences in abundance between depth increments (based off the pairwise Wilcoxon rank sum test) are highlighted by a grey bar with the level of significance marked with asterisks (* = $p < 0.05$, ** = $p < 0.01$, *** $p < 0.001$).

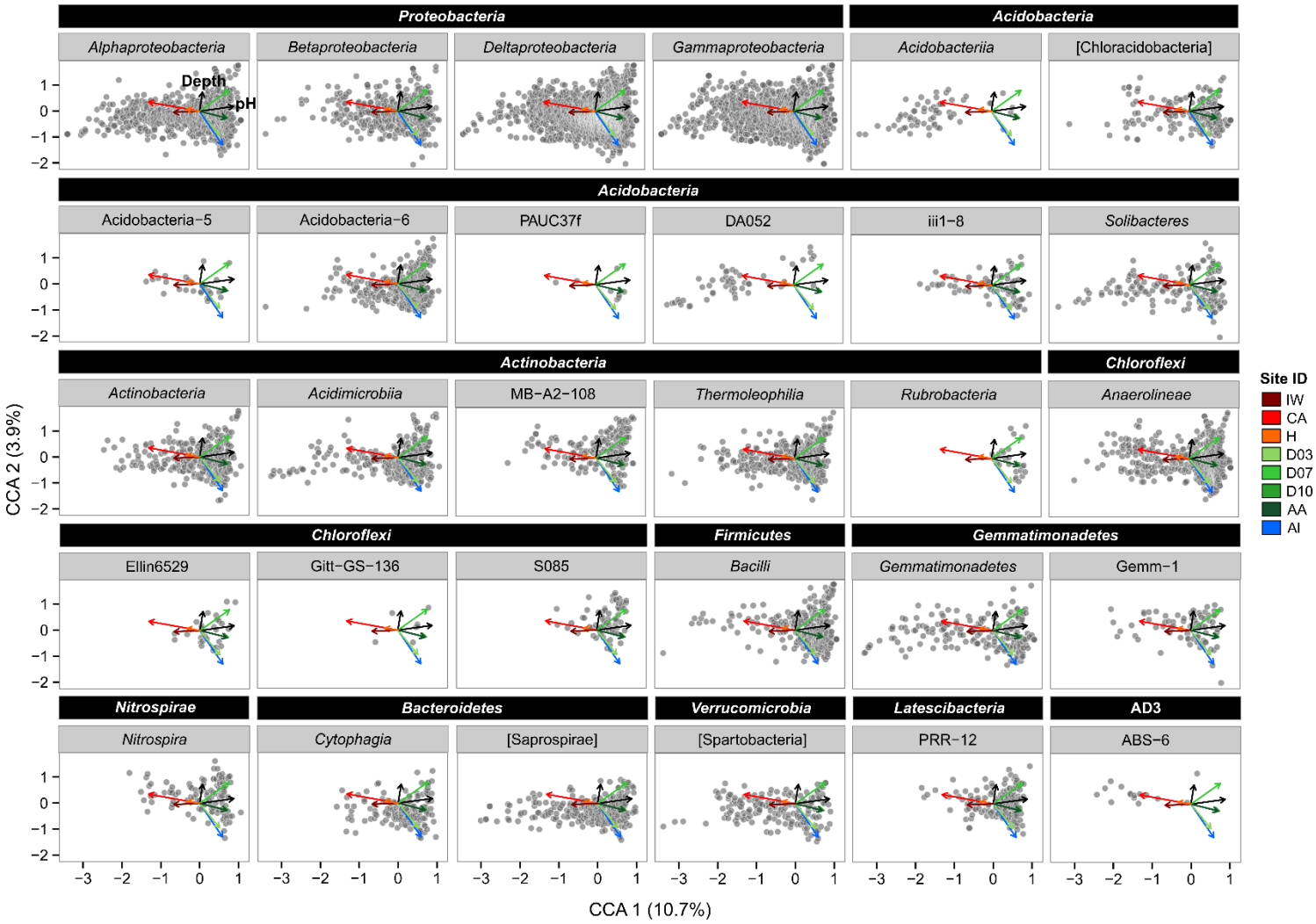

Figure 21: CCA biplot showing OTU distributions among the most abundant classes of bacteria. Significant correlations ($p < 0.001$) of environmental variables (depth, pH, site ID) and soil samples are shown as vectors (arrows). Vectors represent the (increasing) direction of both the pH and depth (labels) as well as land-use (right legend) gradients across ordination space. Percentages on the CCA axes represent the total inertia (variance in species dispersion) explained by pH, depth, and land-use type (site ID).

Table 12: Spearman's rank correlation analysis showing correlated (based on Spearman's ρ) taxa with depth for both the 15 and 5 cm increment datasets. Only taxa that were significantly ($p \leq 0.001$) correlated with depth and had a $\rho \geq 0.5$ are reported. Results are ranked by the average ρ between the 15 and 5 cm increment datasets. Dashes represent taxa that were not found to be correlated in the respective dataset based on the criteria described above.

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Actinobacteria	MB-A2-108	0.57	0.66
NC10	12-24	0.58	0.62
Nitrospirae	*Nitrospira*; *Nitrospirales*; 0319-6A21	-	0.59
Gemmatimonadetes	Gemm-1	0.56	0.60
GN04	MSB-5A5	-	0.58
Chloroflexi	P2-11E	0.56	0.57
Chloroflexi	*Anaerolineae*; SB-34	0.52	0.60
SBR1093		0.54	0.57
Acidobacteria	BPC102	0.54	0.57
GAL15		0.53	0.57
Chloroflexi	SAR202	-	0.54
Acidobacteria	iii1-8	0.57	0.51
Other[c]		0.51	0.53
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Pseudonocardiaceae*; *Pseudonocardia*	-0.69	-0.63
Bacteroidetes	*Sphingobacteriia*; *Sphingobacteriales*	-0.63	-0.62
Proteobacteria	*Gammaproteobacteria*; *Thiotrichales*; *Piscirickettsiaceae*	-0.65	-0.55
Verrucomicrobia	[Pedosphaerae]; [Pedosphaerales]; Ellin517	-0.60	-0.59
Proteobacteria	*Deltaproteobacteria*; *Bdellovibrionales*; *Bdellovibrionaceae*; *Bdellovibrio*	-0.57	-0.59
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Bradyrhizobiaceae*; Other	-0.65	-0.51
Verrucomicrobia	[Pedosphaerae]; [Pedosphaerales]; auto67_4W	-0.57	-
Verrucomicrobia	*Verrucomicrobiae*; *Verrucomicrobiale*s; *Verrucomicrobiaceae*; Other	-0.54	-0.61
Actinobacteria	Actinobacteria; Actinomycetales; *Streptomycetaceae*; *Streptomyces*	-0.57	-
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Burkholderiaceae*; Other	-	-0.57

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Verrucomicrobia	[Pedosphaerae]; [Pedosphaerales]; Other	-0.57	-
Proteobacteria	*Alphaproteobacteria*; *Sphingomonadales*; *Sphingomonadaceae*	-0.56	-
Proteobacteria	*Alphaproteobacteria*; *Rhodobacterales*; *Rhodobacteraceae*; *Amaricoccus*	-0.57	-0.54
Proteobacteria	*Deltaproteobacteria*; *Spirobacillales*	-0.55	-
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Nocardioidaceae*; *Nocardioides*	-0.55	-0.54
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Intrasporangiaceae*; *Terracoccus*	-0.54	-
Proteobacteria	*Deltaproteobacteria*; *Myxococcales*; *Myxococcaceae*; *Myxococcus*	-	-0.54
Proteobacteria	*Deltaproteobacteria*; MIZ46	-	-0.54
Proteobacteria	*Deltaproteobacteria*; *Myxococcales*	-0.57	-0.52
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Comamonadaceae*; *Rubrivivax*	-0.54	-
Bacteroidetes	*Flavobacteriia*; *Flavobacteriales*; *Cryomorphaceae*	-0.55	-0.52
Gemmatimonadetes	*Gemmatimonadetes*; KD8-87	-0.54	-
Acidobacteria	Acidobacteria-6; iii1-15; RB40	-0.53	-
Verrucomicrobia	*Verrucomicrobiae*; *Verrucomicrobiales*; *Verrucomicrobiaceae*	-0.53	-
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Burkholderiaceae*	-0.56	-0.51
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Hyphomicrobiaceae*; Other	-	-0.53
Bacteroidetes	*Cytophagia*; *Cytophagales*; *Cytophagaceae*; *Adhaeribacter*	-0.54	-0.52
Bacteroidetes	[Saprospirae]; [Saprospirales]; *Chitinophagaceae*	-0.54	-0.52
Proteobacteria	*Alphaproteobacteria*; *Rhodospirillales*	-0.53	-
Actinobacteria	*Acidimicrobiia*; *Acidimicrobiales*; C111	-0.55	-0.50
Proteobacteria	*Deltaproteobacteria*; [Entotheonellales]; [Entotheonellaceae]; *Candidatus* Entotheonella	-	-0.52
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Comamonadaceae*; *Ramlibacter*	-0.52	-
Bacteroidetes	*Cytophagia*; *Cytophagales*; *Cytophagaceae*	-0.52	-
Actinobacteria	*Thermoleophilia*; *Gaiellales*; AK1AB1_02E	-0.51	-0.52
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Nocardioidaceae*	-0.51	-
Acidobacteria	Acidobacteria-6; iii1-15; mb2424	-0.51	-
Acidobacteria	[Chloracidobacteria]; RB41; Ellin6075	-0.51	-
Proteobacteria	*Deltaproteobacteria*; *Bdellovibrionales*; *Bacteriovoracaceae*; Other	-	-0.51

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Intrasporangiaceae*	-0.50	-0.52
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Bradyrhizobiaceae*	-0.51	-
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Rhizobiaceae*; *Agrobacterium*	-0.51	-
Bacteroidetes	[Saprospirae]; [Saprospirales]; *Saprospiraceae*	-0.51	-
Proteobacteria	*Gammaproteobacteria*; *Xanthomonadales*; *Sinobacteraceae*	-	-0.51
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Mycobacteriaceae*; *Mycobacterium*	-0.50	-
Proteobacteria	*Betaproteobacteria*; *Nitrosomonadales*; *Nitrosomonadaceae*; *Nitrosovibrio*	-	-0.50
Chloroflexi	TK10; B07_WMSP1	-0.50	-
Bacteroidetes	*Cytophagia*; *Cytophagales*; *Cytophagaceae*; *Sporocytophaga*	-	-0.50
Verrucomicrobia	[Spartobacteria]; [Chthoniobacterales]; [Chthoniobacteraceae]	-0.50	-
Proteobacteria	*Gammaproteobacteria*; Other	-0.50	-

[a]15 cm increment dataset (n=187; 17,549 OTUs); [b]5 cm increment dataset (n=189; 19,640 OTUs); [c]Other refers ambiguous classification (via RDP)

Table 13: Spearman's rank correlation analysis showing correlated (based on Spearman's ρ) taxa with pH for both the 15 and 5 cm increment datasets. Only taxa that were significantly ($p \leq 0.001$) correlated with pH and had a $\rho \geq 0.5$ are reported. Results are ranked by the average ρ between the 15 and 5 cm increment datasets. Dashes represent taxa that were not found to be correlated in the respective dataset based on the criteria described above. Taxa that were also observed to be indicators species (Figure 23) are bolded.

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Nitrospirae	*Nitrospira*; *Nitrospirales*; *Nitrospiraceae*	0.79	0.78
Chloroflexi	*Anaerolineae*; S0208	0.81	0.74
Gemmatimonadetes		0.74	0.77
Nitrospirae	*Nitrospira*; *Nitrospirales*; *Nitrospiraceae*; Other[c]	0.74	0.73
Acidobacteria	BPC102	0.70	0.70
Bacteroidetes	[Saprospirae]; [Saprospirales]; *Chitinophagaceae*; *Niabella*	0.68	0.68
Acidobacteria	S035	0.66	0.69
NC10	12-24	0.71	0.62
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*	0.67	0.66
Proteobacteria	*Gammaproteobacteria*; *Chromatiales*; *Ectothiorhodospiraceae*	0.67	0.66
Actinobacteria	**MB-A2-108; 0319-7L14**	**0.72**	**0.60**
Acidobacteria	BPC102; MVS-40	0.69	0.63
Chloroflexi	S085	0.65	0.67
Gemmatimonadetes	*Gemmatimonadetes*	0.65	-
Chloroflexi	TK10; AKYG885	0.64	0.62
Chloroflexi	*Chloroflexi*; *Chloroflexales*; Other; Other	0.59	0.67
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Glycomycetaceae*; *Glycomyces*	-	0.60
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Micrococcaceae*	-	0.59
Chloroflexi	*Anaerolineae*; SB-34	0.57	0.61
Acidobacteria	Acidobacteria-6; BPC015	0.53	0.64
Chloroflexi	*Anaerolineae*	-	0.59
Proteobacteria	*Gammaproteobacteria*; *Xanthomonadales*; *Xanthomonadaceae*; *Pseudoxanthomonas*	0.59	-
Gemmatimonadetes		-	0.58
Actinobacteria	*Thermoleophilia*; *Gaiellales*; *Gaiellaceae*	0.59	0.57

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Firmicutes	*Bacilli*; *Bacillales*	-	0.58
Proteobacteria	*Gammaproteobacteria*; *Chromatiales*	0.57	0.57
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Microbacteriaceae*; *Agromyces*	0.57	-
NC10	wb1-A12	0.57	-
Proteobacteria	*Gammaproteobacteria*; *Xanthomonadales*; *Xanthomonadaceae*; *Thermomonas*	0.56	-
Proteobacteria	*Alphaproteobacteria*; *Rhodospirillales*; *Rhodospirillacea*e; *Inquilinus*	0.55	-
Proteobacteria	*Betaproteobacteria*; *Nitrosomonadales*; *Nitrosomonadaceae*	-	0.55
Firmicutes	***Bacilli*; *Bacillales*; *Paenibacillaceae*; *Paenibacillus***	**0.55**	**-**
GN04		0.50	0.59
Acidobacteria	Acidobacteria-6; CCU21	0.53	0.56
Chloroflexi	Gitt-GS-136	-	0.54
Actinobacteria	Other	0.54	-
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Oxalobacteraceae*; Other	0.54	-
Chloroflexi	*Chloroflexi*; [Roseiflexales]	0.54	-
Firmicutes	*Bacilli*; *Bacillales*; *Paenibacillaceae*; *Brevibacillus*	-	0.54
Verrucomicrobia	*Opitutae*; *Opitutales*; *Opitutaceae*; Other	0.53	-
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Hyphomicrobiaceae*; *Hyphomicrobium*	0.53	-
Chloroflexi	TK10	-	0.53
Actinobacteria	*Rubrobacteria*; *Rubrobacterales*; *Rubrobacteraceae*	0.53	-
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Actinosynnemataceae*	0.52	-
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Hyphomicrobiaceae*; *Devosia*	0.53	0.51
Gemmatimonadetes	C114	-	0.52
Proteobacteria	*Alphaproteobacteria*; *Sphingomonadales*; *Sphingomonadaceae*; *Kaistobacter*	0.52	-
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Micromonosporaceae*; *Catellatospora*	0.51	-
NC10		0.51	-
Chloroflexi	*Anaerolineae*; H39	0.51	0.50
Proteobacteria		-	0.50
Firmicutes	*Bacilli*; *Bacillales*; *Planococcaceae*; *Paenisporosarcina*	-	0.50

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Chloroflexi	*Anaerolineae*; pLW-97	-	0.50
Planctomycetes		-	0.50
Acidobacteria	*Solibacteres*; *Solibacterales*	-0.73	-0.70
Gemmatimonadetes	*Gemmatimonadetes*; Ellin5290	-0.71	-
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Hyphomicrobiaceae*; *Rhodoplanes*	-0.72	-0.70
Proteobacteria	*Alphaproteobacteria*; Ellin329	-0.74	-0.67
Acidobacteria	***Solibacteres*; *Solibacterales*; *Solibacteraceae*; *Candidatus* Solibacter**	**-0.73**	**-0.67**
Acidobacteria	*Acidobacteriia*; *Acidobacteriales*; *Acidobacteriaceae*	-0.70	-0.70
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Burkholderiaceae*; *Burkholderia*	-0.71	-0.67
Verrucomicrobia	[Spartobacteria]; [Chthoniobacterales]; [Chthoniobacteraceae]; DA101	-0.66	-0.69
Acidobacteria	***Solibacteres*; *Solibacterales*; *Solibacteraceae*; *Candidatus* Solibacter**	**-**	**-0.67**
Acidobacteria	*Acidobacteriia*; *Acidobacteriales*; *Acidobacteriaceae*; *Edaphobacter*	-0.64	-0.70
Verrucomicrobia	[Spartobacteria]; [Chthoniobacterales]; [Chthoniobacteraceae]; Other	-0.66	-
Chloroflexi	TK10; B07_WMSP1; FFCH4570	-	-0.65
Verrucomicrobia	[Pedosphaerae]; [Pedosphaerales]; Ellin515	-0.67	-0.62
Gemmatimonadetes	Ellin5290	-	-0.64
Chloroflexi	*Ktedonobacteria*; *Thermogemmatisporales*; *Thermogemmatisporaceae*	-0.61	-0.67
Acidobacteria	**DA052; Ellin6513**	**-0.66**	**-0.61**
Proteobacteria	*Alphaproteobacteria*; *Caulobacterales*; *Caulobacteraceae*; Other	-0.63	-
Chloroflexi	*Ktedonobacteria*; Other	-0.60	-0.64
Proteobacteria	*Alphaproteobacteria*; BD7-3	-0.61	-
Verrucomicrobia	[Pedosphaerae]; [Pedosphaerales]; [Pedosphaeraceae]; *Pedosphaera*	-0.62	-0.59
Proteobacteria	*Alphaproteobacteria*; *Caulobacterales*; *Caulobacteraceae*	-0.58	-0.61
AD3	JG37-AG-4	-0.62	-0.56
Proteobacteria	*Betaproteobacteria*; *Burkholderiales*; *Oxalobacteraceae*; *Collimonas*	-	-0.59
Actinobacteria	*Actinobacteria*; *Actinomycetales*; *Frankiaceae*	-0.59	-0.59
Proteobacteria	*Alphaproteobacteria*; *Rhodospirillales*; *Acetobacteraceae*; Other	-0.65	-0.51
Acidobacteria	***Acidobacteriia*; *Acidobacteriales*; *Koribacteraceae*; *Candidatus* Koribacter**	**-0.58**	**-0.58**

Taxon		Spearman's ρ	
		15 cm[a]	5 cm[b]
Proteobacteria	*Deltaproteobacteria*; *Myxococcales*; *Myxococcaceae*	-0.58	-
Planctomycetes	*Planctomycetia*; *Gemmatales*; *Isosphaeraceae*	-0.61	-0.53
Chloroflexi	*Thermomicrobia*; Ellin6537	-0.60	-0.54
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Beijerinckiaceae*; Other	-0.56	-0.57
AD3	Other	-0.54	-0.58
AD3	ABS-6	-0.57	-0.55
Chlamydiae	*Chlamydiia*; *Chlamydiales*; *Rhabdochlamydiaceae*; *Candidatus* Rhabdochlamydia	-0.55	-0.56
Gemmatimonadetes	*Gemmatimonadales*; Ellin5301	-	-0.55
Elusimicrobia	*Elusimicrobia*; FAC88	-0.59	-0.50
Elusimicrobia	*Elusimicrobia*; *Elusimicrobiales*	-0.54	-
Chloroflexi	*Ktedonobacteria*; Elev-1554	-0.57	-0.51
Gemmatimonadetes	*Gemmatimonadetes*; KD8-87	-0.53	-
Proteobacteria	*Alphaproteobacteria*; *Rhizobiales*; *Bradyrhizobiaceae*; *Bradyrhizobium*	-	-0.53
Proteobacteria	*Betaproteobacteria*; *Nitrosomonadales*	-	-0.53
Acidobacteria	[Chloracidobacteria]; Ellin7246	-0.51	-0.54
Actinobacteria	*Acidimicrobiia*; *Acidimicrobiales*; EB1017	-0.52	-
Chloroflexi	*Ktedonobacteria*; JG30-KF-AS9	-0.51	-0.52
Bacteroidetes	At12OctB3	-0.50	-

[a]15 cm increment dataset (n=187; 17,549 OTUs); [b]5 cm increment dataset (n=189; 19,640 OTUs); [c]Other refers ambiguous classification (via RDP)

3.5.2 Indicator species

When individual sites were assessed by depth, several putative indicator taxa were observed (Figure 22). *Syntrophobacteraceae* appeared to be indicative of deeper soils for both sites D10 and Al, whereas an unclassified *Betaproteobacteria* member and a member from the candidate class S035 from the *Acidobacteria* phylum were observed to be indicative of the active agricultural field. There were no overlapping depth related taxa when results from Figure 22 were compared to the taxa correlations in Table 12.

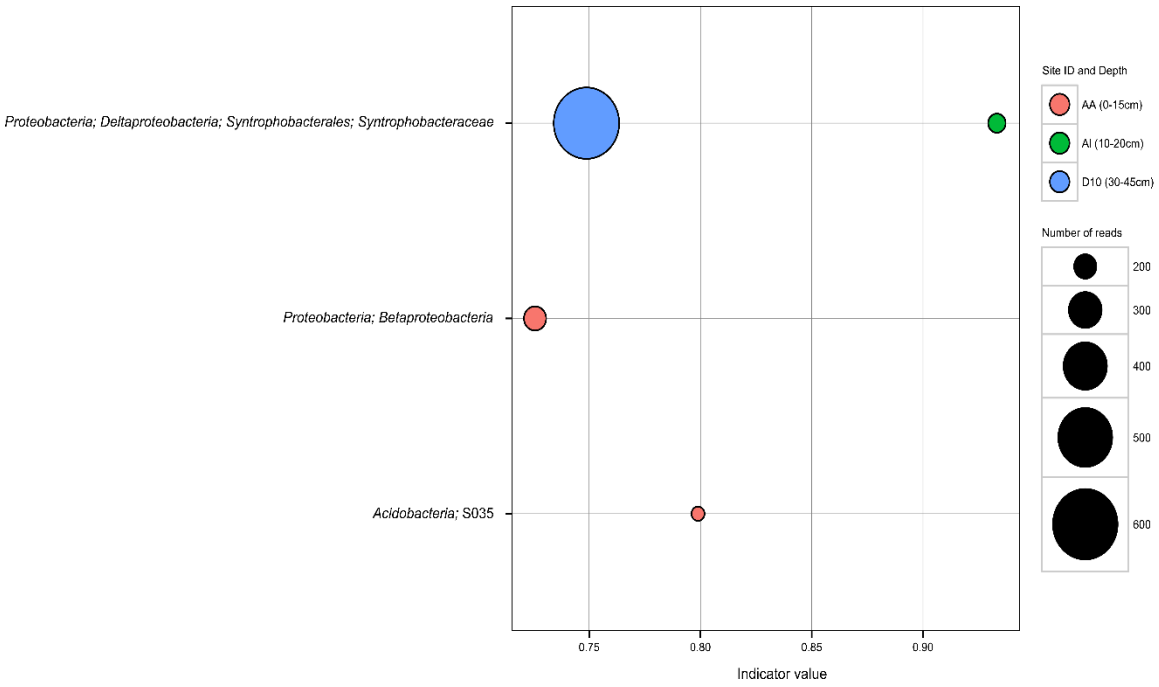

Figure 22: Indicator species analysis for individual sites and depth based on an indicator value cut-off of ≥ 0.7, sequence abundance ≥ 100, $p \leq 0.05$).

Furthermore, because pH played an important role in shaping bacterial distributions (Figure 21, Table 12), indicator species analysis was also conducted with respect to pH. Indicators of alkaline pH soil environments included members from the *Firmicutes* (*Shimazuella*, *Paenibacillus*), *Actinobacteria* (*Solirubrobacterales*, MB-A2-108), *Cyanobacteria*, and the *Latescibacteria* (formerly WS3) phyla (Figure 23). Indicators for acidic soils were exclusively from the *Acidobacteria* phylum, with the exception of a member of the candidate family 0319-6A21 from the *Nitrospirae* phylum (Figure 23). When these results were compared with taxa correlations in Table 13, there were observable consistencies. *Paenibacillus* and MB-A2-108, in addition to being indicative of more alkaline soils (Figure 23), were also found to be positively correlated with pH (Table 13).

Acidobacteria members *Candidatus* Solibacter, Ellin6513, and *Candidatus* Koribacter were found to be negatively correlated with pH in addition to being an observed indicator for low pH soils (Figure 23, Table 13).

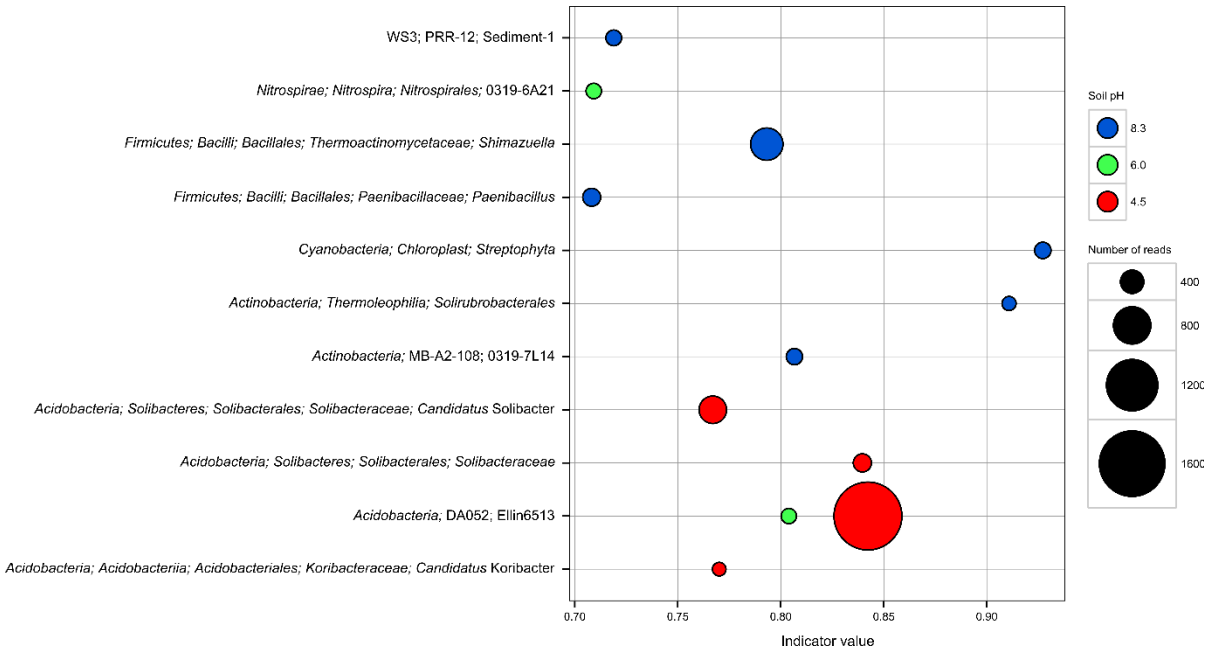

Figure 23: Indicator species analysis for pH based on an indicator value cut-off of ≥ 0.7, sequence abundance ≥ 100, $p \leq 0.05$).

3.6 Bacterial community function predictions

3.6.1 Predicted metagenome shifts with depth and across land-usage

In addition to assessing biogeographical patterns using 16S rRNA genes alone, an alternative approach is to assess potential functional and metabolic capabilities that exist throughout soil profiles based on taxonomy. Here, a broad analysis of the metabolic and functional capabilities throughout soil profiles and across land-use types was carried out using the PICRUSt algorithm, which predicts metagenomes based on existing 16S rRNA gene data (161).

After closed reference OTU picking, ~25% of the sequences were removed from the initial data subsets, generating an OTU table consisting of 15,897 OTUs. Of the 6,909 annotations in the KEGG reference database, ~90% of the genes were observed at least once in the data subsets (6,254). Overall, PICRUSt appeared to perform similarly to soil-based analyses conducted in initial validation efforts by Langille *et al*. (161); all samples had NSTI values within the range of approximately 0.14–0.24 (mean = 0.20; standard deviation = ± 0.02).

The majority of the predicted genes were associated with "metabolism" KEGG pathways (52.9%) including "purine metabolism", "peptidases", and "oxidative phosphorylation". Other dominant predicted KEGG pathways were from the "genetic information processing" (15.6%) and "environmental information processing" (13.0%) groups. Genes associated with "DNA repair and recombinaton proteins", "ribosomes", and "transcription factors" were among the most abundant predictions in the "genetic information processing" group. Genes associated with "(ABC) transporters", "two-component systems", and "bacterial secretion systems" were among the most abundant predictions in the "environmental information processing" group. Less than 5% of the total predicted genes were associated with the KEGG pathways "cellular processes", "human diseases", and "organismal systems". Unclassified genes represented 13.0% of the PICRUSt dataset with the majority of those predictions having unknown functions.

With depth, the PICRUSt results mimicked previously shown trends based on 16S rRNA genes in this study (see section 3.3). Strong depth-specific changes in bacterial community composition were not apparent for all sites (Figure 24A). The field sites showed stronger depth separation than the forested sites (Figure 24B and C). Shifts with depth operated generally along the y-axis (i.e., PC2). Consistent with visual observations, field sites had a higher explained variation by PC2 (Figure 24B), suggesting a stronger depth separation than the forest sites (Figure 24C). Furthermore, PCA plots of predicted metagenomes showed some degree of separation by land-use type (Figure 25). As might be expected, and similar to depth-based patterns, metagenome observations for the land-use type analysis mimicked trends observed based on 16S rRNA gene

analysis (Figure 13). Although the shifts in PC values are relatively small (0.01 for PC1 and 0.005 for PC2), the % variation explained by these principal components were fairly high (88.2% and 6.35%, respectively; 95% of the total explained variation) (Figure 25).

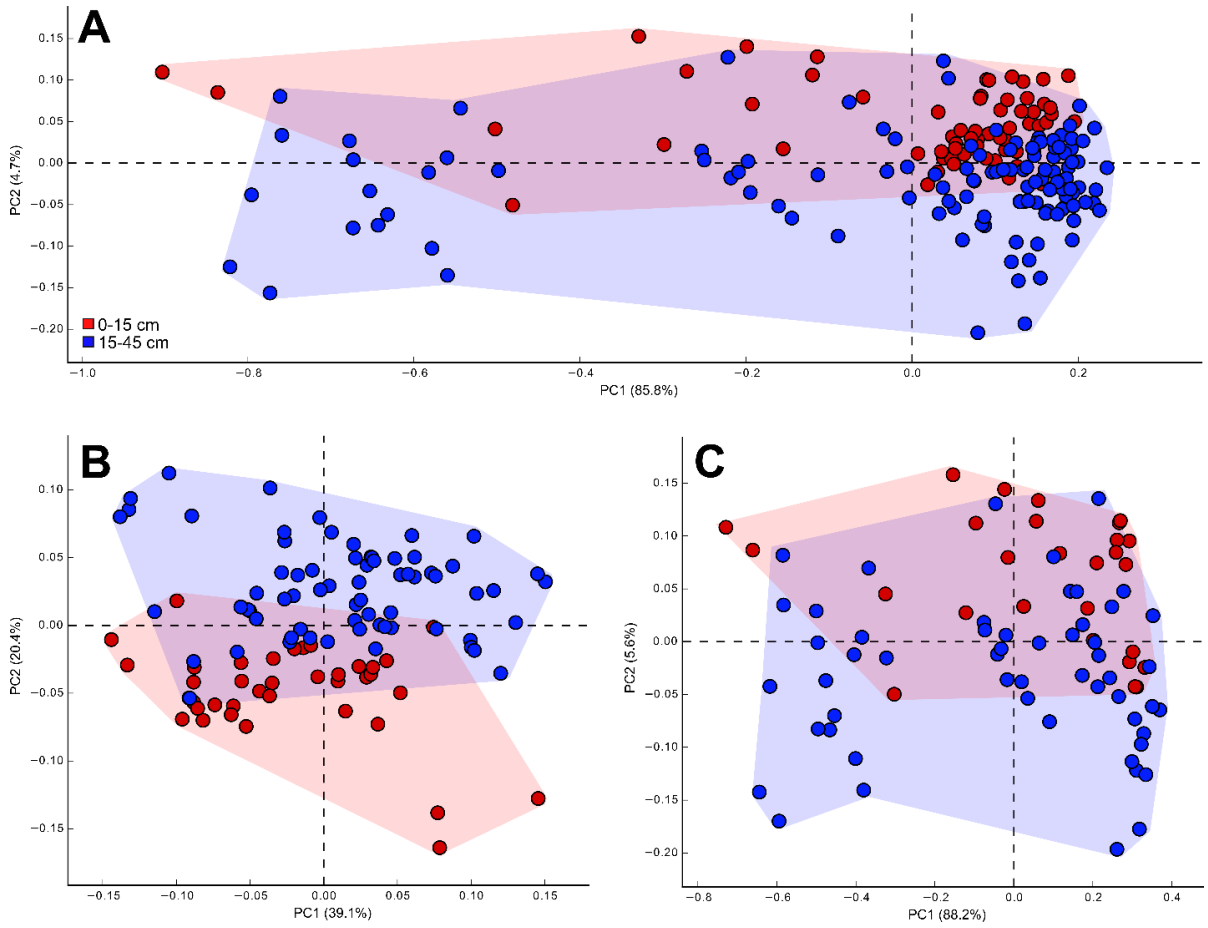

Figure 24: PCA analysis of predicted metagenomes with depth for the A) Full dataset, B) Field sites only, and C) Forest sites only. PC1 and PC2 represent the principal components explaining the variation in the dataset. Results are based off relative abundances of predicted genes at level three in the KEGG hierarchy using level two as the parent class (i.e., abundances of genes at level three relative to sum of genes at level two in the KEGG hierarchy).

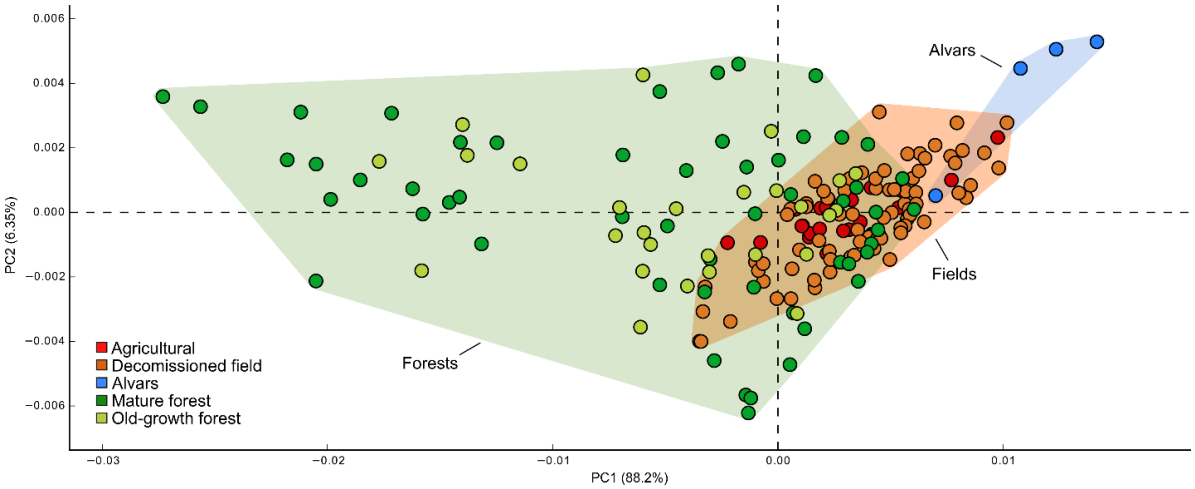

Figure 25: PCA analysis of predicted metagenomes for forests and field sites. PC1 and PC2 represent the principal components explaining the variation in the dataset. Results are based off relative abundances of predicted genes at level three in the KEGG hierarchy using the entire sample as the parent class (i.e., abundances of genes at level three relative to the sum of the total genes predicted in the KEGG hierarchy).

3.6.2 Assessing differences in predicted genes with depth and across land-use type

To assess differences in specific predicted genes with depth and across land-use type, a heat map was constructed for the 15 cm increment dataset (Figure 26). After gene filtering, the most abundant KEGG pathways were “ABC transporters”, “transporters”, and “oxidative phosphorylation” (Figure 26). The majority of the gene predictions appeared to have largely similar abundances across all samples with the exception of genes associated with the pathways “ABC transporters” and “transporters” (from the environmental information processing group) (Figure 26). These groups appeared to correspond with the overall change in land-use type and pH; genes were generally lower in the forest sites (CA, IW, H) than the field sites (D03, D07, D10, AA, Al) as well as in the low pH samples (Figure 26). Notably, the magnitude of changes across land-use type appeared to be more evident for genes associated with “transporters” (Figure 26). Post-hoc analysis found significant albeit small differences in “transporter” genes between agricultural sites and both mature and old-growth forest sites, as well as decommissioned fields and both mature and old-growth forests (Figure 27). As seen in Figure 24, although there was some evidence of depth changes across samples (particularly the field sites), there were no significant differences in specific KEGG pathways as highlighted in Figure 26.

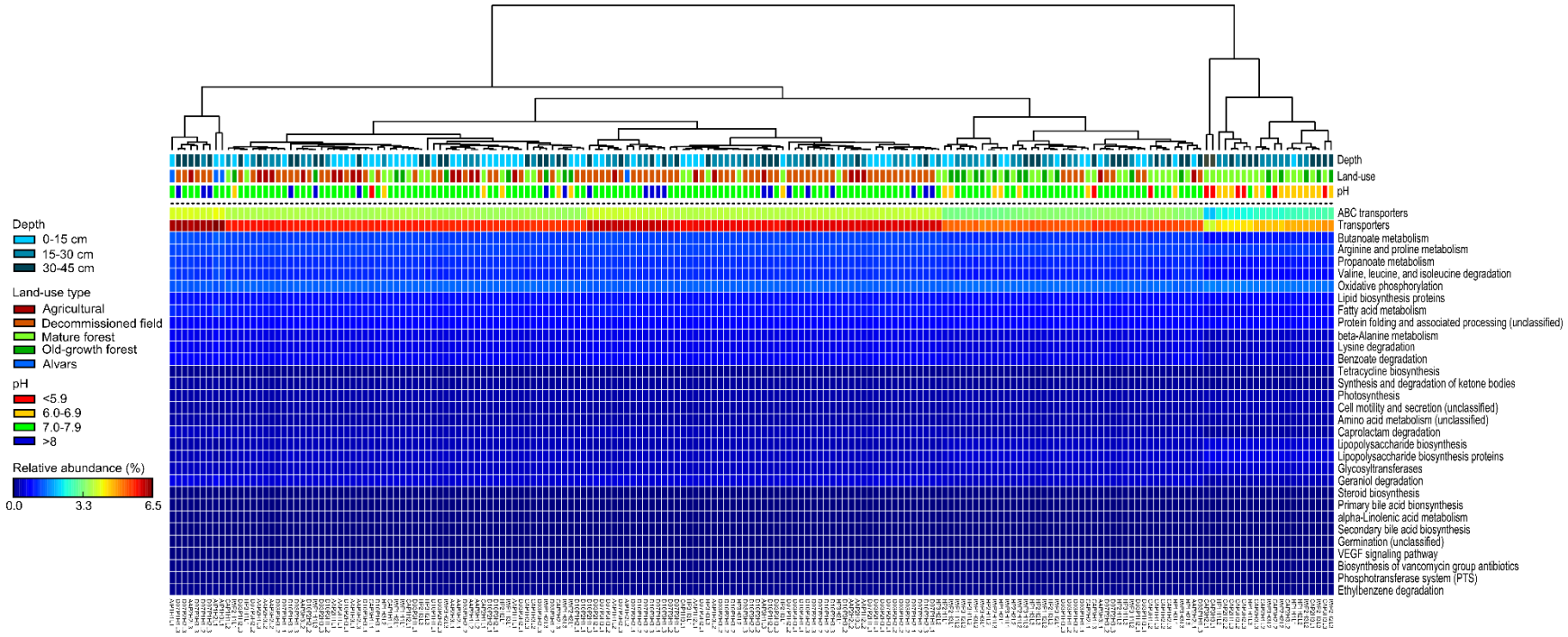

Figure 26: Heat map showing filtered ($q < 0.05$; Benjamini-Hochberg corrected; effect size > 0.35) predicted metagenomes per sample. Dendrogram located on top of theheat map shows land-use relationships (UPGMA, average neighbour clustering). Relative abundance data are reported based off of gene prediction at level three in the KEGG hierarchy using the entire sample as the parent class (i.e., abundances of genes at level three relative to sum of the total genes). Labels located at the bottom of the heat map represent sample IDs (see Table 15, Appendix A).

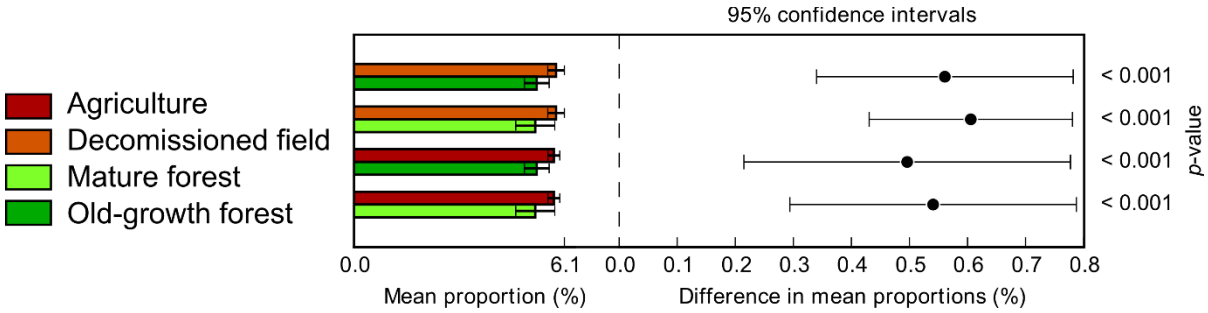

Figure 27: Post-hoc analysis (Kruskal-Wallis H-test, Tukey-Kramer) of predicted genes associated with the KEGG category "transporter" based on land-use type (*p* values are Benjamini-Hochberg corrected).

Chapter 4
Discussion

Soil microorganisms are recognized drivers of a fundamental suite of processes critical for sustaining global biogeochemical cycles (26, 73). Considering their ubiquity and relatively uncharacterized biogeography in deeper soil layers (i.e., >15 cm (25)), there is potential for missing key microorganisms and microbial community dynamics responsible for carrying-out important biogeochemical processes on Earth. This study has broadly evaluated depth-specific effects on bacterial community composition across a range of land-use types across the ***rare*** Charitable Research Reserve. Where previous work used lower resolution techniques (e.g., PLFA, DGGE, RFLP) (2, 88, 137, 148) or focussed on a few depth profiles from similar environment types (25, 138, 146, 150), this study leveraged high-throughput sequencing of 16S rRNA genes obtained from an extensive dataset (376 samples) across distinct land-use types to assess how bacterial communities are affected by depth.

4.1 16S rRNA gene clustering analysis

The results from the 16S rRNA gene clustering comparison analysis were consistent with initial hypotheses; the UPARSE default (e.g., UPARSE+2) clustering approach had the fewest OTUs in comparison to CD-HIT, UCLUST, or UPARSE+1 (Table 10). These results demonstrated that the inclusion of singletons was the main source of difference in OTU counts. In addition, these results were largely consistent with previous studies with OTU counts generated via UCLUST diverging most strongly from UPARSE-clustered OTUs (218). It is important to note that, given the quality of the underlying reads, a successfully assembled sequence is likely correct because PANDAseq uses the *Q*-scores for overall assembled-read quality scoring (172). For all sequencing events in this study, the percentage of bases having *Q*-scores ≥30, 20, and 10 (that is, 99.9, 99, and 90% base calling accuracy, respectively) was >80, >90, or =100%, respectively (data not shown) suggesting accuracy in read assembly. As a result, even a large majority of singletons, which are likely to contain errors (179), can be assembled and included in downstream analyses despite assembly quality threshold stringency. This represents an important consideration for downstream data interpretation when singleton sequences are included in datasets (Table 10). Furthermore, it has been reported that a 0.9 quality threshold, the threshold used in this study on the datasets for downstream analyses, represents an empirical upper limit for paired-end read assembly via PANDAseq and, in fact, increasing this threshold would require unnecessarily higher read quality for sequence reconstruction (172).

Considering that CD-HIT and UCLUST used a *de novo* chimera checking step (UCHIME) different than the UPARSE subsets (UPARSE–OTU), the consistent chimeric sequence counts were expected (210). The marked difference in chimeric sequence counts between the UPARSE+1 datasets and the CD-HIT and UCLUST datasets are also consistent with findings from other studies where UPARSE's inherent chimeric sequence detection was shown to classify more sequences as chimeric than the UCHIME-based method (218). In addition, the observed asymptotic (global maximum) trends of the rarefaction curves corresponding to the UPARSE+2 dataset, suggests that UPARSE+2 surveys the majority of phylotypes present at a lower sequencing depth (Figure 9). These findings are consistent with previous reports (217, 219) although a comprehensive attempt was not made to assess the effects of different clustering strategies on overall alpha- and beta-diversity.

Taken together, these results provide important insight for choosing an appropriate sequence clustering algorithm. Recent studies have shown that the heuristic clustering strategies generate conserved trends in alpha- and beta-diversity estimates (217). As a result, choosing any one of the clustering strategies analyzed in this investigation will likely provide reasonable and largely consistent results (217). However, as mentioned previously, the main factor governing OTU counts in this comparison appeared to be the inclusion of singleton sequences (Table 10). Considering the error-prone nature of these singletons, including them in datasets will undoubtedly increase the amount of spurious OTUs, which may skew data interpretation (179). Based on these results, the UPARSE+2 clustering algorithm appeared to be the most effective (217) considering its default removal of singleton sequences as well as time and memory efficiencies.

4.2 Bacterial community dynamics across the *rare* Charitable Research Reserve

4.2.1 Depth influences

Depth represented a strong environmental gradient influencing soil bacterial communities, consistent with the initial hypothesis and results from previous studies (Figure 13–15, 17) (4, 25, 88, 141, 146, 148). In addition to strongly influencing bacterial richness (i.e., observed decreases with depth; Figure 10), the influence of depth was particularly highlighted in Figure 18 and 19, where many soil physicochemical parameters changed with depth (i.e., likely confounding variables). The magnitude of depth specific responses was stronger in the field sites than in the forest sites, even though shifts in C, NH_4^+, NO_3^-, H_2O, and texture showed generally consistent trends with depth across all pits (Figure 6–8, 15, 17). Considering that the major difference in community composition between field and forest sites was largely a consequence of pH (Figure 16 and 17), it was expected that pH was also the main factor causing the observed differences in depth-based responses in

bacterial communities between environment types. Indeed, further comparison of ordinations in Figure 15A and B with pH highlighted that although individual pits from forested sites showed community separation by depth (largely pit-specific; data not shown), the pH gradient across soils from forested sites (4.5–8) was much stronger in shaping bacterial communities (Figure 17). pH is consistently reported as a predominant factor shaping microbial communities (53, 74, 85, 220). In soils, pH has been found to affect microorganisms across large spatial gradients (i.e., continental scale) and various land-use types (74), in addition to agricultural plots as observed using methods in common with this study (53). Consistent with these studies, the effect of pH was shown to be a principal factor shaping bacterial species richness (Figure 11 and 12) and community composition (Figure 16), even exceeding the effect of depth, as evidenced by the notable differences between environment types (i.e., the forest and field sites) (Figure 17) (74, 81). Interestingly, many studies focussing on depth specific responses have highlighted that SOC quantity and quality are the major factors influencing subsurface community composition in soil profiles. Fierer *et al*. (88) suggested that, considering that soil pH, temperature, O_2, and texture do not generally shift substantially in soil profiles, they likely do not play a major role in defining community structure in soil profiles. The magnitude of pH changes in soil pits are consistent with previous studies (~1.5 pH units in some pits; Figure 6) and pH was found to largely govern differences in bacterial communities throughout depth profiles between sampling locations. This further demonstrates the importance of pH as a predicator of bacterial community composition, not only across surface soils (74), but also throughout the soil subsurface.

As previously highlighted, other factors were also shown to change with depth (Figure 18). SIC explained a relatively high proportion of the variation in the datasets (Figure 17), particularly in the forested environments, and showed specific changes with depth (Figure 7). In general, SIC was higher in the lower depth samples likely as a result of closer proximity to carbonate bedrock material at some sites (Figure 7) (9, 221). Although few studies have addressed the effect of SIC on microbial community composition (88), it likely plays important roles in soil buffering, constraining pH to minimize differences in soil microbial communities (9, 221).

Furthermore, SOC, NH_4^+, and NO_3^- also affected communities, albeit to a lesser extent than pH and SIC (Figure 17). In general, these components appeared to decrease with soil depth in sampled pits (Figure 7 and 8), an observation corroborated in a variety of previous works (89, 222, 223). Together, SOC, NH_4^+, and NO_3^-, are principle resources required for cell growth and consequently represent important “filters” (i.e., gradients causing shifts in taxa abundances due to changes in the environment and resource availability) for microbial diversity and community composition (91). More importantly, changes in the quality of SOC (labile, readily accessible forms)

(224) have been widely reported to represent a stronger constraint on community structure than quantity (25, 148). Although pH was observed to more strongly affect bacterial communities, changes in SOC still represents an underlying and consistent gradient affecting bacterial communities in soils. In this study, SOC quality was not directly assessed. However, as highlighted by Fierer *et al*. (2003), it is likely that shifts in communities between top soils (i.e., ~0–15 cm) and subsoils (i.e., >15 cm) (Figure 13–15) are, in part, a result of differences in labile SOC (e.g., from root exudates and litter inputs) leading to physiologically altered (adapted) organisms capable of persistence in the different SOC conditions (88).

Soil bodies are characterized by distinctive layers (i.e., horizons; see section 1.1.1), which reflect different physicochemical environments and are likely to affect microbial communities (9, 150). Because depth increments were not dictated by soil horizon in the investigation, the potential homogenization of inter-horizon soils may have potentially disrupted distinct communities leading to more overlap in bacterial communities between depth increments. As a result, the gradual shift in bacterial community composition with depth (i.e., the lower *A* values; Figure 13) may be a result of the sampling scheme used in this investigation. Although distinct groupings of bacterial communities by soil horizons have been reported in other studies (138, 142, 150), gradual shifts in bacterial community composition have also been observed (25). Regardless, the effect of depth, even when considering potential horizon homogenization, was still apparent across all sites at both coarse- and fine-scale depth increments (Figure 13).

It is also important to note that, due to the number of samples taken in this investigation, physicochemical analysis was conducted on composite soil samples. That is, soils from each 15 cm depth increment at each subplot were pooled. As a result, even though DNA was extracted from every soil sample, the corresponding physicochemical data is reflective of the overall edaphic features of the specific depth interval within each subplot and therefore may reduce the resolution of trends in the dataset. This effect would probably be particularly apparent in the 5 cm increment dataset considering that the corresponding physicochemical measurements are representative of a larger depth gradient. Although this represents a caveat to data interpretation, the observed trends in the 5 cm increment dataset were largely consistent with those observed throughout the 15 cm increment dataset (Figure 13) and those trends showed consistencies with other studies (see previous paragraphs).

Furthermore, it is important to note that sequences from *Bacteria* were the focus of this study. As a result, there is an inherent limitation of these observations in that they are reflective of only a subset of the microbial community. There are many interactions that occur between different

microbial groups (i.e., archaea, fungi, and viruses) that are likely to govern the overall biogeography of soil microorganisms across the ***rare*** Charitable Research Reserve.

4.2.2 Land-use influences

Visual assessment of ordinations indicated that bacterial communities were also strongly influenced by land-use type (Figure 13B and D, and 14, 15, 17). Differences between bacterial communities in forest and field environments appeared to mimic differences in the plant communities between sites (Table 7). In addition, NPMANOVA highlighted that the dominant plant species and plant richness explained a high proportion of the variation across all samples (i.e., "Full dataset"; Figure 17). Several studies have shown that aboveground influences have major roles in shaping microbial community composition (63, 100, 122). In particular, plant diversity has been suggested to predict the diversity and assemblages of soil microorganisms due to increases in food sources (i.e., root exudates, litter), microhabitat stabilization, and physicochemical conditions of the soil environment (100). The latter is partially a result of compounds released by plants into soils, which can include sugars, amino acids, organic acids, polysaccharides, and enzymes (225). Because plant inputs (i.e., root exudates, litter) are translocated throughout soil profiles over time and affect the physicochemical environment (225), it is likely that plants are important in shaping bacterial communities throughout soil profiles. The relative differences in community composition between top and subsurface soils (i.e., reduced groupings of sites; Figure 14) may partly reflect plant community influences; subsurface soils are presumably less influenced by plant-generated surface effects (i.e., pH changes or root exudates and litter input). This is consistent with previous suggestions by Fierer *et al*. (88) that changes in microbial communities with depth are associated with changes in SOC quality (see section 4.2.1); variations in physicochemistry throughout soil profiles, in part created by the plant communities, presumably add "selective gradients" (e.g., changes in pH, C, N) leading to changes in community members (25, 88). A more robust analysis linking plant-generated physicochemical effects to subsurface bacterial community composition would further help elucidate the effects of plants on soil microbial community composition.

As previously highlighted (see section 2.1), agricultural sites across the ***rare*** Charitable Research Reserve have been subjected to fertilizer use. The Preston Flats agricultural field studied in this investigation received urea as a source of N which represents an important potential constraint on communities across land-use types (226). N fertilization has important effects on soil biogeochemical cycles and plant community composition (226). In a study comparing experimental plots, N fertilization was found to strongly affect bacterial communities even overshadowing pH effects (226). In this study, although N (NH_4^+, and NO_3^-) represented an important edaphic factor shaping

communities (Figure 8 and 17), there was no prominent influence of N fertilization effects considering that the active field was overall similar in bacterial community composition to field sites where no known recent fertilizers were applied. In this study, N-fertilization effects across land-use type were not directly assessed but represent a key consideration as well as an area for future investigations.

4.2.3 Successional responses

Considering that evidence of secondary succession was observed in the above-ground plant community, and that soils as well as agricultural practices across the ***rare*** Charitable Research Reserve have been largely consistent over time (i.e., soil formed from similar parent material, consistent farming practices over the past ~165 years (180, 227)), I hypothesized that successional responses in field sites would be reflected in subsurface bacterial communities. Sites D07 and D10, the most recently decommissioned fields sampled, maintained plant species characteristic of early succession (perennial grasses and forbs). Site D03, the oldest decommissioned field site, had characteristic mid-stage successional plants (i.e., small woody plant, shrub species, as well as forbs and grasses; Table 7) (228). Contrary to the initial hypothesis, there appeared to be no direct indication of bacterial community succession across the field sites sampled in this investigation, as evidenced by lack of site groupings by year since decommissioning; the disturbed site (e.g., active agricultural field) should presumably be distinct from decommissioned sites, which should also be distinct from each other (Figure 15 and 19B). Instead, edaphic factors including soil texture, particularly sand, silt, and clay fractions, as well as moisture appeared to strongly govern changes in overall community composition across the field sites (Figure 17 and 19B). Although these results do not eliminate the possibility of the absence of ongoing or past successional responses of subsurface communities, which may be (or had been) operating on different time scales (i.e., years to decades or days to weeks (229)), it does further demonstrate the importance of the physicochemical environment in structuring bacterial communities.

4.3 Taxa dynamics

4.3.1 Phylum-level trends

At coarse taxonomic resolution, obvious shifts in the dominant phyla with depth likely influenced the overall shifts in community composition with the depth gradient (Figure 13 and 20). Members from the *Proteobacteria* and *Bacteroidetes* phyla were observed to decrease with increasing soil depth (Figure 20), which is an observation noted in previous work (4, 25, 150). The decline in the

relative abundance of these two phyla has been hypothesized to be a result of the general copiotrophic nature of many members within these bacterial groups (4, 25, 150, 230). As a result, closer proximity to labile sources of SOC typical of surface soils (e.g., closer proximity to root systems) may, in part, explain the consistencies in the changes of these bacterial groups observed across depth-related soil studies (230).

Evidence for distinct overall shifts in the relative abundance of *Acidobacteria* and *Actinobacteria* members were not observed despite previous reports for defined distributions throughout soil profiles (Figure 20) (4, 138, 150). However, Eilers *et al.* (25) noted trends similar to those observed in this study and highlighted that members from these phyla are likely strongly dependent on site specific soil profile characteristics. Indeed, acidobacterial subgroups have been shown to demonstrate differential predominance across soil horizons (4). In contrast to observations from Eilers *et al.* (25) and Hansel *et al.* (4), the *Verrucomicrobia* phylum did not exhibit a defined mid-profile peak in relative abundance but instead appeared to be consistently distributed throughout soil profiles (Figure 20). Considering that many members of the *Verrucomicrobia* phylum are free-living and oligotrophic (231), their distributions throughout soil profiles may indicate an adaptation of many members to the low-nutrient environments characteristic of bulk soils (25).

This study also noted increased abundance of members from the *Chloroflexi*, *Gemmatimonadetes*, *Nitrospirae*, and *Latescibacteria* (formerly WS3) phyla with depth (Figure 20). Although reasoning for these patterns remains largely unexplored, previous research has highlighted interesting features about these bacterial groups. For example, the *Chloroflexi* phylum has been shown to increase in abundance throughout soil profiles correlating strongly (and negatively) with total biomass (150). Considering that total biomass has been shown to parallel changes in C quality, this may highlight the ability of a large fraction of *Chloroflexi* members to utilize more recalcitrant C compounds (25). Previous research has highlighted that *Gemmatimonadetes* members are strongly adapted to low soil moisture conditions (232). In this study, moisture was consistently observed to decrease with increasing soil depth (Table 14, Appendix A) across the majority of sampled pits, suggesting a possible factor governing subsurface distributions of *Gemmatimonadetes* members (232). Furthermore, members from the *Nitrospirae* phylum may prefer non-rhizosphere environments (233). Will *et al.* (150) hypothesized that heterotrophic microorganisms that are associated with roots suppress the growth of autotrophic *Nitrospirae*. Considering the general decrease in SOC and N observed in their study, the increased abundance of members form the *Nitrospirae* phylum is likely due to a selective advantage that dark-adapted chemolithoautotrophic members have in subsoil systems (150). The conditions noted by Will *et al.* (150) are similar to those observed in this study providing validity to extending this hypothesis to *Nitrospirae* members observed in this work.

4.3.2 Depth and pH specific taxa correlations

In this study, an important observation was that all OTUs that were positively correlated with depth (i.e., increased in abundance with increasing soil depth) were all uncultured organisms and belonged candidate bacterial groups (Table 12). This finding further highlights the importance of studying subsurface soil environments considering the predominance of uncharacterized organisms that appear to be more abundant in deeper soils (Table 12). In contrast, the larger amount of negatively correlated OTUs with depth is consistent with the previous discussion for factors influencing microorganisms throughout soil profiles (see section 4.2.1) where shifts in the physicochemical environment particularly, SOC, NH_4^+, and NO_3^- strongly governs microbial diversity in soils (91). For example, all known members from the *Pseudonocardiaceae* are aerobic, which likely explains their greater abundance in surface soils where O_2 levels in pore spaces are presumably higher (Table 12) (234). In addition, members from the family *Bradyrhizobiaceae*, which were found to be negatively correlated with depth (Table 12), contain symbiotic N-fixing organisms, which are often predominant in top soils where root densities and presumably root exudates are higher.

Moreover, evidence from Figure 21, 23, and Table 13 further highlights the strong effect of pH in shaping taxa distributions in soils, consistent with previous work (235, 236). More positively correlated taxa with pH were observed, consistent with the general increase in richness at neutral to near-neutral pH ranges as reported in this study (Figure 11 and 12) and in previous work (74, 237). Additionally, *Nitrospirae* and *Chloroflexi* members were among the top strongly and positively correlated taxa with pH (Table 13). The association of members from these phyla in more neutral soils has also been observed by Bartram *et al*. (53) and is consistent with their known ecology; both *Nitrospiraceae* and *Anaerolineae* members are all known to have pH optima at neutral to near-neutral values (238, 239).

Furthermore, consistent with their known ecology, many members from *Acidobacteria* phylum, particularly the *Acidomicrobiia* and *Solibacteres* class (subdivision 1 and 3, respectively), were found to be strongly and negatively correlated with pH (Table 13) (240, 241). This observation has also been noted in previous studies (53, 236). In particular, OTUs that were strongly negatively correlated with pH were from the acidobacterial groups *Koribacteraceae* and *Solibacterales.* These results were also consistent with previously reported pH ranges of these organisms (241). Interestingly, correlation analysis found a strong negative relationship with pH and members from the *Gemmatimonadetes* class Ellin5290 (Table 13). As previously discussed, members from the *Gemmatimonadetes* phylum have been reported to be strongly adapted to low moisture conditions

(232). Results from Table 13 as well as the notable moisture effect on communities observed in Figure 17 may reflect the strong effect of pH in addition to moisture on *Gemmatimonadete* members.

The strong correlation of many taxa to changes in pH further highlights the importance of pH in shaping soil bacterial communities (82). Many taxa are adapted to narrow and specific growth and pH ranges (82, 242). For example, in acidic environments, specific mechanisms are required to maintain homeostasis and can include restricting H^+ entry into the cell, removing excess H^+ from the cytoplasm, and employing proteins (e.g., chaperones) to repair damages to cellular components (i.e., nucleic acids, proteins) from extreme pH shifts (243). A previous report noted a consistent 50% reduction in activity of *in situ* bacterial communities with a 1.5 unit change in pH (244). Consequently, Rousk *et al*. (82) highlighted that out-competition would rapidly occur if growth (activity) decreases by 25% (compared to optimal growth) by bacteria that are less hindered by the changing conditions. These narrow pH ranges resulting from specific adaptations to accommodate pH changes, helps support why pH is a strong factor structuring soil bacterial communities (82).

4.3.3 Indicator species

Contrary to initial hypotheses, robust indicators were not particularly evident for depth, land-use, or pH (Figure 22 and 23). Because by definition an indicator should be specific and exclusive to a particular group, these findings are consistent with the nature of the soil environment; the extensive heterogeneity of soils and the widespread distributions of taxa throughout the subsurface environment reduces indicator strength (i.e., indicator value). Nonetheless, indicator species were observed for specific depth increments at specific sites (Figure 22). The observation that members from *Syntrophobacteraceae* family were putatively indicative of deeper horizons is largely consistent with their known ecology considering that members are strictly anaerobic (Figure 22) (245). However, results suggesting that they are particularly indicative of sites D10 and Al remains unclear as there were no observable similarities between the sites (Figure 22). Furthermore, results suggesting an unclassified *Betaproteobacteria* member and a member from the candidate class S035 (*Acidobacteria* phylum) as indicators for surface soils from site AA also remains largely uncertain considering the lack of information regarding these organisms.

Several bacterial indicators were associated primarily with neutral soil samples (Figure 23). Members from the *Firmicutes* (*Shimazuella*, *Paenibacillus*), *Actinobacteria* (*Solirubrobacterales*, MB-A2-108), *Cyanobacteria*, and the WS3 (also known as *Latescibacteria*) phyla were found to be generally indicative of alkaline environments, largely consistent with their known ecology (246) (Figure 23). The prevalence of members from the genera *Shimazuella* (247) and *Paenibacillus* (248) in neutral to near-neutral environments are supported by studies suggesting their preference for

neutral and alkaline environments. In contrast, many members from the *Solirubrobacterales* group, a member from the *Actinobacteria* phylum and an indicator for alkaline environments reported in this study, have been shown to be found in acidic environments (249). This observation likely highlights the widespread known ecological breadth of actinobacterial members (38, 250). Other neutral pH soil indicators included *Cyanobacteria* and WS3 members, consistent with their known preference for higher pH soil environments (220, 251). When indicator results were compared to correlation analysis, there was some evidence of consistencies (i.e., prevalence generally at similar pHs) for some taxa particularly for pH (Figure 22 and 23, Table 12 and 13). Whereas consistencies (i.e., high indicator value and correlation to a specific category) between these results would further highlight and reinforce the influence of depth and pH on specific taxa, the lack thereof can be explained by the differences in the approaches. Indicator species analysis takes into account specificity and fidelity of a taxon to a specific group whereas correlation analysis measures dependence between two parameters (208, 252). Although many taxa have dependence (i.e., are correlated) with depth or pH, their distribution may not be specific enough to be considered an indicator. This may ultimately be a result of heterogeneity of the soil environment where microorganisms are distributed widely throughout soil profiles.

4.4 PICRUSt metagenome predictions

The PICRUSt predictions reflected observations in the taxonomy-based analyses, even though the majority of the predictions appeared to be largely similar (i.e., PCoA ordinations in Figure 13, 15, 24, and 25). This finding represents an important caveat to the predictions generated by PICRUSt; predictions are limited by the number of metagenomes in the reference database (161). This is due to the fact that the PICRUSt algorithm is highly dependent on tree topology and OTU distance to a reference genome (161, 253). Therefore, OTUs will always have a nearest neighbour, even if the phylogenetic distances are large, thereby linking all OTUs in the dataset (253). As a result, PICRUSt predictions are undoubtedly under-representative of the full extent of the functional diversity that is present in the environment particularly in poorly studied systems (161). Additionally, it is important to note that the PICRUSt results in this study only capture the functional diversity for bacteria (161). As a result, PICRUSt provides only a partial assessment of the overall metagenome considering the predominance of other microorganisms in soils such archaea, fungi, and viruses, which also contribute to the overall functional diversity (161).

Furthermore, PICRUSt is fundamentally limited by the biases associated with the primers used to amplify portions of the 16S rRNA gene (161). The primers in this study amplify a region of the 16S rRNA gene reported to minimize bias (254). Other primers however have been shown to

incompletely capture members of certain groups of organisms. For example, primers 27–F and 338–R targeting the V1-V2 region of the 16S rRNA gene have been shown to underestimate members from the *Verrucomicrobia* phylum (231). As a result, application of PICRUSt to a dataset generated with such primers would not accurately represent genes contributed by those organisms (161). With that being said, the NSTI scores observed for both datasets were within similar and expected ranges reported by Langille *et al*. (161) for soils. Although further assessment of whether the PICRUSt results are accurate would require empirical analysis through actual metagenomic sequencing of samples, Langille *et al*. (161) observed good correlation of PICRUSt's predictions and "full" metagenomic sequencing for soil datasets at similar NSTI scores (NSTI = 0.17; standard deviation = ± 0.02; Spearman's $\rho = 0.81$). Consequently, despite caution in the interpretation of PICRUSt-generated results, the approach still remains suitable as a preliminary exploratory method allowing both corroboration and potential for reformulation of hypotheses.

PICRUSt predicted genes associated with a broad range of functional categories. The majority of the gene predictions were associated with the group "metabolism", which is consistent with immense breadth of functional diversity present in soils considering the range of metabolic functions associated with this category (24). Interestingly, a relatively high proportion of the PICRUSt predicted genes were unclassified, highlighting the further need for characterization of the soil environment (255). This represents a particularly important area for future investigations.

The effect of depth was not particularly observed in the PICRUSt analyses (Figure 24 and 26). However, land-use type appeared to play a role particularly with respect to changes in genes associated with "transporters" (Figure 26). The apparent difference in transporter gene abundances between field and forest sites may, in part, be a result of the physicochemical environment, particularly pH (Figure 26 and 27). As previously discussed (section 4.3.2), because organisms that are adapted to low pH require specific mechanisms such unique transporters (i.e. H^+ ATPases, antiporters and symporters) to maintain homeostasis (243), the lower abundance of transporter associated genes in the forest sites may reflect the importance of specific transporter groups essential for life in acidic environments (Figure 11 and 26). This would also be observed in the species richness for lower pH sites (i.e., similar groups of strongly adapted acidophiles). Not surprisingly, site CA, which had the lowest observed pHs (Figure 6) as well as the lowest species richness among sampled sites, was also found to have the lowest amount of predicted genes associated with the "transporters" KEGG pathway. Further analysis (e.g., metagenome sequencing) would be needed to confirm this initial observation.

4.5 Microbiogeographical implications

Defining the influence of the contemporary and historical environment in shaping the distributions of microorganisms on Earth represents a challenging and active area of investigation in microbial ecology (118, 123). Indeed, studies assessing factors affecting the spatial-temporal distributions of microorganisms have found evidence supporting both the "everything is everywhere" hypothesis (205), as well as the hypothesis that site-history plays a major role in shaping soil microbial communities (see section 1.2.1) (122). In this work, evidence suggested the stronger influence of the contemporary physicochemical factors rather than historical legacies in shaping bacterial communities. This is highlighted by the importance of physicochemical parameters, particularly pH- and depth-specific factors, in governing overall bacterial community diversity and composition (Figure 13, 14, 16, and 17) (237, 256).

Perhaps a stronger line of evidence for the importance of the immediate physicochemical environment is the lack of successional responses observed after agricultural decommissioning (see section 4.2.3). Consistent with this hypothesis, Kuramae and colleagues noted that soil pH (256) as well as other soil physicochemical parameters, such as K and NH_4^+ (257), could strongly affect the successional trajectory of soil bacteria rather than the aboveground plant richness and diversity. Succession can be considered a process that imparts lasting historical characteristics on the environment resulting from land-use disturbances (258). For example, a microcosm experiment evaluating legacy effects of intensively managed lands found that regardless of the restoration attempt, the historical legacies of past land-usages had lasting effects on soil quality consequently affecting soil organisms (258, 259). Of course, there a variety of caveats to the biogeographic interpretation of the results in this study, including that succession was approximated using different sites across the ***rare*** Charitable Research Reserve (i.e., chronosequence analysis) (260). With that being said, the apparent effect of SIC (Figure 17) on community composition may highlight historical constraints. SIC increased within deeper soils, presumably a result of closer proximity to dolostone ($CaMg(CO_3)_2$) bedrock material formed millions of years ago.

To better assess the interplay between contemporary and historical influences, designing an experiment that assesses these components explicitly would be necessary (123). This includes defining specifically "contemporary" and "historical" influences as well as controlling for the effect of these factors on microbial community dynamics (123). Moreover, the importance of contemporary processes (i.e. physiochemical environment) and historical effects is likely dependent on the physiology of individual taxa, habitat type, geographic scale, as well as taxonomic resolution (123). As a result, taking into account the effect of these components is also necessary (123). Although this

study did not attempt to assess these components directly, it does add to the growing body of literature helping to further assess the patterns of the distributions of microorganisms on Earth.

Chapter 5
Conclusions & Future Directions

5.1 Contributions and perspectives

The goal of microbial ecology is to better understand how microorganisms interact with each other and their environment. The ability to conduct microbial ecological research has been impacted by technological advancements in 'omics-based approaches and techniques for characterization of nucleic acid biomarkers (e.g., metagenomics, transcriptomics, high-throughput sequencing) (1, 24, 261, 262). These advances have enabled cultivation-independent assessments of microbial community composition in a range environments that can be coupled with bioinformatic and statistical analyses that facilitate functional and phylogenetic inferences in relation to microbial physiology and ecology. As a result, fundamental questions and hypotheses that were once beyond the reach of microbial ecologists are at the forefront of modern investigations in microbiology. Some of the current important themes and key research areas in microbial ecology include: comparative community metagenomics (263), rare biosphere and microbial "dark matter" exploration (255, 264), and applications of macroecological theory to microorganisms (126, 152, 265). Ultimately, all of these emerging areas strive to address the fundamental goal of microbial ecology by addressing questions such as: Who is there? What do they do? How do they interact with each and their surroundings? And, what processes shape their distribution and diversity in the environment? The latter of these themes includes the field of microbial biogeography, which has gained renewed interest since its conception (section 1.2.1) considering the growing number of studies providing direct evidence for defined microbial distributions in the environment (118, 152).

My thesis research helped address the overall goal of microbial ecology by contributing insight into mechanisms shaping bacterial biogeography. This study focused on the relatively "uncharted territory" of the subsurface soil environment using an extensive dataset of 16S rRNA gene sequences. Specifically, the research presented here examined bacterial community composition throughout soil profiles across a range of land-use types. In addition to assessing depth-specific biogeographical patterns, this research has set up baseline observations of bacterial community dynamics at the ***rare*** Charitable Research Reserve, paving the way for future investigations. Considering that the ***rare*** Charitable Research Reserve is an uncontaminated region hosting a range of land usages with consistent underlying geology, there is enormous potential for utilizing this land for hypothesis-driven future investigations in microbial ecology. These investigations could include assessing the impacts of warming and elevated CO_2, simulating potential impacts of climate change,

or may further investigate phylogenetic novelty for potential applied benefits for humans. Importantly, this research also represents first steps in the direction of pioneering a Canadian "microbial observatory", similar to the Long Term Ecological Research (LTER) network established in the United States (http://www.lternet.edu/), with the capability of supporting long-term extensive research in microbial ecology.

In addition to baseline characterization of the ***rare*** Charitable Research Reserve, the central goal of this study was to investigate depth-specific biogeographical patterns in soils. In this respect, my thesis has contributed insight into patterns and factors affecting subsurface microbial communities, expanding on the few studies that included depth as an environmental gradient. In particular, this work confirmed that depth is a major determinant of bacterial community composition in soils (Figure 13, 14, 18, and 19), consistent with observations of previous investigations of terrestrial microbiology, including Eilers *et al.* (25), Fierer *et al.* (88), and Hansel *et al.* (4). In addition, this study highlighted changes in the abundance of specific bacterial groups in relation to depth, reflecting ecological niches that are inhabited by specific bacteria throughout the subsurface environment (Figure 20) (25). Moreover, a notable discovery in this study was the increased abundance of uncharacterized and uncultured organisms with increasing soil depth (Table 12), which is a result that emphasizes the need for future investigations in assessing the uncharacterized phylogenetic novelty that exists within subsurface soil environments. The results in this study have also demonstrated that pH is among the top edaphic factors governing microbial community dynamics in soils at the ***rare*** Charitable Research Reserve, which is consistent with previous work, most notably by Lauber *et al.* (74), showing that pH affected microbial communities at continental scales. Along with previous investigations of pH-controlled agricultural plots (53, 82) and continental-scale biogeography work, this study demonstrates that major global patterns of bacterial biogeography, such as those that relate to pH, operate in a consistent and predictable manner even across local scales with consistent underlying geology.

Results of this study also highlight the influence of the contemporary physicochemical environment, rather than historical legacies, in shaping bacterial communities. This was particularly evident when considering the influence of physicochemical parameters, particularly pH, in shaping soil bacterial communities (Figure 13, 14, 16, and 17) (237, 256). Nonetheless, historical effects also influenced bacterial communities in this dataset. Specifically, the effect of SIC (Figure 17), which increased within deeper soils presumably as a result of closer proximity to dolostone ($CaMg(CO_3)_2$) bedrock, highlights potential influences of geology acting over millions of years. Although this study did not attempt to assess these components directly, my research adds to the growing body of literature helping to further assess biogeographical distributions of microorganisms on Earth.

5.2 Future directions

An extension of the work presented in this thesis would be to assess how other groups of microorganisms, such as archaea and fungi, are structured by depth and land-use type. These additional analyses would allow for a more complete assessment of the overall microbial community dynamics at the ***rare*** Charitable Research Reserve. The Neufeld Lab anticipates follow-up investigations looking into biogeographical patterns of archaea across the ***rare*** Charitable Research Reserve. Additionally, investigation of the rare biosphere, particularly in deeper soils, will be an important area for future investigation. Assessing the distribution of rare taxa, as well as characterizing genes and unknown phylogenetic groups, will further help to explore the diversity and function of microorganisms in the environment.

Another important area for further research, building on initial analysis of the ***rare*** Charitable Research Reserve, would involve a more robust and strategic analysis linking plant-generated physicochemical effects to subsurface bacterial community composition. Such a study would further elucidate the effects of plants on soil microbial community composition. This could include long-term (i.e., over years) monitoring of changes in soil characteristics generated from altering dominant plant members and tracking (e.g., monitoring fate of substrates via gas chromatography or stable isotopes) those effects throughout the subsurface.

Given that microbial ecologists are beginning to better understand how traditional macroecological theories apply to microbial world, this represents an important area for future research considering that there is great need for synthesizing and developing holistic theories in microbial ecology. Such theories would allow generation of more predictive ways of describing microorganisms on Earth (266). As discussed by Prosser *et al*. (266), the field of microbial ecology becomes an "accumulation of situation-bound statements" with little overarching insight for microbiologists without this area of development. Additional areas for future investigations include more direct investigations of the interplay between the immediate physicochemical environment and past legacy effects in shaping microbial communities (118), as well as further assessment of the validity and application of the intermediate disturbance hypothesis (130, 267) to microbial community stability.

Caveats to using 16S rRNA genes have been previously discussed (see section 1.3.3.1) and represent key areas for future work to reinforce and elaborate on observed trends. The selection of accurate primer pairs for amplicon generation is arguably the most significant step in 16S rRNA gene studies and represents an important consideration (163). Although the primer pairs (341–F and 806–R) used in this investigation have been reported to reduce bias in taxonomic representation there are still some taxonomic groups that may be under- or over-represented (268). Further work in

developing more robust primer pairs for microbial groups (i.e., bacteria, archaea) capable of further reducing taxonomic bias is critical for more accurate representation of community composition.

Furthermore, using rRNA or discriminatory agents, such propidium monoazide to distinguish between dead and live cells (167), would strengthen this research by highlighting active community members. With that being said there are inherent limitations of these approaches as well, such as the non-uniform scaling of rRNA concentration and growth rate, or the assumption that membrane integrity is analogous to cellular activity (166, 167). However, studies have been able to detect distinct patterns and make inferences based on these techniques (269, 270). Such inferences include the recognition that many low abundance microbial taxa can be highly active members of communities in aquatic and terrestrial environment (271, 272).

Another important consideration for data in this study is the interpretation of trends based on relative abundances which provides no direct information on absolute taxa abundances (273). This can limit data interpretation considering that two samples that have a comparable relative abundance (based on 16S rRNA genes) of a certain bacterial group could actually have a widely different absolute abundance in each sample if the total bacterial cell counts differ substantially (273). Such may be particularly apparent in the PICRUSt analysis where metagenome predictions for each sample relies on multiplying gene family abundances by a taxon's relative abundance. Although qPCR can be used to assess 16S rRNA gene copy number, the increased time constraints as well as the biases that can be introduced via DNA extraction and amplification efficiencies can also limit the applicability of this approach in 16S rRNA HTP studies. Recent work by Smets *et al*. (273) suggests that the inclusion of an internal standard (i.e., DNA from a known organism not typically found in soils) during DNA extraction represents a simple step that can help ascertain absolute taxa abundances by enabling the calculation of 16S rRNA gene reads in each sample (273). Including this step in future analyses will help further explore the biogeography of soils across ***rare*** Charitable Research Reserve by allowing direct assessment of absolute abundance shifts across soils from different environments.

Furthermore, although I assessed predicted metagenomes based on the PICRUSt approach, the limitations of those predictions (section 4.4) warrants further empirical analysis via metagenomic sequencing. Metagenomic sequencing has the potential to reveal significant insight into the structure of microbial communities by revealing dominant genes responsible for important life strategies and biogeochemical functions in soils (274). Considering the abundance of predicted genes that were unclassified, metagenomic sequencing may also further reveal uncharacterized genes responsible for important and unique functions in soils. Additionally, a recently published functional community profiling algorithm, Tax4Fun (253), may represent an alternative cost-effective approach to

metagenomic sequencing. Similar to PICRUSt, Tax4Fun uses 16S rRNA gene datasets to predict metagenomes but relies on homology searches rather than a phylogenetic tree to make predictions (253, 275). Validation and performance efforts for the Tax4Fun algorithm highlighted a stronger correlation of whole genomes to 16S rRNA gene based metagenomic predictions than those observed via PICRUSt (253).

In conclusion, this study explored biogeographical patterns of soil bacteria across the ***rare*** Chartable Research Reserve, highlighting the strong effects of depth and soil physicochemistry. These results add to the growing body of research documenting how microorganisms are distributed on Earth, demonstrates the potential for increased phylogenetic novelty in relation to soil depth, and provides a subsurface microbial biogeography baseline for the ***rare*** Charitable Research Reserve that will enable future research efforts in soil microbial ecology.

Bibliography

1. Pepper IL, Gerba CP, Gentry TJ. 2015. Environmental Microbiology, 3rd ed. Academic Press, San Diego, California.

2. Agnelli A, Ascher J, Corti G, Ceccherini MT, Nannipieri P, Pietramellara G. 2004. Distribution of microbial communities in a forest soil profile investigated by microbial biomass, soil respiration and DGGE of total and extracellular DNA. Soil Biol Biochem **36**:859–868.

3. Simonson RW. 1959. Outline of a generalized theory of soil genesis. Soil Sci Soc Am J **23**:152–156.

4. Hansel CM, Fendorf S, Jardine PM, Francis CA. 2008. Changes in bacterial and archaeal community structure and functional diversity along a geochemically variable soil profile. Appl Environ Microbiol **74**:1620–1633.

5. Pedrós-Alió C. 2012. The rare bacterial biosphere. Ann Rev Mar Sci **4**:449–466.

6. Whitman WB, Coleman DC, Wiebe WJ. 1998. Prokaryotes: the unseen majority. Proc Natl Acad Sci **95**:6578–6583.

7. Crocker RL. 1952. Soil genesis and the pedogenetic factors. Q Rev Biol **27**:139–168.

8. Albiach R, Canet R, Pomares F, Ingelmo F. 2001. Organic matter components and aggregate stability after the application of different amendments to a horticultural soil. Bioresour Technol **76**:125–129.

9. Brady NC, Weil RR. 2010. Elements of the Nature and Properties of Soils, 3rd ed. Prentice Hall, Upper Saddle River, New Jersey.

10. Agriculture and Agri-Food Canada. 1998. The Canadian System of Soil Classification, 3rd ed. NRC Research Press, Ottawa, Ontario.

11. Carroll D. 1959. Ion exchange in clays and other materials. Geol Soc Am Bull **70**:749–780.

12. Stotzky G. 1966. Influence of clay minerals on microorganisms: effect of particle size, cation exchange capacity, and surface area on bacteria. Can J Microbiol **12**:1235–1246.

13. Mummey DL, Stahl PD. 2004. Analysis of soil whole- and inner-microaggregate bacterial communities. Microb Ecol **48**:41–50.

14. Cruz-Martínez K, Rosling A, Zhang Y, Song M, Andersen GL, Banfield JF. 2012. Effect of rainfall-induced soil geochemistry dynamics on grassland. Appl Environ Microbiol **78**:7587–7595.

15. Pimentel D, Stachow U, Takacs DA, Brubaker HW, Dumas AR, Meaney JJ, Onsi DE, Corzilius DB. 1992. Conserving biological diversity in agricultural/forestry systems. Bioscience **42**:354–362.

16. Sylvia DM, Fuhrmann J, Hartel PG, Zuberer DA. 2005. Principles and Applications of Soil Microbiology, 2nd ed. Prentice Hall, Upper Saddle River, New Jersey.

17. Frouz J, Elhottová D, Kuráž V, Šourková M. 2006. Effects of soil macrofauna on other soil biota and soil formation in reclaimed and unreclaimed post mining sites: results of a field microcosm experiment. Appl Soil Ecol **33**:308–320.

18. Powlson DS, Prookes PC, Christensen BT. 1987. Measurement of soil microbial biomass provides an early indication of changes in total soil organic matter due to straw incorporation. Soil Biol Biochem **19**:159–164.

19. Tisdall J, Oades J. 1982. Organic matter and water-stable aggregates in soils. J Soil Sci **33**:141–163.

20. Summers S, Whiteley AS, Kelly LC, Cockell CS. 2013. Land coverage influences the bacterial community composition in the critical zone of a sub-Arctic basaltic environment. FEMS Microbiol Ecol **86**:381–393.

21. Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. 2013. Going back to the roots: the microbial ecology of the rhizosphere. Nat Rev Microbiol **11**:789–99.

22. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. 2006. The role of root exudates in rhizosphere interactions with plants and other organisms. Annu Rev Plant Biol **57**:233–266.

23. Rillig MC, Mummey DL. 2006. Mycorrhizas and soil structure. New Phytol **171**:41–53.

24. Torsvik V, Øvreås L. 2002. Microbial diversity and function in soil: from genes to ecosystems. Curr Opin Microbiol **5**:240–245.

25. Eilers K, Debenport S, Anderson S, Fierer N. 2012. Digging deeper to find unique microbial communities: the strong effect of depth on the structure of bacterial and archaeal communities in soil. Soil Biol Biochem **50**:58–65.

26. Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science **320**:1034–1038.

27. Schimel JP, Schaeffer SM. 2012. Microbial control over carbon cycling in soil. Front Microbiol **3**:1–11.

28. Rumpel C, Kögel-Knabner I. 2010. Deep soil organic matter—a key but poorly understood component of terrestrial C cycle. Plant Soil **338**:143–158.

29. D'Orazio V, Traversa A, Senesi N. 2014. Forest soil organic carbon dynamics as affected by plant species and their corresponding litters: a fluorescence spectroscopy approach. Plant Soil **374**:473–484.

30. Cabello P, Roldán MD, Moreno-Vivián C. 2004. Nitrate reduction and the nitrogen cycle in archaea. Microbiology **150**:3527–3546.

31. Kertesz MA, Mirleau P. 2004. The role of soil microbes in plant sulphur nutrition. J Exp Bot **55**:1939–1945.

32. Kim K, Jordan D, McDonald G. 1997. Effect of phosphate-solubilizing bacteria and vesicular-arbuscular mycorrhizae on tomato growth and soil microbial activity. Biol Fertil Soils **26**:79–87.

33. Molina JAE, Rovira AD. 1964. The influence of plant roots on autotrophic nitrifying bacteria. Can J Microbiol **10**:249–257.

34. Tourna MJ, Stieglmeier M, Spang A, Könneke M, Schintlmeister A, Urich T. 2011. *Nitrososphaera viennensis*, an ammonia oxidizing archaeon from soil. Proc Natl Acad Sci U S A **108**:1–6.

35. Claus G, Kutzner HJ. 1985. Physiology and kinetics of autotrophic denitrification by *Thiobacillus denitrificans*. Appl Environ Microbiol **22**:283–288.

36. Southam G, Beveridge TJ. 1992. Enumeration of Thiobacilli within pH-neutral and acid mine tailings and their role in the development of secondary mineral soil. Appl Environ Microbiol **58**:1904–1912.

37. Lie TA, Akkermans ADL, van Egeraat AWSM. 1984. Natural variation in symbiotic nitrogen-fixing *Rhizobium* and *Frankia* spp. Antonie van Leeuwenhoek J Microbiol **50**:489–503.

38. Piao Z, Yang L, Zhao L, Yin S. 2008. Actinobacterial community structure in soils receiving long-term organic and inorganic amendments. Appl Environ Microbiol **74**:526–530.

39. Stevenson IL. 1956. Antibiotic activity of actinomycetes in soil and their controlling effects on root-rot of wheat. J Gen Microbiol **14**:440–448.

40. McSpadden Gardener BB. 2004. Ecology of *Bacillus* and *Paenibacillus* spp. in agricultural systems. Phytopathology **94**:1252–1258.

41. Konopka A, Oliver L, Turco RF. 1998. The use of carbon substrate utilization patterns in environmental and ecological microbiology. Microb Ecol **35**:103–115.

42. Weimer PJ, Zeikus JG. 1977. Fermentation of cellulose and cellobiose by *Clostridium thermocellum* in the absence and presence of *Methanobacterium thermoautotrophicum*. Appl Environ Microbiol **33**:289–297.

43. Matsen JB, Yang S, Stein LY, Beck D, Kalyuzhnaya MG. 2013. Global molecular analyses of methane metabolism in Methanotrophic alphaproteobacterium, *Methylosinus trichosporium* OB3b. Part I: Transcriptomic study. Front Microbiol **4**:1–16.

44. Shukla AK, Vishwakarma P, Upadhyay SN, Tripathi AK, Prasana HC, Dubey SK. 2009. Biodegradation of trichloroethylene (TCE) by methanotrophic community. Bioresour Technol **100**:2469–2474.

45. Ghiorse WC. 1984. Biology of iron- and manganese-depositing bacteria. Annu Rev Microbiol **38**:515–550.

46. Laanbroek HJ. 1990. Bacterial cycling of minerals that affect plant growth in waterlogged soils: a review. Aquat Bot **38**:109–125.

47. Gregory E, Staley JT. 1982. Widespread distribution of ablity to oxidize manganese among freshwater bacteria. Appl Environ Microbiol **44**:509–511.

48. Summers A, Silver S. 1978. Microbial transformations of metals. Annu Rev Microbiol **32**:637–672.

49. Kavamura VN, Esposito E. 2010. Biotechnological strategies applied to the decontamination of soils polluted with heavy metals. Biotechnol Adv **28**:61–69.

50. Ye J, Singh A, Ward OP. 2004. Biodegradation of nitroaromatics and other nitrogen-containing xenobiotics. World J Microbiol Biotechnol **20**:117–135.

51. Johnson CH, Patterson AD, Idle JR, Gonzalez FJ. 2012. Xenobiotic metabolomics: major impact on the metabolome. Annu Rev Pharmacol Toxicol **52**:37–56.

52. Nannipieri P, Ascher J, Ceccherini MT, Landi L, Pietramellara G, Renella G. 2003. Microbial diversity and soil functions. Eur J Soil Sci **54**:655–670.

53. Bartram AK, Jiang X, Lynch MDJ, Masella AP, Nicol GW, Dushoff J, Neufeld JD. 2014. Exploring links between pH and bacterial community composition in soils from the Craibstone Experimental Farm. FEMS Microbiol Ecol **87**:403–415.

54. Eichorst SA, Breznak JA, Schmidt TM. 2007. Isolation and characterization of soil bacteria that define *Terriglobus* gen. nov., in the phylum *Acidobacteria*. Appl Environ Microbiol **73**:2708–2717.

55. Zhou J, Xia B, Treves DS, Wu L, Marsh TL, Neill RVO, Palumbo AV, Tiedje JM. 2002. Spatial and resource factors influencing high microbial diversity in soil. Appl Environ Microbiol **68**:326–334.

56. Treves DS, Xia B, Zhou J, Tiedje JM. 2003. A two-species test of the hypothesis that spatial isolation influences microbial diversity in soil. Microb Ecol **45**:20–28.

57. Wildung RE, Garland TR, Buschbom RL. 1975. The interdependent effects of soil temperature and water content on soil respiration rate and plant root decomposition in arid grassland soils. Soil Biol Biochem **7**:373–378.

58. Zhang X, Zhao L, Xu SJ, Liu YZ, Liu HY, Cheng GD. 2012. Soil moisture effect on bacterial and fungal community in Beilu River (Tibetan Plateau) permafrost soils with different vegetation types. J Appl Microbiol **114**:1054–1065.

59. Lüdemann H, Arth I, Liesack W, Lu H. 2000. Spatial changes in the bacterial community structure along a vertical oxygen gradient in flooded paddy soil cores. Appl Environ Microbiol **66**:754–762.

60. Andrews JA, Matamala R, Westover KM, Schlesinger WH. 2000. Temperature effects on the diversity of soil heterotrophs and the delta ^{13}C of soil-respired CO_2. Soil Biol Biochem **32**:699–706.

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

62. Lozupone CA, Knight R. 2007. Global patterns in bacterial diversity. Proc Natl Acad Sci U S A **104**:11436–11440.

63. Kowalchuk GA, Buma DS, de Boer W, Klinkhamer PGL, van Veen JA. 2002. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. Antonie van Leeuwenhoek J Microbiol **81**:509–520.

64. Fontaine S, Mariotti A, Abbadie L. 2003. The priming effect of organic matter: A question of microbial competition? Soil Biol Biochem **35**:837–843.

65. Foster RC. 1988. Microenvironments of soil microorganisms. Biol Fertil Soils **6**:189–203.

66. Bailey VL, Fansler SJ, Stegen JC, McCue LA. 2013. Linking microbial community structure to β-glucosidic function in soil aggregates. ISME J **7**:2044–2053.

67. Ranjard L, Richaume A. 2001. Quantitative and qualitative microscale distribution of bacteria in soil. Res Microbiol **152**:707–716.

68. Kilbertus G. 1980. Etude des microhabitats contenus dans les agrégats du sol. Leur relation avec la biomasse bactérienne et la taille des procaryotes présents. Rev d'écologie Biol du Sol **17**:543–557.

69. Hassink, J., Bouwman, L.A., Zwart, K.B., Brussaard L. 1993. Relationships between habitable pore space, soil biota and mineralization rates in grassland soils. Soil Biol Biochem **25**:47–55.

70. Vos M, Wolf AB, Jennings SJ, Kowalchuk GA. 2013. Micro-scale determinants of bacterial diversity in soil. FEMS Microbiol Rev **37**:936–954.

71. Chenu C, Hassink J, Bloem J. 2001. Short-term changes in the spatial distribution of microorganisms in soil aggregates as affected by glucose addition. Biol Fertil Soils **34**:349–356.

72. Sessitsch A, Weilharter A, Martin H, Kirchmann H, Kandeler E, Gerzabek MH. 2001. Microbial population structures in soil particle size fractions of a long-term fertilizer field experiment. Appl Environ Microbiol **67**:4215–4224.

73. Serna-Chavez HM, Fierer N, Van Bodegom PM. 2013. Global drivers and patterns of microbial abundance in soil. Glob Ecol Biogeogr **22**:1162–1172.

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

75. Haynes RJ. 1983. Soil acidification induced by leguminous crops. Grass Forage Sci **38**:1–11.

76. Bouman OT, Curtin D, Campbell CA, Biederbeck VO, Ukrainetz H. 1995. Soil acidification from long-term use of anhydrous ammonia and urea. Soil Sci Soc Am J **59**:1488–1494.

77. Adams F. 1984. Crop response to lime in southern United States, p. 211–265. *In* Adams, F (ed.), Soil Acidity and Liming, Agronomy Monograph, 2nd ed. ASA, CSSA, and SSSA, Madison, Wisconsin.

78. Ritchie GSP, Dolling PJ. 1985. The role of organic matter in soil acidification. Aust J Soil Res **23**:569–576.

79. Jones D. 1998. Organic acids in the rhizospere - a critical review. Plant Soil Soil **205**:25–44.

80. Blake L, Goulding KWT, Mott CJB, Johnston AE. 1999. Changes in soil chemistry accompanying acidification over more than 100 years under woodland and grass at Rothamsted Experimental Station, UK. Eur J Soil Sci **50**:401–412.

81. Fierer N, Jackson RB. 2006. The diversity and biogeography of soil bacterial communities. Proc Natl Acad Sci U S A **103**:626–631.

82. Rousk J, Bååth E, Brookes PC, Lauber CL, Lozupone C, Caporaso JG, Knight R, Fierer N. 2010. Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J **4**:1340–1351.

83. Nemergut DR, Costello EK, Hamady M, Lozupone C, Jiang L, Schmidt SK, Fierer N, Townsend AR, Cleveland CC, Stanish L, Knight R. 2011. Global patterns in the biogeography of bacterial taxa. Environ Microbiol **13**:135–144.

84. Bengtson P, Sterngren AE, Rousk J. 2012. Archaeal abundance across a pH gradient in an arable soil and its relationship to bacterial and fungal growth rates. Appl Environ Microbiol **78**:5906–5911.

85. Blagodatskaya EV, Anderson T. 1998. Interactive effects of pH and substrate quality on the fungal-to-bacterial ratio and QCO_2 of microbial communities in forest soils. Soil Biol Biochem **30**:1269–1274.

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

87. Peng YL, Yang MN, Cai XB. 2010. Influence of soil factors on species diversity of arbuscular mycorrhizal (AM) fungi in Stipa steppe of Tibet Plateau. J Appl Ecol **21**:1258–1263.

88. Fierer N, Schimel JP, Holden PA. 2003. Variations in microbial community composition through two soil depth profiles. Soil Biol Biochem **35**:167–176.

89. Trumbore S. 2000. Age of soil organic matter and soil respiration: radiocarbon constraints on belowground C dynamics. Ecol Appl **10**:399–411.

90. Zhang X, Liu W, Schloter M, Zhang G, Chen Q, Huang J, Li L, Elser JJ, Han X. 2013. Response of the abundance of key soil microbial nitrogen-cycling genes to multi-factorial global changes. PLoS One **8**:e76500.

91. Sun B, Roberts DM, Dennis PG, Caul S, Daniell TJ, Hallett PD, Hopkins DW. 2013. Microbial properties and nitrogen contents of arable soils under different tillage regimes. Soil Use Manag **30**:152–159.

92. Matejek B, Huber C, Dannenmann M, Kohlpaintner M, Gasche R, Göttlein A, Papen H. 2010. Microbial nitrogen-turnover processes within the soil profile of a nitrogen-saturated spruce forest and their relation to the small-scale pattern of seepage-water nitrate. J Plant Nutr Soil Sci **173**:224–236.

93. Rütting T, Boeckx P, Müller C, Klemedtsson L. 2011. Assessment of the importance of dissimilatory nitrate reduction to ammonium for the terrestrial nitrogen cycle. Biogeosciences **8**:1779–1791.

94. Colliver BB, Stephenson T. 2000. Production of nitrogen oxide and dinitrogen oxide by autotrophic nitrifiers. Biotechnol Adv **18**:219–232.

95. Schnurer J, Clarholm M, Bostrom S, Rosswall T. 1986. Effects of moisture on soil-microorganisms and nematodes - a field experiment. Microb Ecol **12**:217–230.

96. Schellenberger S, Drake HL, Kolb S. 2011. Functionally redundant cellobiose-degrading soil bacteria respond differentially to oxygen. Appl Environ Microbiol **77**:6043–6048.

97. Wichern J, Wichern F, Joergensen RG. 2006. Impact of salinity on soil microbial communities and the decomposition of maize in acidic soils. Geoderma **137**:100–108.

98. Pathak H, Rao DLN. 1998. Carbon and nitrogen mineralization from added organic matter in saline and alkali soils. Soil Biol Biochem **30**:695–702.

99. Prescott CE, Grayston SJ. 2013. Tree species influence on microbial communities in litter and soil: Current knowledge and research needs. For Ecol Manage **309**:19–27.

100. Prober SM, Leff JW, Bates ST, Borer ET, Firn J, Harpole WS, Lind EM, Seabloom EW, Adler PB, Bakker JD, Cleland EE, DeCrappeo NM, DeLorenze E, Hagenah N, Hautier Y, Hofmockel KS, Kirkman KP, Knops JMH, La Pierre KJ, MacDougall AS, McCulley RL, Mitchell CE, Risch AC, Schuetz M, Stevens CJ, Williams RJ, Fierer N. 2015. Plant diversity predicts beta but not alpha diversity of soil microbes across grasslands worldwide. Ecol Lett **18**:85–95.

101. Bakker MG, Schlatter DC, Otto-Hanson L, Kinkel LL. 2014. Diffuse symbioses: roles of plant-plant, plant-microbe and microbe-microbe interactions in structuring the soil microbiome. Mol Ecol **23**:1571–1583.

102. Kinkel LL, Bakker MG, Schlatter DC. 2011. A coevolutionary framework for managing disease-suppressive soils. Annu Rev Phytopathol **49**:47–67.

103. **S**tams AJM, Plugge CM. 2009. Electron transfer in syntrophic communities of anaerobic bacteria and archaea. Nat Rev Microbiol **7**:568–577.

104. Wright DA, Killham K, Glover LA, Prosser JI. 1995. Role of pore size location in determining bacterial activity during predation by protozoa in soil. Appl Environ Microbiol **61**:3537–3543.

105. O'Malley MA. 2007. The nineteenth century roots of "everything is everywhere". Nat Rev Microbiol **5**:647–651.

106. Ranjard L, Dequiedt S. 2010. Biogeography of soil microbial communities: a review and a description of the ongoing french national initiative. Agron Sustain Dev **30**:359–365.

107. Zavarzin GA. 2006. Winogradsky and modern microbiology. Microbiology **75**:581–592.

108. Dworkin M. 2012. Sergei Winogradsky: a founder of modern microbiology and the first microbial ecologist. FEMS Microbiol Rev **36**:364–379.

109. Darwin C. 1883. The Variation of Animals and Plants under Domestication, 2nd ed. Appleton & Co, New York, New York.

110. Van Iterson G, den Dooren de Jong LE, Kluyver AJ. 1983. Martinus Willem Beijerinck: His Life and His Work. Science Tech, Madison, Wisconsin.

111. De Wit R, Bouvier T. 2006. "Everything is everywhere, but, the environment selects"; what did Baas Becking and Beijerinck really say? Environ Microbiol **8**:755–758.

112. Baas Becking LGM. 1934. Geobiologie of Inleiding Tot de Milieukunde. Van Stockum & Zoon, The Hague, South Holland.

113. van Niel CB. 1949. The Delft School and the rise of general microbiology. Bacteriol Rev **13**:161–174.

114. La Rivière JWM. 1997. The Delft School of Microbiology in historical perspective. Antonie van Leeuwenhoek J Microbiol **71**:3–13.

115. Staley JT, Gosink JJ. 1999. Poles apart: biodiversity and biogeography of sea ice bacteria. Annu Rev Microbiol **53**:189–215.

116. Papke RT, Ramsing NB, Bateson MM, Ward DM. 2003. Geographical isolation in hot spring cyanobacteria. Environ Microbiol **5**:650–659.

117. Whitaker RJ, Grogan DW, Taylor JW. 2003. Geographic barriers isolate endemic populations of hyperthermophilic archaea. Science **301**:2002–2004.

118. Martiny J, Bohannan BJM, Brown JH, Colwell RK, Fuhrman JA, Green JL, Horner-Devine MC, Kane M, Krumins JA, Kuske CR, Morin PJ, Naeem S, Ovreås L, Reysenbach A, Smith VH, Staley JT. 2006. Microbial biogeography: putting microorganisms on the map. Nature **4**:102–112.

119. Martiny AC, Treseder K, Pusch G. 2013. Phylogenetic conservatism of functional traits in microorganisms. ISME J **7**:830–838.

120. de Candolle AP. 1820. Essai Élémentaire de Geographie Botanique. F. G. Levrault, Paris, France.

121. Horner-Devine CM, Lage M, Hughes JB, Bohannan BJM. 2004. A taxa-area relationship for bacteria. Nature **432**:750–753.

122. Jangid K, Williams MA, Franzluebbers AJ, Schmidt TM, Coleman DC, Whitman WB. 2011. Land-use history has a stronger impact on soil microbial community composition than aboveground vegetation and soil properties. Soil Biol Biochem **43**:2184–2193.

123. Hanson CA, Fuhrman JA, Horner-Devine CM, Martiny J. 2012. Beyond biogeographic patterns: processes shaping the microbial landscape. Nat Rev Microbiol **10**:497–506.

124. van Loosdrecht MCB, Norde W, Lyklema J, Zehnder AJB. 1990. Hydrophobic and electrostatic parameters in bacterial adhesion. Aquat Sci **52**:103–114.

125. Abu-Ashour J, Joy DM, Lee H, Whiteley HR, Zelin S. 1994. Transport of microorganisms through soil. Water, Air Soil Pollut **75**:141–158.

126. Locey KJ. 2010. Synthesizing traditional biogeography with microbial ecology: the importance of dormancy. J Biogeogr **37**:1835–1841.

127. Wilkinson DM, Koumoutsaris S, Mitchell EAD, Bey I. 2012. Modelling the effect of size on the aerial dispersal of microorganisms. J Biogeogr **39**:89–97.

128. Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. Proc Natl Acad Sci U S A **91**:6808–6814.

129. Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. Nature **394**:69–72.

130. Kim M, Heo E, Kang H, Adams J. 2013. Changes in soil bacterial community structure with increasing disturbance frequency. Microb Ecol **66**:171–181.

131. Nekola JC, White PS. 2007. The distance decay of similarity in biogeography and ecology. J Biogeogr **26**:867–878.

132. Cho JC, Tiedje JM. 2000. Biogeography and degree of endemicity of fluorescent *Pseudomonas* strains in soil. Appl Environ Microbiol **66**:5448–5456.

133. Staddon WJ, Trevors JT, Duchesne LC, Colombo CA. 1998. Soil microbial diversity and community structure across a climatic gradient in western Canada. Biodivers Conserv **7**:1081–1092.

134. Buckley HL, Miller TE, Ellison AM, Gotelli NJ. 2003. Reverse latitudinal trends in species richness of pitcher-plant food webs. Ecol Lett **6**:825–829.

135. Yergeau E, Newsham KK, Pearce DA, Kowalchuk GA. 2007. Patterns of bacterial diversity across a range of Antarctic terrestrial habitats. Environ Microbiol **9**:2670–2682.

136. Fierer N. 2008. Microbial biogeography: patterns in microbial diversity across space and time, p. 95–115. *In* Zengler, K (ed.), Accessing Uncultivated Microorganisms: from the Environment to Organisms and Genomes and Back. ASM Press, Washington, D.C.

137. Blume E, Bischoff M, Reichert JM, Moorman T, Konopka A, Turco RF. 2002. Surface and subsurface microbial biomass, community structure and metabolic activity as a function of soil depth and season. Appl Soil Ecol **20**:171–181.

138. Hartmann M, Lee S, Hallam SJ, Mohn WW. 2009. Bacterial, archaeal and eukaryal community structures throughout soil horizons of harvested and naturally disturbed forest stands. Environ Microbiol **11**:3045–3062.

139. Buss H, Bruns M, Schultz M. 2005. The coupling of biological iron cycling and mineral weathering during saprolite formation, Luquillo Mountains, Puerto Rico. Geobiology **3**:247–260.

140. Fierer N, Chadwick OA, Trumbore SE. 2005. Production of CO_2 in soil profiles of a California annual grassland. Ecosystems **8**:412–429.

141. Deng J, Gu Y, Zhang J, Xue K, Qin Y, Yuan M, Yin H, He Z, Wu L, Schuur EAG, Tiedje JM, Zhou J. 2015. Shifts of tundra bacterial and archaeal communities along a permafrost thaw gradient in Alaska. Mol Ecol **24**:222–234.

142. Gittel A, Bárta J, Kohoutová I, Mikutta R, Owens S, Gilbert J, Schnecker J, Wild B, Hannisdal B, Maerz J, Lashchinskiy N, Capek P, Santrůčková H, Gentsch N, Shibistova O, Guggenberger G, Richter A, Torsvik VL, Schleper C, Urich T. 2014. Distinct microbial communities associated with buried soils in the Siberian tundra. ISME J **8**:841–53.

143. Potthoff M, Steenwerth KL, Jackson LE, Drenovsky RE, Scow KM, Joergensen RG. 2006. Soil microbial community composition as affected by restoration practices in California grassland. Soil Biol Biochem **38**:1851–1860.

144. Oehl F, Sieverding E, Ineichen K. 2005. Community structure of arbuscular mycorrhizal fungi at different soil depths in extensively and intensively managed agroecosystems. New Phytol **165**:273–283.

145. Ekelund F, Rønn R, Christensen S. 2001. Distribution with depth of protozoa, bacteria and fungi in soil profiles from three Danish forest sites. Soil Biol Biochem **33**:475–481.

146. Lipson DA, Raab TK, Parker M, Kelley ST, Brislawn CJ, Jansson J. 2015. Changes in microbial communities along redox gradients in polygonized Arctic wet tundra soils. Environ Microbiol Rep **7**:649–657.

147. Kim HM, Jung JY, Yergeau E, Hwang CY, Hinzman L, Nam S, Hong SG, Kim OS, Chun J, Lee YK. 2014. Bacterial community structure and soil properties of a subarctic tundra soil in Council, Alaska. FEMS Microbiol Ecol **89**:465–475.

148. Griffiths RI, Whiteley AS, O'Donnell AG, Bailey MJ. 2003. Influence of depth and sampling time on bacterial community structure in an upland grassland soil. FEMS Microbiol Ecol **43**:35–43.

149. Kemnitz D, Kolb S, Conrad R. 2007. High abundance of Crenarchaeota in a temperate acidic forest soil. FEMS Microbiol Ecol **60**:442–448.

150. Will C, Thürmer A, Wollherr A, Nacke H, Herold N, Schrumpf M, Gutknecht J, Wubet T, Buscot F, Daniel R. 2010. Horizon-specific bacterial community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl Environ Microbiol **76**:6751–6759.

151. Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA. 2005. Isolation of an autotrophic ammonia-oxidizing marine archaeon. Nature **437**:543–546.

152. Ramette A, Tiedje JM. 2007. Biogeography: an emerging cornerstone for understanding prokaryotic diversity, ecology, and evolution. Microb Ecol **53**:197–207.

153. Gaunt M, Sall AA, de Lamballerie X, Falconar KI, Dzhiranian TI, Gould EA. 2001. Phylogenetic relationships of aviviruses correlate with their epidemiology, disease association and biogeography. J Gen Virol **82**:1867–1876.

154. Rossello-Mora R, Amann R. 2001. The species concept for prokaryotes. FEMS Microbiol Ecol **25**:39–67.

155. Konstantinidis KT, Ramette A, Tiedje JM. 2006. The bacterial species definition in the genomic era. Philos Trans R Soc B **361**:1929–1940.

156. Hill GT, Mitkowski NA, Aldrich-Wolfe L, Emele LR, Jurkonie DD, Ficke A, Maldonado-Ramirez S, Lynch ST, Nelson EB. 2000. Methods for assessing the composition and diversity of soil microbial communities. Appl Soil Ecol **15**:25–36.

157. Thomas T, Gilbert J, Meyer F. 2012. Metagenomics - a guide from sampling to data analysis. Microb Inform Exp **2**:1–12.

158. Jovanovich SB, Gates CM, Feldman RA, Spudich JL, Spudich EN, Delong EF. 2000. Bacterial rhodopsin: evidence for a new type of phototrophy in the sea. Science **289**:1902–1907.

159. Nicol GW, Schleper C. 2006. Ammonia-oxidising Crenarchaeota: important players in the nitrogen cycle? Trends Microbiol **14**:207–212.

160. Wang J, Qi J, Zhao H, He S, Zhang Y, Wei S, Zhao F. 2013. Metagenomic sequencing reveals microbiota and its functional potential associated with periodontal disease. Sci Rep **3**:1–10.

161. Langille MGI, Zaneveld J, Caporaso JG, McDonald D, Knights D, Reyes JA, Clemente JC, Burkepile DE, Vega Thurber RL, Knight R, Beiko RG, Huttenhower C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat Biotechnol **31**:814–821.

162. DeSantis TZ, Hugenholtz P, Larsen N, Rojas M, Brodie EL, Keller K, Huber T, Dalevi D, Hu P, Andersen GL. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Appl Environ Microbiol **72**:5069–5072.

163. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, Horn M, Glöckner FO. 2013. Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res **41**:1–11.

164. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res **41**:1–7.

165. Levy-Booth DJ, Campbell RG, Gulden RH, Hart MM, Powell JR, Klironomos JN, Pauls KP, Swanton CJ, Trevors JT, Dunfield KE. 2007. Cycling of extracellular DNA in the soil environment. Soil Biol Biochem **39**:2977–2991.

166. Blazewicz SJ, Barnard RL, Daly RA, Firestone MK. 2013. Evaluating rRNA as an indicator of microbial activity in environmental communities: limitations and uses. ISME J **7**:2061–2068.

167. Nocker A, Sossa-Fernandez P, Burr MD, Camper AK. 2007. Use of propidium monoazide for live/dead distinction in microbial ecology. Appl Environ Microbiol **73**:5111–5117.

168. Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol **26**:1135–1145.

169. Metzker ML. 2010. Sequencing technologies - the next generation. Nat Rev Genet **11**:31–46.

170. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C. 2014. Ten years of next-generation sequencing technology. Trends Genet **30**:418–426.

171. Branton D, Deamer DW, Marziali A, Bayley H, Benner SA, Butler T, Di Ventra M, Garaj S, Hibbs A, Huang X, Jovanovich SB, Krstic PS, Lindsay S, Ling XS, Mastrangelo CH, Meller A, Oliver JS, Pershin Y V., Ramsey JM, Riehn R, Soni GV., Tabard-Cossa V, Wanunu M, Wiggin M, Schloss JA. 2008. The potential and challenges of nanopore sequencing. Nat Biotechnol **26**:1146–1153.

172. Masella AP, Bartram AK, Truszkowski JM, Brown DG, Neufeld JD. 2012. PANDAseq: paired-end assembler for illumina sequences. BMC Bioinformatics **13**:1–7.

173. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: A fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics **30**:614–620.

174. Liu B, Yuan J, Yiu SM, Li Z, Xie Y, Chen Y, Shi Y, Zhang H, Li Y, Lam TW, Luo R. 2012. COPE: An accurate k-mer-based pair-end reads connection tool to facilitate genome assembly. Bioinformatics **28**:2870–2874.

175. Chen W, Zhang CK, Cheng Y, Zhang S, Zhao H. 2013. A comparison of methods for clustering 16S rRNA sequences into OTUs. PLoS One **8**:e70837.

176. **S**chloss PD, Westcott SL. 2011. Assessing and improving methods used in operational taxonomic unit-based approaches for 16S rRNA gene sequence analysis. Appl Environ Microbiol **77**:3219–3226.

177. Li W, Godzik A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics **22**:1658–1659.

178. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics **26**:2460–2461.

179. Edgar RC. 2013. UPARSE: highly accurate OTU sequences from microbial amplicon reads. Nat Methods **10**:996–998.

180. Schneider B. 2012. Agricultural Practices at rare Charitable Research Reserve. Cambridge, Ontario.

181. Presant EW, Wicklund RE. 1971. Soils of Waterloo County. Canada Department of Agriculture, Ottawa, Ontario.

182. Yilmaz P, Kottmann R, Field D, Knight R, Cole JR, Amaral-Zettler L, Gilbert JA, Karsch-Mizrachi I, Johnston A, Cochrane G, Vaughan R, Hunter C, Park J, Morrison N, Rocca-Serra P, Sterk P, Arumugam M, Bailey M, Baumgartner L, Birren BW, Blaser MJ, Bonazzi V, Booth T, Bork P, Bushman FD, Buttigieg PL, Chain PSG, Charlson E, Costello EK, Huot-Creasy H, Dawyndt P, DeSantis T, Fierer N, Fuhrman JA, Gallery RE, Gevers D, Gibbs RA, San Gil I, Gonzalez A, Gordon JI, Guralnick R, Hankeln W, Highlander S, Hugenholtz P, Jansson J, Kau AL, Kelley ST, Kennedy J, Knights D, Koren O, Kuczynski J, Kyrpides N, Larsen R, Lauber CL, Legg T, Ley RE, Lozupone CA, Ludwig W, Lyons D, Maguire E,

Methé BA, Meyer F, Muegge B, Nakielny S, Nelson KE, Nemergut D, Neufeld JD, Newbold LK, Oliver AE, Pace NR, Palanisamy G, Peplies J, Petrosino J, Proctor L, Pruesse E, Quast C, Raes J, Ratnasingham S, Ravel J, Relman DA, Assunta-Sansone S, Schloss PD, Schriml L, Sinha R, Smith MI, Sodergren E, Spo A, Stombaugh J, Tiedje JM, Ward D V., Weinstock GM, Wendel D, White O, Whiteley A, Wilke A, Wortman JR, Yatsunenko T, Glöckner FO. 2011. Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications. Nat Biotechnol **29**:415–420.

183. Hendershot WH, Lalande H, Duquette M. 1993. Soil Reaction and Exchangeable Acidity, p. 141–142. *In* Carter, MR (ed.), Soil Sampling and Methods of Analysis, 1st ed. Lewis Publishers, Boca Raton, Florida.

184. Maynard DG, Kalra YP. 1993. Nitrate and Exchangeable Ammonium Nitrogen, p. 25–38. *In* Carter, MR (ed.), Soil Sampling and Methods of Analysis, 1st ed. Lewis Publishers, Boca Raton, Florida.

185. U. S. Environmental Protection Agency. 1983. Nitrogen, Ammonia, Method 350.1 (Colorimetric, Automated, Phenate), p. 350–1.1–350–1.4. *In* Methods for Chemical Analysis of Water and Wastes, 600/4–79th–0 ed. National Service Center for Environmental Publications, Cincinnati, Ohio.

186. U. S. Environmental Protection Agency. 1993. Nitrogen, Nitrate-Nitrite, Method 353.2 (Colorimetric, Automated, Cadmium Reduction), p. 353.2–1–353.2–7. *In* Methods for Chemical Analysis of Water and Wastes, 600/R93/10 ed. National Service Center for Environmental Publications, Cincinnati, Ohio.

187. U. S. Environmental Protection Agency. 1993. Nitrogen, Nitrite, Method 354.1 (Spectrophotometric), p. 354.1–1–354.1–2. *In* Methods for Chemical Analysis of Water and Wastes, 600/4–79th–0 ed. National Service Center for Environmental Publications, Cincinnati, Ohio.

188. Nelson DW, Sommers LE. 1982. Total Carbon, Organic Carbon, and Organic Matter, p. 549–552. *In* Pages, AL (ed.), Methods of Soil Analysis, Part 2, Chemical and Microbiological Properties, 2nd ed. American Society of Agronomy, Inc., Soil Science Society of America, Inc., Madison, Wisconsin.

189. Sheldrick BH, Wang C. 1993. Particle Size Distribution, p. 499–507. *In* Carter, MR (ed.), Soil Sampling and Methods of Analysis, 1st ed. Lewis Publishers, Boca Raton, Florida.

190. Petersen RG, Calvin LD. 1965. Sampling, p. 44–72. *In* Klute, A (ed.), Methods of Soil Analysis: Part 1 Physical and Mineralogical Methods, 2nd ed. American Society of Agronomy, Madison, Wisconsin.

191. Bartram AK, Lynch MDJ, Stearns JC, Moreno-Hagelsieb G, Neufeld JD. 2011. Generation of multimillion-sequence 16S rRNA gene libraries from complex microbial communities by assembling paired-end illumina reads. Appl Environ Microbiol **77**:3846–3852.

192. Lynch MD, Masella AP, Hall MW, Bartram AK, Neufeld JD. 2013. AXIOME: automated exploration of microbial diversity. Gigascience **2**:1–5.

193. Wang Q, Garrity GM, Tiedje JM, Cole JR. 2007. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Appl Environ Microbiol **73**:5261–5267.

194. McDonald D, Price MN, Goodrich J, Nawrocki EP, DeSantis TZ, Probst A, Andersen GL, Knight R, Hugenholtz P. 2012. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J **6**:610–618.

195. Gower JC. 1966. Distance properties of latent root and vector methods used in multivariate analysis. Biometrika Trust **53**:325–338.

196. Bray C, Curtis JT. 1957. An ordination of the upland forest communites of Southern Wisconsin. Ecol Monogr **27**:325–349.

197. Mielke PW, Berry KJ, Johnson ES. 1976. Multi-response permutation procedures for a priori classifications. Commun Stat Theory Methods **5**:1409–1424.

198. Clarke KR. 1993. Non-parametric multivariate analyses of changes in community structure. Aust J Ecol **18**:117–143.

199. McCune B, Grace J. 2002. MRPP (Multi-response permutation procedures) and related techniques, p. 304. *In* Analysis of Ecological Communities. MjM Software Design, Gleneden Beach, Oregon.

200. Anderson MJ. 2001. A new method for non-parametric multivariate analysis of variance. Austral Ecol **26**:32–46.

201. **J**ari Oksanen, F. Guillaume Blanchet, Roeland Kindt, Pierre Legendre, Peter R. Minchin, R. B. O'Hara, Gavin L. Simpson, Peter Solymos, M. Henry H. Stevens HW. 2013. vegan: Community Ecology Package R package. 2.3-0.

202. R Core Team. 2014. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

203. Legendre P, Anderson MJ. 1999. Distance-based redundancy analysis: testing multispecies responses in multifactorial ecological experiments. Ecol Monogr **69**:1–24.

204. Ramette A. 2007. Multivariate analyses in microbial ecology. FEMS Microbiol Ecol **62**:142–160.

205. Kim M, Cho A, Lim HS, Hong SG, Kim JH, Lee J, Choi T, Ahn TS, Kim OS. 2015. Highly heterogeneous soil bacterial communities around Terra Nova Bay of Northern Victoria Land, Antarctica. PLoS One **10**:e0119966.

206. McMurdie PJ, Holmes S. 2013. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One **8**:e61217.

207. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI, Huttley GA, Kelley ST, Knights D, Koenig JE, Ley RE,

Lozupone CA, McDonald D, Muegge BD, Pirrung M, Reeder J, Sevinsky JR, Turnbaugh PJ, Walters WA, Widmann J, Yatsunenko T, Zaneveld J, Knight R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nat Methods **7**:335–336.

208. Dufrêne M, Legendre P. 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol Monogr **67**:345–366.

209. Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B **57**:289–300.

210. Edgar RC, Haas BJ, Clemente JC, Quince C, Knight R. 2011. UCHIME improves sensitivity and speed of chimera detection. Bioinformatics **27**:2194–2200.

211. Zaneveld JR, Lozupone C, Gordon JI, Knight R. 2010. Ribosomal RNA diversity predicts genome diversity in gut bacteria and their relatives. Nucleic Acids Res **38**:3869–3879.

212. Markowitz VM, Chen IMA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, Anderson I, Billis K, Varghese N, Mavromatis K, Pati A, Ivanova NN, Kyrpides NC. 2014. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res **42**:560–567.

213. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. 2012. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **40**:109–114.

214. Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20**:289–290.

215. Parks DH, Tyson GW, Hugenholtz P, Beiko RG. 2014. STAMP: Statistical analysis of taxonomic and functional profiles. Bioinformatics **30**:3123–3124.

216. Tukey JW. 1949. Comparing individual means in the analysis of variance. Biometrics **5**:99–114.

217. Sinclair L, Osman OA, Bertilsson S, Eiler A. 2015. Microbial community composition and diversity via 16S rRNA gene amplicons: evaluating the Illumina platform. PLoS One **10**:e0116955.

218. Schmidt TSB, Matias Rodrigues JF, von Mering C. 2014. Limits to robustness and reproducibility in the demarcation of operational taxonomic units. Environ Microbiol **17**:1689–1706.

219. Pylro VS, Roesch LFW, Morais DK, Clark IM, Hirsch PR, Tótola MR. 2014. Data analysis for 16S microbial profiling from different benchtop sequencing platforms. J Microbiol Methods **107**:30–37.

220. Nayak S, Prasanna R. 2007. Soil pH and its role in cyanobacterial abundance and diversity in rice field soils. Appl Ecol Environ Res **5**:103–113.

221. Van Breemen N, Mulder J, Driscoll CT. 1983. Acidification and alkalinization of soils. Plant Soil **75**:283–308.

222. Yang YH, Fang JY, Guo DL, Ji CJ, Ma WH. 2010. Vertical patterns of soil carbon, nitrogen and carbon: nitrogen stoichiometry in Tibetan grasslands. Biogeosciences Discuss **7**:1–24.

223. Richter DD, Markewitz D. 1995. How deep is soil? Bioscience **45**:600–609.

224. Fissore C, Giardina CP, Kolka RK, Trettin CC, King GM, Jurgensen MF, Barton CD, Mcdowell SD. 2008. Temperature and vegetation effects on soil organic carbon quality along a forested mean annual temperature gradient in North America. Glob Chang Biol **14**:193–205.

225. Garbeva P, van Veen JA, van Elsas JD. 2004. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. Annu Rev Phytopathol **42**:243–270.

226. Ramirez KS, Lauber CL, Knight R, Bradford MA, Fierer N. 2010. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. Ecology **91**:3414–3463.

227. Li Y, Wen H, Chen L, Yin T. 2014. Succession of bacterial community structure and diversity in soil along a chronosequence of reclamation and re-vegetation on coal mine spoils in China. PLoS One **9**:e115024.

228. Lavorela S, McIntyre S, Landsberg J, Forbes TDA. 1997. Plant functional classifications: from general groups to specific groups based on response to disturbance. Trends Ecol Evol **5347**:474–478.

229. Fierer N, Nemergut D, Knight R, Craine JM. 2010. Changes through time: integrating microorganisms into the study of succession. Res Microbiol **161**:635–642.

230. Fierer N, Bradford MA, Jackson RB. 2007. Toward an ecological classification of soil bacteria. Ecology **88**:1354–1364.

231. Bergmann GT, Bates ST, Eilers KG, Lauber CL, Caporaso G, Walters WA, Knight R, Fierer N. 2012. The under-recognized dominance of Verrucomicrobia in soil bacterial communities. Soil Biol Biochem **43**:1450–1455.

232. DeBruyn JM, Nixon LT, Fawaz MN, Johnson AM, Radosevich M. 2011. Global biogeography and quantitative seasonal dynamics of Gemmatimonadetes in soil. Appl Environ Microbiol **77**:6295–6300.

233. Dunbar J, Barns SM, Ticknor LO, Kuske CR. 2002. Empirical and theoretical bacterial diversity in four Arizona soils. Appl Environ Microbiol **68**:3035–3045.

234. Platas G, Morón R, González I, Collado J, Genilloud O, Peláez F, Diez MT. 1998. Production of antibacterial activities by members of the family Pseudonocardiaceae: influence of nutrients. World J Microbiol Biotechnol **14**:521–527.

235. Shen C, Xiong J, Zhang H, Feng Y, Lin X, Li X, Liang W, Chu H. 2013. Soil pH drives the spatial distribution of bacterial communities along elevation on Changbai Mountain. Soil Biol Biochem **57**:204–211.

236. Jones RT, Robeson MS, Lauber CL, Hamady M, Knight R, Fierer N. 2009. A comprehensive survey of soil acidobacterial diversity using pyrosequencing and clone library analyses. Int Soc Microb Ecol **3**:442–453.

237. Kuramae EE, Yergeau E, Wong LC, Pijl AS, van Veen JA, Kowalchuk GA**.** 2012. Soil characteristics more strongly influence soil bacterial communities than land-use type. FEMS Microbiol Ecol **79**:12–24.

238. Daims H. 2014. The Family Nitrospiraceae, p. 733–749. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes, 4th ed. Springer, New York, New York.

239. Yamada T, Sekiguchi Y, Hanada S, Imachi H, Ohashi A, Harada H, Kamagata Y. 2006. *Anaerolinea thermolimosa* sp. nov., *Levilinea saccharolytica* gen. nov., sp. nov. and *Leptolinea tardivitalis* gen. nov., sp. nov., novel filamentous anaerobes, and description of the new classes *Anaerolineae classis* nov. and *Caldilineae classis* nov. in the. Int J Syst Evol Microbiol **56**:1331–1340.

240. George IF, Hartmann M, Liles MR, Agathos SN. 2011. Recovery of as-yet-uncultured soil *Acidobacteria* on dilute solid media. Appl Environ Microbiol **77**:8184–8188.

241. Ward NL, Challacombe JF, Janssen PH, Henrissat B, Coutinho PM, Wu M, Xie G, Haft DH, Sait M, Badger J, Barabote RD, Bradley B, Brettin TS, Brinkac LM, Bruce D, Creasy T, Daugherty SC, Davidsen TM, DeBoy RT, Detter JC, Dodson RJ, Durkin AS, Ganapathy A, Gwinn-Giglio M, Han CS, Khouri H, Kiss H, Kothari SP, Madupu R, Nelson KE, Nelson WC, Paulsen I, Penn K, Ren Q, Rosovitz MJ, Selengut JD, Shrivastava S, Sullivan SA, Tapia R, Thompson S, Watkins KL, Yang Q, Yu C, Zafar N, Zhou L, Kuske CR. 2009. Three genomes from the phylum *Acidobacteria* provide insight into the lifestyles of these microorganisms in soils. Appl Environ Microbiol **75**:2046–2056.

242. Rosso L, Lobry JR, Bajard S. 1995. Convenient model to describe the combined effects of temperature and pH on microbial. Appl Environ Microbiol **61**:610–616.

243. Baker-Austin C, Dopson M. 2007. Life in acid: pH homeostasis in acidophiles. Trends Microbiol **15**:165–171.

244. Fernández-Calviño D, Bååth E. 2010. Growth response of the bacterial community to pH in soils differing in pH. FEMS Microbiol Ecol **73**:149–156.

245. Kuever J, Rainey FA, Widdel F. 2005. The deltaproteobacterial orders *Desulfovibrionales*, *Desulfobacterales*, *Desulfarcales*, and *Syntrophobacterales*, p. 925–1040. *In* Garrity, G (ed.), Bergey's Manual of Systematic Bacteriology. Springer, New York, New York.

246. Albuquerque L, da Costa MS. 2014. The Family *Gaiellaceae*, p. 357–360. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes, 4th ed. Springer, New York, New York.

247. Carrillo L, Benítez-Ahrendts MR. 2014. The Family *Thermoactinomycetaceae*, p. 389–411. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes, 4th ed. Springer, New York, New York.

248. Shanmugam M, Stackebrandt E. 2014. The Family *Paenibacillaceae*, p. 268–280. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes, 4th ed. Springer, New York, New York.

249. Albuquerque L, da Costa MS. 2014. The Families *Conexibacteraceae*, *Patulibacteraceae* and *Solirubrobacteraceae*, p. 185–200. *In* Rosenberg, E, DeLong, EF, Lory, S, Stackebrandt, E, Thompson, F (eds.), The Prokaryotes, 4th ed. Springer, New York, New York.

250. Goodfellow M, Williams ST. 1983. Ecology of *Actinomycetes*. Annu Rev Microbiol **37**:189–216.

251. Youssef NH, Farag IF, Rinke C, Hallam SJ, Woyke T, Elshahed MS. 2015. In silico analysis of the metabolic potential and niche specialization of candidate phylum "*Latescibacteria*" (WS3). PLoS One **10**:e0127499.

252. Best DJ, Roberts DE. 1975. Algorithm AS 89: the upper tail probabilities of Spearman's rho. J R Stat Soc Ser C **24**:377–379.

253. Aßhauer KP, Wemheuer B, Daniel R, Meinicke P. 2015. Tax4Fun: predicting functional profiles from metagenomic 16S rRNA data. Bioinformatics **31**:2882–2884.

254. Cai L, Ye L, Tong AHY, Lok S, Zhang T. 2013. Biased diversity metrics revealed by bacterial 16S pyrotags derived from different primer sets. PLoS One **8**:e53649.

255. Lynch MDJ, Neufeld JD. 2015. Ecology and exploration of the rare biosphere. Nat Rev Microbiol **13**:217–229.

256. Kuramae EE, Gamper HA, Yergeau E, Piceno YM, Brodie EL, Desantis TZ, Andersen GL, van Veen JA, Kowalchuk GA. 2010. Microbial secondary succession in a chronosequence of chalk grasslands. ISME J **4**:711–715.

257. Kuramae E, Gamper H, Van Veen J, Kowalchuk G. 2011. Soil and plant factors driving the community of soil-borne microorganisms across chronosequences of secondary succession of chalk grasslands with a neutral pH. FEMS Microbiol Ecol **77**:285–294.

258. Liiri M, Häsä M, Haimi J, Setälä H. 2012. History of land-use intensity can modify the relationship between functional complexity of the soil fauna and soil ecosystem services - a microcosm study. Appl Soil Ecol **55**:53–61.

259. Jangid K, Whitman WB, Condron LM, Turner BL, Williams MA. 2013. Soil bacterial community succession during long-term ecosystem development. Mol Ecol **22**:3415–3424.

260. Johnson EA, Miyanishi K. 2008. Testing the assumptions of chronosequences in succession. Ecol Lett **11**:419–431.

261. Ogilvie LA. 2012. Microbial Ecological Theory: Current Perspectives. Horizon Scientific Press, Norfolk, United Kingdom.

262. Singh J, Behal A, Singla N, Joshi A, Birbian N, Singh S, Bali V, Batra N. 2009. Metagenomics: concept, methodology, ecological inference and recent advances. Biotechnol J **4**:480–494.

263. Allen EE, Banfield JF. 2005. Community genomics in microbial ecology and evolution. Nat Rev Microbiol **3**:489–498.

264. Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W, Eisen JA, Hallam SJ, Kyrpides NC, Stepanauskas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. Nature **499**:431–437.

265. Caroline H, Buckley M. 2008. The uncharted microbial world: microbes and their activities in the environment, p. 1–41. *In* The Uncharted Microbial World: Microbes and their Activities in the Environment. American Society of Microbiology, Seattle, Washington.

266. Prosser JI, Bohannan BJM, Curtis TP, Ellis RJ, Firestone MK, Freckleton RP, Green JL, Green LE, Killham K, Lennon JJ, Osborn M, Solan M, van der Gast CJ, Young JPW. 2007. The role of ecological theory in microbial ecology. Nat Rev Microbiol **5**:384–392.

267. Sheil D, Burslem DFRP. 2013. Defining and defending Connell's intermediate disturbance hypothesis: a response to Fox. Trends Ecol Evol **28**:571–572.

268. Mori H, Maruyama F, Kato H, Toyoda A, Dozono A, Ohtsubo Y, Nagata Y, Fujiyama A, Tsuda M, Kurokawa K. 2014. Design and experimental application of a novel non-degenerate universal primer set that amplifies prokaryotic 16S rRNA genes with a low possibility to amplify eukaryotic rRNA genes. DNA Res **21**:217–227.

269. Anderson IC, Parkin PI. 2007. Detection of active soil fungi by RT-PCR amplification of precursor rRNA molecules. J Microbiol Methods **68**:248–253.

270. Bae S, Wuertz S. 2009. Discrimination of viable and dead fecal *Bacteroidales* bacteria by quantitative PCR with propidium monoazide. Appl Environ Microbiol **75**:2940–2944.

271. Baldrian P, Kolařík M, Štursová M, Kopecký J, Valášková V, Větrovský T, Žifčáková L, Šnajdr J, Rídl J, Vlček Č, Voříšková J. 2012. Active and total microbial communities in forest soil are largely different and highly stratified during decomposition. ISME J **6**:248–258.

272. Campbell BJ, Yu L, Heidelberg JF, Kirchman DL. 2011. Activity of abundant and rare bacteria in a coastal ocean. Proc Natl Acad Sci U S A **108**:12776–12781.

273. Smets W, Leff JW, Bradford MA, Mcculley RL, Lebeer S, Fierer N. 2015. A method for simultaneous measurement of soil bacterial abundances and community composition via 16S rRNA gene sequencing. Peer J **3**:e1622.

274. Souza RC, Cantão ME, Vasconcelos ATR, Nogueira MA, Hungria M. 2013. Soil metagenomics reveals differences under conventional and no-tillage with crop rotation or succession. Appl Soil Ecol **72**:49–61.

275. Aßhauer KP. 2015. Statistical models for large-scale comparative metagenome analysis. Georg August University Göttingen.

Appendix A

Table 14: Summary of physicochemical data. Physicochemical results are based on composite soils obtained from each depth increments from triplicate pits from each subplot.

Sample	Site ID*	Plot	Depth (cm)	TC[a]	SIC[b]	SOC[c]	H_2O	NH_4^+	NO_3^-	NO_2^-	Grav[d]	Sand	Silt	Clay	Texture	pH
				(% dry)				(mg kg^{-1} dry)			(%)					
IW-P1-L1C	IW	1	0-15	1.81	0.00	1.81	21.66	13.7	3.53	<0.03	0.2	73.8	19.4	6.7	Fine sandy loam	7.5
IW-P1-L2C			15-30	0.489	0.00	0.489	13.80	5.49	0.855	<0.03	1.2	76.5	19.8	3.7	Loamy fine sand	7.7
IW-P1-L3C			30-45	1.43	0.980	0.450	11.07	3.95	0.712	<0.03	3.6	77.6	16.3	6.1	Loamy fine sand	8.0
IW-P2-L1C		2	0-15	2.32	0.00	2.32	23.71	13.2	1.45	<0.03	2.7	40.5	49.7	9.8	Loam	6.6
IW-P2-L2C			15-30	0.991	0.00	0.991	19.12	5.52	0.718	<0.03	4.8	41.8	51.3	6.8	Silt loam	6.4
IW-P2-L3C			30-45	0.593	0.00	0.593	15.78	4.89	0.728	<0.03	4.0	45.5	41.6	12.9	Loam	6.6
IW-P3-L1C		3	0-15	1.81	0.00	1.81	25.88	12.3	4.17	<0.03	4.5	50.4	39.1	10.5	Loam	7.1
IW-P3-L2C			15-30	0.712	0.00	0.712	18.54	8.69	1.34	<0.03	4.9	53.0	28.3	18.7	Sandy Loam	7.0
IW-P3-L3C			30-45	1.12	0.959	0.161	15.29	7.49	1.28	<0.03	12.3	61.8	22.6	15.6	Fine sandy loam	7.2
H-P1-L1C	H	1	0-15	2.53	0.00	2.53	17.63	18.5	1.16	<0.03	4.0	31.0	58.1	10.9	Silt loam	6.2
H-P1-L2C			15-30	0.861	0.00	0.861	14.87	7.42	0.701	<0.03	3.2	34.6	53.5	12.0	Silt loam	6.8
H-P1-L3C			30-45	0.542	0.00	0.542	14.45	4.45	1.12	<0.03	5.3	45.1	36.3	18.6	Loam	7.1
H-P2-L1C		2	0-15	3.10	0.131	2.97	13.01	20.5	2.06	<0.03	5.6	58.9	30.7	10.4	Sandy loam	6.5
H-P2-L2C			15-30	0.998	0.00	0.998	10.70	8.41	1.03	<0.03	9.5	58.0	28.4	13.6	Sandy loam	7.0
H-P2-L3C			30-45	3.32	2.70	0.620	6.72	5.00	0.876	<0.03	29.0	67.5	16.5	16.1	Gravelly Sandy Loam	7.6
H-P3-L1C		3	0-15	2.56	0.566	1.99	11.15	5.91	4.63	1.03	10.0	48.0	35.3	16.6	Loam	7.5
H-P3-L2C			15-30	2.00	1.34	0.660	9.90	6.03	1.51	0.106	4.7	55.9	29.7	14.4	Fine sandy loam	7.7
H-P3-L3C			30-45	1.08	0.757	0.323	9.92	4.12	1.19	<0.03	9.8	44.6	38.1	17.4	Loam	7.7
CA-P1-L1C	CA	1	0-15	9.30	0.888	8.41	45.99	22.1	12.3	5.78	0.0	65.0	26.3	8.6	Fine sandy loam	7.4
CA-P1-L2C			15-30	3.13	2.16	0.970	16.36	4.47	2.06	0.204	3.9	74.9	16.1	9.1	Fine sandy loam	7.8
CA-P1-L3C			30-45	3.42	2.73	0.690	13.74	4.11	1.24	<0.03	1.0	75.6	17.4	7.0	Fine sandy loam	7.8
CA-P2-L1C		2	0-15	2.42	0.00	2.42	21.51	10.9	3.78	0.0750	0.0	75.7	17.5	6.8	Fine sandy loam	6.2
CA-P2-L2C			15-30	1.33	0.00	1.33	16.39	4.80	1.87	<0.03	0.0	80.9	16.1	3.0	Loamy fine sand	6.0
CA-P2-L3C			30-45	0.295	0.00	0.295	11.29	4.32	1.32	<0.03	0.0	80.7	13.4	5.9	Loamy fine sand	6.5
CA-P3-L1C		3	0-15	7.84	0.00	7.84	13.12	13.8	1.41	<0.03	0.0	48.7	40.3	11.0	Loam	4.5
CA-P3-L2C			15-30	2.51	0.00	2.51	11.53	4.44	1.06	<0.03	0.0	49.8	40.4	9.9	Loam	5.4
CA-P3-L3C			30-45	1.42	0.00	1.42	8.66	4.08	0.538	<0.03	0.0	56.8	37.5	5.7	Fine sandy loam	5.8
AA-P1-L1C	AA	1	0-15	1.98	0.672	1.31	8.66	2.79	5.62	0.0760	3.3	74.5	16.5	9.1	Sandy loam	7.5
AA-P1-L2C			15-30	2.06	1.62	0.440	6.78	3.05	2.83	<0.03	6.8	80.7	12.7	6.6	Loamy sand	7.8
AA-P2-L1C		2	0-15	1.37	0.00	1.37	9.57	2.50	7.32	<0.03	2.6	77.2	14.9	7.9	Fine sandy loam	7.5

Sample	Site ID*	Plot	Depth (cm)	TC[a]	SIC[b]	SOC[c]	H_2O	NH_4^+	NO_3^-	NO_2^-	Grav[d]	Sand	Silt	Clay	Texture	pH
				(% dry)				(mg kg^{-1} dry)			(%)					
AA-P2-L2C			15-30	1.90	0.208	1.69	12.15	3.67	5.23	0.241	9.2	61.9	29.0	9.1	Fine sandy loam	7.7
AA-P2-L3C			30-45	2.03	0.400	1.63	10.34	4.19	2.85	<0.03	7.5	62.3	27.2	10.5	Fine sandy loam	7.8
AA-P3-L1C			0-15	2.11	0.953	1.16	9.14	2.85	10.4	<0.03	0.7	61.6	27.2	11.2	Fine sandy loam	7.8
AA-P3-L2C		3	15-30	2.34	2.17	0.170	7.43	3.24	4.58	<0.03	0.1	63.3	22.9	13.8	Fine sandy loam	7.9
AA-P3-L3C			30-45	3.76	3.30	0.460	8.70	2.71	2.92	<0.03	0.0	67.6	19.3	13.0	Fine sandy loam	8.0
D03-P1-L1C			0-15	3.18	0.335	2.85	27.01	9.86	12.7	2.19	0.5	32.6	52.7	14.7	Silt loam	7.6
D03-P1-L2C		1	15-30	2.13	0.154	1.98	26.49	5.24	6.19	1.03	1.2	32.3	55.6	12.2	Silt loam	7.8
D03-P1-L3C			30-45	1.78	0.00	1.78	28.51	7.06	4.50	0.677	5.7	38.0	49.1	12.9	Loam	7.7
D03-P2-L1C			0-15	2.39	0.00	2.39	14.30	9.34	6.73	1.03	1.6	36.4	45.8	17.8	Loam	7.3
D03-P2-L2C	D03	2	15-30	1.62	0.00	1.62	16.55	5.35	3.74	0.561	0.2	34.7	50.9	14.4	Silt loam	7.5
D03-P2-L3C			30-45	1.03	0.00	1.03	16.63	2.49	2.06	0.282	3.5	35.2	53.6	11.2	Silt loam	7.5
D03-P3-L1C			0-15	2.88	0.112	2.77	18.96	3.85	11.1	2.53	3.6	32.3	49.1	18.6	Loam	7.3
D03-P3-L2C		3	15-30	2.03	0.245	1.79	18.65	2.72	6.42	0.523	1.1	30.4	55.1	14.5	Silt loam	7.6
D03-P3-L3C			30-45	1.73	0.134	1.60	18.58	3.79	3.54	0.263	7.6	36.7	51.3	12.0	Silt loam	7.8
D07-P1-L1C			0-15	1.13	0.397	0.733	7.04	3.32	3.97	0.801	4.3	84.4	9.9	5.8	Loamy fine sand	8.0
D07-P1-L2C		1	15-30	2.10	1.88	0.220	4.85	1.42	1.38	0.0330	0.0	93.1	4.0	2.8	Fine sand	8.1
D07-P1-L3C			30-45	4.02	3.84	0.180	4.25	1.49	1.39	0.0350	0.0	95.5	3.7	0.8	Fine sand	8.3
D07-P2-L1C			0-15	1.61	0.340	1.27	12.50	7.09	5.59	1.41	0.2	70.4	21.4	8.1	Fine sandy loam	7.4
D07-P2-L2C	D07	2	15-30	1.12	0.132	0.988	11.57	4.94	2.16	0.688	8.8	67.2	23.6	9.3	Fine sandy loam	7.6
D07-P2-L3C			30-45	1.71	1.16	0.550	9.81	2.99	1.85	0.0820	1.6	73.3	19.6	7.1	Fine sandy loam	7.8
D07-P3-L1C			0-15	1.69	0.00	1.69	10.26	6.53	6.82	0.590	1.6	70.6	19.7	9.7	Fine sandy loam	7.4
D07-P3-L2C		3	15-30	1.44	0.832	0.608	7.74	3.95	2.08	0.0730	4.9	78.6	15.6	5.8	Loamy fine sand	7.9
D07-P3-L3C			30-45	2.15	2.10	<0.1	7.89	2.55	1.52	<0.03	4.5	81.7	11.8	6.5	Loamy fine sand	7.9
D10-P1-L1C			0-15	2.41	0.280	2.13	17.61	2.82	7.21	0.573	1.4	40.3	47.0	12.7	Loam	7.6
D10-P1-L2C		1	15-30	2.19	1.04	1.15	16.69	3.42	1.86	0.100	5.4	47.5	41.3	11.2	Loam	7.9
D10-P1-L3C			30-45	3.93	3.81	0.120	9.06	2.01	1.41	<0.03	17.4	71.4	24.0	4.6	Sandy loam	8.1
D10-P2-L1C			0-15	2.65	0.389	2.26	14.80	1.56	12.3	<0.03	0.9	42.0	41.2	16.8	Loam	7.5
D10-P2-L2C	D10	2	15-30	1.47	0.530	0.940	14.96	6.01	3.20	0.113	1.1	42.7	43.7	13.6	Loam	7.8
D10-P2-L3C			30-45	2.83	2.25	0.580	11.56	1.49	1.69	<0.03	10.9	61.4	31.4	7.2	Fine sandy loam	8.0
D10-P3-L1C			0-15	2.78	0.685	2.10	15.53	3.09	8.42	0.974	0.7	38.7	43.5	17.8	Loam	7.5
D10-P3-L2C		3	15-30	2.06	0.392	1.67	15.94	3.72	3.49	0.0830	1.2	36.6	50.1	13.3	Silt loam	7.9
D10-P3-L3C			30-45	3.18	2.39	0.790	10.28	1.50	1.89	0.0350	9.4	57.0	34.1	8.9	Fine sandy loam	8.1
Alvars	Al	1	0-20	7.51	1.74	5.77	22.89	24.9	26.5	0.805	8.3	36.9	46.7	16.4	Loam	7.5

[a]Total C; [b]Soil Inorganic C; [c]Soil Organic C; [d]Gravel;*See section 2.1 for description

Table 15: Soil horizon characterization for each soil sample across ***rare*** Charitable Research Reserve. Horizon classification was made using a combination of soil physicochemical results, visual analysis of the soil profile, information from a previous survey of soils across the ***rare*** Charitable Research Reserve (181), as well as guidelines from the Canadian system of soil classification (10).

Sample ID[a]	Site ID	Depth (cm)	Plot	Horizon(s)
IWP1H1L1	IW	15	1	Ah/Ae
IWP1H1L2	IW	30	1	Ae
IWP1H1L3	IW	45	1	Ae/Bt
IWP1H2L1	IW	15	1	Ah/Ae
IWP1H2L2	IW	30	1	Ae
IWP1H2L3	IW	45	1	Ae/Bt
IWP1H3L1	IW	15	1	Ah/Ae
IWP1H3L2	IW	30	1	Ae
IWP1H3L3	IW	45	1	Ae/Bt
IWP2H1L1	IW	15	2	Ah
IWP2H1L2	IW	30	2	Ae/Bt
IWP2H1L3	IW	45	2	Bt/Ck
IWP2H2L1	IW	15	2	Ah/Ae
IWP2H2L2	IW	30	2	Ae/Bt
IWP2H2L3	IW	45	2	Bt/Ck
IWP2H3L1	IW	15	2	Ah/Ae
IWP2H3L2	IW	30	2	Ae/Bt
IWP2H3L3	IW	45	2	Bt/Ck
IWP3H1L1	IW	15	3	Ah/Ae
IWP3H1L2	IW	30	3	Ae/Bt
IWP3H1L3	IW	45	3	Bt/Ck
IWP3H2L1	IW	15	3	Ah/Ae
IWP3H2L2	IW	30	3	Ae/Bt
IWP3H2L3	IW	45	3	Bt/Ck
IWP3H3L1	IW	15	3	Ah/Ae
IWP3H3L2	IW	30	3	AeBt
IWP3H3L3	IW	45	3	Bt/Ck
HP1H1L1	H	15	1	Ah/Ae
HP1H1L2	H	30	1	Ae/Bt
HP1H1L3	H	45	1	Bt/Ck
HP1H2L1	H	15	1	Ah/Ae
HP1H2L2	H	30	1	Ae/Bt
HP1H2L3	H	45	1	Bt/Ck
HP1H3L1	H	15	1	Ah
HP1H3L2	H	30	1	Ah/Ae
HP1H3L3	H	45	1	Ae/Bt
HP2H1L1	H	15	2	Ah/Ae

Sample ID[a]	Site ID	Depth (cm)	Plot	Horizon(s)
HP2H1L2	H	30	2	Ae/Bt
HP2H1L3	H	45	2	Bt/Ck
HP2H2L1	H	15	2	Ah/Ae
HP2H2L2	H	30	2	Ae/Bt
HP2H2L3	H	45	2	Bt/Ck
HP2H3L1	H	15	2	Ah
HP2H3L2	H	30	2	Ah/Bt
HP2H3L3	H	45	2	Bt/Ck
HP3H1L1	H	15	3	Ah/Ae
HP3H1L2	H	30	3	Ae
HP3H1L3	H	45	3	Bt
HP3H2L1	H	15	3	Ah/Ae
HP3H2L2	H	30	3	Ae
HP3H2L3	H	45	3	Bt
HP3H3L1	H	15	3	Ah/Ae
HP3H3L2	H	30	3	Ae
HP3H3L3	H	45	3	Bt
CAP1H1L1	CA	15	1	Ah
CAP1H1L2	CA	30	1	Ah/Bm
CAP1H1L3	CA	45	1	Bm/Ck
CAP1H2L1	CA	15	1	Ah/Bm
CAP1H2L2	CA	30	1	Bm
CAP1H2L3	CA	45	1	Bm/Ck
CAP1H3L1	CA	15	1	Ah
CAP1H3L2	CA	30	1	Ae/Bm
CAP1H3L3	CA	45	1	Bm/Ck
CAP2H1L1	CA	15	2	Ah/Ae
CAP2H1L2	CA	30	2	Ae/Bm
CAP2H1L3	CA	45	2	Bm
CAP2H2L1	CA	15	2	Ah/Ae
CAP2H2L2	CA	30	2	Ae/Bm
CAP2H2L3	CA	45	2	Bm
CAP2H3L1	CA	15	2	Ah/Ae
CAP2H3L2	CA	30	2	Ae/Bm
CAP2H3L3	CA	45	2	Bm
CAP3H1L1	CA	15	3	Ah/E/Ae
CAP3H1L2	CA	30	3	Ae
CAP3H1L3	CA	45	3	Ae
CAP3H2L1	CA	15	3	Ah/E/Ae
CAP3H2L2	CA	30	3	EAe
CAP3H2L3	CA	45	3	Ae
CAP3H3L1	CA	15	3	Ah/E/Ae

Sample ID[a]	Site ID	Depth (cm)	Plot	Horizon(s)
CAP3H3L2	CA	30	3	Ae
CAP3H3L3	CA	45	3	Ae
AAP1H1L1	AA	15	1	Ap
AAP1H1L2	AA	30	1	Bt/Ck
AAP1H2L1	AA	15	1	Ap/Bt
AAP1H2L2	AA	30	1	Bt/Ck
AAP1H3L1	AA	15	1	Ap/Bt
AAP1H3L2	AA	30	1	Bt/Ck
AAP2H1L1	AA	15	2	Ap/Bt
AAP2H1L2	AA	30	2	Bt/Ck
AAP2H1L3	AA	45	2	Bt/Ck
AAP2H2L1	AA	15	2	Ap
AAP2H2L2	AA	30	2	Ap/Bt
AAP2H2L3	AA	45	2	Bt/Ck
AAP2H3L1	AA	15	2	Ap/Ae
AAP2H3L2	AA	30	2	Ae/Bt
AAP2H3L3	AA	45	2	Ae/Bt
AAP3H1L1	AA	15	3	Ap/Ae
AAP3H1L2	AA	30	3	Ae/Bt
AAP3H1L3	AA	45	3	Ae/Bt
AAP3H2L1	AA	15	3	Ap
AAP3H2L2	AA	30	3	Ap/Ae
AAP3H2L3	AA	45	3	Ae
AAP3H3L1	AA	15	3	Ap
AAP3H3L2	AA	30	3	Ap/Ae
D03P1H1L1	D03	15	1	Ah
D03P1H1L2	D03	30	1	Ah
D03P1H1L3	D03	45	1	Ah
D03P1H2L1	D03	15	1	Ah
D03P1H2L2	D03	30	1	Ah
D03P1H2L3	D03	45	1	Ah
D03P1H3L1	D03	15	1	Ah
D03P1H3L2	D03	30	1	Ah/Bt
D03P1H3L3	D03	45	1	Bt
D03P2H1L1	D03	15	2	Ah
D03P2H1L2	D03	30	2	Ah
D03P2H1L3	D03	45	2	Ah/Bt
D03P2H2L1	D03	15	2	Ah
D03P2H2L2	D03	30	2	Ah
D03P2H2L3	D03	45	2	Ah/Bt
D03P2H3L1	D03	15	2	Ah
D03P2H3L2	D03	30	2	Ah

Sample ID[a]	Site ID	Depth (cm)	Plot	Horizon(s)
D03P2H3L3	D03	45	2	Ah/Bt
D03P3H1L1	D03	15	3	Ah
D03P3H1L2	D03	30	3	Ah/Bt
D03P3H1L3	D03	45	3	Bt
D03P3H2L1	D03	15	3	Ah
D03P3H2L2	D03	30	3	Ah/Bt
D03P3H2L3	D03	45	3	Bt
D03P3H3L1	D03	15	3	Ah
D03P3H3L2	D03	30	3	Ah/Bt
D03P3H3L3	D03	45	3	Bt
D07P1H1L1	D07	15	1	Ah/Ae
D07P1H1L2	D07	30	1	Ae
D07P1H1L3	D07	45	1	Ae/Bt
D07P1H2L1	D07	15	1	Ah/Ae
D07P1H2L2	D07	30	1	Ae/Bt
D07P1H2L3	D07	45	1	Bt
D07P1H3L1	D07	15	1	Ah
D07P1H3L2	D07	30	1	Ae
D07P2H1L1	D07	15	2	Ah
D07P2H1L2	D07	30	2	AhBt
D07P2H1L3	D07	45	2	Bt
D07P2H2L1	D07	15	2	Ah
D07P2H2L2	D07	30	2	Ah/Bt
D07P2H2L3	D07	45	2	Bt
D07P2H3L1	D07	15	2	Ah
D07P2H3L2	D07	30	2	Ah/Bt
D07P2H3L3	D07	45	2	Bt
D07P3H1L1	D07	15	3	Ah
D07P3H1L2	D07	30	3	Ah/Bt
D07P3H1L3	D07	45	3	Bt
D07P3H2L1	D07	15	3	Ah
D07P3H2L2	D07	30	3	Ah/Bt
D07P3H2L3	D07	45	3	Bt
D07P3H3L1	D07	15	3	Ah
D07P3H3L2	D07	30	3	Ah/Bt
D07P3H3L3	D07	45	3	Bt
D10P1H1L1	D10	15	1	Ah/Bt
D10P1H1L2	D10	30	1	Bt/Ck
D10P1H2L1	D10	15	1	Ah/Ae
D10P1H2L2	D10	30	1	Ae/Bt
D10P1H2L3	D10	45	1	Bt/Ck
D10P1H3L1	D10	15	1	Ah

Sample ID[a]	Site ID	Depth (cm)	Plot	Horizon(s)
D10P1H3L2	D10	30	1	Bt
D10P1H3L3	D10	45	1	Ck
D10P2H1L1	D10	15	2	Ah
D10P2H1L2	D10	30	2	Bt
D10P2H1L3	D10	45	2	Bt/Ck
D10P2H2L1	D10	15	2	Ah
D10P2H2L2	D10	30	2	Bt
D10P2H2L3	D10	45	2	Bt/Ck
D10P2H3L1	D10	15	2	Ah
D10P2H3L2	D10	30	2	Bt
D10P2H3L3	D10	45	2	Bt/Ck
D10P3H1L1	D10	15	3	Ah
D10P3H1L2	D10	30	3	Ah/Bt
D10P3H1L3	D10	45	3	Bt/Ck
D10P3H2L1	D10	15	3	Ah
D10P3H2L2	D10	30	3	Ah/Bt
D10P3H2L3	D10	45	3	Bt/Ck
D10P3H3L1	D10	15	3	Ah
D10P3H3L2	D10	30	3	Bt
D10P3H3L3	D10	45	3	Ck
AlP1H1L1	Al	10	1	Ah
AlP1H2L1	Al	10	1	Ah
AlP1H2L2	Al	20	1	Ah
AlP1H3L1	Al	10	1	Ah

[a]Sample IDs are presented as a code consisting of 7–9 digits where the first 1–3 characters indicate site ID (see section 2.1 for details); the number after "P" refers to the subplot number; the number after "H" refers to the pit number; the numbers after "L" refers to the depth increment where 1, 2, and 3 refers to the 0–15, 15–30, and 30–45 depth increment, respectively.

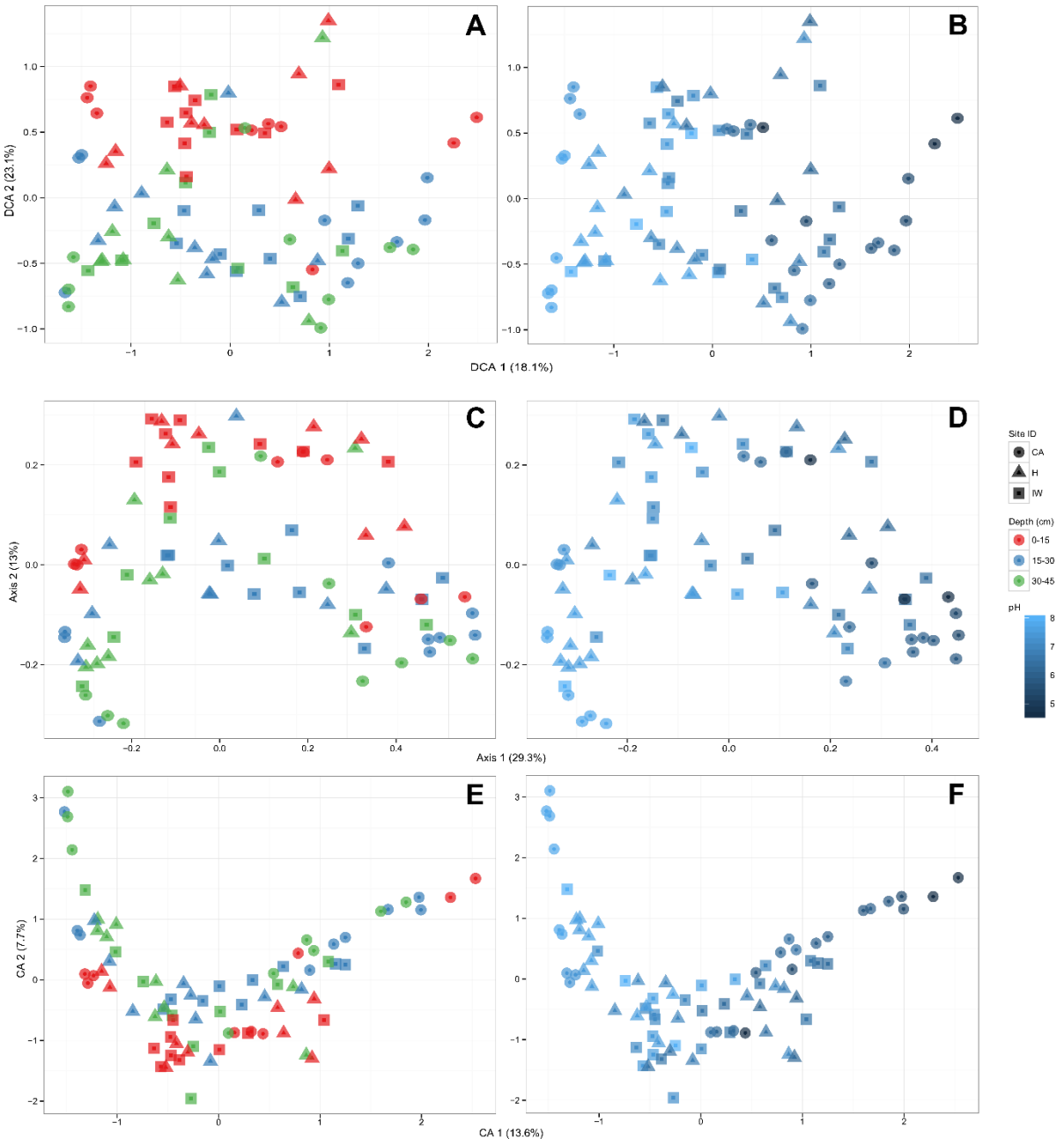

Figure 28: DCA for forest sites showing the reduction of the "arch effect" coloured by A) depth and B) pH. For comparison, panels C) and D) shows the original PCoA and panels E) and F) show the correspondence analysis, an intermediate step in DCA.