DOCTOR OF PHILOSOPHY

Responses of soil biological and functional diversity to anthropogenic change from plot- to national-scales

George, Paul

Award date:
2020

Awarding institution:
Bangor University

DOCTOR OF PHILOSOPHY

Responses of soil biological and functional diversity to anthropogenic change from plot- to national-scales

George, Paul

Award date:
2020

Awarding institution:
Bangor University

Responses of soil biological and functional diversity to anthropogenic change from plot- to national-scales

Paul B. L. George

A thesis submitted to Bangor University in candidature for the degree Philosophiae Doctor

February (2020)

School of Natural Sciences,
Bangor University,
Bangor,
Gwynedd,
LL57 2UW

PRIFYSGOL BANGOR UNIVERSITY

Canolfan Ecoleg a Hydroleg y DU
UK Centre for Ecology & Hydrology

STARS
Soils Training And Research Studentships

Thesis Summary

Soil biota account for ~25% of global biodiversity and underpins a wide range of ecosystem services. Here defined as soil-dwelling bacteria, archaea, fungi, protists, animals, and viruses, soil biota have often been overlooked and generalised in biodiversity surveys and studies of ecosystem service provision. However, studying the response of the whole community is highly likely to obscure important biodiversity trends in the constituent fractions of soil biota. The aims of this thesis were to determine how belowground biodiversity and community structure are related to land use and soil physicochemical properties at the national-scale and investigate shifts of organisms with important functional roles. Also, the fate of soil communities and impacts on nutrient cycling under long-term carbon deprivation at the plot-level were investigated. To address the first aspect of the thesis, biological (invertebrate specimens and environmental DNA sequences) and environmental (soil and environmental properties) data were collected as part of the Glastir Monitoring & Evaluation Programme, an assessment of the impacts of the Glastir agri-environment scheme on soils across Wales, UK. Using this data, I showed that diversity and abundance of mesofauna, and richness of soil animals generally, from metabarcoding analyses, but supported by traditional taxonomy are reduced in arable land. I suggest therefore that mesofauna could be valuable biological indicators, due to the congruence between results obtained from morphological and molecular analyses. Metabarcoding data also revealed a trend of declining richness from high-productivity arable sites to low-productivity upland habitats shared by bacteria, fungi, and protists. Archaea showed an opposing trend. All groups were strongly influenced by pH and carbon-to-nitrogen ratio. A comparison of primer choice for fungi (ITS1 vs. 18S) revealed biases stemming from primer and database choice that influenced functional diversity but not the overarching trend in fungal richness in response to land use. Using 18S primers detected Glomeromycetes and other groups that greatly influenced functional diversity across land uses. All of these investigations determined soil type was a poor predictor of soil biota metrics. The distribution of sulphate-reducing bacteria (SRB) was also investigated. I found that richness of these bacteria was relatively constant across land uses. Concurrent analysis of common generalist anaerobic taxa followed the overarching trend across land use productivity mentioned previously. There was a shift in proportional abundance of SRB to generalist anaerobes along the productivity gradient, indicating that competitive forces may be at play, like niche separation. The results of a long-term carbon-deprivation experiment comprise the final chapter of this thesis. I found that following ten years of suspension of carbon inputs, richness of all soil microbes and viruses declined along with a range of measures of soil chemical and physical quality. Functional genes shifted to anaerobic and recalcitrant energy sources. The work has provided diverse, essential information on patterns of soil biota and the physicochemical and land use factors governing the distributions of the many fractions of soil biodiversity. The thesis is important in understanding the natural history of Welsh and temperate soil biota in general. It also provides an important framework for future analyses and projections of the response of soil biota, associated function and ecosystem services in the context of predicted environmental change.

Acknowledgments

I must thank my supervisory team of Davey Jones, David Robinson, and Si Creer for all their assistance over the course of this endeavour. David, I will always remember your warm welcome on my first day. Thank you for your support especially in navigating the logistics of getting established in a new country. Si, thank you for all your help and encouragement as I entered the world of molecular ecology. Davey, your constant support and guidance has made my time at Bangor an amazing experience. You challenged me to reach goals I did not think I could attain and have pushed me to be a better scientist.

I would also like to thank Aidan Keith and Rob Griffiths from CEH, Lancaster, for their patience, support, and collaboration with me throughout my PhD. Thanks also go to the members of ECW past and present for their friendship, understanding, and help throughout this journey. Special thanks go to Fiona Seaton for all her help throughout the course of our PhD journeys. Thanks also to all the other members of STARS for your friendship and the trips abroad.

Mae rygbi yn rhan fawr o fy mywyd i, mae bod yn awlod o dîm gyda Clwb Rygbi Bethesda wedi bod yn ardderchog. Rydw i yn gwario fy amser ym Mhethesda yn chwarae rygbi, dysgu Cymraeg, ac hefyd yn gwneud ffrindiau gyda phobl newydd. Mae fy amser yng Nghymru wedi bod yn mwy na dim ond prifysgol. Diolch yn fawr hogia.

Finally, thank you Mum, Dad, and Ross, for encouraging and supporting me in my decision to pursue this amazing opportunity. It's not always easy being so far from home but knowing you're cheering me on makes a world of difference. Thank you so much and see you soon.

Table of Contents

Abbreviations

°C – Degrees Celsius

β-diversity – Beta diversity

µL – Microliter(s)

Al – Aluminium

AMF – Arbuscular mycorrhizal fungi

AOA – Ammonia oxidising archaea

AOB – Ammonia oxidising bacteria

ANOVA – Analysis of variance

AIC – Akaike information criterion

ASV – Amplicon sequence variant

AVC – Aggregate vegetation class

C – Carbon

Ca – Calcium

$CaCl_2$ – Calcium chloride

$CaCO_3$ – Calcium carbonate

CAP – Canonical analysis of principal coordinates

CEC – Cation exchange capacity

CEH – Centre for Ecology & Hydrology

cm – Centimetre(s)

CS – Countryside Survey

DNA – Deoxyribonucleic acid

eDNA – Environmental DNA

EIDC – Environmental information data centre

EMF – Ectomycorrhizal fungi

ERAMMP – Environment and Rural Affairs Monitoring & Modelling Programme

EU – European Union

g - Gram

GMEP – Glastir Monitoring & Evaluation Programme

H' – Shannon-Weiner diversity index

H_2S – Hydrogen sulphide

Ha – Hectare(s)

IDH – Intermediate disturbance hypothesis

ITS – Internal transcribed spacer

km – Kilometre(s)

LOI – Loss-on-ignition

m – Metre(s)

mg – milligram(s)

mL – Millilitre(s)

mm – Millimetre(s)

min – Minute(s)

N – Nitrogen

N_2 – Atmospheric nitrogen

NERC – National Environment Research Council

NH_3 – Ammonia

NH_4^+ – Ammonium

NO_3^- – Nitrate

NMDS – Non-metric multidimentional scaling

OTU – Operational taxonomic unit

P – Phosphorus

PCR – Polymerase chain reaction

PERMANOVA – Permutational multivariate analysis of variance

PLFA – Phospholipid fatty acid

PLS – Partial least squares

qPCR – Quantitative polymerase chain reaction

rDNA – Ribosomal deoxyribonucleic acid

rRNA – Ribosomal ribonucleic acid

S – Sulphur

s – Second(s)

SCW – Supercomputing Wales

SO_x – Sulphates

SRB – Sulphate-reducing bacteria

SIMPER – Similarity percentages

SOM – Soil organic matter

SWR – Soil water repellency

UK – United Kingdom

VIP – Variable importance in projection

VWC – Volumetric water content

List of Tables

List of Figures

List of Articles

This thesis is based on the following articles:

I. Paul B. L. George, Aidan M. Keith, Simon Creer, Gaynor L. Barrett, Inma Lebron, Bridget A. Emmett, David A. Robinson, David L. Jones (2017) Evaluation of mesofauna as soil quality indicators in a national-level monitoring programme. *Soil Biology & Biochemistry* **115**: 537-546.

II. Paul B. L. George, Delphine Lallais, Simon Creer, Fiona M. Seaton, Robert I. Griffiths, John G. Kenny, Richard M. Eccles, Inma Lebron, Bridget A. Emmett, David A. Robinson, David L. Jones (2019) Divergent national-scale trends of microbial and animal biodiversity across diverse temperate soil ecosystems. *Nature Communications* **10**:1107. DOI: 10.1038/s41467-019-09031-1

III. Paul B. L. George, Simon Creer, Robert I. Griffiths, Bridget A. Emmett, David A. Robinson, David L. Jones (2019) Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey. *Frontiers in Environmental Science* **7**:173. DOI: 10.3389/fens.201900173.

IV. Paul. B. L. George, Katia P. Coelho, Simon Creer, Inma Lebron, Bridet A. Emmett, David A. Robinson, David L. Jones (2019) Inverse relationship between generalist anaerobes and sulphate-reducing bacteria is driven by pH across land uses in temperate soils. (*In Review*)

V. Paul B. L. George, David B. Fidler, Jonathan A. Atkinson, Sacha J. Mooney, Joy Van Nostrand, Simon Creer, Robert I. Griffiths, James MacDonald, David A. Robinson, Davey L. Jones (2019) Shifts in soil microbial biological and functional diversity under long-term carbon depletion. (*In Preparation*)

I also contributed to the following paper, which is referred to within this thesis:

I. Fiona M. Seaton, David L. Jones, Simon Creer, **Paul B. L. George**, Simon M. Smart, Inma Lebron, Gaynor L. Barrett, Bridget A. Emmett, David A. Robinson (2019) Plant and soil communities are associated with the response of soil water repellency to environmental stress. *Science of the Total Environment* **687**, 929-938.

Chapter 1: Introduction

1.1 General introduction

Soils are central to the delivery of a wide range of ecosystem services that are necessary for earth system functioning. In addition, nearly a quarter of Earth's biodiversity resides within soil habitats. Soil biota include representatives of nearly every major evolutionary lineage, including: bacteria, archaea, fungi and protists, as well as animals and viruses (FAO & ITPS, 2015). Famously, soils have been dubbed "the poor man's tropical rainforest" (Usher *et al*., 1979) owing to the fact that soils represent the most species-rich terrestrial habitats (Wall *et al*., 2005). Soil biota perform a wide range of roles within soil trophic networks. The majority of soil biota are involved in the decomposition of soil organic matter (SOM). Some organisms, bacteria in particular, are integral to biogeochemical cycles (e.g. carbon (C), nitrogen (N), iron (Fe), phosphorus (P), and sulphur (S) cycling). Others, such as mycorrhizal fungi and N_2 fixing bacteria, form integral symbiotic relationships with plants. A large fraction of soil biota also parasitise plants, which are of serious concern in agricultural management. Soils are also a reservoir for human and livestock pathogens (Coleman *et al.,* 2018) as well as biological compounds that comprise modern pharmaceuticals (Robinson *et al*., 2014).

The small size and extreme biodiversity of soil biota, has made characterising soil communities a historically difficult task. Techniques such as morphological identification or culturing of soil microbes have been staples of quantifying soil biodiversity. However, such techniques cannot reveal a complete picture of

belowground biodiversity. Taxonomic identification of soil organisms requires much expertise (i.e. nematodes, Chen *et al*., 2010) and only a relatively small fraction of soil microbes can be cultured in the laboratory (Islam & Wright, 2006). Yet, modern molecular-based techniques have overcome these impediments and sparked a step-change in the characterisation of all components of belowground communities (George *et al*., 2019 and examples therein). Attempts to quantify and characterise soil biota at the global scale are now being published (Tedersoo *et al*., 2014; Delgao-Baquerizo *et al*., 2018) and are greatly improving our understanding of Earth's total biodiversity (Cameron *et al*., 2019).

In recent years, the prominent role of soils in providing ecosystem services has become increasingly recognised (Ford *et al*., 2012; Robinson *et al*., 2014). Soil is responsible for the provision of a vast array of ecosystem services including the maintenance of water and nutrient cycles, gas exchange, plant growth and thereby crop and fibre production (Barrios, 2007; Robinson *et al*., 2014). Soil biota are directly involved in the delivery of many ecosystem services, such as nutrient cycling and indirectly like the medical use of antibiotic compounds derived from soil organisms (Robinson *et al*., 2014). This is most apparent in the use of agri-environment schemes to meet policy goals. Such recognition of the importance of soil natural capital is essential to ensure sustainable development in the face of climate change and other anthropogenic perturbations.

However, soil biota face mounting threats from anthropogenic pressures. Land use change can cause significant changes to soil communities and so impact the ecosystem services they provide. A shift to more intensive agriculture, for instance,

is known to reduce species richness of earthworms, certain mites, and Collembola as well as overall food web complexity (Tsiafouli *et al*., 2015). In addition, the interactions between soil biota and their environment are so complex that the underlying mechanisms governing ecosystem functions are not wholly understood, leading many to refer to soils as a "black box" (Cortois & De Deyn, 2012). This, combined with the lack of inclusion of soil biota in biodiversity conservation initiatives (Cameron *et al*., 2019) has put great impetus on integrating the identification of soil biota and maintaining the ecosystem services they provide.

Therefore, this thesis will explore the interactions of soil biodiversity with soil physicochemical and environmental properties as well as land use at the national scale. In addition, research at the field-scale will highlight the effects of extreme C-deprivation on biodiversity and provision of soil ecosystem services, namely, nutrient cycling.

1.2 Outline of thesis

This thesis is comprised of a further 7 chapters, beginning with a literature review that describes the various components of soil biota, their interactions with soil physicochemical properties, roles in the provision of ecosystem services, and how such services and biodiversity are generally affected by land use change. The following data chapters (Chapter 3-7) are presented as scientific articles. Each chapter contains the authorship details and contributions in addition to their publishing status as presented in the List of Articles (page xi).

The first 4 data chapters (Chapter 3-6) present findings from work on the Welsh Government Glastir Monitoring & Evaluation Programme (GMEP) dataset. As such,

there is an unavoidable repetition of certain introductory materials, methodology, and references. Chapters 3 and 4 outline the composition of mesofauna (Chapter 3) and total soil biota (Chapter 4) across Wales, and their relationships with land use, as well as environmental and physicochemical properties. Chapter 5 explores how methodological choices may have impacted the results presented in Chapter 4. This chapter focuses on fungi and how findings of richness and functional groups differ based on the primers chosen to construct amplicon libraries for metabarcoding. Chapter 6 highlights the distribution and prevalence of bacterial groups that are integral to nutrient cycling within the previously studied sites. The chapter focuses on sulphur-reducing bacteria (SRB) across Wales. Particular attention is paid to the differences in SRB and generalist anaerobe populations and how they compare land uses. Chapter 7 describes the results of a decade-long experiment investigating the changes in belowground communities and the subsequent impacts on nutrient cycles (particularly C, N, S, and P) in sites where nutrient inputs to soils have been prevented.

A general discussion is provided in Chapter 8. Here, the results of all data chapters are presented in the context of the aims and objectives of this thesis. It will also present the main conclusions and highlight areas for future research. It is important to note that Chapters 3-6 represent only part of the soils component within GMEP; other research projects rely on data generated as part of this thesis. These findings will be introduced where appropriate as part of the general discussion. Finally, the appendices include supplementary material from those data chapters that have been published in scientific journals, as well as, further information pertaining to the experiment described in Chapter 7, and a co-authored paper that uses data generated in Chapter 4.

1.3 Thesis aims and objectives

Broadly, the aim of this thesis is to determine the importance of soil biodiversity across a wide range of different soil types and land uses. Specific objectives of this work include:

1) To reveal shifts in the diversity and community structure of soil biota and determine how they relate to soil physical properties and heterogeneous land uses at the national-scale (Chapter 3-5).

2) To investigate the potential effects of changing land uses on organisms important to the provision of ecosystem services, *in situ* (Chapter 5-6).

3) To reveal the fate of soil communities under long-term stress and the consequences for nutrient cycling (Chapter 7).

1.4 References

Barrios, E., 2007. Soil biota, ecosystem services and land productivity. Ecol. Econom. 64, 269-285.

Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., Gottschall, F. et al., 2019. Global mismatches in aboveground and belowground biodiversity. Conserv. Biol. 33, 1187-1192.

Chen, D., Wang, Y., Lan, Z., Li, J., Xing, W., Hu, S., Bai, Y., 2015. Biotic community shifts explain the contrasting responses of microbial and root respiration to experimental soil acidification. Soil Biol. Biochem. 90, 139-147.

Coleman, D.C., Callahan, M., Crossley, D. Jr., 2018. Fundamentals of Soil Ecology, third ed. Academic Press, London.

Cortois, R., De Deyn, G.B., 2012. The curse of the black box. Plant Soil 350, 27-33.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., et al., 2018. A global atlas of dominant bacteria found in soil. Science 359, 320-325.

FAO & ITPS, 2015. Status of the World's Soil Resources (SWSR) – Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils. Rome, Italy.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I. et al., 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soils ecosystems. Nat. Commun. 10, 1107 doi: 10.1038/s41467-019-09031-1

Islam, R., Wright, S.W., 2006. Microbial biomass measurement methods. In: Lal, R. (Ed.), Encyclopedia of Soil Science, second ed. Talyor and Francis, New York, pp. 1067-1070.

Koch, A., McBratney, A., Lal, R., 2012. Put soil security on the global agenda. Nature 492, 186.

Robinson, D.A., Fraser, I., Dominati, E.J., Davi∂sdóttir, B., Jónsson, J.O.G., Jones, L., Jones, S.B. et al., 2014. On the value of soil resources in the context of natural capital and ecosystem service delivery. Soil Sci. Soc. Am. J. 78, 685-700.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Villarreal Ruiz, L. et al., 2014. Global diversity and geography of soil fungi. Science 346, 1256688.

Tsiafouli, M.A., Thébault, E., Sgardelis, S.P., de Ruiter, P.C., van der Putten, W.H., Birkhofer, K., Hemerik, L. et al., 2015. Intensive agriculture reduces soil biodiversity across Europe. Global Change Biol. 21, 973-985.

Usher, M.B., Davis, P., Harris, J., Longstaff, B., 1979. A profusion of species? Approaches towards understanding the dynamics of the populations of microarthropods in decomposer communities. In: Anderson, R.M., Turner, B.D., Taylor, L.R. (Eds.), Population Dynamics. Blackwell Scientific Publications, Oxford, pp. 359-384.

Wall, D.H., Bardgett, R.D., Kelly, E.F., 2010. Biodiversity in the dark. Nat. Geosci. 3, 297-298.

Chapter 2: Literature Review

2.1 Introduction to soil biota

Approximately 25% of Earth's biodiversity lives within the soil (FAO & ITPS, 2015). It has been estimated that 1 g of soil can support 1 x 10^{10} bacterial cells from up to 50,000 operational taxonomic units (OTUs) (Raynaud & Nunan, 2014). That same gram of soil may also support between $10^5 – 10^6$ fungi (Maier & Pepper, 2000), more than 10^4 protists (Adl & Coleman, 2005; Geisen *et al*., 2014), and hundreds of animals (Song *et al*., 2017). It is now known that soils support a much higher number of archaea than expected, even in non-extreme environments (Timonen & Bomberg, 2009). Viral diversity in soils is estimated to exceed the total species counts (Fierer *et al*., 2007); this is true of many other organismal groups as well (Barrios, 2007; Fierer *et al*., 2007). It is therefore no surprise that Usher *et al*. (1979) dubbed soils "the poor man's tropical rainforest". Accordingly, soil biodiversity is often referred to as a "black box" which is slowly being illuminated as our understanding of soil biota develops.

Aboveground communities ultimately rely on soils as a substrate, within which soil biota play an essential role in maintaining terrestrial ecosystems through processes including decomposition, nutrient cycling, trophic energy transfer, and as pathogens (Barrios, 2007; Bardgett & van der Putten, 2014; Robinson *et al*., 2014). Unlike aboveground food webs, which are based on a single energy pathway, soil trophic networks include multiple energy pathways including the decomposition of organic matter by both bacteria and fungi (Murray *et al.,* 2009; Crotty *et al.,* 2011) as well as a distinct parasitic pathway wherein, parasites infect or directly consume plant roots and tubers (Bongers & Bongers, 1998; Bardgett & van der Putten, 2014). This has led to the

proposal of complex branching interactions between these channels at higher trophic levels, creating intricate and complicated networks (Crotty *et al.,* 2011).

Concerted efforts to disentangle soil trophic interactions have historically focused on feeding preference. This has been measured either directly through prey choice (Jonas *et al*., 2007), or indirectly through tracing digestive enzymes (Berg *et al*., 2004) or fatty acids (Ruess & Chamberlain, 2010). More recent advances have led to the use of stable isotope analyses to trace the bacterial energy channel (Crotty *et al*., 2011; 2012; Ruess & Chamberlain, 2010). These studies have continually shown that the application of traditional trophic dynamics oversimplifies soil systems. The following section will briefly outline the major fractions of soil biota and discuss their interactions within soil trophic networks, as well as important processes in which they are involved.

2.1.1 Soil biodiversity

Viruses are important but massively overlooked regulators of soil communities (Kimura *et al*., 2008). Although they will not be a focus of this thesis, it is important to acknowledge their presence in soil as our knowledge of viral diversity in soils is severely lacking. Assessments of viral diversity in soils have found that true viral diversity vastly outnumbers current species counts (Fierer *et al*., 2007). Soils are a reservoir for an incredible number of inactive viruses – up to 1.2×10^9 viral particles per gram – many of which are important pathogens of bacteria and fungi known as bacteriophages and mycophages respectively (Swanson *et al*., 2009). The relationship between bacteria and bacteriophages is of growing interest due to the important role bacteria play in providing ecosystem services (Vos *et al*., 2009) and as a genomic reservoir (Kimura *et al*., 2008). Research into the coevolution of these groups has shown that this soil "arms race" has the

potential to continue indefinitely, with higher degrees of resistance in bacteria to current strains of bacteriophages, than their past and future iterations (Gómez & Buckling, 2011). Such interaction highlights the important regulatory effect of viruses in the ultimate community composition of soils. However, despite their importance in community dynamics, viruses in soils have been consistently ignored in major soil community composition analyses.

Prokaryotes (bacteria and archaea) are the most common and diverse group of organisms found in soils. Indeed, estimates of bacterial diversity range from 1×10^3 up to 1×10^6 unique strains or species per single gram of soil (Fierer *et al*., 2007). Although both of these groups are prolific in soils, estimates of biodiversity are poor due to difficulties isolating and culturing species (Daniel, 2004). These organisms play essential roles in ecosystem function. Soil prokaryotes includes largely benign species that breakdown detritus as well as important symbionts and pathogens of crops, livestock, and humans (Mendes *et al*., 2013).

Prokaryotes are important primary decomposers of organic matter in soils. They can quickly colonise litter and dead organisms, secreting digestive enzymes and consuming the resulting nutrients through osmotrophy, whereby nutrients in water are then taken up through simple diffusion of small molecules as well as facilitated diffusion and active transport of larger ones (Adl, 2003). However, due to physiological limitations, such digestive enzymes are specialised, meaning many prokaryotes exhibit a degree of substrate preference. The use of a suboptimal substrate may be necessary in some cases, but switching substrates can be complicated, requiring environmental cofactors or the production of different digestive enzymes (Adl, 2003).

Certain prokaryotes are essential components of biochemical cycling. For example, many prokaryotes are involved in N-cycling. In particular, ammonia oxidising archaea and bacterial (AOA and AOB) are nitrifiers, meaning that they convert ammonium (NH_4^+) or ammonia (NH_3) to nitrate (NO_3^-) and are important in the degradation of nitrogenous wastes in natural systems (Banning *et al*., 2015). Elsewhere in the N cycle, Rhizobia, a paraphyletic group of Alpha- and Betaproteobacteria, have a well-documented symbiotic relationship with legumes (Moulin *et al*., 2001). Rhizobia are capable of performing N fixation, taking atmospheric N (N_2), which is inaccessible to most organisms, and converting it to forms accessible to the plant, such as NH_3, and exchanging this with their host plant for C (Ratcliff *et al*., 2008). The S cycle is also strongly influenced by bacteria. Sulphates are reduced to sulphates (H_2S) by sulphate reducing bacteria (SRB), which can be utilised by other microbes, incorporated into mineral formation, or released into the atmosphere (Muyzer & Stams, 2008).

Soil prokaryotes comprise the primary level of the bacterial energy channel. This trophic channel is best understood through the flow of energy from saprotrophic bacteria to higher-level consumers (Adl, 2003). This energy channel has a high turnover rate (Moore *et al*., 2005). Pathogenic prokaryotes can be considered as indirect contributors to this trophic network as their prevalence leads to increased host mortality rates and therefore more substrate for saprotrophs (Adl, 2003).

Soil fungi are most often the dominant eukaryotic life found in soil samples (Peay *et al*., 2008) and perform essential roles in soil trophic networks as decomposers, plant symbionts, pathogens, and predators (Adl, 2003). Soil fungi include unicellular forms, microscopic filaments as well as large conspicuous forms with complex reproductive

strategies (Peay *et al*., 2008). Reproduction in fungi can occur via simple asexual processes or through complex sexual strategies involving multiple stages and mating types (Kück & Pöggeler, 2009). Such complex life histories, in addition to their relatively simple morphology, have led to difficulties in delineating individual fungal species (Harrington & Rizzo, 1999).

Fungi perform two major functions in soils. Firstly, fungi are important saprotrophs. Like saprotrophic bacteria, they colonise and consume organic matter within the soil by secreting large amounts of digestive proteins and consuming nutrients through osmotrophy. However, fungi are able to simultaneously produce multiple digestive enzymes in large amounts to decompose complex plant polymers (e.g. cellulose, protein, lignin), providing access to a greater range of soil organic matter (SOM) (Adl, 2003). Conversely, some fungi, known as mycorrhizal fungi, are important symbionts of plants. Mycorrhizal fungi form important associations with plant roots and provide plants better access to water and facilitate mineral uptake in exchange for immediate and continuous access to carbohydrates within the plant that are produced through photosynthesis (Wang & Qui, 2006). There are two major groups of fungi that form these associations. Arbuscular mycorrhizal fungi (AMF) are the most common (Schüβler *et al*., 2001) and oldest lineage of these symbionts (Wang & Qui, 2006). Recent research has shown that AMF is a polyphyletic group dominated by Glomeromycetes, though recent research suggests some species belong to the Mucuromycotina (Orchard *et al*., 2017). They are obligate symbionts, which infect the cell membranes of their host's roots (Genre *et al*., 2005). Ectomycorrhizal fungi (EMF) are a polyphyletic group that commonly forms associations with woody plants. These fungi do not penetrate the cells of their hosts;

rather, hyphae cover host roots in a thick sheath between and form a complex lattice between root cells (Tedersoo *et al*., 2010). Mycorrhizal fungi are strongly influenced by aboveground plant communities, though pH is also a major driver (Barnes at al., 2016).

Traditionally, soil trophic networks have focused on fungi as primary consumers of SOM. With their slow biomass turnover (Rousk & Bååth, 2007), symbiotic and saprotrophic fungi form the base of the fungal energy channel. They are especially an important food source for specialised nematodes as well as many microarthropod grazers (Scheu, 2002; Adl, 2003). However, in reality this view is simplistic and ignores some very unique life-history strategies. For instance, many fungi are consumers of living soil-dwelling organisms. Most notable amongst this group are those fungi that actively prey upon nematodes and plant pathogens. Active predation of nematodes and similar sized animals has been observed in several fungal groups. Various predatory strategies have been identified, including the use of hyphal traps, adhesive spores, toxins, and colonisation by cells in a dispersal stage. However, the relative importance of predation in energy acquisition by such fungal groups is unclear (Adl, 2003). There are also many significant fungal pathogens that reside in soil and cause important diseases of crop and horticultural plants (McCartney *et al*., 2003; Barnes *et al*., 2016).

Although often overlooked, protists are ubiquitous in soils (Geisen *et al*., 2018). Since their discovery by van Leeuwenhoek (1677), protists have proven a difficult group to of organisms to study. Even now they represent a vast swathe of largely undescribed biodiversity (Geisen *et al*., 2018), whose evolutionary relationships have only recently been defined (Adl *et al*., 2012). Indeed, the word protist is a catchall term designed for ease of use to refer to all eukaryotes that are not plants, fungi, or animals (Geisen *et al*.,

2018). Protists can be found throughout soil trophic networks forming groups of consumers, saprotrophs, filter-feeders of suspended particles of SOM and pathogens (Adl & Gupta, 2006). Protistan consumers are also very efficient at processing prey, meaning that bacterivores disproportionately stimulate nutrient cycling, mineralisation, respiration, and pollutant decomposition rates within soil communities (Adl & Gupta, 2006). Yet, many protists can acquire energy through both autotrophic and heterotrophic means, referred to as 'mixotrophs' (Geisen *et al*., 2018), making sweeping commentary on exact protistan contributions to soil trophic networks difficult. Indeed, Geisen *et al*. (2018) highlight that despite being commonly treated as bacterivores, protists can be categorised into a wide range of trophic groups including fungivores, saprotrophs, parasites, mutualists, and phototrophs. This range of complex protistan functional and taxonomic diversity begs for further research.

Soil fauna play essential roles in numerous ecosystem services and as ecosystem engineers (i.e. earthworms). The smallest soil animals including rotifers, tardigrades, and many nematode species are primarily bacterial or fungal grazers and microbiovores (Bongers & Bongers, 1998; Adl, 2003; Crotty *et al*., 2012). An enormous range of arthropods fills many trophic roles in soils including predators and decomposers feeding on SOM and fungi (Adl, 2003). It is thought that the mechanical destruction of detritus primarily by arthropods (Zimmer, 2002; Schädler & Brandl; 2004) and the consumption of soil and excretion of large quantities of waste by earthworms allow bacteria and fungi access to otherwise difficult to obtain nutrients (Adl, 2003). Plant parasites such as certain species of nematodes and arthropods (Bongers & Bongers, 1998; Bardgett & van der Putten, 2014) can also be found within the soil fauna. Due to their relatively large

size, these animals represent the link between soil trophic networks and aboveground consumers, where energy enters the greater landscape (Bardgett & van der Putten, 2014).

Many types of soil fauna play important functional roles in soils, though our understanding of life history characteristics, or functional traits, which influence ecosystem functions or mitigate environmental change (Suding *et al*., 2008), in soil fauna are poorly studied (Turnbull *et al*., 2014). Functional traits have also been pioneered for earthworms (Bartlett *et al*., 2010), nematodes (Bongers, 1990), Collembola, Oribatid mites, and some insects (Briones, 2014). The functional roles of earthworms are perhaps the most famous of the soil fauna. Earthworms are considered so-called 'ecosystem engineers' for their role in soil formation as well as the facilitation of water flow and aeration within soils though their burrows. Additionally, their consumption of soil, including microbes, increases mineralisation rates and provides a method of dispersal for inactive microbes. Earthworms can be considered in three broad functional categories: epigeic species, which live close to the soil surface, endogeic species that reside deep in the soil, and anecic species, which represent and intermediate group, burrowing up to the soil surface (Bartlett *et al*., 2010). Yet other soil fauna play equally important roles in soils.

Nematodes are an exceptional indicator taxon in soils due to their ubiquitous distribution, high abundance, ease of sampling, and functional diversity (Ferris *et al*., 2001). Bongers (1990) created the Maturity Index to assess changes in soil quality based on the composition of the free-living nematode community using the relative abundances of nematodes with reproductive characteristics. Briefly, nematodes are ranked from 1 to 5 on the 'coloniser-persistor' scale, with lower values given to smaller, generalist, highly

fecund (r-selected) species and higher values given to larger, less fecund, and more specialised (K-selected) species (Bongers, 1990; Vonk *et al*., 2013).

Some attempts have been made to assess soil fauna functions using taxonomy-independent metrics such as body size spectra for macro-invertebrate communities (Hocking *et al*., 2013), Collembola (Turnbull & Lindo, 2015), and nematodes (George & Lindo, 2015; Liu *et al*., 2015). Body size can be a good indicator of resource utilisation and extinction risk as species generally increase in size moving up trophic networks and larger consumers are more likely to be specialised, with smaller population sizes and therefore more prone to environmental stochasticity (Gonzalez & Chaneton, 2002; Cardillo, 2003; Turnbull *et al*., 2014). Relatively large soil fauna are known to show negative responses in abundance and functional diversity in response to agricultural intensification (Tsiafouli *et al*., 2015). However, even supposedly taxonomy-independent methods require some degree of sample identification and measurement, which can prove difficult when dealing with the extremely small sizes of most soil invertebrates (Turnbull *et al*., 2014) and indeed microbes – i.e. >0.5 μm diameter for bacteria (Christensen *et al*., 1999). As a result, many methods have been developed to capture important functions and characteristics of soil biota effectively and efficiently (Zornoza *et al*., 2015; Bouchez *et al*., 2016).

2.2 Characterising soil biota

Characterising soil biota has always been limited by methodological constraints. Indeed the scales in which this biodiversity exists have limited our understanding of the composition and dynamics governing belowground communities. However, data on soil biota identity, functional traits, and biomass underpin the field of soil ecology. A recent

meta-analysis by Cameron *et al*. (2019) highlights the inconsistencies in our understanding of extent of global belowground biodiversity. As such, many methods have been developed to describe and assess soil biodiversity. Historically assessments of soil biodiversity required the extraction of soil fauna from soils by creating uncomfortable conditions for organisms – i.e. Tullgren funnel (Winter & Behan-Pelletier, 2007) – or by culturing microbes on growth media in the laboratory (Carini, 2019). However, culturing techniques cannot capture total soil biodiversity (Islam & Wright, 2006; Geisen, 2016) and skilled taxonomists are increasingly hard to come by (Mora *et al*., 2011). Many soil ecologists have readily embraced the use of DNA-based community assessments as a method to overcome the shortcomings of culture-dependent methods. There is a wealth of such of biological assays that use approaches to directly sample the soil (Bouchez *et al*., 2016). The following subsections will briefly describe some of the methodologies used in this thesis – although there are many more – to study soil biota, including their uses, positives, and shortcomings.

2.2.1 Morphometrics

As the majority of soil biota are microscopic, species identification has required great expertise in taxonomy and microscopy. Morphological identifications are necessary to delineate new species (e.g. Lindo, 2015), and are common in many experimental manipulations (e.g. Berg *et al*., 1998; Räty & Huhta, 2003; Lindo *et al*., 2012) and surveys (e.g. Huhta *et al*., 1967; Keith *et al*., 2012; Tsiafouli *et al*., 2015). They also serve as a starting point for some functional metrics as discussed previously. Such methods involve much practice and experience to master, and are both labour-intensive and time consuming, although they are critical as species estimates of soil biota far

exceed current inventories (Barrios, 2007; Fierer *et al*., 2007). This state of knowledge has been referred to as the *taxonomic impediment*, as our understanding of biodiversity is incomplete. Further, the temporal costs of species descriptions and identifications and a decline in funding and expertise have compounded the taxonomic impediment (Coleman, 2015). There has been great concern of a decline in trained taxonomists generally (Mora *et al*., 2011). In recent years, there has been a push for ecologists and taxonomists to better collaborate (Halme *et al*., 2015) and for modernisation in taxonomic practices (Coleman, 2015), although there are still worries those taxonomic experts are not being adequately replaced, especially as many unknown species face extinction prior to scientific description (Bacher, 2012).

2.2.2 Phospholipid fatty acid analyses

Assessments of microbial biomass can be derived from cultured microbes – i.e. growing microbes in laboratories under controlled conditions although a large component of soil biodiversity cannot be cultured (Islam & Wright, 2006; Geisen, 2016). The development of phospholipid fatty acid (PLFA) analyses by White *et al*. (1979) has allowed for the direct quantification of microbial biomass from soils. The use of PFLA analysis has become widespread in soils research (Buyer & Sasser, 2012; Quideau *et al*., 2016). Briefly, PLFA analysis works by using gas chromatography to analyse lipids extracted from soils using a chloroform: methanol buffer solution and further fractionated into relevant groupings (Kaur *et al*., 2005). All organisms have a phospholipid layer in their cell membranes, which rapidly decomposes after death, meaning PLFA analysis can provide an accurate estimate of live biomass in soils (White *et al*., 1979). Different fractions of soil communities can be identified by their PLFA profiles allowing data on

proportions of soil microbial biomass to be analysed, such as bacteria, including Gram-negative and –positive groups, as well as fungi. However, there is growing uncertainty that such profiles are unique for all groups (Kaur *et al*., 2005). Drawbacks to PLFA are the high complexity of analysis and financial costs (Islam & Wright, 2006), though recent improvements have improved throughput time (Buyer & Sasser, 2012) and many laboratories offer commercial analysis services making outsourcing an appealing option (Quideau *et al*., 2016).

2.2.3 Quantitative polymerase chain-reaction

The development of the polymerase chain reaction (PCR) has been recognised as one of the most important advancements in biological sciences (Bartlett & Stirling, 2003). This method allows for the amplification of a target sequence of DNA across a number (usually 25-35) of cycles at various temperatures. There are three major steps: denaturation of target DNA, annealing of primer sequences to target DNA, and extension of this region to an appropriate length using DNA polymerase (Mullis *et al*., 1987). There have been a number of modifications on this technique as new reagents become available and as more complex questions are asked of the methodology. Quantitative PCR (qPCR) is a popular variation on standard PCR methodology. In this case, the number of amplicons is monitored in real-time during the PCR as opposed to after the process has ended using fluorescent markers that increase in intensity as more products are produced (Bouchez *et al*., 2016). Quantitative PCR can give a measure of absolute or relative microbial abundance (Raeymaekers, 2000). In soil ecology, qPCR is commonly used to determine abundance of broad microbial groups or measure gene expression, such as those involved in N cycling (Bouchez *et al*., 2016). Despite its frequent use and declining

costs, qPCR presents a number of challenges, especially to first-time users, as expensive, specialised equipment and technical knowledge to optimise primer choice, thermocycling parameters, and data analysis are needed to perform qPCR analyses effectively.

2.2.4 Metabarcoding

Technological advances and decreasing costs have led to the development of high-throughput sequencing methodologies. Certain high-throughput sequencing techniques have been developed in order to work directly with the genetic material of environmental samples. DNA barcoding is a popular method for identifying species based on unique gene sequences from short, standardised regions, popularly referred to as DNA barcodes, to delineate species (Herbert & Gregory, 2005). These must be short sequences of highly conserved DNA that can be sequenced with very conserved DNA sequence, a barcode, and also contain enough variation for the primers to show a meaningful difference between species but little within species (Taberlet *et al*., 2007). In prokaryotes the 16S rRNA gene subunit is recommended for barcoding analyses (Caporaso *et al*., 2011) as are the internal transcribed spacer (ITS) region in fungi (Schoch *et al*., 2012) and the 18S rRNA gene region for most protists (Behnke *et al*., 2011). The barcode sequences are then amplified using PCR.

Metabarcoding refers to the application of DNA barcoding sequences targeted and amplified from environmental samples and has been used effectively in soils (Orgiazzi *et al*., 2015). Metabarcoding surveys are popular in soil ecology. For example, metabarcoding has been used in surveys of soil biodiversity in European nations (Terrat *et al*., 2015; Dupont *et al*., 2016) and even across the globe (Tedersoo *et al*., 2014;

Delgado-Baquerizo *et al*., 2018). This precedence along with continually falling costs makes it an attractive option for characterizing diverse soil communities.

However, there are limitations. Metabarcoding will amplify all matching DNA, including that of dead organisms (Epp *et al*., 2012) and the nature of the data does not allow for quantification or abundance measures. Problems also arise from incomplete taxa inventories in published databases (Orgiazzi *et al*., 2015). Programmes that can annotate functional data to metabarcoding outputs have been developed (Langille *et al*., 2013; Nguyen *et al*., 2016). Yet these must be used cautiously, as metabarcoding does not provide information on actual gene expression and improper analyses will render functional data invalid. As metabarcoding can produce huge amounts of data, management and processing of sequencing data requires that researchers must be literate in computational practices. Indeed, the size of individual sequences in addition to their abundance necessitate specific algorithms be written to properly analyse datasets (Creer *et al*., 2010).

2.2.5 Metagenomics

Metagenomics, the study of genomes extracted directly from environmental samples, is a significant advancement as it allows the identification of previously unculturable and unknown organisms (Creer *et al*., 2010; Bouchez *et al*., 2016). There are many metagenomic techniques but they can be simply explained as extracting DNA or RNA from individual cells directly from the soil itself. Shotgun sequencing is one approach. In this method, as first described by Staden (1979), DNA sequences are broken up at random into many smaller segments, or reads. This process is repeated to produce

many overlapping reads. Ultimately, computer algorithms are used to collate overlapping reads into continuous sequences.

Shotgun sequencing has been integrated with high-throughput sequencing to produce and analyse whole genomes of microbial communities in relatively short time periods. In environmental samples, shotgun sequencing can be used to discern biodiversity or functional genes (Fierer *et al*., 2012; Mendes *et al*., 2015; Orellana *et al*., 2018). However, there is a potential for error in this method as the creation of chimeric sequences is high, and could lead to inaccurate sequencing and diversity measures (Prosser, 2015). Nevertheless, shotgun sequencing of environmental samples has proven effective in a variety of habitats and is effective for working with soil communities (Fierer *et al*., 2012; Mendes *et al*., 2015; Orellana *et al*., 2018). As with metabarcoding, this method produces huge amounts of data, which must be correctly filtered and analysed and also uses increasingly complex software, further highlighting the need for interdisciplinary training in computer sciences.

2.2.7 Microarrays

The use of microarrays is a method of studying functional genes that can be performed independent of PCR amplification (Bouchez *et al*., 2016). They have a long history of use in studies of pure cultures and in medicine (Zhou, 2003). GeoChip is a microarray developed for use in studying microbial functions and roles in biogeochemical processes (He *et al*., 2007). As with other microarrays, GeoChip works by hybridising fluorescently labelled DNA with probes made of short complementary nucleotides. The subsequent strength of fluorescence on bound probes is analysed to quantify relative gene copy numbers (He *et al*., 2007; Bouchez *et al*., 2016). GeoChip

data has appeared in a wide range of studies of soil microbial communities (Bai *et al*., 2013; Wei *et al*., 2016) and has been used to study soil microbial community responses to climate change (Xue *et al*., 2016). Probe number and design limit the number of genes targeted by a microarray, though due to their low cost, ease of use, and flexible design, microarrays are a useful tool for both quantitative and qualitative genetic assessments (Bouchez *et al*., 2016). Thus microarrays, such as GeoChip, and metagenomic techniques represent an important bridge between assessing biodiversity and the functional roles they fill in natural ecosystems.

2.3 Soil biota and ecosystem services

The key ecosystem processes outlined previously, such as nutrient cycling and decomposition, can be considered standalone or contributing components to ecosystem services. The term ecosystem services has become popular as a catchall for the benefits provided by natural systems to human societies. Although the term has been defined multiple times since its inception in the late 1960s, de Groot *et al*. (2002) define ecosystem services as "the internal functioning of the ecosystem and sometimes the benefits derived by humans from the properties and processes of those ecosystems", which has proven to be a fitting definition for the modern world. The important role of soils in providing ecosystem services has become increasingly studied in recent years (Ford *et al*., 2012; Robinson *et al*., 2014). Soils are essential to the maintenance of water and nutrient cycles, gas exchange, plant growth, and as habitat for an incredible variety of life (Barrios, 2007; Robinson *et al*., 2014). This realisation has led to calls for increased soil security worldwide (Koch *et al*., 2012; Jones *et al*., 2013).

There is growing evidence that belowground biodiversity exerts a strong influence on ecosystem services (Barrios, 2007; Robinson *et al*., 2014; de Sosa *et al*., 2018). Table 2.1 outlines the functions of soil biota that are ultimately beneficial to human society ranging from the obvious, food production/security, to less well known services, such as antibiotic reservoirs, and even more abstract, such as the maintenance of aesthetically and culturally important landscapes (Robinson *et al*., 2014). The positive effect of soil biodiversity on disease suppression is well known (Schnitzer *et al*., 2011). In this way, greater belowground biodiversity can increase crop yields (Barrios, 2007). Subsequently, fiscal values have been assigned to these ecosystem services in an effort to relate their importance to a wider audience and attempt to quantify the consequences of their degradation in more immediate terms (de Groot *et al*., 2002). This has only begun to be applied to soils in the past 25 years (Robinson *et al*., 2014). From some perspectives, this practice and indeed the term ecosystem services itself, is controversial and remains philosophically challenging and raises practical concerns regarding valuation of ecosystem services (reviewed in Salles, 2011). Moreover, measuring some ecosystem services can be difficult. Nutrient cycling, for instance, involves biogeochemical processes that although important, can be difficult to assign fiscal values too. Thus, in this thesis, the term ecosystem services will be used *sensu* de Groot *et al*. (2002), to denote those processes that promote a properly functioning ecosystem and their potentially beneficial outputs.

Recent research has attempted to identify soil organisms that can serve as indicators of ecosystem condition (Keith *et al*., 2012; Maxwell *et al*., 2017). However, these efforts are not always effectively relayed to the public, especially policy-makers, hindering their

effectiveness. When only the scientific community is aware of these discoveries, they cannot be effectively utilised. Conversely, policy-makers are accountable to the people they represent and may resist changes that appear risky or poorly understood. Therefore, greater co-operation between the scientific community and policy-makers is needed to ensure the maintenance of ecosystem processes from the landscape- to global-scales.

Table 2.1 Description of major ecosystem services provided by soil biota following the framework of de Groot et al. (2002).

Category	Service	
Regulation	Waste processing	Decomposition of waste materials returns nutrients to their respective cycles. Soils also reduce or eliminate human pathogenic agents in waste applied to land.
	Climate regulation	Soil is a C sink; changes can result in reducing and/or offsetting greenhouse gas emissions.
	Hydrological regulation	Water filtration and storage occurs in soils.
	Hazard regulation	Soil structure can prevent erosion and mitigate natural disasters.
	Nutrient cycling	Nutrients are mineralised and transformed through biotic and abiotic processes in soils.
	Pollution regulation	Pollutants are broken down in a similar manner to nutrients.
	Soil formation	Decomposition of organic matter within soils generates new soils.
	Biological control	Predation and competition between soil biota can prevent disease outbreaks.
Production	Food production	Soil fertility is crucial to producing food supplies
	Raw materials production	Materials like timber and natural fibres must be grown or raised with the use of fertile soils.
	Biomedical reservoir	Coevolution amongst soil biota can be harnessed to make novel antibiotics and pharmaceuticals.
	Novel industrial processes	Novel chemicals and enzymes can be harnessed from soil biota.
Habitat	Physical structure	Soil heterogeneity determines local biodiversity both below- and aboveground.
	Biodiversity	Soils support a wide array of biota, with both intrinsic and economic value.
Information	Education	The study of soils has led to many discoveries.
	Recreation	Soils provide natural, semi-natural, and artificial landscapes for human activities.

Note: Adapted from de Groot *et al*. (2002), Barrios (2007) and Robinson *et al*. (2014).

2.4 Impacts of land use change on soil biota and ecosystem services

2.4.1 Land use change

The conversion of natural ecosystems to suit human needs, or more simply land use change, has far-reaching consequences for life on Earth. Altering natural systems through activities including forestry, urban development, and agriculture has led to habitat fragmentation and degradation, increasing global C emissions, culminating in a loss of biodiversity worldwide (Foley *et al*., 2005). Although some of the changes resulting from land use change are intuitive (e.g. loss of habitat from deforestation), many of the most important changes occur on within the soil at scales undetectable to casual observation. For example, following conversion of natural systems to agricultural fields, levels of SOM decline rapidly leading to the removal of approximately 30% of the C in the top 100 cm of soil within 10 years (Post & Kwon, 2006). Similarly, clear-cutting forest stands changes local and regional hydrological regimes, potentially leading to removal of soil through run-off, as well as changes in other characteristics including pH, compaction, and biologically regulated processes such as decomposition (Keenan & Kimmins, 1993). In turn, these changes can greatly alter the communities of below- and aboveground biota (Huhta *et al*., 1967; Keenan & Kimmins, 1993; Marshall, 2000).

Such changes can have significant effects upon the functioning of natural systems and human development at both local and global scales. Since land use change is expected to increase over the next 40 years, these effects may become significant contributors to both the global climate crisis and socioeconomic instability within many regions. Indeed, Schmitz *et al*. (2014) expect the conversion of natural systems to croplands alone to increase by 200 – 300 million ha by 2050. Land use change can have

unexpected impacts, for example it has been linked to increase rates of human disease emergence and outbreaks (Foley *et al*., 2005). Soil C loss is of great concern globally (Lal, 2004) and occurs at a loss of 0.6% annually in England and Wales (Bellamy *et al.,* 2005). The C sequestration ability of soils remains unclear, with a number of factors including land use, agricultural practices, and microbial function influencing the total amount of C retained in soils (Gosling *et al*., 2017)

Conversion of natural systems to those used in food production, especially from recent agricultural intensification in both the developing and developed world, is a major contributor to global climate change (Foley *et al*., 2005). Numerous studies have shown that agricultural intensification has led to changes in local soil dynamics ultimately resulting in increased nutrient runoff, soil erosion, and overall declines in soil fertility (Stoate *et al*., 2009), as well as reductions in local biodiversity, especially insect pollinators, beetles, amongst other invertebrates (Tsiafouli *et al*., 2015; Sánchez-Bayo & Wyckhuys, 2019). Obviously, such drastic changes in biodiversity may have serious implications for the effectiveness and long-term function of intensive agriculture (Tsiafouli *et al*., 2015), though the effects of such changes on soil biota are not always straightforward.

2.4.2 Land use change and soil biota

Land use strongly influences the structure of soil trophic networks (de Vries *et al*., 2013; Tsiafouli *et al*., 2015). However, the effects of land use change on local biodiversity vary based both within and between taxa on the intensity of change and the ability of local communities to withstand it. Bacterial diversity has been found to increase following the conversion of rainforest soils for slash-and-burn agriculture in the Amazon

(da C. Jesus *et al*., 2009), whereas AMF communities change little following the same disturbance (Aguilar-Fernández *et al*., 2009). Gosling *et al.* (2017) found that conversion of agricultural soils to grasslands resulted in a community with increased biomass with a loss of bacterial dominance. Ultimately, this conversion created low-productivity grasslands due to low nitrogen (N) availability. Generally, conventional agricultural practices are detrimental to soil biodiversity. Tsiafouli *et al.* (2015) for instance, found that agricultural intensification across Europe largely negatively affected soil food webs, with losses in functional diversity across many trophic groups. Interestingly, whilst nematodes were not greatly impacted by agricultural methods, larger animals were. On the contrary, no-till agriculture promotes local biodiversity (Souza-Andrade *et al*., 2003; Adl *et al*., 2006), which can have subsequent positive influences on crop management, such as pest control (Lal *et al*., 2007). Thus soil communities can exhibit a wide range of responses even within the same group of organisms to similar disturbances (Table 2.2) due to the complexities of soil community structuring and impacts of anthropogenic change.

Table 2.2 Some recorded effects of various types of land use change on species richness, abundance, community composition, and/or diversity in major groups of soil biota.

Taxon	Land-use change	Response	Reference
Bacteria	Forest clearing for slash-and-burn agriculture	Increase in diversity; shift in community composition	da C. Jesus *et al*. (2009)
Archaea	Comparison of forest, cropland, restored, and degraded soils	Shifts in community composition	Shen *et al*. (2013)
AMF	Agricultural intensification	Declining species richness	Oehl *et al*. (2003)
Various protists	Agricultural extensification	Declinin gabundances	Mills & Adl (2006)
Nematodes	Clear-cutting Agricultural intensification	Declining abundance No change in species richness	Huhta *et al*. (1967) Tsiafouli *et al*. (2015)
Collembola	Comparisons of forest, mixed use, and agricultural land	Declining diversity and species richness with intensification	Sousa *et al*. (2006)
Oribatid Mites	Agricultural intensification	Declining species richness with intensification	Tsiafouli *et al*. (2015)
Earthworms	Agricultural intensification	Declining abundance, diversity, and species richness	Postma-Blaauw *et al*. (2010); Tsiafouli *et al*. (2015)

Such inconsistency can be attributed to two main sources. First, differences between soil habitats are likely contributors to this variation. Soil type and climatic variables can strongly influence community composition (Sousa *et al*., 2006). It must also be remembered that soils are complex systems in which a wide array of biotic and abiotic factors are at play. Huhta *et al*. (1967) refer to this as the so-called "prevailing situation" and suggest that the effects of anthropogenic disturbances must in a sense compete with those of natural variables. Biotic responses can also vary significantly because certain groups may be more resistant or susceptible to change due to evolutionary constraints like dispersal, reproductive capacity, or feeding preferences (Wall *et al*., 2010).

2.4.3 Impacts on ecosystem services

Conversion of natural systems for agricultural production has led to an increase in soil erosion globally. As sediment is moved to areas of deposition, organic C within the soils is made available and topsoils are buried (Van Oost *et al*., 2007). As a result, C and N cycles may be altered, leading to cascading effects for plant production, nutrient cycling and extreme environmental degradation through biodiversity loss (Quinton *et al*., 2010). Since land use change can alter local biodiversity, ecosystem services may be impeded or lost from the system due to shifts in the belowground community. For example, agricultural intensification across Europe has been shown to change food web structure (de Vries *et al*., 2013; Tsiafouli *et al*., 2015). Past studies have shown that soil communities dominated by bacteria, such as those associated with intensive agriculture, have higher rates of N loss (de Vries *et al*., 2012) and lower C sequestration (Six *et al*., 2006). Whereas, more extensively managed systems in which fungi dominate over bacteria have more efficient nutrient cycling (de Vries *et al*., 2012) and greater C sequestration (Six *et al*., 2006). However, this has not been observed on at the landscape level. de Vries *et al.* (2013) found no change in the bacterial-to-fungal ratio in a study of agricultural intensification across Europe. In this case, biomass across both channels was equally reduced. Additionally, there are worries that novel biomedical or industrially important compounds will be removed from soil systems (Daniel, 2004) either through direct extinction or shifts in community structure that make these compounds disadvantageous.

Land use change can also have far-reaching effects temporally. Wooded areas previously cultivated by the Romans in central France still retain noticeably higher

nutrient levels when compared to pristine woodlands (Diedhiou *et al*., 2009). These sites also have a distinct EMF community (Diedhiou *et al*., 2009), which means that much of Europe's present belowground biodiversity and its distribution could be representative of ancient anthropogenic disturbances instead of a pristine condition. Other examples of this phenomenon include medieval mottes, which have higher soil fertility than nearby uncultivated land (Closset-Kopp & Decocq, 2015). Anthropogenic changes in soil physical properties can take even longer to recover. Evidence of prehistoric irrigation in the Sonora Desert can be seen today where silt was added to sandy substrates supporting distinct plant communities (Hall *et al*., 2013).

With the strong evidence that land use change can cause serious problems for society in both the immediate and distant future, efforts to mitigate the negative impacts of development are being recognised as necessary steps to prevent long-term environmental destruction. Across Europe, efforts have been made to mitigate the environmental impacts of intensive farming, by encouraging farmers to implement less intensive management practices in exchange for economic incentives through agri-environment schemes. Agri-environment schemes identify country-specific targets such as reducing greenhouse gas emissions and land abandonment or restoring native biodiversity (Kleijn & Sutherland, 2003). In the UK, agri-environment schemes have historically focused on restoring native biodiversity (Kleijn & Sutherland, 2003); however, as of 2014, member states of the European Union are required to address the ecological challenges presented by food and energy production in their Rural Development Schemes (National Assembly for Wales, 2011). By incorporating measures to mitigate climate change, local habitat

degradation and biodiversity loss, new agri-environment schemes aim to provide a foundation for sustainable land use.

Agri-environment schemes are therefore becoming popular tools for governments and scientists to determine how the negative effects of land-use change can be overcome in their jurisdictions. However, the efficacy of these programmes is not always determined before implementation. Sometimes they may achieve their goals but produce other problems. For example, outwintering livestock is economically profitable and can positively influence bird biodiversity, but this practice severely damages the soil, leading to increased erosion and long-term declines in fertility (Jones *et al*., 2012). Therefore, agri-environment schemes must be comprehensively planned and actively monitored to ensure positive outcomes. Additionally, the financial incentives common to many European agri-environment frameworks are prone to abuse. Thus it is necessary to actively monitor such initiatives to mitigate unintended side effects and ensure positive outcomes.

2.4.4 Agri-environment schemes

The effects of past European agri-environment schemes have been extensively studied. However, these results are not uniform. For example, data from Switzerland have shown that agri-environment schemes largely benefit local biodiversity (Herzog *et al*., 2005; Knop *et al*., 2006). Yet Kleijn *et al*. (2001) found that species richness and diversity of birds and vascular plants was not affected by Dutch agri-environment schemes. Such discrepancy is likely due to a combination of inadequate study design (Kleijn & Sutherland, 2003) and inherent difficulties in implementing national-level schemes where smaller-scale plans may be more effective (Feehan *et al*., 2005). It must

also be remembered that these schemes need to incorporate the economic needs of farmers (Emery & Franks, 2012) as they move towards wider scale goals, including the inclusion of ecosystem services (Arnott *et al*., 2019).

Glastir was the Welsh Government's latest agri-environment scheme. It was launched in 2012 in order to update and align previous Welsh agri-environment schemes with the most recent changes to the European Union's (EU) Common Agriculture Policy, which took effect in 2014 (National Assembly for Wales, 2011). The objectives of Glastir were to: (1) provide balance between the need to produce food and protect the environment, (2) be accessible to all, (3) support biodiversity, climate change and water outputs, and (4) spread money for implementing agri-environment work more widely among farmers. Glastir was developed in order to meet the new EU sustainable development goals (National Assembly for Wales, 2011). These objectives are shared in other agri-environment and rural development plans across the EU (Dwyer *et al*., 2007). Building on the previous agri-environment schemes, strict guidelines were set to ensure Glastir participants meet scheme objectives (National Assembly for Wales, 2011). In addition to the scheme itself, an ongoing monitoring programme was established to actively evaluate the effectiveness of Glastir.

2.4.5 Glastir Monitoring & Evaluation Programme

Data was collected throughout the course of Glastir through the Glastir Monitoring and Evaluation Programme (GMEP), at the time the largest and most in-depth monitoring programme within the EU (Emmett & the GMEP team, 2017). This project was maintained through the collaboration of the Welsh Government with British government agencies including the Centre for Ecology and Hydrology (CEH) and British

Geological Survey, universities (Bangor University, St Andrews University, Staffordshire University, University of Aberdeen, University of Southampton, and Victoria University of Wellington), private sector scientific consultants (ADAS, APEM, Biomathematics and Statistics Scotland, Bowburn Consultants, ECORYS, Edwards Ecological Services), and charitable organisations (British Trust for Ornithology, Butterfly Conservation Wales, and Freshwater Habitats Trust). It was also a progressive step for agri-environment schemes in Wales. Previously, monitoring programmes released their results after the completion of the scheme. Data collection for GMEP follows a holistic ecosystem approach. A rolling annual survey is conducted across the affected areas in addition to the use of previously collected data from past agri-environment schemes and national-level research projects. Data presented in this thesis are direct output from GMEP (Emmett & the GMEP team, 2017).

The GMEP sampling design was based on previous experience with the UK Countryside Survey (CS) (Emmett *et al*., 2010). Over the course of summer 2013, survey teams took samples and measurements from 5 soil samples, which were taken from each of the 5 subplots that make up all 60 x 1 km^2 plots at 15 cm depth for physicochemical properties and at 8 cm from which mesofauna were extracted. This number was increased in year two (90 1 km^2 plots, 450 samples). Some samples were allocated for resampling in 4 years time (Garbutt *et al*., 2014). Sampling continued following the same design in 2015 and 2016; however, only data from 2013 and 2014 will feature in this thesis.

Some of the major findings from GMEP thus far are reviewed here. Soil organic C levels were stable in improved land and woodlands across Wales (Emmett & the GMEP team, 2017). However, this conflicts strongly with other national

monitoring schemes (Bellamy *et al.*, 2005), but may reflect the high prevalence of grasslands in the GMEP data (Emmett & the GMEP team, 2017). Soil pH increased due to recent reductions in acid deposition but remains higher than recommended levels for long-term productivity in improved systems (Emmett & the GMEP team, 2017). As such, many Welsh fields are too acidic for optimal growth due to reduced liming by farmers (Gibbons *et al*., 2014). Measures of soil fertility are difficult to assess but there is no evidence of significant changes in N or net P though there has been a shift from mineral to organic sources of P across Wales in this time. Soil N has become stable in improved and woodland habitats and is declining in natural habitats, which may benefit native plant communities (Emmett & the GMEP team, 2017). Also of interest are preliminary results of biodiversity surveys, in particular the reported declines of ants, centipedes and isopods and increase in millipedes and some beetles (Smart *et al*., 2014). Further conclusions from GMEP are still being compiled and analysed as part of various research projects including this thesis and those of other PhD students.

2.5 Conclusions

In summary, this literature review highlights the large extent of soil biota, some popular methods for its assessment, its importance in the provision of ecosystem services, the impacts of land use change on such processes and the biodiversity that underpins them, as well as the role of agri-environment schemes in managing sustainable development with a focus on Glastir and GMEP. From this discussion the following conclusions can be drawn. Soil biota underpin critically important ecosystem functions, yet their roles are often hard to determine. The choice of methodology for soil biological

assessments is critical when designing experiments. Some methods cannot be used to assess functional roles and investigations in which function is of interest may be better served by using multiple techniques. The GMEP framework is designed to accommodate the study of a wide range of ecosystem services and underlying biodiversity, though further analyses are required to integrate the two. Utilising both the GMEP framework and a long-term experimental trial, this thesis will present research that seeks to bridge the gap between our understanding how land use change impacts soil biodiversity and associated ecosystem services.

2.6. References

Adl, S.M., 2003. The Ecology of Soil Decomposition. CABI Publishing, Wallingford.

Adl, S.M., Coleman, D.C., 2005. Dynamics of soil protozoa using a direct count method. Biol. Fertil. Soils 42, 168-171.

Adl, S.M., Coleman, D.C., Read, F., 2006. Slow recovery of soil biodiversity in sandy loam soils of Georgia after 25 years of no-tillage management. Agr. Ecosyst. Environ. 114, 323-334.

Adl, S.M., Gupta, V.V.S.R., 2006. Protists in soil ecology and forest nutrient cycling. Can. J. For. Res. 36, 1805-1817.

Adl, S.M., Simpson, A.G.B., Lane, C.E., Lukeš, J., Bass, D., Bowser, S.S., Brown, M.W. et al., 2012. The revised classification of eukaryotes. J. Eukaryot. Microbiol. 59, 429-493.

Aguilar-Fernández, M., Jaramillo, V.J., Varela-Fregoso, L., Gavito, M.E., 2009. Short-term consequences of slash-and-burn practices on the arbuscular mycorrhizal fungi of a tropical dry forest. Mycorrhiza 19, 179-186.

Arnott, D., Chadwick, D., Harris, I., Koj, A., Jones, D.L., 2019. What can management option uptake tell us about ecosystem services delivery through agri-environment schemes? Land Use Policy 81, 194-208.

Bacher, S., 2012. Still not enough taxonomists: reply to Joppa *et al*. Trends Ecol. Evol. 27, 65-66.

Bai, S., Li, J., He, Z., Van Nostrand, J., Tian, Y., Lin, G., Zhou, J. et al., 2013. GeoChip-based analysis of the functional gene diversity and metabolic potential of soil microbial communities in mangroves. Appl. Microbiol. Biotechnol. 97, 7035-7048.

Banning, N.C., Maccarone, L.D., Fisk, M.L., Murphy, D.V., 2015. Ammonia-oxidising bacteria not archaea dominate nitrification activity in semi-arid agricultural soil. Sci. Rep.-UK 5, 11146.

Bardgett, R.D., van der Putten, W.H., 2014. Belowground biodiversity and ecosystem functioning. Nature 515, 505-511.

Barnes, C.J., van der Gast, C.J., Burns, C.A., McNamara, N.P., Bending, G., D., 2016. Temporally variable geographical distance effects contribute to the assembly of root-associated fungal communities. Front. Microbiol. 7:195.

Barrios E., 2007. Soil biota, ecosystem services and land productivity. Ecol. Econ. 64, 269-285.

Bartlett, J.M.S., Stirling, D., 2003. A short history of the polymerase chain reaction. In: Bartlett, J.M.S., Stirling, D., (Eds.), PCR Protocols, second ed. Humana Press, Totowa, pp. 3-7.

Bartlett, M.D., Briones, M.J.I., Neilson, R., Schmidt, O., Spurgeon, D., Creamer, R.E., 2010. A critical review of current methods in earthworm ecology: from individuals to populations. Eur. J. Soil Biol. 46, 67-73.

Behnke, A., Engel, M., Christen, R., Nebel, M., Kelin, R.R., Stoeck, T., 2010. Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. Environ. Microbiol. 13, 340-349.

Bellamy, P.H., Loveland, P.J., Bradley, R.I., Lark, R.M., Kirk, G.J.D., 2005. Carbon losses from all soils across England and Wales 1978-2003. Nature 437, 245-248.

Berg, M.P., Kniese, J.P., Bedaux, J.J.M., Verhoef, H.A., 1998. Dynamics and stratification of functional groups of micro- and mesoarthropods in the organic layer of a Scots pine forest. Biol. Fertil. Soils 26, 268-284.

Bongers, T., 1990. The maturity index: an ecological measure for environmental disturbance based on nematode species composition. Oecologia 83, 14-19.

Bongers, T., Bongers, M., 1998. Functional diversity of nematodes. App. Soil Ecol. 10, 239-251.

Bouchez, T., Blieux, A.L., Dequiedt, S., Domaizon, I., Dufresne, A., Ferreira, S., Godon, J.J. et al., 2016. Molecular microbiology methods for environmental diagnosis. Environ. Chem. Lett. 14, 423-441.

Briones, M.J.I., 2014. Soil fauna and soil functions: a jigsaw puzzle. Front. Environ. Sci. 2, 1-22.

Buyer, J.S., Sasser, M., 2012. High throughput phospholipid fatty acid analysis of soils. Appl. Soil Ecol. 61, 127-130.

Da C. Jesus, E., Marsh, T.L., Tiedje, J.M., de S. Moreira, F., 2009. Changes in land use alter the structure of bacterial communities in Western Amazon soils. ISME J. 3, 1004-1011.

Cameron, E.K., Martins, I.S., Lavelle, P., Mathieu, J., Tedersoo, L., Bahram, M., Gottschall, F. et al., 2019. Global mismatches in aboveground and belowground biodiversity. Conserv. Biol. 33, 1187-1192.

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Louzupone, C.A., Turnbaugh, P.J., Fierer, N. et al., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proc. Nat. Acad. Sci. 108, 4516-4522.

Cardillo, M., 2003. Biological determinants of extinction risk: why are smaller species less vulnerable? Anim. Conserv. 6, 63e69.

Carini, P., 2019. A “cultural” renaissance: genomics breathes new life into an old craft. mSystems 4: e00092-19.

Coleman, C.O., 2015. Taxonomy in the times of the taxonomic impediment – examples from the community of experts on amphipod crustaceans. J. Crustacean Biol. 35, 720-740.

Christensen, H., Hansen, M., Sørensen, J., 1999. Counting and size classification of active soil bacteria by fluorescence *in situ* hybridization with an rRNA oligonucleotide probe. Appl. Environ. Microb. 65, 1753-1761.

Closset-Kopp, D., Decocq, G., 2015. Remnant artificial habitats as biodiversity islets into forest oceans. Ecosystems 18, 507-519.

Creer, S., Fonseca, V.G., Porazinska, D.L., Giblin-Davis, R.M., Sung, W., Power, D.M., Packer, M. et al., 2010. Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Mol. Ecol. 19 (Suppl. 1), 4-20.

Crotty, F.V., Adl, S.M., Blackshaw, R.P., Murray, P.J., 2012. Using stable isotopes to differentiate trophic feeding channels within soil food webs. J. Eukaryot. Microbiol. 0, 1-7.

Crotty, F.V., Blackshaw, R.P., Murray, P.J., 2011. Tracking the flow of bacterially derived ^{13}C and ^{15}N through soil faunal feeding channels. Rapid Commun. Mass Sp. 25, 1503-1513.

Daniel, R., 2004. The soil metagenome – a rich resource for the discovery of novel natural products. Curr. Opin. Biotech. 15, 199-204.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T. et al., 2018. A global atlas of the dominant bacteria found in soil. Science 359, 320-325.

Diedhiou, A.G., Dupouey, J.-L., Buée, M., Dambrine, E., Laüt, L., Garbaye, J., 2009. Response of ectomycorrhizal communities to past Roman occupation in an oak forest. Soil Biol. Biochem. 41, 2206-2213.

Dupont, A.Ö.C., Griffiths, R.I., Bell, T., Bass, D., 2016. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. Environ. Microbiol. 18, 2010-2024.

Dwyer, J., Ward, N., Lowe, P., Baldock, D., 2007. European rural development under the Common Agricultural Policy's 'second pillar': institutional conservatism and innovation. Reg. Stud. 41, 873-887.

Emery, S.B., Franks, J.R., 2012. The potential for collaborative agri-environment schemes in England: can a well-designed collaborative approach address farmers' concerns with current schemes. J. Rural Stud. 28, 218-231.

Emmett, B.E., the GMEP team, 2017. Glastir Monitoring & Evaluation Programme. Final Report to Welsh Government – Executive Summary (Contract reference: C147/2010/11). NERC/Centre for Ecology & Hydrology (CEH Projects: NEC04780/NEC05371/NEC05782).

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z. et al., 2010. Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology, 192pp. (CS Technical Report No. 9/07, CEH Project Number: C03259).

Epp, L.S., Boessenkool, S., Belleain, E.P., Haile, J., Esposito, A., Riaz, T., Erséus, C. et al., 2012. New environmental metabarcodes for analyzing soil DNA: potential for studying past and present ecosystems. Mol. Ecol. 21, 1821-1833.

FAO & ITPS, 2015. Status of the World's Soil Resources (SWSR) – Main Report. Food and Agriculture Organization of the United Nations and Intergovernmental Technical Panel on Soils. Rome, Italy.

Feehan, J., Gillmor, D.A., Culleton, N., 2005. Effects of an agri-environment scheme on farmland biodiversity in Ireland. Agri. Ecosyst. Environ. 107, 275-286.

Ferris H., Bongers T., de Goede, R.G.M., 2001. A framework for soil food web diagnostics: extension of the nematode faunal analysis concept. App. Soil Ecol. 18, 13-29.

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M. et al., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microb. 73, 7059-7066.

Fierer, N., Leff, J.W., Adams, B.J., Nielsen, U.N., Bates, S.T., Lauber, C.L., Owens, S. et al., 2012. Cross-biome metagenomic analyses of soil microbial communities and their functional attributes. Proc. Natl. Acad. Sci. USA. 109, 21390-21395.

Foley, J.A., DeFries, R., Asner, G.P., Barford, C., Bonan, G., Carpenter, S.R., Chapin, F.S. et al., 2005. Global consequences of land use. Science 309, 570-574.

Ford, H., Garbutt, A., Jones, D.L., Jones, L., 2012. Impacts of grazing abandonment on ecosystem service provision: coastal grassland as a model system. Agr. Ecosyst. Environ. 162, 108-115.

Garbutt, A., Henrys, P., Astbury, S., Biggs, J., Brown, M., Cosby, B.J., Edwards, F. et al., 2014. Field Survey Design and Implementation. In: Emmett, B.E., and the GMEP team, 2014. Glastir Monitoring & Evaluation Programme. First Year Annual Report to Welsh Government (Contract reference: C174/2010/11). NERC/Centre for Ecology & Hydrology (CEH Prject: NEC04780), pp. 102-126.

Geisen, S., 2016. The bacterial-fungal energy channel concept challenged by enormous functional versatility of soil protists. Soil Bio. Biochem. 102, 22-25.

Geisen, S. Bandow, C., Römbke, J., Bondowski. M., 2014. Soil water availability strongly alters the community composition of soil protists. Pedobiologia 57, 205-213.

Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D. et al., 2018. Soil protists: a fertile frontier in soil biology research. FEMS Microbiol. Rev. 42, 293-323.

Genre, A., Chabaud, M., Timmers, T., Bonfante, P., Barker, D.G., 2005. Arbuscular mycorrhizal fungi edit a novel intracellular apparatus in *Medicago truncatula* root epidermal cells before infection. Plant Cell 17, 3489-3499.

George, P.B.L., Lindo, Z., 2015a. Application of body size spectra to nematode trait-index analyses. Soil Biol. Biochem. 84, 15-20.

Gibbons, J.M., Williamson, J.C., Williams, A.P., Withers, P.J.A., Hockley, N., Harris, I.M., Hughes, J.W. et al. 2014. Sustainable nutrient management at field, farm and regional level: soil testing, nutrient budgets and the trade-off between lime application and greenhouse gas emissions. Agr. Ecosyst. Environ. 188, 48-56.

Gómez, P., Buckling, A., 2011. Bacteria-phage antagonistic coevolution in soil. Science 106-109.

Gonzalez, A., Chaneton, E., 2002. Heterotrophic extinction, abundance and productivity in a fragmented microecosystem. J. Anim. Ecol. 71, 594-603.

Gosling, P., van der Gast, C., Bending, G., 2017. Converting highly productive arable cropland in Europe to grassland: - a poor candidate for carbon sequestration. Sci Rep. UK 7:10493.

de Groot, R.S., Wilson, M.A., Boumans, R.M.J., 2002. A typology for the classification, description and valuation of ecosystem functions, goods and services. Ecol. Econ. 41, 393-408.

Hall, S.J., Trujillo, J., Nakase, D., Strawhacker, C., Kruse-Peeples, M., Schaafsma, H., Briggs, J., 2013. Legacies of prehistoric agricultural practices within plant and soil properties across an arid ecosystem. Ecosystems 16, 1273-1293.

Halme, P., Kuusela, S., Juslén, A., 2015. Why taxonomists and ecologists are not, but should be, carpooling? Biodivers. Conserv. 24, 1831-1836.

Harrington, T.C., Rizzo, D.M., 1999. Defining species in the fungi. In: Worrall, J.J. (ed.) Structure and Dynamics of Fungal Populations, Kluwer Press, Dordrecht, pp. 43-71.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z., et al. 2007. GeoChip: a comprehensive microarray for investigated biogeochemical, ecological and environmental processes. ISME J. 1, 67-77.

Herbert, P.D.N., Ratnasingham, S., deWaard, J.R., 2003. Barcoding animal life: cytochrome *c* oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B (Suppl.) 270, S96-S99.

Herzog, F., Dreier, S., Hofer, G., Marfurt, C., Schüpbach, B., Spiess, M., Walter, T., 2005. Effect of ecological compensation areas on floristic and breeding bird diversity in Swiss agricultural landscapes. Agri. Ecosyst. Environ. 108, 189-204.Hocking, M.D., Dulvy, N.K., Reynolds, J.D., Ring, R.A., Reimchen, T.E., 2013. Salmon subsidize an escape from a size spectrum. Proc. R. Soc. B 280: 2-122433.

Huhta, V., Karppinen, E., Nurminen, M., Valpas, A., 1967. Effect of silvicultural practices upon arthropod, annelid and nematode populations in coniferous forest soil. Ann. Zool. Fenn. 4, 87-145.

Islam, R., Wright, S.W., 2006. Microbial biomass measurement methods. In: Lal, R. (Ed.), Encyclopedia of Soil Science, second ed. Talyor and Francis, New York, pp. 1067-1070.

Jones, D.L., Cross, P., Withers, P.J.A., DeLuca, T.H., Robinson, D.A., Quilliam, R.S., Harris, I.M. et al., 2013. Nutrient stripping: the global disparity between food security and soil nutrient stocks. J. Appl. Ecol. 50, 851-862.

Jones, D.L., DeLuca, T.H., Quilliam, R.S., Williams, A.P., Newell-Price, J.P., 2012. Assessing the environmental sustainability of forage crops for out-wintering livestock in Wales. Welsh Government.0, 851-862.

Jonas, J.L., Wilson, G.W.T., White, P.M., Joern, A., 2007. Consumption of mycorrhizal and saprophytic fungi by Collembola in grassland soils. Soil Biol. Biochem. 39, 2594-2602.

Kaur, A., Chaudhary, A., Kaur, A., Choudhary, R., Kaushik, R., 2005. Phospholipid fatty acid – a bioindicator of environment monitoring and assessment in soil ecosystem. Curr. Sci. India 89, 1103-1112.

Keenan, R.J., Kimmins, J.P., 1993. The ecological effects of clear-cutting. Environ. Rev. 1, 121-144.

Keith, A.M., Boots, B., Hazard, C., Niechoj, R., Arroyo, J., Bending, G.D., Bloger, T. et al., 2012. Cross-taxa congruence, indicators and environmental gradients in soils under agricultural and extensive land management. Eur. J. Soil Biol. 49, 55-62.

Kleijn, D., Berendse, F., Smit, R., Gilissen, N., 2001. Agri-environment schemes do not effectively protect biodiversity in Dutch agricultural landscapes. Nature 413, 723-725.

Kleijn, D., Sutherland, W.J., 2003. How effective are European agri-environment schemes in conserving and promoting biodiversity? J. Appl. Ecol. 40, 947-969.

Kimura, M., Jia, Z.-J., Nakayama, N., Asakawa, S., 2008. Ecology of viruses in soils: past, present and future perspectives. Soil Sci. Plant Nutr. 54, 1-32.

Knop, E., Kleijn, D., Herzog, F., Schmid, B., 2006. Effectiveness of the Swiss agri-environment scheme in promoting biodiversity. J. Appl. Ecol. 43, 120-127.

Koch, A., McBratney, A., Lal, R., 2012. Put soil security on the global agenda. Nature 492, 186.

Kück, U., Pöggeler, S., 2009. Cryptic sex in fungi. Fungal Biol. Rev. 23, 86-90.

Langille, M., Zaneveld, J., Caporaso, J.G., McDonald, D., Knights, D., Reyes, J. A., Clements, J.C. et al., 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. Nat. Biotechnol. 31, 814-821.

van Leewenhoeck, A., 1677. Observations, communicated to the publisher by Mr. Antony van Leewenhoek, in a Dutch letter of the 9 Octob. 1676 here English'd concerning little animals by him observed in rain-well-sea and snow water; as alosi in water wherein pepper had lain infused. Phil. Trans. 12, 821-831.

Lal, R., 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304, 1623-1627.

Lal, R., Reicosky, D.C., Hanson, J.D., 2007. Evolution of the plow over 10,000 years and the rationale for no-till farming. Soil Tillage Res. 93, 1-12.

Lindo, Z., 2015. A rare new species of *Metrioppia* (Acari: Oribatida: Peloppiidae) from a Pacific Northwest temperate rainforest. Can. Entomol. 147, 553-563.

Lindo, Z., Whiteley, J.A., Gonzalez, A., 2012. Traits explain community disassembly and trophic contraction following experimental environment change. Glob. Change. Biol. 18, 2448-2457.

Liu, T., Guo, R., Ran W., Whalem J.K., Li, H.X., 2015. Body size is a sensitive trait-based indicator of soil nematode community response to fertilization in rice and wheat agroecosystems. Soil Biol. Biochem. 88, 275-281.

Maier, R.M., Pepper, I.A., 2000. Terrestrial Environments. In: Maier, R.M., Pepper, I., Gerba, C.P. (Eds.), Environmental Microbiology. Academic Press, San Diego, pp. 61-90.

Marshall, V.G., 2000. Impacts of forest harvesting on biological processes in northern forest soils. Forest Ecol. Manag. 133, 43-60.

Maxwell, D., Robinson, D.A., Thomas, A., Jackson, B., Maskell, L., Jones, D.L., Emmett, B.A., 2017. Potential contribution of soil diversity and abundance metrics to identifying high nature value farmland (HNV). Geoderma 305, 417-432.

McCartney, H.A., Foster, S.J., Fraaije, B.A., Ward, E., 2003. Molecular diagnostics for fungal plant pathogens. Pest Manag. Sci. 59, 129-142.

Mendes, R., Garbeva, P., Raaijmakers, J.M., 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic mircoorganisms. FEMS Microbiol. Rev. 37, 634-663.

Mendes, L.W., Tsai, S.M., Navarrete, A.A., de Hollander, M., van Veen, J.A., Kuramae, E.E., 2015. Soil-borne microbiome: linking diversity to function. Microb. Ecol. 70, 255-265.

Mills, A., Adl, S.M., 2006. The effects of land use intensification on soil biodiversity in the pasture. Can. J. Plant. Sci. 86 (special issue), 1339-1343.

Moore, J.C., McCann, K., de Ruiter, P.C., 2005. Modeling trophic pathways, nutrient cycling, and dynamic stability in soils. Pedobiologia 49, 499-510.

Mora, C., Tittensor, D.P., Adl, S. Simpson, A.G.B., Worm, B., 2011. How many species are there on earth and in the ocean? PLoS Biol. 9: 1-8.

Moulin, L., Munive, A., Dreyfus, B., Boivin-Masson, C., 2001. Nodulation of legumes by members of the β-subclass of Proteobacteria. Nature 411, 948-950.

Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., Erlich, H., 1986. Specific enzymatic amplication of DNA in vitro: the polymerase chain reaction. Cold Spring Harb. Symp. Quant. Biol. 51, 263-273.

Murray, P.J., Clegg, C.D., Crotty, F.V., de la Fuente Martinez, N., Williams, J.K., Blackshaw, R.P., 2009. Dissipation of bacterially derived C and N through the meso- and macrofauna of a grassland soil. Soil Biol. Biochem. 41, 1146-1150.

Muyzer, G., Stams, A.J.M., 2008. The ecology and biotechnology of sulphate-reducing bacteria. Nat. Rev. Microbiol. 6, 441-454.

National Assembly for Wales, 2011. An Introduction to Glastir and other UK agri-environment schemes. Paper number 11/012.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S. et al., 2016. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild.

Oehl, F., Sieverding, E., Ineichen, K., Mäder, P., Boller, T., Wiemken, A., 2003. Impact of land use intensity of the species diversity of arbuscular mycorrhizal fungi in agroecosystems of Central Europe. Appl. Environ. Microb. 69, 2816-2824.

Orchard, S., Hilton, S., Bending, G.D., Dickie, I.A., Standish, R.I., Gleeson, D.B., Jeffery, R.P. et al., 2017. Fine endophytes (*Glomus tenue*) are related to Mucoromycotina, not Glomeromycota. New Phyt.213, 481-486.

Orellana, L.H., Chee-Sanford, J.C., Sanford, R.A., Löffler, F.E., Konstantinidis, K.T., Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl. Environ. Microbiol. 84, e01646-17.

Orgiazzi, A., Dunbar, M.B., Panagos, P., de Groot, G.A., Lemanceau, P., 2015. Soil biodiversity and DNA barcodes: opportunities and challenges. Soil Biol. Biochem. 80, 244-250.

Peay, K.G., Kennedy, G., Bruns, T.D., 2008. Fungal community ecology: a hybrid beast with a molecular master. BioScience 58, 799-810.

Post, W.M., Kwon, K.C., 2000. Soil carbon sequestration and land-use change: processes and potential. Glob. Change Biol. 6, 317-328.

Postma-Blaauw, M.B., de Goede, R.G.M., Bloem, J., Faber, J.H., Brussaard, L., 2010. Soil biota community structure and abundance under agricultural intensification and extensification. Ecology 91, 460-473.

Prosser, J.I., 2015. Dispersing misconceptions and identifying opportunities for the use of 'omics' in soil microbial ecology. Nat. Rev. 13, 439-446.

Quideau, S.A., McIntosh, A.C.S., Norris, C.E., Lloret, E., Swallow, M.J.B., Hannam, K., 2016. Extraction and analysis of microbial phospholipid fatty acids in soils. J. Vis. Exp. 114, e54360.

Raeymaekers, L., 2000. Basic principles of quantitative PCR. Mol. Biotechnol. 15, 115-122.

Räty, M., Huhta, V., 2003. Earthworms and pH affect communities of nematodes and enchytraeids in forest soil. Biol. Fertil. Soils 38, 52-58.

Ratcliff, W.C., Kadam, S.V., Denison, R.F., 2008. Poly-3-hydroxybutyrate (PHB) supports survival and reproduction in starving rhizobia. FEMS Microbiol. Rev. 65, 391-399.

Raynoud, X., Nunan, N., 2014. Spatial ecology of bacteria at the microscale in soil. PLoS ONE 9, e87217.

Robinson, D.A., Fraser, I., Dominati, E.J., Davi∂sdóttir, B., Jónsson, J.O.G., Jones, L., Jones, S.B. et al., 2014. On the value of soil resources in the context of natural capital and ecosystem service delivery. Soil Sci. Soc. Am. J. 78, 685-700.

Rousk, J., Bååth, E., 2007. Fungal biomass production and turnover in soil estimated using the acetate-in-ergosterol technique. Soil Biol. Biochem. 39, 2173-2177.

Ruess, L., Chamberlain, P.M., 2010. The fat that matters: soil food web analysis using fatty acids and their carbon stable isotope signature. Soil Biol. Biochem. 42, 1898-1910.

Salles, J.-M., 2011. Valuing biodiversity and ecosystem services: why put a value on Nature? C.R. Biologies 334, 469-482.

Sánchez-Bayo, F., Wyckhuys, K.A.G., Worldwide decline of the entomofauna: a review of its drivers. Biol. Conserv. 232, 8-27.

Schädler, M., Brandl, R., 2005. Do invertebrate decomposers affect the disappearance rate of litter mixtures? Soil Biol. Biochem. 37, 329-337.

Scheu, S., 2002. The soil food web: structure and perspectives. Eur. J. Soil Biol. 38, 11-20.

Schoch, C.L., Seifert, K.A., Huhndorf, S., Robert, V., Spouge, J.L., Levesque, C.A., Chen, W., Fungal Barcoding Consortium, 2012. Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for *Fungi*. P. Natl. Acad. Sci. USA. 109, 6241-6246.

Schmitz, C., van Meijl, H., Kyle, P., Nelson, G.C., Fujimori, S., Gurgel, A., Havlik, P. et al., 2014. Land-use change trajectories up to 2050: insights from a global agro-economic model comparison. Agr. Econ. 45, 1-16.

Schnitzer, S.A., Klironomos, J.N., HilleRisLambers, J., Kinkel, L.L., Reich, P.B., Xiao, K., Rillig, M.C. et al., 2011. Soil microbes drive the classic plant diversity-productivity pattern. Ecology 92, 296-303.

Schüβler, A., Schwarzott, D., Walker, C., 2001. A new fungal phylum, the *Glomeromycota*: phylogeny and evolution. Mycol. Res. 105, 1413-1421.

Shen, J.-P., Cao, P., Hu, H.-W., He, J.-Z., 2013. Differential response of archaeal groups to land use change in an acidic red soil. Sci. Total Environ. 461-462, 742-749

Six, J., Frey, S.D., Thiet, R.K., Batten, K.M., 2006. Bacterial and fungal contributions to carbon sequestration in agroecosystems. Soil Sci. Soc. Am. J. 70, 555-569.

Smart, S.M., Astbury, S., August, T., Botham, M., Cooper, J., Emmett, B.A., Hall, J. et al., Biodiversity. In: Emmett, B.E., and the GMEP team, 2014. Glastir Monitoring & Evaluation Programme. First Year Annual Report to Welsh Government (Contract reference: C174/2010/11). NERC/Centre for Ecology & Hydrology (CEH Prject: NEC04780), pp. 128-171.

Song, D., Pan, K., Tariq, A., Sun, F., Li, Z., Sun, X., Zhang, L. et al., 2017. Large-scale patterns of distribution and diversity of terrestrial nematodes. Appl. Soil Ecol. 114, 161-169.

de Sosa, L.L., Glanville, H.C., Marshall, M.R., Williams, A.P., Jones, D.L., 2018. Quantifying the contribution of riparian soils to the provision of ecosystem services. Sci. Total Environ. 624, 807-819.

Sousa, J.P., Bolger, T., da Gama, M.M., Lukkari, T., Ponge, J.-F., Simón, C., Traser, G. et al., 2006. Changes in Collembola richness and diversity along a gradient of land-use intensity: a pan European study. Pedobiologia 50, 147-156.

Souza-Andrade, D., Colozzi-Filho, A., Giller, K.E., 2003. The soil microbial community and soil tillage. In: El Titi, A., (Ed.), Soil Tillage in Agroecosystems. CRC Press, Boca Raton, pp. 51-81.

Staden, R., 1979. A strategy of DNA sequencing employing computer programs. Nucleic Acids Res. 6, 2601-2610.

Stoate, C., Báldi, A., Beja, P., Boatman, N.D., Herzon, I., van Doorn, A., de Snoo, G.R. et al., 2009. Ecological impacts of early 21st century agricultural change in Europe – a review. J. Environ. Manag. 91, 22-46.

Suding, K.N., Lavorel, S., Chapin III, F.S., Cornelissen, J.H.C., Díaz, S., Garnier, E.G., Hooper, D.U. et al., 2008. Scaling environmental change through the community-

level: a trait-based response-and-effect framework for plants. Glob. Change Biol. 14, 1125-1140.

Swanson, M.M., Fraser, G., Daniell, T.J., Torrance, L., Gregory, P.J., Taliansky, M., 2009. Viruses in soils: morphological diversity and abundance in the rhizosphere. Ann. Appl. Biol. 155, 51-60.

Taberlet, P., Coissac, E., Pompanon, F., Gielly, L., Miquel, C., Valentini, A., Vermat, T. et al., 2007. Power and limitations of the chloroplast *trn*L (UAA) intron for plant DNA barcoding. Nucleic Acids Res. 35, *e*14.

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Villarreal Ruiz, L. et al., 2014. Global diversity and geography of soil fungi. Science 346, 1256688.

Tedersoo, L., May, T.W., Smith, M.E., 2010. Ectomycorrhizal lifestyle in fungi: global diversity, distribution, and evolution of phylogenetic lineages. Mycorrhiza 20, 217-263.

Terrat, S., Dequiedt, S., Horrigue, W., Lelievre, M., Cruaud, C., Saby, N.P.A., Jolivet, C. et al., 2015. Improving soil bacterial taxa-area relationship assessment using DNA meta-barcoding. Heredity 114, 468-475.

Timonen, S., Bomberg, M., 2009. Archaea in dry soil environments. Phytochem. Rev. 8, 505-518.

Tsiafouli, M.A., Thébault, E., Sgardelis, S.P., de Ruiter, P.C., van der Putten, W.H., Birkhofer, K., Hemerik, L. et al., 2015. Intensive agriculture reduces soil biodiversity across Europe. Global Change Biol. 21, 973-985.

Turnbull, M.S., George, P.B.L., Lindo, Z., 2014. Weighing in: size spectra as a standard tool in soil community analyses. Soil Biol. Biochem. 68, 366-372.

Turnbull, M.S., Lindo, Z., 2015. Combined effects of abiotic factors on Collembola communities reveal precipitation may act as a disturbance. Soil Biol. Biochem. 82, 36-43.

Usher, M.B., Davis, P., Harris, J., Longstaff, B., 1979. A profusion of species? Approaches towards understanding the dynamics of the populations of microarthropods in decomposer communities. In: Anderson, R.M., Turner, B.D.,

Taylor, L.R. (Eds.), Population Dynamics. Oxford University Press, Oxford, pp. 359-384.

Vonk, J.A., Breure, A.M., Mulder, C., 2013. Environmentally-driven dissimilarity of trait-based indicies of nematodes under different agricultural management and soil types. Agr. Ecosyst. Environ. 179, 133-138.

Vos, M., Birkett, P.J., Birch, E., Griffiths, R.I., Buckling, A., 2009. Local adaptation of bcateriophages to their bacterial hosts in soil. Science 325. 833.

de Vries, F.T., Bloem, J., Quirk, H., Stevens, C.J., Bol, R., Bardgett, R.D., 2012. Extensive management promotes plant and microbial nitrogen retention in temperate grassland. PLoS One 7, e51201.

de Vries, F.T., Thébault, E., Lirri, M., Birkhofer, K., Tsiafouli, M.A., Bjørnlund, L., Jørgensen, H.B. et al., 2013. Soil food web properties explain ecosystem services across European land use systems. Proc. Natl. Acad. Sci. USA. 110, 14296-14301.

Wall, D.H., Bardgett, R.D., Kelly, E.F., 2010. Biodiversity in the dark. Nature Geosci. 3, 297-298.

Wang, B., Qiu, Y.-L., 2006. Phylogeneitc distribution and evolution of mycorrhizas in land plants. Mycorrhiza 16, 299-363.

Wei, S.T.S., Lacap-Bugler, D.C., Lau, M.C.Y., Caruso, T., Rao, S., de los Rios, A., Archer, S.K. et al., 2016. Taxonomic and functional diversity of soil and hypolithic microbial communities in Miers Valley, McMurdo Dry Valleys, Antacritca. Front. Microbiol. 7, 1642.

White, D.C., Davis, W.M., Nickels, J.S., King, J.D., Bobbie, R.J., 1979. Determination of sedimentary microbial biomass by extractable lipid phosphate. Oecologia 40, 51-62.

Winter, J.P., Behan-Pelletier, V.M., 2007. Microarthropods. In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis, second ed. CRC press, Boca Raton, pp. 399-414.

Xue, K., Yuan, M.M., Shi, Z.J., Qin, Y., Deng, Y., Cheng, L., Wu, L. et al., 2016. Tundra soil carbon is vulnerable to rapid microbial decomposition under climate warming. Nat. Clim. Change 6, 595-600.

Zhou, J., 2003. Microarrays for bacterial detection and microbial community analysis. Curr. Opin. Microbiol. 6, 288-294.

Zimmer, M., 2002. Is decomposition of woodland leaf litter influenced by its species richness? Soil Biol. Biochem. 34, 277-284.

Zornoza, R., Acosta, J.A., Bastida, F., Domínguez, S.G., Toledo, D.M., Faz, A., 2015. Identification of sensitive indicators to assess the interrelationship between soil quality, management practices and human health. Soil 1, 173-185.

Chapter 3: Evaluation of mesofauna communities as soil quality indicators in a national-level monitoring programme

This manuscript has been published in *Soil Biology & Biochemistry*

George, P.B.L.[1,2,*], Keith, A.M.[3], Creer, S.[4], Barrett, G.L.[2], Lebron, I.[2], Emmett, B.A.[2], Robinson, D.A.[2], Jones, D.L.[1], 2017. Evaluation of mesofauna communities as soil quality indicators in a national-level monitoring programme. *Soil Biol. Biochem*. 115, 537-546.

[1] School of Environment, Natural Resources & Geography, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] Centre for Ecology & Hydrology, Lancaster, Bailrigg, Lancaster, LA1 4AP, UK

[4] School of Biological Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[*] Corresponding author

A.M.K., D.L.J., D.A.R., and B.A.E. conceived this project. A.M.K. extracted meosfauna. P.B.L.G. and A.M.K. identified mesofauna. B.L.G. and I.L. processed the soil samples and collected data. Statistical analyses were led by P.B.L.G. with assistance from A.M.K. P.B.L.G. wrote the first draft of the manuscript and A.M.K., D.A.R., S.C., and D.L.J. contributed to subsequent revisions. All authors read and approved the final manuscript.

Abstract

Mesofauna underpin many ecosystem functions in soils. However, mesofauna communities are often overlooked when discussing these functions on large scales. They have been proposed as bioindicators of soil quality and ecosystem health. This study aimed to evaluate differences between mesofauna communities, particularly mites and Collembola, across multiple habitat and soil types as well as in relation to an organic matter gradient and their relationships with soil characteristics, on a national-scale. Soil cores were collected from 685 locations in a nationwide soil monitoring programme of Wales and the mesofauna extracted and identified. Plant community composition, soil type, as well as physical and chemical variables, including pH, total C and N, were also measured at these locations. Mesofauna were extracted using a Tullgren funnel technique. Mites were separated into predatory (mesostigmatid) and decomposer (oribatid) groups; Collembola were separated according morpho-type. Abundances of mesofauna were consistently lowest in arable sites and highest in lowland woodlands, except for predatory mites. Differences between similar habitat types (e.g. Fertile and Infertile grasslands) could not be detected with the national-level dataset and differences in mesofauna communities between soil types were not clear. Relationships between mesofauna groups and soil organic matter class, however, were much more informative. Oribatid abundances were lowest in mineral soils and correlated with all soil properties except total phosphorus. Collembola and predatory mite abundances were likely negatively influenced by increased moisture levels in upland peat habitats where their abundances were lowest. These groups also had low abundances in heathlands, and this was reflected in low diversity values. Together, these findings show that this national-

level soil survey can effectively identify differences in mesofauna communities and correlations with soil properties. The use of broad-groups of mesofauna in national-level soil monitoring is encouraged to better understand the ecological context of changes in soil properties.

Key words: *Soil biodiversity; Vegetation class; Microarthropods; Hydrophobicity; Wales; Acari*

3.1 Introduction

Mesofauna represent a major component of soil biological communities and play a critical role in maintaining soil quality and a range of ecosystem functions (Barrios, 2007). Indeed, soil invertebrates support primary production, nutrient cycling, and soil formation, as well as facilitating water supply and regulating local erosion and climate (Lavelle et al., 2006: Barrios, 2007). Such functions are key components soil health (Doran and Zeiss, 2000). Mites (Gulvik, 2007) and Collembola (Rusek, 1998) are considered the most abundant and important groups of mesofauna. Collembola in soils are generally important consumers of microbial films and fungal hyphae or of larger plant detritus, and can be important architects of soil structure in some systems (Rusek, 1998). Important mite groups include the Oribatida and Mesostigmata. Oribatids are the most numerous and diverse sub-order of mites and are usually slow moving, heavily armoured, with comparatively low fecundity, relatively long lifespans (Gulvik, 2007) and commonly consume organic matter as well as fungi (Schneider et al., 2005). Mesostigmatids are commonly important predators within soils, consuming a wide range of invertebrate fauna (Gulvik, 2007)

Such life-history characteristics as well as their small size, varied ecological preferences, relatively high fecundity, and ease of sampling, have led to mesofauna being proposed as bioindicators of soil quality and ecosystem health (Gerlach et al., 2013). At the broad level, abundances of mites and Collembola are useful for understanding how ecosystems respond to the impacts and intensity of land-use (Black et al., 2003; Rutgers et al. 2009; Nielsen et al., 2010a; Arroyo et al., 2013), in addition to studying the effects of anthropogenic disturbance (Tsiafouli et al., 2015). Whille mesofauna are often overlooked, with recent emphasis on ecosystem-level surveys of soil microbial biodiversity with modern metagenomic techniques (Graham et al., 2016), it is encouraging that attempts to use mesofauna as bioindicators have been implemented in a number of large-scale soil assessment and ecosystem monitoring programmes across Europe.

In the Netherlands, abundances of mesofauna specifically in agricultural and horticultural sites were found to decline in areas with high disturbance and increase in areas where disturbance was minimal (Rutger et al., 2009). Cluzeau et al. (2012) also found that greater abundances of Collembola indicated the use of organic fertilisers and level of agricultural management. Ireland's Crébeo soil biodiversity assessment found certain mite species were only effective indictors of a small number of pasture sites with unusually high concentrations of phosphorous (P) and calcium (Keith et al., 2012). Soil invertebrate measures were added as bioindicator metrics to the UK Countryside Survey in 1998. Black et al. (2003) found mites, especially oribatids, preferred highly organic, moist soils as well as undisturbed upland habitats including moors, heaths, bogs, and

woods, whereas Collembola made up a greater proportion of mesofauna communities in grasslands and deciduous woodlands.

The fact that such monitoring programmes are undertaken at a national-scale means that trends can be observed for wide geographic areas, offering a range of benefits for ecological synthesis. Firstly, broad, intensive sampling contributes to a national taxonomic inventory for soil biota. Secondly, large-scale soil monitoring programmes provide a spatially varied dataset ideal for identifying biological indicators in relation to ecosystem health. Thirdly, such datasets also offer an opportunity to develop and test large-scale hypotheses on, agricultural practices, land remediation, and pollution in relation to ecosystem services and health. Finally, soils have been described as a resource as critical to sustaining human life as air and water (Havlicek, 2010). This importance is slowly becoming recognised through policy with, for example, the government of Wales adopting soil carbon (C) as a national status indicator of progress (Welsh Government, 2016).

The effectiveness of mesofauna as indicators of soil health on a national-scale is unclear, since contemporary surveys to date lack extensive detail on mesofauna trends. Of particular concern is whether differences between mesofauna communities are indicative of functional processes at the level of habitat or soil type. Understanding if broad-groups of mesofauna can show consistent nationwide trends or if they are limited to certain environmental characteristics is needed to realise their application as effective bioindicators.

Here, we present findings on mesofauna community metrics collected over a 2-year period as part of a nation-wide monitoring programme. Specifically, we aim to evaluate

how mesofauna communities, including abundances of various groups of mites and Collembola, differ between habitats and soils with diverse physico-chemical properties across an intensively sampled national landscape including many diverging habitats. We hypothesise that mesofauna will be more abundant and diverse with decreasing disturbance and specifically, that biodiversity will be lowest in frequently disturbed agricultural sites and highest in less-disturbed sites like woodlands. We also explore relationships between various mesofaunal groups and soil physical and chemical parameters. We expect organic matter (positive), pH, (positive) and moisture content (negative) to be most strongly correlated with mesofauna abundances. The ultimate aim of the work was to establish whether important mesofauna groups effectively delineate habitat and environmental differences for a national-scale assessment of ecosystem health.

3.2 Materials and methods

3.2.1 Study design

In Wales, UK, Glastir is a national-level agri-environment scheme, encompassing 4,911 landowners and an area of 3,263 km^2. It is the main way that the Welsh Government and the European Union (EU) pays for environmental goods and services. The Glastir Monitoring and Evaluation Programme (GMEP) was established to evaluate the scheme's effectiveness. GMEP collected evidence for six intended outcomes from the Glastir scheme; climate change mitigation, improvement to soil and water quality, a halt in the decline of biodiversity, improved woodland management and greater access to the welsh landscape and condition of historic features (Emmett & the GMEP team, 2015). Throughout its duration, GMEP was the largest and most in-depth active soil monitoring

programme measuring environmental state and change in the EU (Emmett & the GMEP team, 2014). For a detailed description of GMEP see Appendix 1.1.

As part of GMEP, survey teams travelled across Wales taking soil samples. The methodology used was established previously in the UK Countryside Survey (CS) (Emmett et al., 2010). Briefly, randomly allocated 1 km^2 squares, each containing 5 plot locations, were monitored across Wales. The habitat of each plot was classified using an Aggregate Vegetation Class (AVC), a classification based on a high-level aggregation of vegetation types derived from plant species data in each plot. There are eight categories of AVC: Crops/weeds, Tall grassland/herb, Fertile grassland, Infertile grassland, Lowland wood, Upland wood, Moorland-grass mosaic, and Heath/bog (Bunce et al., 1999; for detailed description see Appendix 1 Table 1). Soil type was categorised following the Main Group classifications of the National Soil Map (Avery, 1990; for detailed description see Appendix 1.2 and Appendix 1 Table 2). In addition, an organic matter classification was derived from loss-on-ignition (LOI) categories namely: mineral (0-8% LOI), humus-mineral (8-30% LOI), organo-mineral (30-60% LOI), and organic (60-100% LOI). The LOI classification was previously established as part of the 2007 Countryside Survey (Emmett et al., 2010).

Soils were sampled from late spring until early autumn in 2013 and 2014, with cores being taken at each plot (8 cm depth, 4 cm diameter) for subsequent mesofauna extraction, co-located with cores for soil chemical and physical parameters. These were taken from 60 x 1 km^2 squares in 2013 and 90 x 1 km^2 in 2014 (Fig. 3.1), with 684 samples ultimately included in analyses. Cores were kept in cool boxes or fridges at 4ºC

and then posted overnight to the Centre for Ecology & Hydrology (CEH), Lancaster for mesofauna extraction.

Soil physical and chemical characteristics were conducted on the additional soil cores from each site. We chose standard soil quality indicators including bulk density (g/cm^3), pH (measured in 0.01 M $CaCl_2$), volumetric water content (m^3/m^3), total phosphorous (P) (mg/kg), total C (%), total nitrogen (N) (%), and soil water repellency (as water drop penetration time in seconds). Mean values for each variable are presented in Appendix 1 Table 3. These analyses were conducted following Countryside Survey 2007 protocols (Emmett et al., 2010).

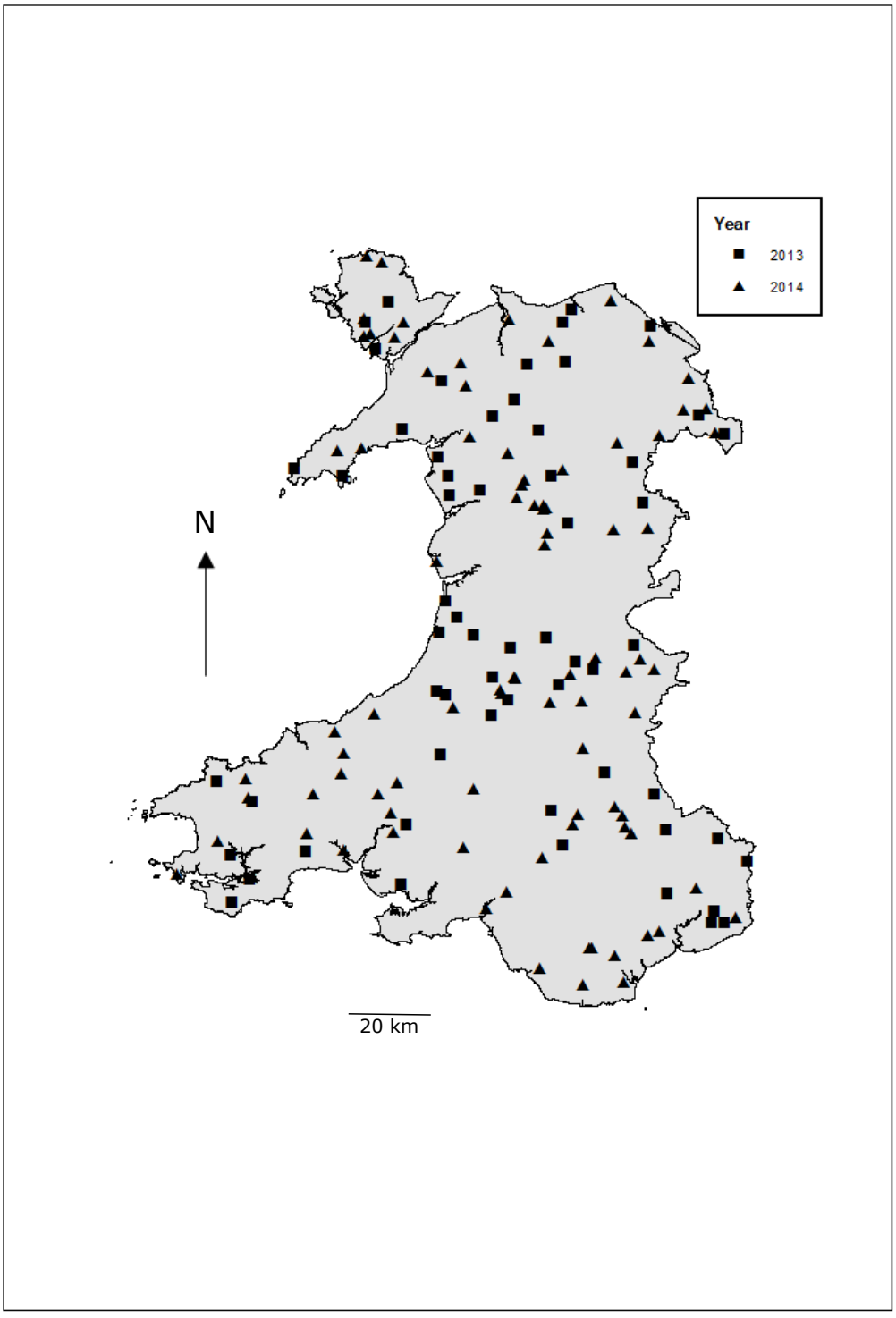


Figure 3.1 Map of 1 km^2 squares selected for GMEP monitoring. Sites are randomly offset by 10 km to protect landowner anonymity.

3.2.2 Mesofauna extraction and identification

Mesofauna were extracted from soil cores using a Tullgren funnel technique over the course of five days and collected in tubes containing 70% ethanol (Winter and Behan-Pelletier, 2007). Extracts were sorted to identify and enumerate mesofauna at broad-level groups. Due to their importance and proportional dominance in soils, mites and Collembola were of primary interest. Mites (Acari) were identified to Order (Mesostigmata) and Sub-order (Oribatida and Prostigmata) following Crotty and Shepherd (2014). Collembola were identified to Order (Symphypleona) or Superfamily (Entobryoidea and Poduroidea) following Hopkin (2007). Other animals identified included Araneae, Chilopoda, Coleoptera, Dermaptera, Diplura, Diptera, Hemiptera, Isopoda, Oligochaeta, Protura, Pseudoscorpiones, and Thysanoptera. Counts of extracted invertebrates were combined to determine total mesofauna (mites and Collembola abundances) as well as total invertebrate catch (abundance of all invertebrates extracted). Shannon's diversity (H') was calculated on abundance data of mesofauna groups.

2.2.3 Statistical analyses

Differences in community composition were assessed using non-metric dimensional scaling (NMDS) with subsequent analysis of multivariate homogeneity of group variances (*betadisper* function), followed by ANOVA with Tukey's HSD *post-hoc* tests, and similarity percentages (SIMPER), using the R software package "vegan" (Oksansen et al., 2016). The significance of changes in mesofauna abundances, total catch, and diversity were tested with linear mixed models using the "nlme" package (Pinheiro et al., 2016) with R version 3.1.1 (R Core Team, 2016) following log +1 transformations. The terms "identifier" (to denote who identified the mesofauna) and

"square" (the 1 km^2 square from which each sample was taken) were included as random-effects in the models. Where significant, the data were subjected to Tukey's HSD *post-hoc* testing to determine significant differences in mesofauna metrics between individual AVCs, soil types, and LOI classes. Correlations between mesofauna abundance and soil properties were determined using Spearmann's rank correlation coefficient and modified versions of the previously described linear mixed models with pseudo-R^2 values calculated with the "piecewiseSEM" package (Lefcheck, 2015).

3.3 Results

3.3.1 Mesofauna composition

Oribatid mites were generally the most common mesofauna group accounting for between 20 and 44% of the individuals recorded across AVC types. The Entomobryoidea were the most common group of Collembola encountered, especially in Upland and Lowland Woods, where they accounted for approximately 15-25% of mesofauna. Symphypleona (Collembola) were the rarest mesofauna group in all AVCs, representing less than 4% of the individuals recorded. While NMDS analysis revealed no distinct clusters in similarity of community composition (Appendix 1 Fig. 1), significant differences in homogeneity of variance across AVC types ($F_{7,677} = 3.113$, $p = 0.003$) were reflected through differences in the variation in mesofauna composition of Fertile grasslands and both Upland wood ($p = 0.04$) and Heath/bog ($p = 0.02$). Based on SIMPER analysis, this was likely driven by differences in proportional abundances of total Collembola and Mesostigmata. Mesostigmata accounted for approximately 21% and 18% of the dissimilarity when Fertile grassland was compared to Heath/bog and Upland

wood, respectively. Collembola accounted for approximately 33% and 36% of the dissimilarities between these groups.

3.3.2 Abundance and diversity measures

Differences between AVC types

AVC had a significant effect on total mesofauna abundance ($F_{7, 515} = 5.646$, $p < 0.001$), which was highest in Lowland wood and lowest in Crops/weeds (Fig. 3.2A). Total mesofauna abundances in Crops/weeds were significantly lower than in Lowland ($p < 0.001$) and Upland wood ($p = 0.003$), Infertile grassland ($p < 0.001$), and Moorland-grass mosaic ($p = 0.029$). Total mesofauna abundance in Lowland wood abundances was also greater than Heath/bog ($p = 0.039$; Fig. 3.2A). The effect of AVC on total invertebrate catch (mesofauna plus others) was also highly significant ($F_{7, 515} = 5.487$, $p < 0.001$), following the same trends.

As with total mesofauna, AVC had a significant effect on oribatid mites ($F_{7, 515} = 13.352$, $p < 0.001$). Again, the highest abundances of oribatid mites were found in Lowland wood, and lowest in Crops/weeds. Abundances were significantly lower in Crops/weeds and Fertile grassland than all other AVCs except Tall grass and herb ($p = 1.0$; $p = 0.995$, respectively). Additionally, abundances were significantly greater in Lowland wood than in Tall grass and herb ($p = 0.025$) and Infertile grassland ($p = 0.005$) AVCs (Fig. 3.2B). Though abundances of Mesostigmata differed significantly by AVC ($F_{7, 515} = 8.874$, $p < 0.001$), such differences were not consistent with the overall trend (Fig. 3.2C). Numbers of mesostigmatid mites were significantly lower in Moorland-grass mosaic and Heath/bog than Fertile (both $p < 0.001$) and Infertile grassland (both $p <$

0.001), as well as Upland wood ($p = 0.023$, $p < 0.001$, respectively). Abundances in Heath/bog were also significantly lower than in Lowland wood ($p = 0.014$).

Abundances of Collembola also did not follow the overall trend. The abundances in Entomobryoidea and Poduroidea abundances did show similar differences between AVCs ($F_{7, 515} = 5.716$, $p < 0.001$; $F_{7, 515} = 5.966$, $p < 0.001$, respectively). Abundances of Symphypleona were negligible across all AVC's. Entomobryoidea abundances were significantly greater in Lowland wood than in Fertile ($p = 0.037$) and Infertile grassland ($p = 0.047$), Moorland-grass mosaic ($p = 0.018$), Crops/weeds ($p = 0.001$), and Heath/bog ($p < 0.001$). Abundances in Crops/weeds ($p = 0.027$) and Heath/bog ($p < 0.001$) were significantly lower than in Upland wood. Additionally, abundances in Heath/bog were also significantly lower than Infertile ($p = 0.006$), and Fertile grassland ($p = 0.042$; Fig. 3.2D). Abundances of Poduroidea were significantly lower in Crops/Weeds ($p = 0.008$), Moorland-grass mosaic ($p = 0.017$), and Heath/bog ($p < 0.001$) AVCs than Lowland wood. Furthermore, abundances in Heath/bog were also significantly lower than both grasslands (both $p < 0.001$), Moorland-grass mosaic ($p = 0.01$), and Upland wood ($p = 0.001$; Fig. 3.2E).

AVC had a significant ($F_{7, 515} = 14.1816$, $p < 0.001$) effect on of H' diversity values (Fig 3.2F). This was likely driven by changes in Collembola and mesostigmatid abundance. Values were significantly lower in Crops/weeds and Heath/bog when compared both Fertile ($p = 0.033$; $p < 0.001$, respectively) and Infertile grassland ($p = 0.029$; $p < 0.001$, respectively). Additionally H' values in Infertile grassland were significantly greater than in Moorland-grass mosaic ($p < 0.001$). Lowland wood values were significantly greater than Heath/bog ($p < 0.001$) and Moorland-grass mosaic

(0.012). Heath/bog H' values were also significantly lower than Moorland-grass mosaic ($p = 0.007$) and Upland wood ($p < 0.001$).

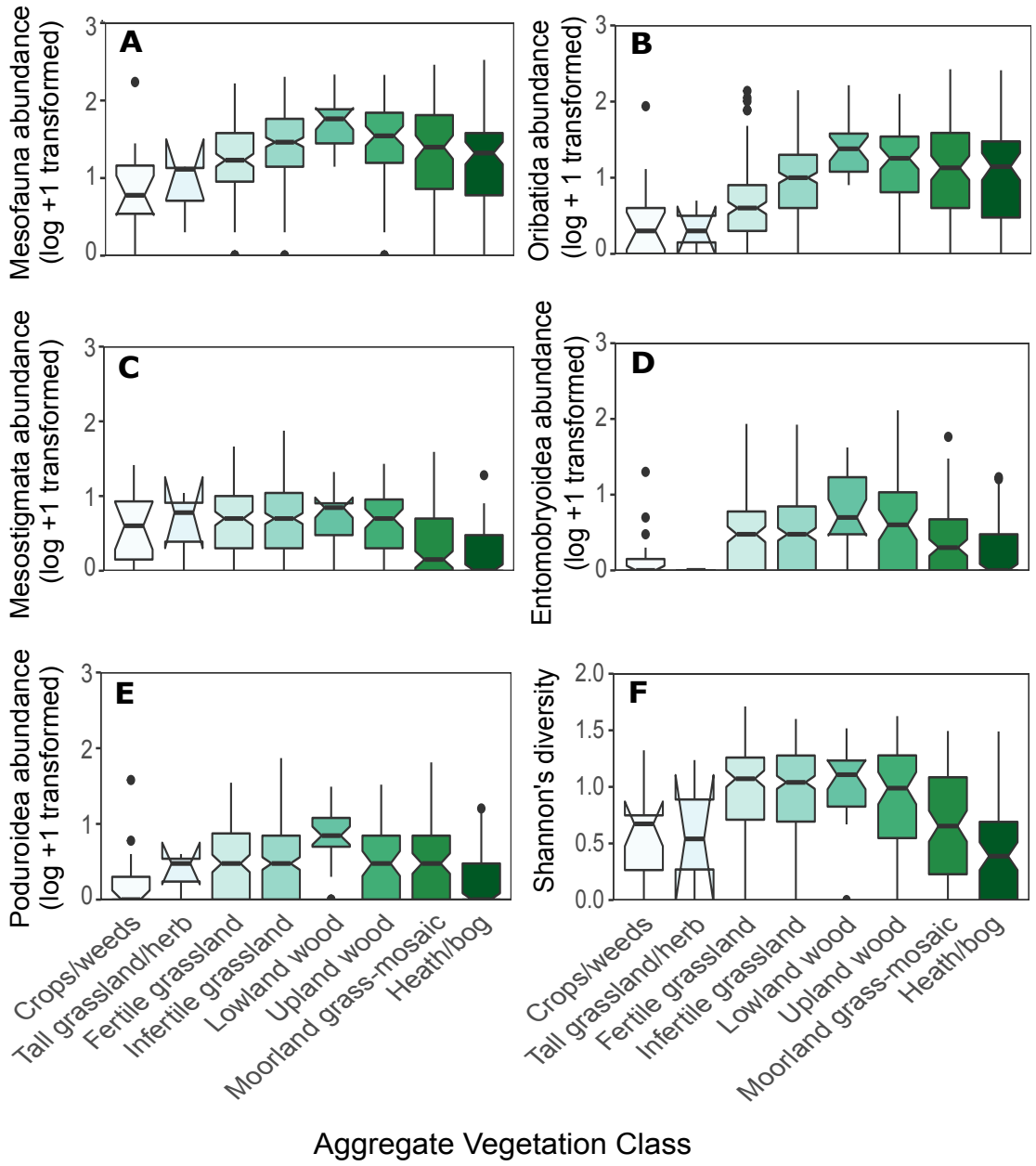


*Figure 3.2 Boxplots of **A)** total mesofauna; **B)** Oribatida; **C)** Mesostigmata; **D)** Entomobryoidea; **E)** Poduroidea; **F)** Shannon's diversity plotted against Aggregate Vegetation Class. All abundances are log_{10} plus one transformed. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers. AVC's are ordered from most (Crops/weeds) to least (Heath/bog) productive.*

Differences between soil types and LOI classes

Soil type had detectable effects on the Mesostigmata ($F_{6, 516} = 4.344$, $p < 0.001$), Entomobryoidea ($F_{6, 516} = 3.098$, $p = 0.006$), and Poduroidea ($F_{6, 516} = 2.340$, $p = 0.031$; Fig. 3). Mesostigmatid abundances were greater in brown soils than peat ($p < 0.001$) and surface-water gley soils ($p = 0.005$; Fig. 3.3C). Entomobryidea and Poduroidea abundances were also higher in brown soils than in peats ($p = 0.009$; $p = 0.043$, respectively; Fig 3.3D, 3.3E). These differences are reflected in H' values ($F_{6, 516} = 5.955$, $p < 0.001$), where the same differences can be seen (brown soils-peats: $p < 0.001$; brown soils-surface-water gleys: $p = 0.004$), in addition to a significant difference between podzolic and peat soils ($p = 0.002$) (Fig. 3.3F).

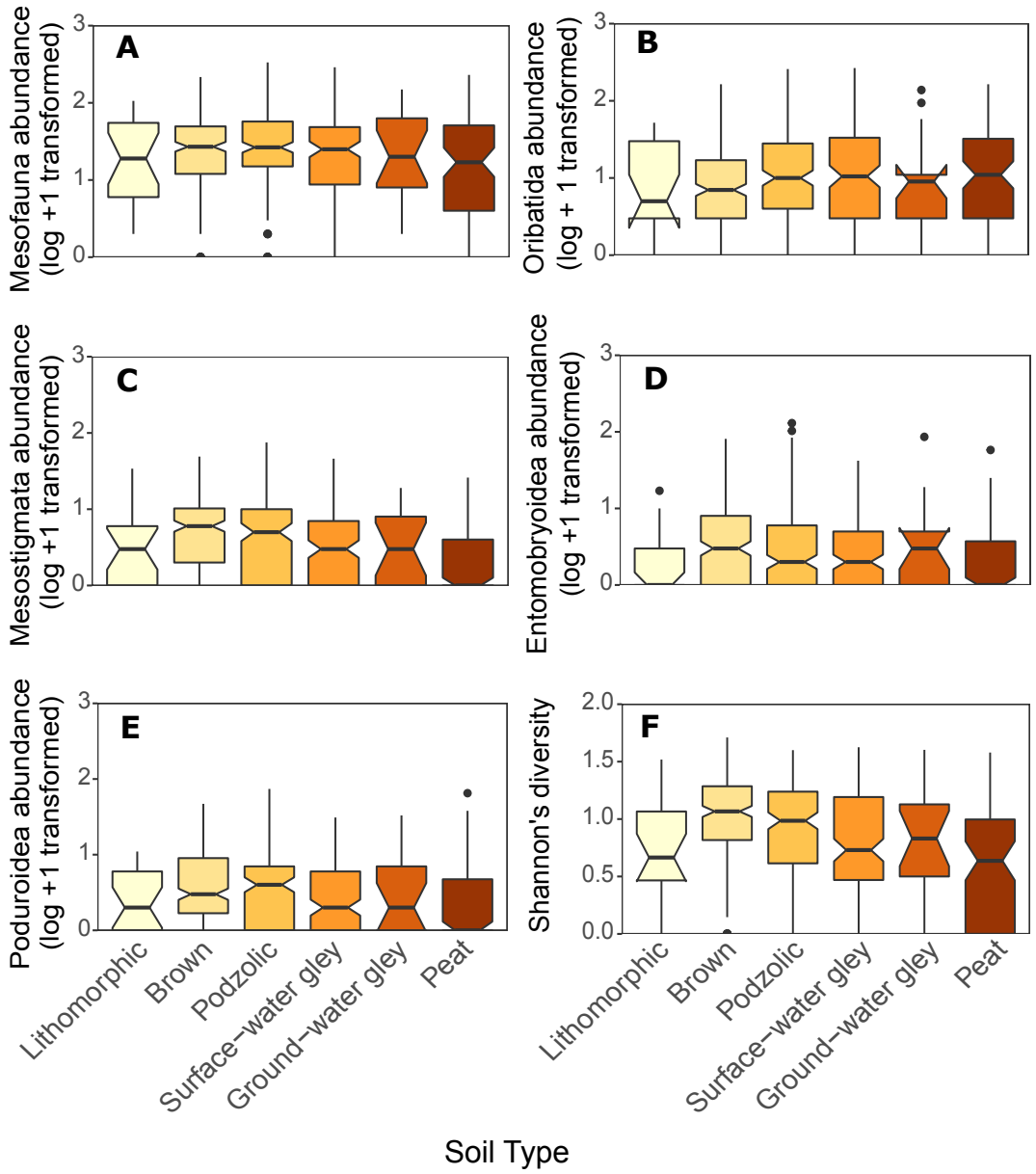


*Figure 3.3 Boxplots of **A)** total mesofauna; **B)** Oribatid mites; **C)** Mesostigmatid mites; **D)** Entomobryoidea; **E)** Poduroidea; **F)** Shannon's diversity plotted against soil type. All abundances are log_{10} plus one transformed. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers. Soils are listed in approximate order of increasing soil moisture content.*

Differences in mesofauna abundance between LOI classes were more informative. Differences were observed for total mesofauna ($F_{3, 518}$ = 3.973, p = 0.008;

Fig. 4A), total invertebrates ($F_{3, 518} = 3.985$, $p = 0.008$), and oribatid mites $F_{3, 518} = 7.742$, $p < 0.001$). Here, abundances were significantly higher in humus-mineral than in mineral soils ($p = 0.026$; $p = 0.030$; $p < 0.001$, respectively). Oribatid abundances were also significantly greater in organo-mineral soils than mineral soils ($p = 0.007$) and in lower organic than mineral soils ($p < 0.001$; Fig. 4B).

The effect of LOI class on abundance was the same for Mesostigmata ($F_{3, 518} = 11.979$, $p < 0.001$) and Entomobryoidea ($F_{3, 518} = 7.358$, $p < 0.001$). Here, abundances were significantly lower in organic soils than humus-mineral, mineral (all $p < 0.001$), and organo-mineral ($p = 0.022$, $p = 0.036$, respectively) soils (Fig. 4B, 4C). A similar trend was observed in Poduroidea abundances ($F_{3, 518} = 9.964$, $p < 0.001$). However, in this case, organic soils only had significantly lower abundances than humus-mineral ($p < 0.001$) and mineral ($p = 0.009$) soils (Fig. 4E). LOI class significantly ($F_{3, 518} = 29.093$, $p < 0.001$) affected diversity values, being significantly greater in humus-mineral ($p < 0.001$), mineral, ($p < 0.001$) and organo-mineral ($p = 0.009$) soils than in organic soils. There were also significant differences between organo-mineral soils and both mineral ($p = 0.022$) and humus-mineral soils ($p = 0.001$; Fig. 4F).

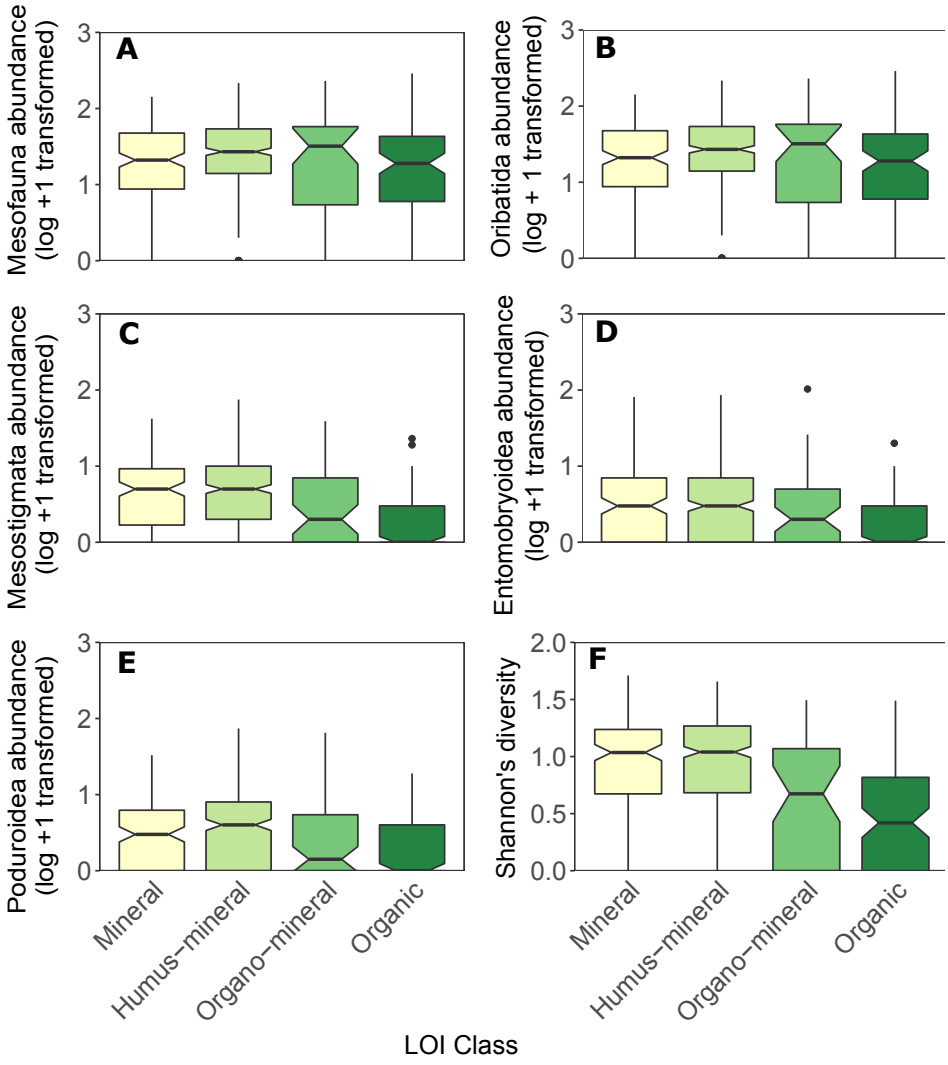


*Figure 3.4 Boxplots of **A)** mesofauna; **B)** Oribatid mites; **C)** Mesostigmatid mites; **D)** Entomobryoidea; **E)** Poduroidea; **F)** Shannon's diversity for each loss-on-ignition (LOI) class. All abundances are log_{10} plus one transformed. Notches indicate confidence interval around the median. Black dots are outliers. Overlapping notches are a proxy for non-significant differences between medians. LOI classes are listed in order of increasing soil organic matter content.*

3.3.3 Correlates with soil physical and chemical variables

Oribatid mites showed significant correlations with every soil property analysed except soil moisture content (Table 3.1). Positive relationships were found between oribatid mites and total C, total N, C:N ratio, and soil water repellency; negative relationships were found between oribatid mites and pH and total P (Table 3.1). The

oribatids were the only group to have a significant relationship with soil water repellency (Fig. 3.5). Total mesofauna correlated negatively with moisture content and pH, and positively with soil water repellency. Mesostigmata had significant positive relationships with bulk density and pH, and had significant negative relationships were present for total C, total N, C:N ratio, and moisture content (Table 3.1). Entomobryoidea and Poduroidea displayed negative relationships with total C, total N, C:N ratio, and soil moisture content. Both groups only had significant positive correlations with bulk density (Table 3.1).

*Table 3.1 Spearman correlations rho values for correlations with abundance of mesofauna groups with soil physical and chemical variables. *** indicates $p < 0.001$, ** $0.001 > p < 0.01$, * $0.01 > p < 0.05$, blank indicates $p > 0.05$.*

Soil variable	Total mesofauna	Oribatida	Mesostigmata	Entomobryoidea	Poduroidea
Total C (%)	0.018	0.190***	-0.244***	-0.151***	-0.123***
Total N (%)	-0.007	0.144***	-0.233***	-0.152***	-0.123***
C:N ratio	0.051	0.244***	-0.256***	-0.120***	-0.132***
pH (0.01 M $CaCl_2$)	-0.122***	-0.317***	0.118***	0.043	0.024
Bulk density (g/cm)	-0.056	-0.233***	0.216***	0.126***	0.089*
Soil water repellency[a]	0.142***	0.267***	-0.060	-0.48	-0.029
Volumetric water content (m^3/m^3)	-0.175***	0.006	-0.381***	-0.235***	-0.286***
Total P (mg/kg)	-0.054	-0.127***	0.054	0.001	0.022

[a] Soil water repellency was derived from median water drop penetration times (s) and log transformed

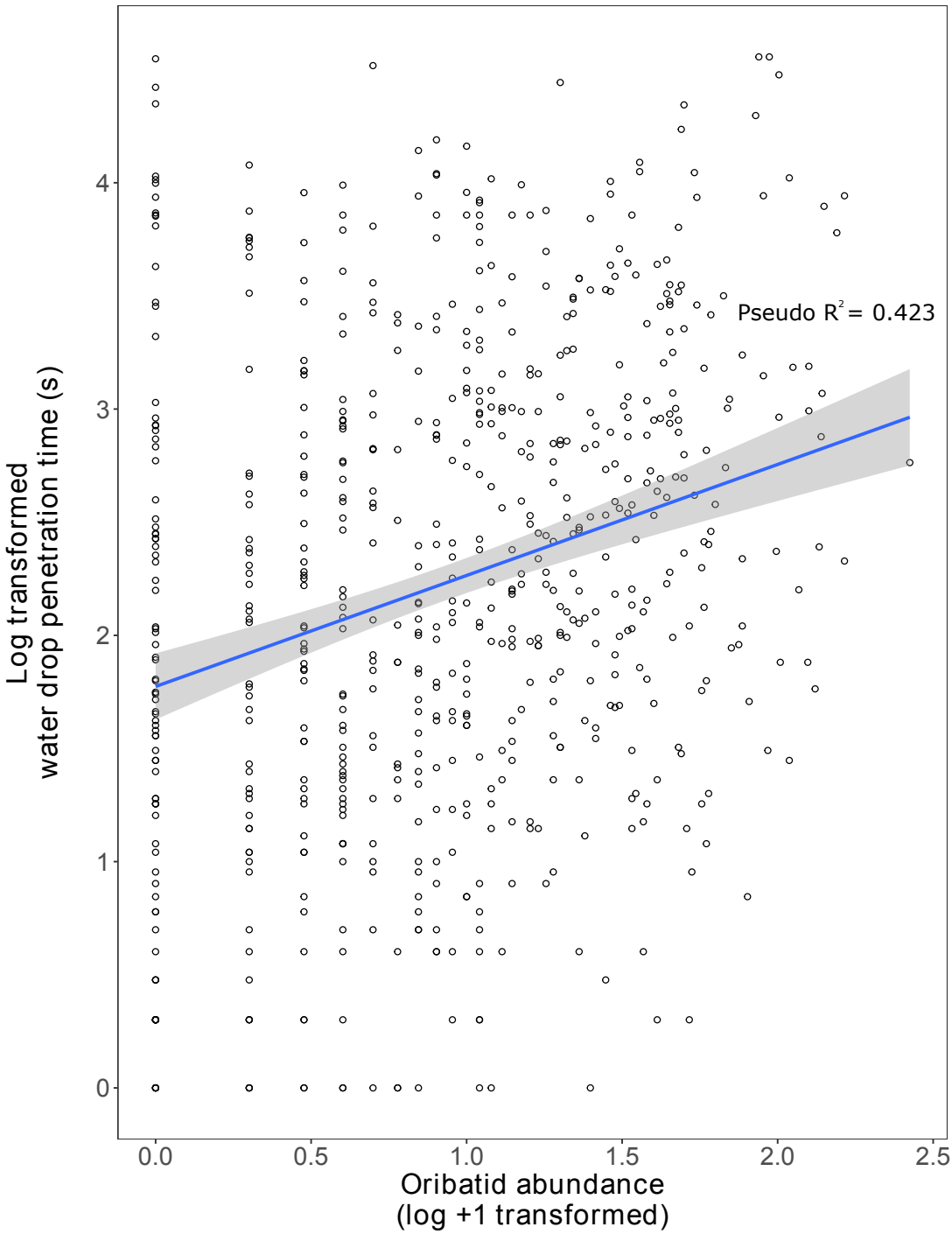


Figure 3.5 Scatterplot and linear regression line of log_{10} plus one transformed oribatid abundances versus log-transformed soil water repellency ($log_{10}(s)$) from all sample sites. Grey area around regression line represents 95% confidence interval. Pseudo-R^2 value was calculated using the R package "Piecewise SEM" (Lefcheck, 2015).

3.4 Discussion

3.4.1. Trends in mesofauna communities

Total abundance and diversity values were consistently lower in arable sites. A range of studies has shown that mite and Collembola abundances decline in agricultural habitats when compared to more extensive habitats (de Vries et al., 2013; Arroyo et al., 2013; Tsiafouli et al., 2015). For example, Tsiafouli et al. (2015) demonstrated that mite and Collembola diversity and biomass decline with increasing agricultural land-use intensity across a range of European sites. Mites and Collembola (Behan-Pelletier, 2003; Tsiafouli et al., 2015) are generally susceptible to disturbance, which has been seen across Europe (Postma-Blaauw et al., 2010; de Vries et al., 2013), North America (Behan-Pelletier, 2003), and Australia (Osler and Murphy, 2005).

LOI classification was found to be more informative than soil type when examining differences in mesofauna. This is likely an artefact of the resolution and accuracy of how soils were classified. Soil types were inferred from the major groups defined by Avery (1990) associated with the series listed for each sample location on the National Soil Map (see Supplementary Material). Alternatively, LOI classification was derived from co-located plot data and may provide more important ecological trends than traditional mapped soil taxonomy.

3.4.2. Soil properties and oribatid mites

The negative correlation in oribatid mite abundance with pH and bulk density, in addition to the positive relationship with soil organic matter observed in the present study

is consistent with results from Ireland (Arroyo et al., 2013). Oribatid mites are sensitive to agricultural practices, primarily due to life-history characteristics such as low fecundity, and relatively long generation times (Behan-Pelletier, 1999). Soil compaction and litter removal have been shown to lower oribatid densities in forest plantations (Battigelli et al., 2004) and both processes commonly occur under conventional agricultural management.

Oribatid mites were the only group to correlate with soil water repellency. Although soil water repellency is not commonly studied in relation to mesofauna, it is known that soils rich in fungi are often hydrophobic (Hallett et al., 2001; Rillig et al., 2010). Many species of oribatid mites are fungivorous (Behan-Pelletier, 1999). We suspect that this relationship may be indicative of soils with high fungal abundance. Further research using microbial data could therefore be used to explore if there is a similar correlation of between soil hydrophobicity and fungi, likely to be driven by filamentous species (Rillig, 2005).

Oribatid mites had significant, positive relationships with total N and C:N ratio. This is contrary to research by Cole et al. (2008), who found positive interactions with Collembola abundance and total N, and no relationship with oribatid abundance. However, others have shown that many oribatid taxa may be tolerant of increased soil N at low or medium levels (Seniczak et al., 1998). The positive relationship with C:N ratio suggests that Welsh oribatid populations are predominantly fungivorous, whereas the other groups are either obviously predatory or might favour bacterivorous or omnivorous diets (Osler and Sommerkorn, 2007). Oribatids were negatively correlated with total P, which is consistent with a previous study by Schon et al. (2011), who found P additions

led to declining oribatid abundances. This relationship may be indicative of a shift towards intensive agriculture.

3.4.3 Trends in Mesostigmata and Collembola populations

Interestingly, Collembola abundances were as low in Heath/bog sites as they were in Crops/weeds. Most Heath/bog sites were located in upland regions. The Welsh uplands include at-risk habitats such as peatlands, which are sensitive to disturbance (Reed et al., 2009), tend to be colder and have a higher frequency of precipitation as well as N deposition than lowland habitats (Kirkham, 2001). Temperature and moisture level have been identified as stressors for Collembola communities. Choi et al. (2002) found development of temperate Collembola can be halted by temperatures lower than 5 °C. Sustained elevated N deposition in American forest soils has been shown to reduce Collembola densities (Gan et al., 2013). Increased frequency and severity of precipitation has also been demonstrated to drive down Collembola richness and abundance in both mesocosm (Turnbull and Lindo, 2015) and *in situ* experiments (Tsiafouli et al., 2005). Furthermore, populations of peatland Collembola have been shown to increase when peatlands are drained for forestry (Silvan et al., 2000).

Unexpectedly, mesostigmatid abundances only declined in Moorland-grass mosaic and Heath/bog sites. Mesostigmatids had the same trends with bulk density, and moisture content as Collembola but were also positively correlated with pH. It is likely that prey availability has the strongest influence on mesostigmatid abundance (Nielsen et al., 2010a; Nielsen et al., 2010b). Decreased prey abundance (i.e. Collembola, Nielson et al., 2010a) could have limited their populations in moist upland habitats. Higher abundances in primarily agricultural areas may have been maintained through

consumption of unsurveyed prey such as nematodes (Koehler, 1997), as predatory mites in arable habitats are often generalists or omnivores (Postma-Bloouw et al., 2010).

3.4.3. Implications for national-level soil monitoring

The approach employed by GMEP is efficient and cost-effective and the addition of a separate mesofauna core for each site does not add considerably to sampling effort for a monitoring programme (Emmett and the GMEP Team, 2014; 2015). This study shows that meaningful conclusions can be drawn from a nationwide mesofauna dataset collected using a relatively simple, standardised methodology. Yet, an effective bioindicator must be sufficiently informative and comparable between disparate research groups whilst overcoming inherent trade-offs with required sampling effort and expertise.

Trends observed in the present study highlight some important shortcomings of using broad-groups of mesofauna as bioindicators. Differences between AVCs were most commonly observed in those with extreme differences in disturbance levels or plant communities, such as Crops/weeds, Lowland wood, and Heath/bog. Indeed, using small, subterranean fauna to inform habitat classifications is likely an over-complicated methodology, when aboveground plant community assessments are easier and more informative. Indeed, our methodology could not consistently detect community changes amongst the grassland and agricultural AVCs. This means that results of agricultural interventions focused on conversion to semi-natural grassland or extensification may not be evident in national soil surveys. Furthermore, the relationship between abundances and soil type was not clear and challenging to interpret.

Comparing trends from nationwide data sets to the literature also presents challenges. The majority of the research published on the interaction of mesofauna and soil properties focuses on the habitat or microhabitat scale. The trends presented here represent an entire habitat gradient and in many cases are driven by specific AVCs. For instance, the relationship between soil water repellency and oribatid mite abundance is driven by grassland AVCs. It should also be noted that working on a national-scale leads to discrepancies in replication. For example, in our dataset, the Tall grass/herb AVC was only observed three times, making any trends in this habitat unreliable. Conversely, an overabundance of habitat types in a national survey may obscure interesting trends in unique or rare systems. Thus, it may be necessary to subsample data from national surveys by habitat to find comparable data.

Broad-groups of mesofauna were however, informative of relationships using locally derived soil data. Relationships with soil properties, though potentially obscured when taken as a whole, allow for important insights into the ecological implications of changes in the environment. Similarly, comparing mesofauna groups to soils classified by LOI percentage on a national-level revealed trends that better inform us of the ecological meaning behind distributions of traditional soil taxonomic classifications. It is possible that further classification of mesofauna to species-level could be more informative. We chose not to do this, as previous British surveys have focused on trends in broad-groups only (Black et al., 2003). National monitoring has an added benefit of creating a national inventory of taxonomic specimens from which further research can be conducted and from which more species may be described. Additionally, reference collections of identified mesofauna provide a strong starting-point for studies using

metabarcoding style analyses (Creer et al., 2016). Greater confidence can be given to database matches of mesofauna from community and environmental DNA and DNA from identified specimens can be uploaded to databases (Ratnasingham and Hebert, 2007) to build more complete reference libraries. It is important to remember that comparisons of new molecular datasets will require reference to historical taxonomic data strengthening the case for acquiring new reference materials as part of monitoring. Therefore, we suggest that the addition of broadly grouped mesofauna surveys to any national soil monitoring programme will be an important compliment to the assessments of soil properties and biodiversity.

3.5 Conclusions

Our results show that at the national-level, mesofauna populations have the potential to be effective environmental indicators, through their consistent sensitivity to differences in habitat, plot-level soil class, and soil physical characteristics. This research represents an important first step to assess agri-environment schemes and land-use change. The present study was one of the most extensive nationwide surveys of mesofauna in Europe. The results show that conventional stresses from agriculture can be observed across the country with relatively low sampling effort. Additionally, it has revealed trends in Collembola and mesostigmatid mites in highly sensitive upland areas. Such results may be of use to policy-makers and land-managers actively trying to maintain a balance between rural development and natural values. The sampling design used here has been shown to be effective for analysis of disparate habitat types. However, further refinements are needed to separate similar habitat types and to understand relationships with soil type as defined by the National Soil Map. We encourage that this

approach be included in regional- to national-level soil monitoring programmes to better inform researchers of the ecological implications of changing soil properties. With more adoption of mesofauna in monitoring plans, a more complete picture of the roles of mesofauna as bioindicators will be made.

3.6 Acknowledgements

This research was supported under the Glastir Monitoring & Evaluation Programme. (Contract reference: C147/2010/11) and by the UK Natural Environment Research Council (NERC) through the RC/Centre for Ecology & Hydrology (CEH Project: NEC04782). We thank the GMEP team for their contribution in collecting the data. PBLG is funded by Bangor University through the NERC STARS PhD Partnership Studentship.

3.7 References

Arroyo, J., Keith, A.M., Schmidt, O., Bolger, T., 2013. Mite abundance and richness in an Irish survey of soil biodiversity with comments on some newly recorded species. Irish Naturalists' Journal 33, 19-27.

Avery, B.W., 1990. Soils of the British Isles. CAB International, Wallingford.

Barrios, E., 2007. Soil biota, ecosystem services and land productivity. Ecological Economics 64, 269-285.

Battigelli, J.P., Spence, J.R., Langor, D.W., Berch, S.M., 2004. Short-term impact of forest soil compaction and organic matter removal on soil mesofauna density and oribatid mite diversity. Canadian Journal of Forest Research 34, 1136-1149.

Behan-Pelletier, V.M., 1999. Oribatid mite biodiversity in agroecosystems: role for biodindication. Agriculture, Ecosystems and Environment 74, 411-423.

Behan-Pelletier, V.M., 2003. Acari and Collembola biodiversity in Canadian agricultural soils. Canadian Journal Soil Science 83, 279-288.

Black, H.I.J., Parekh, N.R., Chaplow, J.S., Monson, F., Watkins, J., Creamer, R., Potter, E.D., Poskitt, J.M., Rowland, P., Ainsworth, G., Hornung, M., 2003. Assessing soil

biodiversity across Great Britain: national trends in the occurrence of heterotrophic bacteria and invertebrates in soil. Journal of Environmental Management 67, 255-266.

Bunce, R.G.H., Barr, C.J., Gillespie, M.K., Howard, D.C.H., Scott, W.A., Smart, S.M., van der Poll, H.M., Watkins, J.W., 1999. Vegetation of the British countryside – the Countryside Vegetation System. ECOFACT Volume 1. DETR, London, 224 pp.

Choi, W.L., Ryoo, M.I., Kim, J.-G., 2002. Biology of *Parnonychiurus kimi* (Collembola: Onychiuridae) under the influence of temperature, humidity and nutrition. Pedobiologia 46, 548-577.

Cole, L., Buckland, S.M., Bardgett, R.D., 2008. Influence of disturbance and nitrogen addition on plant and soil animal diversity in grassland. Soil Biology & Biochemistry 40, 505-514.

Cluzeau, D., Guernion, M., Chaussod, R., Martin-Laurent, F., Villenave, C., Cortet, J., Ruiz-Camacho, N., Pernin, C., Mateille, T., Philippot, L., Bellido, A., Rougé, L., Arrouays, D., Bispo, A., Pérès, G., 2012. Integration of biodiversity in soil quality monitoring: baselines for microbial and soil fauna parameters for different land-use types. European Journal of Soil Biology 49, 63-72.

Creer, S., Deiner, K., Frey, S., Porazinska, D., Taberlet, P., Thomas, W.K., Potter, C., Bik, H.M., 2016. The ecologist's field guide to sequence-based identification of biodiversity. Methods in Ecology and Evolution 7, 1008-1018.

Crotty, F.V., Shepherd, M., 2014. A Key to the Soil Mites of Britain and Ireland (Test Version 5). Field Studies Council, Telford.

Doran, J.W., Zeiss, M.R., 2000. Soil health and sustainability: managing the biotic component of soil quality. Applied Soil Ecology 15, 3-11.

Emmett, B.A., the GMEP team, 2014. Glastir Monitoring & Evaluation Programme. First Year Annual Report to Welsh Government (Contract reference: C174/2010/11). NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780), pp. 442.

Emmett, B.A., the GMEP team, 2015. Glastir Monitoring & Evaluation Programme, Second Year Annual Report to Welsh Government (Contract Refernce: C147/2010/11). NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780), pp 316.

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J., Potter, E., Robinson, D.A., Scott, A., Wood, C., Woods, C., 2010. Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology, 192pp. (CS Technical Report No. 9/07, CEH Project Number: C03259).

Gan, H., Zak, D.R., Hunter, M.D., 2013. Chronic nitrogen deposition alters the structure and function of detrital food webs in a northern hardwood ecosystem. Ecological Applications 23, 1311-1321.

Gerlach, J., Samways, M., Pryke, J., 2013. Terrestrial invertebrates as bioindicators: an overview of available taxonomic groups. Journal of Insect Conservation 17, 831-850.

Graham, E.B., Knelman, J.E., Schindlbacher, A., Siciliano, S., Breulmann, M., Yannarell, A., Beman, J.M., Abell, G., Philippot, L., Prosser, J., Foulquier A., Yuste, J.C., Glanville, H.C., Jones, D.L., Angel, R., Salminen, J., Newton, R.J., Bürgmann, H., Ingram, L.J., Hamer, U., Siljanen, H.M.P., Peltoniemi, Potthhast, K., Bañeras, L. Hartmann, M., Banerjee, S., Y. Yu R.-Q., Nogaro, G., Richter, A., Koranda, M., Castle, S.C., Goberna, M., Song, B., Chatterjee, A., Nunes, O.C., Lopes, A.R., Cao, Y., Kaisermann, A., Hallin, S., Strickland, M.S., G.-P., J., Barba, J., Kang, H., Isobe, K., Papaspyrou, S., Pastorelli, R., Lagomarsino, A., Lindström, E.S., Basiliko, N., Nemergut, D.R., 2016. Microbes as engines of ecosystem function: when does community structure enhance predictions of ecosystem processes? Frontiers in Microbiology 7, 214.

Gulvik, M.E., 2007. Mites (Acari) as indicators of soil biodiversity and land use monitoring: a review. Polish Journal of Ecology 55, 415-440.

Hallett, P.D., Ritz, K., Wheatley, R.E., 2001. Microbial derived water repellency in golf course soil. International Turfgrass Society Research Journal 9, 518-524.

Havlicek, E., 2012. Soil biodiversity and bioindication: from complex thinking to simple acting. European Journal of Soil Biology 49, 80-84.

Hopkin, S.P., 2007. A Key to the Collembola (Springtails) of Britain and Ireland. Field Studies Council, Telford, UK.

Keith, A.M., Boots, B., Hazard, C., Niechoj, R., Arroyo, J., Bending, G.D., Bolger, T., Breen, J., Clipson, N., Doohan, F.M., Griffin, C.T., Schmidt, O., 2012. Cross-taxa congruence, indicators and environmental gradients in soils under agricultural and extensive land management. European Journal of Soil Biology 49, 55-62.

Kirkham, F.W., 2001. Nitrogen uptake and nutrient limitation in six hill moorland species in relation to atmospheric nitrogen deposition in England and Wales. Journal of Ecology 89, 1041-1053.

Koehler, H.H., 1997. Mesostigmata (Gamasina, Uropodina), efficient predators in agroecosystems. Agriculture, Ecosystems and Evironment 62, 105-117.

Lavelle, P., Decaëns, T., Aubert, M., Barot, S., Blouin, M., Bureau, F., Margerie, P., Mora, P., Rossi, J.-P., 2006. Soil invertebrates and ecosystem services. European Journal of Soil Biology 42, S3-S15.

Lefcheck, J.S., 2015. piecewiseSEM: piecewise structural equation modelling in R for ecology, evolution, and systematics. Methods in Ecology and Evolution 7, 573-579.

Nielsen, U.N., Osler, G.H.R., Campbell, C.D., Neilson, R., Burslem, D.F.R.P., van der Wal, R., 2010a. The influence of vegetation type, soil properties and precipitation on the composition of soil mite and microbial communities at the landscape scale. Journal of Biogeography 37, 1317-1328.

Nielsen, U.N., Osler, G.H.R., Campbell, C.D., Neilson, R., Burslem, D.F.R.P., van der Wal, R., 2010b. The enigma of soil animal species diversity revisited: the role small-scale heterogeneity. PLoS One 5, e11567.

Oksanen, J., Blanchet, F.G., Friendly, M., Kindt, R., Legendre, P., MacGlinn, D., Minchin, P.R., O'Hara, R.B., Simpson, G.L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., Wagner, H., 2016. vegan: Community Ecology Package. R package version 2.4-0. URL: https://CRAN.R-project.org/package=vegan.

Osler, G.H.R., Murphy, D.V., 2005. Oribatid mite species richness and soil organic matter fractions in agricultural and native vegetation soils in Western Australia. Applied Soil Ecology 29, 93-98.

Osler, G.H.R., Sommerkorn, M., 2007. Towards a complete soil C and N cycle: incorporating the soil fauna. Ecology 88, 1611-1621.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2016. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-128, URL: http://CRAN.R-project.org/package=nlme.

Postma-Blaauw, M.B., de Goede, R.G.M., Bloem, J., Faber, J.H., Brussard, L., 2010. Soil biota community structure and abundance under agricultural intensification and extensification. Ecology 91, 460-473.

R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/.

Ratnasingham, S., Hebert, P.D.N., 2007. BOLD: the barcode of life data system (www.barcodinglife.org). Molecular Ecology Notes 7, 355-364.

Reed, M.S., Bonn, A., Slee, W., Beharry-Borg, N., Birch, J., Burt, T.P., Chapman, D., Chapman, P.J., Clay, G.D., Cornell, S.J., Fraser, E.D.G., Glass, J.H., Holden, J., Hodgson, J.A., Hubacek, K., Irvine, B., Jin, N., Kirkby, M.J., Kunin, W.E., Moore, O., Moseley, D., Prell, C., Price, M.F., Quinn, C.H., Redpath, S., Reid, C., Stagl, S., Stringer, L.C., Termansen, M., Thorp, S., Towers, W., Worrall, F., 2009. The future of the uplands. Land Use Policy 265, S204-S216.

Rillig, M.C., 2005. A connection between fungal hydrophobins and soil water repellncy? Pedobiologia, 49, 395-399.

Rillig, M.C., Mardatin, N.F., Leifheit, E.F., Antunes, P.M., 2010. Mycelium of arbuscular mycorrhizal funi increases soil water repellency and is sufficient to maintain water-stable soil aggregates. Soil Biology & Biochemistry 42, 1189-1191.

Rutgers, M., Schouten, A.J., Bloem, J., van Eekeren, N., de Goede, R.G.M., Jagers op Akkerhuis, G.A.J.M., van der Wal, A., Mulder, C., Brussaard, L., Breure, A.M., 2009. Biological measurements in a nationwide soil monitoring network. European Journal of Soil Science 60, 820-832.

Rusek, J., 1998, Biodiversity of Collembola and their functional role in the ecosystem. Biodiversity and Conservation 7, 1207-1219.

Schon, N.L., Mackay, A.D., Gray, R.A.J., Minor, M.A., 2011. Influence of phosphorus inputs and sheep treading on soil macrofauna and mesofauna in hill pastures. New Zealand Journal of Agricultural Resarch 54, 83-96.

Schneider, K., Renker, C, Maraun, M., 2005. Oribaitd mite (Acari, Oribatida) feeding on ectomycorrhizal fungi. Mycorrhiza 16, 67-72.

Seniczak, S., Dabrowski, J., Klimek, A., Kaczmarek, S., 1998. Effects of air pollution produced by a nitrogen fertilizer factory on the mites (Acari) associated with young Scots pine forests in Poland. Applied Soil Ecology 9, 453-458.

Silvan, N., Laiho, R., Vasander, H., 2000. Changes in mesofauna abundance in peat soils drained for forestry. Forest Ecology and Management 133, 127-133.

Tsiafouli, M.A., Kallimanis, A.S., Katana, E., Stamou, G.P., Sgardelis, S.P., 2005. Responses of soil microarthropods to experimental short-term manipulations of soil moisture. Applied Soil Ecology 29, 17-26.

Tsiafouli, M.A., Thébault, E., Sgardelis, S.P., de Ruiter, P.C., van der Putten, W.H., Birkhofer, K., Hemerik, L., de Vries, F.T., Bardgett, R.D., Brady, M.V., Bjornlund, L., Jørgensen, H.B., Christensen, S., d'Hertefeldt, T., Hotes, S., Hol, W.H. G., Frouz, J., Liiri, M., Mortimer, S.R., Setälä, H., Tzanopoulos, J., Uteseny, K., Pižl, V., Stray, J., Wolters, V., Hedlund, K., 2015. Intensive agriculture reduces soil biodiversity across Europe. Global Change Biology 21, 973-985.

Turnbull, M.S., Lindo, Z., 2015. Combined effects of abiotic factors on Collembola communities reveal precipitation may act as a disturbance. Soil Biology & Biochemistry 82, 36-43.

de Vries, F.T., Thébault, E., Liiri, M., Birkhofer, K., Tsiafouli, M.A., Bjørnlund, L., Jørgensen, H.B., Brady, M.V., Christensen, S., de Ruiter, P.C., d'Hertefeldt, T., Frouz, J., Hedlund, K., Hemerik, L., Hol., W.H.G., Hotes, S., Mortimer, S.R., Setälä, H., Sgardelis, S.P., Uteseny, K., van der Putten, W.H., Wolters, V., Bardgett, R.D., 2013. Soil food web properties explain ecosystem services across European land use systems. Proceedings of the National Academy of Sciences 110, 14296-14301.

Welsh Government, 2016. How to measure a nation's progress? National indicators for Wales. Hawlfraint y Goron, WG26817, ISBN Digidol 978 1 4734 6309 7.

Winter, J.P., Behan-Pelletier, V.M., 2007. Microarthropods In: Carter, M.R., Gregorich, E.G. (Eds.), Soil Sampling and Methods of Analysis 2nd edition. CRC press, Taylor & Francis Group LLC, Boca Raton, pp. 399-414.

Chapter 4: Diverging trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems

This manuscript has been published in *Nature Communications*

George, P.B.L.[1,2*], Lallias, D.[3], Creer, S.[1], Seaton, F.M.[2,1], Kenny, J.G.[4], Eccles, R.M.[4], Griffiths, R.I.[2], Lebron, I.[2], Emmett, B.A.[2], Robinson, D.A.[2], Jones, D.L.[1,5], 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nat. Commun.* 10, 1107. doi: 10.1038/s41467-019-09031-1

[1] School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] GABI, INRA, AgroParis Tech, Université Paris-Saclay, 78350 Jouy-en-Josas, France

[4] Centre for Genomic Research, University of Liverpool, Liverpool, L69 7ZB, UK

[5] School of Agriculture and Environment, University of Western Australia, Crawley, WA 6009, Australia

[*] Corresponding author

D.L.J., D.A.R., S.C., and B.A.E. conceived this project. G.L.B., D.L., D.L.J., and I.L. processed the soil samples and collected data with quality assurance by F.M.S.. D.L. and S.C. led the DNA extraction, primer design, library generation, and established the bioinformatics pipelines. J.K. and R.E. led sequence data generation and assisted with library preparations. Bioinformatics and statistical analyses were led by P.B.L.G. with assistance from S.C., R.I.G., D.L., and F.M.S.. P.B.L.G. wrote the first draft of the manuscript and F.M.S., S.C., D.L., D.A.R., and D.L.J. contributed to subsequent revisions. All authors read and approved the final draft of the manuscript.

Abstract

Soil biota account for ~25% of global biodiversity and is vital to nutrient cycling and primary production. There is growing momentum to study total belowground biodiversity across large ecological scales to understand how habitat and soil properties shape belowground communities. Microbial and animal components of belowground communities follow divergent responses to soil properties and land use intensification; however, it is unclear whether this extends across heterogeneous ecosystems. Here, a national-scale metabarcoding analysis of 436 locations across 7 different temperate ecosystems shows that belowground animal and microbial (bacteria, archaea, fungi, and protists) richness follow divergent trends, whereas β-diversity does not. Animal richness is governed by intensive land use and unaffected by soil properties, while microbial richness was driven by environmental properties across land uses. Our findings demonstrate that established divergent patterns of belowground microbial and animal diversity are consistent across heterogeneous land uses and are detectable using a standardised metabarcoding approach.

Key words: *Soil biodiversity; eDNA; Metabarcoding; Protists; Archaea*

4.1 Introduction

Soil biota, including bacteria, archaea, protists, fungi and animals underpin globally important ecosystem functions. Fundamental functions of soil communities include nutrient and hydrological cycling, decomposition, pollution mitigation, and supporting terrestrial primary production, which are inextricably linked to global food security, climate regulation and other ecosystem services[1-2]. Nevertheless, until recently, characterising soil biodiversity (popularly referred to as a "black box") has been constrained by our inability to identify typically intractable levels of diversity using either traditional or molecular approaches. High-throughput sequencing has however resulted in a step change, facilitating the characterisation of bacteria[3-7], archaea[6-8], fungi[9-10], protists[11-13], and animals[14] within the belowground biosphere. Increasingly, efforts have been made to investigate the total biodiversity of the soil biosphere across large ecological[15-17] and taxonomic scales[15-16,18-19].

Understanding the response of the total soil biosphere to changes in land use and environmental drivers has become an important research focus in regional soil monitoring programmes[15-16,19] and in small-scale field[20-21] and mesocosm experiments[18,20]. Yet despite the move towards unified study of soil biota, fundamental challenges of technique and scale remain. Often such studies require the comparison of soil biota metrics captured through both traditional and modern molecular techniques[15,19-21]. To our knowledge, relatively few studies have attempted to assess all components of belowground communities using a multi-marker metabarcoding approach[22].

There is mounting evidence that the microbial and animal fractions of soil communities may respond differentially to land use change. Microbial richness

increases[15], whereas richness of soil fauna declines in response to more intense land use[15,23-24]. However, these findings come from relatively homogenous landscapes, such as grasslands[15]. It is unclear whether the differential responses of soil microbes and fauna extend across heterogeneous land uses. For example, across heterogeneous landscapes of Wales, UK, α-diversity of mesofauna is both lowest in agricultural and bog systems, which are the most- and least-intensively managed systems in the country, respectively[23]. Changes in soil properties may further dictate declines of common soil fauna in low-intensity land uses. Therefore, it is critical to assess whether the positive effect of increasing land use intensity on microbial richness is consistent across regions made up of markedly diverse ecosystems and land uses. Similarly, the importance of individual soil properties in shaping belowground communities has also proven difficult to disentangle. Many studies have demonstrated the consistent dominance of pH in shaping belowground community composition at national [23,25-28] and global scales[4-5,9,29]. However, climatic factors[9,30] and other soil properties, including organic matter, nitrogen (N) availability, and the carbon (C)-to-N ratio[9] are also recognised as important drivers of belowground community composition yet consistent trends remain elusive[30]. Therefore it is unclear whether the total soil biosphere responds to changes in land use and soil properties in the same manner across heterogeneous landscapes.

Here, we sought to assess whether divergent responses to land use and soil properties in the microbial and animal fractions of soil communities persist across heterogeneous systems at the national-scale using a standardised metabarcoding approach. We present a national-scale analysis of soil biodiversity across Wales, UK, from the micro-to-macro scale including all major groups of soil microbes in addition to

animals, from 436 sites over 2 years across a diverse array of oceanic-temperate ecosystems, including grasslands, forests, bogs, and managed systems. Biotic metrics come from high-throughput sequencing of prokaryotic, fungal, microbial eukaryotic and soil animal communities using 16S, ITS, and 18S rRNA marker genes; these are complemented by an extensive suite of co-located abiotic soil properties and vegetation cover data. Specifically, we investigate how richness and β-diversity of all major fractions of subterranean life respond to land use type and prevailing soil properties (e.g. organic matter, pH, and N) to explore which lineages play a demonstrable role in determining belowground community structures across large and complex ecological gradients. Our results demonstrate that across a gradient of heterogeneous land uses, richness of soil animals is governed more by land use regime rather than intrinsic soil properties. In contrast, microbial richness is driven by soil properties and demonstrates a largely linear trend of decreasing richness along a productivity gradient of land use based on decreasing soil nutrient availability.

4.2 Results

4.2.1 Sequencing results

Illumina sequencing and environmental data was collected from across Wales as part of the Glastir Monitoring and Evaluation Programme (GMEP)[31]. Sample sites were categorised into Aggregate Vegetation Classes (AVCs) based on plant species assessments using established criteria (see Appendix 2.1). An explanation of the composition of AVCs is described in Appendix 2 Table 1. Briefly, the 7 AVCs used in the current study were established by clustering samples based on an assessment of vegetation data using a detrended correspondence analysis[32]. The ordination of the

detrended correspondence analysis has shown that the land use categories follow a gradient of soil nutrient content[32] from which soil productivity and management intensity can also be inferred (see Appendix 2.1 and Appendix 2 Table 1). The AVCs in descending order of productivity are: Crops/weeds, Fertile grassland, Infertile grassland, Lowland wood, Upland wood, Moorland grass-mosaic, and Heath/bog.

In total, 29,690 bacterial and 156 archaeal operational taxonomic units (OTUs) were identified from 16S reads. Overall, the most abundant class was Alphaproteobacteria (Fig. 4.1A). Proportional abundances (OTU n/total x 100) of Acidobacteria increased in less-productive land use types from its lowest in Crops/weeds to its highest in Heath/bog AVCs. In contrast abundances of Actinobacteria followed the exact opposite trend, as did Spartobacteria and Bacilli (Fig. 4.2A). For archaea, Nitrososphaeria was the most abundant class overall (Fig. 4.1D); however, the proportion of Thermoplasmata became dominant in less productive AVCs (Fig. 4.2D).

There were 7,582 OTUs recovered from ITS1 sequences. Agaricomycetes were the most abundant class of fungi overall. There were also a large proportion of Sordariomycetes (Fig. 4.1B). Proportionate abundances of Sordariomycetes and Agaricomycetes followed contrasting trends, with the dominance of the former replaced by the later in lower productivity AVCs (Fig. 4.2B).

In total, 8,683 protist OTUs were recovered from the 18S reads. Chloroplastida (green algae) was by far the most abundant protist group, followed by Rhizaria, Stramenopiles, and then Alveolates (Fig. 4.1C). Green algae, largely comprised of unidentified sequences (Appendix 2 Fig. 1A), were least abundant in Crops/weed and Heath/bog sites (Fig. 4.2C). Proportions of Rhizaria were relatively constant across

AVCs (Fig. 4.2C) and entirely comprised of Cercozoa (Appendix 2 Fig. 1B). Among Stramenopiles proportions of Ochrophyta were also largely consistent, while those of Oomycetes and Bicosoecida followed contrasting trends across the productivity gradient of AVCs, declining and increasing, respectively (Appendix 2 Fig. 1C). Ciliates were the most common Alveolates in most AVCs; however, the proportion of Apicomplexa was greater in the Lowland wood and grassland AVCs (Appendix 2 Fig. 1D). The proportion of Amoebozoa was surprisingly low (Fig. 4.1C), potentially due to primer bias in our study when compared to other studies[12,15]. Across AVCs Tublulinea was consistently dominant among the Amoebozoa, though divergent trends in Gracilipodida and Discosea can be seen along the productivity/intensity gradient (Appendix 2 Fig. 1E).

In the animal dataset, 1,138 OTUs were recovered. Nematode OTUs were the most abundant animal group across all samples (Fig. 4.1E). Annelids and arthropods followed opposing trends in proportionate abundance, increasing and decreasing respectively, across the productivity gradient. Proportions of platyhelminthes and tardigrades also increased in less-productive AVCs (Fig. 4.2E).

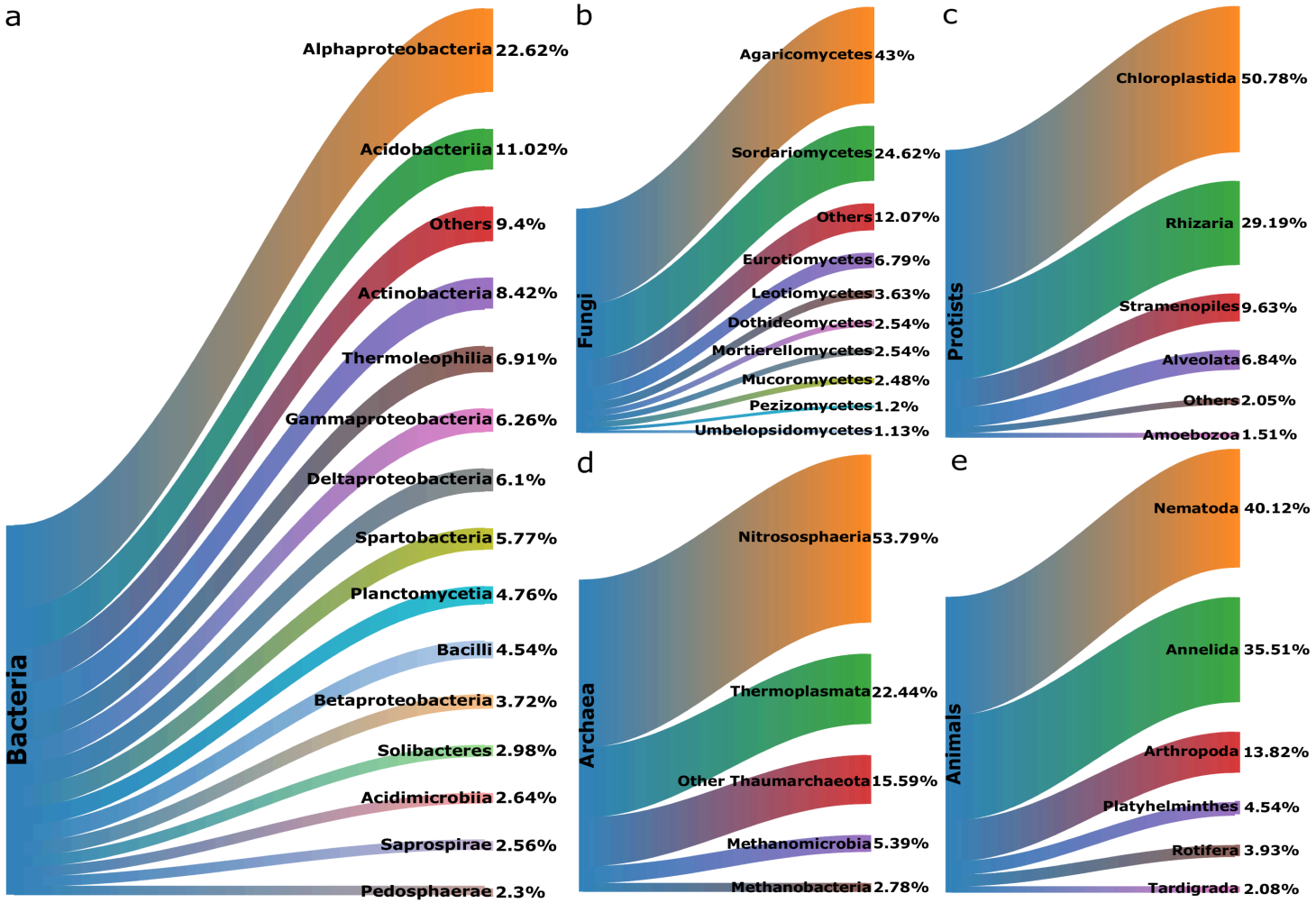


*Figure 4.1 Sankey diagrams of proportional abundances of OTUs from all samples for major soil biota groups. Arms denote proportions of OTUs at the class-level for **A**) bacteria; **B**) fungi; of major lineages of **C**) protists; class-level for **D**) archaea; and at the phylum-level for **E**) animals. For information on how this figure was created, please see Appendix 2.3.*

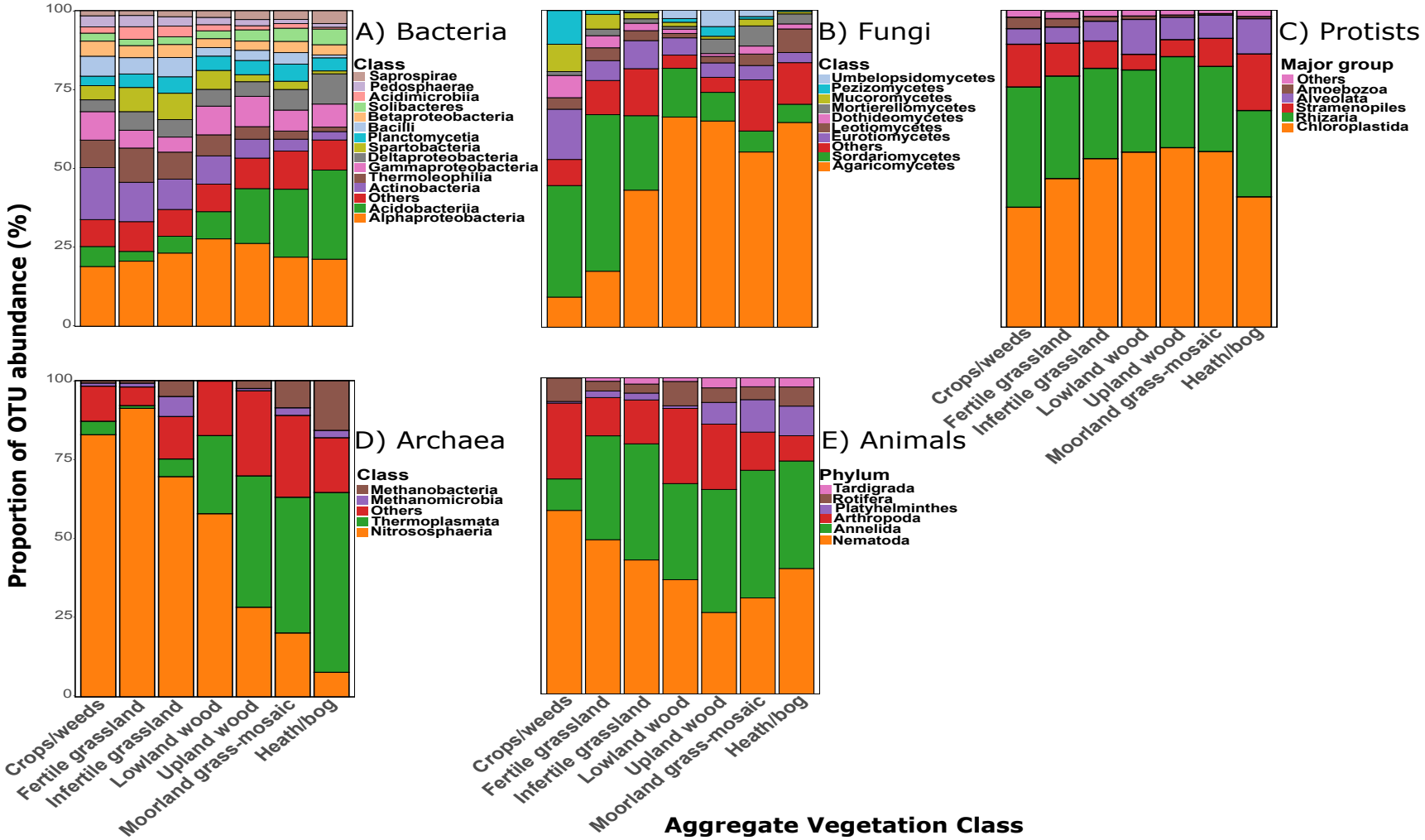


*Figure 4.2 Proportionate abundances of OTUs for major soil biota groups within each Aggregate Vegetation Class. Land uses are ordered from most (Crops/weeds) to least (Heath/bog) using the same divisions as Fig. 4.1 for **A**) bacteria; **B**) fungi; **C**) protists; **D**) archaea; and **E**) animals.*

4.2.2 Effect of land use on belowground richness

We found significant differences in biodiversity trends across land use types. There was a marked shift along the productivity gradient of Crops/weeds-to-Heath/bog in all organismal groups, except animals (Fig. 4.3). Significant differences in the mean richness of bacterial OTUs were prominent ($F_{6, 264} = 78.47$, $p < 0.0001$) following ANOVA. Bacterial richness decreased in AVCs across the productivity gradient with highest values in the most productive Crops/weeds and grasslands and lowest in the low productivity land uses (i.e. Moorland grass-mosaic, Heath/bog) (Fig. 4.3A). The same trend was also observed in fungi ($F_{6, 248} = 48.98$, $p < 0.001$; Fig. 4.3B), and protists ($F_{6, 249} = 59.86$, $p < 0.001$; Fig. 4.3C). For individual pair-wise comparisons see Appendix 2.2. Richness of archaeal OTUs had an opposing trend to that of other microbial groups. Archaeal OTU richness was significantly lower ($F_{6, 185} = 24.37$, $p < 0.001$) in higher-productivity AVCs and highest in the least-productive land-use types (Fig. 4.3D). In the Crops/weeds AVC richness of archaeal OTUs was significantly lower than Upland wood ($p = 0.01$), Moorland grass-mosaic ($p = 0.005$), and Heath/bog sites ($p < 0.001$) based on Tukey's *post hoc* tests, with the remaining land uses displaying intermediate OTU richness values.

Animal OTU richness did not follow the trends observed in microbial communities. Differences observed with ANOVA were significant ($F_{6, 244} = 6.25$, $p < 0.001$) but plateaued after the grassland AVCs, as opposed to the sloped trend of microbial groups across the productivity gradient (Fig. 4.3E). Richness in the Infertile grasslands was significantly greater than in Crops/weeds ($p = 0.008$), Heath/bog ($p = 0.003$), and Upland wood ($p = 0.02$) based on Tukey's *post hoc* tests. Richness was

lowest in the most intensively management Crops/weeds sites and was shown to be significantly lower than richness of Lowland woods ($p = 0.04$) with Tukey's test. Collectively the results demonstrate a strong divergence between the richness of animal and microbial communities across all AVCs.

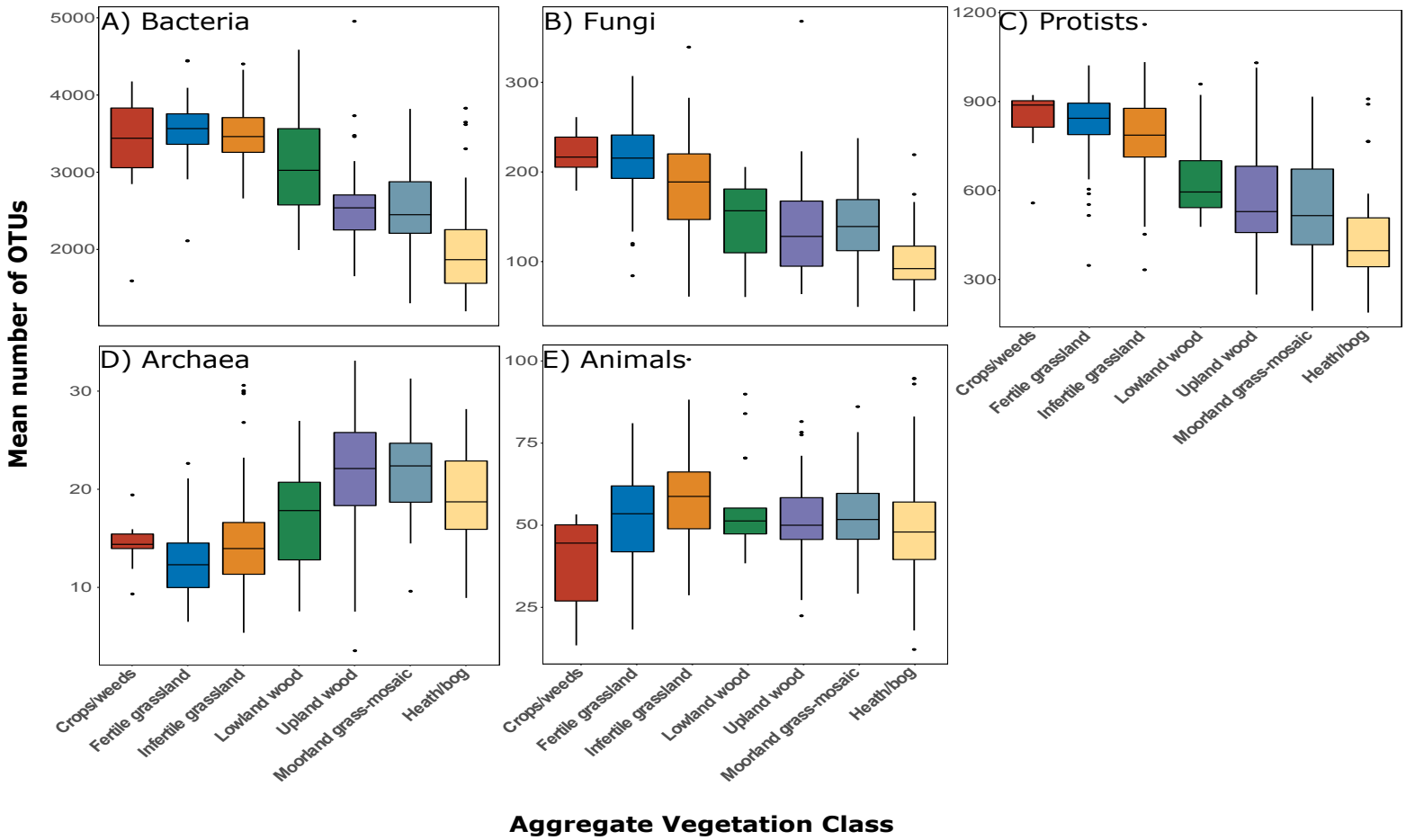


Figure 4.3. Boxplots of OTU richness for each organismal group. Richness of A) bacteria; B) fungi; C) protists; D) archaea; E) animals are plotted against Aggregate Vegetation Class ordered from most (Crops/weeds) to least (Heath/bog) productive. Boxes are bounded on the first and third quartiles; horizontal lines denote medians. Black dots are outliers beyond the whiskers, which denote 1.5X the interquartile range

4.2.3 Relationships of richness between organismal groups

Bacterial richness from the total data set was significantly correlated with all other organismal groups (Appendix 2 Table 2). Such relationships were positive between bacterial richness and richness of fungi, protists, and animals. Similarly there was a positive relationship between protistan richness and both fungal and animal richness. However, archaeal richness demonstrated significant, but negative correlations with all organisms except animals. Indeed animal richness (measured by metabarcoding) was only significantly correlated with animals (measured by taxonomic assessment; Table 4.1) and protists (Appendix 2 Table 2).

4.2.4 Relationships between richness and environmental variables

Partial least squares (PLS) regressions demonstrated that the divergence observed between animal and microbial communities may be due to the effects of soil properties. No soil properties were significantly correlated with richness of soil animal OTUs (Table 4.1). Conversely, there were strong relationships between microbial richness and a range of soil properties. However, although microbes were influenced by the same environmental variables, there were distinct patterns within each group. For example, while pH was the best predictor of bacterial richness, it was ranked as second for fungi and protists and third for archaea. Bulk density and C:N ratio were also major drivers of richness across all microbial groups. Elevation (here closely linked with precipitation and organic matter content) was the most important environmental variable in relation to archaea and protist richness. Organic matter and bulk density were strong predictors of fungal OTU richness. All environmental properties that had positive relationships with OTU richness of bacteria, fungi, and protists had negative relationships with archaea.

*Table 4.1 Results of partial least squares regressions for soil biota against soil properties for richness. Positive relationships are underlined; negative relationships are written in italics. *** indicates $p < 0.001$, ** $0.001 > p < 0.01$, * $0.01 > p < 0.05$, blank indicates $p > 0.05$.*

Soil and environmental variables	**Taxon**				
	Bacteria	Archaea	Fungi	Protists	Animals
Total C[L]	*1.14 ($R^2 = 0.44^{***}$)*	1.21 ($R^2 = 0.13^{***}$)	0.44	*1.3 ($R^2 = 0.35^{***}$)*	0.9
Total N[L]	0.93	0.89	0.93	0.8	1.18
C :N[S]	*1.45 ($R^2 = 0.41^{***}$)*	1.31 ($R^2 = 0.09^{***}$)	*1.64 ($R^2 = 0.28^{***}$)*	*1.67 ($R^2 = 0.35^{***}$)*	0.1
Total P (mg kg ^-1)[S]	0.35	0.59	0.7	0.85	0.67
Organic matter (% LOI)[L]	*1.47 ($R^2 = 0.5^{***}$)*	1.27 ($R^2 = 0.14^{***}$)	*1.13 ($R^2 = 0.29^{***}$)*	1.27 *($R^2 = 0.35^{***}$)*	1.08
pH ($CaCl_2$)	1.98 ($R^2 = 0.51^{***}$)	*1.68 ($R^2 = 0.25^{***}$)*	1.52 ($R^2 = 0.23^{***}$)	1.56 ($R^2 = 0.33^{***}$)	0.9
Soil water repellency[L*]	*1.31 ($R^2 = 0.2^{***}$)*	0.9	*1.23 ($R^2 = 0.13^{***}$)*	0.93	0.98
Volumetric water content (m^3 m^3 ^-1)	0.36	1.33 ($R^2 = 0.13^{***}$)	0.6	0.41	0.4
Soil bound water (g water g dry soil ^-1)	*1.25 ($R^2 = 0.41^{***}$)*	0.83	*1.08 ($R^2 = 0.26^{***}$)*	*1.23 ($R^2 = 0.31^{***}$)*	0.63
Rock volume (mL)	0.25	0.61	0.64	0.27	1.3
Bulk density (g cm^3 ^-1)	1.39 ($R^2 = 0.44^{***}$)	*1.43 ($R^2 = 0.18^{***}$)*	1.41 ($R^2 = 0.29^{***}$)	1.5 ($R^2 = 0.35^{***}$)	1.39
Clay content (%)[A]	0.85	*1.19 ($R^2 = 0.1^{***}$)*	0.84	1.14 ($R^2 = 0.09^{***}$)	0.05
Sand content (%)[A]	0.45	0.16	0.6	0.51	0.78
Elevation (m)	*1.66 ($R^2 = 0.42^{***}$)*	1.7 ($R^2 = 0.27^{***}$)	*1.68 ($R^2 = 0.22^{***}$)*	*1.65 ($R^2 = 0.36^{***}$)*	0.57
Mean annual precipitation (mL)	*1.08 ($R^2 = 0.25^{***}$)*	1.75 ($R^2 = 0.3^{***}$)	*1.44 ($R^2 = 0.18^{***}$)*	*1.48 ($R^2 = 0.27^{***}$)*	0.46
Temperature (°C)	0.51	0.5	0.56	0.58	0.35
Collembola[L1]	0.34	0.06	0.41	0.17	1.14 ($R^2 = 0.03^{***}$)
Mites[L1]	0.49	0.2	*1.17 ($R^2 = 0.03^{***}$)*	0.23	1.74 (R^2 0.08***)
Total mesofauna[L1]	0.44	0.1	*1.03 ($R^2 = 0.01^{*}$)*	0.15	1.71 (R^2 0.08***)

Note: [A] denotes Aitchison's log-ratio transformation; [L] denotes $\log_{10}$-transformation; [L1] denotes $\log_{10}$ plus 1 transformation [S] denotes square-root-transformation; * soil water repellency was derived from median water drop penetration times (s).

4.2.5 Community structure (β-diversity) across land uses

Non-metric multidimensional scaling (NMDS) using Bray-Curtis distances showed consistent differences in β-diversity between AVCs across all organismal groups. Plots show tight clustering of the Crops/Weeds, Fertile Grassland, and Infertile Grassland AVCs, whereas the other AVCs form a more dispersed organismal assemblage (Fig. 4.4 for bacteria and Appendix 2 Fig. 2-5). Results of PERMANOVAs were significant across all groups and analyses of dispersion were also significant (Fig. 4.4 for bacteria and Appendix 2 Fig. 2-5) for all groups except for the dispersion of animals ($F_{6, 401} = 0.67$, $p = 0.68$) owing to the wide range of sample numbers within each AVC (Appendix 2 Fig. 5). We also found that this clustering was present using constrained canonical analyses of principle components (CAP) ordinations for each organismal group (Appendix 2 Fig. 6-10).

pH was the best predictor of β-diversity from linear fitting for all soil organisms (Table 4.2 and Appendix 2 Tables 3-6). The carbon-to-nitrogen (C:N) ratio was the second most important variable in all major groups except animals. Mean C:N values were higher in the Crops/weeds and grassland AVCs and lower in the remaining land use types (Appendix 2 Table 6). Mean pH values and C:N ratios (Appendix 2 Table 6) reflect the distribution of points in NMDS plots, with tight groupings observed in the Crops/Weeds and grasslands AVCs and increasingly more spread out groupings in all other AVCs as pH values decreased and became more varied (Fig. 4.4 for bacteria and Appendix 2 Fig. 2-5). Across all groups, all or nearly all variables were significant following linear fitting; however, most were only weakly correlated with β-diversity values. Other important variables varied in their ranked importance including: elevation,

mean annual precipitation, organic matter content, total C, bulk density, volumetric water content, and clay content of soil (Table 4.2 and Appendix 2 Tables 3-6). The results of linear model fitting for CAP ordinations, though not identical (Appendix 2 Tables 7-11), were highly related to those of the NMDS ordinations (Appendix 2 Fig. 11).

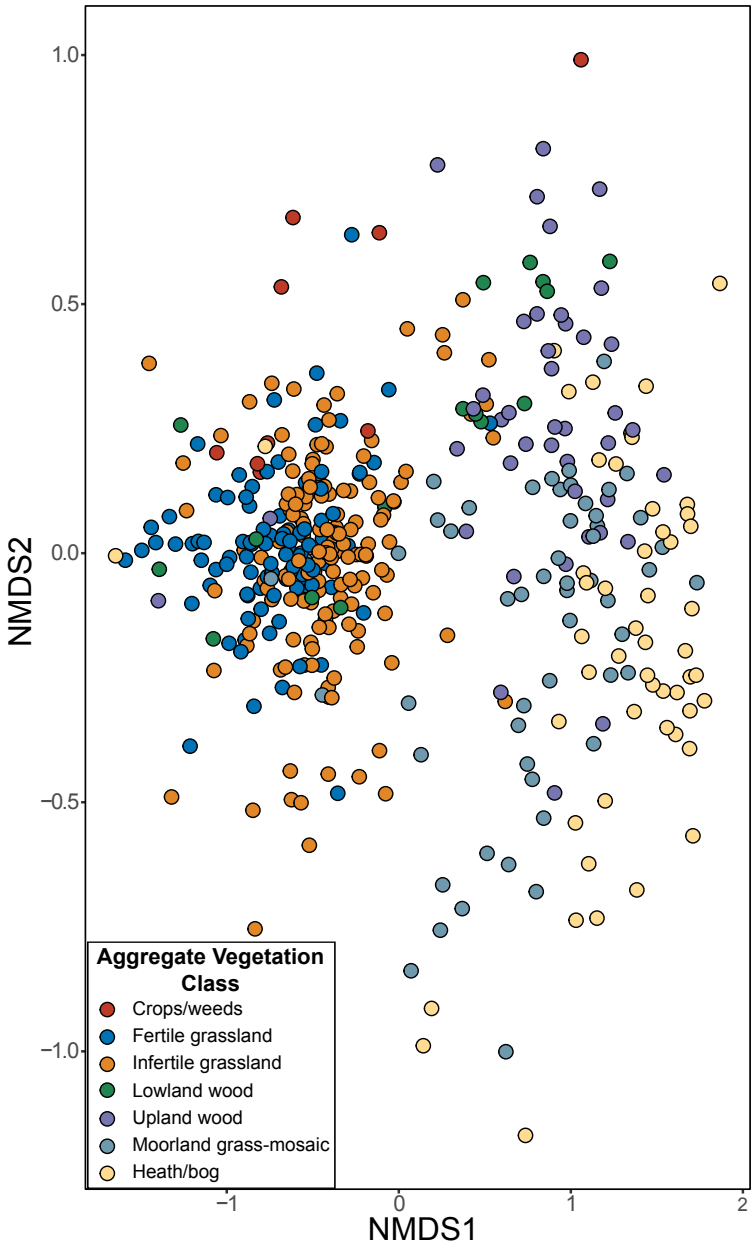


Figure 4.4 Plot of the non-metric dimensional scaling ordination (stress = 0.06) of bacterial community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Results of PERMANOVA ($F_{6,427}$ = 30.76, p = 0.001) and dispersion of variances of groups ($F_{6,427}$ = 10.97, p = 0.001) were significant.

Table 4.2 Summary of relationships amongst environmental factors and bacterial communities.

Soil and environmental variables		**Correlation**		
	R^2	Axis1	Axis2	Axis3
pH ($CaCl_2$)	0.71^{***}	-	-	+
C:N ratio^{S}	0.52^{***}	+	-	+
Volumetric water content (m^3 m^3 ^-1)	0.49^{***}	+	-	+
Bulk density (g cm^3 ^-1)	0.47^{***}	-	+	-
Organic matter (% LOI)^{L}	0.46^{***}	+	-	+
Elevation (m)	0.45^{***}	+	-	-
Mean annual precipitation (mL)	0.43^{***}	+	-	-
Total C^{L}	0.39^{***}	+	-	+
Clay content (%)^{A}	0.33^{***}	-	+	-
Soil bound water (g water g dry soil ^-1)	0.31^{***}	+	-	+
Soil water repellency^{L*}	0.27^{***}	+	-	-
Total N (%)^{L}	0.26^{***}	+	-	+
Sand content (%)^{A}	0.21^{***}	+	+	+
Collembola^{L1}	0.09^{***}	-	+	-
Mites^{L1}	0.06^{***}	+	+	-
Total P (mg kg ^-1)^{S}	0.06^{***}	-	-	-
Total mesofauna^{L1}	0.06^{***}	+	+	-
Rock volume (mL)	0.05^{**}	-	+	-
Temperature (°C)	0.03^{*}	+	+	-

4.3 Discussion

High-throughput sequencing the biosphere amongst heterogeneous soils revealed both expected and novel relationships between soil organisms and environmental drivers. The richness of microbes and animals had notable contrasting trends across land use types. The richness of microbial communities was strongly influenced by both land use and environmental variables, especially pH, C:N ratio, elevation, organic matter, and annual precipitation. Conversely, we found no significant associations between measured environmental variables and animal richness, which was negatively impacted by higher intensity land use, suggesting that richness patterns of microbial and macrobial life fractions

adhere to different ecological determinants. For β-diversity, pH was by far the most important environmental variable in shaping community composition of all organismal groups, yet other drivers were attributable for influencing patterns of α-diversity.

Our findings demonstrate that diverging trends between soil microbes and fauna extend across distinct, heterogeneous land uses. Furthermore, we build on the work of Gossner et al.[15] by demonstrating that microbial richness, with the exception of archaea, increases with greater land use intensity across heterogeneous ecosystems at the national-scale. The divergence between microbes and animals at this scale is supported by previous findings from French soils[17,25]. Across France, bacterial richness[17] and biomass[25] were strongly linked to belowground environmental properties but largely unaffected by aboveground climatic variables, which commonly influence animal and plant biogeography[25,30]. Our findings show that richness of fungi and protists also follow this trend – whereas archaea follow an opposing trend to all other groups.

There are several mechanisms that may explain the relationship between higher microbial richness and intensifying anthropogenic disturbance. One explanation is that consistent nutrient inputs from fertilizers and disturbance under tillage stimulate high α-diversity in these areas[16]. Indeed higher α-diversity has been observed in cropping systems than in forest or grassland sites for both bacteria[16-17], and fungi[16]. Interestingly, high microbial richness in more productive land use types (e.g. arable) may illustrate the intermediate disturbance hypothesis (IDH) within soil ecosystems. Under the IDH, as outlined by Connell[33], diversity reaches its highest levels where succession has been interrupted by intermittent disturbance events. In our sites, microbial richness was highest in AVCs concurrent to disturbances (augmented by nutrient inputs) from agricultural interventions such as

fertilisation, tilling, clearing, and the cultivation of livestock. However, it is also possible that the high diversity observed in the grassland and especially in agricultural land uses stems from organisms that have entered a dormant state after disturbance-induced changes to their environment[13,34]. Disturbance pressures can also lead to high bacterial diversity through the reduction in dominant OTUs, which are replaced by a wide range of weaker competitors. It has been demonstrated that α-bacterial diversity is greater in the phyllosphere of ivy in urban habitats associated with more anthropogenic stressors than in less disturbed sites[35]. Our findings suggest that the phenomenon of greater species richness resulting from the addition of nutrients and non-equilibrium dynamics induced by disturbance may extend to across all microbial groups, with the possible exception of archaea.

Richness of all microbial groups, except archaea, followed the land use productivity/management intensity gradient[32] with higher richness in the highly productive and more disturbed grasslands and arable sites and lower richness in the least productive, relatively undisturbed upland Heath/bog sites. Changes within bacterial and fungal communities reflected expected within-community changes following the shift in soil nutrient quality across land uses. Actinobacteria[36] and Sordariomycetes[37] are known to dominate bacterial and fungal communities in high productivity grasslands as witnessed here. In contrast, Acidobacteria increased in proportion in low productivity, highly acidic AVCs as expected based on previous studies from the UK[27] and across the globe[7]. Likewise, the greater proportion of Agaricomycetes OTUs in low productivity AVCs is intuitive as many Agaricomycete fungi are common in bogs and related low-productivity habitats across Wales[38].

Protists have been chronically overlooked in European soil monitoring programmes (but see[28]), as extracting trends of protist diversity across land uses is difficult. For example,

Gossner et al.[15] were not able to show changes in richness across all protists with land use intensification. We demonstrate that protistan richness follows the trends of bacteria and fungi across land uses, with the highest richness levels in arable land. As with other microbes, there is evidence of increased protist richness at the mesocosm[39] and field[40] level, in response to fertiliser addition. Furthermore, in German grassland soils, protist richness has been shown to increase with land use intensity[41]. Our results show that an association between intensification and protistan richness extends across the national-scale over multiple land uses. Unlike other microbes, archaeal richness was greatest in low productivity AVCs and lowest in highly productive sites (Fig. 3d). Furthermore, our understanding of the extent of soil archaeal diversity and its functional capabilities is continually increasing[6-8]. Recent research has revealed many lineages of Thaumarchaeota are crucial links in the N cycle and methanogenesis in soils[7-8]. Archaeal richness was highest in the Moorland grass-mosaic and Heath/bog AVCs, likely due to the specialised nature of acidophilic lineages. In particular, the Thaumarchaeota[42] and Thermoplasmata[43] are known to proliferate (Fig. 2d) under reduced competition from bacteria.

Animal richness did not change linearly with land use and was not strongly influenced by environmental variables. Our molecular analysis of soil eDNA support recent findings by George et al.[23] based on morphological assessments of coincident soil mesofauna. Both the present work and George et al.[23] demonstrated that animal richness and abundance were lowest in land uses associated with more intensive management. Animal richness peaked in Infertile grasslands and was lowest in Crops/weeds sites (Fig. 3e). Agricultural disturbance negatively affects soil faunal richness and diversity across large geographic scales[14,23-24]. However, in the low-productivity land uses, although proportional abundances of arthropod taxa declined

similarly to the findings of George et al.[23], overall richness was not as strongly affected due to an increase in fractions of annelids, platyhelminthes, and tardigrades. Such an increase in the peat-rich, low-disturbance, higher elevation sites is rather intuitive since annelids, platyhelminthes, and tardigrades are susceptible to desiccation and require moist habitats to be active components of the soil community[44-45]. As soil animals still exhibited expected lower diversity trends in more intensively managed land uses[15,23-24], there are further opportunities for research into understanding the mechanisms underlying the divergent richness trends between microscopic animals and the rest of soil communities.

Soil pH, as evidenced by ordination results, was the most important environmental variable in our study for β-diversity and in most cases richness as has been previously observed across the UK[27-28] and at larger national[25-26] and continental scales[4-6]. pH has been implicated with driving richness of soil Archaea[42-43] and is the most important driver of protist communities in the UK[28]. However, pH only plays a marginal role in shaping soil protist communities globally[11]. Likewise pH is a poor predictor of global fungal biogeography, yet is a good predictor of ectomycorrhizal fungal richness[9], which may contribute to the Agaricomycetes OTUs observed in the present study. Nevertheless, it is important to acknowledge the inconsistent nature of correlations between microbial biodiversity and pH, potentially due to variations in soil properties occurring at scales that do not align with large-scale soil surveys[30].

We also observed a strong effect of C:N ratio in determining richness of microbes and β-diversity of all organismal groups, as has been observed in bacterial[27] and protistan[28] β-diversity across Britain and some fungi globally[9]. Yet C:N ratio is often co-correlated with other soil properties including bulk density, total C, organic matter, elevation, and mean annual

precipitation. Disentangling such related variables is difficult; despite using PLS analyses[46] we could not disentangle co-correlated soil properties. For example, AVCs such as Moorland grass-mosaic and Heath/bog generally had higher elevation, mean annual precipitation, C:N ratio, and both total C and N (Appendix 2 Table 12) owing to their less-disturbed, upland location and often peat-rich soils. Higher C:N ratios are indicative of lower-quality soils[47] and have historically been associated with a shift in microbial biomass from bacterial to fungal dominance[48]. Our results suggest that, with the exception of archaea, microbial richness is equally susceptible to the effect of soil quality degradation. According to our results, archaea, on the contrary, appear to be well adapted to habitats with lower nutrient quality.

We observed strong relationships between soil properties and microbial, but not animal richness. We suspect this is due to the direct effects of soil properties on microbes. For example, shifts in pH towards either a more alkaline or acidic condition inhibit the ability of most non-specialised bacteria to uptake nutrients from their environment[26]. In addition the quality of soil nutrients, as discussed previously, was likely a strong determinant of available nutrient resources and therefore total richness of microbes. We also found strong relationships between soil properties and β-diversity and across all organismal groups. These relationships between Bray-Curtis dissimilarities of and soil properties demonstrate that more dissimilar belowground communities correlate positively with indicators of better quality soils across the breadth of soil biota (Supplementary Table 6). However, associations between nutrient quality and animal community composition are likely the result of nutrients influencing the composition of the aboveground plant community[49] rather than direct interactions with animals. Furthermore, animals are more vagile than microbes and can actively seek out

microhabitats with better resources[50], limiting the direct impact of soil properties on animal richness.

Using an extensive soil sampling programme and metabarcoding, we present perhaps the most comprehensive assessment of the belowground diversity in Europe. Despite uncertainties on the ability of environmental DNA methods using small soil volumes to accurately characterise communities of larger organisms[51], we were still able to detect key differences in larger organisms (i.e. animals) across land uses. Our results highlight the complexity of belowground ecology by demonstrating a divergence of patterns of richness between soil fauna and microorganisms at a national-level. We show that microbial richness is strongly influenced by soil properties in a near-uniform manner, whereas animal richness is not. Rather, animal richness is likely driven by changes in aboveground communities that stem from intensive land use management, while microbial richness was affected by soil properties in addition to land use. A particularly interesting outcome of our analyses is the near-uniform trend of declining microbial richness along a gradient of decreasing land use productivity/management intensity. The data therefore suggest that soil properties strongly affect bacteria, fungi, and protists in a similar manner, whereby richness decreases with soil quality; whereas archaea showed an opposing trend with increasing richness as productivity declined. The richness of animal OTUs, on the contrary, was not affected by soil properties although β-diversity was. Although often considered as ecological 'black boxes', soils continue to provide unique and coherent insights into the differences between interconnected microbial and macrobial assemblages. Our findings also highlight the importance of the dynamics between biotic and abiotic processes that drive the organization of belowground biological diversity.

4.4 Methods

4.4.1 Sampling

Soil samples were collected between late spring and early autumn in 2013 and 2014 as part of GMEP (Appendix 1.1), established to monitor the Welsh Government's agri-environment scheme, Glastir. The scheme covered an area of 3,263 km^2 with 4,911 landowners[31]. Briefly, surveyors collected samples from randomly selected 1 km^2 squares with up to 3 locations within squares, following protocols established by the UK Countryside Survey[52]. As described previously, habitat within plots was classified using plant species assessments into one of seven AVCs[32]: Crops/weeds (n = 9), Fertile grassland (n = 98), Infertile grassland (n = 162), Lowland wood (n = 17), Upland wood (n = 44), Moorland-grass mosaic (n = 54), and Heath/bog (n = 52) (Appendix 2.1; Appendix 2 Table 1). Soil type was derived from the National Soil Map[53] (Appendix 1.2; Appendix 1 Table 2). Organic matter content was classified by loss-on-ignition (LOI) following the protocols of the 2007 Countryside Survey[51].

A total of 436 cores were collected from 1 km^2 squares, with up to 3 samples coming from an individual square based on a randomised sampling design. Cores were transported to the Centre for Ecology and Hydrology, Bangor, United Kingdom, and stored at -80 °C until DNA extraction. Soil physical and chemical properties were taken from 4 cm diameter by 15 cm deep cores co-located with the high-throughput sequencing cores. These included total C (%), N (%), P (mg kg ^-1), organic matter (% LOI), pH (measured in 0.01 M $CaCl_2$), mean soil water repellency (median water drop penetration time in seconds), bulk density (g cm^3 ^-1), volume of rocks (cm^3), soil bound water (g water g dry soil ^-1), volumetric water content (m^3 m^3 ^-1), as well as clay and sand content (%) of soil. Abundances of mesofauna collected as

part of GMEP were taken from George et al.[23] and geographic data including grid eastings, northings, and elevation were also included in our analyses. For complete details on chemical analyses see Emmett et al.[51]. Temperature (°C) and mean annual precipitation (mL) were extracted from the Climate Hydrology and Ecology research Support System dataset[54]. Mean values for each variable were recorded for each AVC (Appendix 2 Table 12) and soil properties were normalised where appropriate.

Soil texture data were measured by laser granulometry with a LS320 13 analyser (Beckman-Coulter). We subsampled approximately 0.5 g of soil taken from 15 cm cores by manual quartering and removed organic C using H_2O_2 and then transferred the sample into 250 mL bottles, added 5 mL of 5 % Calgon ® and shook overnight at 240 rpm. Bottles were emptied manually into the laser diffraction instrument for measuring particle size distribution. Full Mie theory was used to obtain a particle size distribution from the raw measurement data, with the real refractive index set to 1.55 and the absorption coefficient at 0.1 as in Özer et al.[55]. The cut-off points for clay, silt, and sand were: 2.2 μm, 63 μm and 2000 μm respectively. Clay and sand percentages were selected for subsequent analyses and normalised using Aitchison's log-ratio transformation.

4.4.2 DNA extraction

Soils were homogenised by passing through a sterilised 2 mm stainless steel sieve. Sieves were sterilised between samples by rinsing under the tap water using high flow, applying Vircon laboratory disinfectant and UV-treating each side for 5 minutes. DNA was extracted by mechanical lysis and the homogenisation step performed in triplicate from 0.25 g of soil per sample using a PowerLyzer PowerSoil DNA Isolation Kit (MO-BIO). Pre-treatment with 750 μL of 1 M $CaCO_3$ following Sagova-Mareckova et al.[56] was performed as it was

shown to improve PCR performances, especially for acidic soils. Extracted DNA was stored at -20 °C until amplicon library preparation began. To check for contamination in sieves 3 negative control DNA extractions were completed and an additional 2 negative control kit extractions were performed using the same technique but without the $CaCO_3$ solution.

4.4.3 Primer selection and PCR protocols for library preparation

Amplicon libraries were created using primers for rRNA marker genes, specifically for the V4 region of the 16S rDNA gene targeting bacteria and archaea (515F/806R)[57], ITS1 targeting fungi (ITS5/5.8S_fungi)[58], and the V4 region of the 18S rDNA gene (TAReuk454FWD1/TAReukREV3)[59] targeting a wide range of, but not all, eukaryotic organisms. We used a two-step PCR following protocols devised in conjunction with the Liverpool Centre for Genome Research. Amplification of amplicon libraries was run in triplicate on DNA Engine Tetrad® 2 Peltier Thermal Cycler (BIO-RAD Laboratories) and thermocycling parameters for each PCR started with 98 °C for 30 s and terminated with 72 °C for 10 min for final extension and held at 4 °C for a final 10 min. For the 16S locus, first-round PCR amplification followed 10 cycles of 98 °C for 10 s; 50 °C for 30 s; 72 °C for 30 s. For ITS1, there were 15 cycles of 98 °C for 10 s; 58 °C for 30 s; 72 °C for 30 s. For 18S there were 15 cycles at 98 °C for 10 s; 50 °C for 30 s; 72 °C for 30 s. Twelve µL of each first-round PCR product were mixed with 0.1 µL of exonuclease I, 0.2 of µL thermosensitive alkaline phosphatase, and 0.7 µL of water and cleaned in the thermocycler with a programme of 37 °C for 15 min and 74 °C for 15 min and held at 4 °C. Addition of Illumina Nextera XT 384-way indexing primers to the cleaned first round PCR products were amplified following a single protocol which started with initial denaturation at 98 °C for 3 min; 15 cycles of 95 °C for 30 s; 55°C for 30 s; 72 °C for 30s; final extension at 72 °C for 5 min and held at 4 °C. Twenty-five

μL of second-round PCR products were purified with an equal amount of AMPure XP beads (Beckman Coulter). Library preparation for 2013 samples was conducted at Bangor University. Illumina sequencing for both years and library preparation for 2014 samples were conducted at the Liverpool Centre for Genome Research.

4.4.4 Bioinformatics

Bioinformatics analyses were performed on the Supercomputing Wales cluster. A total of 130,219260, 104,276,828, and 98,999,009 raw reads were recovered from the 16S, ITS1, and 18S sequences, respectively. Illumina adapters were trimmed from sequences using Cutadapt[60] with 10% level mismatch for removal. Sequences were then de-multiplexed, filtered, quality-checked, and clustered using a combination of USEARCH v. 7.0[61] and VSEARCH v. 2.3.2[62]. Open-reference clustering (97% sequence similarity) of operational taxonomic units (OTUs) was performed using VSEARCH; all other steps were conducted with USEARCH. Sequences with a maximum error greater than 1 and shorter than 200 bp were removed following the merging of forward and reverse reads for 16S and ITS1 sequences. A cut-off of 250 bp was used for 18S sequences, according to higher quality scores. There were 15,202,313 (16S), 7,242,508 (ITS1), and 9,163,754 (18S) cleaned reads left at the end of these steps. Sequences were sorted and those that only appeared once in the dataset were removed. Briefly, filtered sequences were matched first against a number of different reference databases: Greengenes 13.8[63], UNITE 7.2[64], and SILVA 128[65] for 16S, ITS1, 18S, respectively. Ten per cent of sequences that failed to match were clustered *de novo* and used as a new reference database for failed sequences. Sequences that failed to match with the *de novo* database were subsequently also clustered *de novo*. All clusters were collated and chimeras were removed using the uchime_ref command in VSEARCH.

Chimera-free clusters and taxonomy assignment were used to create an OTU table with QIIME v. 1.9.1[66] using RDP[67] methodology with the GreenGenes database for 16S and UNITE database for ITS1 data. Taxonomy was assigned to the 18S OTU table using BLAST[68] against the SILVA database and OTUs appearing only once or in only 1 sample were removed from each OTU table.

Newick trees were constructed for the 16S and 18S tables using 80% identity thresholds. The trees were combined with their respective OTU tables as part of analyses using the R package phyloseq[69], removing OTUs that did not appear in both the tree and OTU table. OTUs identified as eukaryotes in the 16S OTU table, non-fungi OTUs in the ITS OTU table, as well as OTUs identified as fungi, plants, and non-soil animals were removed from the 18S OTU table. Read counts from each group were normalised using rarefaction. The OTU tables were rarefied 100 times using phyloseq[69] (as justified by Weiss et al.[70]) and the resulting mean richness was calculated for each sample. The read depth used for rarefaction varied for each group (Appendix 2 Table 13). Samples with lower read counts than this cut-off were removed before rarefaction. A summary of number of replicates per AVC is included in Supplementary Table 1.

4.4.5 Statistical analyses

All statistical analyses were run using R v. 3.3.3[71] using the rarefied data sets for each organismal group. The vegan package[72] was used to assess β-diversity via NMDS and CAP ordinations based on Bray-Curtis dissimilarities. A linear model for each environmental variable was fit separately to the ordination using the envfit function, the results are presented ranked according to goodness-of-fit. Results of goodness-of-fit for each variable from both ordination methods were compared using regression analyses to look for congruence. The

values of all variables were plotted against NMDS scores to determine if there were positive or negative relationships with each NMDS axis. Differences in β-diversity amongst AVCs were calculated with PERMANOV. The assumption of homogeneity of dispersion was verified using the betadisper function.

Linear mixed models were constructed using package nlme[73] to test the differences in α-diversity amongst AVCs for each organismal group. Model selection was performed using AVC, soil type, LOI classification, and sample year as fixed factors; sample square identity was the random factor. To determine the best possible model, predictors other than AVC were dropped to find the lowest AIC scores using the AICcmodavg package[74]. For each model, significant differences were assessed by ANOVA and pairwise differences were identified with Tukey's *post-hoc* tests from the multcomp package[75].

Partial least squares regressions found in package pls[76] were used to identify the most important environmental variables for richness. Such analysis is ideal for data where there are many more explanatory variables than sample numbers or where extreme multicollinearity is present[46]. As in Lallias et al.[46], we used the variable importance in projection (VIP) approach[77] to sort the original explanatory variables by order of importance; variables with VIP values > 1 were considered most important. Relationships between important variables and richness values for each group of organisms were investigated by linear regression. Richness was normalised before regression when necessary. Pearson's correlation coefficient was used to directly compare richness of organismal groups.

4.4.6 Data availability

Data associated with this paper will be publically published in the National Environment Research Council (NERC) Environmental Information Data Centre (EIDC).

Sequences with limited sample metadata have been uploaded to The European Nucleotide Archive and can be accessed with the following primary accession codes after the end of data embargo (27 June, 2020): PRJEB27883 (16S), PRJEB28028 (ITS), and PRJEB28067 (18S). Data are also available from the authors upon reasonable request with permission from the Welsh Government. The source data underlying Fig. 4.3A-E is provided as a Source Data file.

4.5 Acknowledgements

This research was supported under the Glastir Monitoring & Evaluation Programme (Contract reference: C147/2010/11) and by the UK Natural Environment Research Council (NERC) through the RC/Centre for Ecology & Hydrology (CEH Project: NEC04782). PBLG was supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council and NERC [Grant number NE/M009106/1]. STARS is a consortium consisting of Bangor University, British Geological Survey, Centre for Ecology and Hydrology, Cranfield University, James Hutton Institute, Lancaster University, Rothamsted Research and the University of Nottingham. We thank the GMEP team for their contribution in collecting the data, especially Aidan Keith, for assistance in enumerating mesofauna, and the laboratory staff of Environment Centre Wales. We appreciate the support and assistance of Supercomputing Wales for bioinformatics analyses. Sequence data generation was performed at the Centre for Genomic Research, University of Liverpool.

4.6 References

1. Barrios, E. Soil biota, ecosystem services and land productivity. *Ecol. Econ.* **64**, 269-285 (2007).

2. Bardgett, R. D. & van der Putten, W. H. Belowground biodiversity and ecosystem functioning. *Nature* **515**, 505-511 (2014).

3. Ramirez, K. S. et al. Detecting macroecological patterns in bacterial communities across independent studies of global soils. *Nat. Microbiol.* **3**, 189-196 (2018).

4. Delgado-Baquerizo, M. et al. A global atlas of the dominant bacteria found in soil. *Science* **359**, 320-325 (2018).

5. Lauber, C., Hamady, M, Knight, R., & Fierer, N. Soil pH as a predictor of soil bacterial community structure at the continental scale: a pyrosequencing-based assessment. *Appl. Environ. Microbial.* **75**, 5111-5120 (2009).

6. Fierer, N. & Jackson, R. B. The diversity and biogeography of soil bacterial communities. *Proc. Nat. Acad. Sci. USA* **103**, 626-631 (2006).

7. Fierer, N. et al. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi and viruses in soil. *Appl. Environ. Microbiol.* **73**, 7059-7066 (2007).

8. Offre, P., Spang, A., & Schleper, C. Archaea in biogeochemical cycles. *Annu. Rev. Microbiol.* **67**, 437-457 (2013).

9. Tedersoo, L. et al. Global diversity and geography of soil fungi. *Science* **346**, 1256688 (2014). DOI: 10.1126/science.1256688

10. Buée, M. et al. 454 Pyrosequencing analyses of forest soils reveal and unexpectedly high fungal diversity. *New Phytol*. **184**, 449-456 (2009).

11. Bates, S. T. et al. Global biogreography of highly diverse protistan communities in soil. *ISME J.* **7**, 652-659 (2013).

12. Mahé, F. et al. Parasites dominate hyperdiverse soil protist communities in Neotropical rainforests. *Nat. Ecol. Evol.* **1**, 0091(2017).

13. Geisen, S. et al. Soil protists: a fertile frontier in soil biology research. *FEMS Microbiol. Rev.* **42**, 293-323 (2018).

14. Wu, T., Ayres, E., Bardgett, R. D., Wall, D. H., Garey, J. R. Molecular study of worldwide distribution and diversity of soil animals. *Proc. Nat. Acad. Sci. USA* **108**, 17720-17725 (2011).

15. Gossner, M. M. et al. Land-use intensification causes multitrophic homogenization of grassland communities. *Nature* **540**, 266-269 (2016).

16. Szoboszlay, M., Dohrmann, A. B., Poeplau, C., Don, A., Tebbe, C. C. Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. *FEMS Microbiol. Ecol.* **93**, fix146 (2017).

17. Terrat, S. et al. Mapping and predictive variations of soil bacterial richness across France. *PLoS One* **12**, e0186766 (2017). DOI: 10.1371/journal.pone.0186766

18. Li, F.-R., Liu, FJ.-L., Liu, L.-L. Land-use change alters patterns of soil biodiversity in arid lands of northwestern China. *Plant Soil* **428**, 371-388 (2018).

19. Keith, A. M. et al. Cross-taxa congruence, indicators and environmental gradients in soils under agricultural and extensive land management. *Eur. J. Soil Biol.* **49**, 55-62 (2012).

20. Leff, J. W. et al. Predicting soil communities from plant community taxonomy, phylogeny, and traits. *ISME J.* **12**, 1794-1805 (2018).

21. Scharroba, A. et al. Effects of resource availability and quality on the structure of the micro-food web of an arable soil across depth. *Soil Biol. Biochem.* **50**, 1-11 (2012).

22. Ramirez, K. S. et al. Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B* **281**, 20141988 (2014). DOI: 10.1098/rspb.2014.1988

23. George, P. B. L. et al. Evaluation of mesofauna communities as soil quality indicators in a national-level monitoring programme. *Soil Biol. Biochem.* **115**, 537-546 (2017).

24. Tsiafouli, M. A. et al. Intensive agriculture reduces soil biodiversity across Europe. *Glob. Change Biol.* **21**, 973-985 (2015).

25. Dequiedt, S. et al. Biogeographical patterns of soil molecular microbial biomass as influenced by soil characteristics and management. *Global Ecol. Biogeogr.* **20**, 641-652 (2011).

26. Tripathi, B. M. et al. Tropical soil bacterial communities in Malaysia: pH dominates in the Equatorial tropics too. *Microb. Ecol.* **64**, 474-484 (2012).

27. Griffiths, R. I. et al. The bacterial biogeography of British soils. *Environ. Microbiol.* **13**, 1642-1654 (2011).

28. Dupont, A. Ö. C., Griffiths, R. I., Bell, T., Bass, D. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. *Environ. Microbiol.* **18**, 2010-2024 (2016).

29. Maestre, F. T. et al. Increasing aridity reduces soil microbial diversity and abundance in drylands. *Proc. Nat. Acad. Sci. USA* **112**, 15684-15689 (2015).

30. Hendershot, J. N., Read, Q. D., Henning, J. A., Sanders, N. J., Classen, A. T. Consistently inconsistent drivers of microbial diversity and abundance at macroecological scales. *Ecology* **98**, 1757-1763 (2017).

31. Emmett, B. A. & the GMEP Team. Glastir Monitoring & Evaluation Programme. Final Report to Welsh Government (Contract reference: C147/2010/11. NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780/NEC05371/NEC05782) (2017).

32. Bunce, R.G.H. et al. Vegetation of the British countryside – the countryside vegetation system. (Department of the Environment, Transport and the Regions, London) (1999).

33. Connell, J.H. Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302-1310 (1978).

34. Wang, G. Mayes, M. A., Gu, L., Schadt, C. W. Representation of dormat and active microbial dynamics for ecosystem modeling. *PLoS One.* **9**, e89252 (2014).

35. Smets, W. et al. Impact of urban land use on the bacterial phyllosphere of ivy (*Hedera sp.*). *Atmos. Environ.* **147**, 376-383 (2016).

36. Nacke, H. et al. Pyrosequencing-based assessment of bacterial community structure along different management types in German forest and grassland soils. *PLoS ONE* **6**, e17000 (2011).

37. Klaubauf, S. et al. Molecular diversity of fungal communities in agricultural soils from Lower Austria. *Fungal Divers.* **44**, 65-75 (2010).

38. O'Hanlon, R. & Harrington, T. J. Diversity and distribution of mushroom-forming fungi (Agaricomycetes) in Ireland. *Biol. Environ.* **111B**, 117-133 (2011).

39. Xiong, W. et al. Soil protist communities form a dynamic hub in the soil microbiome. *ISME J.* **12**, 634-638 (2018).

40. Lentendu, G. et al. Effects of long-term differential fertilization on eukaryotic microbial communities in an arable soil: a multiple barcoding approach. *Mol. Ecol.* **23**, 3341-3355 (2014).

41. Venter, P. C., Nitsche, F., Domonell, A., Heger, P., Arndt, H. The protistan microbiome of grassland soil: diversity in the mesoscale. *Protist* **168**, 546-564 (2017).

42. Oton, V. E., Quince, C., Nicol, G. W., Prosser, J. I., Gubry-Rangin, C. Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeota. *ISME J.* **10**, 85-96 (2016).

43. Lin, X., Handley, K. M., Gilbert, J. A., Kostka, J. E. Metabolic potential of fatty acid oxidation and anaerobic respiration by abundant members of Thaumarchaeota and Thermoplasmata in deep anoxic peat. *ISME J.* **9**, 2740-2744 (2015).

44. Coleman, D. C. Callaham, M., Crossley Jr., D. Fundamentals of Soil Ecology 3rd ed. 233-253 (Academic Press, London, 2018).

45. Cannon, R. J. C., Baker, R. H. A., Taylor, M. C., Moore, J. P. A review of the status of the New Zealand flatworm in the UK. *Ann. Appl. Biol.* **135**, 597-614 (1999).

46. Lallias, D., et al. Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J.* **9**, 1208-1221 (2015).

47. Bloem, J., de Ruiter, P. C., Bouwman, L. A. Soil food webs and nutrient cycling in agroecosystems. In *Modern soil microbiology* (eds. van Elsas, J. D., Jansson, J. K. & Trevors, J. T.), 245–278 (CRC Press, Boca Raton, 1997).

48. de Vries, F. T., Hoffland, E., van Eekeren, N., Brussard, L., Bloem, J. Fungal/bacterial ratios in grasslands with contrasting nitrogen management. *Soil Biol. Biochem.* **38**, 2092-2103 (2006).

49. Schuldt, A. et al. Belowground top-down and aboveground bottom-up effects structure multitrophic community relationships in a biodiverse forest. *Sci. Rep-UK* **7**, 4222 (2017). DOI: 10.1038/s41598-017-04619-3

50. Martiny, J. B. H. et al. Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102-112 (2006).

51. Wiesel, L., Daniell, T. J., King, D., Neilson, R. Determination of the optimal soil sample size to accurately characterise nematode communities in soil. *Soil Biol. Biochem.* **80**, 89-91 (2015).

52. Emmett, B. E. et al. Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology (CS Technical Report No. 9/07, CEH Project Number: C03259) (2010).

53. Avery, B.W. Soil classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden, UK (1980).

54. Robinson, E. L. et al. Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961-2015) [CHESS-met] v1.2. NERC Environmental Information Data Centre. (2017) https://doi.org/10.5285/b745e7b1-626c-4ccc-ac27-56582e77b900

55. Özer, M., Orhan, M., Işik, N. S. Effect of particle optical properties on size distribution of soils obtained by laser diffraction. *Enivron. Engi. Geosci.* **16**, 163-173 (2010).

56. Sagova-Mareckova, M. et al. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl. Environ. Microb.* **74**, 2902-2907 (2008).

57. Caporaso, J.G. et al. Global patters of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Nat. Acad. Sci.* **108 Suppl. 1**, 4516-4522 (2011).

58. Epp, L. S. et al. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* **8**, 1821-1833 (2012).

59. Behnke, A. et al. Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ. Microbiol.* **13**, 340-349 (2011).

60. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011. Date of access 04/07/2017. *EMBnet* **17**, 10-12 (2011). DOI: 10.14806/ej.17.1.200

61. Edgar, R. C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **10**, 2460-2461 (2010).

62. Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584 (2016). DOI: 10.7717/peerj.2584

63. DeSantis, T. Z. et al. Greengenes, a chimera-checked16S rRNA gene database and workbench compatible with ARB. *Environ. Microbiol.* **72**, 5069-5072 (2006).

64. Kõljalg, U. et al. Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* **22**, 5271-5277 (2013).

65. Quast, C. et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* **41 (D1)**, D590-D596 (2013).

66. Caporaso, J. G. et al. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**, 335-336 (2010).

67. Wang, Q., Garrity, G. M., Tiedje, J. M., Cole, J. R. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microb.* **73**, 5261-5267 (2007).

68. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).

69. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013). DOI: 10.1371/journal.pone.0061217

70. Weiss, S. et al. Normailzation and microbial differential abundance strategies depend upon data characteristics. *Microbiome* **5**, 1-18 (2017).

71. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). https://www.R-project.org/.

72. Oksanen, J. et al. Vegan: community ecology package. R package version 2.4-0 https://cran.r-project.org/web/packages/vegan/index.html (2016).

73. Pinheiro, J., et al. nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1e128, https://cran.r-project.org/web/packages/nlme/index.html (2016).

74. Mazerolle, M. J. AICcmodavg: model selection and multimodel inference based on (Q)AIC(c). R package version 2.1-1 https://cran.r-project.org/web/packages/AICcmodavg/index.html (2017).

75. Hothorn, T., Bretz, F., Westfall, P. Simultaneous inference in general Parametric Models. *Biometrical J.* **50**, 346-363 (2008).

76. Mevik, B-H., Wehrens, R., Liland, K. H., Hiemstra, P. pls: Partial Least Squares and Principal Component Regression. R package version 2.6-0 https://cran.r-project.org/web/packages/pls/index.html (2016).

77. Chong, I.-G. & Jun, C.-H. Performance of some variable selection methods when multicollinearity is present. *Chemometr. Intell. Lab. Syst.* **78**, 103-112 (2005).

Chapter 5: Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey

This manuscript has been published in *Frontiers in Environmental Science*

George, P.B.L.[1,2*], Creer, S.[1], Griffiths, R.I.[2], Lebron, I.[2], Emmett, B.A.[2], Robinson, D.A.[2], Jones, D.L.[1,3], 2019. Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey. *Frontiers in Environmental Science* **7**:173. doi: 10.3389/fens.2019.00173

[1] School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] School of Agriculture and Environment, University of Western Australia, Crawley, WA 6009, Australia

[*] Corresponding author

P.B.L.G., D.L.J., D.A.R. and S.C. conceived this project. B.A.E and D.A.R. facilitated the use of the GMEP dataset. Bioinformatics and statistical analyses were led by P.B.L.G. with assistance from S.C. and R.I.G. P.B.L.G. wrote the first draft of the manuscript and S.C., D.A.R., and D.L.J. contributed to subsequent revisions. All authors read and approved the final draft of the manuscript.

Abstract

The internal transcribed spacer (ITS) region is the accepted DNA barcode of fungi. Its use has led to a step-change in the assessment and characterisation of fungal communities from environmental samples by precluding the need to isolate, culture, and identify individuals. However, certain functionally important groups, such as the arbuscular mycorrhizas (Glomeromycetes), are better characterised by alternative markers such as the 18S rRNA region. Previous use of an ITS primer set in a nationwide metabarcoding soil biodiversity survey revealed that fungal richness declined along a gradient of productivity and management intensity. Here, we wanted to discern whether this trend was also present in data generated from universal 18S primers. Furthermore, we wanted to extend this comparison to include measures of functional diversity and establish trends with soil types and soil organic matter (SOM) content. Over the 413 individual sites examined (arable, grassland, woodland, moorland, heathland), we found congruent trends of total fungal richness and β-diversity across land uses, SOM class and soil type with both ITS and 18S primer sets. A total of 24 fungal classes were shared between datasets, in addition to 15 unique to ITS1 and 12 unique to 18S. However, using FUNGUILD, divergent trends of functional group richness became apparent, especially for symbiotrophic fungi, likely driven by an increased detection rate of Glomeromycetes in the 18S dataset. The disparate trends were also apparent when richness and β-diversity were compared to soil properties. Additionally, we found SOM class to be a more meaningful variable than soil type biodiversity analyses because organic matter was calculated for each sample whereas soil type was assigned from a national soil map. We advocate that a combination of fungal primers should be used in large-scale soil biodiversity surveys to capture important groups that can be underrepresented by universal barcodes.

Utilising such an approach can prevent the oversight of ubiquitous but poorly described species as well as critically important functional groups.

Key words: *UNITE; SILVA; identification bias; high-throughput sequencing; Arbuscular mycorrhizal fungi; Archaeorhizomycetes*

5.1 Introduction

Soil fungi are the dominant eukaryotic component of soil communities and are known to perform crucial ecosystem functions (Peay et al., 2008). Characterising the diversity of fungi within the landscape and their response to anthropogenic perturbation therefore represents an important topic within ecology. High-throughput sequencing has allowed the rapid estimation and identification of fungi by overcoming historical limitations of culture isolation and classifying fruiting bodies (Tedersoo et al., 2015). Using these DNA-based approaches it has been estimated that global fungal diversity in soil ranges from 3.5 – 5 million species. Yet at the beginning of the present decade, only around one-tenth of fungal diversity was thought to have been described (Rosling et al., 2011). In terms of ecosystem function, the majority of fungi are important in organic matter turnover and nutrient recycling as they facilitate the conversion of complex organic polymers into forms more readily accessible to other organisms (Peay et al., 2008; Nguyen et al., 2016). Consequently, they play a crucial role in regulating both below- and above-ground productivity (Peay et al., 2008). Many soil fungi also form important interactions with plants. Some form mutualistic relationships, best exemplified by the wide range of mycorrhizas (Wang and Qui, 2006; Smith and Read, 2008; Nguyen et al., 2016), whereas others are pathogens, responsible for numerous plant and animal diseases within agriculture and forestry (Fisher et al., 2012; Nguyen et al., 2016). Depending

on environmental conditions or life stage, fungi are capable of taking on some or all of these roles (i.e. saprotroph, symbiotroph, pathotroph) (Fisher et al., 2012). Despite the recognition that fungi are extremely important in soil ecosystems, characterising fungal communities has remained a challenge, exemplified by the numerous studies on soil bacteria in comparison to fungi.

Fungal barcode sequences are found within the ubiquitous, multicopy ribosomal RNA gene. Within this, the internal transcribed spacer (ITS) region has been accepted as a universal barcode for fungi (Schoch et al., 2012). Recent development of ITS-based databases such as UNITE (Kõjlalg et al., 2013) and Warcup (Deshpande et al., 2016) have overcome limitations in collecting and assigning taxonomic identities to unknown sequences, though database selection may introduce bias into results (Tedersoo et al., 2015; Xue et al., 2019). Yet ITS barcodes exhibit some limitations when dealing with unknown or environmental samples. Generally, the ITS region cannot be aligned above the family-level (Cavender-Bares et al., 2009), making phylogenies based on ITS sequence data unreliable. Importantly, the ITS region has proven unreliable at distinguishing certain fungal groups at the species-level, such as Glomeromycetes (Stockinger et al., 2010). Such inconsistencies mean that ITS primers may not accurately detect target organisms. For instance, Berruti et al. (2017), found that ITS primers underestimated Glomeromycetes in bulk soil. Such uncertainty may confound experimental results and lead to erroneous conclusions.

Despite the widespread use of ITS barcodes, other markers may better capture the diversity of some fungal taxa. Primers targeting the small and large subunits as well as the ITS regions of the rRNA gene have all been applied to fungi (Tedersoo et al., 2015; Xue et al., 2019). For example, early diverging lineages such as Chytridiomycota (Schoch et al., 2012;

Tedersoo et al., 2015) and Glomeromycetes (Tedersoo, et al., 2015) are poorly represented in ITS sequencing. Additionally, advancements in classification have highlighted the shortcomings of environmental DNA barcoding. For example, the Archaeorhizomycetes are a poorly understood but ubiquitous class of soil fungi and their previously unidentifiable sequences have been major components of past soil biodiversity assessments (Anderson et al., 2003; Rosling, et al., 2011). Overlooking these lineages may potentially lead to erroneous assumptions of biological and functional diversity in soils.

Underrepresentation of Glomeromycetes in particular exemplifies this issue. Arbuscular mycorrhizal fungi (AMF) form symbiotic relationships with more than 80% of vascular plant families and have been categorised into the monophyletic Glomeromycetes (Schüβler et al., 2001). Unlike most fungi, the ITS region has consistently demonstrated poor resolution in some closely related AMF species (Stockinger et al., 2010) as it is too hyper-variable (Thiéry et al., 2016). Instead, the 18S region is more commonly used for barcoding AMF, especially in ecological studies (Öpik et al., 2014). Therefore it is important to recognise biases inherent even in supposedly universal barcodes.

We previously undertook a nation-wide assessment of soil biodiversity across Wales, representing a breadth of heterogeneous land uses, which included agricultural land, grasslands, woodlands, and upland bogs. In this case, fungal richness and β-diversity were assessed using soil environmental DNA, utilising ITS1 primers (George et al., 2019). Yet, from the earliest stages of experimental design, we were cognisant that the ITS1 universal primer choice may not account for numerous functionally important fungal groups, particularly AMF. Thus, the primary objective of the present study was to assess whether observed fungal biodiversity (richness and β-diversity) across contrasting land uses from the ITS1 dataset

would differ when compared to a dataset derived from an alternative choice of primer and database. We therefore sought to assess if primer choice influenced fungal biodiversity across land use, soil type, and soil organic matter (SOM) class. Our next aim was to critically evaluate the influence of climatic and edaphic factors (e.g. soil pH, total carbon (C), nitrogen (N), phosphorus (P)) on fungal diversity arising from the use of the two different primer sets. Our final aim was to look for differences in coverage of taxonomic and functional diversity between the two primer sets across the broad range of land uses and soil types evaluated.

5.2. Materials and methods

5.2.1 Study design

Data were collected as part of the Glastir Monitoring & Evaluation Programme (GMEP). The GMEP initiative was established by Welsh Government to monitor their most recent agri-environment scheme, Glastir, which involved 4,911 landowners over an area of 3,263 km^2 (Fig. 5.1). Through the GMEP framework, survey teams collected samples in 2013 and 2014 between April and October in each year (Emmett and the GMEP Team, 2017). Sampling protocols were based on those of the UK-wide ecosystem monitoring programme, Countryside Survey (Emmett et al., 2010). The survey design randomly located 300, 1 km squares across 26 land classes in Wales which survey teams sampled with 5 plots in each square. A subset of samples were then randomly chosen from squares with a maximum of 3 selected in an individual square. A total of 437 samples were collected for biodiversity analyses.

At each sampling location, 2 cores were collected. One was a 15 cm deep by 4 cm diameter core from which measurements of soil physical and chemical properties were taken, including total C (%), N (%), P (mg/kg), organic matter (% loss-on-ignition), pH (measured in

0.01 M $CaCl_2$), mean soil water repellency (water drop penetration time in seconds), bulk density (g/cm^3), volume of rocks (cm^3), volumetric water content (m^3/m^3), as well as percentage sand and clay. For complete details on chemical analyses methodology, see Emmett et al. (2010). Soil texture data were measured by laser granulometry with a LS320 13 analyser (Beckman-Coulter) as described in George et al. (2019). The cut-off points for clay, silt, and sand were: 2.2 μm, 63 μm and 2000 μm respectively. Clay and sand percentages were selected for subsequent analyses and normalised using Aitchison's log_{10}-ratio transformation. Further geographic data including grid eastings, northings, and elevation were also collected. Mean temperature (°C) on date of sample collection and annual precipitation (mL) data were extracted from the Climate Hydrology and Ecology research Support System dataset (Robinson et al., 2017). Environmental variables were normalised (by log_{10} or square root transformation) where appropriate (see Table 5.1).

Each sampling site was assigned to a land use category, soil type, and SOM class (based on percentage organic matter). The land use classification used in this study was originally developed for the UK Countryside Survey in 1990 (Bunce et al., 1999). Briefly, vegetation was recorded by surveyors and used to classify each site into one of the 8 Aggregate Vegetation Classes (AVCs) as described in Bunce et al. (1999; for further details please see Appendix 2.1). The AVCs have been shown to follow a gradient of soil nutrient content from which productivity and management intensity can also be inferred (see Appendix 2.1 and Bunce et al., 1999). There were 7 AVCs identified in the present study. The AVCs in descending order of productivity are: Crops/weeds (including arable land), Fertile grassland, Infertile grassland, Lowland woodland, Upland woodland, Moorland grass-mosaic, Heath/bog (Appendix 3 Table 1). Soil type based on the predominant major soil group classification was

extracted from the National Soil Map (Appendix 3.1; Avery, 1980). Additionally, we classified soils on a per sample basis by organic matter content. Each sample was grouped into one of four organic matter classes based on percent loss-on-ignition (LOI) following the protocols of the 2007 Countryside Survey (Emmett et al., 2010): mineral (0-8% LOI), humus-mineral (8-30% LOI), organo-mineral (30-60% LOI), and organic (60-100% LOI). Mean values for each environmental variable were recorded for each land use, soil organic matter class, and soil type.

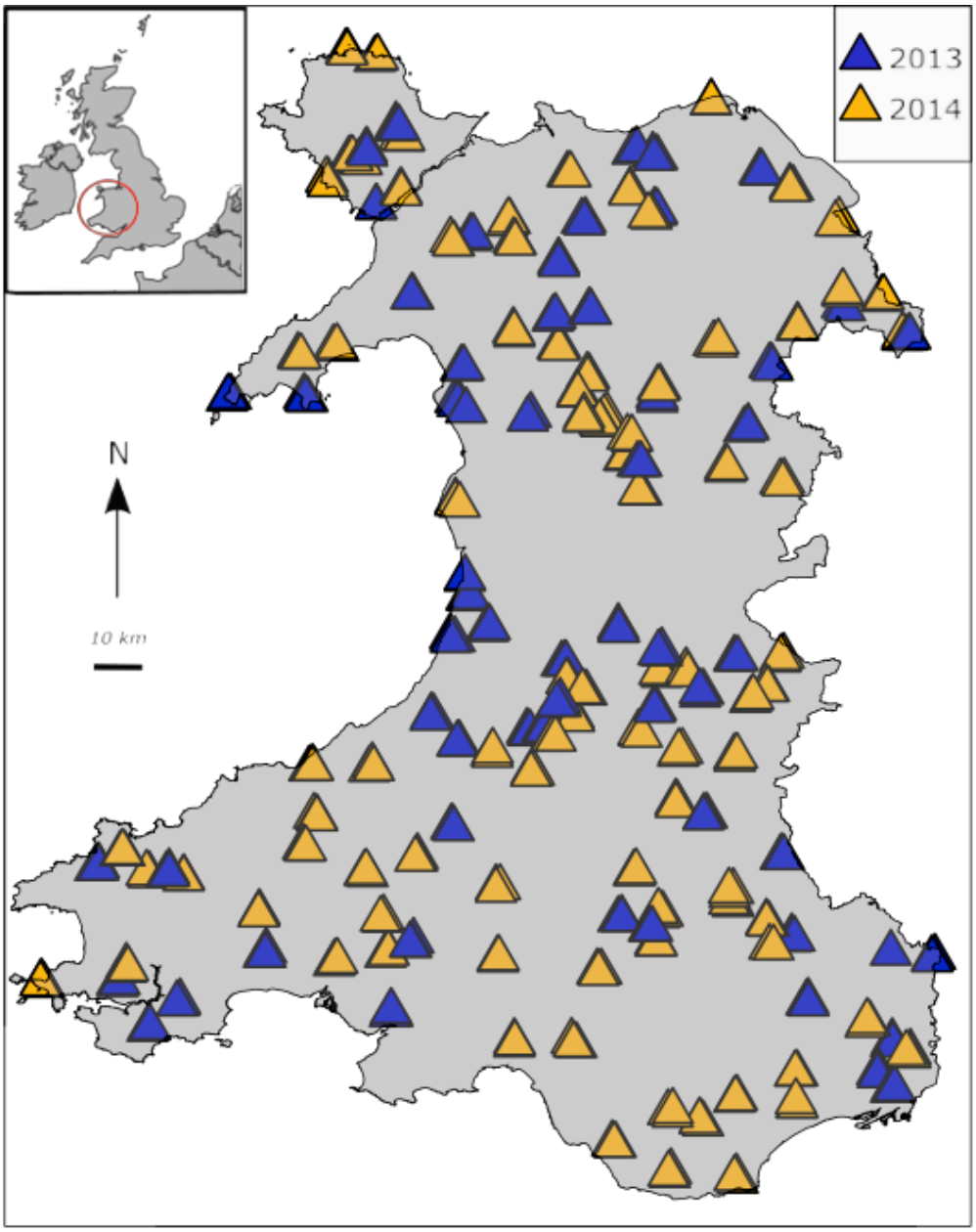


Figure 5.1 Map of sites selected for GMEP monitoring. To protect landowner anonymity, each triangle gives an approximate location of every 1 km^2 plot from which samples were taken.

5.2.2 DNA extraction

Soils used in DNA extraction were collected from 15 cm deep by 8 cm diameter cores. Soil samples were transported in refrigerated boxes; samples were received at Environment Centre Wales, Bangor within an average of 48 h post-extraction and frozen at -80 °C upon

arrival. Soils were then thawed and homogenised as they passed through a sterilised 2 mm stainless steel sieve after which they were returned to a -80 °C freezer until DNA extraction. Sieves were sterilised between samples by rinsing with tap water at high pressure and an application of Vircon® laboratory disinfectant followed by UV-treating each side for 5 minutes. DNA was extracted by mechanical lysis from 0.25 g of soil per sample using a PowerLyzer PowerSoil DNA Isolation Kit (MO-BIO Inc.). Soils were pre-treated with 750 μL of a suspension of $CaCO_3$ (1 M) following Sagova-Mareckova et al. (2008) to improve PCR performances, especially for acidic soils. Extracted DNA was stored at -20 °C until amplicon library preparation began. The extractions and homogenisation steps were performed in triplicate. To check for contamination in sieves, 3 negative control DNA extractions were completed as well as 2 negative control kit extractions using the same technique but without the $CaCO_3$ pre-treatment. Portions of the resultant DNA were used to create amplicon libraries for sequencing with each primer set.

5.2.3 Primer selection and PCR protocols for library preparation

Amplicon libraries were created using primers for the ITS1 (ITS5/5.8S_fungi) area to specifically target fungi (Epp et al., 2012) and the V4 region of the 18S gene (TAReuk454FWD1/TAReukREV3) (Behnke et al., 2011) targeting a wide range of, but not all, eukaryotic organisms, including fungi. A two-step PCR following protocols devised in conjunction with the Liverpool Centre for Genome Research was used as described in George et al. (2019). Amplification of amplicon libraries was run in triplicate on DNA Engine Tetrad® 2 Peltier Thermal Cycler (BIO-RAD Laboratories Inc.) and thermocycling parameters for both PCR protocols started with 98 °C for 30 s and terminated with 72 °C for 10 min for final extension and held at 4 °C for a final 10 min. For the ITS1 locus, there were 15 cycles of 98 °C

for 10 s; 58 °C for 30 s; 72 °C for 30 s. For the 18S locus there were 15 cycles at 98 °C for 10 s; 50 °C for 30 s; 72 °C for 30 s. Twelve µL of each first-round PCR product were mixed with 0.1 µL of exonuclease I, 0.2 of µL thermosensitive alkaline phosphatase, and 0.7 µL of water and cleaned in the thermocycler with a programme of 37 °C for 15 min and 74 °C for 15 min and held at 4 °C. Addition of Illumina Nextera XT 384-way indexing primers to the cleaned first round PCR products were amplified following a single protocol which started with initial denaturation at 98 °C for 3 min; 15 cycles of 95 °C for 30 s; 55°C for 30 s; 72 °C for 30 s; final extension at 72 °C for 5 min and held at 4 °C. Twenty-five µL of second-round PCR products were purified with an equal amount of AMPure XP beads (Beckman Coulter). Library preparation for the 2013 samples was conducted at Bangor University. Illumina sequencing for both years and library preparation for 2014 samples were conducted at the Liverpool Centre for Genome Research.

5.2.4 Bioinformatics

Bioinformatics analyses were performed on the Supercomputing Wales cluster as previously described in George et al. (2019). A total of 104,276,828, and 98,999,009 raw reads were recovered from the ITS1 and 18S sequences, respectively. Illumina adapters were trimmed from sequences using Cutadapt (Martin, 2011) with 10% level mismatch for removal. Sequences were then de-multiplexed, filtered, quality-checked, and clustered using a combination of USEARCH v. 7.0 (Edgar, 2010) and VSEARCH v. 2.3.2 (Rognes et al., 2016). Open-reference clustering (97% sequence similarity) of operational taxonomic units (OTUs) was performed using VSEARCH; all other steps were conducted with USEARCH. Sequences with a maximum error greater than 1 and shorter than 200 bp were removed following the merging of forward and reverse reads for ITS1 sequences. A cut-off of 250 bp was used for

18S sequences, according to higher quality scores. There were 7,242,508 (ITS1) and 9,163,754 (18S) cleaned reads following these steps. Sequences were sorted and those that only appeared once in each dataset were removed.

Remaining sequences were matched first against the UNITE 7.2 (Kõljalg et al., 2013) and SILVA 128 (Quast et al., 2013) databases for the ITS1 and 18S sequences, respectively. Ten per cent of sequences that failed to match were clustered *de novo* and used as a new reference database for failed sequences. Sequences that failed to match with the *de novo* database were subsequently also clustered *de novo*. All clusters were collated and chimeras were removed using the uchime_ref command in VSEARCH. Chimera-free clusters and taxonomy assignment were summarised in OTU tables with QIIME v. 1.9.1 (Caporaso et al., 2010) using RDP (Wang et al., 2007) methodology with the UNITE database for ITS1 data. Taxonomy was assigned to the 18S OTU table using BLAST (Altschul et al., 1990) against the SILVA database and OTUs appearing only once or in only 1 sample were removed from each OTU table. Based on DNA quality and read counts, 413 samples were used for analyses of the ITS1 data and 422 for 18S data (from the total of 438).

A Newick tree was constructed for the 18S tables using 80% identity thresholds and was paired with the 18S OTU table as part of analyses using the R package phyloseq (McMurdie and Holmes, 2013). Non-fungi OTUs were removed from both OTU tables. Read counts from each group were rarefied 100 times using phyloseq (as justified by Weiss et al. (2017)) and the resulting mean richness was calculated for each sample. The ITS1 table was rarefied at a depth of 4,000 reads whereas the 18S table was rarefied to 10,000 reads. A subset of the 18S data was rarefied to 400 reads across 398 samples to analyse Glomeromycetes OTUs separately. Samples with observed lower read counts were removed before rarefaction.

To assess functional diversity, both OTU tables were processed using FUNGUILD (Nguyen et al., 2016) and the resulting matched OTU tables were used to investigate functional roles based on trophic mode. Sequences have been uploaded to The European Nucleotide Archive and can be accessed with the following primary accession codes after the end of the data embargo: PRJEB28028 (ITS1), and PRJEB28067 (18S).

5.2.5 Statistical analysis

All statistical analyses were run using R v. 3.3.3 (R Core Team, 2017) following rarefaction. For each data set, NMDS ordinations using Bray-Curtis dissimilarity were created with the vegan package (Oksanen et al., 2016) to assess β-diversity. Environmental data was fitted linearly onto each ordination of AVCs using the envfit function. NMDS scores were plotted against these values for each variable to determine the direction of associations. Differences in β-diversity amongst AVCs were calculated with PERMANOVA and homogeneity of dispersion was also assessed.

Linear mixed models were constructed using package nlme (Pinheiro et al., 2016) to show the differences in α-diversity amongst AVCs, soil types, and LOI classification, for both ITS1 and 18S fungal data sets. Sample year as fixed factors; sample square identity was the random factor. This methodology was also used for the subsets of data that matched to the FUNGUILD database. For each model, significant differences were assessed by ANOVA and pairwise differences were identified using Tukey's *post-hoc* tests from the multcomp package (Hothorn et al., 2008).

Partial least squares regressions from the pls package (Mevik et al., 2016) were used with the variable importance in projection (VIP) approach (Chong and Jun, 2005) to sort the original explanatory variables by order of importance to identify the most important

environmental variables for richness. Such analysis is ideal for data where there are many more explanatory variables than sample numbers or where extreme multicollinearity is present (Lallias et al., 2015; George et al., 2019). Variables with VIP values > 1 were considered most important. Relationships between important variables and richness values for each group of organisms were investigated by linear regression. Richness was normalised before regression when necessary.

5.3 Results

5.3.1 Soil properties

Soil properties displayed a range of changes across land uses (Table 5.1). Notably, total C ($F_{6, 427} = 89.13$ $p < 0.001$), total N ($F_{6, 427} = 61.03$, $p < 0.001$), C:N ratio ($F_{6, 427} = 94.41$, $p < 0.001$), organic matter content ($F_{6, 428} = 107.02$, $p < 0.001$), elevation ($F_{6, 429} = 78.42$, $p < 0.001$), and mean annual precipitation ($F_{6, 429} = 72.6$, $p < 0.001$), and moisture ($F_{6, 427} = 33.74$, $p < 0.001$) increased with declining land use productivity. We also observed a reduction in pH ($F_{6, 428} = 69.56$, $p < 0.001$), bulk density ($F_{6, 428} = 79.87$, $p < 0.001$), and clay content ($F_{6, 344} = 19.54$, $p < 0.001$) across the land use productivity gradient. Trends in other variables such as soil water repellency ($F_{6, 428} = 22.08$, $p < 0.001$), total P ($F_{6, 424} = 7.1$, $p < 0.001$), sand content ($F_{6, 344} = 5.71$, $p < 0.001$), stone content ($F_{6, 427} = 10.4$, $p < 0.001$), and temperature at time of sampling ($F_{6, 429} = 4.4$, $p < 0.001$), though significant, were less clear across land uses however. These findings were also apparent when samples were grouped from low-to-high organic matter content by organic matter class (Appendix 3 Table 2). Overall, no clear trends were evident across the different soil types (Appendix 3 Table 3).

5.3.2 Sequencing Data

A total of 7,582 and 4,408 fungal OTUs were recovered using the ITS1 and 18S primer sets, respectively. Of these, 5,666 were assigned an identifier at the class-level in the ITS1 dataset while 4,367 were assigned an identifier in the 18S dataset. There were 15 classes that were only found in the ITS1 dataset and 12 unique to the 18S data. Endogonomycetes was the most abundant class found only in the ITS dataset (19 OTUs), whereas Laboulbeniomycetes (17 OTUs) was the most abundant fungal class unique to the 18S data. A total of 24 classes were present in both ITS1 and 18S data (Fig. 5.2A).

As reported in George et al. (2019), Agaricomycetes were the most abundant class of fungi in the ITS1 dataset overall. There were also a large proportion of Sordariomycetes (Fig. 5.2B). Archaeorhizomycetes was the most abundant class in the 18S dataset (Fig. 5.2C). Proportionate abundances of Sordariomycetes and Agaricomycetes followed contrasting trends, with the dominance of the former replaced by the latter in lower productivity AVCs in the ITS1 data, as described previously (Fig. 5.3A). Although Agaricomycetes and Sordariomycetes comprised smaller fractions of the 18S dataset (Fig. 5.2C), this trend was still apparent (Fig. 5.3B). Additionally, the Archaeorhiozmycetes from 18S data generally followed the same trend as the Sordariomycetes (Fig. 5.3B). The preceding trends observed across land uses are also evident across organic matter classes (Appendix 3 Fig. 1) but are not as clear across soil types (Appendix 3 Fig. 2).

When a class was present in both datasets, it was usually much more prevalent in one than the other (Appendix 3 Table 4). For example, there were 1858 Agaricomycetes and 915 Sordariomycetes OTUs in the ITS1, yet these numbers dropped to 646 and 417 OTUs in the 18S dataset. Similarly, Glomeromycetes accounted for 162 of the OTUs in the 18S data, but

only 6 OTUs in the ITS1 dataset. Abundances of classes unique to the ITS1 and 18S datasets can be found in Appendix 3 Table 5 and Appendix 3 Table 6, respectively.

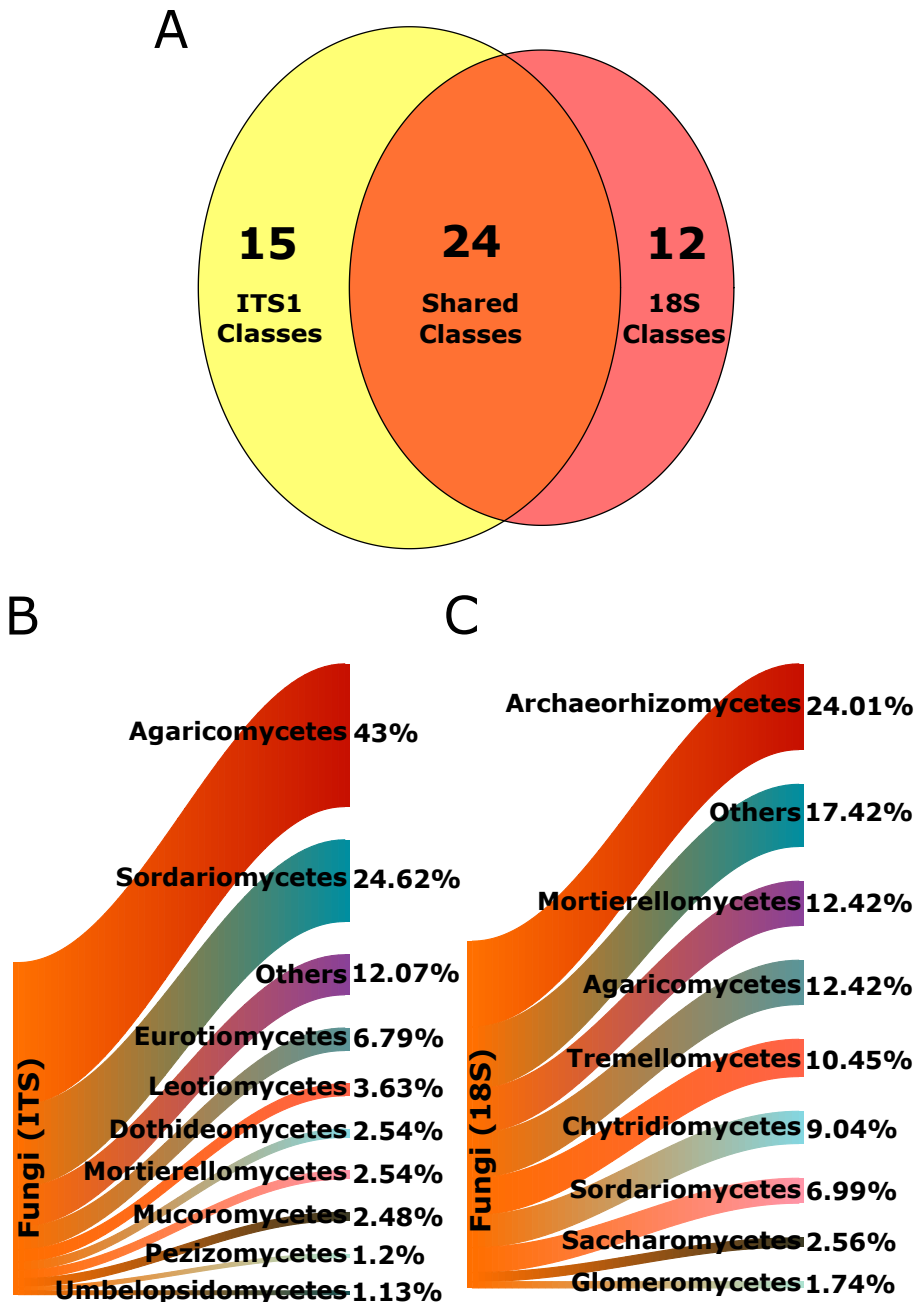


*Figure 5.2 Composition of fungal classes from ITS1 and 18S datasets. **A**) Venn diagram denoting total number of shared and unique classes in each data set, following exclusion of unknown sequences. Sankey diagrams of proportional abundances of fungal OTUs from all samples from **B**) ITS1 data and **C**) 18S data. Arms denote proportions of OTUs of the most populous classes.*

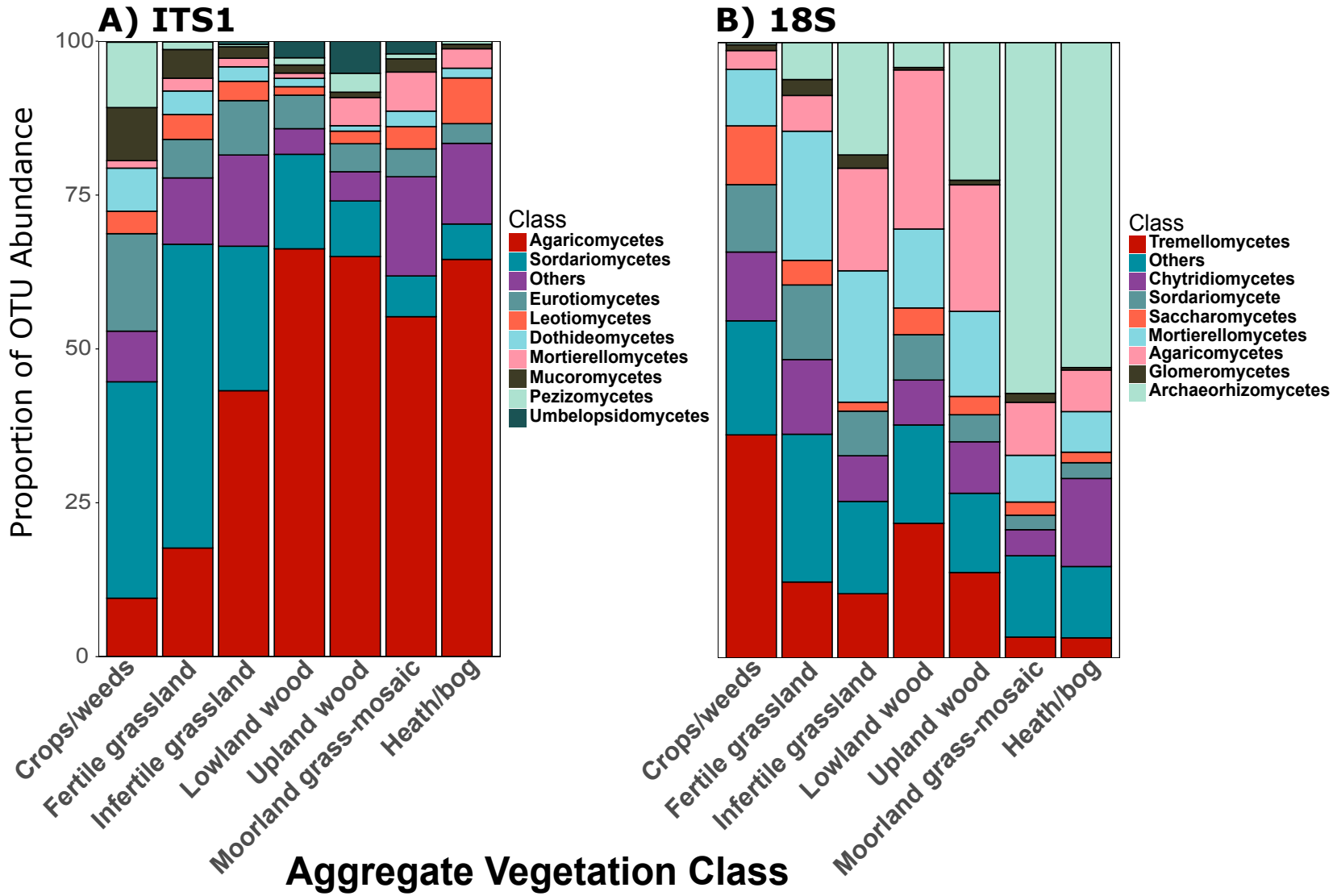


*Figure 5.3 Proportionate abundances of fungal OTUs for **A)** ITS1 and **B)** 18S data across Aggregate Vegetation Class. Aggregate Vegetation Classes are ordered from most (Crops/weeds) to least (Heath/bog) productive.*

Table 5.1 Mean values (± SE) of soil physical and chemical variables for each Aggregate Vegetation Class. Following normalisation on selected variables (see below), ANOVAs and Tukey's post-hoc tests were performed. Numbers followed by the same letter in the same row are not significantly different.

Environmental variable	Crops/weeds	Fertile grassland	Infertile grassland	Lowland wood	Upland wood	Moorland grass-mosaic	Heath/bog
Total C (%)^{L}	3.87 (± 0.83)d	4.75 (± 0.2)d	5.85 (± 0.33)d	5.78 (± 1.07)d	9.7 (± 2.25)c	12.19 (± 2.07)b	23.57 (± 1.88)a
Total N (%)^{L}	0.32 (± 0.05)d	0.45 (± 0.02)d	0.49 (± 0.02)d	0.4 (± 0.06)d	0.58 (± 0.1)c	0.83 (± 0.11)b	1.05 (± 0.09)a
C:N ratio^{S}	11.44 (± 0.81)cd	10.49 (± 0.13)d	11.62 (± 0.27)cd	13.92 (± 0.75)bc	15.86 (± 0.7)b	14.41 (± 0.42)b	20.65 (± 0.94)a
Total P (mg/kg)^{S}	1103.44 (± 145.47)ab	1194.9 (± 45.53)a	1045.5 (± 43.3)ab	601.68 (± 77.68)c	762.45 (± 61.95)bc	930.49 (± 57.5)ab	769.63 (± 50.04)ab
Organic matter (% LOI)^{L}	7.53 (± 1.62)d	9.39 (± 0.34)d	11.25 (± 0.55)d	10.71 (± 1.7)d	18.79 (± 4.16)c	22.99 (± 3.72)b	39.26 (± 3.6)a
pH ($CaCl_2$)	4.73 (± 0.26)b	5.2 (± 0.08)a	4.73 (± 0.05)b	4.31 (± 0.26)b	3.57 (± 0.1)cd	3.85 (± 0.09)c	3.84 (± 0.1)d
Soil water repellency*	4077.56 (± 3990.72)abc	264.01 (± 73.28)c	781.68 (± 137.58)b	2975.47 (± 2108.12)abc	1965.87 (± 698.61)a	4186.13 (± 798.48)a	3186.4 (± 812.15)a
Volumetric water content (m^3/m^3)	0.23 (± 0.03)bc	0.35 (± 0.01)b	0.34 (± 0.01)b	0.22 (± 0.02)c	0.36 (± 0.03)b	0.46 (± 0.02)a	0.52 (± 0.02)a
Rock volume (mL)	3.95 (± 1.11)abc	5.25 (± 0.45)b	5.44 (± 0.42)b	9.13 (± 2.49)a	4.41 (± 0.57)ab	3.25 (± 0.39)c	1.87 (± 0.21)c
Bulk density (g/cm^3)	1.03 (± 0.09)a	0.9 (± 0.02)a	0.8 (± 0.02)b	0.71 (± 0.08)b	0.56 (± 0.04)c	0.5 (± 0.04)c	0.47 (± 0.03)d
Clay content (%)^{A}	22.25 (± 1.85)ab	25.46 (± 0.65)a	23.18 (± 0.64)ab	17.47 (± 1.34)ab	17.82 (± 1.82)ab	18.12 (± 1.27)c	11.76 (± 2.24)d
Sand content (%)^{A}	30.97 (± 4.66)ad	24.88 (± 1.25)d	29.21 (± 1.44)bd	42.99 (± 4.01)ac	40.23 (± 4.15)abc	29.5 (± 3.0)b	45.15 (± 7.61)a
Elevation (m)	88.71 (± 47.69)cd	109.38 (± 8.62)d	167.28 (± 8.65)c	119.06 (± 16.38)cd	297.83 (± 20.62)b	406.63 (± 19.22)a	380.55 (± 19.7)a
Mean annual precipitation (mL)	968.44 (± 69.01)c	1078.19 (± 24.71)c	1177.05 (± 18.91)c	1100.12 (± 52.28)c	1405.33 (± 65.35)b	2027.23 (± 74.39)a	1771.2 (± 58.19)a
Temperature (°C)	12.64 (± 1.18)ab	12.09 (± 0.41)b	13.44 (± 0.29)a	15.80 (± 0.87)a	14.53 (± 0.53)a	14.51 (± 0.36)a	13.87 (± 0.29)a

Note: ^{A} denotes Aitchison's $\log_{10}$-ratio transformation; ^{L} denotes $\log_{10}$-transformation; square-root-transformation; *Soil water repellency was derived from median water drop penetration times (s) and $\log_{10}$ transformed.

5.3.3 Fungal Richness and β-Diversity from ITS1 and 18S Data

We found that fungal richness followed the same trends across land use, irrespective of primer set. As previously demonstrated in George et al. (2019), fungal OTU richness from ITS1 metabarcoding significantly declined ($F_{6, 258} = 39.87$, $p < 0.001$; Fig. 5.4A) from high to low productivity/management intensity. Richness in Fertile grasslands was significantly greater than all other AVCs ($p < 0.001$) except Crops/weeds. In the 18S dataset, richness was also significantly higher ($F_{6,267} = 82.73$, $p < 0.001$) in more productive/managed land uses and declined along this gradient. However, richness in grasslands was highest in this dataset (Fig. 4B). For complete pairwise differences between land uses see Supplementary Material.

The trend of declining richness with productivity was also apparent when samples were categorised by organic matter content (Fig. 5.5). In both datasets, richness was significantly greater ($F_{3, 259} = 48.13$, $p < 0.001$; $F_{3, 269} = 46.71$, $p < 0.001$; for ITS1 and 18S, respectively) in mineral and humus-mineral than all other classifications (ITS1, Fig. 5.5A; 18S, Fig. 5.5B). There was no consistent pattern of richness when soils were categorised by soil type (Appendix 3 Fig. 3). Again, pairwise differences between organic matter classes and soil types are described in Appendix 3.2.

Community composition based on non-metric multidimensional scaling of Bray-Curtis distances also showed consistent trends between the datasets. Plots demonstrate tight clustering of Crops/weeds, and grassland AVCs in both ITS1 (Fig. 5.6A) and 18S (Fig. 5.6B) compared to the wide dispersal of other AVCs. Such results are supported by PERMANOVAs, which show significant differences ($F_{6, 406} = 10.74$, $p = 0.001$; $F_{6, 415} = 15.65$, $p = 0.001$); however, analyses of dispersion were also significant ($F_{6, 406} = 41.30$,

$p = 0.001$; $F_{6, 415} = 10.69$, $p = 0.001$) as a result of the large disparity in replicates between land uses.

When these results are visualised by organic matter classification, the tight clusters are populated by mineral and humus-mineral samples, whereas organo-mineral and organic samples are more common in the widely dispersed areas of the plots (Appendix 3 Fig. 4 and Appendix 3 Fig. 5). Soil types are more widely dispersed but Brown and Surface-water gley soils are more common in the tightly grouped area (Appendix 3 Fig. 6 and Appendix 3 Fig. 7). Again, significant results were observed for both PERMANOVA and dispersion of variance across organic matter classes and soil types in both datasets.

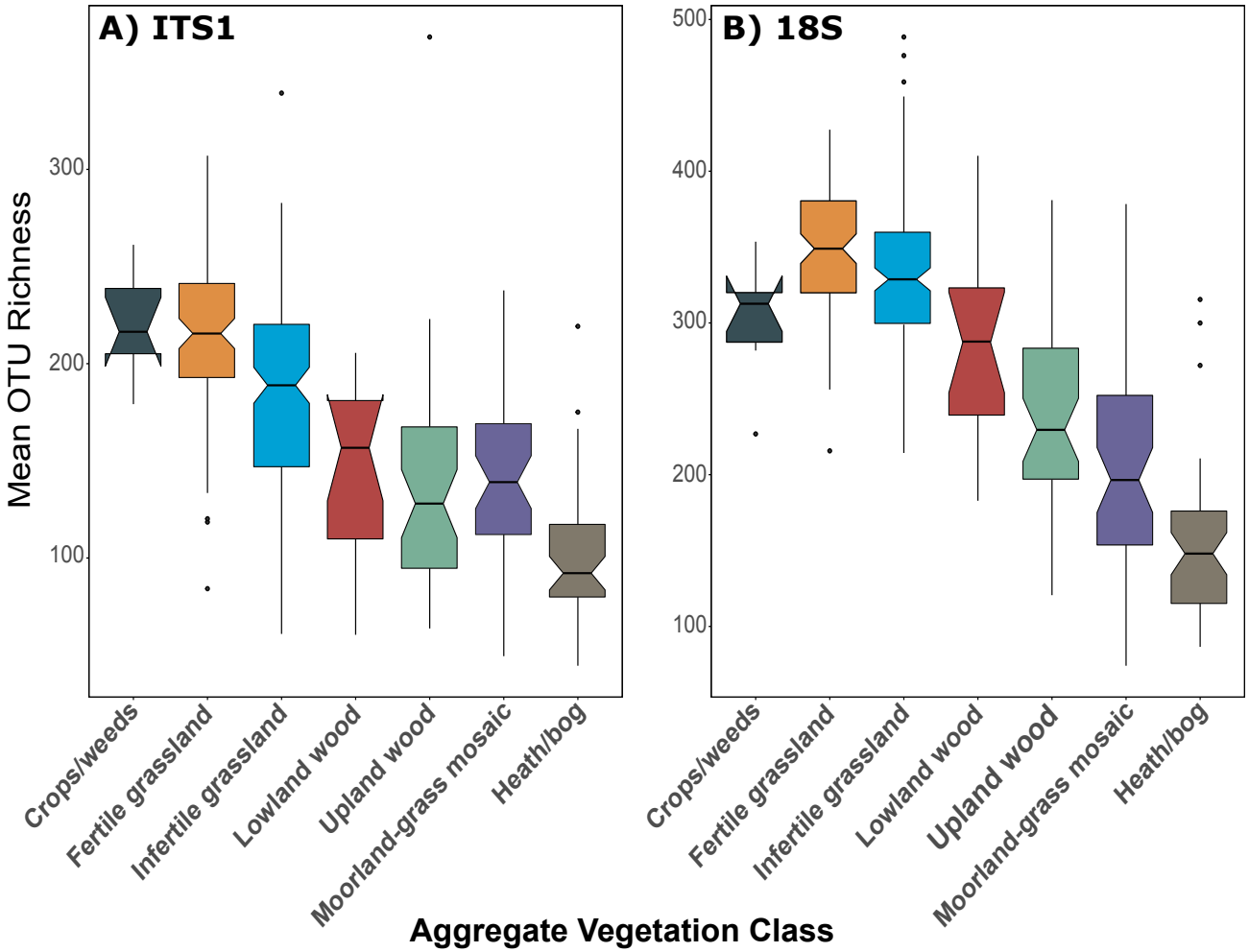


*Figure 5.4 Boxplots of fungal OTU richness for **A**) ITS1 and **B**) 18S datasets plotted against Aggregate Vegetation Class. Aggregate Vegetation Classes are ordered from most (Crops/weeds) to least (Heath/bog) productive. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers.*

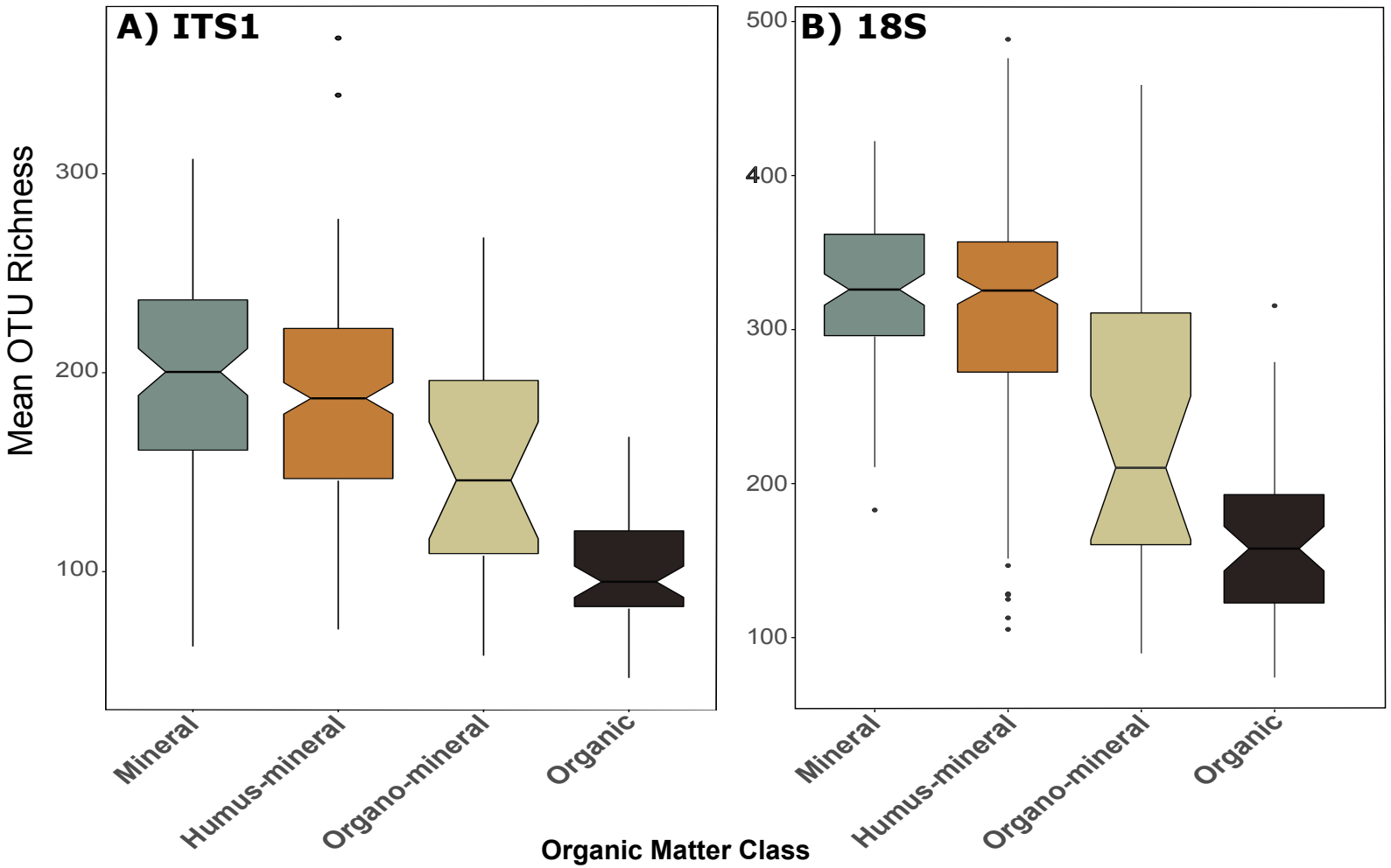


*Figure 5.5 Boxplots of fungal OTU richness for **A)** ITS1 and **B)** 18S datasets plotted against organic matter class. Organic matter classes are listed in order of increasing percent organic matter. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers.*

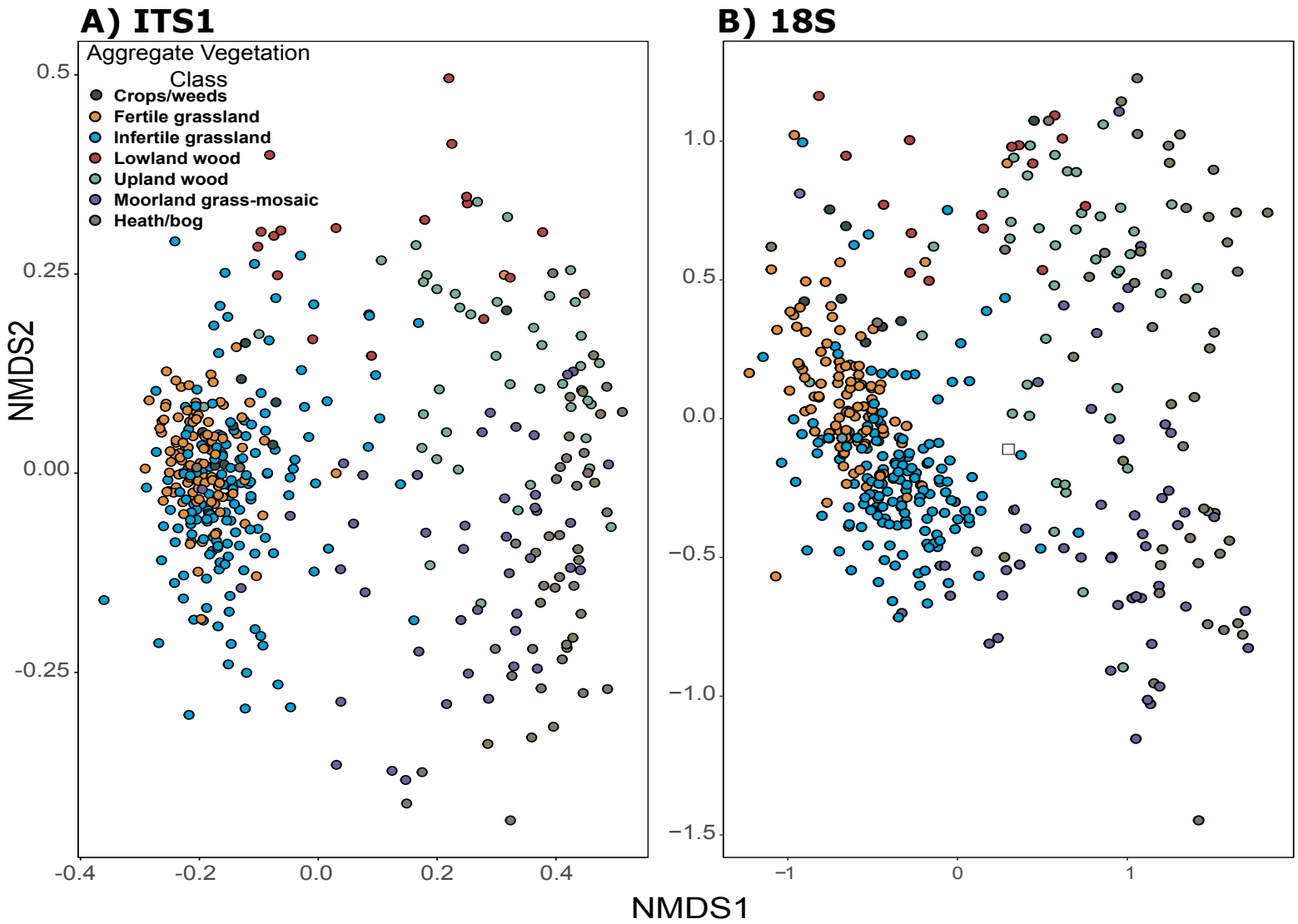


*Figure 5.6 Non-metric dimensional scaling ordinations of fungal community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Data from ITS1 (stress = 0.13) is shown in **A**); data from 18S (stress = 0.11) is shown in **B**).*

5.3.4 Relationships Between Soil Properties and Fungal Biodiversity

Fungal richness showed similar relationships to soil properties in both datasets. Across samples, PLS and VIP analyses highlighted strong correlations between fungal richness and soil properties. There were significant, positive relationships of richness with pH and bulk density; and significant, negative correlations between richness and C:N ratio, organic matter, elevation, and mean annual precipitation (Table 5.2). Although these results followed the same trend in ITS1 and 18S data, however, their relative rankings varied. For example, fungal richness from ITS1 data was most strongly correlated with bulk density and organic matter, while richness from 18S data was more strongly correlated to C:N ratio and elevation in addition to bulk density (Table 5.2). Furthermore, there were some relationships unique to each dataset. Significant negative relationships were observed between richness and soil water repellency. Similarly, richness derived from 18S data was negatively related to total C and sand content of soil but also positively related to clay content.

We found pH was the best predictor of β-diversity from linear fitting for fungi no matter what gene region is amplified (Table 5.3 and Table 5.4). All fitted variables were significantly correlated to β-diversity, though most of these only weakly. It is likely that they did not strongly influence the fungal communities. Variables followed similar rankings in both the ITS1 and 18S data. Elevation, annual precipitation, soil moisture, C:N ratio, organic matter, and bulk density all had R^2 values greater than 0.35, but their relative order differed between datasets (Table 5.3 and Table 5.4).

*Table 5.2 Results of partial least squares regressions for fungal richness against environmental variables. Positive relationships are underlined; negative relationships are written in italics. *** indicates P < 0.001, blank indicates P > 0.05.*

Soil and environmental variables	Fungi (ITS)	Fungi (18S)
Total C^{L}	0.44	*1.03 (R^2 = 0.38***)*
Total N^{L}	0.93	0.56
C:N ratio^{S}	*1.64 (R^2 = 0.28***)*	*1.71 (R^2 = 0.41***)*
Total P^{S}	0.70	0.87
Organic matter (% LOI)^{L}	*1.13 (R^2 = 0.29***)*	*1.17 (R^2 = 0.38***)*
pH ($CaCl_2$)	1.52 (R^2 = 0.23***)	1.55 (R^2 = 0.37***)
Soil water repellency^{L}	*1.23 (R^2 = 0.13***)*	0.82
Volumetric water content (m^3/m^3)	0.60	0.70
Rock volume (mL)	0.64	0.43
Bulk density (g/cm^3)	1.41 (R^2 = 0.29***)	1.33 (R^2 = 0.41***)
Clay content (%)^{A}	0.84	1.19 (R^2 = 0.11***)
Sand content (%)^{A}	0.6	*1.11 (R^2 = 0.1***)*
Elevation (m)	*1.68 (R^2 = 0.22***)*	*1.83 (R^2 = 0.41***)*
Mean annual precipitation (mL)	*1.44 (R^2 = 0.18***)*	*1.52 (R^2 = 0.27***)*
Temperature (°C)	0.56	0.52

Note: ^{A} denotes Aitchison's log_{10}-ratio transformation; ^{L} denotes log_{10}-transformation; ^{S} denotes square-root-transformation.

*Table 5.3 Summary of relationships amongst environmental factors and fungal communities based on ITS data. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, blank indicates P > 0.05.*

Variable		**Correlation**	
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.6^{***}	-	+
C:N ratio[S]	0.47^{***}	+	-
Elevation (m)	0.41^{***}	+	-
Volumetric water content (m^3/m^3)	0.41^{***}	+	-
Mean annual precipitation (mL)	0.39^{***}	+	-
Bulk density (g/cm^3)	0.38^{***}	-	+
Organic matter (% LOI)[L]	0.37^{***}	+	-
Total C[L]	0.31^{***}	+	-
Clay content (%)[A]	0.28^{***}	-	+
Soil water repellency[L]	0.24^{***}	+	-
Total N (%)[L]	0.21^{***}	+	-
Sand content (%)[A]	0.19^{***}	+	+
Total P (mg/kg)[S]	0.11^{***}	-	-
Rock volume (mL)	0.07^{***}	-	+
Temperature (°C)	0.04^{***}	-	+

Note: [A] denotes Aitchison's log_{10}-ratio transformation; [L] denotes log_{10}-transformation; [S] denotes square-root-transformation.

*Table 5.4 Summary of relationships amongst environmental factors and fungal communities based on 18S data. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, blank indicates P > 0.05.*

Variable		**Correlation**	
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.61^{***}	-	+
Elevation (m)	0.50^{***}	+	-
Mean annual precipitation (mL)	0.46^{***}	+	-
Volumetric water content (m^3/m^3)	0.45^{***}	+	-
C:N ratio^S	0.43^{***}	+	+
Organic matter (% LOI)^L	0.43^{***}	+	+
Bulk density (g/cm^3)	0.39^{***}	-	-
Total C^L	0.34^{***}	+	+
Clay content (%)^A	0.30^{***}	-	+
Total N (%)^L	0.28^{***}	+	-
Soil water repellency^L	0.21^{***}	+	-
Sand content (%)^A	0.14^{***}	+	+
Total P (mg/kg)^S	0.10^{***}	-	-
Rock volume (mL)	0.06^{***}	-	+
Temperature (°C)	0.05^{***}	-	+

Note: ^{A} denotes Aitchison's log_{10}-ratio transformation; ^{L} denotes log_{10}-transformation; ^{S} denotes square-root-transformation

5.3.5 Effect of Land Use on Functional Diversity

There was a distinct difference in trophic modes of OTUs that were successfully matched to the FUNGUILD database between ITS1 and 18S datasets. In total, 3,402 and 1,783 OTUs from the ITS1 and 18S datasets respectively were matched to the FUNGUILD database. Overall, saprotrophs were the most abundant trophic mode in both datasets (Fig. 5.6); however, pathotrophs ranked second in ITS1 (Fig. 5.6A) data while the pathotroph-saprotroph-symbiotroph multi-trophic group was second-most abundant in 18S data (Fig. 5.6B). Across land uses, proportions of pathotrophs and pathotroph-saprotroph-symbiotrophs fell with declining productivity (Fig. 5.7). In matches from the ITS1 data, pathotroph-saprotrophs

increased across the productivity gradient (Fig. 5.7A), as did saprotrophs in the 18S data (Fig. 5.7B). The aforementioned trend in proportional abundance of pathotrophs and pathotroph-saprotroph-symbiotrophs was also present across organic matter classes (Appendix 3 Fig. 8). Symbiotrophs appeared to follow an opposite trend, increasing as productivity fell. Interestingly, this was the case for saprotrophs in the 18S (Appendix 3 Fig. 8B) but not the ITS1 (Appendix 3 Fig. 8A) dataset. Proportional abundances of fungal OTUs grouped by trophic modes did not follow a discernable pattern across changing soil types (Appendix 3 Fig. 9). For simplicity, we focused further analyses only on the broadly defined saprotroph, pathotroph, and symbiotroph groups, ignoring all combination groups; pairwise differences for all of the following comparisons are described in Appendix 3.2.

Across land uses, significant differences were observed in the richness of saprotrophic fungi in both the ITS1 ($F_{6,258} = 25.14$, $p < 0.001$) and 18S ($F_{6, 267} = 31.10$, $p < 0.001$) data; however, there were differences between datasets (Fig. 5.8). In the ITS1 dataset, richness followed the same trend as overall fungal richness, with the highest and lowest values in the Crops/weeds and Heath/bog AVCs respectively (Fig. 5.8A). Although this pattern was preserved in the 18S data (Fig. 5.8B), richness of saprotrophs was much more even across AVCs in this case. Indeed, rather than the linear decline of richness along the productivity gradient, there appeared to be 3 distinct levels in the data affiliated with (i) grassland/agricultural sites, (ii) woodlands, and (iii) bogs.

The same pattern was also apparent across organic matter classifications in both datasets (ITS1: $F_{3, 260} = 32.86$, $p < 0.001$; 18S: $F_{3, 269} = 41.13$, $p < 0.001$; Fig. 5.9). In the ITS1 dataset, each class was significantly different from the others (Fig. 5.9A). In the 18S data, saprotroph richness was significantly higher in mineral and humus-mineral soils than organo-

mineral and organic soils (all $p < 0.001$ except mineral – organo-mineral $p = 0.02$) (Fig. 5.9B). Again, the overarching trend of fungal richness was not apparent when samples were grouped by soil type. Although there were significant differences across soil types in both the ITS1 ($F_{5, 259} = 9.7$, $p < 0.001$) and 18S ($F_{5, 268} = 10.73$, $p < 0.001$) datasets, these differences did demonstrate consistent patterns across soil types (Appendix 3 Fig. 10).

In the case of pathotrophic fungi, richness also followed a similar trend to the saprotrophs across both datasets. In the ITS1 data, significantly ($F_{6, 258} = 26.11$, $p < 0.001$) greater richness values were observed in Crops/weeds and grassland samples (Fig. 5.8A). Richness of pathotrophs was significantly highest in Crops/weeds sites. Again, this trend was present, though not as clear, in the 18S dataset (Fig. 5.8B). Significant differences ($F_{6, 267} = 52.26$, $p < 0.001$) were observed between AVCs, with the highest richness of pathotrophs occurring in the Fertile grassland and Crop/weeds land uses.

Across organic matter classes, significant differences were also observed in pathotroph richness in the ITS1 ($F_{3, 250} = 24.91$, $p < 0.001$) and 18S ($F_{3, 269} = 30.49$, $p < 0.001$) datasets. However, in this case the trends were more apparent in the 18S data than the ITS1 data (Fig. 5.9). Pathotroph richness was highest in mineral soils and lowest in organic soils when compared to all other classes in the ITS1 data (Fig. 5.9A). However, all organic matter classifications were statistically different from each other in the 18S data (Fig. 5.9B), in descending order from mineral to peat soils. Again, trends were less clear across soil types (Appendix 3 Fig. 10). Significant differences were observed in the ITS1 data ($F_{5, 259} = 6.93$, $p < 0.001$) with the lowest pathotroph richness found in peat soils (Appendix 3 Fig. 10A). In the 18S data, differences between pathotrophic fungi across soil types were more similar to those observed in other groups (Appendix 3 Fig. 10B). Pathotroph richness was significantly ($F_{5, 268}$

= 13.6, $p < 0.001$) different across soil types with the highest values found in brown soils and the lowest in peats.

The previously described trend of declining richness across the land use productivity gradient (i.e. Fig. 5.4) was not apparent when considering symbiotrophs. Furthermore, although significant differences were apparent in both the ITS1 ($F_{6, 258} = 14.88$, $p < 0.001$) and 18S ($F_{6, 267} = 55.13$, $p < 0.001$) datasets they were by no means identical (Fig. 8). Symbiotroph richness was highest in Lowland wood sites followed by Upland wood. This trend was not apparent in the 18S dataset, however (Fig. 5.8B). Here richness of symbiotrophs was greatest in grassland AVCs and lowest in Heath/bog sites much like the overarching trend of total fungal OTU richness.

When samples were grouped by organic matter class, further discrepancies became apparent between the datasets. Whereas the previously described trend of decreasing richness with increasing organic matter content held true in the 18S data ($F_{3, 269} = 36.28$, $p < 0.001$; Fig. 5.9B), no significant differences were observed in the ITS1 dataset ($F_{3, 260} = 1.88$, $p = 0.13$; Fig 5.9A). In the 18S data, richness of symbiotrophs was greater in mineral and humus-mineral soils when compared to organo-mineral ($p = 0.002$, $p = 0.04$, respectively) and organic ($p < 0.001$) soils (Fig. 5.9B). There were also no significant differences ($F_{5, 259} = 1.43$, $p = 0.21$) in symbiotroph richness across soil types in ITS1 data (Appendix 3 Fig. 10A), though there were in 18S data ($F_{5, 259} = 12.52$, $p < 0.001$; Appendix 3 Fig. 10B). As described previously, richness was lowest in peat soils and highest in brown soils.

We suspected that the differences in functional diversity observed between datasets might be a result of differential coverage of important groups. We were able to confirm this when we analysed the richness of OTUs identified as Glomeromycetes present in the 18S

dataset (Fig. 5.10). All of the 162 Glomeromycetes OTUs were assigned as highly-probable symbiotrophs through FUNGUILD. Across land uses, richness of Glomeromycetes followed similar trends to those of symbiotrophs and saprotrophs from 18S data. There were significant ($F_{6, 244} = 33.47$, $p < 0.001$) differences across land uses, though they appeared, like the saprotroph richness to be tiered between grasslands, woods, and bogs (Fig. 5.10A). Richness of Glomeromycetes was higher in grasslands than all other AVCs except Crops/weeds and lowest in Heath/bog sites. Again, when grouped by organic matter class (Fig. 5.10B) and soil type (Fig. 5.10C) Glomeromycetes richness followed the same trend as saprotrophs and symbiotrophs from the 18S dataset. Richness was significantly ($F_{3, 246} = 37.65$, $p < 0.001$) greater in mineral and humus-mineral soils than all others. Across soil types, richness of Glomeromycetes was significantly ($F_{5, 245} = 8.65$, $p < 0.001$) lower in peat soils when compared to most other soil types.

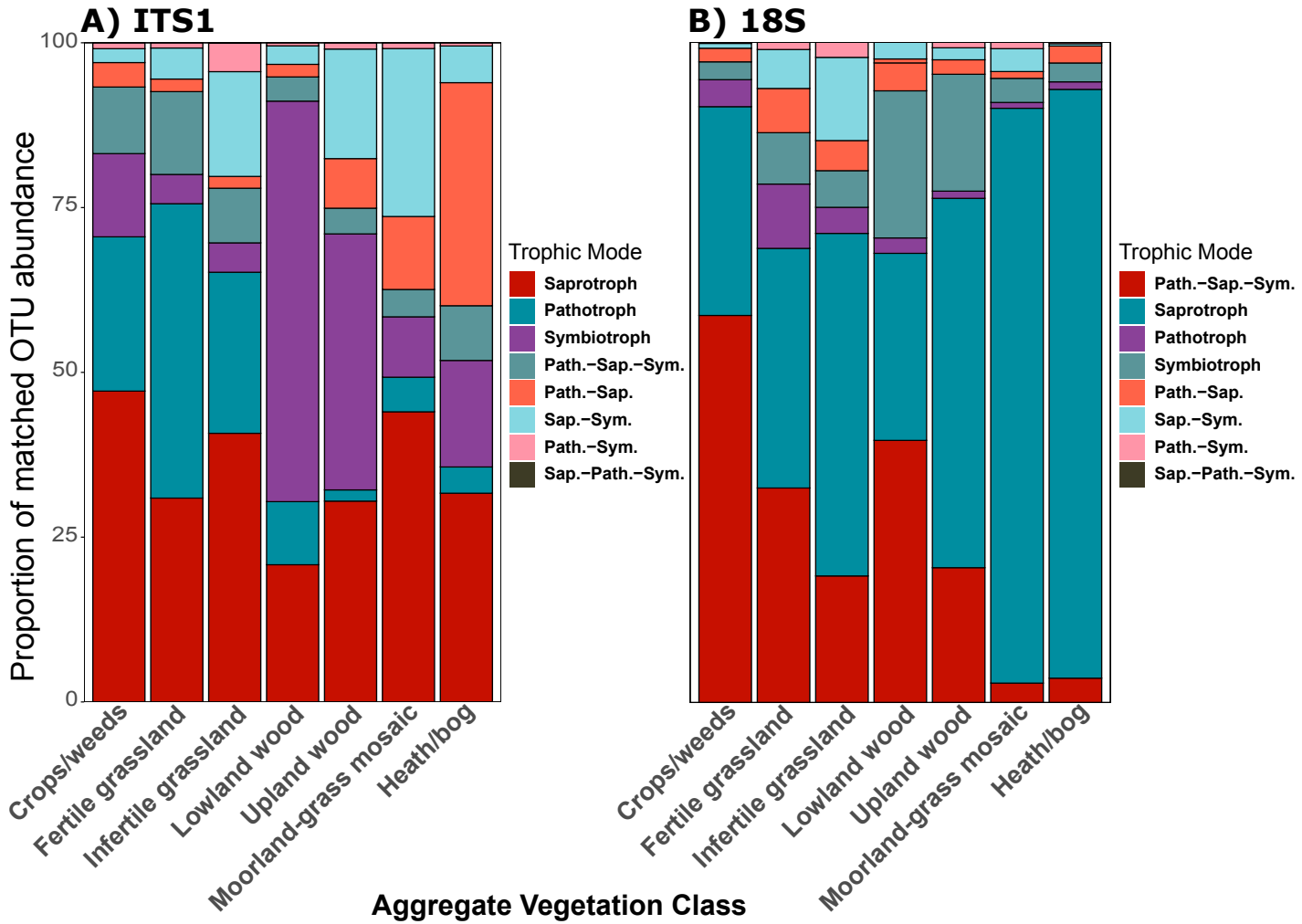


*Figure 5.7 Proportionate abundances of fungal OTUs matched to FUNGuild trophic groups for **A)** ITS1 and **B)** 18S data across Aggregate Vegetation Classes. Aggregate Vegetation Classes are ordered from most (Crops/weeds) to least (Heath/bog) productive. Abbreviations for multi-trophic mode groups are as follows: Path.-Sap. (Pathotroph-Saprotroph); Path.-Sap.-Sym. (Pathotroph-Saprotroph-Symbiotroph); Path.-Sym. (Pathotroph-Symbiotroph); Sap.-Path.-Sym (Saprotroph-Pathotroph-Symbiotroph); Sap.-Sym. (Saprotroph-Symbiotroph).*

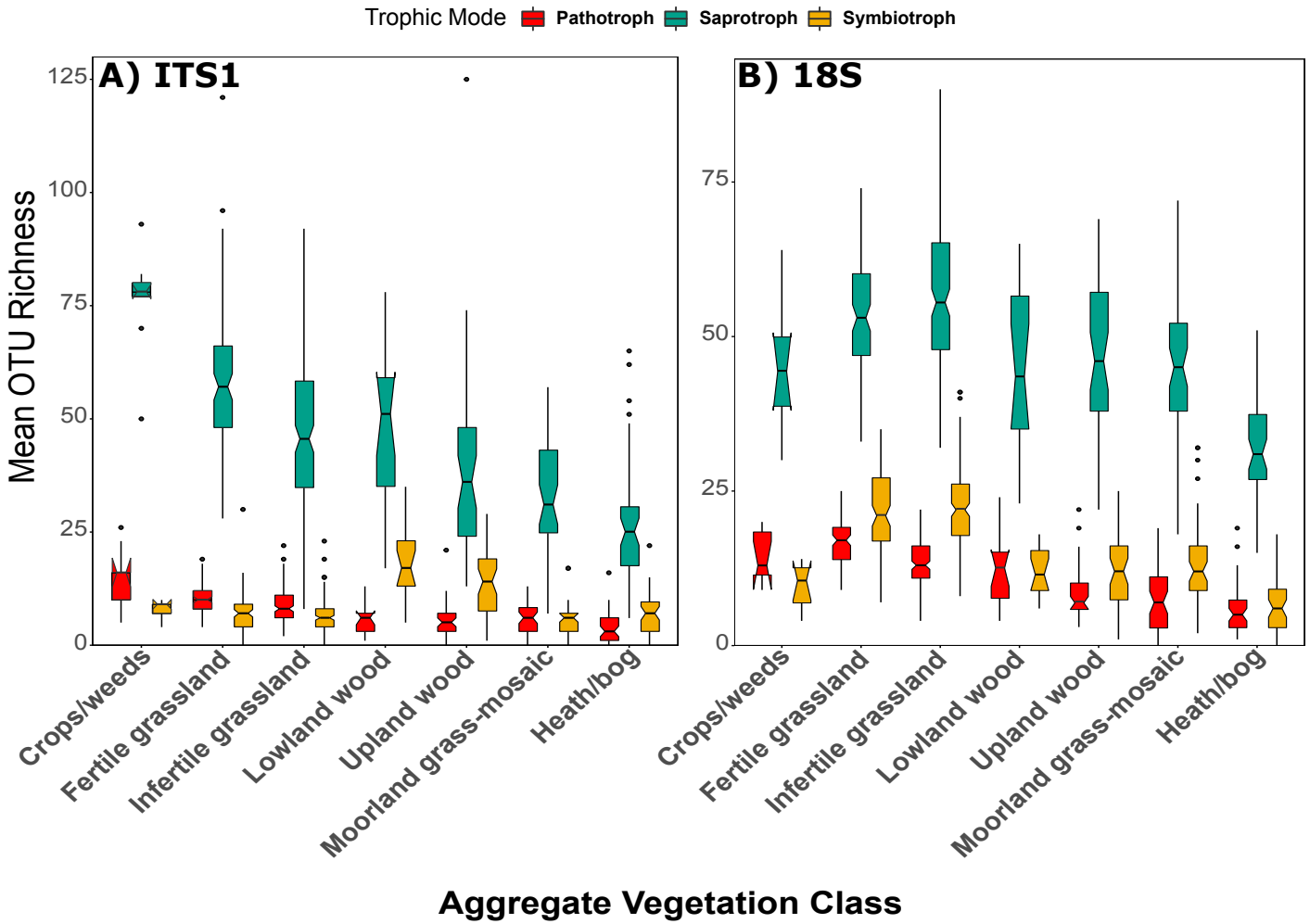


*Figure 5.8 Boxplots of richness of fungal OTUs matched to the pathotrophic, saprotroph, and symbiotroph trophic modes in FUNGuild for **A)** ITS1 and **B)** 18S datasets plotted against Aggregate Vegetation Class. Aggregate Vegetation Classes are ordered from most (Crops/weeds) to least (Heath/bog) productive. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers.*

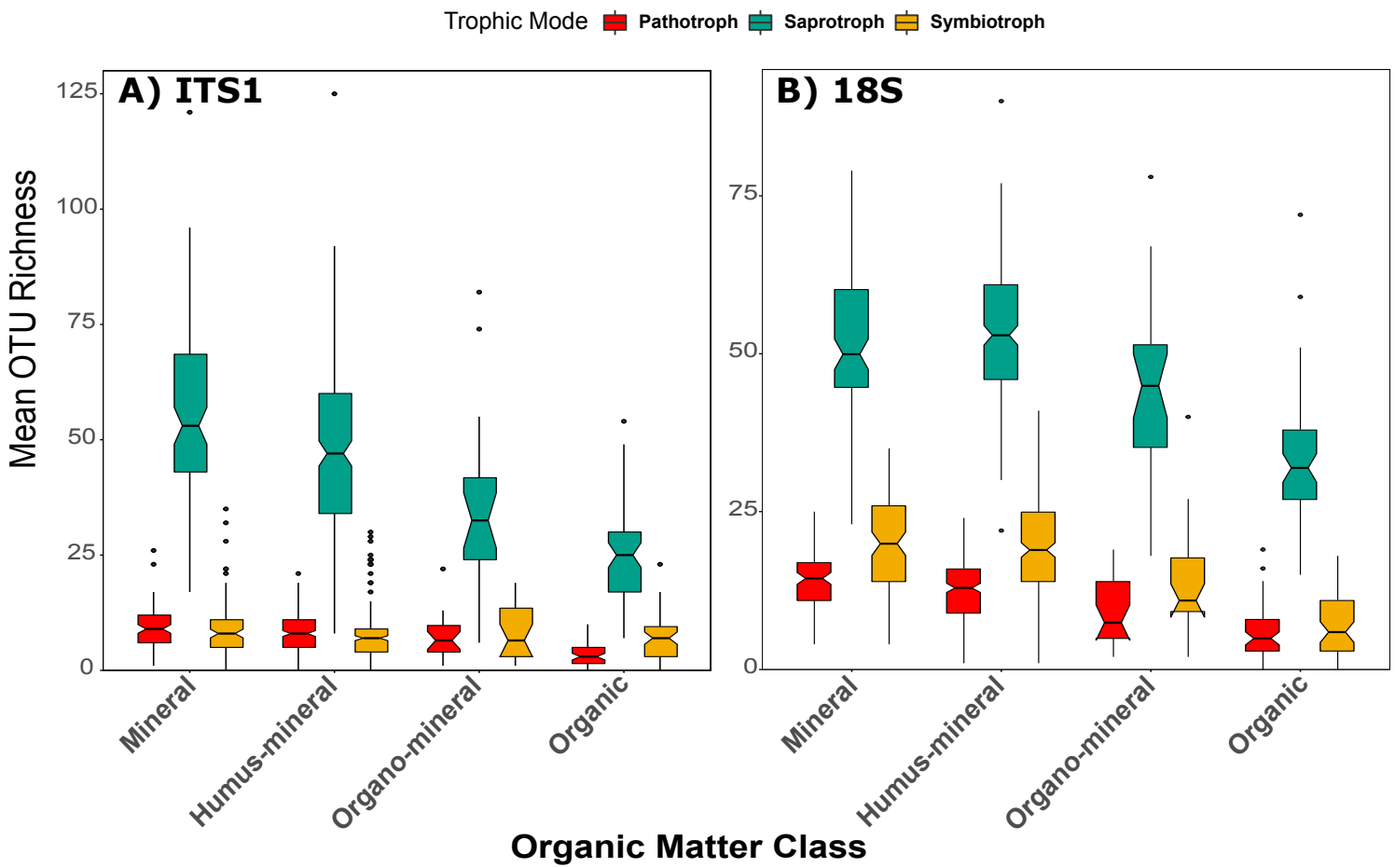


*Figure 5.9 Boxplots of richness of fungal OTUs matched to the pathotrophic, saprotroph, and symbiotroph trophic modes in FUNGuild for **A)** ITS1 and **B)** 18S datasets plotted against organic matter class. Organic matter classes are listed in order of increasing percent organic matter. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers.*

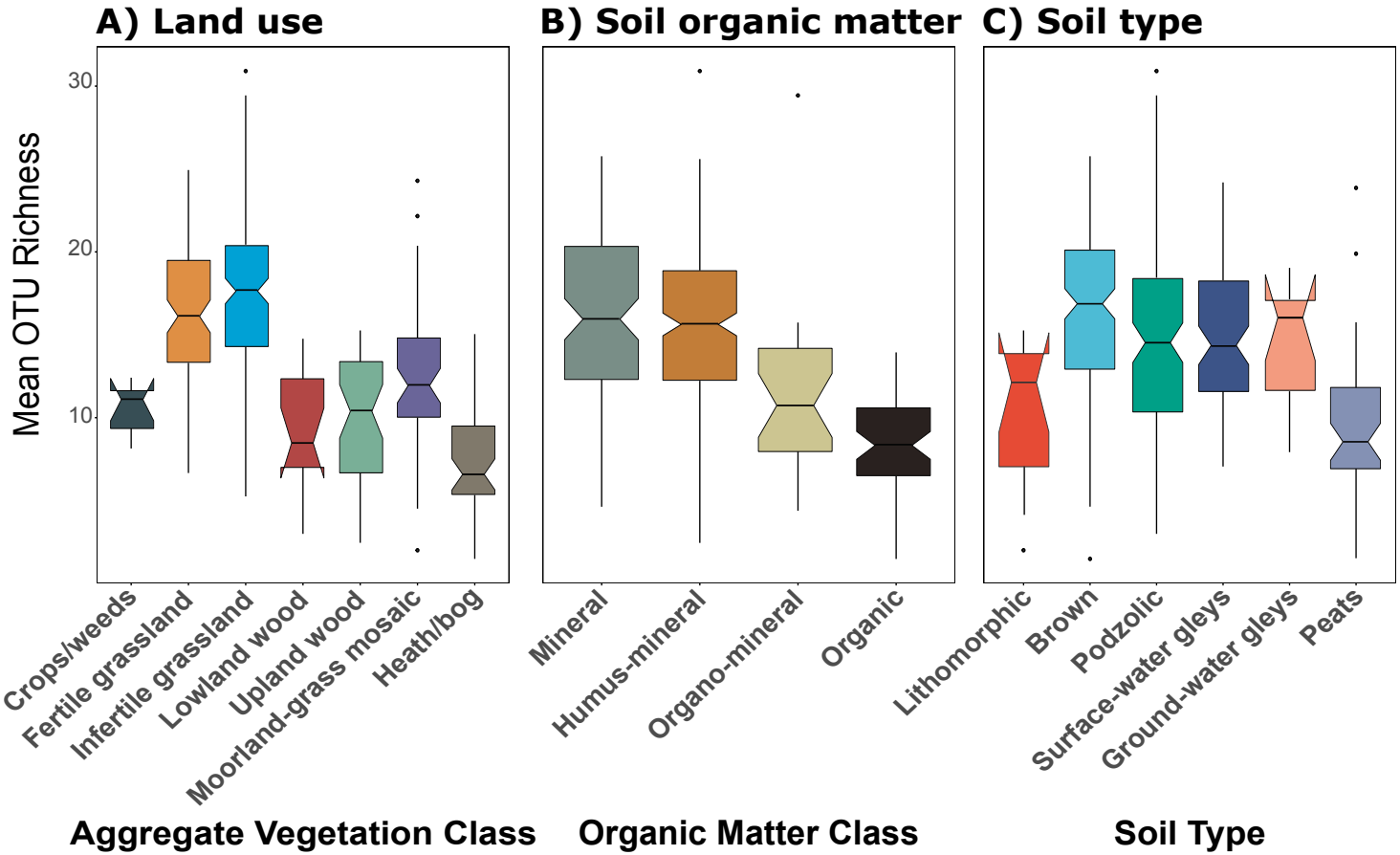


*Figure 5.10 Boxplots of richness of Glomeromycetes OTUs plotted against **A)** Aggregate Vegetation Class; **B)** organic matter class; **C)** soil type. Aggregate Vegetation Classes are ordered from most (Crops/weeds) to least (Heath/bog) productive. Organic matter classes are listed in order of increasing percent organic matter. Soils are listed in increasing order of moisture retention. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range. Notches indicate confidence interval around the median. Overlapping notches are a proxy for non-significant differences between medians. Black dots are outliers.*

5.4 Discussion

5.4.1 Primer Choice and the Total Fungal Community

We observed congruent patterns in total fungal OTU richness across land uses, organic matter classes and soil type when measured with either ITS1 or 18S primer sets. Richness was greater in arable and grassland land uses, which are highly productive, intensively managed and declined in the less productive, largely unmanaged bogs. Although these findings had been previously known from the ITS1 dataset (George et al., 2019), it is important to note that the trend was also present in the fungal OTUs identified from 18S sequencing. A similar trend was observed across organic matter classes. Here, fungal richness fell as organic matter increased. Fungal α-diversity is known to be greater in arable soils than in grasslands or forests (Szoboszlay et al., 2017). Potential mechanisms for this include: (i) increased nutrient availability due to fertiliser input (Szoboszlay et al., 2017), and (ii) beneficial disturbance from tillage and other standard agricultural practices. The latter is consistent with the intermediate disturbance hypothesis whereby high levels of diversity are maintained by consistent interruption of successional processes (Connell, 1978).

Soils rich in organic matter, especially peats, found in upland moors, bogs, and other wetlands across harbour distinct fungal communities from neighbouring habitats (Anderson et al., 2003). Fungi dominate microbial communities in bogs (Thormann and Rice, 2007) although their proportional abundance drops sharply below the first 5 cm of bog habitats (Potter et al., 2017). Yet, richness in bogs is consistently low, perhaps due to environmental pressures such as high acidity, highly recalcitrant SOM, low nutrients and oxygen levels (Rousk et al., 2010; Tedersoo et al., 2014) or reduced competition within the fungal community.

In comparison to AVC and SOM levels, differences in fungal communities were not as clear across soil types as defined by the National Soil Map (Avery, 1980), which is inline with previous work on microbial activity across the UK (Jones et al., 2014). Richness was highest in brown soils and was lowest in peats. Brown soils commonly support grassland communities across Wales (Avery, 1980; Rudeforth et al., 1984). Nearly half of the Fertile and Infertile grasslands surveyed in GMEP were categorised as brown soils. The absence of other major trends besides these may be due to the use of the dominant soil type and lack of resolution for the soil classification. The soils map used in this study simply does not provide enough resolution (1:63, 360; Avery, 1980) for soil type to be an effective category. Furthermore, this system heavily uses subsoil properties to determine soil type (Avery, 1980), while our work only involved the upper 15 cm. However, it is our opinion that the use of organic matter classification is more effective and simple metric that can be easily implemented in large-scale studies in lieu of fine-scale maps.

Results of PLS analyses demonstrates that soil properties and associated environmental factors influencing fungal richness are consistent across ITS1 and 18S datasets. Major drivers included pH, bulk density, C:N ratio, organic matter, elevation, and mean annual temperature (Table 5.2). Such results from 18S data are consistent with previous findings from the ITS1 data (George et al., 2019). However, there were certain properties that were significant in only one of the datasets and the relative importance of these properties does vary between the two datasets. There are several possible explanations for this. Firstly, 9 more samples were used in the 18S dataset (n = 422) than the ITS1 data (n = 413), which may have introduced the discrepancy in relative importance of the data. However, it is much more likely that a differential coverage of fungal groups between the two datasets caused these discrepancies.

Community composition showed consistent clustering across land uses, organic matter classes, and soil types in both data sets. As in George et al. (2019), communities were most similar in the grassland and arable sites and more spread out across woodlands and upland habitats. This was likely driven by environmental factors across Wales. In both datasets, pH was the most important environmental variable influencing community composition and although the remaining properties followed similar patterns, their relative importance again differed in the dataset. The importance of pH, elevation, C:N ratio, and precipitation in determining fungal community composition fits well in the wider context of soil fungi biogeography. Tedersoo et al. (2014) previously highlighted the importance of these variables in the distribution of fungi at the global scale. Furthermore, the strong positive correlation with C:N ratio is indicative of the expected fungal dominance (de Vries et al., 2006) of nutrient-poor, acidic soils (Bloem et al., 1997).

5.4.2 Primer Choice and Fungal Functional Diversity

Differences between richness of trophic modes of fungi, used here as a proxy for functional diversity, showed some discrepancies across land uses and soil classification between data sets. Saprotrophs made up the largest proportion of the 3 functional groups studied and generally exhibited the same trends as total richness across soils and land uses. This was also the case for pathotrophs. Indeed, correlations between environmental variables with pathotroph and saprotroph richness were largely consistent across datasets. However, we observed divergent trends in symbiotroph richness across land uses and soils. Symbiotroph richness was highest in woodlands in the ITS1 dataset whereas it was highest in grasslands according to the 18S data (Fig. 5.7A and 5.7B). A similar increase in richness within grasslands in the 18S data is repeated when Glomeromycetes were considered on their own

(Fig. 5.9); AMF are the predominant mycorrhizal fungi in grassland systems (Smith and Read, 2008). The symbiotroph peak in the ITS1 data may be explained by an increase in coverage of ectomycorrhizas which are the most common group to associate with trees and shrubs (Smith and Read, 2008). Despite these differences, both datasets suggest that symbiotroph richness was low in arable land, which is in line with previous findings demonstrating high susceptibility of mycorrhizal fungi to disturbance, for example tillage (Schnoor, et al., 2011; Säle et al., 2015), and the addition of fertilizers, which decreases the receptiveness of many agricultural plants to mycorrhizal infection (Smith and Read, 2008).

The divergent trend in symbiotroph richness and discrepancies in relationships between functional groups and environmental variables likely stem from primer biases. Primer biases have been well recognised as a confounding factor in categorising communities from environmental DNA (Cai et al., 2013; Elbrecht and Leese, 2015; Tedersoo et al., 2015). Tedersoo et al. (2015) assessed the effectiveness of fungal barcodes from the ITS, 18S, and 28S rDNA regions and found that primer choice did not affect richness or β-diversity results of soil fungi communities from Papua New Guinea, although fewer OTUs were recovered by 18S primers than ITS primers. *In silico* analyses suggests such findings are the result of lumping of sequences in the 18S that may predominantly affect rare sequences, thereby strengthening community matrices. Similarly, results were similar enough for all primers to be suitable for analyses at the class-level (Tedersoo et al., 2015). Although the 18S primers used here were designed to cover the breadth of eukaryotes and may lack specificity to fungi (Behnke et al., 2011), our results show strong congruence to the ITS1 data across total richness and indeed most functional groups.

Unlike Tedersoo et al. (2015) we observed considerable differences in the proportions of fungal classes between the ITS1 and 18S data sets. We suspect that such differences stem from the need to use appropriate databases to assign taxonomy to OTUs to each dataset (Xue et al., 2019). Perhaps only 30%-35% of Glomeromycetes are present in 18S and ITS databases, respectively (Hart et al., 2015), and although sequences are continuously being uploaded to such repositories, it is likely the majority of AMF are not identifiable from environmental samples (but see Öpik et al., 2014). Similarly we suspect that, although not studied in detail, primer choice may lead to biases in other groups. Archaeorhizomycetes accounted for nearly 25% of the 18S sequences but less than 1% from the ITS1 data (Fig. 5.2B). Primer bias has been recognised for Archaeorhizomycetes even before the class' formal description; approximately 19% of 18S sequences collected from Anderson et al. (2003), have been matched to Archaeorhizomycetes, whereas none were recovered from the same samples using ITS primers. Despite its recent description, Archaeorhizomycetes are ubiquitous components of soil communities. Strong associations have been observed with trees, yet precise functional roles of these fungi have yet to be determined (Rosling et al., 2011). Subsequently, such biases likely account for divergent relationships between functional group richness and environmental properties.

5.4.3 Conclusions

Our comparison of the use of ITS1 and 18S primers and their respective databases in a nationwide metabarcoding survey of fungi yielded 3 major findings. First, the congruent findings of total richness and β-diversity across land use and their relationships to environmental variables confirmed our previous research (George et al., 2019). Second, soil organic matter was found to be a more sensitive metric than soil type in our survey design.

Third, biases from the combination of primer and database choice became apparent for certain classes of fungi, including Glomeromycetes and Archaeorhizomycetes, which strongly influenced functional group richness across land uses as well as their relationships with environmental variables. It is therefore important to recognise the sensitivity of metabarcoding to primer choice, even when using universal primers. Without simultaneous analyses of environmental DNA using both primers and databases, the presence of AM fungi as well as the newly characterised Archaeorhizomycetes would have been overlooked and unquantified in this survey. Furthermore, since the majority of soil biodiversity is undescribed (Ramirez et al., 2015), utilising multiple primers will elucidate a more complete picture of belowground biodiversity by revealing shortcomings in existing probes and revealing the presence of as yet undescribed organisms. We therefore advocate that future nation-wide surveys included both a sample-based metric of soil type (i.e. organic matter classification) and multiple primers for fungal biodiversity. Such measures should not be arduous to implement, especially if researchers can identify specific fungal groups of particular interest to accommodate.

5.5 Acknowledgements

This project was supported as part of the Welsh Government funded Glastir Monitoring & Evaluation Programme (Contract reference: C147/2010/11) and by the UK Natural Environment Research Council (NERC) through the RC/Centre for Ecology & Hydrology (CEH Project: NEC04782). PBLG was supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council and NERC [Grant number NE/M009106/1]. STARS is a consortium consisting of Bangor University, British Geological Survey, Centre for Ecology and Hydrology, Cranfield

University, James Hutton Institute, Lancaster University, Rothamsted Research and the University of Nottingham.

We thank the GMEP Team for their contribution in collecting the data and the laboratory staff of Environment Centre Wales, especially Delphine Lallias for sample processing and DNA extraction. Thanks also to Fiona M. Seaton for assistance with statistical analyses. We thank Supercomputing Wales for support and training on their system for bioinformatics analyses. We thank John G. Kenny and Richard M. Eccles for generating sequence data at the Centre for Genomic Research, University of Liverpool.

5.6 References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* 215, 403-410. doi: 10.1016/S0022-2836(05)80360-2

Anderson, I. C., Campbell, C. D., and Prosser, J. I. (2003). Diversity of fungi in organic soils under a moorland – Scots pine (*Pinus sylvestris* L). *Environ. Microbiol.* 5, 1121-1132. doi: 10.1046/j.1462-2920.2003.005522.x

Avery, .W. (1980). Soil classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden, UK.

Behnke, A., Engel, M., Christen, R., Nebel, M., Klein, R. R., and Stoeck, T. (2011). Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariable SSU rRNA gene regions. *Environ. Microbiol.* 13, 340-349. doi: 10.1111/j.1462-2920.2010.02332.x

Berruti, A., Desirò, A., Visentin, S., Zecca, O., Bonfante, P. (2017). ITS fungal barcoding primers versus 18S AMF-specific primers reveal similar AMF-based diversity patterns in roots and soils of three mountain vineyards. *Environ. Microbiol. Rep.* 9, 658-667.

Bloem, J., de Ruiter, P. C., and Bouwman, L. A. (1997). "Soil food webs and nutrient cycling in agroecosystems," in *Modern soil microbiology*, eds. van Elsas, J. D., Jansson, J. K., and Trevors, J. T. (Boca Raton, FL: CRC Press), 245–278.

Bunce, R. G. H., Carr, C. J., Gillespie, M. K., Howard, D. C., Scott, W. A., Smart, S. M., Van de Poll, H. M., and Watkins, J. W. (1999). Vegetation of the British countryside – the countryside vegetation system. (Department of the Environment, Transport and the Regions, London).

Cai, L., Ye, L., Tong, A. H. Y., Lok, S., and Zhang, T. (2013). Biased diversity metrics revealed by bacterial 16S phylotags derived from different primer sets. *PLoS ONE* 8, e53649. doi: 10.1371/journal.pone.0053649

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Wildmann, J., Yaysunenko, T., Zaneveld, J., and Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* 7, 335-336. doi: 10.1038/nmeth.f.303

Cavender-Bares, J., Kozak, K, H,, Fine, P. V. A., Kembel, S. W. (2009) The merging of community ecology and phylogenetic biology. *Ecol. Lett.* 12, 693-715. doi: 10.1111/j.1461-02548.2009.01314.x

Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science* **199**, 1302-1310. doi: 10.1126/science.199.4335.1302

Chong, I.-G., and Jun, C.-H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometr. Intell. Lab. Syst.* 78, 103-112. doi: 10.1016/j.chemolab.2004.12.011

Deshpande, V., Wang, Q., Greenfield, P., Charleston, M., Porras-Alfaro, A., Kuske, C. R., et al. (2016). Fungal identification using a Bayesian classifier and the Warcup training set of internal transcribed spacer sequences. *Mycologia* 108, 1-5. doi: 10.3852/14-293.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **10**, 2460-2461. doi: 10.1093/bioinformatics/btq461

Elbrecht, V., and Leese, F. (2015). Can DNA-based ecosystem assessments quantify species abundance? Testing primer bias and biomass-sequence relationships with an innovative metabarcoding protocol. *PLoS ONE* 10, e0130324. doi: 10.1371/journal.pone.0130324

Emmett, B. A. & the GMEP Team. (2017). Glastir Monitoring & Evaluation Programme. Final Report to Welsh Government (Contract reference: C147/2010/11. NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780/NEC05371/NEC05782).

Emmett, B. A., Reynolds, B., Chamberlain, P. M., Rowe, E., Spurgeon, D., Brittain, S. A., et al. (2010). Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology (CS Technical Report No. 9/07, CEH Project Number: C03259).

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., et al. (2012). New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosystems. *Mol. Ecol.* 8, 1821-1833. doi: 10.1111/j.1365-294X.2012.05537.x

Fisher, M. C., Henk, D. A., Briggs, C. J., Brownstein, J. S., Madoff, L. C., McCraw, S. L., et al. (2012). Emerging fungal threats to animal, plant and ecosystem health. *Nature* 484, 186-194. doi: 10.1038/nature10947

George, P. B. L., Lallias, D., Creer, S., Seaton, F. M., Kenny, J. G., Eccles, R. M., et al. (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nat. Commun.* 10, 1107. doi:10.1038/s41467-019-09031-1

Hart, M. M., Aleklett, K., Chagnon, P.-L., Egan, C., Ghignone, S., Helgasson, T., et al. (2015). Navigating the labyrinth: a guide to sequence-based, community ecology of arbuscular mycorrhizal fungi. *New Phytol.* 207, 235-247. doi: 10.1111/nph.13340

Hothorn, T., Bretz, F., and Westfall, P. (2008). Simultaneous inference in general Parametric Models. *Biometrical J*. 50, 346-363. doi: 10.1002/bimj.200810425

Jones, D. L., Simfukwe, P., Hill, P. W., Mills, R. T. E., Emmett, B. A. (2014). Evaluation of dissolved organic carbon as a soil quality indicator in national monitoring schemes. *PLoS One* 9(3):e90882. doi: 10.1371/journal.pone.0090882

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., et al. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Mol. Ecol.* 22, 5271-5277. doi: 10.1111/mec.12481

Lallias, D., Hiddink, J., G., Fonseca, V. G., Gaspar, J. M., Sung, W., Neill, S. P., et al. (2015). Environmental metabarcoding reveals heterogeneous drivers of microbial eukaryote diversity in contrasting estuarine ecosystems. *ISME J.* **9**, 1208-1221. doi: 10.1038/ismej.2014.213

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011. Date of access 04/07/2017. *EMBnet* **17**, 10-12. doi: 10.14806/ej.17.1.200

McMurdie, P. J., and Holmes, S. (2013). phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217. doi: 10.1371/journal.pone.0061217

Mevik, B-H., Wehrens, R., Liland, K. H., and Hiemstra, P. (2016). pls: Partial Least Squares and Principal Component Regression. R package version 2.6-0. https://CRAN.R-project.org/package=pls

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., et al. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecol.* 20, 241-248. doi: 10.1016/j.funeco.2015.06.006

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., et al. (2016). Vegan: community ecology package. R package version 2.4-0. https://CRAN.R-project.org/package1⁄4vegan.

Öpik, M., Davison, J., Moora, M., and Zobel, M. (2014). DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences[1]. *Botany* 92, 135-147. doi: 10.1139/cjb-2013-0110

Peay, K. G., Kennedy, G., and Bruns, T. D. (2008). Fungal community ecology: a hybrid beast with a molecular master. *BioScience* 58, 799-810. doi: 10.1641/B580907

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., EISPAK authors, Heisterkamp, S., et al. (2016). nlme: Lineral and Nonlinear Mixed Effects Models. R package version 3.1e138. https://CRAN.R-project.org/package1/4nlme

Potter, C., Freeman, C., Golyshin, P. N., Ackermann, G., Fenner, N., McDonald, J. E., et al. (2017). Subtle shifts in microbial communities occur alongside the release of carbon induced by drought and rewetting in contrasting peatland ecosystems. *Sci. Rep.-UK* 7, 11314. doi: 10.1038/s41598-017-11546

Quast, C., Pruesse, E., Yilmaz, P., Gerken, J., Schweer, T., Yarza, P., et al. (2013). The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucl. Acids Res.* 41 (D1), D590-D596. doi: 10.1093/nar/gks1219

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). https://www.R-project.org/.

Ramirez, K. S., Leff, J. W., Barberán, A., Bates, S. T., Betley, J., Crowther, T. W., et al. (2015). Biogeographic patterns in below-ground diversity in New York City's Central Park are similar to those observed globally. *Proc. R. Soc. B.* 281, 20141988. doi: 10.1098/rspb.2014.1988

Robinson, D. A., Astbury, S., Barrett, G., Burden, A., Carter, H., Emmett, B. A., Garbutt, A., et al. (2019). Topsoil physico-chemical properties from the Glastir Monitoring and Evaluation Programme, Wales 2013-2016. NERC Environmental Information Data Centre. doi: 10.5285/0fa51dc6-1537-4ad6-9d06-e476c137ed09

Robinson, E. L., Blyth, E., Clark, D. B., Comyn-Platt, E., Finch, J., and Rudd, A. C. (2017). Climate hydrology and ecology research support system meteorology dataset for Great Britain (1961-2015) [CHESS-met] v1.2. NERC Environmental Information Data Centre.

Rognes, T., Flouri, T., Nichols, B., Quince, C., and Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ* **4**, e2584. doi: 10.7717/peerj.2584

Rosling, A., Cox, F., Cruz-Martinez, K., Ihrmark, K., Grelet, G.-A., Lindahl, B. D., et al. (2011). Archaeorhizomycetes: unearthing an ancient class of ubiquitous soil fungi. *Science* 333, 876-879. doi: 10.1126/science.1206958

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., et al. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J.* 4, 1340-1351. doi: 10.1038/ismej.2010.58

Rudeforth, C. C., Hartnup, R., Lea, J. W., Thompson, T. R. E., and Wright, P. S. (1984). Soils and their use in Wales. Harpenden: Lawes Agricultural Trust (Soil Survey of England and Wales).

Sagova-Mareckova, M., Cermak, L., Novotna, J., Plhackova, K, Forstova, J., and Kopecky, J. (2008). Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Appl. Environ. Microb.* 74, 2902-2907. doi: 10.1128/AEM.02161-07

Säle, V., Aguilera, P., Laczko, E., Mäder, P., Berner, A., Zihlmann, U., et al. (2015). Impact of conservation tillage and organic farming on the diversity of arbuscular mycorrhizal fungi. *Soil Biol. Biochem.* 84, 38-52. doi:10.1016/j.soilbio.2015.02.005

Schnoor, T. K., Lekberg, Y., Rosendahl, S., and Olsson, P. A. (2011) Mechanical soil disturbance as a determinant of arbuscular mycorrhizal fungal communities in semi-natural grassland. *Mycorrhia* 21, 211-220. doi: 10.1007/s00572-010-0325-3

Schoch, C. L., Seifert, K. A., Huhndorf, S., Robert, V., Spouge, J. L., Levesque, C. A., et al. (2012). Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. *P. Natl. Acad. Sci. USA.* 109, 6241-6246. doi: 10.1073/pnas.1117018109

Schüβler, A., Schwarzott, D., and Walker, C. (2001). A new fungal phylum, the *Glomeromycota*: phylogeny and evolution. *Mycol. Res.* 105, 1413-1421. doi: 10.1017/S0953756201005196

Smith, S. E., and Read, D. J. (2008). Mycorrhizal Symbiosis 3rd ed. London: Academic Press.

Stockinger, H., Krüger, M., and Schüβler, A. (2010). DNA barcoding of arbuscular mycorrhizal fungi. *New Phytol.* 187, 461-474. doi: 10.111/j.1469-8137.2010.03262.x

Szoboszlay, M., Dohrmann, A. B., Poeplau, C., Don, A., and Tebbe, C. C. (2017). Impact of land-use change and soil organic carbon quality on microbial diversity in soils across Europe. *FEMS Microbiol. Ecol.* 93, fix146. doi: 10.1093/femsec/fix146

Tedersoo, L., Anslan, S., Bahram, M., Pólme, S., Riit, T., Liiv, I., Kõljalg, U., et al. (2015). Shotgun metagenomes and multiple primer pair-barcode combinations of amplicons reveal biases in metabarcoing analyses of fungi. *MycoKeys* 10, 1-43. doi: 10.3897/mycokeys.104852

Tedersoo, L., Bahram, M., Põlme, S., Kõljalg, U., Yorou, N. S., Wijesundera, R., et al. (2014). Global diversity and geography of soil fungi. *Science* 346:1256688. doi: 10.1126/science.1256688

Thiéry, O., Vasar, M., Jairus, T., Davison, J., Roux, C., Kivistik, P.-A., Metspalu, A., Milani, L., Saks, Ü., Moora, M., Zobel, M., and Öpik, M. Sequence variation in nuclear ribosomal small subunit, internal transcribed spacer and large subunit regions of *Rhizophagus irregularis* and *Gigaspora margarita* is high and isolate-dependent. *Mol. Ecol.* 25, 2816-2832. doi: 10.1111/mec.13655

Thormann, M. N., and Rice, A. V. (2007). Fungi from peatlands. *Fungal Divers.* 24, 241-299.

de Vries, F. T., Hoffland, E., van Eekeren, N., Brussard, L., and Bloem, J. (2006). Fungal/bacterial ratios in grasslands with contrasting nitrogen management. *Soil Biol. Biochem.* 38, 2092-2103. doi: 10.1016/j.soilbio.2006.01.008

Wang, B., and Qui, Y.-L. (2006). Phylogeneitc distribution and evolution of mycorrhizas in land plants. *Mycorrhiza* 16, 299-363. doi: 10.1007/s00572-005-0033-6

Wang, Q., Garrity, G. M., Tiedje, J. M., and Cole, J. R. (2007). Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl. Environ. Microb.* 73, 5261-5267. doi: 10.1128/AEM.00062-07

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., et al. (2017). Normailzation and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 1-18. doi: 10.1186/s40168-017-0237-y

Xue, C., Hao, Y., Pu, X., Penton, C. R., Wang, Q., Zhao, M., et al. (2019). Effect of LSU and ITS genetic markers and reference databases on analyses of fungal communities. *Biol. Fert. Soils* 55, 79-88. doi: 10.1007/s00374-018-1331-4

Chapter 6: Inverse relationship between generalist anaerobes and sulphate-reducing bacteria is driven by pH across land uses in temperate soils

This manuscript has been submitted to the *European Journal of Soil Science*

George, P.B.L.[1,2,*], Coelho, K.P.[3], Creer, S.[1], Lebron, I.[2], Emmett, B.A.[2], Robinson, D.A.[2], Jones, D.L.[1,4], 2019. *In Review*.

[1] School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] Postgraduate program in Agroecology, Maranhão State University, Maranhão, CEP 65055-310, Brazil

[4] SoilsWest, School of Agriculture and Environment, The University of Western Australia, Perth WA 6009, Australia

[*] Corresponding author

P.B.L.G., K.P.C., and D.L.J. and conceived this project using the GMEP framework led by B.A.E. P.B.L.G. and K.P.C. prepared the dataset. P.B.L.G and I.L. prepared and analysed sulphate extracts from soil samples. Statistical analyses were led by P.B.L.G. P.B.L.G. wrote the first draft of the manuscript and K.P.C, D.A.R., S.C., and D.L.J. contributed to subsequent revisions. All authors read and approved the final manuscript.

Abstract

Sulphate-reducing bacteria (SRB) represent a key biological component of the global sulphur cycle and are common in soils where they reduce SO_x to H_2S during the anaerobic degradation of soil organic matter. The factors that regulate their distribution in soil, however, remain poorly understood. We sought to determine the ecological patterns of SRB richness within a nation-wide 16S metabarcoding dataset. Across 436 sites belonging to 7 contrasting temperate ecosystems (e.g. arable, grasslands, woodlands, heathland and bog), SRB richness was relatively constant across land uses but greatest in grasslands and lowest in woodlands and peat-rich soils. The dominant SRB were *Desulfobacca*, *Desulfosporosinus* and *Desulfobulbus*. In contrast, richness of other anaerobic generalist bacterial taxa found in our dataset (e.g. *Clostridium*, *Geobacter*, and *Pelobacter*) followed an established trend of declining richness linked to land use productivity. Overall, the richness of SRBs and generalist anaerobes had strong positive correlations with pH and sulphate concentration and strong negative relationships with elevation, carbon, and carbon-to-nitrogen ratio. It is likely that these results reflect the driving influence of pH and competition for optimal electron acceptors with generalist anaerobic bacteria on SRB richness.

***Key words**:* Anaerobes; Atmospheric deposition; Nutrient cycling; Soil acidity; Dissimilatory sulphate reduction.

6.1 Main text

Sulphate-reducing bacteria (SRB) are common soil organisms, which are capable of transforming sulphate (SO_x) into hydrogen sulphide (H_2S) under anoxic conditions (Hines et al., 1999; Bahr et al., 2005; Xia et al., 2014). Consequently, these organisms play a fundamental role in global sulphur (S) cycling and also in the iron (Fe) cycle through the formation of FeS_2 (Muyzer and Stams, 2008). After waterlogging, soils are often rich in H_2S due to high local abundances of SRB leading to changes in plant metabolism (Stubner, 2004; Lamers et al., 2013; Li et al., 2016). Despite their namesake, SO_x is a poor terminal electron acceptor and SRB are able to use a number of different terminal electron acceptors including other S compounds, nitrate, organic compounds, and even oxygen, whist some are known to be facultative anaerobes (Muyzer and Stams, 2008).

Currently, more than 220 species of SRB have been described with soils often possessing diverse SRB communities (Barton and Fauque, 2009). For example, in landfill cover soils the number of SRB operational taxonomic units (OTUs) has been shown to range from 30 (Scheid and Stubner, 2001) to 70 (Xia et al., 2014). These bacteria may also form relationships with other S-dependent bacteria, such as green phototrophic S bacteria (Overmann and van Gemerden, 2000). Therefore, a strict assumption that SRB communities are directly linked to S or SO_x availability and that SRB rely on strict anaerobic conditions is overly simplistic (Muyzer and Stams, 2008). Competition between SRB and other anaerobic bacteria for C substrates (e.g. acetate) has also been shown to strongly influence SRB distributions in soil (Muyzer and Stams, 2008).

It is unclear which edaphic factors may be most important in determining SRB population distributions. Major SO_x inputs to agricultural land include inorganic fertiliser addition (e.g. ammonium and potassium sulphates), soil amendments (e.g. calcium sulphate), and livestock waste (Carvalho and van Raij, 1997; Allison et al., 2001; Abdelmseeh et al., 2008; Pan et al., 2016). Atmospheric S deposition from anthropogenic and marine sources is a major source of SO_x, especially at higher elevations in wetter climates (Stevens et al., 1997). Subsequently, one might expect richness of SRB to increase with elevation owing to an increase in anaerobic niches in upland habitats and SO_x availability. Indeed, Drenovsky et al. (2010) demonstrated with phospholipid fatty acid analyses that the proportion of *Desulfobacter* biomass increased with soil moisture in Californian soils. It is also possible that SRBs may be used as an environmental indicator of ecosystem recovery from acid deposition (Review of Transboundary Air Pollution, 2012), which is now declining in many industrialised countries (Kirk et al., 2010; Reynolds et al., 2013). This however, requires an understanding of the key factors that regulate SRB communities across a wide range of land uses.

Here, we use a national-level metabarcoding data set to determine the distribution of SRB richness in soil. We hypothesised that habitat would be a major driver of SRB richness and therefore we expected richness to increase in acidic soils and upland anoxic habitats. Since other anaerobic microbes can directly compete with SRBs (Muyzer and Stams, 2008) we also investigated the relationships of some common anaerobic bacterial taxa with land use. We further hypothesised that richness of both SRB and other anaerobic bacteria would be positively correlated to increasing pH and elevation.

This work was undertaken by analysing the metabarcoding dataset of soil biodiversity across Wales, UK, collected as part of the Glastir Monitoring and Evaluation Programme

(GMEP) presented in George et al. (2019). Soil samples were collected across Wales (n = 436) between late spring and early autumn in 2013 and 2014 (Appendix 4 Fig. 1). Sampling protocols followed the UK Countryside Survey (Emmett et al., 2010), whereby samples were collected from randomly selected 1 km^2 squares (see Appendix 1.1). . Within each 1 km^2 square, up to 3 samples were collected; for further details see Emmett et al. (2010). Soil physicochemical properties examined were pH (measured in 0.01 M $CaCl_2$), organic matter (% loss-on-ignition), total C and nitrogen (N) (%), C: N ratio, phosphorous (P) (mg kg^{-1}), bulk density (g cm^{-3}), and moisture content (g water g^{-1}). Geographic coordinates and elevation (m) were also collected. Mean annual precipitation (mL) at each site was extracted from the CHESS dataset (Robinson et al., 2014). Sulphate concentrations (mg kg^{-1}) were determined using 1:5 (w/v) distilled water extracts (Tabatabai, 1996) followed by analysis by ion chromatography (Metrohm Ltd).

At each sample site, habitat was classified using plant species assemblages into one of seven Aggregated Vegetation Classes as described by Bunce et al. (1999), see Appedix 4.1 and Appendix 2 Table 1 for further detail. Maps of S deposition from non-marine (2013-2015) and marine (2014-2016) sources were made by the Centre for Ecology and Hydrology (Appendix 4 Fig. 2). Summarised environmental and soil property data across AVCs from George et al. (2019) are presented in Appendix 4 Table 1.

Collocated soil cores were collected for metabarcoding analyses. DNA extraction, metabarcoding, and bioinformatics analyses are described in George et al. (2019). Briefly, DNA was extracted in triplicate from 0.25 g of homogenised soil, which was passed through sterilised sieves, via mechanical lysis using PowerLyser PowerSoil DNA Isolation Kits (MO-BIO). A pre-treatment of 750 µL of 1 M $CaCO_3$ (Sagova-Mareckova et al., 2008) was used as

this has been shown to improve PCR performances of DNA extracted from acidic soils. Extracted DNA was sequenced using a two-step Illumina Mi-Seq amplicon sequencing protocol. Amplicon libraries were created in triplicate on a DNA Engine Tetrad® 2 Peltier Thermal Cycler (BIO-RAD Laboratories) using the V4 region of the 16S rDNA gene with the 515F/806R universal primers (Caporaso et al., 2011) at Bangor University ad the Liverpool Centre for Genomic Research in 2013 and 2014, respectively. First-round PCR amplification began at 98 °C for 30 s, followed by 10 cycles of 98 °C for 10 s; 50 °C for 30 s; 72 °C for 30s; with a final extension stage of 72 °C for 10 min and held at 4 °C for a further 10 min. For the second-round PCR, 12 µL of first-round product was mixed with 0.1 µL exonucleaseI, 0.2 µL thermosensitive alkaline phosphatase, and 0.7 µL of water and cleaned in the thermocycler with a programme of 37 °C for 15 min and then 74 °C for 15 min followed by a hold at 4 °C. Next, Illumina Netera XT 384-way indexing primers were added and amplified with an initial denaturation at 98 °C for 3 min; followed by 15 cycles of 95 °C for 30 s; 55 °C for 30 s; 72 °C for 30s; and a final extension at 72 °C for 5 min and then held at 4 °C. These products were subsequently purified using an equal volume of AMPure XP beads (Beckman Coulter).

Raw sequences were de-multiplexed, filtered, quality-checked, and clustered using the USEARCH v. 7.0 (Edgar, 2010) and VSEARCH v. 2.3.2 (Rognes et al., 2016) software. Operational taxonomic units (OTUs) were made using open-reference clustering at 97% similarity (George, et al., 2019). Sequences with a maximum error > 1 and shorter than 200 bp were removed from analysis. The subsequent OTU table was generated using QIIME 1.9.1 (Caporaso et al., 2010) and analysed using the phyloseq package (McMurdie and Holmes, 2013) in R v. 3.5.1 (R Core Team, 2018), removing all OTUs identified as chimeras or non-bacterial taxa using the GreenGenes 13.8 database (DeSantis et al., 2006) as well as singletons.

Read counts were normalised through rarefaction. The OTU table was rarefied 100 times at 40,000 read depth and mean richness recorded. Next, we compared SRB taxa from the literature to our dataset and found OTUs identified as 11 SRB genera (*Desulfobacca*, *Desulfobotulus*, *Desulfobulbus*, *Desulfocapsa*, *Desulfococcus*, *Desulfomonile*, *Desulforhabdus*, *Desulfosarcina*, *Desulfosporosinus*, *Desulfotomaculum*, and *Desulfovibrio*) and three generalist anaerobic bacteria (*Clostridium*, *Geobacter*, and *Pelobacter*) were selected for further analysis. Sequences can be accessed on the European Nucleotide Archive (primary accession code: PRJEB27883).

Linear mixed models were created using the package nlme (Pinheiro et al., 2019) and tested using Tukey's HSD test from the multcomp package (Hothorn et al., 2008) to assess differences between richness of SRB and other mentioned anaerobic bacteria across AVCs. Identities of 1 km^2 were used as a random factor in all models. Relationships between environmental variables and the richness of both SRB and anaerobic bacteria were assessed using Spearman's ranked correlation. Soil properties and environmental variables were normalised where appropriate.

In Wales, the areas of highest S-deposition are consistently at high elevation, including the Snowdonia and Brecon Beacons National Parks (Fig. 5.1). Yet sulphate concentrations were highest in Crops/weeds (arable) and woodland AVCs (Appendix Table 2). Richness of SRB and anaerobes showed a significant positive correlation with sulphate concentrations (Table 6.1). There was also a significant negative correlation between SRB richness and C-to-N ratio, organic matter, and elevation. Anaerobic bacteria richness had significant negative correlations with total C, N, C-to-N ratio, organic matter, moisture, elevation, and annual precipitation (Table 6.1).

*Table 6.1 Correlations (Spearman's rho) between richness of both sulphate-reducing bacteria (SRB) and anaerobes and soil properties and environmental variables. Note *** (P < 0.001), ** (0.001 > P < 0.01), * (0.01 > P < 0.05), and NS (P > 0.05).*

	Total C (%)[a]	**Total N (%)[a]**	**C: N ratio[b]**	**Sulphate (mg kg^{-1})[a]**	**pH ($CaCl_2$)**	**Moisture content (g g^{-1})[a]**	**Elevation (m)**	**Mean annual precipitation (mL)**
Richness of SRB	-0.08[NS]	0.03 [NS]	-0.17***	0.15**	0.20***	0.14**	-0.19***	0.03
Richness of anaerobes	-0.45***	-0.38***	-0.53***	0.11*	0.58***	-0.27***	-0.46***	-0.27***

Note: [a] denotes log_{10}-transformation; [b] denotes square-root transformation

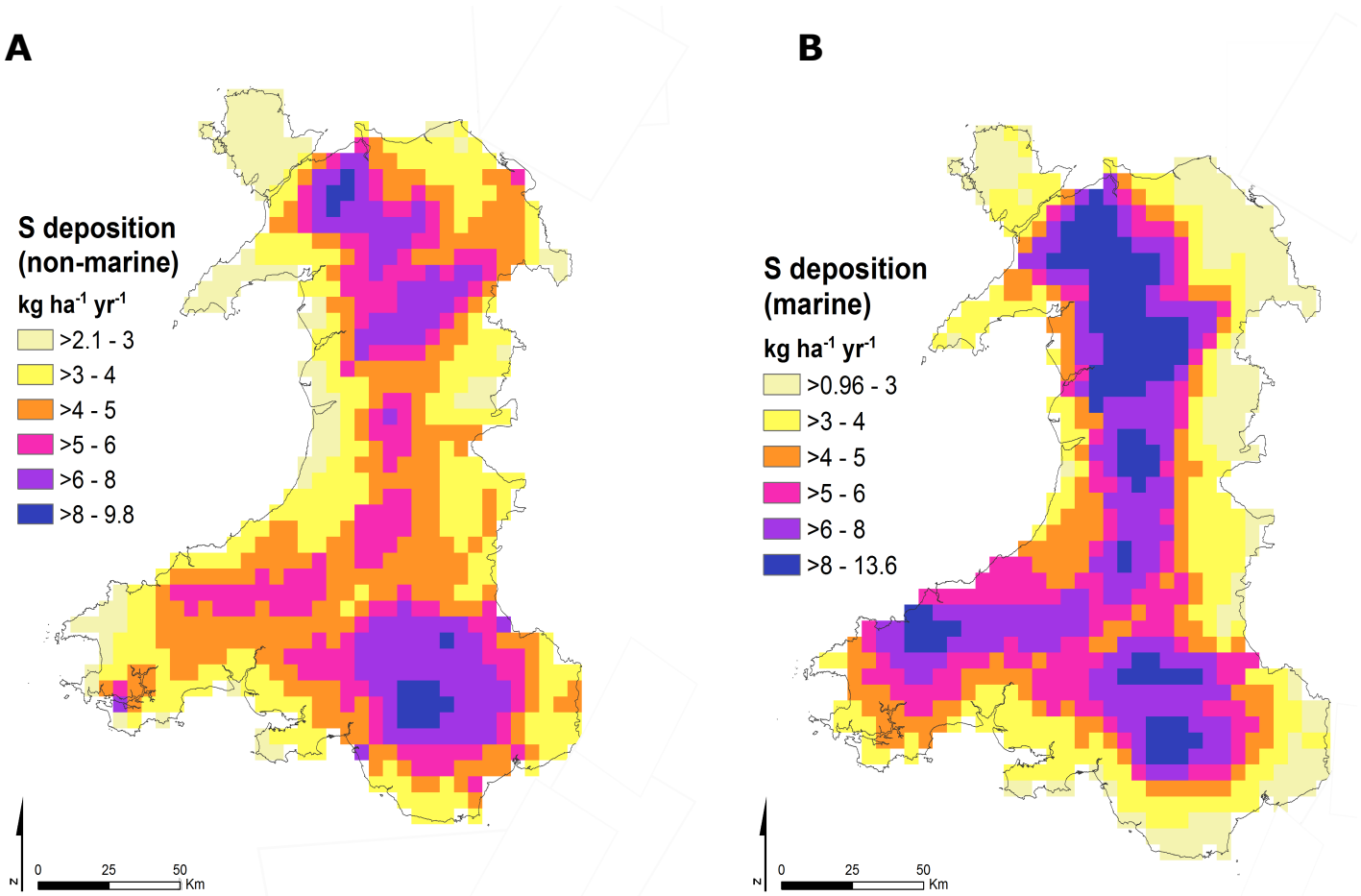


*Figure 6.1 Sulphur deposition maps of Wales from **A**) non-marine (2013-2015) and **B**) marine (2014-2016) sources.*

Richness of SRB was greater ($F_{6, 273} = 5.22$, $P < 0.001$) in Fertile grasslands, Infertile grasslands and Moorland grass-mosaic than in both Lowland and Upland woods (Fig. 6.2A). Richness was significantly lower in Upland wood than Fertile grasslands ($P < 0.001$), Infertile grasslands ($P = 0.005$), and Moorland grass-mosaic ($P = 0.002$). The same trend was observed between Lowland woods and Fertile grasslands ($P = 0.007$), Infertile grasslands ($P = 0.03$), and Moorland grass-mosaic ($P = 0.01$).

Stronger differences were observed between anaerobe richness across AVCs (Fig. 6.2B). Unexpectedly, richness of anaerobes was significantly ($F_{5, 272} = 27.31$, $P < 0.001$) greater in the high productivity AVCs including Crops/weeds and both Fertile and Infertile grasslands, than low productivity AVCs. Anaerobe richness was greater in Crops/weeds than both types of woodland, Heath/bog (all $P < 0.001$) and Moorland grass-mosaic AVCs ($P = 0.03$). These differences were also present between Fertile and Infertile grasslands (all $P < 0.001$). Anaerobe richness was also significantly greater in Moorland grass-mosaic sites than in Heath/bog ($P = 0.01$) samples (Fig. 6.2B).

Across AVCs, proportional abundance of SRB increased along the productivity gradient of Crops/weeds to Heath/bog sites (Fig. 6.2C). This was mirrored by a decline in generalist anaerobes. However the proportion of anaerobes never fell below 25% even in Heath/bog sites dominated by SRB; whereas anaerobes outnumber SRB by ~90% in high productivity AVCs (Fig. 6.2C)

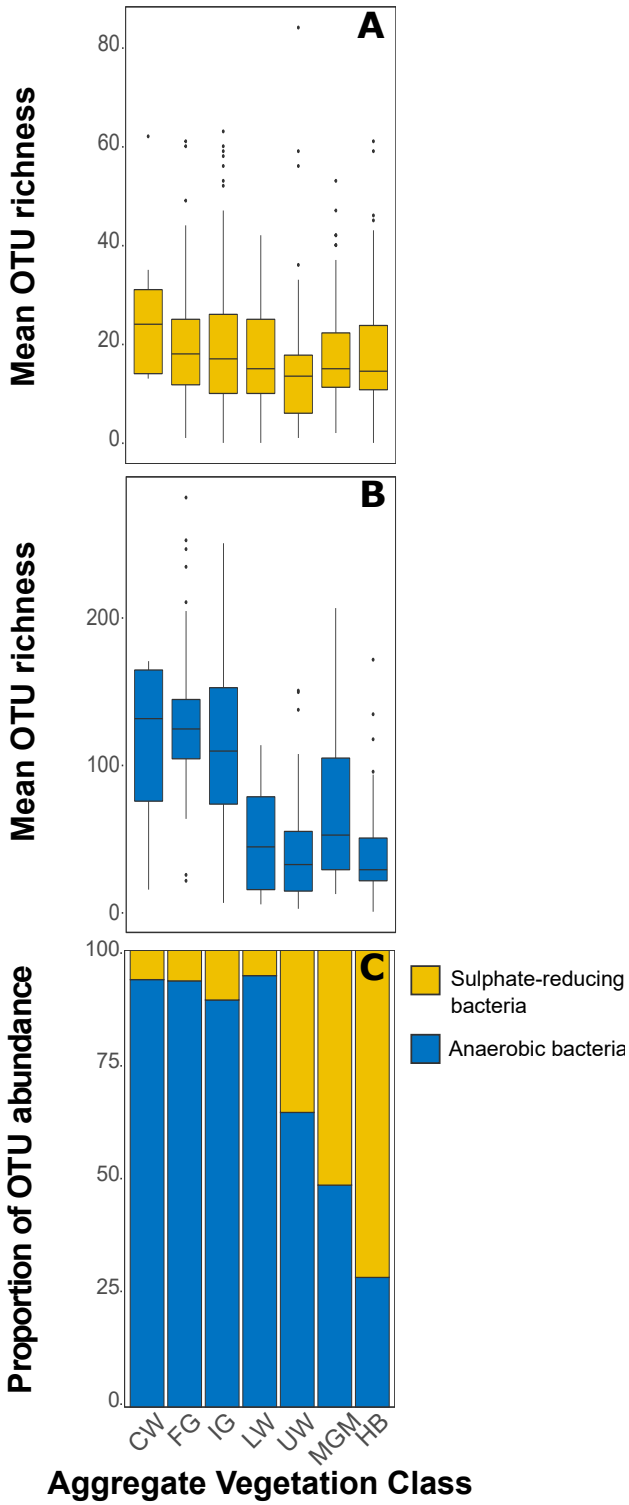


*Figure 6.2 Richness of SRB **A**) and selected generalist anaerobe OTUs **B**), and relative abundances of both groups **C**) across Aggregate Vegetation Classes (AVCs). Abbreviations for AVCs are as follows: Crops/weeds (CW); Fertile grassland (FG); Infertile grassland (IG); Lowland wood (LW); Upland wood (UW); Moorland grass-mosaic (MG); Heath/bog (HB).*

Contrary to our expectations, SO_x concentrations were highest in the Crops/weeds and woodland AVCs, rather than high elevation Moorland grass-mosaic and Heath/bog sites (Appendix 4 Table 1). This was surprising, as SO_x deposition rates are known to increase with elevation (Stevens et al., 1997; Lovett et al., 1999) in addition to observing a significant, positive correlation between SRB and SO_x concentrations (Table 6.1). One explanation for this is that arable sites were subjected to amendment with fertilisers containing SO_x (Allison et al., 2001; Pan et al., 2016); however, without detailed land management histories we cannot be confident in this explanation. There was a marked increase in relative abundance of SRB OTUs in Moorland grass-mosaic and Heath/bog (Fig. 6.1C), despite the negative correlation between SRB and organic matter (Table 6.1). This could also reflect a greater adaptability of certain SRB taxa; *Desulfobulbus*, for instance, is capable of utilising alcohols and alternative organic acids in the absence of sulphates (Biswas et al., 2014).

Unexpectedly, richness of the anaerobic bacteria highlighted in this study did not increase in stereotypically anaerobic habitats, such as Heath/bog. Rather, richness of these taxa followed the overarching trend of microbial richness declining with soil productivity across Wales found by George et al. (2019). This is likely due to the generalist nature of these microbes. For example, *Geobacter* sp. are ubiquitous components of soil bacterial communities as they are able to utilise a wide range of alternative electron acceptors (Lovley et al., 2011). Similarly, *Clostridium* are common constituents of soil communities (Jeong et al., 2004) while *Pelobacter* is common in anoxic waterlogged soils (Masuda, et al., 2018). Furthermore, SRB and anaerobes demonstrated strong congruent relationships with C:N ratio, though this is contrary to previous findings from anoxic systems (Yuan et al., 2019). Previous analysis has confirmed the driving influence of pH on bacterial richness across Wales (George

et al., 2019) and the globe (Lauber et al., 2009; Delgado-Baquerizo et al., 2018). This relationship is clearly evident in the distribution of selected anaerobes in the present study (Fig. 6.1B). Since both anaerobes and SRB demonstrated the same relationship with pH (Table 6.1), the relatively constant richness of SRB across AVCs may indicate that none of the land uses represent their ideal habitat.

Our findings demonstrate a relatively constant richness of SRB across diverse temperate soil ecosystems. In addition, we found that the distribution of anaerobic bacteria followed established trends with our study area. It would be interesting to see if this trend scales towards Mediterranean or sub-Arctic climates in continental Europe. Additionally, the integration of real-time PCR techniques targeting sulphate reductase genes could help elucidate the discrepancies between SRB richness, abundance, activity and S supply. Nevertheless this work highlights the use of national-scale environmental DNA biodiversity inventories in investigating localised microbial populations.

6.2 Acknowledgements

This research was supported under the Glastir Monitoring & Evaluation Programme (Contract reference: C147/2010/11) and by the UK Natural Environment Research Council (NERC) through the RC/Centre for Ecology & Hydrology (CEH Project: NEC04782). PBLG is supported by a Soils Training and Research Studentship (STARS) grant from the Biotechnology and Biological Sciences Research Council and NERC [Grant number NE/M009106/1]. KPC is funded by the Maranhão Foundation for Research, Science and Technology Development (FAPEMA). We thank Sam Tomlinson (CEH) for his assistance generating sulphate deposition maps.

6.3 References

Abdelmseeh, V.A., Jofreit J., Hayward, G., 2008. Sulphate and sulphide corrosion in livestock buildings, part I: concrete deterioration. Biosystems Engineering 99, 372-381.

Allison, M.F., Fowler, J.H., Allen, E,J., 2001. Responses of potato (*Solanum tuberosum*) to potassium fertilizers. Journal of Agricultural Science 136, 407-426.

Bahr, M., Crump, B.C., Klepac-Ceraj, V., Teske, A., Sogin, M.L., Hobbie, J.E., 2005. Molecular characterization of sulfate-reducing bacteria in a New England salt marsh. Environmental Microbiology 7, 1175-1185.

Barton, L.L., Fauque, G.D., 2009. Biochemistry, physiology and biotechnology of sulfate-reducing bacteria. Advances in Applied Microbiology 68, 41-98.

Biswas, K., Taylor, M.W., Turner, S.J., 2014. *dsrAB*-based analysis of sulphate-reducing bacteria in a moving bed biofilm reactor (MBBR) wastewater treatment plants . Environmental Biotechnology 98, 7211-7222.

Carvalho, M.C.S., van Raij, B., 1997. Calcium sulphate, phosphogypsum and calcium carbonate n the amelioration of acid subsoils for root growth. Plant and Soil 192, 37-48.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.W., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T., Singh, B.K., Fierer, N., 2018. A global atlas of the dominant bacterial found in soil. Science 359, 320-325.

Drenovsky, R.E., Steenwerth, K.L., Jackson, L.E., Scow, K.E., 2010. Land use and climatic factors structure regional patterns in soil microbial communities. Global Ecology and Biolgeography 19, 27-39.

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z., Hughes, S., Lawlor, A.J., Poskitt, J., Potter, E., Robinson, D.A., Scott, A., Wood, C., Woods, C., 2010. Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology, 192 pp. (CS technical Report No. 9/07, CEH Project Number: C03259).

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L. 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. Nature Communications 10, 1107.

Hines, M.E., Evans, R.S., Sharak, B.R., Genthner, Willis, S.G., Friedman, Rooney-Varga, J.N., Devereux, R., 1999. Molecular phylogenetic and biogeochemical studies of sulfate-reducing bacteria in the rhizosphere of *Spartina alterniflora*. Applied and Environmental Microbiology 65, 2209-2216.

Hothorn, T., Bretz, F., Westfall, P., 2008. Simultaneous inference in general parametric models. Biometrical Journal 50, 346-363.

Jeong, H., Yi, H., Sekiguchi, Y., Muramatsu, M., Kamagata, Y., Chun, J., 2004. *Clostridium jejunse* sp. nov., isolated from soil. International Journal of Systematic and Evolutionary Microbiology 54, 1465-1468.

Kirk, G.J.D., Bellamy, P.H., Lark, R.M., 2010. Changes in soil pH across England and Wales in response to decreased acid deposition. Global Change Biology 16, 3111-3119.

Lamers, L.P.M., Govers, L.L., Janssen, I.C.J.M., Geurts, J.J.M., Van der Welle, M.E.W., Van Katwijk, M.M., Van der Heide, T., Roelofs, J.G.M., Smolders, A.J.P., 2013. Sulfide as a soil phytotoxin – a review. Frontiers in Plant Science 4, 268 doi: 10.3389/fpls.2013.00268.

Lauber, C.L., Hamady, M., Knight, R., Fierer, N., 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental cale. Applied and Environmental Microbiology 75, 5111-5120.

Li, Z.-G., Min, X., Zhou, Z.-H., 2016. Hydrogen sulphide: a signal molecule in plant cross-adaptation. Frontiers in Plant Science. 7, 1621.

Lovley, D.R., Ueki, T., Zhang, R., Malvankar, N.S., Shrestha, P.M., Flanagan, K.A., Aklujkar, M., Butler, JE., Giloteaux, L., Rotaru, A.-E., Holmes, D.E., Franks, A.E., Orellana, R., Risso, C., Nevin, K.P., 2011. *Geobacter*: the microbe electric's physiology, ecology, and practical applications. Advances in Microbial Physiology 59, 1-100.

Lovett, G.M., Thompson, A.W., Anderson, J.B., Bowser, J.J., 1999. Elevational pttern of sulphur deposition at a site in the Catskill Mountain, New York. Atmospheric Environment 33, 617-624.

Masuda, Y., Itoh, H., Shiratori, Y., Senoo, K., 2018. Metatranscriptomic insights into microbial consortia driving methane metabolism in paddy soils. Soil Science and Plant Nutrition 64, 455-464.

McMurdie, P.J., Holmes, S., 2013. phyloseq: an R package for reproducible interactive analysis and graphics fof Microbiome census data. PLoS ONE 8(4), e61217.

Muzyer, G., Stams, A.J.M., 2008. The ecology and biotechnology of sulphate-reducing bacteria. Nature Reviews Microbiology 6, 441-454.

Overmann, J., van Gemerden, H., 2000. Microbial interactions involving sulphur bacteria: implications for the ecology and evolution of bacterial communities. FEMS Microbiology Reviews 24, 591-599.

Pan, B., Lam, S.K., Moiser, A., Lou, Y., Chen, D., 2016. Ammonia volatilization from synthetic fertilizers and its mitigation strategies: a global synthesis. Agriculture, Ecosystems and Environment 232, 283-289.

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., R Core Team, 2019. "nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-140, https://CRAN.R-project.prg/package=nlme

R Core Team, 2018. R: A language and environment for statistical computing. T Foundation for Statistical Computing. Vienna, Ausgtria. https://www.R-project.org

Review of Transboundary Air Pollution, 2012. Acidification, Eutrophication, Ground Level Ozone and Heavy Metals in the UK. Contract Report to the Department for Environment, Food and Rural Affairs. Centre for Ecology & Hydrology, Wallingford, UK. Accessible at: http://rotap.ceh.ac.uk

Reynolds, B., Chamberlain, P.M., Poskitt, J., Woods, C., Scott, W.A., Rowe, E.C., Robinson, D.A., Frogbrook, Z.L., Keith, A.M., Henrys, P.A., Black, H.I.J., 2013. Countryside Survey: National "Soil Change" 1978–2007 for Topsoils in Great Britain—acidity, carbon, and total nitrogen status. Vadose Zone Journal, 12, doi:10.2136/vzj2012.0114.

Scheid, D., Stubner, S., 2001. Structure and diversity of Gram-negative sulfate-reducing bacteria on rice roots. FEMS Microbiology Ecology 36, 175-183.

Stevens, P.A., Ormerod, S.J., Reynolds, B., 1997. Final Report in the Acid Waters Survey for Wales: Volume 1 Main Text. Institute for Terrestrial Ecology (National Environment Research Council Project T07072R5). Bangor, UK.

Stubner, S., 2004. Quantification of Gram-negative sulphate-reducing bacteria in rice field soil by 16S rRNA gene-targeted real-time PCR. Journal of Microbial Methods 57, 219-230.

Tabatabai, M.A., 1996. Sulfur. In: Sparks, D.L., Page, A.L., Helmke, P.A., Loeppert, R.H., Soltanpour, P.N., Tabatabai, M.A., Johnston, C.T., Sumner, M.E. (Eds.), Methods of Soil Analysis, second edition. Soil Science Society of America, Madison, WI, pp. 921-961.

Xia, F.F., Su, Y., Wei, X.M., He, Y.H., Wu, Z.C., Ghulam, A., He, R., 2014. Diversity and activity of sulphur-oxidising bacteria and sulphate-reducing bacteria in landfill cover soils. Letters in Applied Microbiology 59, 26-34.

Yuan, C.-L., Zhang, L.M., Wang, J.-T., Hu, H.-W., Shen, J.-P., Cao, P., He, J.-Z., 2019. Distributions and environmental drivers of archaea and bacteria in paddy soils. Journal of Soils and Sediments 19, 23-27.

Chapter 7: Shifts in soil microbial biological and functional diversity under long-term carbon deprivation

This manuscript is in preparation to be submitted to *Bioinformatics*

George, P.B.L.[1,2,*], Fidler, D.B.[1], Van Nostrand, J.D.[3], Atkinson, J.A.[4], Mooney, S.J.[4], Creer, S.[1], Griffiths, R.I.[2], McDonald, J.E.[1], Robinson, D.A.[2], Jones, D.L.[1,5], 2019. *In Preparation*.

[1] School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] Institute for Environmental Genomics, The University of Oklahoma, Norman, OK 73019, USA

[4] School of Biosciences, The University of Nottingham, Sutton Bonington Campus, LE12 5RD, UK

[5] SoilsWest, UWA School of Agriculture and Environment, The University of Western Australia, Perth WA 6009, Australia

[*] Corresponding author

P.B.L.G., D.B.F., S.C., R.I.G., J.E.M., and D.L.J. conceived and developed this project. P.B.L.G. and K.P.C. prepared the dataset. P.B.L.G and D.L.J. performed physicochemical analyses. J.A.A. and S.J.M. performed x-ray imagining. J.D.V.N. generated GeoChip data, and assisted P.B.L.G. with their analysis. P.B.L.G. performed bioinformatics and statistical analyses with assistance from D.B.F. P.B.L.G. wrote the first draft of the manuscript. All authors contributed to subsequent revisions, read, and approved the final manuscript.

Abstract

Soil organisms are crucial to the support and maintenance of ecosystems through the provision of ecosystem services, such as nutrient cycling and decomposition processes. Many of these organisms are reliant on nutrient inputs from plants, which may be disrupted by anthropogenic disturbance. However, some organisms thrive under disturbance and could mitigate their impacts. We assessed soil community composition and functional genes under a long-term carbon deprivation field experiment. Two sets of paired soils were deprived of carbon inputs for 10 years and 1 year, respectively. Soil DNA was extracted for metabarcoding and GeoChip microarray analyses. Richness of all soil organisms fell under carbon deprivation after 10 years, but not after 1 year. There was a simultaneous reduction in log-fold change of most functional genes, though gene copies increased for phytase as well as for genes involved in decomposing recalcitrant carbon and methanogenesis under carbon deprivation. Several taxa were identified as indicators of both normal and deprived soils. Bioindicator and biomass analyses also suggest a differential loss in fungi under long-term carbon deprivation. Clear differences in pore structure were also observed between vegetated and 10 year carbon deprived soils. Our results highlight a concurrent loss in soil structure and biodiversity following carbon deprivation. Bioindicator and biomass analyses also suggest a differential loss in fungi under long-term carbon deprivation. We have shown that carbon depletion has a profound impact on soils, requires long timescales and may be prevented with timely interventions.

Key words: *GeoChip; Metabarcoding; Bioindicators; Carbon cycling; Nitrogen cycling; Methanogenesis; Anaerobic respiration; Functional genes*

7.1 Introduction

Soil carbon (C) loss is a serious concern for global agricultural management in addition to climate change (1). Many models predict that rising global temperatures will stimulate C loss from soils (2, 3, 4), especially at higher latitudes (5). Yet, empirical evidence shows that despite initial increases, release of soil-borne carbon dioxide (CO_2) returns to ambient levels within a matter of years (6, 7). Such attenuation suggests that as available soil C is consumed, microbes may be forced to use alterative energy sources. Yet it is unclear how the whole soil microbial community (i.e. prokaryotes, fungi, and protists) respond to long-term removal of C sources.

Soil C reflects a large range of organic compounds with varying levels of accessibility and degradability to soil organisms. A continuum of C forms can be defined on these criteria as labile or recalcitrant (8) based on residence time in soil (9). A wide range of organisms readily consume labile compounds such as starch and cellulose (10, 11), whereas more recalcitrant forms, like lignin and polyphenolics (11) can be decomposed by only a relatively few taxa (10). Environmental factors strongly influence soil C processing and may promote or inhibit decomposition (12), which can also influence microbial community structure (13, 14). Few organisms are able to utilise multiple substrates, which creates niche partitioning and promotes mutualistic associations (10). Thus, as soil C is depleted, it is expected that organisms dependent on labile forms will not persist.

Anaerobic microbes utilise electron receptors that are not derived from oxygen for cellular respiration. Though anaerobic microbes are commonly associated with less productive or extreme environments, they are prevalent across a wide range of habitats (15). Many play important roles in soils as part of nutrient cycles, i.e. nitrogen (N) (16). By-products from

anaerobic respiration are often potent greenhouse gases (GHG). Rates of such GHG from soils may be also be promoted by climate change in a positive feedback loop (17). For example, methane (CH_4) has 28-times greater global warming potential than CO_2 (18). As anaerobic respiration can be obligatory or facultative (15), identifying important anaerobic organisms in non-extreme environments may go far to inform GHG emissions models.

Bellamy et al. (19) reported that soil organic C is being lost at a rate of 0.6% per year in England and Wales. Long-term fallow experiments typically demonstrate an increase in acidity (20), loss of soil aggregate stability (21) and a loss of labile C (22; 23) as soils are left barren. Bacq-Labreuil et al. (24) recently showed that long-term C loss (>50 years) resulted in reduced soil porosity at the μm scale in comparison to vegetated treatments. This research complements findings on the biological component of these same English long-term fallow soils. Bacterial biomass is considerably reduced under long-term fallow (25, 26). Interestingly, these same experiments found minimal differences in diversity measures (25, 26), potentially due to the high retention of inactive DNA in fallow soils (25), suggesting that our understanding of soil community responses to long-term fallow is incomplete. Identifying taxa indicative of long-term C deprivation may also prove an effective tool in bioremediation projects. So far, we have discussed functions and processes relating primarily to prokaryotes. However, soils also support a wide range of viruses, fungi, and protists (27, 28). These groups are often overlooked and could offer exciting insights into soil community response to stress.

Here, we used covered bare-fallow and vegetated soil to investigate the impact of long-term C deprivation on major components of soil microbial communities. Comparisons were made in a set of plots, which had been maintained for 1 year before sampling and a set maintained for 10 years. We also explored differences in soil structure for the 10-year fallow

soils only. Our aims were to: (i) determine the consequences of long-term C depletion on soil structure; (ii) determine if long-term C depletion would cause a shift in microbial functional and biological diversity; and (iii) identify taxa indicative of C deprived soils. We expected that soil structure and microbial biodiversity would be significantly lower after long-term C depletion and these deprived soils would be dominated by indicator taxa that increase in functional genes involved in methanogenesis and processing recalcitrant C.

7.2 Results

7.2.1 Soil properties

Differences in soil properties between vegetated and fallow soils for both age classes are presented in Table 7.1. There were few significant differences between treatments in 1 year-old soil across studied soil physicochemical properties. No differences were observed between pH, electrical conductivity (EC), ammonium (NH_4^+), available phosphorous (P), calcium (Ca), potassium (K), magnesium (Mg), cation exchange capacity (CEC), total cations, C, and N, or C: N ratio (Table 7.1). There was however, a significantly higher level of sodium (Na) in vegetated soils ($P < 0.001$) as well as significantly lower moisture content in addition to concentration of nitrate (NO_3^-) ($P = 0.04$). Soil CO_2 flux was also significantly ($P < 0.001$) higher in vegetated soils (Table 7.1).

In the 10 year-old soils, differences in physicochemical properties were more prevalent. Both pH and EC (both $P = 0.01$) were significantly different between vegetated and fallow soils, with lower pH and higher EC in fallow soils. Soil CO_2 flux, CEC, and concentrations of all cations, except aluminium (Al), were significantly greater in vegetated soils. Interestingly, Al concentrations were significantly higher in fallow soils after 10 years of C deprivation ($P <$

0.001). Total C (P = 0.04) and N (P = 0.03) were also significantly greater in vegetated soils, though C: N did not change between treatments (Table 7.1).

X-ray CT imaging revealed clear differences in soil structure between 10-year fallow and vegetated soils (Fig. 7.1). We present 3D images of pore architecture from the vegetated and fallow soils at the column (Fig. 7.1c, d) and aggregate (Fig. 7.1g, h) scales. In general, the vegetated soils appear more porous with large pores that were more connected, though this was more apparent at the column scale than the aggregate scale. Pore morphology measurements revealed a decline of all porosity measures in fallow soils noting a large Euler number relates to reduced pore connectivity (Appendix 5 Table 1). At the column scale, total porosity and total pore area of the fallow samples were lower (Fig. 7.1a, b) than vegetated soils though not statistically significant (both: P = 0.07). However, mean pore size was significantly reduced in fallow samples (P < 0.001). Pore size distribution (Appendix 5 Fig. 2a) showed a reduction in the size of pores across all classes in the fallow in comparison with the vegetated soils, with a greater reduction generally in the number of larger pore size classes. This is demonstrated by the coefficient of uniformity value of c.110 in the vegetated soil versus c.58 in the 10-year fallow.

A similar pattern was observed at the aggregate-level, with total porosity and pore area being reduced in fallow samples (Fig. 7.1e, f), although to a lesser extent than at column scale (P = 0.25 and p = 0.22 respectively). Again, mean pore size was significantly lower in fallow soils (P < 0.001) with a mean value of 0.05 mm^2 in the vegetated soil compared to 0.01 mm^2 in the fallow soil. Pore size distributions were similar between fallow and vegetated treatments at the aggregate scale in the smaller pore size classes, except for in larger pore size classes where more large pores were recorded in the vegetated plots (Appendix 5 Fig. 2b). Interestingly, the

converse for pore connectivity was found at the aggregate scale with higher pore connectivity recorded in the 10-year fallow plots (observed as a higher Euler number) though not significantly so (Appendix 5 Table 1).

*Table 7.1 Soil physicochemical properties from soils subjected to carbon deprivation for 1 year and 10 years. Mean values (± SE) are presented for both treatments within each age class. Significantly greater values within each age class are indicated by: *** ($P < 0.001$), ** ($0.001 > P < 0.01$), *($0.01 > P < 0.05$), and blank ($P > 0.05$). Aluminium was not determined (N.D.) in 1 year-old samples.*

Soil physicochemical properties	**1 year**		**10 years**	
	Vegetated	Fallow	Vegetated	Fallow
pH	5.04 (± 0.05)	4.79 (± 0.04)	5.46 (± 0.04)*	4.67 (± 0.06)
Electrical conductivity (µS cm^{-1})	118.0 (± 7.45)	207.0 (± 26.31)	104.67 (± 2.19)	149.33 (± 2.85)*
Soil CO_2 flux (µmol m^{-2} S^{-1})	6.85 (± 0.58)***	2.04 (± 0.22)	4.97 (± 0.29)***	0.99 (± 0.24)
Moisture (% dry weight)	23.96 (± 0.45)	43.06 (± 2.83)***	35.82 (± 1.53)	34.37 (± 0.10)
Total C (%)	3.11 (± 0.16)	3.57 (± 0.18)	3.73 (± 0.17)*	2.54 (± 0.34)
Total N (%)	0.31 (± 0.01)	0.34 (± 0.01)	0.32 (± 0.01)*	0.23 (± 0.02)
C: N ratio	9.96 (± 0.16)	10.39 (± 0.36)	11.81 (± 0.28)	10.75 (± 0.56)
Nitrate (mg NO_3^- kg^{-1})	2.09 (± 0.53)	30.45 (± 10.49)*	0.88 (± 0.19)	0.35 (± 0.12)
Ammonium (mg NH_4^+ kg^{-1})	43.51 (± 10.93)	38.01 (±9.96)	1.29 (± 0.14)	1.86 (± 1.14)
Available P (mg P kg^{-1})	3.36 (± 0.30)	3.28 (± 0.09)	4.46 (± 0.68)	3.21 (± 1.05)
Calcium (mmol Ca kg^{-1})	19.78 (±1.59)	18.76 (± 0.71)	18.28 ± (2.91)**	4.25 (± 0.77)
Potassium (mmol K kg^{-1})	2.77 (± 0.20)	4.10 (± 0.34)	0.82 (± 0.07)*	0.54 (± 0.03)
Sodium (mmol Na kg^{-1})	0.96 (± 0.07)**	0.61 (± 0.19)	6.48 (± 0.65)**	0.99 (± 0.07)
Magnesium (mmol Mg kg^{-1})	4.28 (± 0.29)	3.86 (± 0.16)	0.01 (± 5.3 x 10^4)***	0.004 (± 3.86 x 10^{-4})
Aluminium (mmol Al kg^{-1})	N.D.	N.D	0.001 (± 0.001)	0.03 (± 0.003)***
Total cations (mmol kg^{-1})	48.52 (± 3.49)	43.09 (± 1.36)	43.89 (± 6.31)**	10.45 (± 1.47)
Cation exchange capacity (mmol NH_4 kg^{-1})	974.48 (± 29.40)	917.75 (± 49.10)	56.68 (± 4.52)*	40.80 (± 3.22)

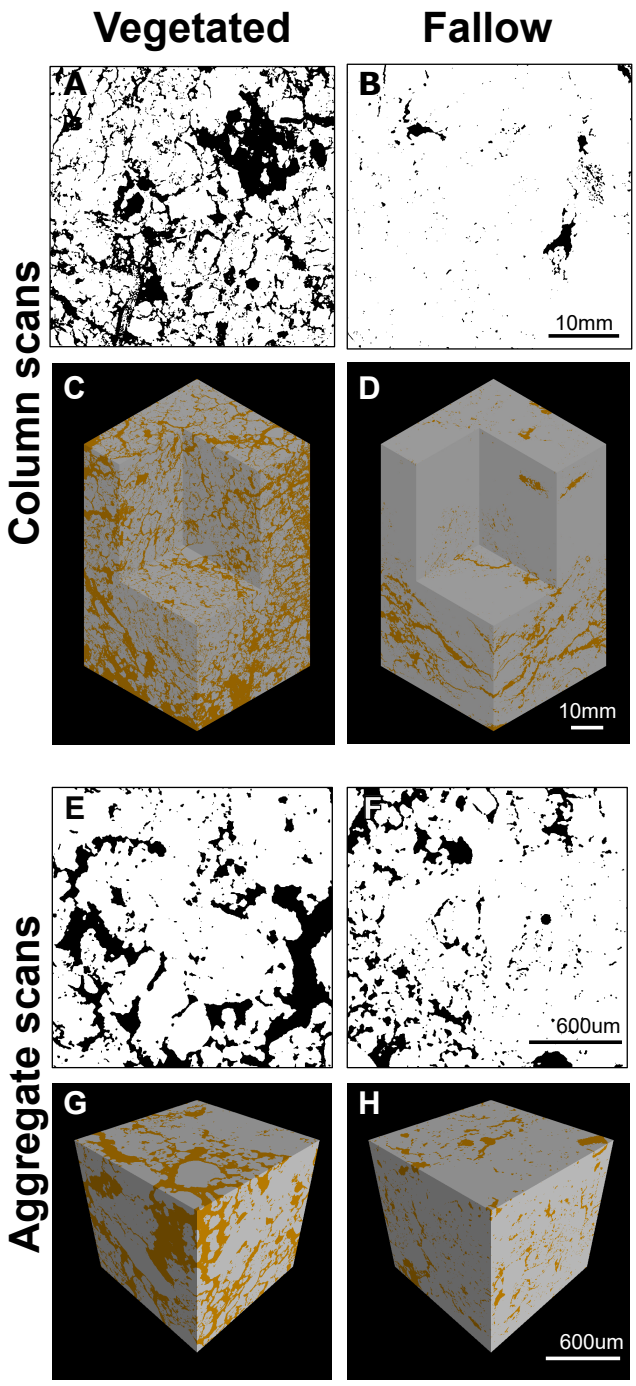


Figure 7.1 Example soil porosity images collected using x-ray µCT. ***A-B****) 2D binary images collected from the centre of a vegetated* ***A****) or fallow* ***B****) column. Pore space shown in black, soil in white.* ***C-D****) 3D reconstruction of soil columns from vegetated* ***C****) and fallow* ***D****) columns.* ***E-F****) 2D binary images collected from the centre of a vegetated* ***E****) or fallow* ***F****) aggregate. Pore space shown in black, soil in white.* ***G-H****) 3D reconstruction of soil aggregates from vegetated* ***G****) and fallow* ***H****) aggregates. Pore space shown in brown, soil in grey.*

7.2.2 Microbial biomass

Total microbial biomass inferred from PLFA analysis revealed broadly similar trends (Table 2). In 1 year-old soils, total PLFA, the proportion of fungal and bacterial PLFAs, in addition to the fungi: bacteria ratio, were all significantly greater in vegetated soils. The proportion of anaerobes in these soils did not meet detection thresholds (Table 7.2). In 10 year-old soils, total PLFA was significantly lower ($P = 0.03$) in fallow soils (Table 7.2). The fungi: bacteria PLFA ratio ($P < 0.001$) also followed this trend. When assessed by proportional contribution to the total PLFA, per cent fungi ($P = 0.002$) and anaerobes were significantly greater ($P = 0.01$) in vegetated sites, though proportions of Gram-positive bacteria, Actinomycetes (both $P = 0.3$), and total bacteria ($P < 0.001$) were higher in fallow soil (Table 7.2). The qCO_2 was greater in vegetated soils in both age classes ($P < 0.001$, $P = 0.01$, respectively).

*Table 7.2 Microbial biomass fractions from phospholipid fatty acid analysis (PLFA) from soils subjected to carbon deprivation for 1 year and 10 years. Mean values (± SE) are presented for both treatments within each age class. Significantly greater values within each age class are indicated by: *** (P < 0.001), ** (0.001 > P < 0.01), *(0.01 > P < 0.05), and blank (P > 0.05). Percentage of anaerobes was not determined (N.D.) in 1 year-old samples.*

Microbial biomass from PLFA analyses	**1 year**		**10 years**	
	Vegetated	Fallow	Vegetated	Fallow
Total PLFA (nmol g^{-1})	246.03 (± 7.87)*	212.23 (± 2.98)	190.81 (± 24.52)**	107.83 (± 6.18)
Gram-negative (%)	43.11 (± 0.19)	43.04 (± 0.27)	48.69 (± 0.15)	47.82 (± 0.54)
Gram-positive (%)	29.73 (± 1.65)	37.82 (± 2.01)	26.57 (± 0.21)	28.27 (± 0.3)**
Actinomycetes (%)	11.17 (± 1.78)	10.73 (± 2.08)	13.53 (± 0.22)	15.33 (± 0.47)*
Anaerobes (%)	N.D.	N.D.	1.04 (± 0.02)*	0.81 (± 0.08)
Fungi (%)	6.35 (± 0.11)**	4.85 (± 0.11)	5.19 (± 0.21)	7.17 (± 0.3)**
Bacteria (%)	91.59 (± 0.16)*	90.03 (± 0.20)	89.83 (± 0.31)	92.23 (± 0.42)**
Fungi: bacteria ratio	0.08 (± 0.002)**	0.06 (±0.001)	0.10 (± 0.004)***	0.07 (± 0.003)
qCO_2 (µmol CO_2 _ $m^{2\,-s1}$ / nmol g^{-1})	0.03 (± 0.002)***	0.001 (± 0.001)	0.03 (± 0.003)*	0.01 (± 0.003)

7.2.3 Microbial biodiversity

A total of 973 prokaryotic, 336 fungal, and 1,638 protistan OTUs were identified across all samples. There were no significant differences between richness or H' index between vegetated and fallow in 1 year-old sites across all organismal groups (Fig. 7.2). However, differences in richness were evident in all groups in 10 year-old soils. In all cases, richness was lower in fallow soils (Fig. 7.2a-c; all $P = 0.01$). This was also the case for H' index values for prokaryotes (Fig. 7.2d; $P < 0.003$) but not for fungi (Fig. 7.2e; $P = 0.59$) or protists (Fig. 7.2f; $P = 0.17$). Viral gene richness derived from GeoChip data shows the same trend as microbial richness. In 10 year-old soils, viral gene richness was significantly lower ($P = 0.001$) in fallow than vegetated soils (Appendix 5 Fig. 2).

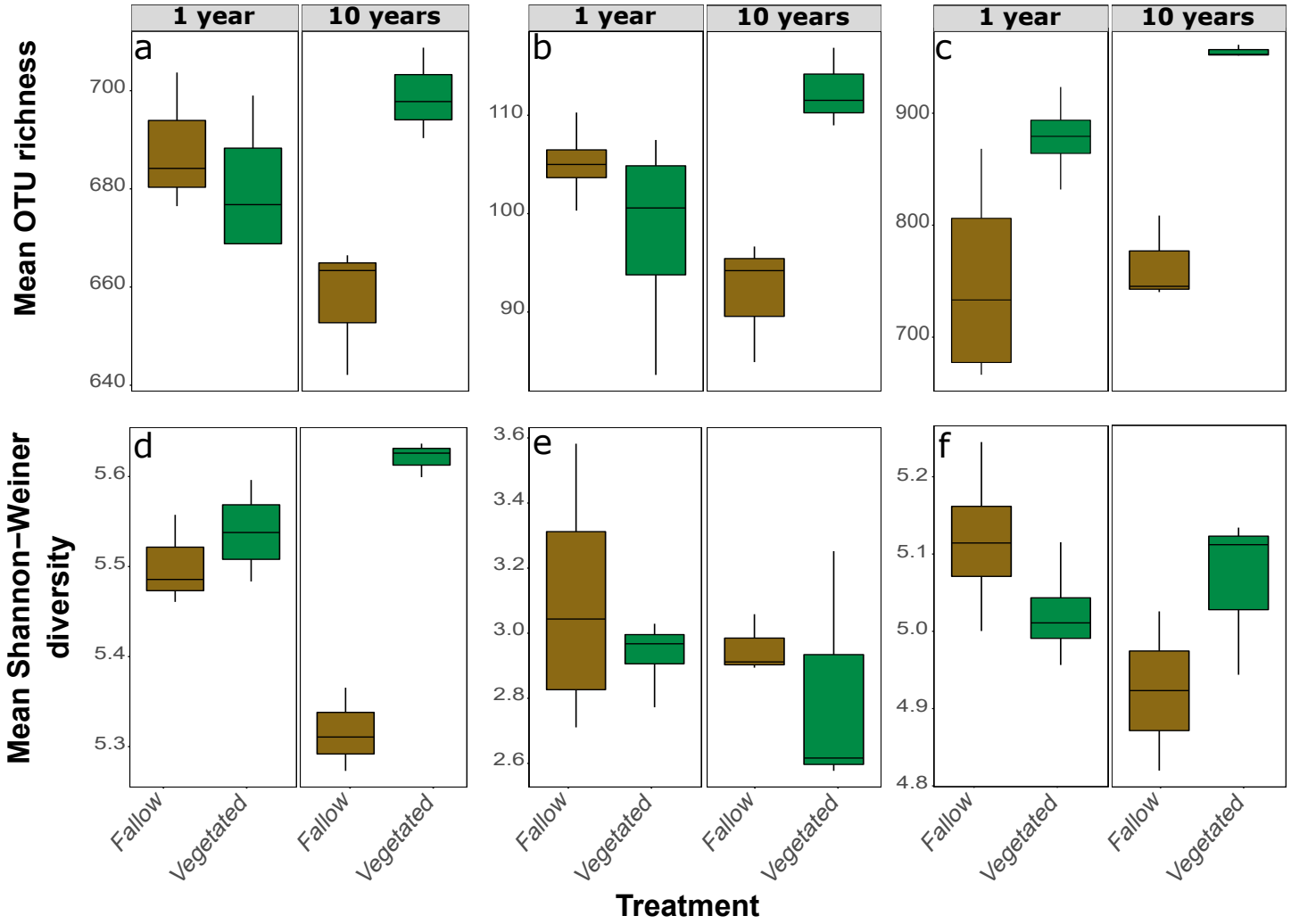


*Figure 7.2 Measures of α-diversity of soil microbial community fractions in covered and vegetated soils. Each plot is split by age. Plots **a**), **b**), and **c**) show richness of bacteria, fungi, and protists, respectively. Plots **d**), **e**), and **f**) show Shannon-Weiner diversity values of bacteria, fungi, and protists, respectively.*

7.2.4 Microbial functional diversity

Absolute signal intensity was almost universally greater in vegetated soils than fallow soils; however, comparisons of changes in proportional signal intensity revealed important trends. Most C degradation genes (55.3%) were significantly greater in ($P < 0.05$) in vegetated soils than fallow soils (Fig. 7.3). These included the more labile C forms (sugar, starch, cellulose), but also the majority of genes involved in degrading more recalcitrant forms. Yet, there was a significant increase in proportional signal intensity under fallow conditions for 8 genes, which are involved in the degradation of recalcitrant forms of C, including pectin (pectinase, pel_Cdeg, and PME) and aromatics (camDCAB, tannase, and vanA). Yet interestingly there was also a significant increase in the proportion of the hemicellulose degrading genes ara ($P = 0.02$) and mannanase ($P = 0.001$) (Fig. 7.3). Of note, variation of signal intensities for some genes (i.e. pulA and apu) meant that significant differences could not be identified, despite clear differences in absolute values (Fig. 7.3). Data from fungal xylose reductase was omitted from Fig. 7.3 as it was not detected or did not meet the criteria for analyses in all fallow samples, but was present in all vegetated samples ($P < 0.001$).

As with C degradation genes, the majority (61.1%) of genes involved in anaerobic respiration processes increased significantly ($P < 0.05$) in vegetated samples (Fig. 7.4). This included 8 genes (ACS, cdhC_methane, mtmC, mttB, Hmd, mtaC, mtbB, and MT2) involved in methanogenesis that are not presented in Fig. 7.4. Although significantly greater in vegetation ($P < 0.001$), these genes were omitted for clarity; ACS, cdhC_methane, and mtmC were not detected or did not meet the criteria for analyses in one or more fallow samples. There was evidence of Hmd, mtaC, mtbB, in only 1 fallow sample, and of MT2 in 2. Abundances of genes involved in acetogenisis (FTHFS), N cycling (nasA, napA, nosZ), and

the reduction of phosphate and sulphur (S) compounds were also greater in vegetated soils. However, there were some genes with greater differential abundance in fallow soils. This included 4 genes involved in methanogenesis (fmdB_fwdB, Ftr, hdrB, and mcrA) and phytase (Fig. 7.4).

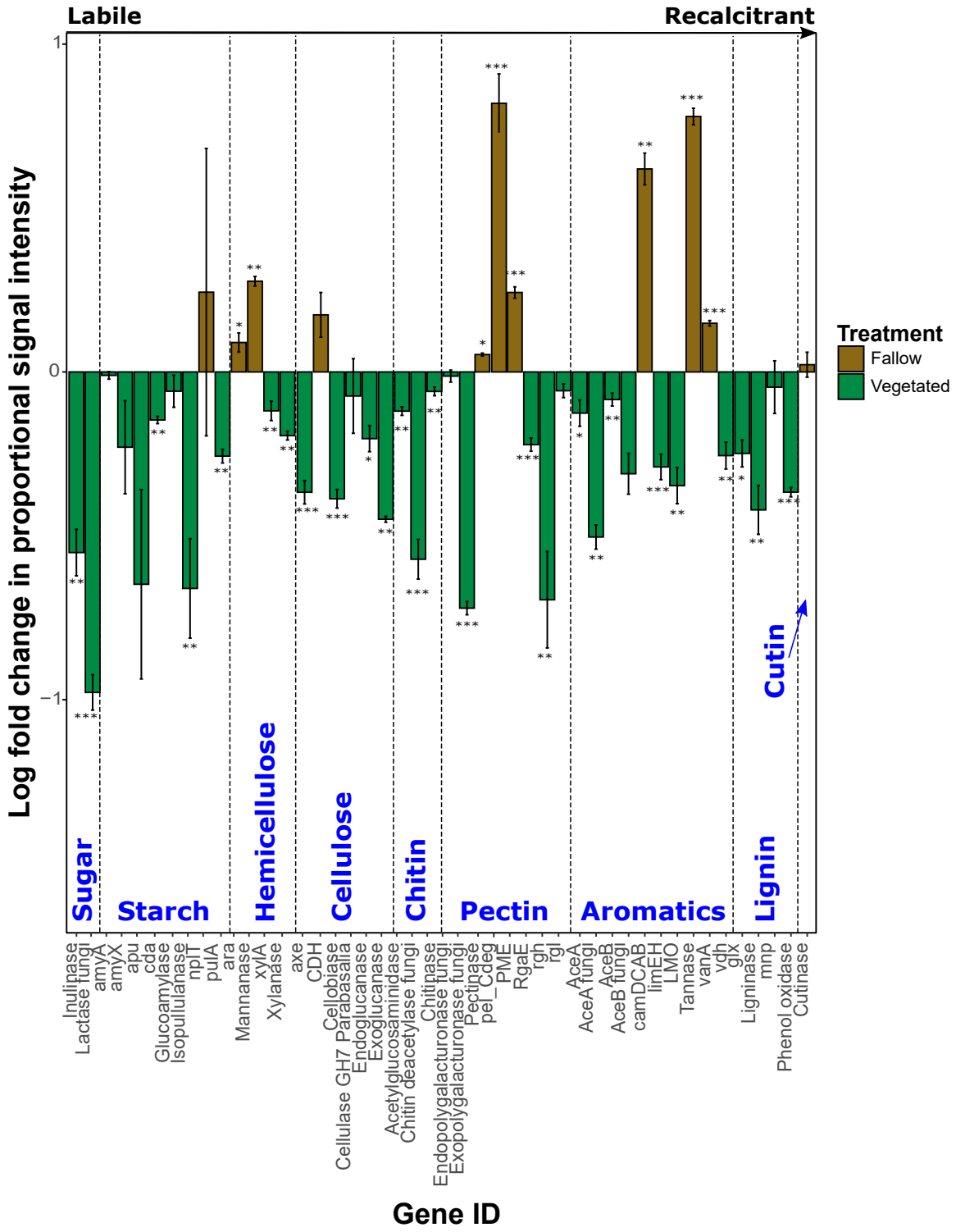


*Figure 7.3 Logarithmically transformed fold change in signal intensity from GeoChip data of carbon degradation genes. Genes are ordered from labile to recalcitrant. Error bars denote standard error. Significant differences are indicated by *. *** indicates $P < 0.001$, ** $0.001 > P < 0.01$, * $0.01 > P < 0.05$, blank indicates $P > 0.05$.*

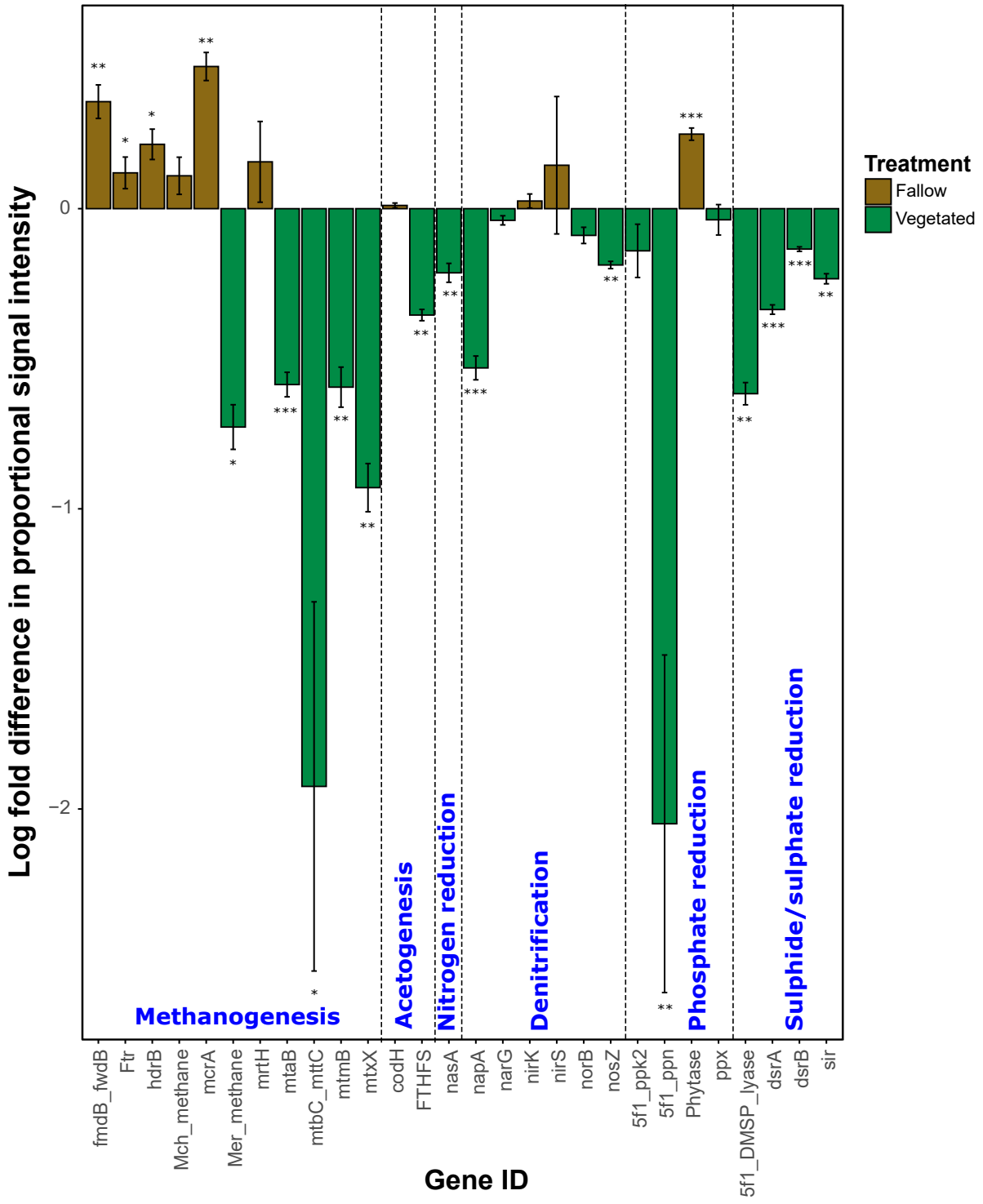


*Figure 7.4 Logarithmically transformed fold change in signal intensity from GeoChip data of carbon degradation genes. Genes are ordered from labile to recalcitrant. Error bars denote standard error. Significant differences are indicated by *. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.*

7.2.5 Bioindicators

We used linear discriminant analyses (LDA) with LDA effect size (LEfSe) methodology and differential abundance analyses to identify taxa OTU bioindicators of vegetated or fallow soils, respectively, in 1 and 10 year-old treatments. Since we identified a large number of bioindicators, we have chosen to focus on only some of the findings here. More detailed discussion can be found in Appendix 5.1; complete inventories are presented in Appendix 5 Tables 2-7.

For prokaryotes, there were more indicator taxa in 10 year-old soils than 1 year-old soils (Fig. 7.5a, c). In the 1 year-old dataset there were 11 indicator taxa, of which, 5 were characteristic of vegetated and 6 of fallow soils (Fig. 7.5a). *Rhodoplanes* and *Flavobacterium* were the best bacterial indicator taxa of vegetated 1 year-old sites. *Nitrosotalea devanaterra* and *Rhodanobacter* were the best indicators for 1 year-old fallow sites (Fig. 7.5a). Nakamurellaceae, *Nitrospira*, and *Solirubrobacter* were the best indicators of 10 year-old vegetated soils. *Methylosinus* and *N. devanaterra* were the strongest indicators of 10 year-old fallow (Appendix 5 Fig. 7.5c).

Identification of differentially abundant OTUs through DESeq2 revealed some congruence with indicator taxa from LDA data. For example, *N. devanaterra* (OTU_1863) and *Rhodanobacter* OTUs (OTU_865, OTU_870) are highlighted as being differentially abundant in 1 year-old fallow soils (Fig. 7.5b). Additionally, OTUs identified as *Streptomyces* (OTU_932, OTU_504), were more abundant in fallow sites (Appendix 5 Fig. 7.5a, b). In 10 year-old soils, an OTU identified as *Solirubrobacter* (OTU_1472) matched vegetated marker taxa from LDA analyses (Fig. 7.5c,d). *Paenbacillus* was also differentially abundant in fallow soils (Fig. 7.5d; Appendix 5 Table 3), which support LDA results (Fig. 7.5c; Appendix 5 Fig.

3). Unexpectedly, some differentially abundant OTUs matched to indicator taxa from opposing treatments. For example, *Rhodoplanes* was an indicator of vegetated soil in LDA data from 1 year-old soils (Fig. 7.5a) but some OTUs (OTU_1627, OTU_1633) were differentially abundant in both 1 year-old and 10 year-old fallow soils (Fig. 7.5b, d, Appendix 5 Fig. 3).

There were much fewer fungal indicator taxa and differentially abundant OTUs for both 1 year-old and 10 year-old sites (Fig. 7.6). Indeed only 4 indicator taxa (2 vegetated and 2 fallow) from 1 year-old soils and 9 indicator taxa (3 vegetated and 6 fallow) from 10 year-old taxa were identified by LDA (Fig. 7.6a, c). In 1 year-old soils, Orbiliomycetes and *Coprinopsis brunneofibrillosa* were indicative of the vegetated treatment and *Cotylidia undulate* and *Mucor heimalis* were indicative of fallow (Fig. 7.6a). Orbiliomycetes was also the strongest indicator of vegetated soils in 10 year-old soils. *Onygenales* and *C. candidolanata* were the strongest indicators of fallow (Fig. 7.6c).

The differentially abundant fungal OTUs in vegetated soils did not match the LDA-identified indicator taxa. Rather, these included 3 OTUs identified as Stephanosporaceae (OTU_301, OTU_272, OTU_246), 2 as different *Mycena* species (OTU_46, OTU_560), 1 as *Conocyb*e, (OUT_216) and an unnamed Ascomycota. There were no differentially abundant OTUs for 1 year-old fallow sites (Fig. 7.6b; Appendix 5 Table 4). The only match between differentially abundant OTUs and taxa from LDA data in the 10 year-old sites was OTU_221, identified as Orbiliomycetes, from vegetated soils (Fig. 7.6c, d). *C. fuscimarginata* was differentially abundant in 10 year-old vegetated soils (Fig. 7.6d; Appendix 5 Table 5). OTUs indicative of fallow in 10 year-old soils included *Trichomerium foliicola*, (Fig. 7.6d; Appendix 5 Table 5).

There were 20 protistan indicator taxa (9 vegetated and 11 fallow) for 1 year-old soils (Fig. 7.7; Appendix 5 Fig. 4) from LDA data. *Spongomonas* was the strongest protistan indicators of vegetated soils in 1 year-old soils (Fig. 7.7a). The families Vampyrellidae and Thaumatomonadidae (Fig. 7.7a) as well as ambiguously identified cercozoans and Oomycetes (Appendix 5 Fig. 5) were indicative of 1 year fallow. In the 10 year-old soils, there were 60 taxa indicative of vegetated soils and 29 taxa indicative of fallow soils (Fig. 7.7c; Appendix 5 Fig. 5). *Chloroidium* and *Spumella* were the strongest indicators of vegetated soils (Fig. 7.7c). Cercomonads, Trebouxiophyceae (green algae), and MAST_12C group Stramenopiles were characteristic of vegetated soils (Fig. 7.7c).

Congruence was found between differentially abundant protistan OTUs such as *Spongomonas* (OTU_1120) and *Hypotrichia* (OTU_2052), and vegetated indicator taxa from 1 year-old vegetated soils (Fig. 7.7a,b; Appendix 5 Table 6). Similarly, differentially abundant OTUs belonging to Vampyrellidae (OTU_1530) and MAST_12C group Stramenopiles appear in the 1 year-old fallow soils (Fig. 7b). *Heteromita* (OTU_9307) and members of the order Euglyphida were differentially abundant in 10 year-old fallow soils (Fig. 7.7d; Appendix 5 Table 7).

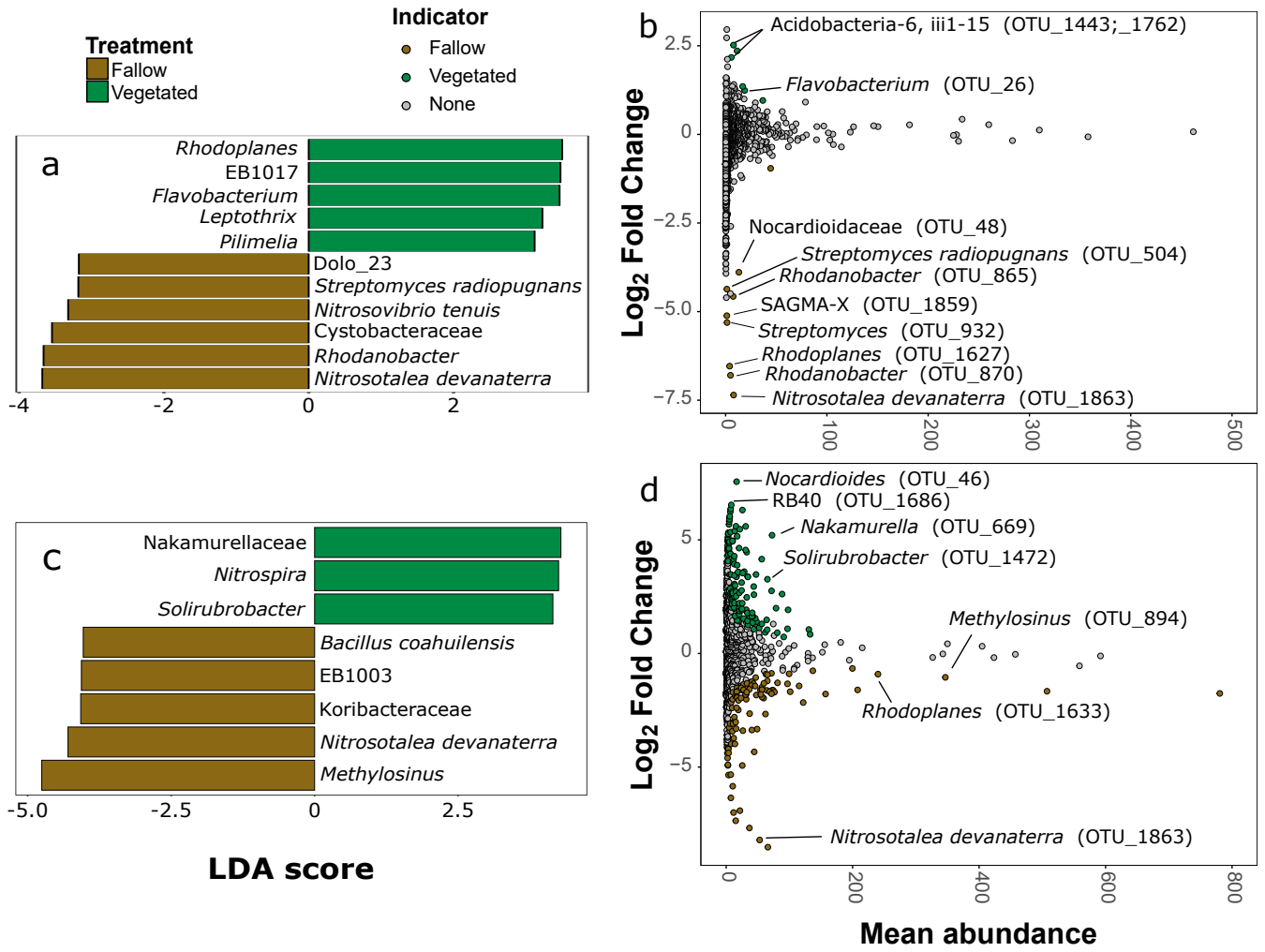


*Figure 7.5 Differential abundance of prokaryotic taxa and OTUs between covered and vegetated soils. Indicator taxa identified using linear discriminant analyses (LDA) for **a**) 1 year old soils and **c**) 10 year old soils. Only data with an LDA score greater than 4 are presented. OTUs indicative of treatment based on corrected P values from DESeq2 analysis identified in **b**) 1 year old and **d**) 10 year old soils. Labeled OTUs are discussed in the main text and/or Appendix 5.*

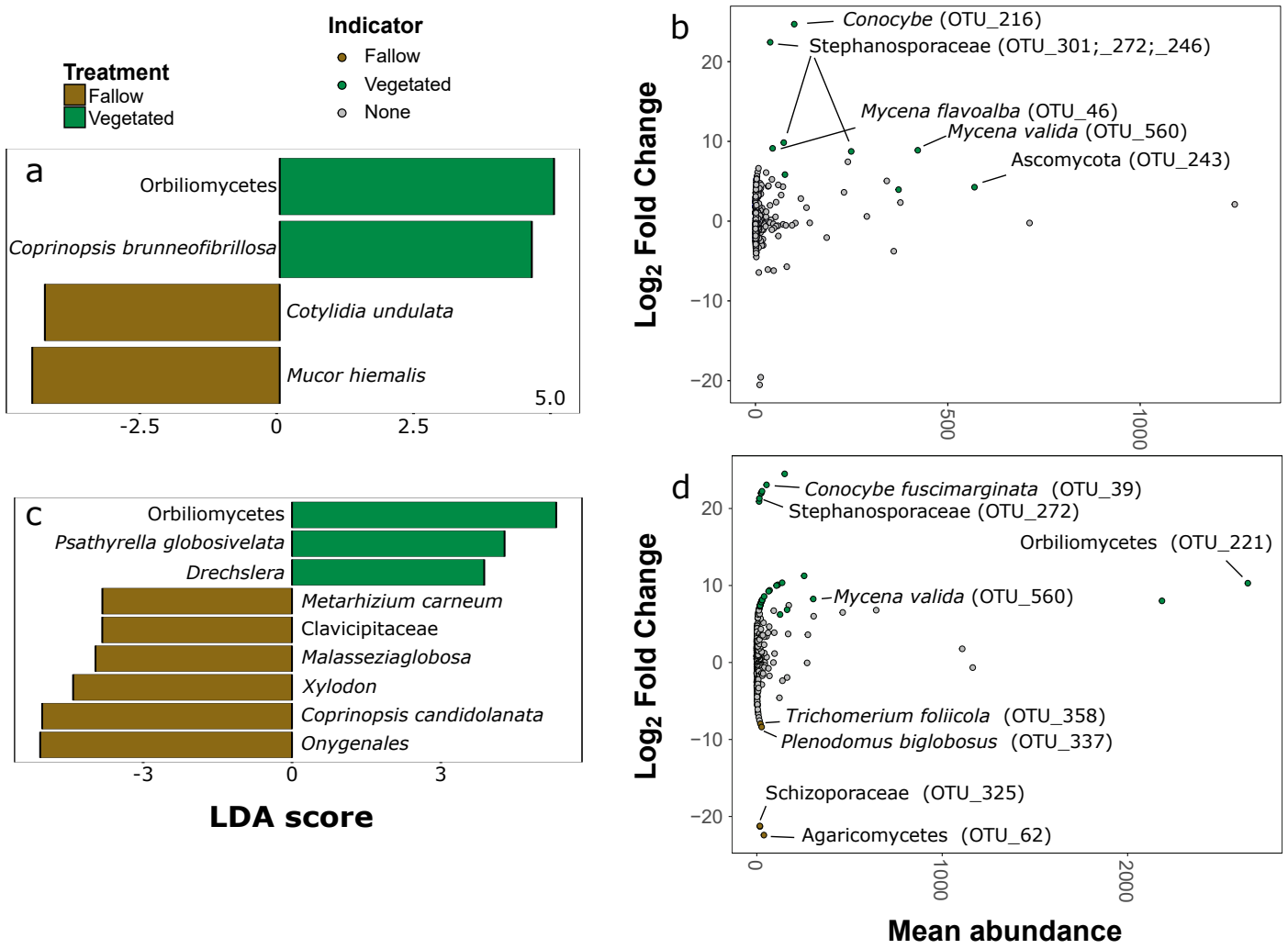


*Figure 7.6 Differential abundance of fungal taxa and OTUs between covered and vegetated soils. Indicator taxa identified using linear discriminant analyses (LDA) for **a)** 1 year old soils and **c)** 10 year old soils. Only data with an LDA score greater than 2.5 are presented. OTUs indicative of treatment based on corrected P values from DESeq2 analysis identified in **b)** 1 year old and **d)** 10 year old soils. Labeled OTUs are discussed in the main text and/or Appendix 5.*

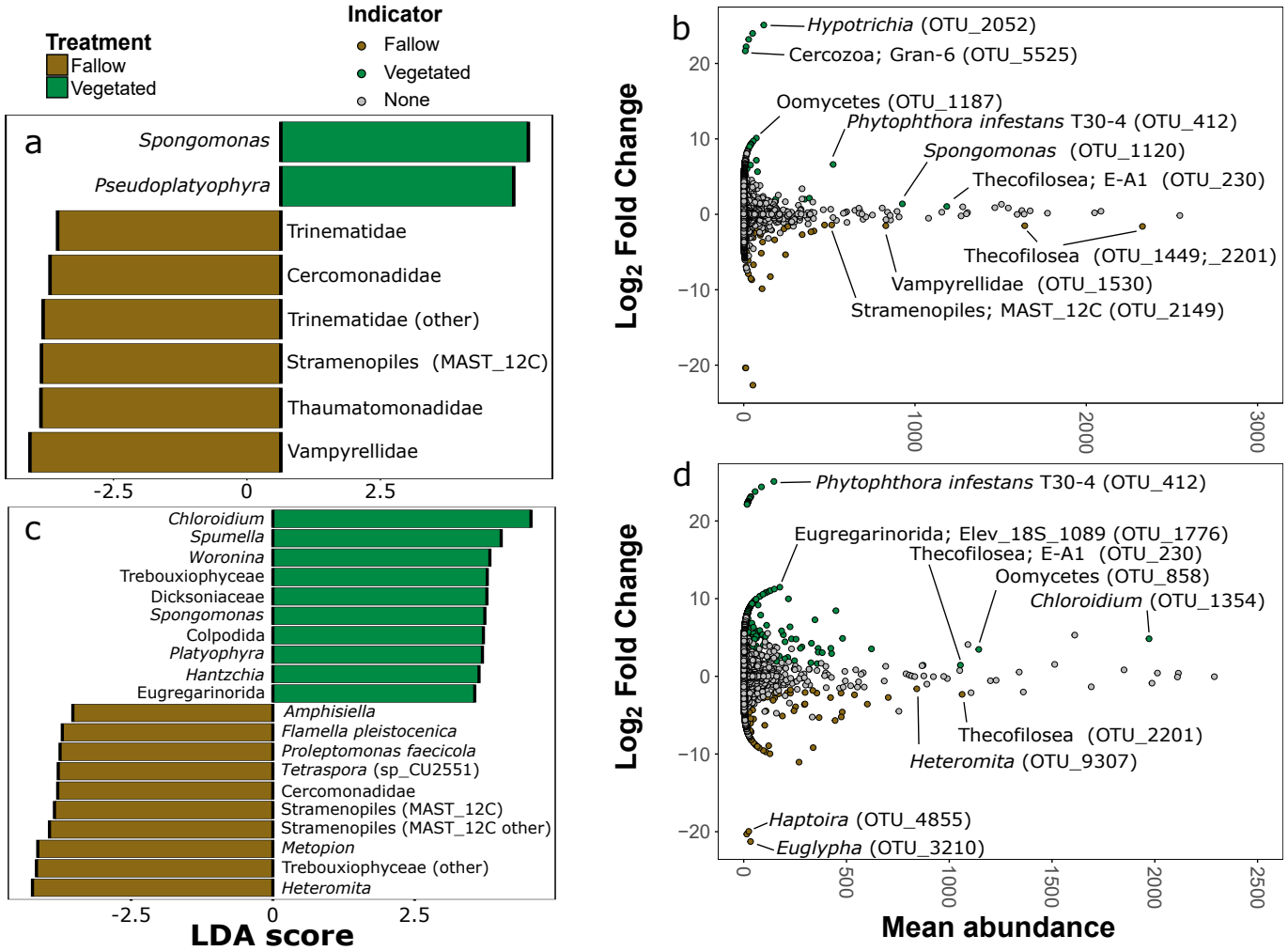


*Figure 7.7 Differential abundance of protistan taxa and OTUs between covered and vegetated soils. Indicator taxa identified using linear discriminant analyses (LDA) for **a**) 1 year old soils and **c**) 10 year old soils. Only data with an LDA score greater than 3.5 are presented. OTUs indicative of treatment based on corrected P values from DESeq2 analysis identified in **b**) 1 year old and **d**) 10 year old soils. Labeled OTUs are discussed in the main text and/or Appendix 5.*

7.3 Discussion

7.3.1 Effects of long-term fallow on soil structure and communities

We observed very few differences between vegetated and fallow soils after 1 year. There were no changes in microbial richness or H' values (Fig. 7.2). The only soil properties to exhibit treatment effects in 1 year-old soils were nitrate, moisture content, and Na (Table 7.1). The reduction of Na in fallow soils after 1 year was expected since Na is the first cation lost from cation exchange sites when increased leaching occurs upon removal of plants. The difference in nitrate may be due to capillary action drawing nitrate into the topsoil under moist conditions (29), which may have been influenced by increased moisture retention in the decomposing plant matter present in fallow soils (30). Though it is also possible that more ephemeral processes, such as localised urine patches from transient rodents, and incongruent climatic conditions between age-class sampling such as recent precipitation may have caused these changes. Any of these processes may also account for the marked difference in nitrate and ammonium concentrations between soil age classes.

Changes in soil properties in the 10 year-old soils aligned with those of other long-term bare fallow experiments. Notably, the soils were significantly more acidic and displayed an increase in acidic cations (e.g. Al^{3+}), in line with Paradelo et al. (20) in France. Total C and N displayed significant losses following C deprivation, as expected based on previous research at the Rothamsted Highfield experiment (26, 31), due to a loss of organic matter inputs from plant matter and therefore loss of soil C stock and N release.

Marked differences were identified between the structure in 10 year fallow soils versus vegetated soils with the former characterised by a reduced porosity, pore size and pore connectivity at the column scale (40 μm). However at high resolution (3 μm), the differences

were less pronounced especially for total porosity and pore connectivity, which was actually greater (i.e. more negative) under 10-year fallow conditions. These findings are consistent with those Bacq-Labreuil et al. (24), who described the pore structure of the Rothamsted Highfield soils. However, the trend toward increasing pore connectivity was unexpected since we expected a collapse in soil structure. Pore connectivity increases with increasing complexity of plant communities (32); our incongruent findings could be vestiges of macropores that have previously been lost due to consumption of labile organic matter lattices over time (33).

We observed an expected reduction in microbial biomass under both fallow timespans, similar to what has been observed at the Highfield Experiment (25, 34) and elsewhere (35, 36). However there were interesting shifts within the studied PLFA fractions between lengths of fallow. We detected a shift in greater proportional abundance of total bacterial biomass to 10 year-old fallow soils, driven by relative increases in Gram-positive bacteria and Actinomycetes with a corresponding loss in fungal PLFA markers (Table 7.2). This shift may indicate that fungi were detrimentally affected by a loss of plants (hosts to symbiotic and parasitic species) and not able to effectively compete with bacteria under long-term fallow. The relative increase in proportion of Actinomycetes strengthens this assumption as these bacteria form hyphae-like colonies and many species feed on cellulose and lignin, much like fungi, and are important decomposers of fungal necromass (37). Furthermore, certain Actinomycetes, such as *Streptomyces*, which were present in our samples, are known to produce antifungal compounds (38). The loss in soil CO_2 flux and qCO_2 further demonstrate the loss of microbial biomass under fallow, even after only 1 year.

Yet, our results on the effect of long-term C deprivation on richness diverge from those of previous experiments. Carson et al. (39) suggest that bacterial diversity increases with

decreasing pore connectivity, speculating that such reduction in connectivity creates many pockets of localised distinct bacterial populations. Our data refutes this hypothesis. Hirsch et al. (25,26) posit that soil biodiversity is resilient to 50 years of fallow; whereas, we show that richness of bacteria, fungi, protists, and viruses all fell after 10 years of fallow. Such a reduction in species richness indicates collapse of all fractions of the soil biosphere in response to a loss of C inputs. This likely stems from a loss in host plants for pathogenic and symbiotrophic organisms under fallow (40, 41). We also observed a reduction in H' of prokaryotes in 10 year-old fallow soils; though this trend was not present in fungi or protists and could not be assessed for viruses. Diversity of eukaryotic organisms may be more resistant to fallow because these organisms have a greater adaptive capacity to utilise multiple C-sources (10) or the employment of resistant structures or simply greater motility to wait for ideal conditions to exploit increasingly distant food sources (28).

7.3.2 Changes in soil functional diversity

The genes that exhibited differentially greater abundance in fallow soils from GeoChip analyses are mostly involved in methanogenesis (Fig. 7.3) and the degradation of more recalcitrant C forms (Fig. 7.4). The greater proportion of certain methanogenesis genes is likely the result of an increase in methanogens in anaerobic habitats due to a loss of soil structure. Smaller pores, like those in our 10 year-old fallow soils, generally experience more prolonged anoxia (42). Similarly there was an increase in proportion of the methyl coenzyme M reductase (*mcrA*) gene in fallow soils, which is a marker gene for methanogens (43). It is however, important to note that other methanogenesis genes were expressed more strongly in vegetated soils. Such inconsistency may indicate that although there was an increase in anaerobic conditions under fallow, vegetated soils may provide suitable conditions for the

reduction of CO_2 to CH_4, the most common methanogenesis pathway (44). As the vegetated soils had stronger soil structure, it is also possible that an increase in water-filled pores (42) promoted methanogen communities by reducing local redox potential (45).

The proportionally greater abundance of certain genes involved in hemicellulose, pectin, and aromatic degradation suggests that the microbial community has been exploiting some of the recalcitrant plant biomass after 10 years of fallow. Plant cell walls are composed of hemicelluloses, pectins (46), and certain aromatic compounds (47). Such compounds are difficult to break down and increasingly scarce food source as in the absence of fresh plant biomass (10). In particular, these results show the absence of accessible low molecular weight C compounds, such as root exudates (48), under long-term fallow. Phytase also displayed greater log-fold prevalence in fallow soils (Fig. 7.4). This suggests that the microbial community has been forced to access poorly available P forms, specifically phytic acid (IP6), that is commonly found in soils (49) and in seeds (50). Together, the collective data suggests that the fallow communities have been forced to utilise the recalcitrant energy sources in the absence of fresh organic matter inputs.

7.3.3 Bioindicators

We identified many bioindicators of fallow and vegetated soils. As was expected, there were more bioindicator taxa in the 10 year-old soils than 1 year-old soils (Fig. 7.5-7). Indeed, there were no differentially abundant fungal OTUs indicative of fallow and only 2 taxa identified for either treatment by LDA from ITS1 data in 1 year-old soils and there were rather few bioindicators of fungi in 10 year-old soils as well (Fig. 7.6). This further demonstrates that fungi were negatively affected by fallow conditions, as shown in richness and biomass data. There was a strong congruence between indicator taxa in both soil age classes. Orbiliomycetes

was identified by LDA as differentially abundant in vegetated soils in both 1 and 10 year-old soils and by DESeq2 in 10 year-old vegetated soils (Fig 7.6). These fungi are commonly saprophytic or actively consume invertebrates (51) making their preference for vegetated soils expected. Other fungal indicators of vegetated soils included representatives of the Stephanosporaceae, *Conocybe*, and *Mycena,* which all prefer habitats rich in plant matter (52,53,54). The bioindicators of fallow soils had less certain affinities and cannot be investigated in detail or are genera that may be exploiting local nutrient hotspots such as ammonia for *Coprinopsis* (55) or isolated organic matter inputs as may be the case for *Trichomerium* (56).

The prokaryotic and protistan communities supported a large number of indicator taxa. Of these, *Rhodanobacter*, *Streptomyces,* were some of the strongest indicators of fallow. As discussed previously, *Streptomyces* and related species are possibly outcompeting fungi for increasingly scarce resources. Similarly the dominance of N-cycling prokaryotes such as *Rhodanobacter, Nitrosotalea devanterra*, and *Nistrososvibrio tenius* (57, 58, 59) is a strong indicator of the anaerobic environments present in fallow soils despite the absence of clear N-cycling gene data. The members of the Nakamurellaceae, *Flavobacterum*, and *Solirubrobacter* were all indicators of vegetated soils, which is expected given their common affinity for soil habitats (60, 61, 62). Interestingly, *Nitrospira* was also an indicator taxa of 10 year vegetated soils and are likely to be important in N cycling (59).

Protistan bioindicators of vegetated soils included plant pathogens, such as *Phytopthora* and other Oomycetes as well as Thecofilisean flagellates and the green algae *Choloroidium*. Interestingly, there were a surprising number of MAST-12 Stramenopiles that were indicators of both ages of fallow (Fig. 7.7); these sequences could come from previously

undescribed species (63). Other protistan fallow indicators included various Cermonads, and various other flagellates such as *Thecofilosea* and *Metopoin*, testate amoebae (*Euglypha*), and Haptorian ciliates. Another interesting indicator of 1 year-old fallow soils were Vampyrellid amoebae, which are predators of fungi (64), suggesting that there may be a shift in food webs deprived of C at this early stage.

However, the identification of bioindicators did not always align between methodologies. For example, *Rhadoplanes* was identified as differentially abundant in 1 year-old vegetated soils by LDA but as an indicator of 10 year-old fallow by DESeq2 analysis (Fig. 7.5a, d). The most likely explanation is that these taxa support a high level of diversity at the species- and strain-level that exploit different niches. This is supported by our identification of various *Rhodoplanes* and Trebouxiophyceae green algae in both fallow and vegetated soils, suggesting that different species are exploiting different niches. However, further validation using strain-level metagenomics is needed.

7.3.4 Conclusions

Taken together, our comprehensive analyses of microbial biomass, diversity, and functional genes challenge previously held assumptions about the interactions of soil biology and structure under long-term fallow. The overall reduction in diversity metrics after 10 years of C deprivation lends support to the notion that soil pores, especially those > 30 μm support the majority of microbial activity and biodiversity proposed by Kravchenko et al. (32). The relative increase in certain methanogenesis, recalcitrant C, and IP6 degrading genes suggests that microbial communities have exhausted all low molecular weight C resources and are using anaerobic energy strategies. However the mixed responses of methanogenesis indicates that confounding factors may also prevail; for future work, measurements of emissions of both CO_2

and CH_4 should be included along with metagenome assemblies to determine the identity of the organisms associated with methanogenesis genes. Both LDA and DESeq2 analyses highlighted a number of bioindicator taxa and OTUs. Future metagenomics work may also further elucidate the patterns described here by identifying these taxa at the strain-level. Nevertheless our results show that long-term C deprivation causes widespread shifts in soil communities and structure that are dominated by anaerobic prokaryotes and motile protists.

7.4 Materials and methods

7.4.1 Experimental design

An experimental trial was established at Bangor University's Henfaes Research Centre, Abergwyngregyn, UK (53.24°N, 4.02°W; EL: 12 m) commencing in 2005. At this time, six 9 m^2 plots were established on Eutric Cambisol soil; half of which were left open (vegetated) and half covered (fallow). Plots were established in a field previously used for sheep grazing, which was fenced to exclude grazers, and demarcated with plastic frames embedded 25 cm into the soil, with 5-8 cm protruding aboveground (Appendix 5 Fig. 6a). Within each frame used for fallow plots, 1 layer of thick black landscaping fabric were attached to prevent plant growth (Appendix 5 Fig. 6b). These replicates were left for 10 years with annual weeding of fallow plots and annual mowing of vegetated plots. In 2015, a further eight plots (4 control and 4 fallow) were established adjacent to the initial trial plots, within the same field (Appendix 5 Fig. 6a). A total of 10 subsamples for biological and physicochemical analyses were collected in spring 2015 (10 year) and 2016 (1 year) using 1 cm diameter soil cores at 10 cm depth within each replicate. These subsamples were pooled and homogenised for subsequent analyses.

7.4.2 Soil properties

Soil properties were analysed using standard operating protocols. Electrical conductivity (μS cm^{-1}) and pH data were collected using a 1:2.5 soil-to-distilled water extraction. Samples for total organic C (%) and N (%) analyses were air dried and milled to a < 0.2 mm powder and subsequently burned in a TruSpec® Analyzer (Leco Corp., St Joseph, MI). Both NH_4^+ and nitrate NO_3^- were calculated through colourimetric methods using a 1:5 potassium sulphate (K_2SO_4) extraction (65, 66). Available P was also analysed in this way but in a distilled water extraction (67). These nutrients were all reported as mg kg^{-1}. Cations (nmol kg^{-1}) including K, Na, Mg, and Al in addition to CEC were obtained in a 1:5 ammonium acetate ($NH_4CH_3CO_2$) extraction. Cation exchange capacity (mmol $NH_4^+ kg^{-1}$) was calculated using ammonium acetate colourimetric methodology (68), whereas individual cations were analysed using inductively coupled plasma analyses. With the exception of total cation analysis for 1-year-old plots, which was performed at the Centre for Ecology and Hydrology, Lancaster, all other soil analyses were conducted at Bangor University.

In situ CO_2 flux measurements were recorded using an automated LI-8150 multiplexer CO_2 flux system (LI-COR Inc., Lincoln, NE). Polyvinyl chloride collars were inserted ~5 cm into the soil to house 20.3 cm diameter dark chambers (LI-COR LI-8100-104) in each plot. Soil CO_2 flux was measured every 2 h (141 measurements total) using an automated infrared gas analyser (LI-COR LI-8100) attached to the multiplexer system for 7 days in June, 2015.

Soil structure was assessed by soil porosity analyses on 10 year-old soils only at the University of Nottingham's Hounsfield Facility. Undisturbed soil cores of 7 cm diameter and ~16 cm depth were scanned using a v|tome|x M 240 kV X-ray Computed Tomography (μCT) scanner (GE Sensing & Inspection Technologies GmbH, Wunstorf, Germany) with an electron

acceleration energy of 170 kV, current of 200 mA, and a resolution of 40 μm. A total of 2400 projection images were collected during each scan, which lasted 140 minutes. Reconstruction was performed using Datos|Rec software (GE Sensing and Inspection Technologies GmbH, Wunsdorf, Germany) and 2272 images were collected for each sample. The data was then subsampled to cuboid volumes, sized 40 mm x 40 mm x 1700 images to avoid any edge effects caused during sample collection or in the subsequent analysis. Pore space was separated from the surrounding soil matrix using the Li global automatic threshold algorithm (69) as described in Helliwell et al., (70). Total porosity, pore size distribution, mean pore size, total pore area and pore connectivity (expressed as the Euler number) was then analysed using ImageJ software (71). The coefficient of uniformity (a ratio of the pore size distribution expressed by d_{60}:d_{10}) (72) was calculated as a simple way of expressing the pore size distribution.

Following initial scanning, cores were also cut into three equal parts (top, middle, and bottom) and air-dried for two days. An aggregate of approximately 4 mm diameter was randomly selected from each dried sample, giving three aggregates per core representing each layer of the sample. These aggregates were then scanned using a Phoenix Nanotom 180NF scanner (GE Sensing and Inspection Technologies, Wunstorf, Germany) with an electron acceleration energy of 90 kV, current of 70 mA, a resolution of 3 μm and with 1800 projection images being collected over 133 minutes. For analysis, scans were subsampled to a 1.4 mm x 1.4 mm x 600 slice cube from the centre of each aggregate. As with the whole column scans, pore space was identified using the Li global automatic threshold algorithm (69) and the same measurements undertaken.

7.4.3 Phospholipid fatty acid analysis

Phospholipid fatty acid analyses (PLFA) were performed by Microbial-ID Inc. (Newark, DE, USA). In total, 10 g from each sample were freeze-dried and sent for analysis. Data generated included: total PLFA (nmol g^{-1}), total bacteria and fungi, as well as Actinomycetes, anaerobes, Gram-positive, and Gram-negative bacteria. The microbial metabolic quotient (qCO_2) was derived from total PLFA and soil CO_2 flux measurements (reported as μmol $CO_2 \_ m^{2\ -s1}$ / nmol g^{-1}).

7.4.4 DNA extraction and sequencing

DNA was extracted from 0.25 g of soil per each homogenised sample using PowerLyzer PowerSoil DNA Isolation Kits (MO-BIO). Amplicon libraries were created and processed through a two-step library preparation protocol followed by Illumina DNA sequencing at the Centre for Genome Research, University of Liverpool. The primer combinations used for the first round were 515F/806R (V4 16S) for 16S libraries (73), ITS5/5.8S_fungi for ITS1 libraries (74), and TAReuk454FWD1/TAReukREV3 (75) for 18S libraries following George et al. (76). Illumina Nextera XT 384-way indexing primers were then used on the first-round products to create final sequences, which were purified with AMPure XP beads (Beckman Coulter). Sequences have been uploaded to the European Nucleotide Archive (Primary Accession: PRJEB33898).

All bioinformatics were performed on the Supercomputing Wales system following George et al. (76). Taxonomy was assigned using QIIME 1.9.1 (77) with RDP methodology (78) using the GreenGenes v. 13_8 (79), UNITE v. 7.2 (80), and SILVA 128 (81) databases for the 16S, ITS1, and 18S OTU tables, respectively. Singletons and OTUs appearing in only 1 sample were removed. Eukaryotic OTUs were removed from the 16S OTU table, as were non-

fungi and non-protist OTUs, which were removed from the ITS1 and 18S tables, respectively. Samples were rarefied 100 times for each OTU table using phyloseq (82) and the rounded mean used for all analyses. Read depths in rarefied OTU tables were: 19,311 for prokaryotes, 3,243 for fungi, and 69,908 for protists.

7.4.5 GeoChip analyses

The 10 year-old soils were selected for functional gene analyses. DNA from these plots was sent to Glomics Inc. (Norman, OK, USA) for GeoChip analysis. Briefly, DNA extracts were fluorescently labelled hybridized to GeoChip4.0 microarrays in 50 μL solutions. Non-binding DNA was washed away after 16 h. Average signal intensity was measured and standardised to remove background signal (signal-to-noise ratio < 2). We used these data to identify genes involved in C degradation and anaerobic respiration, including N, S, and P cycling.

7.4.6 Statistical analyses

Within each age class, soil properties, PLFA ratios, OTU richness, and Shannon-Weiner diversity index (H') values were compared for differences between control and blackout treatments using two-tailed student's T-tests in R 3.3.3 (83). Pore morphology measurements were assessed by a one-tailed student's T-test, also in R. We looked for microbial indicator taxa using linear discriminant analyses (LDAs) following the LDA Effect Size (LEFSe) (84) method in Galaxy (85), with default parameters. Similarly, OTU-level indicators were identified using differential abundance analyses from the R package DESeq2 (86) on unrarefied OTU tables. For the GeoChip data, fold-changes in signal frequency for genes of interest were identified and subsequently log transformed. Again, student's T-tests with Bonferroni correction were used to compare differences between treatments.

7.5 Acknowledgements

PBLG and DBF are supported by Soils Training and Research Studentship (STARS) grants from the Biotechnology and Biological Sciences Research Council and NERC [Grant number NE/M009106/1]. STARS is a consortium consisting of Bangor University, British Geological Survey, Centre for Ecology and Hydrology (CEH), Cranfield University, James Hutton Institute, Lancaster University, Rothamsted Research and the University of Nottingham. JAA and SJM are supported by the Biotechnology and Biological Sciences Research Council Designing Future Wheat Cross-Institute Strategic Programme [grant number BB/P016855/1]. We thank the staff of Environment Centre Wales and the Centre for Genomic Research, University of Liverpool for their contributions to DNA extraction and sequencing. We thank Gloria dos Santos Pereira at CEH Lancaster for assistance with cation analyses. We acknowledge the technical expertise in image processing of Dr. Brian Atkinson from the Hounsfield Facility at the University Of Nottingham.

7.6 References

1. Lal R. 2004. Soil carbon sequestration impacts on global climate change and food security. Science 304:1623-1627. https://doi.org/10.1126/science.1097396
2. Davidson EA, Jenssens IA. 2006. Temperature sensitivity of soil carbon decomposition and feebacks to climate change. Nature 440:165-173. https://doi.org/10.1038/nature04514
3. Friedlingstein P, Cox P, Betts R, Bopp L, von Bloh W, Brovkin V, Cadule P, Doney S, Eby M, Fung I, Bala G, John J, Jones C, Joos F, Kato T, Kawamiya M, Knorr W, Lindsay K, Matthews HD, Raddatz T, Rayenr P, Reick C, Roeckner E, Schnitzler K-G, Schnur R, Strassmann K, Weaver AJ, Yoshiawa C, Zeng N. 2006. Climate-carbon cycle feedback analysis: results from the C^4MIP model intercomparison. J Clim 19:3337-3353. https://doi.org/10.1175/JCL13800.1
4. Crowther TW, Todd-Brown KEO, Rowe CW, Wieder WR, Carey JC, Machmuller MB, Snoek LB, Fang S, Zhou G, Allison SD, Blair JM, Bridgham SD, Burton AJ, Carrillo Y, Reich PB, Clark JS, Classen AT, Dijkstra FA, Elberling B, Emmett B,

Estiarte M, Frey SD, Guo J, Harte J, Jiang L, Johnson BR, Kröel-Dulay G, Larsen KS, Laudon H, Lavallee JM, Luo Y, Lupascu M, Ma LN, Marhan S, Michelsen A, Mohan J, Niu S, Pendall E, Peñuelas J, Pfeifer-Meister L, Poll C, Reinsch S, Reynolds LL, Schmidt LK, Sistla S, Sokol NW, Templer PH, Treseder KK, Welker JM, Bradford MA. 2006. Quantifying global soil C losses in response to warming. Nature 540:104-108. https://doi.org/10.1038/nature20150

5. Carey JC, Tang J, Templer PH, Kroeger KD, Crowther TW, Burton AJ, Dukes JS, Emmett B, Frey SD, Heskel MA, Jiang L, Machmuller MB, Mohan J, Panetta AM, Reich PB, Reinsch S, Wang X, Allison SD, Bamminger C, Bridgham S, Collins SL, de Dato, G, Eddy WC, Enquist BJ, Estiarte M, Harte J, Henderson A, Johnson BR, Larsen KS, Luo Y, Marhan S, Melillo JM, Peñuelas J, Pfeifer-Meister L, Poll C, Rastetter E, Reinmann AB, Reynolds LL, Schmidt IK, Shaver GR, Strong AL, Suseela V, Tietema A. 2016. Temperature response of soil respiration largely unaltered with experimental warming. Proc Natl Acad Sci USA 113:13797-13802. https://doi.org/10.1073/pnas.1605365113
6. Luo Y, Wan S, Hui D, Wallace LL. 2001. Acclimatization of soil respiration to warming in a tall grass prairie. Nature 413:622-625. https://doi.org/10.1038/35098065
7. Melillo JM, Steudler PA, Aber JD, Newkirk K, Lux H, Bowles FP, Catricala C, Magill A, Ahrens T, Morrisseau S. 2002. Soil warming and carbon-cycle feedbacks to the climate system. Science 298:2173-2175. https://doi.org/10.1126/science.1074153
8. Rovira P, Vallejo VR. 2002. Labile and recalcitrant pools of carbon and nitrogen in organic matter decomposing at different depths in soil: an acid hydrolysis approach. Geoderma 107:109-141. https://doi.org/10.1016/S0016-7061(01)00143-4
9. Cheng L, Leavitt SW,Kimball BA, Pinter PJ Jr, Ottman MJ, Matthias A, Wall GW, Brooks T, Williams DG, Thompson TL. 2007. Dynamics of labile and recalcitrant soil carbon pools in sorghum free-air CO_2 enrichment (FACE_ agroecosystem. Soil Biol Biochem 39:2250-2263. https://doi.org/10.1016/j.soilbio.2007.03.031
10. Adl, SM. 2003. The Ecology of Soil Decomposition. CABI Publishing, Oxford, UK.
11. Romero-Olivares AL, Allison SD, Treseder KK, Hui D. 2017. Decomposition of recalcitrant carbon under experimental warming in boreal forest. PLoS One 12:e0179674. https://doi.org/10.1371/jounral.pone.0179674
12. Chen S, Huang Y, Zou J, Shi Y. 2013. Mean residence time of global topsoil organic carbon depends on temperature, precipitation and soil nitrogen. Global Planet Change 100:99-108. https://doi.org/10.1016/j.gloplacha.2012.10.006
13. 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. http://doi.org/10.1128/AEM.003355-09
14. Tedersoo L, Bahram M, Põlme S, Kõljalg U, Yorou NS, Wijesundera R, Ruiz LV, Vasco-Palacios AM, Thu PQ, Suija A, Smith ME, Sharp C, Saluveer E, Saitta A, Rosas M, Riit T, Ratkowsky D, Pritsch K, Põldmaa K, Piepenbring M, Phosri C, Peterson M,

Parts K, Pärtel K, Otsing E, Nouhra E, Njouonkou AL, Nilsson RH, Morgado LN, Mayor J, May TW, Majuakim L, Lodge DJ, Lee SS, Larsson K-H, Kohout P, Hosaka K, Hiiesalu I, Henkel TW, Harend H, Guo L-D, Greslebin A, Grelet G, Geml J, Gates G, Dunstan W, Dunk C, Drenkham R, Dearnaley J, De Kesel A, Dang T, Chen X, Buegger F, Brearkey FQ, Bontio G, Anslan S, Abell S, Abarenkov K. 2014. Global diversity and geography of soil fungi. Science 346:125668. https://doi.org/10.1126/science.1256688

15. Lovley DR, Coates JD. 2000. Novel forms of anaerobic respiration of environmental relevance. Curr Opin Microbiol 3:252-256. https://doi.org/10.1016/51369-5274(00)00085-0

16. Falkowski PG, Fenchel T, Delong EF. 2008. The microbial engines that drive Earth's biogeochemical cycles. Science 320:1034-1039. https://doi.org/10.1126/science.1153213

17. Oertel C, Matschullat J, Zurba K, Zimmermann F, Erasmi S. 2016. Greenhouse gas emissions from soils – a review. Geochemistry 76:327-352. https://doi.org/10.1016/j.chemer.2016.04.002

18. IPCC 2014. Climate Change 2014: Synethesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmentla Panel on Climate Change. IPCC, Geneva, Switzerland

19. Bellamy PH, Loveland PJ, Bradley RI, Murray LR, Kirk GJD. 2005. Carbon losses from all soils across England and Wales 1978-2003. Nature 437:245-248. https://doi.org/10.1038/nature/04038

20. Paradelo R, van Oort F, Chenu C. 2013. Water-dispersible clay in bare fallow soils after 80 years of continuous fertilizer addition. Geoderma 200-201:40-44 https://doi.org/10.1016/j.feoderma.2013.01.014

21. Paradelo R, van Oort F, Billiou D, Chenu C. 2016. Soil organic matter stabilization at the pluri-decadal scale: insight from bare fallow soils with contrasting physicochemical properties and macrostructures. Geoderma 275:48-54. https://doi.org/10.1016/j.geoderma.2016.04.009

22. Coleman K, Jenkinson DS, Crocker GJ, Grace PR, Klír J, Körschens M, Poulton PR, Richter DD. 1997. Geoderma 81:29-44. https://doi.org/10.1016/S0016-7061(97)00079-7

23. Barré P, Eglin T, Christensen BT, Ciais P, Houot S, Kätterer T, van Oort F, Peylin P, Poulton PR, Romanenkov V, Chenu C. 2010. Quntifying and isolating stable soil organic carbon using long-term bare fallow experiments. Biogeosciences 7:3839-3850. https://doi.org/10.5194/bg-7-3839-2010

24. Bacq-Labreuil A, Crawford J, Mooney SJ, Neal AL, Akkari E, McAuliffe C, Zhang X, Redmile-Gordon M, Ritz K. 2018. Effects of cropping systems upon the three-dimensional architecture of soil systems are modulated by texture. Geoderma 332:73-83. https://doi.org/10.1016/j.geoderma.2018.07.002

25. Hirsch PR, Gilliam LM, Sohi SP, Williams JK, Clark IM, Murray PJ. 2009. Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. Soil Biol Biochem 41:2021-2024. https://doi.org/10.1016/j.soilbio.2009.07.011

26. Hirsch PR, Jhurreea D, Williams JK, Murray PJ, Scott T, Misselbrook TH, Goulding KWT, Clark IM. 2016. Soil resilience and recovery: rapid community responses to management changes. Plant Soil 412:283-297. https://doi.org/10.1007/s11104-016-3068-x

27. Fierer N, Breitbart M, Nulton J, Salamon P, Lozupone C, Jones R, Robeson M, Edwards RA, Felts B, Rayhawk S, Knight R, Rohwer F, Jackson RB. 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl Environ Microbiol 73:7059-7066. https://doi.org/10.1128/AEM.00358-07

28. Geisen S, Mitchell EAD, Adl SM, Bonkowski M, Dunthorn M, Ekelund F, Fernández LD, Jousset A, Krashevska V, Singer D, Spiegel FW, Walochnik J, Lara E. 2018. FEMS Microbiol Ecol 42:293-323. https://doi.org/10.1093/femsre/fuy006

29. Simpson JR. The mechanism of surface nitrate accumulation on a barefallow soil in Uganda. J Soil Sci 11:45-60. https://doi.org/10.1111/j.1365-2389.1960.tb02200.x

30. McGuire AM, Bryant DC, Denison RF. Wheat yields, nitrogen uptake, and soil moisture following winter legume cover crop vs fallow. Agron J 90:404-410. https://doi.org/10.2134/agronj1998.00021962009000030015x

31. Gregory AS, Dungait AJ, Watts W, Bol R, Dixon ER, White RP, Whitmore AP. 2016. Long-term management changes topsoil and subsoil organic carbon and nitrogen dynamics in a temperate sagricultural system. Eur J Soil Biol 67:421-430. https://doi.org/10.1111/ejss.12359

32. Kravchenko AN, Guber AK, Razavi BS, Koestel J, Quigley MY, Robertson GP, Kuzyakov Y. 2019. Microbil spatial footprint as a driver of soil carbon stabilization. Nat Commun 10:3121. https://doi.org/10.1038/s41467-019-11057-4

33. Xie W-L, Li P, Zhang M-S, Cheng T-E, Wany Y. 2018. Collapse behaviour and microstructural evolution of loess soils from the Loess Plateau of China. J Mt Sci 15:1642-1657. https://doi.org/10.1007/s11629-018-50062

34. Wu Y, Kemmitt, S, White RP, Jianming X, Brookes PC. 2012. Carbon dynamics in a 60 year fallowed loamy-sand soil compared to that in a 60 year permanent arable or permanent grassland UK soil. Plant Soil 352:51-63. https://doi.org/10.1007/s11104-011-0979-4

35. Steenworth KL, Jackson LE, Calderon FJ, Stromberg MR, Scow KM. 2002. Soil microbial community composition and land use history in cultivated and grassland ecosystems of coastal California. Soil Biol Biochem 34:1599-1611. https://doi.org/10.1016/S0038-0717(02)00144-X

36. Asuming-Brempong S, Gartner S, Adiku SGK, Archer G, Edusei V, Tiedje JM. 2008. Changes in the biodiversity of microbial populations in tropical soils under different fallow treatments. Soil Biol Biochem 40:2811-2818. https://doi.org/10.1016/j.soilbio.2008.08.010

37. de Boer W, Folam LB, Summerbell RC, Boddy L. 2005. Living in a fungal world: impact of fungi on soil bacterial niche development. FEMS Microbiol Rev 29:795-811. https://doi.org/10.1016/femsre.2004.11.005

38. Doumbou CL, Hamby Salove, MK, Crawford DL, Beaulieu C. 2001. Actinomycetes, promising tools to control plant diseases and to promote plant growth. Phytoprotection 82:85-101. https://doi.org/10.72027.706219ar

39. Carson JK, Gonzalez-Quiñones V, Murphy DV, Hinz C, Shaw JA, Gleeson DB. 2010. Low pore connectivity increases bacterial diversity in soil, Appl Environ Microbiol 76:3936-3942. https://doi.org/10.1128/AEM.03085-09

40. Hulugalle NR, Entwistle PC, Cooper JL, Allen SJ, Nehl DB. 1998. Effect of long-fallow on soil quality and cotton lint yield in an irrigated self-mulching, grey Vertosol in the central-west of New South Wales. Aust J Soil Res 36:621-639. https://doi.org/10.1071/597111

41. Pankhurst CE, Stirling GR, Magarey RC, Blair BL, Holt JA, Bell MJ, Garside AL. 2005. Quantification of the effects of rotation breaks on soil biological properties and their impacts on yield decline in sugarcane. Soil Biol Biochem 37:1121-1130. https://doi.org/10.1016/j.soilbio.2004.11.011

42. Keiliweit M, Wanzek T, Kleber M, Nico P, Fendorf S. 2017. Anaerobic microsites have an unaccounted role in soil carbon stabilization. Nat Commun 8:1771. https://doi.org/10.1038/s41467-017-01406-6

43. Morris R, Schauer-Gimenez A, Bhattad U, Kearney C, Struble CA, Zitomer D, Maki JS. 2014. Methyl coenzyme M reductase (*mcrA*) gene abundance correlates with activity measurements of methanogenic H_2/CO_2-enriched anaerobic biomass. Microb Biotechnol 7:77-84. https://doi.org/10.1111/1751-7915.12094

44. Jones WJ, Donnelly MI, Wolfe RS. 1985. Evidence of a common pathway of carbon dioxide reduction to methane in methanogens. J Bacteriol 163:126-131.

45. Dalal RC, Allen DE, Livesley SJ, Richards G. 2008. Magnitude and biophysical regulators of methane emission and consumption in the Australian agricultural, forest, and submerged landscapes: a review. Plant Soil 309:43-76. https://doi.org/10.1007/s11104-007-9446-7

46. Ogden M, Hoefgen R, Roessner U, Persson S, Khan G. 2018. Feeding the walls: how does nutrient availability regulate cell wall composition? Int J Mol Sci 19:2691. https://doi.org/10.3390/ijms19092691

47. Akin DE. 2008. Plant cell wall aromatics: influence on degradation of biomass. Biofuel Bioprod Bior 2:288-303. https://doi.org/10.1002/bbb.76

48. Renella G, Egamberdiyeva D, Landi L, Mench M, Nannipieri P. 2005. Microbial activity and hydrolase activities during decomposition of root exudates released by an artificial root surface in Cd-contaminated soils. Soil Biol Biochem 38:702-708. https://doi.org/10.1016/j.soilbio.2005.06.021

49. Richardson AE, Hadobas PA. 1997. Soil isolates of *Pseudomonas* spp. That utilize inositol phosphates. Can J Microbiol 43:509-516. https://doi.org/10.1139/m97-073

50. Schlemmer U, Frølich W, Prieto RM, Grases F. 2009. Phytate in foods and significance for humans. Mol Nutr Food Res 53:S330-S375. https://doi.org/10.1002/mnfr.200900099

51. Pfister DH, 2015. 2 Pezizomcotina: Pezizomycetes, Orbiliomycetes, p 35-55. *In* McLaughlin DJ, Spatafora JW (ed), The Mycota VII Part B: Systematics and Evolution. Springer-Verlag, Berlin, Heidelberg.

52. Hallen HE, Watling R, Adams GC. 2003. Taxonomy and toxicity of *Conocybe lactea* and related species. Mycol Res 107:969-979. https://doi.org/10.1017/S0953756203008190

53. Hunt J, Boddy L, Randerson PF, Rogers HJ. 2004. An evaluation of 18S rDNA approaches for the study of fungal diversity in grassland soils. Microb Ecol 47:385-395. https://doi.org/10.1007/s00248-003-2018-3

54. Cannon PF, Kirk PM. 2007. Fungal Families of the World. CAB International, Wallingford, UK

55. Raut JK, Suzuki A, FukiharuT, Shimizu K, Kawamoto S, Tanaka C. 2011. *Coprinopsis neophlyctidospora* sp. no., a new ammonia fungus from boreal forests in Canada. Mycotaxon 115:227-238. https://doi.org/10.5248/115.227

56. Chonmunti P, Bhat DJ, Jones EBG, Chukeatirote E, Bahkali AH, Hyde KD. 2012. Trichomeriaceae, a new sooty mould family of Chaetothyriales. Fung Divers 56:63-76. https://doi.org/10.1007/s13225-012-0197-2

57. Kostka JE, Green SJ, Rishishwar L, Prakash O, Katz LS, Mariño-Ramírez L, Jordan IK, Munk C, Ivanova N, Mikhailova N, Watson DB, Brown SD, Palumbo AV, Brooks SC. 2012. Genome sequences for six *Rhodanobacter* strains, isolated from soil and the terrestrial subsurface, with variable denirification capabilities. J Bacteriol 194:4461-4462. https://doi.org/10.1128/JB.00871-12

58. Norman JS, Lin L, Barrett JE. 2015. Paired carbon and nitrogen metabolism by ammonia-oxidizing bacteria and archaea in temperate forest soils. Ecosphere 6:art176. https://doi.org/10.1890/ES14-00299.1

59. Shah V, Jones J, Dickman J, Greenman S. 2014. Response of soil bacterial community to metal nanoparticles in biosolids. J Hazard Mater 274:399-403. https://doi.org/10.1016/j.hazmat.2014.04.003

60. Bernardet J-F, Segers P, Vancanneyt M, Berthe F, Kersters K, Vandamme P. 1996. Cutting a Gordian Knot: emended classification and description of the genus *Flavobacterium*, emended description of the gamily Flavobacteriaceae, and proposal of

Flavobacterium hydatis no. nov. (Basonym, *Cytophaga aquatilis* Strohl and Tait 1978). Int J Syst Bacteriol 46:128-148. https://doi.org/10.1099/00207713-46-1-128

61. Kim KK, Lee J-S. 2014. The Family Nakamurellaceae p 577-585. *In* Rosenberg E, DeLong EF, Lor S, Stachebrandt E, Thompson F (ed). The Protaryotes: Actinobacteria fourth ed. Springer, Heidelberg, Germany.

62. Whitman WB. 2015. *Solirubribacter*. *In* Whitma WB, deVos P, Dedysh S, Hedlund B, Kämpfler P, Rainey F, Trujillo ME, Bowman DR, Glöckner FO, Oren A, Paster BJ, Wade W, Ward N, Busse H-J, Reysenbach A-L (ed). Bergey's Manual ® of Systematics Archaea and Bacteria. Springer, New York.

63. Massana R, del Campo J, Sieracki ME, Audic S, Logares R. 2014. Exploring the uncultured microeukaryote majority in the oceans: reevaluation of ribogroups within stramenopiles. ISME J 8:854-866. https://doi.org/10.1038/ismej.2013.204

64. Neher DA. 1995. Biological diversity in soils of agricultural and natural ecosystems p 55-72. *In* Olson RK, Francis A, Kaffka S (ed), Exploring the role of diversity in sustainable agriculture. American Society of Agronomy, Madison, WI.

65. Mulvaney RL. 1996. Nitrogen – inorganic forms p 1123-1184. *In* Sparks DL (ed). Methods of Soil Analysis. Part 3. Soil Science Society of America, Madison, WI.

66. Miranda KM, Epsey MG, Wink, DA. 2001. A rapid, simple, spectrophotometric method for simultaneous detection of nitrate and nitrite. Nitric Oxide 5:62-71. http://doi.org/10.1006/niox.2000.0319

67. Murphy J, Riley JP. 1967. A modified single solution method for the determination of phosphate in natural waters. Anal Chim Acta 27:31-36. http://doi.org/10.1016/S0003-2670(00)88444-5

68. Summer ME, Miller WP. 1996. Cation Exchange Capacity Coefficients and Exchange Coefficients. *In* Sparks DL (ed). Methods of Soil Analysis. Part 3. Soil Science Society of America, Madison, WI.

69. Li CH, Tam, PKS. 1998. An iterative algorithm for minimum cross entropy thresholding. Pattern Recogm Lett 19:771-776. https://doi.org/10.1016/S0167-8655(98)00057-9

70. Helliwell JR, Miller AJ, Whailey WR, Mooney SR, Sturrock CJ. 2014. Quantifying the impact of microbes on structural development and behaviour in wet soils. Soil Biol Biochem 74:138-147. https://doi.org/10.1016/j.soilbio.2014.03.009

71. Schneider CA, Rasband WS, Eliceiri KW. 2012. NIH Image to ImageJ: 25 years of image analysis. Nat Methods 9:671-675. https://doi.org/10.1038/nmeth.2089

72. Kézdi Á. 1974. Handbook of Soil Mechanics Vol 1: Soil Physics. Elsevier Scientific Publishing, London, UK.

73. Caporaso JG, Laubr CL, Walters WA, Berg-Lyons CA, Turnbaugh PJ, Fierer N, Knight R. 2011. Global patterns of 16S rRNA diversity at a depth of meillions of sequences per sample. Proc Natl Acad Sci USA 108:(suppl. 1) 4516-4522. https://doi.org/10.1073/pnas.1000080107

74. Epp LS, Boessenkool S, Bellemain EP, Halle J, Riaz T, Ersèus C, Gusarov VI, Edwards ME, Johnsen A, Stenøien HK, Hassel K, Kauserud H, Yoccoz NG, Bråthen KA, Willerslev, Taberlet P, Coissac E, Brochmann C. 2012. New environmental metabarcodes for analysing soil DNA: potential for studying past and present ecosytems. Mol Ecol 21:1821-1833. https://doi.org/10.1111/j.1365-294X.2012.05537.x

75. Behnke A, Engel M, Christen R, Nebel M, Klein RR, Stoeck T. 2011. Depicting more accurate pictures of protistan community complexity using pyrosequencing of hypervariabale SSU rRNA gene regions: protistan community complexity. Environ Microbiol 13:340-349. https://doi.org/10.1111/j.1462-2920.2010.02332.x

76. George PBL, Lallias D, Creer S, Seaton FM, Kenny JG, Eccles RM, Griffiths RI, Lebron I, Emmett BA, Robinson DA, Jones DL. 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. Nat Commun 10:1107. https://doi.org/10.1038/s41467-019-09031-1

77. 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. https://doi.org/10.1038/nmeth.f.303

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

79. 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. https://doi.org/10.1128/AEM.03006-05

80. Kõljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme, J, Callaghan TM, Douglas B, Drenken T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Põldmaa K, Saag I, Schüβler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of fungi. Mol Ecol 22:5271-5277. https://doi.org/10.1111/mec.12481

81. Quast C, Pruesse E, Yilmaz P, Gerken 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:D590-D596. https://doi.org/10.1093/nar/gks1219McMurdie PJ, Holmes S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One 8:e61217. https://doi.org/10.1371/journal.pone.0061217

82. R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria https://www.R-project.org/

83. Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. 2011. Metagenomic biomarker discovery and explanation. Genome Biol 12:R60. https://doi.org/10.1186/gb-2011-12-6-r60

84. Afgan E, Baker D, van den Beek M, Blankenberg D, Bouvier D, Čech M, Chilton J, Clements D, Coraor N, Eberhard C, Grüning B, Guerler A, Hillman-Jackson J, Von Kuster G, Rasche E, Soranzo N, Nitesh T, Taylor J, Nekrutenko A, Goecks J. 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. Nucleic Acids Res 44:W3-W10. https://doi.org/10.1093/nar/gkw343

85. Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8

Chapter 8: Discussion

8.1 Introduction

This chapter reviews the data presented in Chapters 3-7, summarising the main findings and discussing them in the context of both this thesis and the wider literature. Detailed discussions of the results from each study are presented in their respective chapters. Summaries of the strengths and weaknesses of the techniques and experimental approaches are also presented. Finally, an outline of future research goals that build on the themes of this thesis is described.

8.2 General synthesis

The work presented in this thesis aims to determine i) shifts in belowground diversity and community structure and determine how they relate to both soil physical properties and heterogeneous land uses at the national-scale; ii) potential effects of changing land uses on the provision of ecosystem services, *in situ*; and iii) the fate of soil communities under long-term stress and the consequences for nutrient cycling. Data generated through GMEP was used to meet the first two objectives whilst data from a long-term C deprivation experiment was used to answer objective iii.

The results of Chapters 3 and 4 showed how all fractions of the belowground community, bar viruses, are distributed across Wales. Trends in mesofauna presented in Chapter 3 include a reduction in abundance and diversity of all mesofauna groups, except Mesostigmatid mites, in arable land and a reduction of overall diversity in addition to Collembola abundance in heath and bog sites. There was also a clear reduction in abundance and diversity in peat soils. These findings are reflective of previous work across Europe (Arroyo *et al*., 2013; Tsiafouli *et al*., 2015) and highlight the potential use of mesofauna as

environmental indicators in the assessment of agri-environment schemes at the national-scale. The morphological diversity findings also represent a ground truth for trends observed by eDNA data presented in Chapter 4. Such methodological congruence is encouraging and shows that a combined morphological and eDNA approach is ideal for assessing such a large amount of biodiversity (as observed for soil fauna) at the national-scale (Will *et al*., 2005).

Further, the eDNA analysis presented in Chapter 4 revealed an overarching trend in soil microbes whereby bacteria, fungi, and protists decline in richness from high productivity arable and grassland habitats to low productivity upland moorland, heath, and bogs. Archaea followed a converse trend, with richness increasing with falling productivity. These trends may be evidence of relic/dormant populations (Tsiafouli *et al*., 2015; Geisen *et al*., 2018), the intermediate disturbance hypothesis theory (Cornell, 1978) in soils, or other competitive processes, which require further investigation. In addition, this work confirms the importance of pH in governing microbial α- and β-diversity in soils (Griffiths *et al*., 2011; Dupont *et al*., 2016; Oton *et al*., 2016; Delgado-Bacquerizo *et al*., 2018). Other edaphic properties, such as C:N ratio, were also identified as important factors, as previously shown for bacteria (Griffiths *et al*., 2011), fungi (Tedersoo *et al*., 2014), and protists (Dupont *et al*., 2016). Since these factors are closely related and often easily manipulated through agricultural management (Subbaro *et al*., 2012; Bai *et al*., 2018), further disentanglement was not possible based on the analyses presented.

In combination, these results represent one of the most comprehensive assessments of soil biodiversity and confirm that it is strongly influenced by both anthropogenic land use and edaphic physicochemical factors. However, the relationships between soil biota and physicochemical properties explored here are only cursory. Targeted analyses of these data

have revealed detailed interactions between soil biota and their environment. Current work shows a novel differential effect of soil texture on bacterial and fungal richness, with increasing heterogeneity positively influencing bacterial richness (Seaton *et al*., *In submission*). This is in agreement with current understanding of microbial dynamics in pore space, which suggest greater bacterial activity in small pores (Kravchenko *et al*., 2019) and that fungi are less affected by pore heterogeneity (Chiu *et al*., 2006). This data has also been used to investigate the interactions of fungal and bacterial richness with soil water repellency. Seaton *et al*. (2019; Appendix 6) found that plant and microbial communities strongly influenced water repellency, with increasing repellency linked to bacterial community composition. In addition, interactions between organismal groups were not assessed in this thesis. Such analyses using this data are now underway. For example, early results show that there is a strong correlation between richness of bacteria and protists across Wales and current analyses are attempting to draw links to aboveground communities nationwide (e.g. birds, plants, insects etc.; Seaton *et al*., *In Prep.*). Other researchers have shown that belowground diversity is linked to ecosystem function, with higher richness being linked to better function (Philippot *et al*., 2013; Delgado-Baquerizo *et al*., 2016), partially due to the increase in functional redundancy insulating against the impacts of species loss (Loreau, 2004).

Such interactions may have implications for measures of functional diversity. I attempted to address this issue by investigating the distributions of various fungal trophic groups as discussed in Chapter 5. Here it was shown that our initial methodology and eDNA findings may not cover the entirety of Welsh soil biodiversity. Rather, the choice in primer and barcoding database hold biases to different organisms. Whilst this did not affect overall trends in richness, it did impact findings of functional groups across land uses, soil types, and organic

matter classes. Richness and proportionate abundance of several fungal classes changed between primer sets, in particular, 18S primers detected Glomeromycetes and Archaeorhizomycetes OTUs that were completely absent in the ITS1 data. Subsequent analyses of trophic guilds revealed differences between datasets, with an increased detection of symbiotrophs, likely Glomeromycetes, in grasslands from the 18S data. Indeed, as discussed, Glomeromycetes are consistently detected better using 18S primers when compared to standard universal 18S primers (Öpik *et al*., 2014). Also, as in Chapter 3, soil type was a poor predictor of biodiversity compared to organic matter classification, likely due to poor resolution from the national soils map used and our focus on topsoils only (Avery, 1980).

Furthermore interspecific interactions are an important determinant of ecosystem service provision (Wagg *et al*., 2019). When SRB richness was compared to that of generalist anaerobes it was shown that both groups dominate different ends of the land use productivity gradient, although absolute richness values followed different trends. Richness of SRB was relatively constant across land uses, whilst anaerobes followed the general trend outlined in Chapter 4. These results, along with those from Chapter 5, illustrate the complex interactions between organisms and their environment that leads to the differential provision of soil ecosystem services. Such findings address the second objective of this thesis by highlighting a shift in functional groups (Chapter 5) and their interactions with other constituent groups of the soil community (Chapter 5 & 6).

The results highlighted above all reveal *in situ* trends of actual biological communities. Experimental manipulations can reveal the trajectories of such communities when pressed to extremes. The long-term C depletion experiment revealed that biomass and diversity of all fractions of soil biodiversity, bar animals, was negatively affected under C starvation and that

this leads to a decline in a range of chemical and physical soil quality indicators. Such results conform with and build on previous findings of microbial biodiversity under long-term fallow (Hirsch *et al*., 2009: Wu *et al*., 2012; Hirsch *et al*., 2016) and lend support to current theories on the preference of soil microbes for pore space >30 μm in diameter (Kravchenko *et al*., 2019). Bioindicator taxa and PLFA analysis demonstrated a concurrent loss of fungal biomass and an increase in Actinomycetes as well as a general loss of plant associated taxa. Analyses of functional genes further supported these observations of a loss of fungal taxa by highlighting a shift to prokaryote genes involved in the decomposition of recalcitrant C forms and methanogenesis. The latter finding is particularly relevant since decreasing pore size results in prolonged anoxia (Keiluweit *et al*., 2017).

As a whole, these results exemplify the depth of soil biodiversity and some of its roles in ecosystem functioning. There is a growing effort to harness soil biodiversity to improve the delivery of ecosystem services (Robinson *et al*., 2014; Bender *et al*., 2016). The research presented here adds to ongoing efforts by demonstrating a near-unifying trend of soil microbes across Wales and the long-term shifts in soil communities under stressful conditions. The data generated in this thesis shows that long-term and national-scale monitoring can effectively capture meaningful trends of soil organisms. Further refinements may yet yield more in-depth results. Some of the GMEP study sites will be revisited as part of the Environment and Rural Affairs Monitoring & Modelling Programme (ERAMMP) recently established by Welsh Government to predict future outcome for rural Wales (Emmett *et al*., 2019). This new project may provide an opportunity to predict future changes in a changing climate. Regardless, this thesis provides an important baseline for such future work and may inform trends at the European-scale.

8.3 Methodological strengths

This work brings together a wide range of experimental methods as is necessary to answer the overarching questions of capturing, enumerating, and characterising, soil biodiversity and its functions. A conceptual illustration of the methods employed by GMEP is presented in Fig. 8.1. The data analysed as part of GMEP was collected using the same methodology as the CS (Emmett *et al*., 2010; Emmett & the GMEP team, 2017). This methodology is proven to be reliable at capturing national trends in soil properties (Emmett *et al*., 2010; Emmett & the GMEP team, 2017), land uses (Bunce *et al*., 1999), and microbial populations (Griffiths *et al*., 2011; Dupont *et al*., 2016) and is comparable to other national soil monitoring projects across Europe (Fay *et al*., 2007; Saby *et al*., 2009). Though it should be noted that some protocols call for sampling from each soil horizon, whereas the CS design does not. This process has the benefit of sampling communities stratified by distinct habitats (Will *et al*., 2010), but adds significantly to processing time.

Furthermore, the GMEP data presented here is only a fraction of one component of the overall project. There was a conscious effort to quantify the state of the Welsh countryside in general, meaning that surveys covered a wide range of features. Other aspects of GMEP included the state of: aboveground biodiversity (i.e. birds and pollinators), freshwater, woodland, climate change mitigation, soil physicochemical properties, and landscape including maintenance of historical features and its accessibility (Emmett & the GMEP team, 2017). In soils, the metabarcoding presented in this work encompasses all fractions of soil biodiversity, expect viruses (see Chapter 3 & 4) and using multiple primer sets for fungi (Chapter 5) and complementary morphological and molecular identification of soil mesofauna (Chapter 3 & 4) that accounted for discrepancies between certain groups. Taken together, this

makes GMEP perhaps one of the most comprehensive surveys of soil biodiversity as yet undertaken.

The C-deprivation experiment presented in Chapter 7 utilises a wide variety of experimental techniques, providing a near-complete picture of the biological and physical impacts of long-term fallow. The combination of metabarcoding, PLFA, and microarray data provides a comprehensive assessment of the biomass, biodiversity, and functional diversity of the belowground communities of fallow and vegetated soils. Furthermore, collaborations with soil physicists allowed for the x-ray CT analyses of soil porosity. Taken together this dataset allows for meaningful conclusions to be made about the complete state of soil biota and its impacts on soil structure under an extreme manipulation.

8.4 Methodological limitations

The methods utilised by GMEP did impose some limitations in the analyses presented here. Firstly, in Chapter 3, morphological identification of soil fauna can introduce subjective inaccuracies due to inexperience, lack of knowledge, and misinterpretation of diagnostic features (Hopkin, 2007; Chen *et al*., 2010). This becomes especially concerning when multiple persons are involved as was the case here, though steps were taken to mitigate misidentification including using identifier as a random factor in mixed effect models. Soil sampling occurred only once at each sampling site during the summer months (Emmett & the GMEP team, 2017). Therefore, the data presented here represents only a snapshot of soil biodiversity at a single time point. Additionally, the GMEP/CS sampling design is limited in that it cannot account for all variations of land use that may exist across Wales. Unlike CS, microbial biomass (e.g. bacteria, Black *et al*., 2003) was not measured under GMEP. As such, some aspects of Welsh biogeography may have been unaccounted for in our analysis. For

example, earthworms are a critical marker of soil quality (Bai *et al*., 2018) and are recognised as ecosystem engineers for their roles maintaining soil water and air flow as well as processing waste and increasing mineralisation rates (Adl, 2003) but were not included in GMEP survey design. The Irish CréBeo soil biodiversity survey found earthworm abundance negatively correlated with that of bacteria and positive correlations between earthworm and fungal richness and community structure (Keith *et al*., 2012). Projections based on this and other European earthworm surveys suggest Wales is home to a high species richness of earthworms with abundances comparable to those of Germany, northern France and the Benelux countries (Rutgers *et al*. 2016). Inclusion of earthworm data in the GMEP framework could have greatly informed this work as Rutgers *et al*. (2016) note that their projections lack any data from the UK outside of Scotland.

A defining aspect of GMEP was its aim to actively monitor impacts of the Glastir agri-environment scheme. This limited the sampling sites to land uses and habitats in priority areas identified by the Welsh Government (Emmett & the GMEP team, 2017). The uptake and extent of Glastir interventions on GMEP sampling sites is not included in this analysis. Arnott *et al*. (2019) highlighted a disparity between socio-economic status of landowners and their participation in Glastir interventions. Furthermore, the duration of Glastir options was varied from short- to long-term and included those that maintained *status quo* conditions (Arnott *et al*., 2019). Therefore, Glastir participation, choice of intervention, and socio-economic factors affecting landowners may be confounding factors affecting soil biodiversity that are not included in this analysis.

Despite the wide breadth of biodiversity markers used in GMEP, viruses were not included. Viruses are often overlooked but represent a critically important component of

biological processes in soil (Kimura *et al*., 2008). Recent work has highlighted their potential roles as regulators of soil C cycling (Trubl *et al*., 2018). Also, current assumptions of the importance of soil viruses currently focus on bacteriophages ignoring those viruses that infect the other fractions of soil biota (Emerson, 2019). However, the methodologies needed to examine viral communities are often complex and require specialist expertise and facilities (Williamson *et al*., 2003; Thurber *et al*., 2009; Trubl *et al*., 2018), making their integration into the GMEP framework an unfortunate impossibility.

A major limitation of the GMEP chapters is a reliance on metabarcoding data. Despite being an increasingly common tool in soil ecology (e.g. Terrat *et al*., 2015; Delgado-Baquerizo *et al*., 2018), metabarcoding is by no means a silver bullet. Indeed, the method as used in this thesis is prone to a number of shortcomings due to its reliance on OTU clustering. The algorithms used to cluster sequences into OTUs are known to introduce significant bias. Furthermore, there are some hard limitations for interpreting metabarcoding data as well. Firstly, metabarcoding data cannot be used for inferring absolute organismal abundances. Inferring functional roles from metabarcoding can be problematic since there is no guarantee the identified organisms are actively expressing key functional genes. Therefore, aside from certain distinct functional and taxonomic groups, it is difficult to state that functions such as key steps in nutrient cycles are being performed. Some evidence for this is even seen in Chapter 6 where SRB appeared to be overshadowed by generalist anaerobes and may be exploiting alternative electron receptors (Muyzer & Stams, 2008). Additionally, in metabarcoding all matching DNA sequences are amplified, including dead organisms (Epp *et al*., 2012), meaning that there is no guarantee the organisms being identified by metabarcoding are actually participating in soil functions.

There is a recent push to avoid OTU-related biases altogether by using amplicon sequence variants (ASVs) to differentiate between unique sequences. In ASV analyses, the biological sequences within samples are inferred before amplification and sequence errors are introduced which means sequences that can be differentiated by even just a single nucleotide (Callahan *et al*., 2017). This push comes from recognition that conventional OTU methodologies are prone to introducing artificial OTUs through overestimations of community diversity due to improper quality controls (Edgar, 2017) and changes in OTU α-diversity across multiple clustering iterations (He *et al*., 2015). Subsequently, many experts have called for a shift away from OTU methods altogether replacing such analyses with clustering-free methodologies including ASVs (Callahan *et al*., 2017 and references therein).

Similarly, continued reductions in monetary and computational expenses have made metagenomics and whole-genome sequencing attractive options for characterising soil communities (Mendes *et al*., 2015; Orellana *et al*., 2018). As with ASV methodologies, modern metagenomic techniques reduce the number of PCR steps and therefore PCR-introduced bias in an analysis (McDonald *et al*., 2016). In many instances, complete genome assemblies are not needed for robust metagenomic analyses, provided that large fragments of target genes are well represented (Hugenholtz & Tyson, 2008). Although there are methodological and financial barriers, metagenomic analyses can be applied to the viral (Trubl *et al*., 2018), prokaryotic, and eukaryotic (Hugenholtz & Tyson, 2008) fractions of microbial communities. Recent soil metagenome analyses have revealed crucial discoveries in the processes and organisms involved in the decomposition of organic matter (Wilhelm *et al*., 2019) and hydrocarbon pollutants (Duarte *et al*., 2017).

Ultimately, the use of more modern approaches such as these would have allowed for more meaningful insights into soil communities. Particularly, there is a need to determine the presence and activity of important functional groups within soils. The GMEP design was focused on capturing the core microbiome of Welsh soils but overlooked many functional metrics. The use of functional genes for identifying important microbial taxa is well established. For example, the ammonia monooxygenase and gene has been used to detect AOA and AOB in soils (i.e. de Sosa *et al*., 2018). Similarly, the dissimilatory sulphate reductase gene is commonly used in studies of SRB (Vigneron *et al*., 2018). Both qPCR (Bouchez *et al*., 2016) and metagenomic (Mendes *et al*., 2015; Orellana *et al*., 2018) techniques could have been used to detect these functionally important organisms and quantify their prevalence across the Welsh landscape. Such analyses would have provided more meaningful data than the functional roles inferred from metabarcoding, such as through the FUNGuild methodology (Nguyen *et al*., 2016).

Some major limitations of the independent experimental work in Chapter 7 are the low replicate numbers and the lack of 1 year-old soils for GeoChip and x-ray CT analyses. The entire C-deprivation project was started over 10 years ago, and the addition of the 1 year-old fallow plots was added as the experiment and analyses carried on. By the time I became involved, the collaborating experts in GeoChip and x-ray CT had already begun their work and the soils collected from 1 year-old fallow plots were almost entirely consumed for analyses and improperly stored for x-ray CT analyses. Unfortunately, this meant that these aspects of the study could not be compared between 1 and 10 years of fallow. Yet given the lack of change in measures of microbial α-diversity between 1- and 10-year fallow soils and the

abundance of nearly all genes detected by GeoChip microarrays was greater in vegetated soils it is likely that no differences would be detected in GeoChip data either.

8.5 Future research

The data generated in this thesis have advanced the understanding of the factors governing belowground microbial community structure at the national-scale. However, the study design and methodological limitations discussed previously warrant future investigation. These gaps in our research are detailed below and their potential place in future research is presented in the conceptual diagram (Fig. 8.1).

i) Methodological and technical constraints limited the ability to quantify the unknown quotient of belowground biodiversity. As outlined previously (Chapter 5 & Chapter 8.4) the choice of primer and database can influence the results generated by metabarcoding analyses. As new databases and primers are developed, there is a possibility that re-analysis of both the GMEP and long-term C-deprivation DNA extracts will yield greater insights into the composition of belowground communities. Similarly, the successful use of mesofauna as indicators of land use change (Chapter 3) and congruent trends from eDNA (Chapter 4) warrant experimentation with other faunal groups. Specifically, a Welsh earthworm survey, potentially complemented by earthworm-specific primers (e.g. Bienert *et al*., 2010), could reveal important trends in soil quality that are accessible and understandable to land-managers.

ii) As described previously, microbial biomass was not measured within GMEP. A lack of quantitative data on the absolute abundance of microbes meant that we were unable to make important inferences from the data. For example, managed

grasslands soils commonly have high bacterial: fungal biomass ratios based on PLFA data (Bardgett *et al*., 1999), whereas those in the opposite state are commonly undisturbed grasslands (Bardgett *et al*., 1996) and more acidic soil systems, such as woodlands (Bååth & Anderson, 2003). This data can be generated using a variety of methods including qPCR if supported by future funding. Such efforts may provide a more complete picture of interactions between the various fractions of belowground communities as they relate to edaphic factors and anthropogenic land use as well as complement mesofaunal abundance data presented in Chapter 3.

iii) Our measures of functional diversity do not reflect realised gene and population functions. Rather, the data inferred from the GeoChip microarray and FunGuild taxonomic assignments indicate the potential for these functions to exist. For example, the aggregated trophic groups identified by FunGuild (Chapter 5) indicate the potential ecological roles played by various fungal taxa (Nguyen *et al*., 2016) though we lack data (i.e. host plant identity, infection rates) that indicates which of these roles they were actively performing This impediment could be overcome to some extent by using microarrays to target functional genes (He *et al*., 2007) as in our long-term C-deprivation experiment (Chapter 7), through targeting specific genes for qPCR analyses (Bouchez *et al*., 2016), for example, stress tolerance and nutrient cycling genes, or with metagenomic techniques (Mendes *et al*., 2015; Orellana *et al*., 2018). Such efforts may help determine the distribution and presence of important soil organisms across Wales.

iv) As discussed above, many GMEP sampling sites will be revisited under ERAMMP. The addition of sampling at multiple time points will facilitate the development of predictive models under ERAMMP. Provisionally, the ERAMMP sampling design will make use of 240 GMEP sites (Emmett *et al*., 2019). Incorporating biological analyses into ERAMMP could lead to the development of powerful models of soil community responses to interventions and land use changes. I would recommend such analyses focus on important functional groups, including AOA, AOB, and SRB, or soil-borne pathogens of humans, livestock, and crops. An ERAMMP time series could be also applied to tracking invasive species or evaluate native seed banks. Ultimately, such analyses must be able to develop quantitative soil quality indicators if they are to be of use for policy-makers or land managers. Such projects would require detailed land use history, including data on agri-environment scheme uptake. (Arnott *et al*., 2019) as well as a newer soil map that is robust at finer-scale resolution to make soil type a more effective indicator than the Avery (1980) map was during the present analyses.

v) Most importantly, there have been major advancements in the use of environmental DNA across all fields of ecology. Whilst it remains an important tool for the modern soil ecologist, advances in metagenomic techniques have made many of the uses of a metabarcoding approach redundant. Recent work has highlighted the efficacy of metagenomic approaches in facilitating culture-independent analyses of soil function (Mendes *et al*., 2015; Orellana *et al*., 2018). Furthermore, there has been a progressive shift away from the use of OTUs across the field of microbial ecology. As discussed previously, comprehensive studies have shown that OTU

clustering methods are prone to unstable results through multiple clustering iterations (He *et al*., 2015), and overestimations of community diversity due to improper quality controls (Edgar, 2017). In its place, ASVs have become a popular new method. The DADA2 R package (Callahan et al., 2016) is commonly used to study ASVs and such analyses have been incorporated into studies of soil ecology (Yuan *et al*., 2018; Grządziel & Gałązka, 2019; Renault *et al*., 2019). Revisiting our analyses with these techniques may reveal different trends as in theory such analyses will be more accurate. However, we made the decision to forgo this, as the main GMEP results presented in Chapter 4 and built upon thereafter, are consistent with established regional (Gossner *et al*., 2016), national (Griffiths *et al*., 2011; Dupont *et al*., 2016; Terrat *et al*., 2017), and global (Fierer *et al*., 2007; Delgado-Bacquerizo *et al*., 2018) trends. Nevertheless, it would be exciting to see how a more refined technique may alter our findings and reveal new insights into belowground communities.

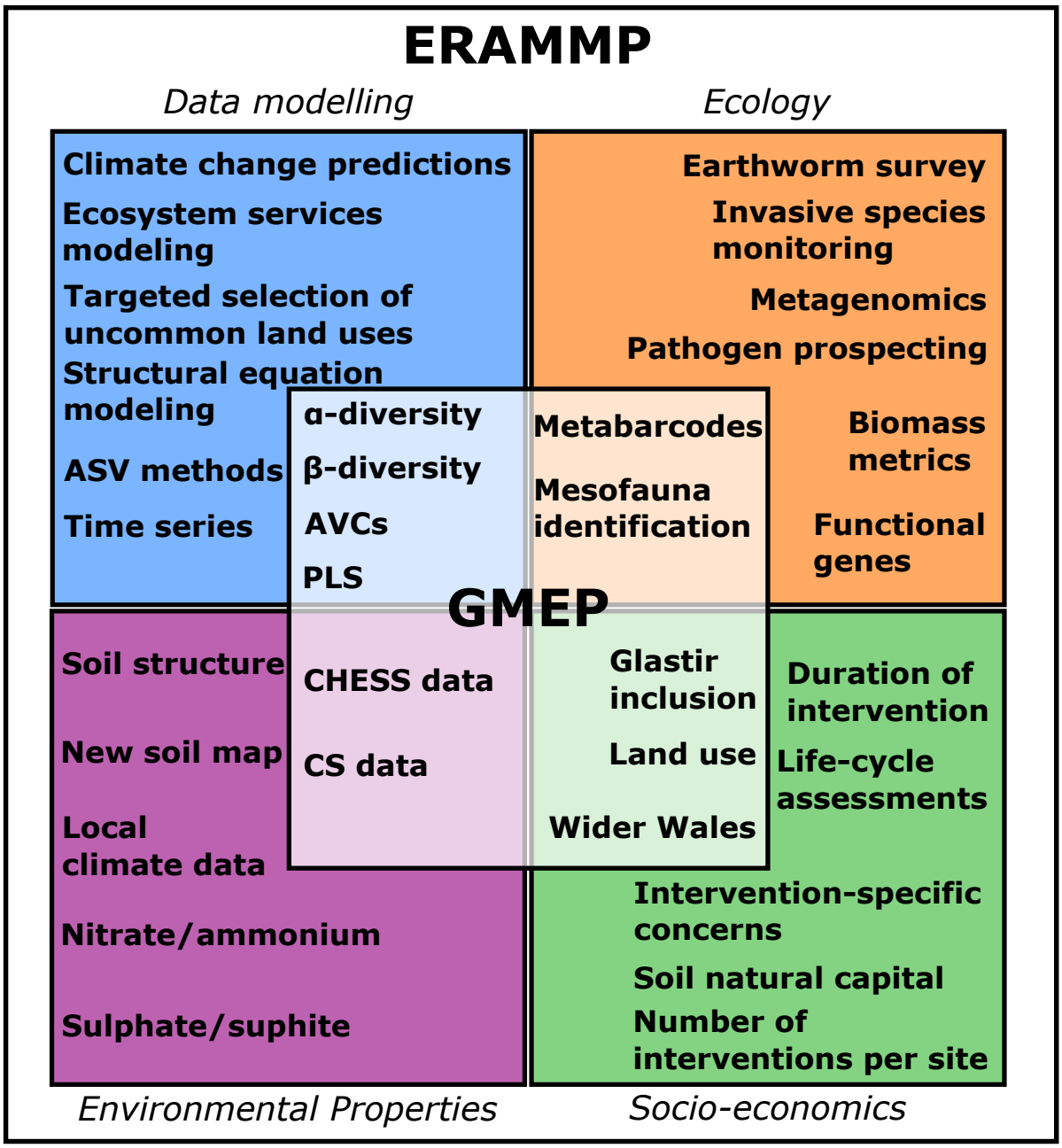


Figure 8.1 A conceptual diagram of methods, concepts, and outcomes of national soil as classified by data modelling (blue), ecology (orange), environmental properties (purple), and socio-economics (green). Aspects that were included as part of GMEP are presented in the inner (white) box; recommended considerations for ERAMMP are presented in the outer box.

8.6 General conclusions

This work sought to answer fundamental questions related to soil biological and functional diversity across heterogeneous land uses and soil types at the national-scale. National-scale trends of soil biota relating to land use and productivity were discovered. Methodological congruence between metabarcoding and morphometric identification of mesofauna was an important finding, signifying the relevance of these invertebrates as indicator taxa. Incongruences between primer sets in analyses of fungi exemplified the need to utilise multiple primers to capture the breath of soil biodiversity. The detection of various fungal functional groups and SRB from the GMEP dataset may indicate trends across land uses at the continental-scale. Long-term C deprivation caused a collapse in soil physical structure and biological diversity, with a shift to organisms reliant on recalcitrant C and anaerobic energy sources. In conclusion, this thesis provided essential information on the natural history and ecology of Earth's most diverse, and arguably, most important biological communities. The results presented here will aid in future soil biological monitoring in Wales, the UK, and across the globe.

8.7 References

Arnott, D., Chadwick, D., Harris, I., Koj, A., Jones, D.L., 2019. What can management option uptake tell us about ecosystem services delivery through agri-environment schemes? Land Use Policy 81, 194-208.

Arroyo, J., Keith, A.M., Schmidt, O., Bolger, T., 2013. Mite abundance and richness in an Irish survey of soil biodiversity with comments on some newly recorded species. Ir. Nat. J. 33, 19-27.

Avery, W., 1980. Soil Classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden.

Bååth, E., Anderson, T.-H., 2003. Comparison of soil fungal/bacterial rations in a pH gradient using physiological and PLFA-based techniques. Soil Biol. Biochem. 35, 955-963.

Bai, Z., Caspari, T., Gonzales, M.R., Batjes, N.H., Mäder, P., Bünemann, E.K., de Goede, R. et al., 2018. Effects of agricultural management practices on soil quality: a review of long-term experiments for Europe and China. Agr. Ecosyst. Environ. 265, 1-7.

Bardgett, R.M., Hobbs, P.J., Frostegård, Å., 1996. Changes in soil fungal:bacterial biomass ratios following reductions in the intensity of management of an upland grassland. Biol. Fertil. Soils 22, 261-264.

Bardgett, R.M., McAlister, E., 1999. The measurement of soil fungal:bacterial biomass ratios as an indicator of ecosystem self-regulation in temperate meadow grasslands. Biol. Fertil. Soils 29, 282-290.

Bender, S.F., Wagg, C., van der Heijden, M.G.A., 2016. An underground revolution: biodiversity and soil ecological engineering for agricultural sustainability. Trends Ecol. Evol. 31, 440-452.

Bienert, F., de Danieli, S., Miquel, C., Coissac, E., Poillot, C., Brun, J.-J., Taberlet, P., 2012. Tracking earthworm communities from soil DNA. Mol. Ecol. 21, 2017-2030.

Bouchez, T., Blieux, A.L., Dequiedt, S., Domaizon, I., Dufresne, A., Ferreira, S., Godon, J.J. et al., 2016. Molecular microbiology methods for environmental diagnosis. Environ. Chem. Lett. 14, 423-441.

Callahan, B.J., McMurdie, P.J., Holmes, S.P., 2017. Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. ISME J. 11, 2639-2643.

Callahan, B.J., McMurdie, P.J., Rosen, M.J., Han, A.W., Johnson, A.J.A., Holmes, S.P., 2016. DADA2: high resolution sample inference from Illumina amplicon data. Nat. Methods 13, 581-583.

Chen, D., Wang, Y., Lan, Z., Li, J., Xing, W., Hu, S., Bai, Y., 2015. Biotic community shifts explain the contrasting responses of microbial and root respiration to experimental soil acidification. Soil Biol. Biochem. 90, 139-147.

Chiu, C.-Y., Chen, T.-H., Imberger, K., Tian, G., 2006. Particle size fractionation of fungal and bacterial biomass in subalpine grassland and forest soils. Geoderma 130, 265-271.

Coleman, D.C., Callahan, M., Crossley, D. Jr., 2018. Fundamentals of Soil Ecology, third ed. Academic Press, London.

Cornell, J.H., 1978. Diversity in tropical rain forests and coral reefs. Science 199, 1302-1310.

Delgado-Baquerizo, M, Maestre, F.T., Reich, P.B., Jeffries, T.C., Gaitan J.J., Encinar, D., Berdugo, M. et al., 2016. Microbial diversity drives multifunctionality in terrestrial ecosystems. Nat. Commun. 7: 10541.

Delgado-Baquerizo, M., Oliverio, A.M., Brewer, T.E., Benavent-González, A., Eldridge, D.J., Bardgett, R.D., Maestre, F.T. et al., 2018. A global atlas of the dominant bacteria found in soil. Science 359, 320-325.

Duarte, M., Nielsen, A., Camarinha-Silva, A., Vilchez-Vargas, R., Bruls, T., Wos-Oxley, M.L., Jauregui, R. et al., 2017. Functional soil metagenomics: elucidation of polycyclic aromatic hydrocarbon degradation potential following 12 years of *in situ* bioremediation. Environ. Microbiol. 19, 2992-3011.

Dupont, A.Ö.C., Griffiths, R.I., Bell, T., Bass, D., 2016. Differences in soil micro-eukaryotic communities over soil pH gradients are strongly driven by parasites and saprotrophs. Environ. Microbiol. 18, 2010-2024.

Edgar, R.C., 2017. Accuracy of microbial community diversity estimated by closed- and open-references OTUs. PeerJ 5: e3889. doi: 10.7717/peerj.3889

Emerson, J.B., 2019. Soil viruses: a new hope. mSystems 4:e00120-19. doi: 10.1128/mSystems.00120-19

Emmett, B.A., Alexander, M., Alison, J., Anthony, S., Ballinger, R., Bell, C., Bowgen, K. et al., 2019. ERAMMP Year 1 Report 11: Year 1 Summary. Welsh Government Contract C210/2016/2017.

Emmett, B.A., the GMEP team, 2017. Glastir Monitoring & Evaluation Programme. Final Report to Welsh Government – Executive Summary (Contract reference: C147/2010/11). NERC/Centre for Ecology & Hydrology (CEH Projects: NEC04780/NEC05371/NEC05782).

Emmett, B.A., Reynolds, B., Chamberlain, P.M., Rowe, E., Spurgeon, D., Brittain, S.A., Frogbrook, Z. et al., 2010. Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology. (CS Technical Report No. 9/07, CEH Project Number: C03259).

Epp, L.S., Boessenkool, S., Belleain, E.P., Haile, J., Esposito, A., Riaz, T., Erséus, C. et al., 2012. New environmental metabarcodes for analyzing soil DNA: potential for studying past and present ecosystems. Mol. Ecol. 21, 1821-1833.

Fay, D., McGrath, D., Zhang, C., Carrigg, C., O'Flaherty, V., Kramers, G., Carton, O.T. et al., 2007. Towards A National Soil Database. Environmental Protection Agency, Ireland, 19 pp. (Synthesis Report 2001-CD/S2-M2).

Fierer, N., Breitbart, M., Nulton, J., Salamon, P., Lozupone, C., Jones, R., Robeson, M. et al., 2007. Metagenomic and small-subunit rRNA analyses reveal the genetic diversity of bacteria, archaea, fungi, and viruses in soil. Appl. Environ. Microb. 73, 7059-7066.

Geisen, S., Mitchell, E.A.D., Adl, S., Bonkowski, M., Dunthorn, M., Ekelund, F., Fernández, L.D. et al., 2018. Soil protists: a fertile frontier in soil biology research. FEMS Microbiol. Res. 42, 293-323.

Gosner, M.M., Lewinsohn, T.M., Kahl, T., Grassein, F., Boch, S., Prati, D., Birkhofer, K. et al., 2016. Land-use intensification causes multitrophic homogenization of grassland communities. Nature 540, 266-269.

Griffiths, R.I., Thompson, B.C., James, P., Bell, T., Bailey, M., Whiteley, A.S., 2011. The bacterial biogeography of British soils: mapping soil bacteria. Environ. Microbiol. 13, 1642-1654

Grządziel, J., Gałązka, A., 2019. Fungal biodiversity of the most common types of Polish soil in a long-term microplot experiment. Front. Microbiol. 10, 6.

He, Y., Caporaso, J.G., Jiang, X.-T., Sheng, H.-F., Huse, S.M., Rideout, J.R., Edgar, R.C. et al., 2015. Stability of operational taxonomic units: an important but neglected property for analyzing microbial diversity. Microbiome 3, 20.

He, Z., Gentry, T.J., Schadt, C.W., Wu, L., Liebich, J., Chong, S.C., Huang, Z. et al., 2007. GeoChip: a comprehensive microarray for investigated biogeochemical, ecological and environmental processes. ISME J. 1, 67-77.

Hirsch, P.R., Jhurreea, D., Williams, J.K., Murray, P.J., Scott, T., Misselbrook, T.H., Goulding, K.W.T. et al., 2016. Soil resilience and recovery: rapid community responses to management changes. Plant Soil 412, 283-297.

Hirsch, P.R., Gilliam, L.M., Sohi, S.P., Williams, J.K., Clark, I.M., Murray, P.J. 2009. Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. Soil Biol. Biochem. 41, 2021-2024.

Hopkin, S.P., 2007. A Key to the Collembola (Springtails) of Britain and Ireland. Field Studies Council, Telford, UK.

Hugenholtz, P., Tyson, G.W., 2008. Microbiology – Metagenomics. Nature 455: 481-483.

Keiluweit, M, Wanzek, T., Kleber, M, Nico, P., Fendorf, S. 2017. Anaerobic microsites have an unaccounted role in soil carbon stabilization. Nat. Commun. 8:1771.

Keith, A.M., Boots, B., Hazard, C., Niechoj, R., Arroyo, J., Bending, G.D., Bolger, T. et al., 2012. Cross-taxa congruence, indicators and environmental gradients in soils under agricultural and extensive land management. Eur. J. Soil Biol. 49, 55-62.

Ktavchenko, A.N., Guber, A.K., Razavi, B.S., Koestel, J., Quigley, M.Y., Robertson, G.P., Kuzyakov, Y. 2019. Microbial spatial footprint as a driver of soil carbon stabilization. Nat Commun. 110:3121.

Kimura, M., Jia, Z.-J., Nakayama, N., Asakawa, S., 2008. Ecology of viruses in soils: past, present and future perspectives. Soil Sci. Plant Nutr. 54, 1-32.

Loreau, M. 2004. Does functional redundancy exist? Oikos 3, 606-611.

McDonald, J.E., Larsen, N., Pennington, A., Connolly, J., Wallis, C., Rooks, D.J., Hall, N. et al., 2016. Characterising the canine oral microbiome by direct sequencing reverse-transcribed rRNA molecules. PLoS ONE 11:e0157046.

Mendes, L.W., Tsai, S.M., Navarrete, A.A., de Hollander, M., van Veen, J.A., Kuramae, E.E., 2015. Soil-borne microbiome: linking diversity to function. Microb. Ecol. 70, 255-265.

Muyzer, G., Stams, A.J., 2008. The ecology and biotechnology of sulphate-reducing bacteria. Nat. Rev. Microbiol. 6, 441-454.

Nguyen, N.H., Song, Z., Bates, S.T., Branco, S., Tedersoo, L., Menke, J., Schilling, J.S. et al., 2016. FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. Fungal Divers. 20, 241-248.

Öpik, M., Davison, J., Moora, M., Zobel, M., 2014. DNA-based detection and identification of Glomeromycota: the virtual taxonomy of environmental sequences. Botany 92, 135-147.

Orellana, L.H., Chee-Sanford, J.C., Sanford, R.A., Löffler, F.E., Konstantinidis, K.T., 2018. Year-round shotgun metagenomes reveal stable microbial communities in agricultural soils and novel ammonia oxidizers responding to fertilization. Appl. Environ. Microbiol. 84, e01646-17.

Oton, V.E., Quince, C., Nicol, G.W., Prosser, J.I., Gubry-Rangin, C., 2016. Phylogenetic congruence and ecological coherence in terrestrial Thaumarchaeta. ISME J. 10, 85-96.

Philippot, L., Spor, A., Bru, D., Bizouard, F., Jones, C.M., Maron, P.-A., 2013. Loss in microbial diversity affects nitrogen cycling in soil. ISME J. 7, 1609-1619.

Robinson, D.A., Fraser, I., Dominati, E.J., Davi∂sdóttir, B., Jónsson, J.O.G., Jones, L., Jones, S.B. et al., 2014. On the value of soil resources in the context of natural capital and ecosystem service delivery. Soil Sci. Soc. Am. J. 78, 685-700.

Renault, S., Masse, J., Norrie, J.P., Blal, B., Hijri, M., 2019. A commercial seaweed extract structured microbial communities associated with tomato and pepper roots and significantly increased crop yield. Microb. Biotechnol. *Published online before print.*

Rutgers, M., Orgiazzi, A., Gardi, C., Römbke, J., Jänsch, S., Keith, A.M., Neilson, R. et al., 2016. Mapping earthworm communities in Europe. Appl. Soil Ecol. 97, 98-111.

Saby, N.P.A., Thioulouse, J., Jolivet, C.C., Ratié, C., Boulnne, L., Bispo, A., Arrouays, D., 2009. Multivariae analysis of spatial patterns of 8 trace elements using the French soil monitoring network data. Sci. Tot. Environ. 407, 5644-5652.

Seaton, F.M., Jones, D.L., Creer, S., George, P.B.L., Smart, S., Lebron, I., Barrett, G. et al., 2019. Plant and soil communities are associated with the response of soil water repellency to environmental stress. Sci. Total Environ. 687, 929-938.

de Sosa, L.L., Glanville, H.C., Marshall, M.R., Williams, A.P., Abadie, M., Clark, I.M., Blaud, A. et al., 2018. Spatial zoning of microbial functions and plant-soil nitrogen dynamics across a riparian area in an extensively grazed livestock system. Soil Biol. Biochem. 120, 153-164.

Subbarao, G.V., Shrawat, K.L., Nakahara, K., Ishikawa, T., Kishii, M., Rao, I.M., Hash, C.T. et al., 2012. Chapter six – Biological nitrification inhibition – a novel strategy to regulate nitrification in agricultural systems. In: Sparks, D.L., ed. Advances in Agronomy 114, Academic Press, Cambridge. pp. 249-302.

Tedersoo, L., Mohammad, B., Põlme, S., Kõljalg, U., Yorou, N.S., Wijesundera, R., Ruiz, L.V. et al., 2014. Global diversity and geography of soil fungi. Science 346, 1256688.

Terrat, S., Dequiedt, S., Horrigue, W., Lelievre, M., Cruaud, C., Saby, N.P.A., Jolivet, C. et al., 2015. Improving soil bacterial taxa-area relationship assessment using DNA meta-barcoding. Heredity 114, 468-475.

Thurber, R.V., Haynes, M., Breitbart, M., Wegley, L., Rohwer, F., 2009. Laboratory procedures to generate viral metagenomes. Nat. Protoc. 4, 470-483.

Tsiafouli, M.A., Thébault, E., Sgardelis, S.P., de Ruiter, P.C., van der Putten, W.H., Birkhofer, K., Hemerik, L. et al., 2015. Intensive agriculture reduces soil biodiversity across Europe. Glob. Change Biol. 21, 973-985.

Trubl, G., Jang, H.B., Roux, S., Emereson, J.B., Solonenko, N., Vik, D.R., Solden, L. et al., 2018. Soil viruses are underexplored players in ecosystem carbon processing. mSystems 3:e00076-18.

Vigneron, A., Cruaud, P., Alsop, E., de Rezende, J.R., Head, I.M., Tsesmetzis, N., 2018. Beyond the tip of the iceberg; a new view of the diversity of sulphite- and sulphate-reducing microorganisms. ISME J. 12, 2096-2099.

Wagg, C., Schlaeppi, K., Banerjee, S., Kuramae, E.E., Marcel, G.A., van der Heijden, M.G.A., 2019. Fungal-bacterial diversity microbiome complexity predict ecosystem functioning. Nat. Commun. 10:4841.

Wilhelm, R.C., Singh, R., Eltis, L.D., Mohn, W.W., 2019. Bacterial contributions to delignification and lignocellulose degradation in forest soils with metagenomic and quantitative stable isotope probing. ISME J. 13, 413-429.

Will, C., Thürmer, A., Wollherr, A., Nacke, H., Herold, N., Schrumpf, M., Gutknecht, J. et al., 2010. Horizon-specific bacteria community composition of German grassland soils, as revealed by pyrosequencing-based analysis of 16S rRNA genes. Appl. Environ. Microbiol. 76, 6751-6759.

Will, K.W., Mishler, B.D., Wheeler, Q.D., 2005. The perils of DNA barcoding and the need for integrative taxonomy. Syst. Biol. 54, 844-851.

Williamson, K.E., Wommack, K.E., Radosevich, M., 2003. Sampling natural viral communities from soil culture-independent analyses. Appl. Environ. Microbiol. 69, 6628-6633.

Wu, Y., Kemmitt, S., White, RoP, Xu, J, Brookes, P.C. 2016. Carbon dynamics in a 60 year fallowed loamy-sand soil compared to that in a 60 year permanent arable or permanent grassland UK soil. Plant Soil 352: 51-63.

Yuan, J., Zhao, J., Wen, T., Zhao, M., Li, R., Goossens, P., Huang, Q. et al., 2018. Root exudates drive the soil-borne legacy of aboveground pathogen infection. Microbiome 6, 156.

Appendix 1

Supplementary Material for Chapter 3

George, P.B.L., Keith, A.M., Creer, S., Barrett, G.L., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L., 2019. Evaluation of mesofauna communities as soil quality indicators in a national-level monitoring programme. *Soil Biol. Biochem*. 115, 537-546.

Appendix 1.1: Glastir Monitoring and Evaluation Programme

The Glastir Monitoring and Evaluation Programme (GMEP) has been designed to assess the outcomes of implementing the Welsh Government's Glastir agri-environment scheme. GMEP is a collaboration funded by the Welsh Government and the European Union. The GMEP programme is run by the NERC Centre for Ecology and Hydrology and is a collaboration between specialists from public research centres, universities, voluntary bodies, and consultancies. When active, GMEP was the largest and most in-depth monitoring programme measuring environmental state and change within the European Union (Emmett and the GMEP Team, 2014). It also represents a progressive step for national scale agri-environment schemes. GMEP follows a holistic ecosystem approach with a rolling annual survey conducted across areas both participating in and abstaining from Glastir. The results of the field survey were combined with national data and models to produce findings that inform stakeholders. Summaries of findings from Year 1 (2013) and Year 2 (2014) have been published and are accessible to the public (Emmett and the GMEP Team 2014; 2015).

Appendix 1.2: Soil maps in Wales

The soils at each sampling point were assigned to soil type using the National Soil Map and Soil Classification (Cranfield University, 2004). This map and classification scheme is derived from Avery (1980) with revisions from Clayden and Hollis (1984). Soils were assigned to groups based on published soil maps and reconnaissance mapping of previously unsurveyed sites (for more detail see Cranfield University, 2004). Generally, soils in Wales are known to map poorly, however, due to the high level of local heterogeneity.

Supplementary Tables

Table 1. Description of Aggregate Vegetation Classes identified in this study. Adapted from Smart et al. (2003).

Aggregate Vegetation Class	Description
Crops/weeds	Communities on disturbed or cultivated land, including weedy, horticultural, and species-poor arable land.
Tall grassland/herb	Tall herbaceous communities common in field edges, "old field" communities, plus road- and stream-sides.
Fertile grassland	Improved or semi-improved grassland. Usually with high nutrient inputs and cut more than once a year.
Infertile Grassland	Semi-improved to unimproved, less productive grasslands, species-rich grasslands including wet or dry and acidic to basic variations.
Lowland wood	Dominated by trees and shrubs in neutral or basic lowlands, scrublands, and hedgerows.
Upland wood	Commonly acidic conifer plantations, scrubland and semi-natural broadleaved woods in the uplands.
Moorland grass/mosaic	Grass-dominated upland pasture, commonly with a long history of livestock grazing.
Heath/bog	Heather dominated, commonly upland landscapes, including dry heath and bogs.

Table 2. UK soil groups listed with their complementary classification in the FAO World Reference Base Classification (WRB, 2006). Soils are listed in alphabetical order.

Major UK soil group	World Reference Base
Brown	Primarily Cambisols plus some Luvisols and Acrisols
Lithomorphic	Leptosols with some Regosols
Surface- and ground-water gleys	Primarily Gleysols, Planosols, and some Fluvisols/Luvisols
Podzolic	Podzols
Peat	Histosols
Man made	Anthrosols

Table 3. Mean values (± SE) of soil physical and chemical variables for each Aggregate Vegetation Class.

Soil variable	Crops/weeds	Tall grassland/herb	Fertile grassland	Infertile grassland	Lowland wood	Upland wood	Moorland grass/mosaic	Heath/bog
Total C (%)	3.51 ± (0.600)	4.42 ± (2.210)	4.88 ± (0.161)	5.99 ± (0.289)	5.60 ± (0.775)	17.25 ± (1.916)	21.25 ± (1.548)	41.33 ± (1.663)
Total N (%)	0.28 ± (0.037)	0.38 ± (0.171)	0.46 ± (0.013)	0.50 ± (0.017)	0.40 ± (0.046)	0.91 ± (0.084)	1.23 ± (0.076)	1.72 ± (0.074)
C:N ratio	13.66 ± (1.87)	10.96 ± (0.630)	10.49 ± (0.112)	11.73 ± (0.191)	13.80 ± (0.563)	17.13 ± (0.643)	16.39 ± (0.460)	24.50 ± (0.817)
pH (0.01 M $CaCl_2$)	5.10 ± (0.277)	5.30 ± (0.237)	5.17 ± (0.064)	4.73 ± (0.039)	4.38 ± (0.234)	3.51 ± (0.077)	3.69 ± (0.058)	3.25 ± (0.077)
Bulk density (g/cm^3)	1.11 ± (0.070)	1.02 ± (0.187)	0.87 ± (0.016)	0.80 ± (0.013)	0.72 ± (0.063)	0.43 ± (0.035)	0.37 ± (0.030)	0.17 ± (0.237)
Soil water repellency[a]	0.88 ± (0.334)	0.78 ± (0.398)	1.64 ± (0.072)	2.07 ± (0.055)	2.16 ± (0.226)	6.12 ± (0.312)	2.91 ± (0.098)	2.99 ± (0.096)
Volumetric water content (m^3/m^3)	0.25 ± (0.025)	0.37 ± (0.095)	0.35 ± (0.010)	0.35 ± (0.009)	0.22 ± (0.019)	0.36 ± (0.021)	0.49 ± (0.017)	0.61 ± (0.017)
Total P (mg/kg)	994.59 ± (115.564)	645.71 ± (75.786)	1234.98 ± (40.592)	1025.97 ± (40.592)	624.26 ± (64.522)	872.65 ± (48.002)	1034.41 ± (47.926)	934.51 ± (39.172)

[a] Soil water repellency was derived from median water drop penetration times (s) and log transformed.

Table 4. Mean abundances (± SD) of mesofauna groups in each Aggregate Vegetation Class.

Aggregate Vegetation Class	Total mesofauna	Total invertebrates	Oribatida	Mesostigmata	Other mites	Entomobryoidea	Poduroidea	Symphypleona
Crops/weeds	18.60 ± (43.21)	19.20 ± (43.11)	7.87 ± (21.86)	5.40 ± (7.44)	0.20 ± (0.56)	1.73 ± (4.91)	3.20 ± (9.46)	0.20 ± (0.77)
Tall grassland/herb	8.67 ± (6.66)	8.67 ± (6.66)	1.67 ± (2.08)	5.00 ± (5.00)	0.33 ± (0.58)	0.00 ± (0.00)	1.67 ± (1.53)	0.00 ± (0.00)
Fertile grassland	26.92 ± (29.49)	27.80 ± (30.07)	9.13 ± (19.92)	6.34 ± (7.64)	2.33 ± (5.32)	4.47 ± (9.57)	4.46 ± (5.84)	0.19 ± (0.56)
Infertile grassland	40.14 ± (37.99)	41.26 ± (38.68)	16.49 ± (22.86)	8.02 ± (10.46)	4.89 ± (11.73)	5.35 ± (10.61)	5.22 ± (8.45)	0.17 ± (0.63)
Lowland wood	60.92 ± (44.81)	61.84 ± (45.63)	33.12 ± (34.11)	5.76 ± (5.29)	1.88 ± (3.75)	11.04 ± (12.35)	8.20 ± (7.53)	0.92 ± (2.22)
Upland wood	47.95 ± (45.03)	49.35 ± (45.67)	24.15 ± (25.40)	5.55 ± (6.21)	2.17 ± (3.69)	9.85 ± (21.06)	5.92 ± (8.81)	0.32 ± (0.95)
Moorland grass/mosaic	41.95 ± (49.86)	42.55 ± (50.07)	26.76 ± (39.03)	2.89 ± (5.36)	3.29 ± (9.67)	4.44 ± (8.67)	4.52 ± (8.16)	0.06 ± (0.37)
Heath/bog	32.83 ± (49.32)	33.21 ± (49.69)	25.42 ± (40.45)	1.56 ± (2.92)	2.82 ± (8.65)	1.44 ± (3.13)	1.53 ± (3.04)	0.06 ± (037)

Table 5. Mean abundances (± SD) of mesofauna groups in each soil type.

Soil type	Total mesofauna	Total invertebrates	Oribatida	Mesostigmata	Other mites	Entomobryoidea	Poduroidea	Symphypleona
Lithomorphic	30.19 ± (30.03)	30.24 ± (30.00)	15.24 ± (17.32)	3.86 ± (7.14)	6.57 ± (13.68)	1.86 ± (3.90)	2.67 ± (3.01)	0.00 ± (0.00)
Brown	35.70 ± (34.06)	36.52 ± (34.68)	12.66 ± (18.64)	7.43 ± (8.79)	3.90 ± (11.06)	6.25 ± (10.19)	5.19 ± (7.01)	0.28 ± (0.98)
Podzolic	41.36 ± (45.13)	42.22 ± (45.80)	19.77 ± (29.34)	6.49 ± (9.49)	3.31 ± (8.81)	6.20 ± (15.44)	5.44 ± (8.94)	0.15 ± (0.61)
Surface-water gley	38.10 ± (45.40)	39.52 ± (45.94)	23.90 ± (37.04)	4.60 ± (7.33)	2.84 ± (5.91)	3.18 ± (5.66)	3.42 ± (5.22)	0.15 ± (0.59)
Ground-water gley	40.12 ± (45.98)	41.35 ± (45.75)	21.59 ± (38.29)	4.29 ± (5.64)	1.59 ± (3.28)	8.06 ± (20.37)	4.35 ± (7.88)	0.24 ± (0.56)
Peat	36.28 ± (47.64)	36.93 ± (47.89)	23.90 ± (33.80)	2.51 ± (4.36)	2.69 ± (6.81)	3.00 ± (7.62)	4.07 ± (9.43)	0.12 ± (0.45)

Table 6. Mean abundances (± SD) of mesofauna groups in each loss-on-ignition (LOI) class.

LOI class	Total mesofauna	Total invertebrates	Oribatida	Mesostigmata	Other mites	Entomobryoidea	Poduroidea	Symphypleona
Mineral	32.44 ± (32.85)	33.34 ± (33.25)	12.42 ± (19.61)	6.74 ± (8.68)	3.03 ± (7.64)	6.16 ± (11.12)	3.89 ± (5.52)	0.21 ± (0.73)
Humus-mineral	39.26 ± (38.84)	40.26 ± (39.43)	17.38 ± (24.43)	6.66 ± (8.92)	4.04 ± (10.44)	5.28 ± (9.98)	5.70 ± (8.33)	0.21 ± (0.83)
Organo-mineral	42.71 ± (51.08)	43.79 ± (52.12)	24.63 ± (29.67)	4.33 ± (6.98)	2.65 ± (5.40)	6.00 ± (15.84)	4.94 ± (11.29)	0.17 ± (0.52)
Organic	33.81 ± (46.00)	34.32 ± (46.19)	27.18 ± (41.23)	1.89 ± (3.62)	1.49 ± (4.41)	1.19 ± (2.41)	1.99 ± (3.63)	0.08 ± (0.39)

References

Avery, B.W., 1980. Soil classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden, UK.

Clayden, B., Hollis, J.M., 1984. Criteria for Differentiating Soil Series. Soil Survey Technical Monograph No 17. Harpenden, UK.

Cranfield University, 2004. The National Soil Map and Soil Classification. Cranfield Soil and Agrifood Institute, Cranfield University, Bedfordshire, UK.

Emmett, B.A., the GMEP team, 2014. Glastir Monitoring & Evaluation Programme. First Year Annual Report to Welsh Government (Contract reference: C174/2010/11). NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780), pp. 442.

Emmett, B.A., the GMEP team, 2015. Glastir Monitoring & Evaluation Programme, Second Year Annual Report to Welsh Government (Contract Reference: C147/2010/11). NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780), pp 316.

Smart, S.M., Robertson, J.C., Shield, E.J., Van De Poll, H.M., 2003. Locating eutrophication effects across British vegetation between 1990 and 1998. Global Change Biology 9, 1763-1774.

World Reference Base for Soil Resources, 2006. Report 103. Food and Agriculture Organization of the United Nations, Rome, Italy.

Supplementary Figures

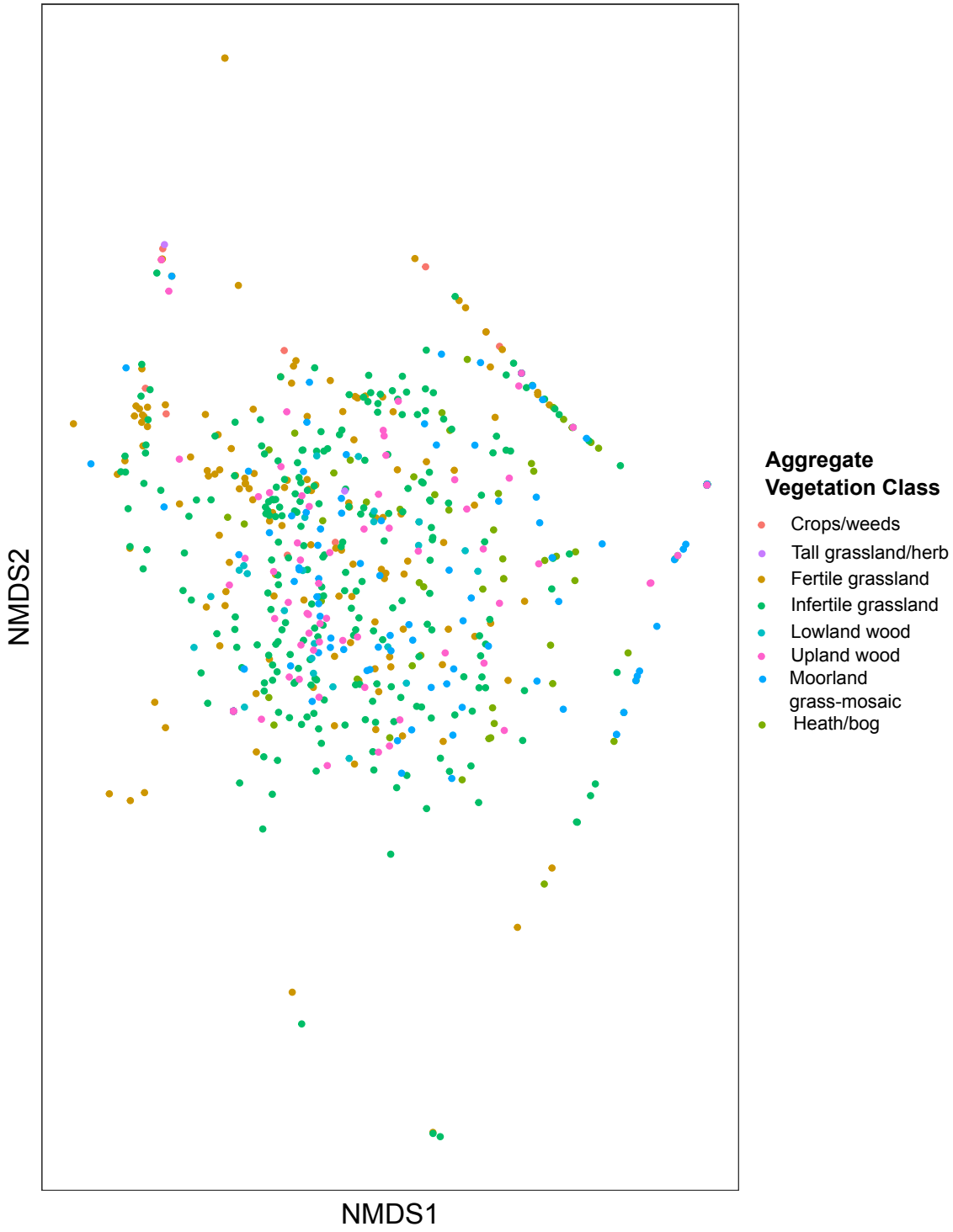


Fig 1. Non-metric dimensional scaling of mesofauna communities (log + 1 transformed abundance) in each Aggregate Vegetation Class.

Appendix 2

Supplementary Material for Chapter 4

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L., 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nat. Commun.* 110, 1107. doi: 10.1038/s41467-019-09031-1

Appendix 2.1 Creation of Aggregate Vegetation Classes

The land use classification used in this study was originally developed for the UK Countryside Survey in 1990. In short, vegetation data was collected from 508 1 km randomly selected squares across the UK. Within each square, vegetation was recoreded in a number of plots placed either placed randomly or targeted to cover semi-natural habitats and along various landscape features such as field boundaries, hedges, and roads. This vegetation data was grouped into 100 vegetation classes using the TWINSPAN programme[1]. Then, detrended correspondence analysis using DECORANA[2] clustered these 100 vegetation classes into 8 Aggregate Vegetation Classes (AVCs), of which 7 were identified in the current study (Appendix 2 Table 1). The AVCs are ordered according to soil nutrient content[3], from the high-nutrient crops to the low-nutrient bogs, the order is listed in Appendix 2 Table 1. Such a decline in soil nutrient content also implies both productivity and management intensity gradients.

Appendix 2.2 Pair-wise comparisons of bacterial, fungal, and archaeal richness

Linear mixed-models showed significant trends across land use types. Bacterial richness decreased ($F_{6, 264} = 78.47$, $p < 0.0001$) in AVCs across the productivity gradient, with highest values in the most productive Crops/weeds and grasslands and lowest in the low productivity land uses (Fig 4.3a). Specifically, richness in Heath/bog sites was

significantly lower than all other AVCs except Upland wood ($p = 0.003$ for Moorland grass-mosaic; $p = 0.002$ for Lowland wood; $p < 0.0001$ for rest). Richness in the Crops/weeds, Fertile grassland, and Infertile grassland AVCs was also significantly greater than Upland wood ($p = 0.01$ for Crops/weeds; $p < 0.001$ for both grasslands). Additionally, higher levels of bacterial OTU richness were observed in Fertile grassland and Infertile grassland AVCs when compared to both Lowland wood ($p = 0.002$; $p < 0.001$, respectively) and Moorland grass-mosaic sites ($p < 0.001$ for both).

Fungi ($F_{6, 248} = 48.98$, $p < 0.001$; Fig. 4.3b) and protists ($F_{6, 249} = 59.86$, $p < 0.001$; Fig. 4.3c) followed the same trend as bacteria. For fungi, richness in Crops/weeds was significantly higher than Moorland grass-mosaic ($p = 0.002$), Heath/bog, and Lowland as well as Upland wood ($p < 0.001$). Heath/bog ($p < 0.001$), Moorland grass-mosaic ($p = 0.01$), Lowland ($p = 0.006$) and Upland wood ($p < 0.001$) all had significantly lower richness values. Richness of Fertile grassland sites was also higher than all other AVCs ($p < 0.001$) except Crops/weeds. For richness of protists (Fig. 4.3c), again, the productive Crops/weeds and grassland sites had significantly greater richness than the woodland and upland sites (all $p < 0.001$). Protist richness of Fertile and Infertile grasslands and Lowland wood, Upland wood, Moorland grass-mosaic, were all significantly greater than in Heath/bog as well (all $p < 0.001$).

Appendix 2.3 Supplementary Methods

Sankey diagrams were produced in R[4] using the riverplot package[5]. In brief, proportional abundances at the class-level were calculated on rarefied OTU tables of each organismal group (i.e. bacteria, archaea, etc.) using package phyloseq[6]. Proportions of each class were then assigned to a data frame of "edges". The data are treated such that

the value denotes the distance between "node 1" (i.e. bacteria) and "node 2" (i.e. Proteobacteria, etc.). The names of these nodes are extracted into a new data frame in, which the horizontal and vertical locations of the nodes are determined. Colour was also assigned within this data frame. Finally, the nodes and edges are coerced into a list and converted to an "rp" class object and then presented with the plot function.

Supplementary Tables

Table 1. Description of Aggregate Vegetation Classes identified in this study. Adapted from Smart et al.[11].

Aggregate Vegetation Class	Description
Crops/weeds (n = 9)	Communities on disturbed or cultivated land, including weedy, horticultural, and species-poor arable land.
Fertile grassland (n = 98)	Improved or semi-improved grassland. Usually with high nutrient inputs and cut more than once a year.
Infertile Grassland (n = 162)	Semi-improved to unimproved, less productive grasslands, species-rich grasslands including wet or dry and acidic to basic variations.
Lowland wood (n = 17)	Dominated by trees and shrubs in neutral or basic lowlands, scrublands, and hedgerows.
Upland wood (n = 44)	Commonly acidic conifer plantations, scrubland and semi-natural broadleaved woods in the uplands.
Moorland grass/mosaic (n = 54)	Grass-dominated upland pasture, commonly with a long history of livestock grazing.
Heath/bog (n = 52)	Heather dominated, commonly upland landscapes, including dry heath and bogs.

Table 2. Pearson's correlation coefficients of the relationship between richness of major groups of soil biota. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Taxon	Bacteria	Archaea	Fungi	Protists
Bacteria				
Archaea	-0.33***			
Fungi	0.65***	-0.29***		
Protists	0.82***	-0.38***	0.65***	
Animals	0.20***	0.04	0.07	0.20***

Table 3. Summary of relationships amongst environmental factors and archaea communities derived from NMDS ordination and linear fitting with the envfit function. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Variable	Correlation		
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.57***	+	-
C:N ratio^S	0.49***	-	+
Elevation (m)	0.48***	-	+
Bulk density (g cm^3 ^-1)	0.41***	+	+
Mean annual precipitation (mL)	0.35***	-	-
Organic matter (% LOI)^L	0.34***	-	-
Total C (%)^L	0.34***	-	-
Clay content (%)^A	0.31***	+	+
Volumetric water content (m^3 m^3 ^-1)	0.3***	-	-
Soil bound water (g water g dry soil ^-1)	0.24***	-	-
Soil water repellency^{L*}	0.24***	-	+
Total N (%)^L	0.17***	-	-
Total P (mg kg ^-1)^S	0.12***	+	-
Sand content (%)^A	0.1***	-	+
Collembola^{L1}	0.06***	+	+
Total mesofauna^{L1}	0.05**	-	+
Mites^{L1}	0.05**	-	+
Temperature (°C)	0.05**	-	+
Rock volume (mL)	0.04*	+	+

Note: ^{A} denotes Aitchison's log-ratio transformation; ^{L} denotes log_{10}-transformation; ^{L1} denotes log_{10} plus 1 transformation ^{S} denotes square-root-transformation; * soil water repellency was derived from median water drop penetration times (s).

Table 4. Summary of relationships amongst environmental factors and fungal communities derived from NMDS ordination and linear fitting with the envfit function. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Variable	Correlation		
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.6^{***}	-	+
C:N ratio[S]	0.47^{***}	+	-
Elevation (m)	0.41^{***}	+	-
Volumetric water content (m^3 m^3 ^-1)	0.41^{***}	+	-
Mean annual precipitation (mL)	0.39^{***}	+	-
Bulk density (g cm^3 ^-1)	0.38^{***}	-	+
Organic matter (% LOI)[L]	0.37^{***}	+	-
Total C (%)[L]	0.31^{***}	+	-
Clay content (%)[A]	0.28^{***}	-	+
Soil bound water (g water g dry soil ^-1)	0.26^{***}	+	-
Soil water repellency[L*]	0.24^{***}	+	-
Total N (%)[L]	0.21^{***}	+	-
Sand content (%)[A]	0.19^{***}	+	+
Collembola[L1]	0.15^{***}	-	+
Total mesofauna[L1]	0.12^{***}	+	+
Total P (mg kg ^-1)[S]	0.11^{***}	-	-
Mites[L1]	0.1^{***}	+	+
Rock volume (mL)	0.07^{***}	-	+
Temperature (°C)	0.04^{***}	-	+

Note: [A] denotes Aitchison's log-ratio transformation; [L] denotes log_{10}-transformation; [L1] denotes log_{10} plus 1 transformation [S] denotes square-root-transformation:

* soil water repellency was derived from median water drop penetration times (s).

Table 5. Summary of relationships amongst environmental factors and protistan communities derived from NMDS ordination and linear fitting with the envfit function. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Variable	Correlation		
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.6^{***}	-	-
C:N ratio^{S}	0.45^{***}	+	-
Elevation (m)	0.43^{***}	+	-
Mean annual precipitation (mL)	0.42^{***}	+	-
Total C (%)^{L}	0.4^{***}	+	-
Organic matter (% LOI)^{L}	0.39^{***}	+	-
Bulk density (g cm^3 ^-1)	0.37^{***}	-	+
Volumetric water content (m^3 m^3 ^-1)	0.37^{***}	+	-
Clay content (%)^{A}	0.28^{***}	-	+
Total N (%)^{L}	0.26^{***}	+	-
Soil water repellency^{L*}	0.22^{***}	+	-
Soil bound water (g water g dry soil ^-1)	0.2^{***}	+	-
Sand content (%)^{A}	0.14^{***}	+	+
Collembola^{L1}	0.12^{***}	-	+
Total mesofauna^{L1}	0.09^{***}	+	+
Mites^{L1}	0.07^{***}	+	+
Total P (mg kg ^-1)^{S}	0.07^{***}	-	-
Rock volume (mL)	0.06^{**}	-	+
Temperature (°C)	0.03^{*}	+	+

Note: ^{A} denotes Aitchison's log-ratio transformation; ^{L} denotes $\log_{10}$-transformation;
^{L1} denotes $\log_{10}$ plus 1 transformation ^{S} denotes square-root-transformation;
* soil water repellency was derived from median water drop penetration times (s).

Table 6. Summary of relationships amongst environmental factors and animal communities derived from NMDS ordination and linear fitting with the envfit function. +/- signify the direction of association between each variable and respective NMDS axes. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Variable	Correlation		
	R^2	Axis1	Axis2
pH ($CaCl_2$)	0.48^{***}	-	+
Bulk density (g cm^3 ^-1)	0.43^{***}	-	-
C:N ratio[S]	0.35^{***}	+	+
Organic matter (% LOI)[L]	0.35^{***}	+	+
Volumetric water content (m^3 m^3 ^-1)	0.32^{***}	+	+
Total C (%)[L]	0.29^{***}	+	+
Elevation (m)	0.28^{***}	+	+
Soil water repellency[L*]	0.27^{***}	+	-
Mean annual precipitation (mL)	0.24^{***}	+	+
Clay content (%)[A]	0.22^{***}	-	-
Total N (%)[L]	0.2^{***}	+	+
Soil bound water (g water g dry soil^-1)	0.2^{***}	+	+
Mites[L1]	0.11^{***}	+	-
Total mesofauna[L1]	0.1^{***}	+	-
Sand content (%)[A]	0.08^{***}	+	-
Rock volume (mL)	0.08^{***}	-	-
Total P (mg kg ^-1)[S]	0.06^{***}	-	+
Collembola[L1]	0.05^{**}	-	-
Temperature (°C)	0.03^{*}	+	-

Note: [A] denotes Aitchison's log-ratio transformation; [L] denotes log_{10}-transformation; [L1] denotes log_{10} plus 1 transformation [S] denotes square-root-transformation:

* soil water repellency was derived from median water drop penetration times (s).

Table 7. Summary of relationships amongst environmental factors and bacterial communities derived from CAP ordination and linear fitting with the envfit function. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Soil and environmental variables	R^2
pH ($CaCl_2$)	0.66***
Mean annual precipitation (mL)	0.51***
C:N ratio^S	0.48***
Elevation (m)	0.47***
Volumetric water content (m^3 m^3 ^-1)	0.46***
Bulk density (g cm^3 ^-1)	0.44***
Organic matter (% LOI)^L	0.39***
Total C (%)^L	0.32***
Clay content (%)^A	0.29***
Soil bound water (g water g dry soil ^-1)	0.26***
Soil water repellency^L	0.26***
Sand content (%)^A	0.22***
Total N (%)^L	0.22***
Total P (mg kg ^-1)^S	0.09***
Collembola^{L1}	0.09***
Total mesofauna^{L1}	0.08***
Mites^{L1}	0.08***
Rock volume (mL)	0.05***
Temperature (°C)	0.04**

Note: ^{A} denotes Aitchison's log-ratio transformation;
^{L} denotes log_{10}-transformation; ^{L1} denotes log_{10} plus 1 transformation
^{S} denotes square-root-transformation:
* soil water repellency was derived from median water drop penetration times (s).

Table 8. Summary of relationships amongst environmental factors and fungal communities derived from CAP ordination and linear fitting with the envfit function. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Soil and environmental variables	R^2
C:N ratio^{S}	0.43***
Elevation (m)	0.35***
pH ($CaCl_2$)	0.35***
Volumetric water content (m^3 m^3 ^-1)	0.34***
Bulk density (g cm^3 ^-1)	0.30***
Mean annual precipitation (mL)	0.22***
Sand content (%)^{A}	0.20***
Organic matter (% LOI)^{L}	0.20***
Clay content (%)^{A}	0.20***
Total C (%)^{L}	0.18***
Soil water repellency^{L}	0.18***
Soil bound water (g water g dry soil ^-1)	0.14***
Collembola^{L1}	0.09***
Total P (mg kg ^-1)^{S}	0.09***
Total mesofauna^{L1}	0.07***
Mites^{L1}	0.07***
Total N (%)^{L}	0.07***
Rock volume (mL)	0.06**
Temperature (°C)	0.06***

Note: ^{A} denotes Aitchison's log-ratio transformation;
^{L} denotes log_{10}-transformation; ^{L1} denotes log_{10} plus 1 transformation
^{S} denotes square-root-transformation:
* soil water repellency was derived from median water drop penetration times (s).

Table 9. Summary of relationships amongst environmental factors and protistan communities derived from CAP ordination and linear fitting with the envfit function. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Soil and environmental variables	R^2
pH ($CaCl_2$)	0.59***
C:N ratio^S	0.46***
Total C (%)^L	0.41***
Organic matter (% LOI)^L	0.40***
Bulk density (g cm^3 ^-1)	0.40***
Elevation (m)	0.38***
Mean annual precipitation (mL)	0.33***
Clay content (%)^A	0.27***
Volumetric water content (m^3 m^3 ^-1)	0.27***
Total N (%)^L	0.26***
Soil water repellency^L	0.24***
Soil bound water (g water g dry soil ^-1)	0.24***
Sand content (%)^A	0.24***
Total P (mg kg ^-1)^S	0.15***
Total mesofauna^{L1}	0.10***
Collembola^{L1}	0.10***
Mites^{L1}	0.09***
Rock volume (mL)	0.03*
Temperature (°C)	0.03*

Note: ^{A} denotes Aitchison's log-ratio transformation;

^{L} denotes log_{10}-transformation; ^{L1} denotes log_{10} plus 1 transformation

^{S} denotes square-root-transformation:

* soil water repellency was derived from median water drop penetration times (s).

Table 10. Summary of relationships amongst environmental factors and archaeal communities derived from CAP ordination and linear fitting with the envfit function. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Soil and environmental variables	R^2
pH ($CaCl_2$)	0.60***
Elevation (m)	0.45***
Bulk density (g cm^3 ^-1)	0.44***
C:N ratio[S]	0.42***
Total C (%)[L]	0.35***
Organic matter (% LOI)[L]	0.35***
Mean annual precipitation (mL)	0.33***
Clay content (%)[A]	0.30***
Volumetric water content (m^3 m^3 ^-1)	0.28***
Soil bound water (g water g dry soil ^-1)	0.27***
Soil water repellency[L]	0.24***
Total N (%)[L]	0.21***
Total P (mg kg ^-1)[S]	0.10***
Sand content (%)[A]	0.06**
Collembola[L1]	0.06***
Mites[L1]	0.06**
Total mesofauna[L1]	0.05**
Temperature (°C)	0.05**
Rock volume (mL)	0.02

Note: [A] denotes Aitchison's log-ratio transformation;
[L] denotes log_{10}-transformation; [L1] denotes log_{10} plus 1 transformation
[S] denotes square-root-transformation:
* soil water repellency was derived from median water drop penetration times (s).

Table 11. Summary of relationships amongst environmental factors and animal communities derived from CAP ordination and linear fitting with the envfit function. *** indicates P < 0.001, ** 0.001 > P < 0.01, * 0.01 > P < 0.05, blank indicates P > 0.05.

Soil and environmental variables	R^2
pH ($CaCl_2$)	0.47***
Volumetric water content (m^3 m^3 ^-1)	0.35***
C:N ratio[S]	0.29***
Bulk density (g cm^3 ^-1)	0.26***
Elevation (m)	0.26***
Organic matter (% LOI)[L]	0.25***
Total C (%)[L]	0.21***
Soil water repellency[L]	0.20***
Clay content (%)[A]	0.20***
Mean annual precipitation (mL)	0.19***
Total N (%)[L]	0.14***
Sand content (%)[A]	0.13***
Soil bound water (g water g dry soil ^-1)	0.13***
Total mesofauna[L1]	0.11***
Mites[L1]	0.10***
Collembola[L1]	0.08***
Total P (mg kg ^-1)[S]	0.08***
Temperature (°C)	0.07***
Rock volume (mL)	0.06***

Note: [A] denotes Aitchison's log-ratio transformation;
[L] denotes log_{10}-transformation; [L1] denotes log_{10} plus 1 transformation
[S] denotes square-root-transformation:
* soil water repellency was derived from median water drop penetration times (s).

Table 12. Mean values (± SE) of soil physical and chemical variables of each Aggregate Vegetation Class (AVC). Following normalisation on selected variables (see notes of Supplementary Tables 2-5) ANOVAs and Tukey's *post-hoc* tests were performed. Results are as follows: total C ($F_{6, 427} = 89.13$ $p < 0.001$), total N ($F_{6, 427} = 61.03$, $p < 0.001$), C :N ratio ($F_{6, 427} = 94.41$, $p < 0.001$), total P ($F_{6, 424} = 7.1$, $p < 0.001$), organic matter ($F_{6, 428} = 107.02$, $p < 0.001$), pH ($F_{6, 428} = 69.56$, $p < 0.001$), soil water repellency ($F_{6, 428} = 22.08$, $p < 0.001$), volumetric water content ($F_{6, 427} = 33.74$, $p < 0.001$), soil bound water ($F_{6, 428} = 79.87$, $p < 0.001$), rock volume ($F_{6, 427} = 10.4$, $p < 0.001$), bulk density ($F_{6, 427} = 90.99$, $p < 0.001$), clay content ($F_{6, 344} = 19.54$, $p < 0.001$), sand content ($F_{6, 344} = 5.71$, $p < 0.001$), elevation ($F_{6, 429} = 78.42$, $p < 0.001$), mean annual precipitation ($F_{6, 429} = 72.6$, $p < 0.001$), and temperature ($F_{6, 429} = 4.4$, $p < 0.001$).

Environmental variable	Crops/weeds	Fertile grassland	Infertile grassland	Lowland wood	Upland wood	Moorland grass-mosaic	Heath/bog
Total C (%)	3.87 (± 0.83)d	4.75 (± 0.2)d	5.85 (± 0.33)d	5.78 (± 1.07)d	9.7 (± 2.25)c	12.19 (± 2.07)b	23.57 (± 1.88)a
Total N (%)	0.32 (± 0.05)d	0.45 (± 0.02)d	0.49 (± 0.02)d	0.4 (± 0.06)d	0.58 (± 0.1)c	0.83 (± 0.11)b	1.05 (± 0.09)a
C :N ratio	11.44 (± 0.81)cd	10.49 (± 0.13)d	11.62 (± 0.27)cd	13.92 (± 0.75)bc	15.86 (± 0.7)b	14.41 (± 0.42)b	20.65 (± 0.94)a
Total P (mg kg^-1)	1103.44 (± 145.47)ab	1194.9 (± 45.53)a	1045.5 (± 43.3)ab	601.68 (± 77.68)c	762.45 (± 61.95)bc	930.49 (± 57.5)ab	769.63 (± 50.04)ab
Organic matter (% LOI)	7.53 (± 1.62)d	9.39 (± 0.34)d	11.25 (± 0.55)d	10.71 (± 1.7)d	18.79 (± 4.16)c	22.99 (± 3.72)b	39.26 (± 3.6)a
pH ($CaCl_2$)	4.73 (± 0.26)b	5.2 (± 0.08)a	4.73 (± 0.05)b	4.31 (± 0.26)b	3.57 (± 0.1)cd	3.85 (± 0.09)c	3.84 (± 0.1)d
Soil water repellency*	4077.56 (± 3990.72)abc	264.01 (± 73.28)c	781.68 (± 137.58)b	2975.47 (± 2108.12)abc	1965.87 (± 698.61)a	4186.13 (± 798.48)a	3186.4 (± 812.15)a
Volumetric water content (m^3 m^3 ^-1)	0.23 (± 0.03)bc	0.35 (± 0.01)b	0.34 (± 0.01)b	0.22 (± 0.02)c	0.36 (± 0.03)b	0.46 (± 0.02)a	0.52 (± 0.02)a
Soil bound water (g water g dry soil ^-1)	2.19 (± 0.32)c	2.74 (± 0.11)c	2.89 (± 0.11)c	2.92 (± 0.34)c	3.7 (± 0.49)b	4.45 (± 0.46)b	6.03 (± 0.47)a
Rock volume (mL)	3.95 (± 1.11)abc	5.25 (± 0.45)b	5.44 (± 0.42)b	9.13 (± 2.49)a	4.41 (± 0.57)ab	3.25 (± 0.39)c	1.87 (± 0.21)c
Bulk density (g cm^3 ^-1)	1.03 (± 0.09)a	0.9 (± 0.02)a	0.8 (± 0.02)b	0.71 (± 0.08)b	0.56 (± 0.04)c	0.5 (± 0.04)c	0.47 (± 0.03)d
Clay content (%)	22.25 (± 1.85)ab	25.46 (± 0.65)a	23.18 (± 0.64)ab	17.47 (± 1.34)ab	17.82 (± 1.82)ab	18.12 (± 1.27)c	11.76 (± 2.24)d
Sand content (%)	30.97 (± 4.66)ad	24.88 (± 1.25)d	29.21 (± 1.44)bd	42.99 (± 4.01)ac	40.23 (± 4.15)abc	29.5 (± 3.0)b	45.15 (± 7.61)a
Elevation (m)	88.71 (± 47.69)cd	109.38 (± 8.62)d	167.28 (± 8.65)c	119.06 (± 16.38)cd	297.83 (± 20.62)b	406.63 (± 19.22)a	380.55 (± 19.7)a
Mean annual precipitation (mL)	968.44 (± 69.01)c	1078.19 (± 24.71)c	1177.05 (± 18.91)c	1100.12 (± 52.28)c	1405.33 (± 65.35)b	2027.23 (± 74.39)a	1771.2 (± 58.19)a
Temperature (°C)	12.64 (± 1.18)ab	12.09 (± 0.41)b	13.44 (± 0.29)a	15.8 (± 0.87)a	14.53 (± 0.53)a	14.51 (± 0.36)a	13.87 (± 0.29)a

*Soil water repellency was derived from median water drop penetration times (s) and log transformed.

Table 13. Rarefaction depth and a breakdown of replicate numbers for each taxonomic group.

Taxon	Rarefaction depth (reads)	Replicates per Aggregate Vegetation Class
Bacteria	40,000	Crops/weeds = 9 Fertile grassland = 96 Infertile grassland = 157 Lowland wood = 17 Upland wood = 43 Moorland grass-mosaic = 54 Heath/bog = 52
Archaea	200	Crops/weeds = 9 Fertile grassland = 87 Infertile grassland = 91 Lowland wood = 15 Upland wood = 42 Moorland grass-mosaic = 48 Heath/bog = 51
Fungi	4,000	Crops/weeds = 9 Fertile grassland = 97 Infertile grassland = 156 Lowland wood = 17 Upland wood = 43 Moorland grass-mosaic = 44 Heath/bog = 47
Protists	15,000	Crops/weeds = 9 Fertile grassland = 98 Infertile grassland = 160 Lowland wood = 17 Upland wood = 40 Moorland grass-mosaic = 46 Heath/bog = 42
Animals	1,000	Crops/weeds = 7 Fertile grassland = 91 Infertile grassland = 144 Lowland wood = 17 Upland wood = 44 Moorland grass-mosaic = 53 Heath/bog = 52

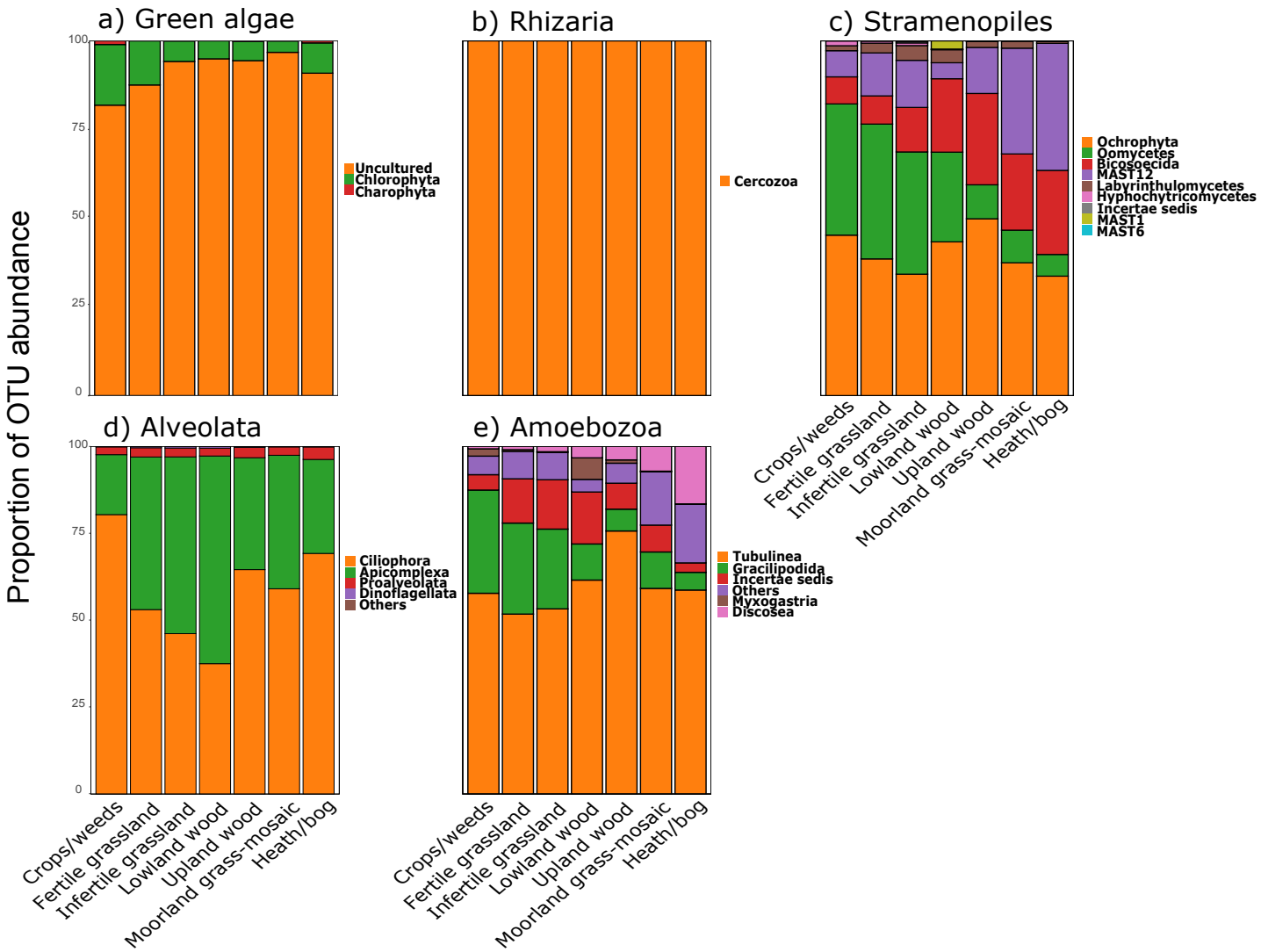


Fig. 1. Proportionate abundances of OTUs for major protist groups within each Aggregate Vegetation Class ordered from most (Crops/weeds) to least (Heath/bog) productive for **a)** Chloroplastida; **b)** Rhizaria; **c)** Stramenopiles; **d)** Alveolates; and **e)** Amoebozoa.

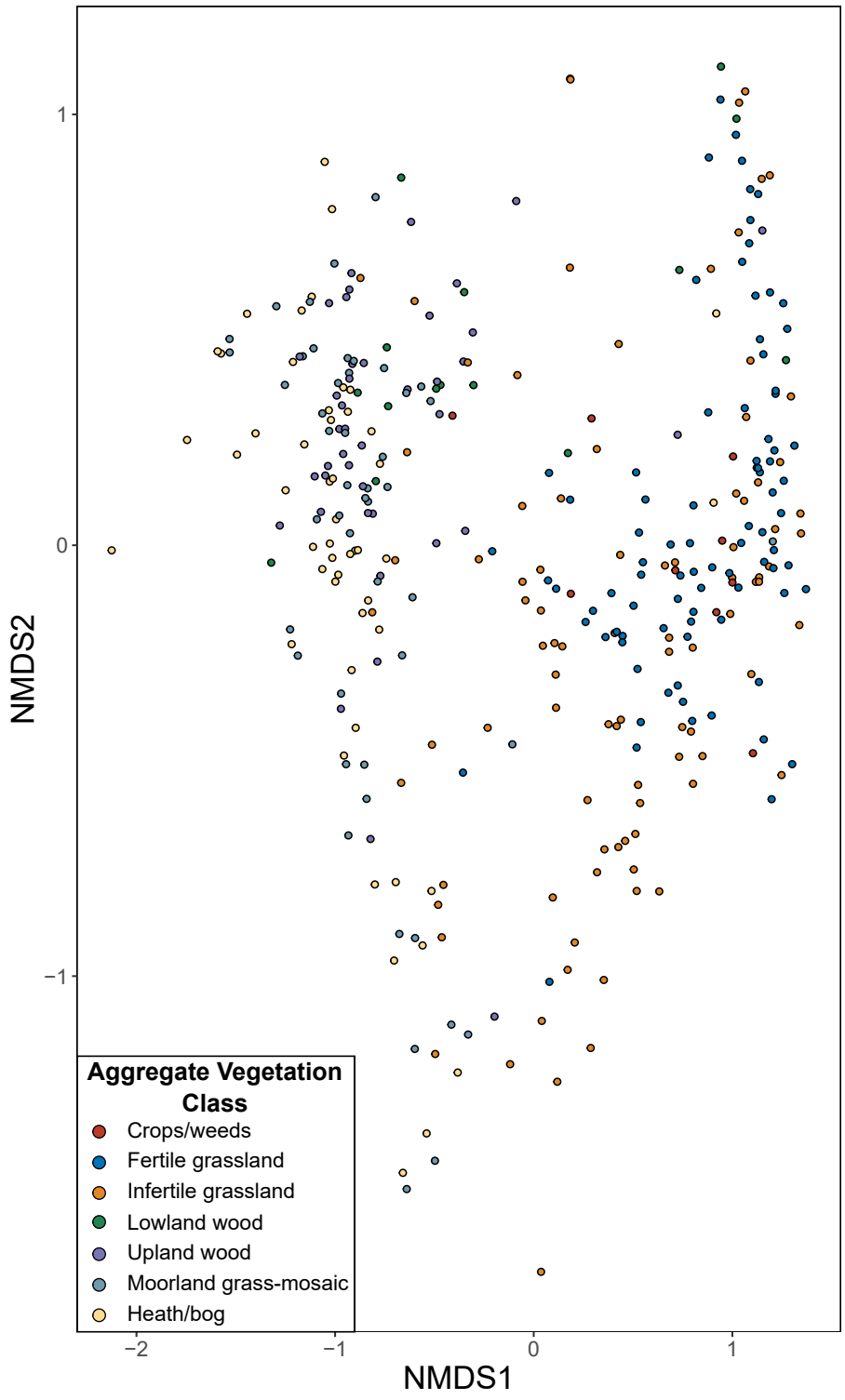


Fig. 2. Plot of the non-metric dimensional scaling ordination (stress = 0.11) of archaea community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Results of PERMANOVA ($F_{6,336} = 15.32$, $p = 0.001$) and of dispersion of variances ($F_{6,336} = 8.52$, $p = 0.001$) were significant.

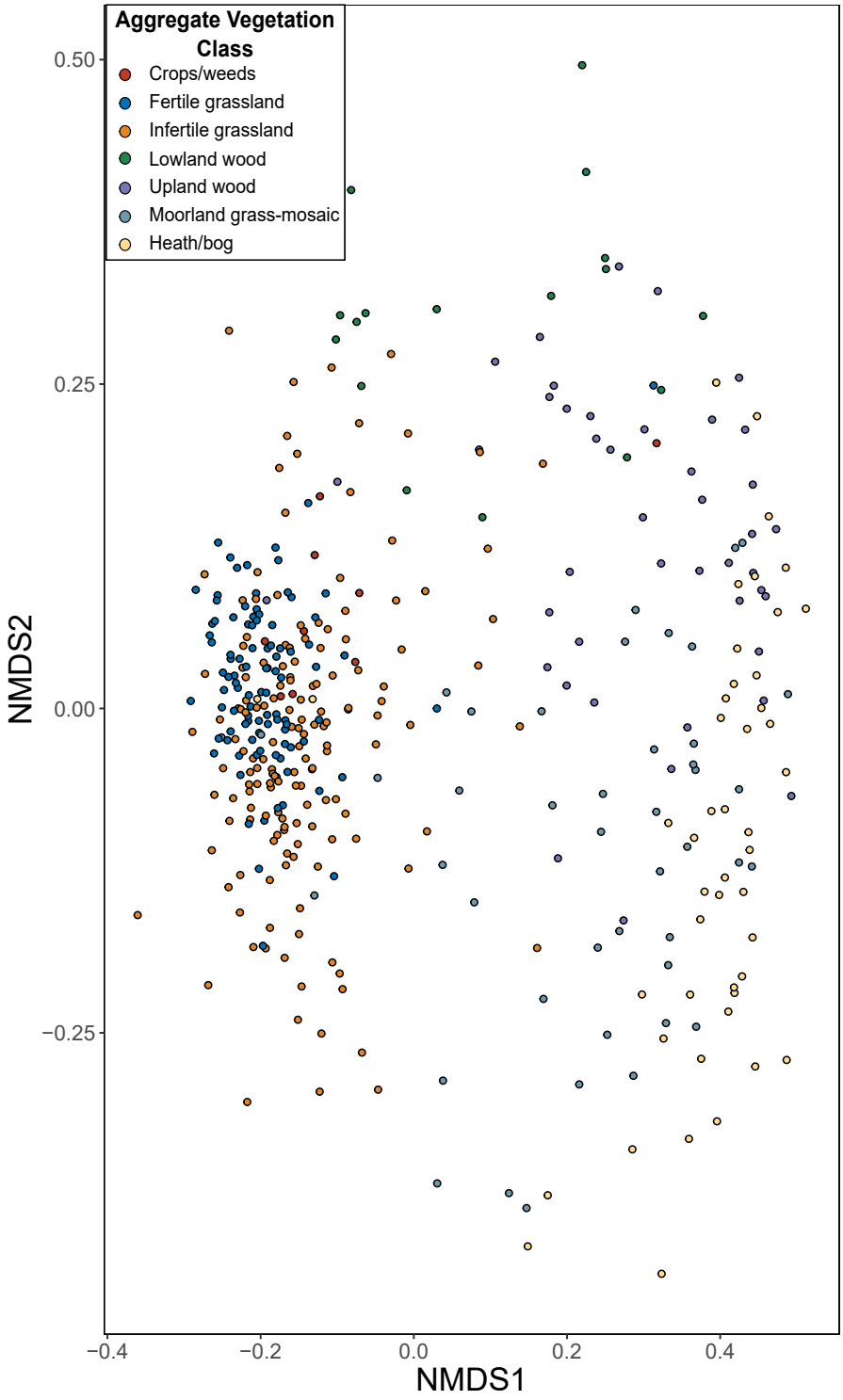


Fig. 3 Plot of the non-metric dimensional scaling ordination (stress = 0.13) of fungi community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Results of PERMANOVA ($F_{6,406}$ = 10.74, p = 0.001) and of dispersion of variances ($F_{6,406}$ = 41.30, p = 0.001) were significant.

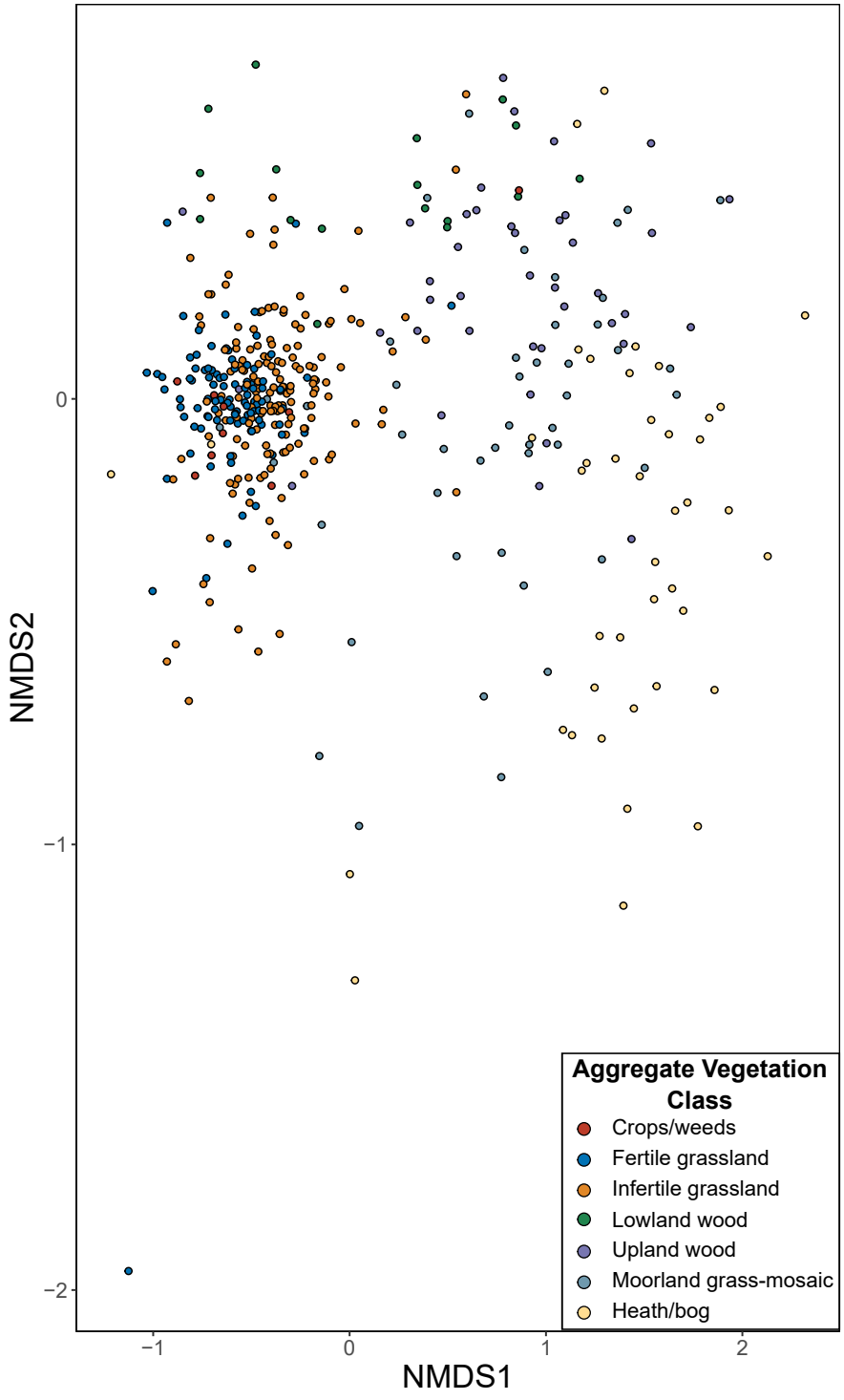


Fig. 4. Plot of the non-metric dimensional scaling ordination (stress = 0.08) of protist community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Results of PERMANOVA ($F_{6,405}$= 31.60, p = 0.001) and of dispersion of variances ($F_{6,405}$ = 17.63, p = 0.001) were significant.

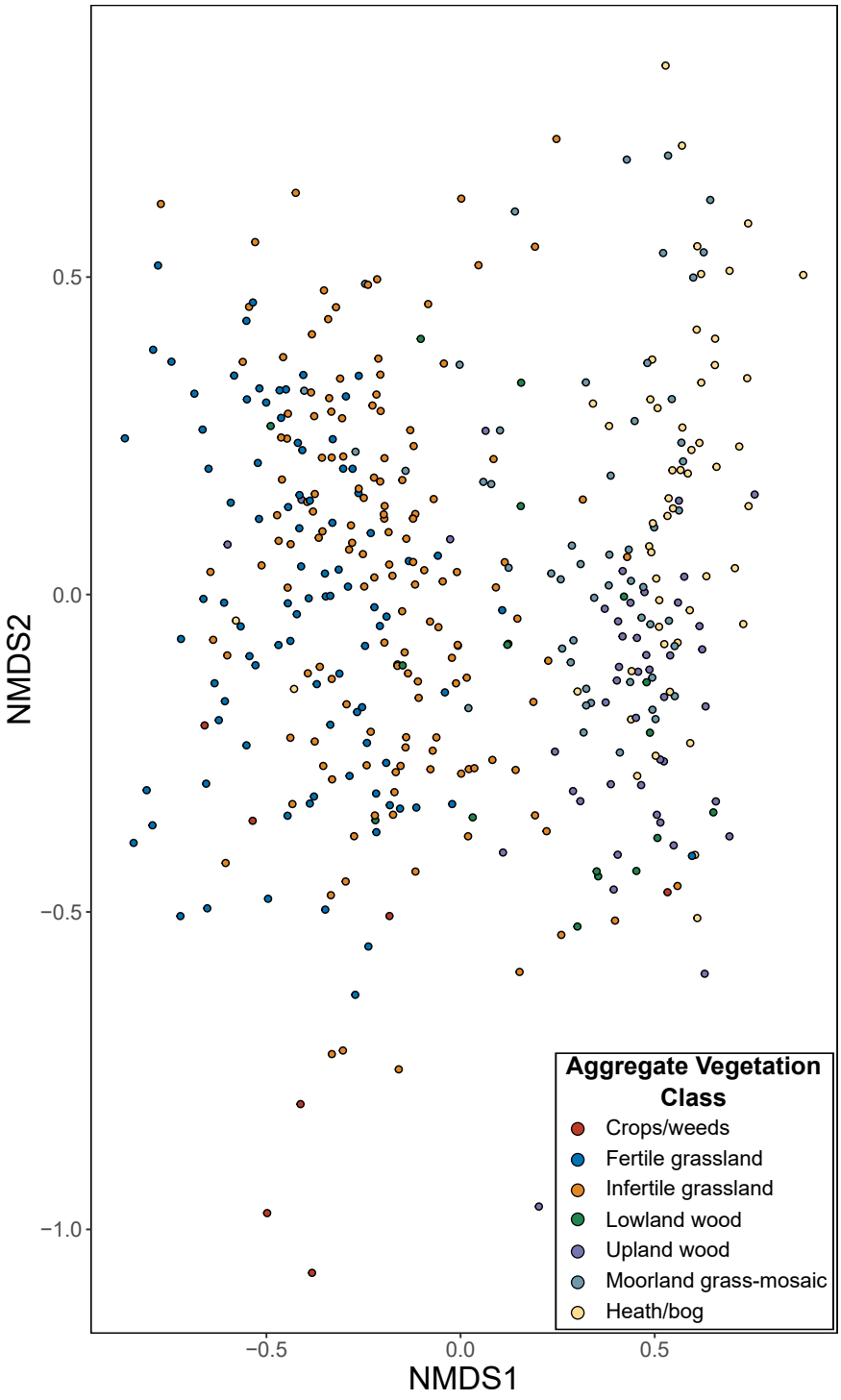


Fig. 5. Plot of the non-metric dimensional scaling ordination (stress = 0.19) of animal community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. The PERMANOVA was significant ($F_{6,401}$ = 7.4, p = 0.001) but not significant differences in dispersion of variances ($F_{6,401}$ = 8.52, p = 0.58) were observed.

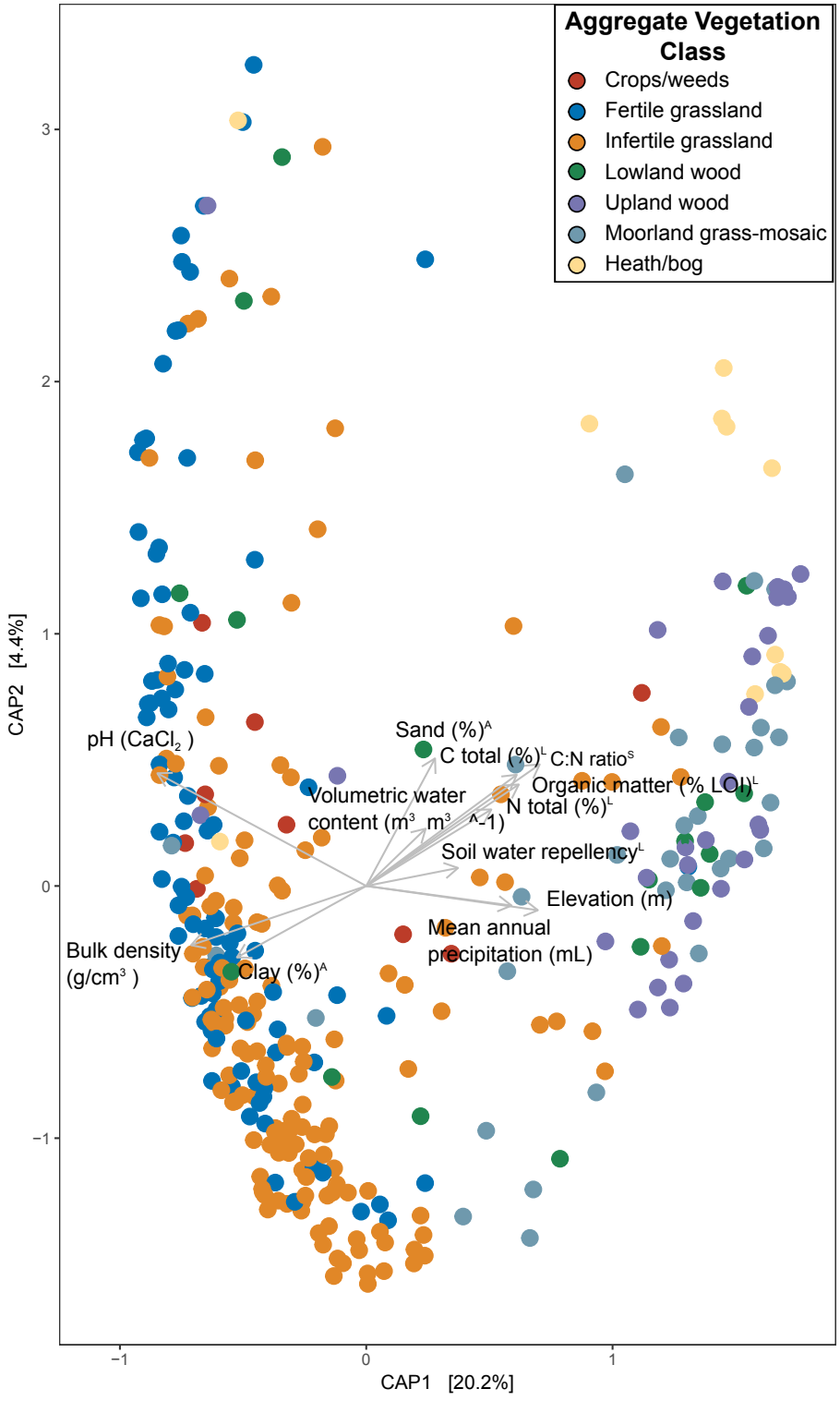


Fig. 6. Vector-loading plot of the canonical analysis of principle coordinates constrained ordination of bacterial community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Only variables with $R^2 > 0.2$ from linear fitting were mapped on this ordination.

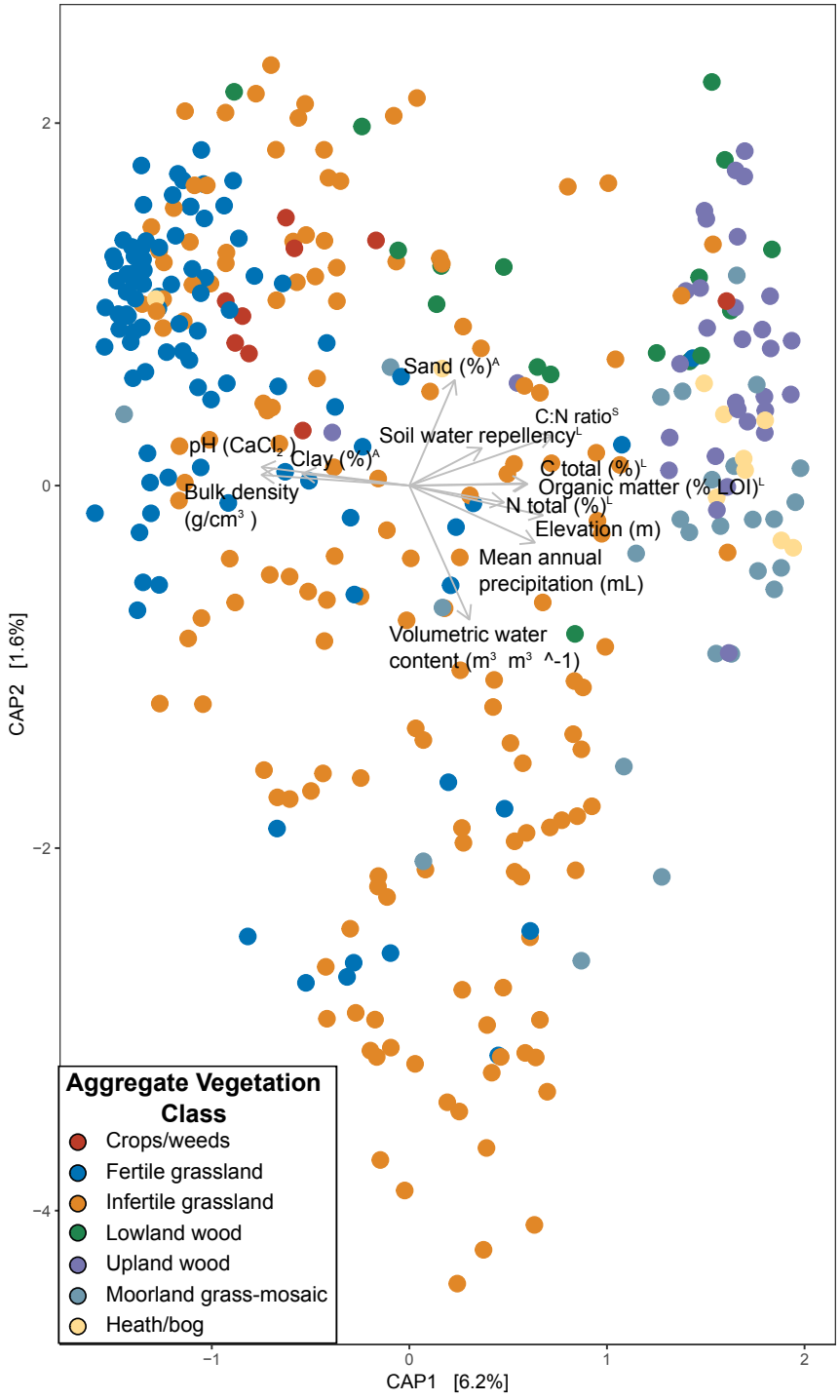


Fig. 7. Vector-loading plot of the canonical analysis of principle coordinates constrained ordination of fungal community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Only variables with $R^2 > 0.2$ from linear fitting were mapped on this ordination.

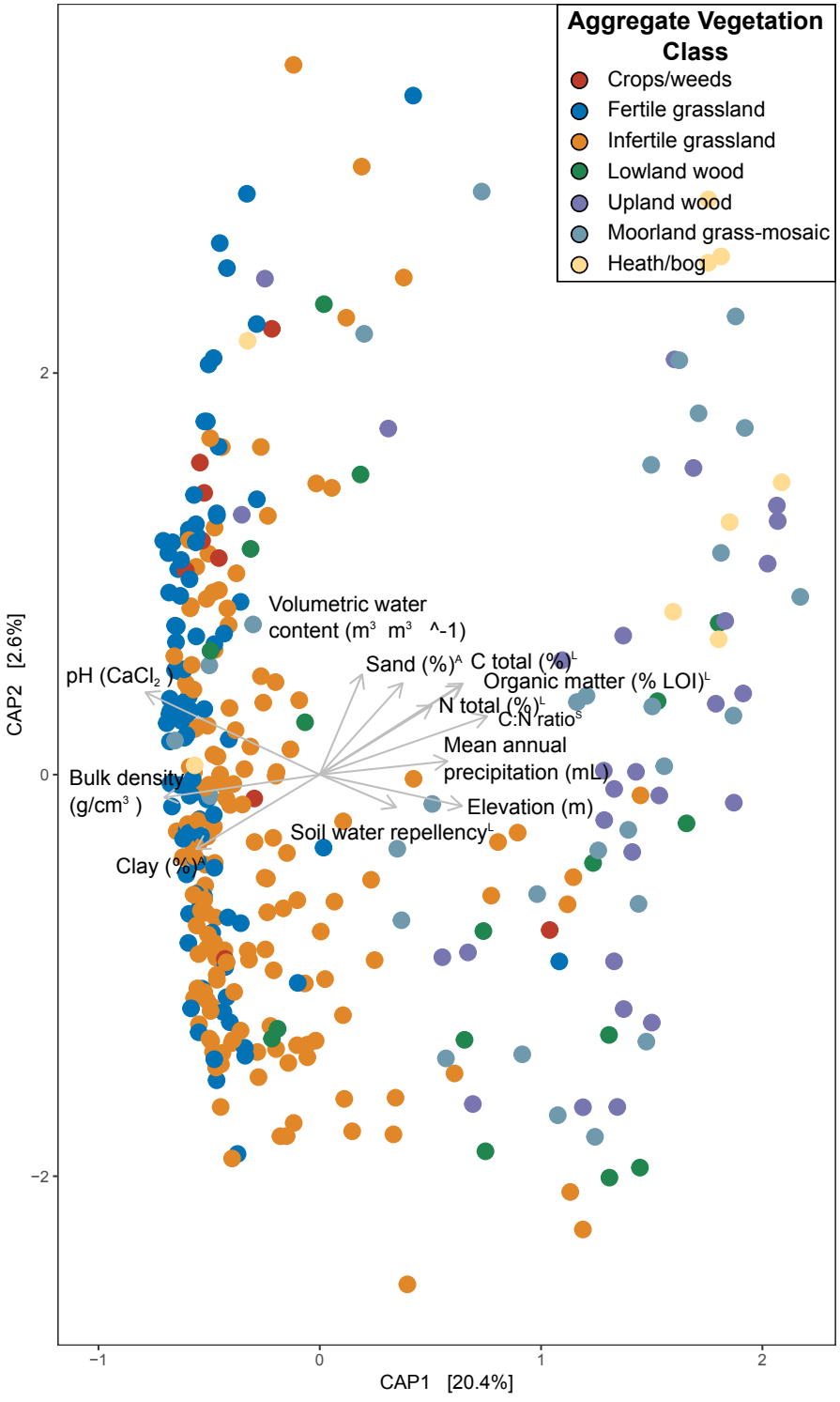


Fig. 8. Vector-loading plot of the canonical analysis of principle coordinates constrained ordination of protistan community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Only variables with $R^2 > 0.2$ from linear fitting were mapped on this ordination.

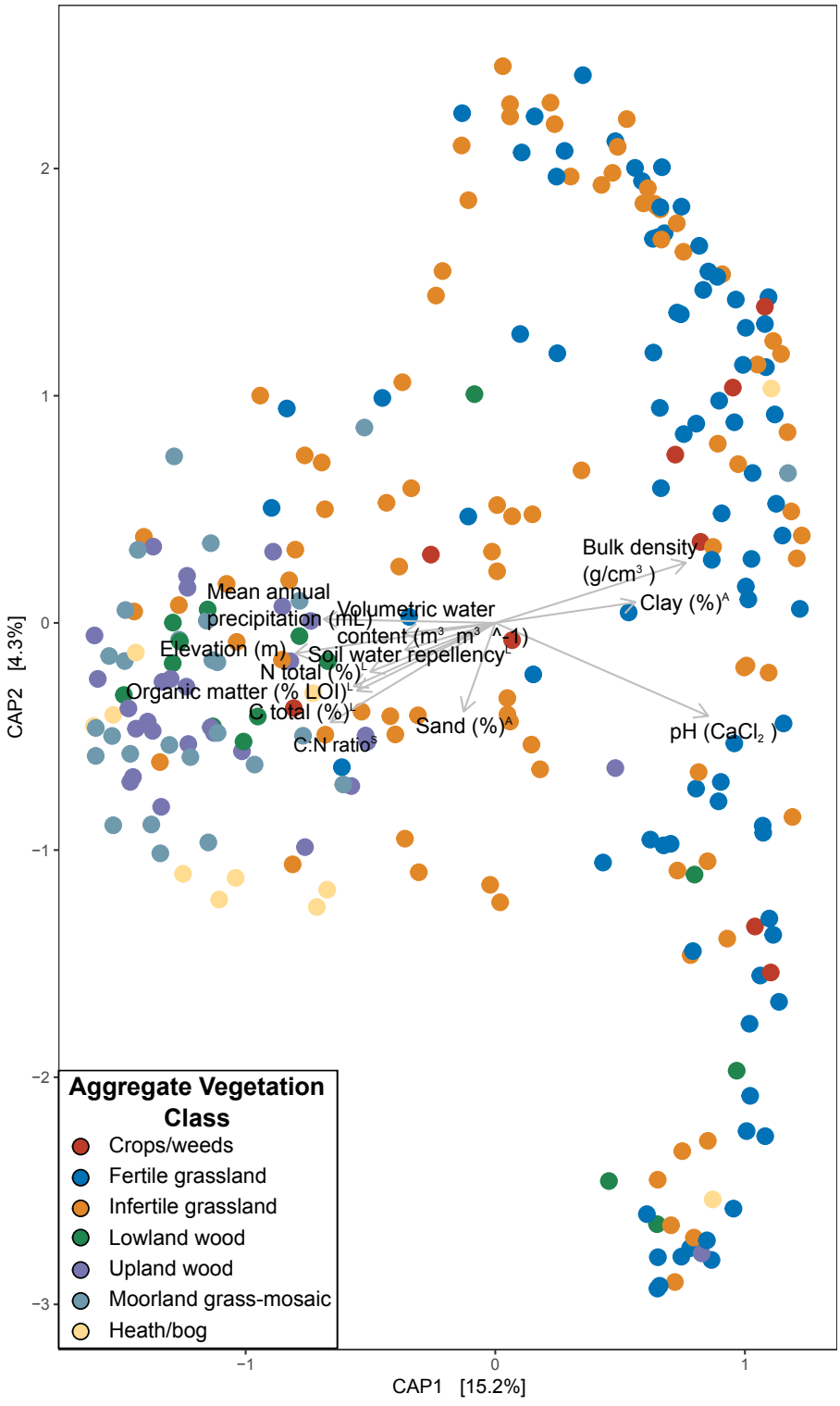


Fig. 9. Vector-loading plot of the canonical analysis of principle coordinates constrained ordination of archaeal community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Only variables with $R^2 > 0.2$ from linear fitting were mapped on this ordination.

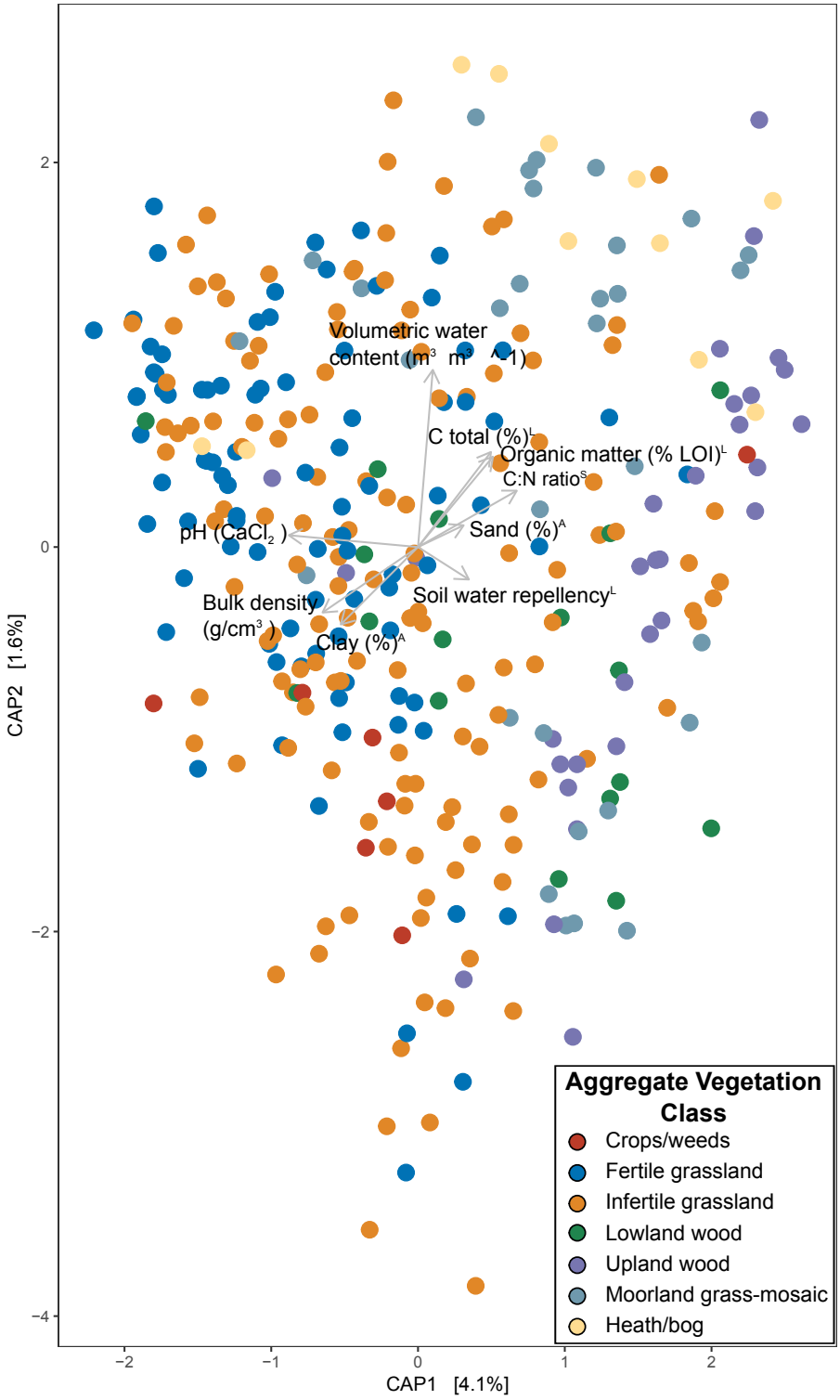


Fig. 10. Vector-loading plot of the canonical analysis of principle coordinates constrained ordination of animal community composition across GMEP sites. Samples are coloured by Aggregate Vegetation Class. Only variables with $R^2 > 0.2$ from linear fitting were mapped on this ordination.

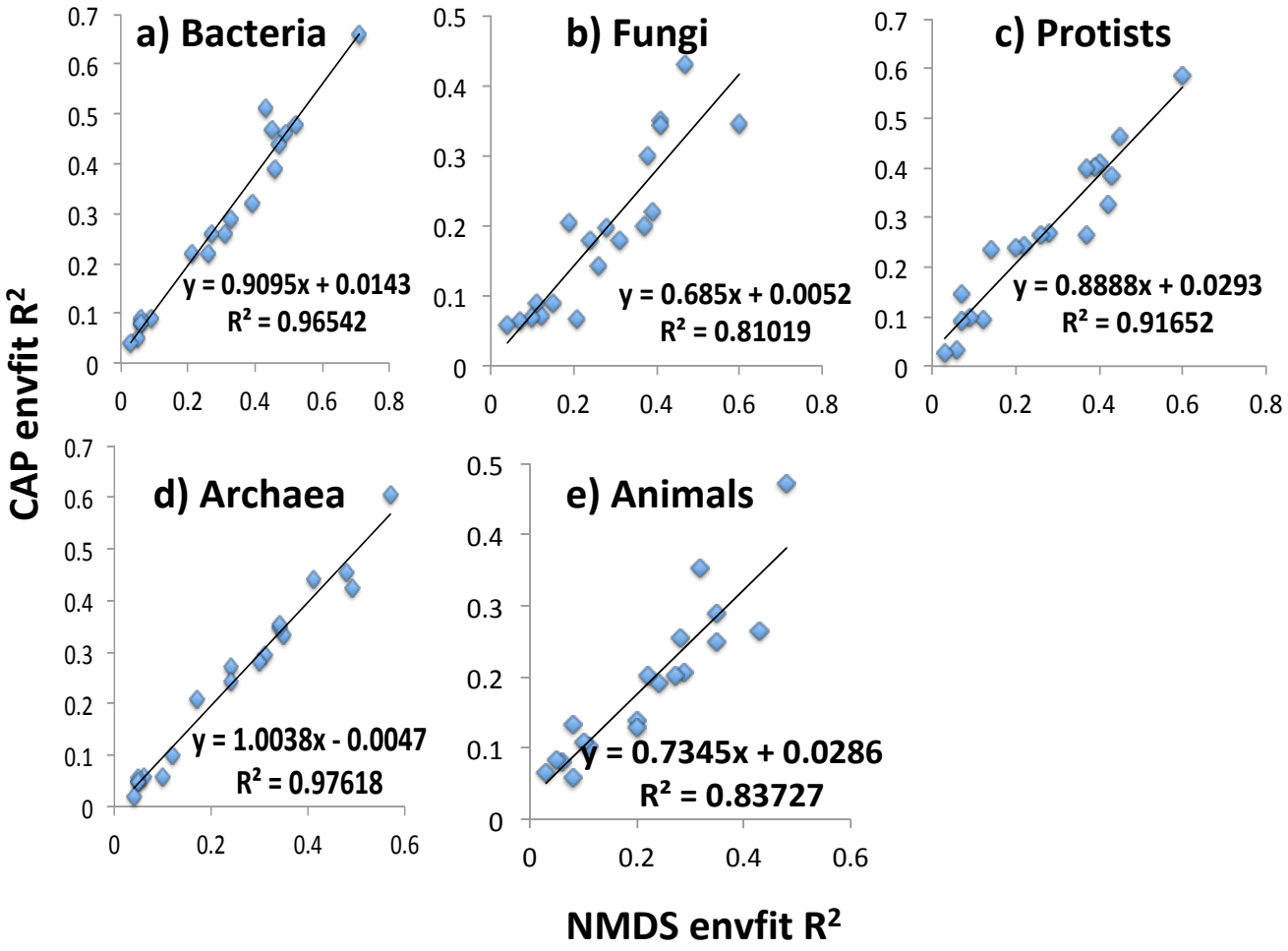


Fig. 11. Regressions of goodness-of-fit values (R^2) of environmental variables calculated from linear fitting to NMDS ordinations versus those from CAP ordinations. Equations and R^2 values are shown for **a**) bacteria; **b**) fungi, **c**) protists; **d**) archaea; and **e**) animals.

Supplementary References

1. Hill, M.O. TWINSPAN – a FORTRAN program for arranging multivariate data in an ordered two-way table by classification of the individuals and attributes. (Cornell University, Ithaca, 1979).

2. Hill, M.O. DECORANA – a FORTRAN program for detrended correspondence analysis and reciprocal averaging (Cornell University, Ithaca, 1979).

3. Bunce, R.G.H. et al. Vegetation of the British countryside – the countryside vegetation system. (Department of the Environment, Transport and the Regions, London, 1990).

4. R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria (2017). https://www.R-project.org/.

5. Weiner, J. riverplot: Sankey or Ribbon Plots. R package version 0.6 https://CRAN.R-project.org/package=riverplot (2017).

6. McMurdie, P. J. & Holmes, S. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS ONE* **8**, e61217 (2013). DOI: 10.1371/journal.pone.0061217

7. Emmett, B. A. & the GMEP Team. Glastir Monitoring & Evaluation Programme. Final Report to Welsh Government (Contract reference: C147/2010/11. NERC/Centre for Ecology & Hydrology (CEH Project: NEC04780/NEC05371/NEC05782) (2017).

8. Cranfield University. The National Soil Map and Soil Classification. Cranfield Soil and Agrifood Institute, Cranfield University, Bedfordshire, UK (2004).

9. Avery, B.W. Soil classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden, UK (1980).

10. Clayden, B. & Hollis, J.M. Criteria for Differentiating Soil Series. Soil Survey Technical Monograph No 17. Harpenden, UK (1984).

11. Smart, S.M., Robertson, J.C., Shield, E.J., Van de Poll, H.M. Locating eutrophication effects across British vegetation between 1990 and 1998. *Glob. Change Biol.* **9**, 1763-1774 (2003).

Appendix 3

Supplementary Material for Chapter 5

George, P.B.L., Creer, S., Griffiths, R.I., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L., 2019. Primer and database choice affect fungal functional but not biological diversity findings in a national soil survey. *Frontiers in Environmental Science* **7**:173. doi: 10.3389/fens.2019.00173

Appendix 3.1 Soil maps in Wales

The soils at each sampling point were assigned to soil type using the National Soil Map and Soil Classification system (Cranfield University, 2004). This map and classification scheme is derived from Avery (1980) with revisions from Clayden and Hollis (1984). Soils were assigned to groups based on published soil maps and reconnaissance mapping of previously unsurveyed sites (for more detail see Cranfield University, 2004). Generally, soils in Wales are known to map poorly due to the high level of local spatial heterogeneity. We analysed soils at the major soil group level. There were 6 soil types that appeared across the GMEP sampling, which have been listed in increasing order of approximate moisture content throughout this study. They are: lithomorphic (ITS1 n = 13; 18S n = 13), brown (ITS1 n = 155; 18S n = 155), podzolic (ITS1 n = 109; 18S n = 113), surface-water gleys (ITS1 n = 80; 18S n = 82), ground-water gleys (ITS1 n = 12; 18S n = 11), and peats (ITS1 n = 44; 18S n = 48). In addition, each soil sample was categorised into an organic matter class based on loss-on-ignition (LOI) following the protocols of the 2007 Countryside Survey (Emmett et al., 2010) into four categories: mineral (0-8% LOI), humus-mineral (8-30% LOI), organo-mineral (30-60% LOI), and organic (60-100% LOI). The categories are listed in order of increasing

organic matter content as follows: mineral (ITS1 n = 103; 18S n = 104), humus-mineral (ITS1 n = 228; 18S n = 232), organo-mineral (ITS1 n = 22; 18S n = 26), and organic (ITS1 n = 59; 18S n = 59).

Appendix 3.2 Pairwise differences of fungal OTU richness between land uses, soil organic matter, and soil type

As previously demonstrated in George et al. (2019), fungal OTU richness from ITS1 metabarcoding significantly ($F_{6, 258}$ = 39.87, $p < 0.001$; Fig. 5.3A) declined from high to low productivity/management intensity. Fungal richness in Fertile grasslands was significantly greater than all other AVCs ($p < 0.001$) except Crops/weeds. Richness in Crops/weeds and Infertile grassland were significantly greater than all other land uses (all $p < 0.01$, except for Lowland wood $p = 0.002$ and $p = 0.01$, respectively). Fungal richness in Heath/bog was also significantly lower than that of Moorland grass-mosaics ($p < 0.001$) and both Lowland ($p = 0.02$) and Upland ($p = 0.007$) woodland AVCs.

In the 18S dataset, richness was also significantly ($F_{6,267}$ = 82.73, $p < 0.001$) higher in more productive/managed land uses and declined along this gradient. Significantly greater fungal richness was observed in grassland AVCs than in all other AVCs ($p <0.001$ for all but Infertile grassland-Lowland wood $p = 0.03$) except Crops/weeds in the 18S dataset. Richness in Crops/weeds was also greater than Upland wood ($p = 0.02$). Heath/bog had the lowest richness, which was significantly lower than all other AVCs ($p < 0.001$). In addition, richness in both Crops/weeds, Lowland wood, and Upland wood was significantly greater than Moorland grass-mosaic ($p < 0.001$, $p = 0.002$, and $p = 0.04$, respectively).

For soil organic matter content, richness from both datasets richness was significantly greater ($F_{3, 259}$ = 48.13, $p < 0.001$; $F_{3, 269}$ = 46.71, $p < 0.001$; for ITS1 and

18S, respectively) in mineral and humus-mineral than all other classifications (Fig. 5.4). Specifically, fungal richness in mineral and humus-mineral soils was greater than that of organic (all $p < 0.001$) as well as organo-mineral soils ($p = 0.03$ and $p < 0.001$, for ITS1 and 18S, respectively). In both datasets, richness in organic soils was lower than that of organo-mineral soils ($p < 0.001$).

In the ITS1 dataset, fungal richness was significantly ($F_{5, 258} = 10.8$, $p < 0.001$) lower in peats than brown, podzolic, and both surface-water (all $p < 0.001$) and ground-water gley ($p = 0.002$) soils when soil type was assessed (Appendix 3 Fig. 3A). However, aside from greater richness in brown soils than podzolic ($p = 0.009$) and surface-water gley ($p = 0.04$) soils, other differences were not apparent. A similar trend ($F_{5, 268} = 14.4$, $p < 0.001$) was observed in the 18S dataset (Appendix 3 Fig. 3B). Here, fungal richness was again lower in Peats compared to brown, podzolic, ground- and surface-water gley soils ($p < 0.001$). Brown soils had the highest fungal richness, which was also greater than that of podzolic ($p < 0.001$) and lithomorphic ($p = 0.02$) soils.

Across land uses, significant differences were observed in the richness of saprotrophic fungi in both the ITS1 ($F_{6,258} = 25.14$, $p < 0.001$) and 18S ($F_{6, 267} = 31.10$, $p < 0.001$) data; however, there were differences between datasets (Fig. 5.8). In the ITS1 dataset, richness followed the same trend as overall fungal richness, with the highest and lowest values in the Crops/weeds and Heath/bog AVCs respectively (Fig. 8A). Saprotroph richness was significantly greater in Crops/weeds than Infertile grassland ($p = 0.001$), Lowland wood, Upland wood, Moorland grass-mosaic, and Heath/bog (all $p < 0.001$). Similarly, saprotroph richness was significantly greater than all groups except Lowland wood (all $p < 0.001$). Richness of saprotophs in Moorland grass-mosaic and

Heath/bog sites was also significantly lower than that of Infertile grassland areas (both $p < 0.001$). Although this pattern was preserved in the 18S data, richness of saprotrophs was much more even in this case. Indeed, rather than the linear decline of richness along the productivity gradient, there appeared to be 3 distinct levels in the data associated with grassland/agricultural sites, woodlands, and bogs. Saprotroph richness was significantly lower in Heath/bog sites than all other AVCs (all $p < 0.001$) and highest in grasslands. There were significant differences between saprotroph richness in grasslands and Moorland grass-mosaic and wood AVCs (all $p < 0.001$ except Fertile grassland – Lowland wood $p = 0.046$ and Infertile grassland – Lowland wood $p = 0.001$).

In the ITS1 dataset, each organic matter class was significantly ($F_{3, 260} = 32.86$, $p < 0.001$) different from the others (Fig. 5.9). Mineral soils had the highest saptrotroph richness when compared to all others ($p < 0.001$ except for humus-mineral $p = 0.003$). Saprotroph richness in humus-mineral soils was greater than both organo-mineral ($p = 0.03$) and organic soils ($p < 0.001$) and richness in organo-mineral soils greater than organic ($p = 0.03$) soils (Fig. 5.9A). In the 18S data, saprotroph richness was significantly ($F_{3, 269} = 41.13$, $p < 0.001$) higher in mineral and humus-mineral soils than organo-mineral and organic (all $p < 0.001$ except mineral – organo-mineral $p = 0.02$) soils (Fig. 5.9B). Again, the overarching trend of fungal richness was not apparent when samples were grouped by soil type. Although there were significant differences across soil types in both the ITS1 ($F_{5, 259} = 9.7$, $p < 0.001$) and 18S ($F_{5, 268} = 10.73$, $p < 0.001$) datasets, these differences did demonstrate consistent patterns across soil types (Appendix 3 Fig. 10). Richness of saprotrophs mirrored the exact trend as total richness in the ITS1 dataset. Saprotroph richness was significantly lower in peats than brown ($p < 0.001$),

podzolic ($p = 0.045$), and both surface-water and ground-water gley (both $p = 0.01$) soils (Appendix 3 Fig. 10A). Saprotroph richness in was higher in brown soils than Podzolic ($p = 0.009$) and surface-water gley ($p = 0.03$) soils, other differences were not apparent. In the 18S dataset, saprotroph richness was lower than in all other soils except lithomorphic (all $p < 0.001$; Appendix 3 Fig. 10B).

In the ITS1 data, significantly ($F_{6, 258} = 26.11$, $p < 0.001$) greater pathotroph richness values were observed in Crops/weeds and grassland samples in comparison to the other AVC categories (Fig. 5.8A). Richness of pathotrophs was significantly greater in Crops/weeds sites than all other AVCs (all $p < 0.001$, except Fertile grassland $p = 0.02$). Similarly, richness was greater in the Fertile grasslands than all other remaining land uses (all $p < 0.001$, except Infertile grassland $p = 0.002$). Richness of pathrotrophs in Infertile grasslands was also significantly greater than in all remaining AVCs except Heath/bog ($p < 0.001$). Again, this trend was present, though not as stark, in the 18S dataset (Fig. 5.8B). Significant differences ($F_{6, 267} = 52.26$, $p < 0.001$) were observed between AVCs, with the highest richness of pathotrophs occurring in the Fertile grassland and Crop/weeds land uses. Pathotroph richness was greater in Fertile grasslands than all other AVCs (all $p < 0.001$, except Lowland wood $p = 0.001$) but Crops/weeds. Richness in Crops/weeds, Infertile grassland, and Lowland wood samples was greater than Moorland grass-mosaic and (all $p < 0.001$, except Lowland wood – Moorland grass-mosaic $p = 0.002$), Upland wood (all $p < 0.001$, except Lowland wood – Upland wood $p = 0.04$), and Heath/bog (all $p < 0.001$). Additionally, pathotroph richness in Heath/bog sites was lower than Moorland grass-mosaic samples ($p = 0.01$).

When grouped by organic matter class, significant differences were also observed in pathotroph richness in the ITS1 ($F_{3, 250} = 24.91$, $p < 0.001$) and 18S ($_{F3, 269} = 30.49$, $p < 0.001$) datasets. However, in this case the trends were more apparent in the 18S data than the ITS1 data (Chapter 5 Fig. 9). Pathotroph richness was significantly greater in mineral than humus-mineral ($p = 0.03$) soils and was significantly lower in organic soils when compared to all others (all $p < 0.001$) in the ITS1 data (Fig. 5.9A). However, all organic matter classifications were statistically different from each other in the 18S data (Fig. 8B), in descending order from mineral to peat soils (all $p < 0.001$, except organo-mineral – organic $p = 0.03$, humus-mineral – mineral and organo-mineral both $p = 0.001$). Again, trends were less clear across soil types (Appendix 3 Fig. 10). Peat soils had significantly ($F_{5, 259} = 6.93$, $p < 0.001$) lower pathotroph richness than brown ($p < 0.001$), podzolic, ground-water gley (both $p = 0.002$), and surface-water gley ($p = 0.007$) soils (Appendix 3 Fig. 10A) in the ITS1 data. Differences between pathotrophic fungi across soil types were more similar to those observed in other groups in the 18S data (Appendix 3 Fig. 10B). Pathotroph richness was significantly ($F_{5, 268} = 13.6$, $p < 0.001$) greater in brown soils than lithomorphic ($p = 0.03$), podzolic ($p = 0.02$), and peat ($p < 0.001$) soils. Richness in peats was again significantly lower than that of podzolic, surface- and ground-water gley (all $p < 0.001$) soils.

Although significant differences were apparent in both the ITS1 ($F_{6, 258} = 14.88$, $p < 0.001$) and 18S ($F_{6, 267} = 55.13$, $p < 0.001$) datasets they were by no means identical (Fig. 5.7). Symbiotroph richness was higher in Lowland wood sites in the ITS1 than all other AVCs (all $p < 0.001$ except Upland wood $p = 0.046$. Symbiotroph richness was also higher in Upland wood sites than the Infertile grassland, Moorland grass-mosaic, and

Heath/bog AVCs (all $p < 0.001$; Fig. 5.8A). This trend was not apparent in the 18S dataset however (Chapter 5 Fig. 8B). Here richness of symbiotrophs was significantly greater in grasslands AVCs than all other AVCs (all $p < 0.001$). Similarly, richness of symbiotrophs from Heath/bog sites was significantly lower than those of Lowland wood ($p = 0.02$), Upland wood, Crops/weeds, and Moorland grass-mosaic (all $p < 0.001$).

Across organic matter classes, the previously described trend of decreasing richness with increasing organic matter content held true in the 18S data ($F_{3,\ 269} = 36.28$, $p < 0.001$; Fig. 8B), with no significant differences observed in the ITS1 dataset ($F_{3,\ 260} = 1.88$, $p = 0.13$; Fig 5.9A). In the 18S data, richness of symbiotrophs was significantly greater in mineral and humus-mineral soils when compared to organo-mineral ($p = 0.002$, $p = 0.04$, respectively) and organic ($p < 0.001$) soils (Fig. 5.9B). There were also no significant differences ($F_{5,\ 259} = 1.43$, $p = 0.21$) in symbiotroph richness across soil types in ITS1 data (Appendix 3 Fig. 10A), though there were in 18S data ($F_{5,\ 259} = 12.52$, $p < 0.001$; Fig. S#B). As described previously, in this case richness was lower in peat soils than in ground-water gley ($p = 0.02$), surface-water gley, podzolic, and brown (all $p < 0.001$) soils. Additionally, symbiotroph richness was higher in brown soils than in podzolic ($p = 0.02$) and lithomorphic ($p = 0.046$) soils.

There were significant ($F_{6,\ 244} = 33.47$, $p < 0.001$) differences in richness of Glomeromycota across land uses, though they appeared, like the saprotroph richness to be tiered between grasslands, woods, and bogs (Fig. 5.10A). Richness of Glomeromycetes was higher in Fertile and Infertile grasslands than all other AVCs (all $p < 0.001$, except Moorland grass-mosaic $p = 0.04$ and $p = 0.008$, respectively) except Crops/weeds. Richness in Heath/bog sites was significantly lower than Moorland grass-

mosaic, Lowland wood (both $p < 0.001$), and Upland wood ($p = 0.01$). In addition, significant differences were observed between Upland wood and Moorland-grass ($p < 0.001$).

Again, when grouped by organic matter class (Fig. 5.10B) and soil type (Fig. 5.10C) Glomeromycetes richness followed the same trend saprotrophs and symbiotrophs from the 18S dataset. Richness was significantly ($F_{3, 246} = 37.65$, $p < 0.001$) greater in mineral and humus-mineral soils than all other organic matter classes ($p < 0.001$). Richness was also lower in organic soils than organo-mineral soils ($p = 0.002$). Across soil types, richness of Glomeromycetes was significantly ($F_{5, 245} = 8.65$, $p < 0.001$) lower in peat soils when compared to brown, podzolic, surface-water (all $p < 0.001$) and ground-water gley ($p = 0.004$) soils.

References

Avery, .W. (1980). Soil classifications for England and Wales (Higher Categories). Soil Survey Technical Monograph No 14. Harpenden, UK.

Clayden, B. and Hollis, J. M. (1984). Criteria for Differentiating Soil Series. Soil Survey Technical Monograph No 17. Harpenden, UK.

Cranfield University. (2004). The National Soil Map and Soil Classification. Cranfield Soil and Agrifood Institute, Cranfield University, Bedfordshire, UK.

Emmett, B. A., Reynolds, B., Chamberlain, P. M., Rowe, E., Spurgeon, D., Brittain, S. A., Frogbrook, Z., Hughes, S., Lawlor, A. J., Poskitt, J., Potter, E., Robinson, D. A., Scott, A., Wood, C., and Woods, C. (2010). Countryside Survey: Soils Report from 2007. NERC/Centre for Ecology and Hydrology (CS Technical Report No. 9/07, CEH Project Number: C03259).

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L. (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. Nature Communications 10:1107.

Supplementary Tables

Table 1. Description of Aggregate Vegetation Classes identified in this study. Adapted from Smart et al. (2003).

Aggregate Vegetation Class	Description
Crops/weeds (ITS1 n = 9; 18S n = 8)	Communities on disturbed or cultivated land, including weedy, horticultural, and species-poor arable land.
Fertile grassland (ITS1 n = 97; 18S n = 96)	Improved or semi-improved grassland. Usually with high nutrient inputs and cut more than once a year.
Infertile grassland (ITS1 n = 156; 18S n = 158)	Semi-improved to unimproved, less productive grasslands, species-rich grasslands including wet or dry and acidic to basic variations.
Lowland wood (ITS1 n = 17, 18S n = 16)	Dominated by trees and shrubs in neutral or basic lowlands, scrublands, and hedgerows.
Upland wood (ITS1 n = 43; 18S n = 43)	Commonly acidic conifer plantations, scrubland and semi-natural broadleaved woods in the uplands.
Moorland grass/mosaic (ITS1 n = 44; 18S n = 53)	Grass-dominated upland pasture, commonly with a long history of livestock grazing.
Heath/bog (ITS1 n = 47; 18S n = 48)	Heather dominated, commonly upland landscapes, including dry heath and bogs.

Table 2. Mean values (± SE) of soil physical and chemical variables of each organic matter class. Following normalisation on selected variables (see below) ANOVAs and Tukey's *post-hoc* tests were performed. Results are as follows: total C ($F_{3\ 431} = 613.22$ $p < 0.001$), total N ($F_{3\ 431} = 564.38$, $p < 0.001$), C :N ratio ($F_{3,\ 431} = 175.81$, $p < 0.001$), total P ($F_{3,\ 428} = 8.46$, $p < 0.001$), organic matter ($F_{3,\ 432} = 1358.5$, $p < 0.001$), pH ($F_{3,\ 432} = 83.53$, $p < 0.001$), soil water repellency ($F_{3\ 432} = 41.39$, $p < 0.001$), volumetric water content ($F_{3,\ 431} = 61.93$, $p < 0.001$), soil bound water ($F_{3\ 432} = 626.58$, $p < 0.001$), rock volume ($F_{3,\ 431} = 19.55$, $p < 0.001$), bulk density ($F_{3,\ 431} = 485.08$, $p < 0.001$), clay content ($F_{3,\ 347} = 44.86$, $p < 0.001$), sand content ($F_{3,\ 347} = 21.56$, $p < 0.001$), elevation ($F_{3,\ 432} = 100.34$, $p < 0.001$), mean annual precipitation ($F_{3,\ 432} = 69.38$, $p < 0.001$), and temperature ($F_{3\ 432} = 0.69$, $p = 0.56$).

Environmental variable	Mineral	Humus-mineral	Organo-mineral	Organic
Total C (%)^{L}	3.14 (± 0.10)d	6.80 (± 0.20)c	39.38 (± 1.23)a	25.27 (± 0.76)b
Total N (%)^{L}	0.29 (± 0.01)d	0.55 (± 0.01)c	1.39 (± 0.06)b	1.83 (± 0.06)a
C:N ratio^{S}	11.14 (± 0.25)d	12.22 (± 0.22)c	18.5 (± 0.89)b	22.1717 (± 0.88)a
Total P (mg/kg)^{S}	855.67 (± 36.98)b	1085.32 (± 34.16)a	1208.15 (± 96.40)a	915.82 (± 47.6)a
Organic matter (% LOI)^{L}	6.18 (± 0.13)d	13.06 (± 0.31)c	44.2 (± 1.58)b	75.64 (± 1.31)a
pH ($CaCl_2$)	5.04 (± 0.08)a	4.55 (± 0.05)b	3.46 (± 0.16)c	3.46 (± 0.06)d
Soil water repellency*	208.08 (± 42.02)c	1589.51 (± 292.96)b	3757.2 (± 688.25)a	3939.5 (± 759.14)a
Volumetric water content (m^3/m^3)	0.29 (± 0.01)c	0.36 (± 0.01)b	0.47 (± 0.04)a	0.65 (± 0.02)a
Soil bound water (g water per g of dry soil)^{L}	2.0 (± 0.06)d	3.26 (± 0.07)c	7.3 (± 0.47)b	9.18 (± 0.28)a
Rock volume (mL)	4.12 (± 0.39)b	5.82 (± 0.36)a	2.84 (± 0.69)bc	1.11 (± 0.2)c
Bulk density (g/cm^3)	1.02 (± 0.01)a	0.71 (± 0.01)b	0.24 (± 0.02)c	0.13 (± 0.01)d
Clay content (%)^{A}	20.43 (± 0.67)a	24.2 (± 0.52)a	10.49 (± 0.88)b	8.14 (± 4.94)c
Sand content (%)^{A}	36.68 (± 1.94)b	25.56 (± 0.97)c	45.24 (± 3.81)b	66.08 (± 12.4)a
Elevation (m)	106.04 (± 8.5)d	201.39 (± 9.18)c	355.3 2 (± 34.17)a	325.18 (± 13.45)b
Mean annual precipitation (mL)	1072.06 (± 29.86)c	1279.39 (± 27.91)b	1884.47 (± 93.43)a	1806.5 (± 39.78)a
Temperature (°C)	13.79 (± 0.41)a	13.14 (± 0.23)a	14.22 (± 0.57)a	12.07 (± 0.3)a

Note: ^{A} denotes Aitchison's log_{10}-ratio transformation; ^{L} denotes log_{10}-transformation; square-root-transformation; *Soil water repellency was derived from median water drop penetration times (s) and log_{10} transformed.

Table 3. Mean values (± SE) of soil physical and chemical variables of each soil type. Following normalisation on selected variables (see below) ANOVAs and Tukey's *post-hoc* tests were performed. Results are as follows: total C ($F_{5, 429} = 52.5$ $p < 0.001$), total N ($F_{5, 429} = 43.8$, $p < 0.001$), C :N ratio ($F_{5, 429} = 38.12$, $p < 0.001$), total P ($F_{5, 426} = 1.89$, $p = 0.1$), organic matter ($F_{5, 430} = 61.01$, $p < 0.001$), pH ($F_{5, 430} = 34.51$, $p < 0.001$), soil water repellency ($F_{5, 430} = 10.16$, $p < 0.001$), volumetric water content ($F_{5, 429} = 23.07$, $p < 0.001$), soil bound water ($F_{5\ 430} = 56.94$, $p < 0.001$), rock volume ($F_{5, 429} = 7.31$, $p < 0.001$), bulk density ($F_{5, 429} = 48.6$, $p < 0.001$), clay content ($F_{5, 346} = 4.18$, $p = 0.01$), sand content ($F_{5, 346} = 3.42$, $p = 0.01$), elevation ($F_{5, 431} = 61.73$, $p < 0.001$), mean annual precipitation ($F_{5, 431} = 43.76$, $p < 0.001$), and temperature ($F_{5, 431} = 1.38$, $p = 0.23$).

Environmental variable	Lithomorphic	Brown	Podzolic	Surface-water gleys	Ground-water gleys	Peats
Total C (%)^{L}	12.55 (± 4.98)b	4.68 (± 0.33)d	9.0 (± 1.31)b	7.62 (± 1.12)c	8.26 (± 1.68)bcd	9.74 (± 2.46)a
Total N (%)^{L}	0.69 (± 0.26)d	0.41 (± 0.02)d	0.65 (± 0.05)d	0.55 (± 0.06)d	0.68 (± 0.08)c	0.61 (± 0.1)b
C:N ratio^{S}	15.66 (± 1.39)b	11.21 (± 0.24)c	12.98 (± 0.52)b	12.92 (± 0.52)b	11.56 (± 1.4)bc	15.5 (± 0.9)a
Total P (mg/kg)^{S}	1027.75 (± 191.25)a	1043.18 (± 37.88)a	1146.16 (± 50.2)a	879.98 (± 46.33)a	974.55 (± 76.97)a	699.94 (± 59.7)a
Organic matter (% LOI)^{L}	22.98 (± 8.73)b	9.27 (± 0.67)d	17.16 (± 0.56)b	13.76 (± 1.7)c	15.27 (± 4.22)bcd	18.7 (± 3.72)a
pH ($CaCl_2$)	4.33 (± 0.22)bcd	4.89 (± 0.07)a	4.21 (± 0.07)b	4.73 (± 0.1)ab	5.05 (± 0.29)a	3.96 (± 0.1)d
Soil water repellency^{*L}	1261.43 (± 1035.2)ac	651.35 (± 144.0)c	1623.78 (± 419.91)ab	1329.46 (± 332.21)bc	4203.08 (± 2721.78)ab	3941.17 (± 1025.53)a
Volumetric water content (m^3/m^3)	0.47 (± 0.05)ac	0.31 (± 0.01)d	0.37 (± 0.02)c	0.38 (± 0.02)bc	0.4 (± 0.04)bcd	0.43 (± 0.02)a
Soil bound water (g water per g of dry soil)	3.98 (± 0.89)b	2.66 (± 0.1)d	3.62 (± 0.29)bc	3.15 (± 0.25)cd	4.26 (± 0.54)bd	3.68 (± 0.51)a
Rock volume (mL)	4.05 (± 0.92)ab	5.52 (± 0.43)a	6.21 (± 0.47)a	3.49 (± 0.49)b	5.0 (± 1.34)ab	2.85 (± 0.32)b
Bulk density (g/cm^3)	3.98 (± 0.09)cd	2.66 (± 0.02)b	3.62 (± 0.03)c	3.15 (± 0.03)a	4.26 (± 0.08)bcd	3.68 (± 0.04)d
Clay content (%)^{A}	15.27 (± 5.33)bc	21.99 (± 0.57)b	22.76 (± 0.92)ac	22.9 (± 0.92)ab	26.28 (± 3.18)ac	19.29 (± 3.17)c
Sand content (%)^{A}	51.49 (± 12.5)a	32.01 (± 1.41)b	27.33 (± 1.8)b	29.09 (± 2.01)b	25.33 (± 2.96)b	27.89 (± 7.52)b
Elevation (m)	219.18 (± 39.49)bc	125.19 (± 7.91)d	268.59 (± 14.32)b	177.78 (± 15.52)c	33.57 (± 15.72)d	363.58 (± 17.46)a
Mean annual precipitation (mL)	1643.0 (± 144.42)ab	1093.36 (± 14.6)d	1525.41 (± 46.68)b	1223.14 (± 48.45)c	949.58 (± 34.2)b	1352.08 (± 66.26)a
Temperature (°C)	15.7 (± 0.63)a	13.24 (± 0.36)a	13.62 (± 0.29)a	13.05 (± 0.33)a	15.2 (± 0.94)a	11.71 (± 0.31)a

Note: ^{A} denotes Aitchison's $\log_{10}$-ratio transformation; ^{L} denotes $\log_{10}$-transformation; square-root-transformation; *Soil water repellency was derived from median water drop penetration times (s) and $\log_{10}$ transformed.

Table 4. Richness of OTUs at the class-level that appear in both the ITS1 and 18S datasets.

Class	Number of OTUs in ITS1	Number of OTUs in 18S
Agaricomycetes	1858	646
Agaricostilbomycetes	1	36
Archaeorhizomycetes	6	129
Arthoniomycetes	1	5
Chytridiomycetes	2	1001
Cystobasidiomycetes	2	2
Dothideomycetes	326	91
Eurotiomycetes	472	86
Exobasidiomycetes	4	40
Geoglossomycetes	7	1
Glomeromycetes	2	162
Lecanoromycetes	22	29
Leotiomycetes	422	66
Microbotryomycetes	40	65
Monoblepharidomycetes	5	27
Orbiliomycetes	31	6
Pezizomycetes	79	49
Pucciniomycetes	2	29
Saccharomycetes	23	106
Sordariomycetes	915	417
Tremellomycetes	25	181
Ustilaginomycetes	3	7

Table 5. Richness of OTUs at the class-level that appear in only the ITS1 dataset.

Class	Number of OTUs
Archaeosporomycetes	1
Calcarisporiellomycetes	6
Endogonomycetes	19
Geminibasidiomycetes	2
GS17	1
Kickxellomycetes	1
Malasseziomycetes	23
Mortierellomycetes	128
Mucoromycetes	55
Mucoromycotiunidentified_cls_Incertae_sedis	2
Paraglomeromycetes	3
Rhizophlyctidomycetes	5
Rhizophydiomycetes	11
Spizellomycetes	8
Umbelopsidomycetes	17
Unidentified	1125
Xylonomycetes	10
Zoopagomycetes	2

Supplementary Table 6. Richness of OTUs at the class-level that appear in only the 18S dataset.

Class	Number of OTUs
Ambiguous taxa*	208
Atractiellomycetes	8
Basidomycetes	3
Dacrymycetes	6
Incertae Sedis	715
Laboulbeniomycetes	17
Lichinomycetes	5
LKM11	182
Microsporidia	1
Neocallimastigomycetes	6
Pneumocystidomycetes	1
Schizosaccharomycetes	3
Taphrinomycetes	9

*This includes OTUs identified as “ambiguous taxa”, “Amb-18S-784”, and “Acomycota sp. MUT 4926”.

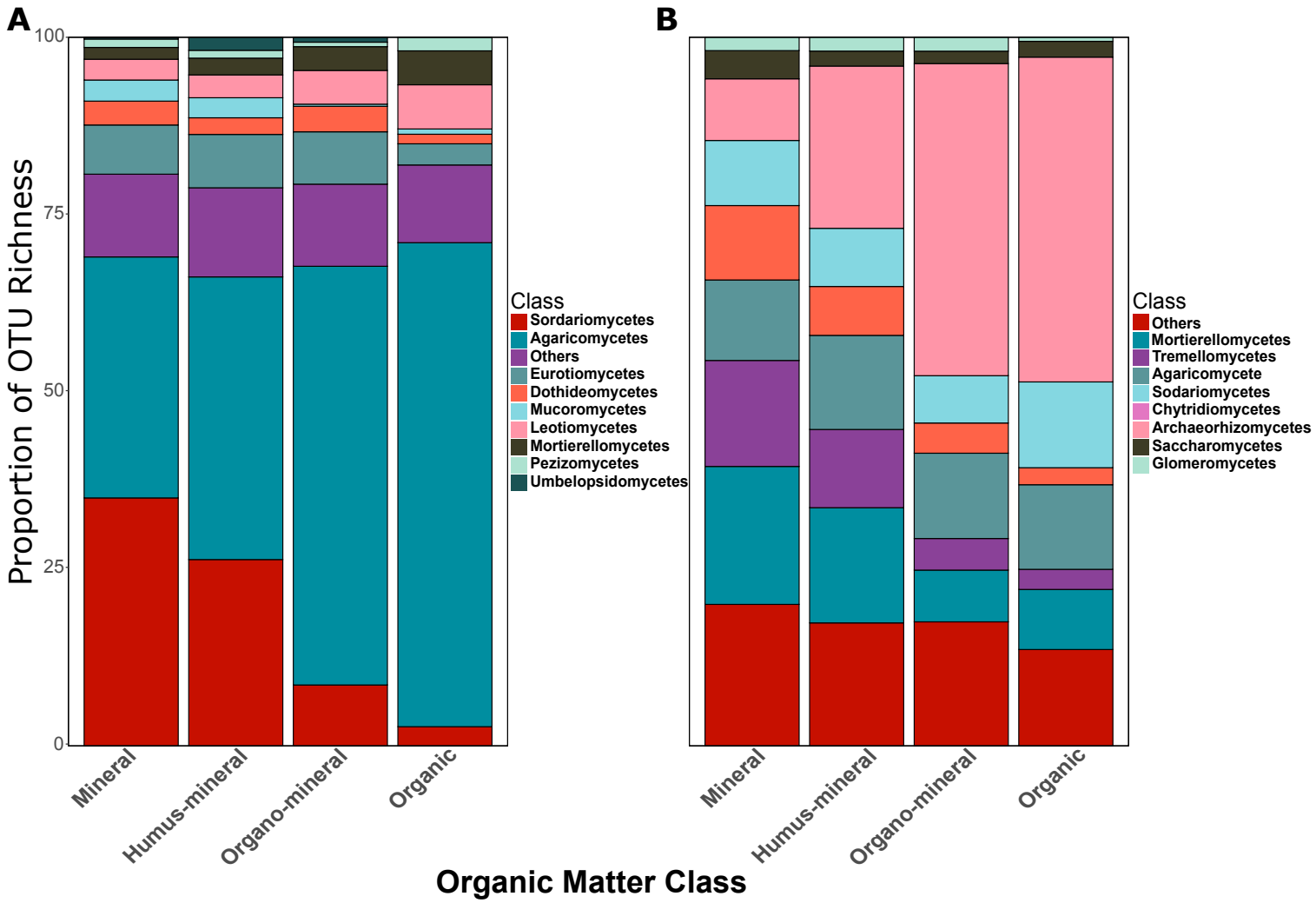


Fig. 1. Proportionate abundances of fungal OTUs for **A)** ITS1 and **B)** 18S data across the different organic matter classes. Organic matter classes are ordered by increasing percent organic matter.

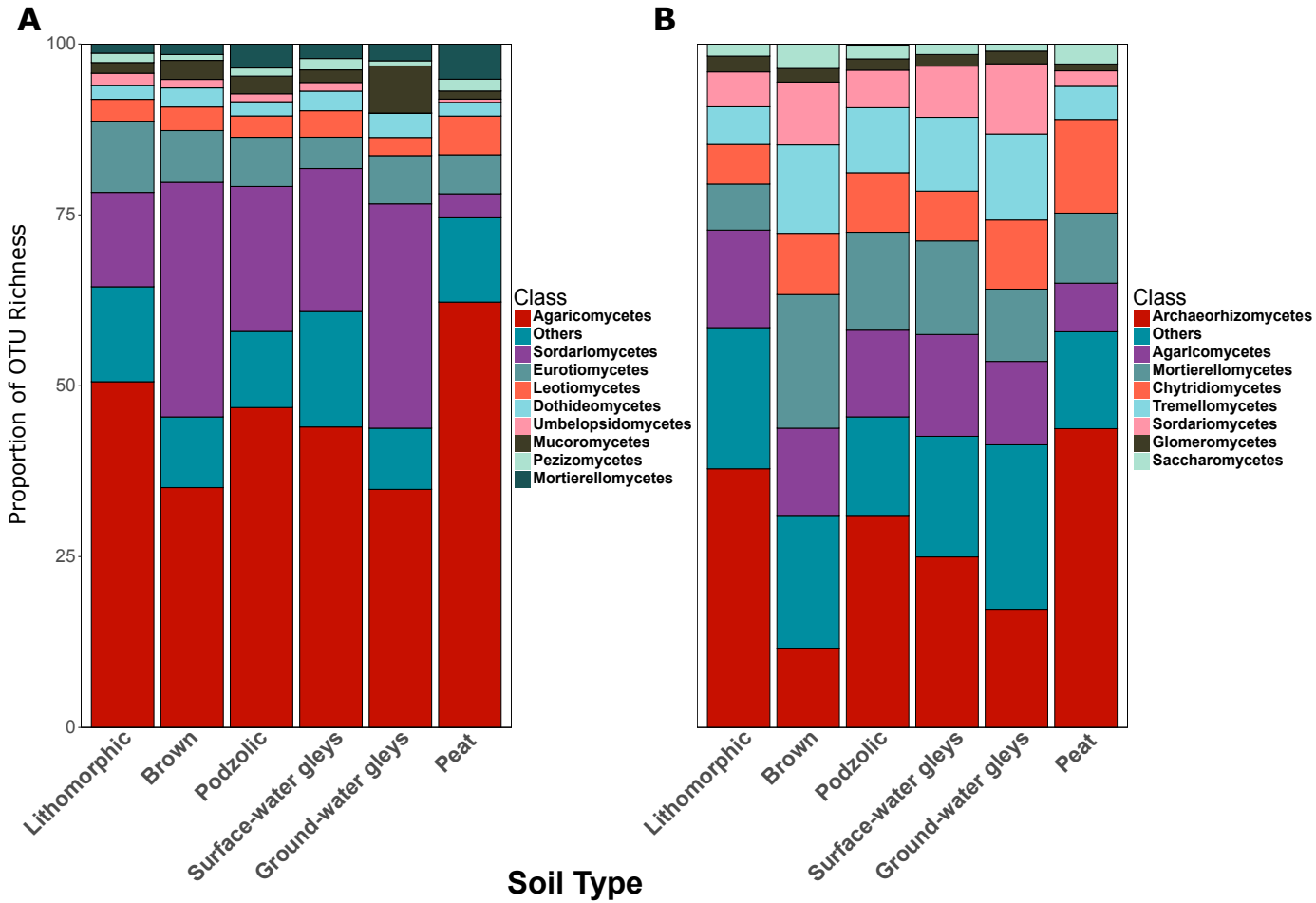


Fig. 2. Proportionate abundances of fungal OTUs for **A)** ITS1 and **B)** 18S data across the different soil types. Soil types are ordered by increasing approximate percent moisture content.

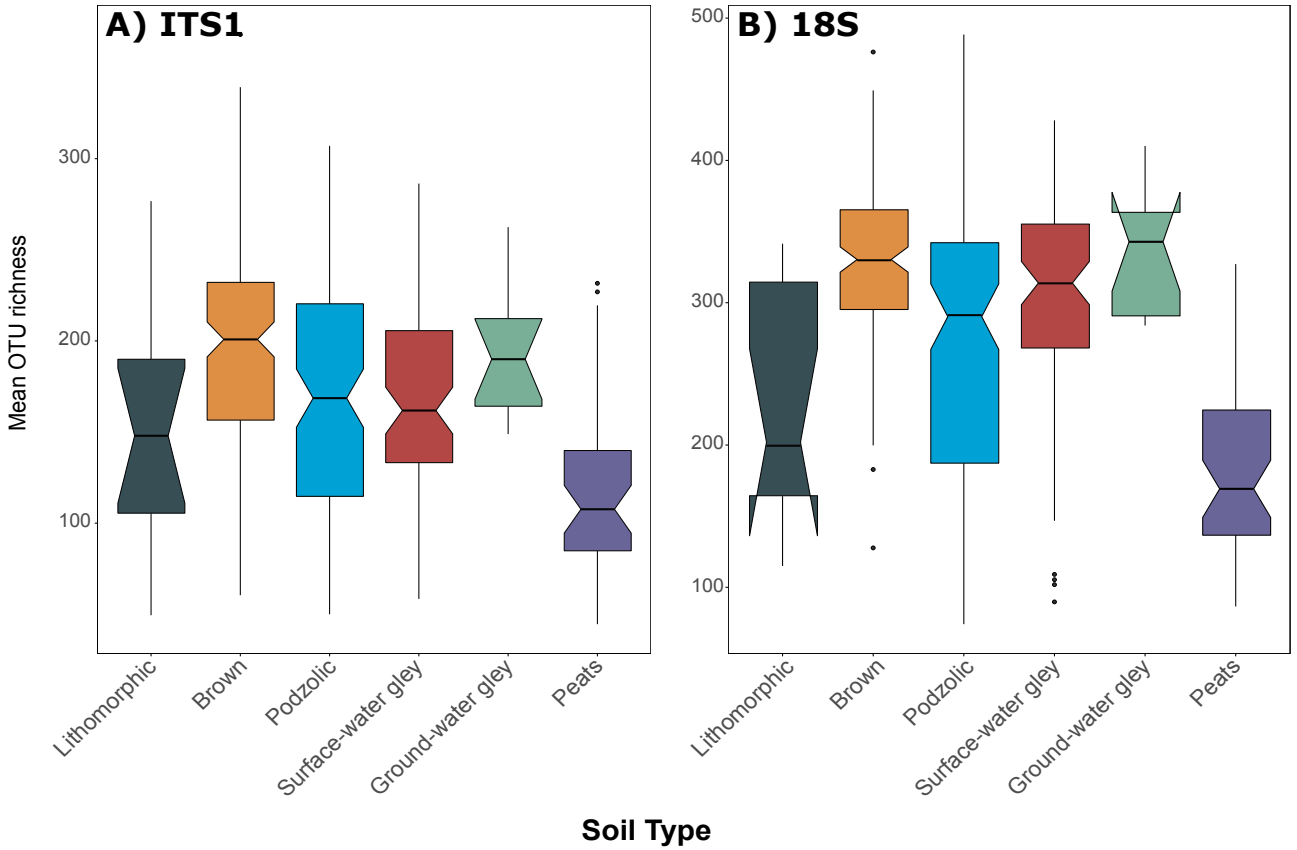


Fig. 3. Boxplots of fungal OTU richness for **A)** ITS1 and **B)** 18S datasets plotted against soil type. Soil types are ordered by increasing approximate moisture content. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range.

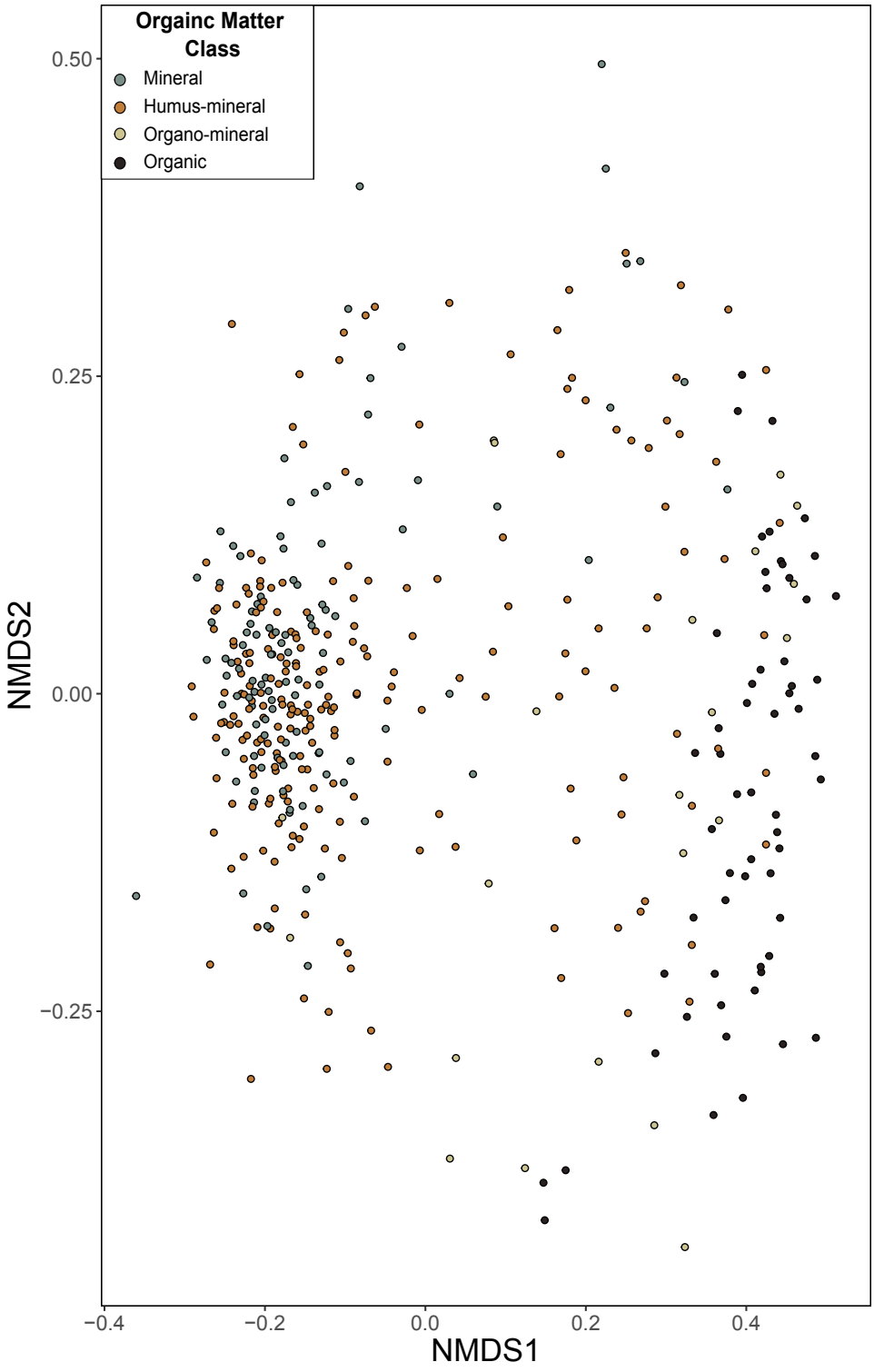


Fig. 4. Non-metric dimensional scaling ordinations of fungal community composition from ITS1 data across GMEP sites (stress = 0.13). Samples are coloured by organic matter class. Results of both PERMANOVA ($F_{3,408}$ = 9.34, p = 0.001) and of testing dispersion of variances ($F_{3,408}$ = 10.66, p = 0.001) were significant.

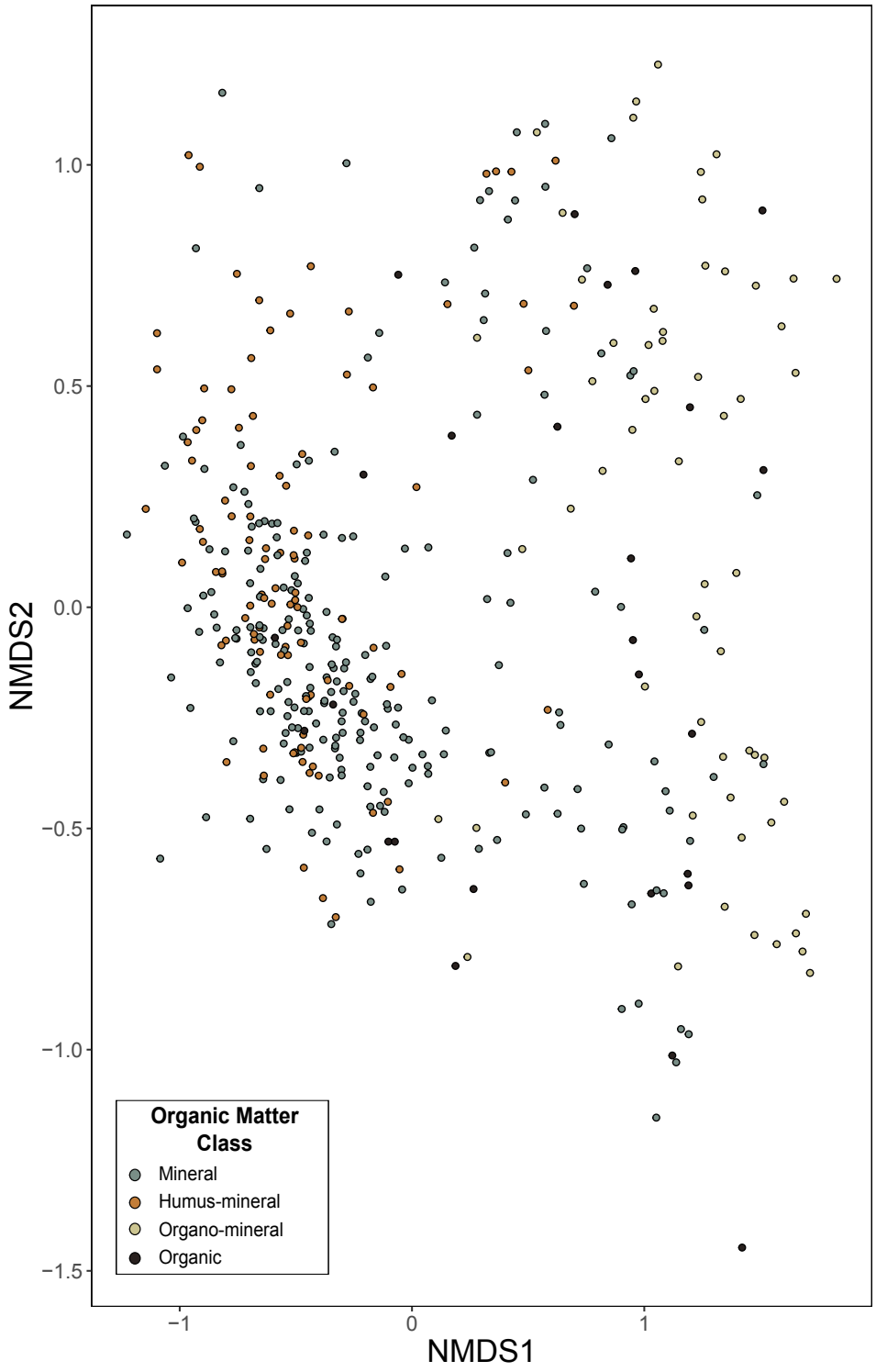


Fig. 5. Non-metric dimensional scaling ordinations of fungal community composition from 18S data across GMEP sites (stress = 0.11). Samples are coloured by organic matter class. Results of both PERMANOVA ($F_{3,417}$ = 13.06, p = 0.001) and of testing dispersion of variances ($F_{3,417}$ = 8.69, p = 0.001) were significant.

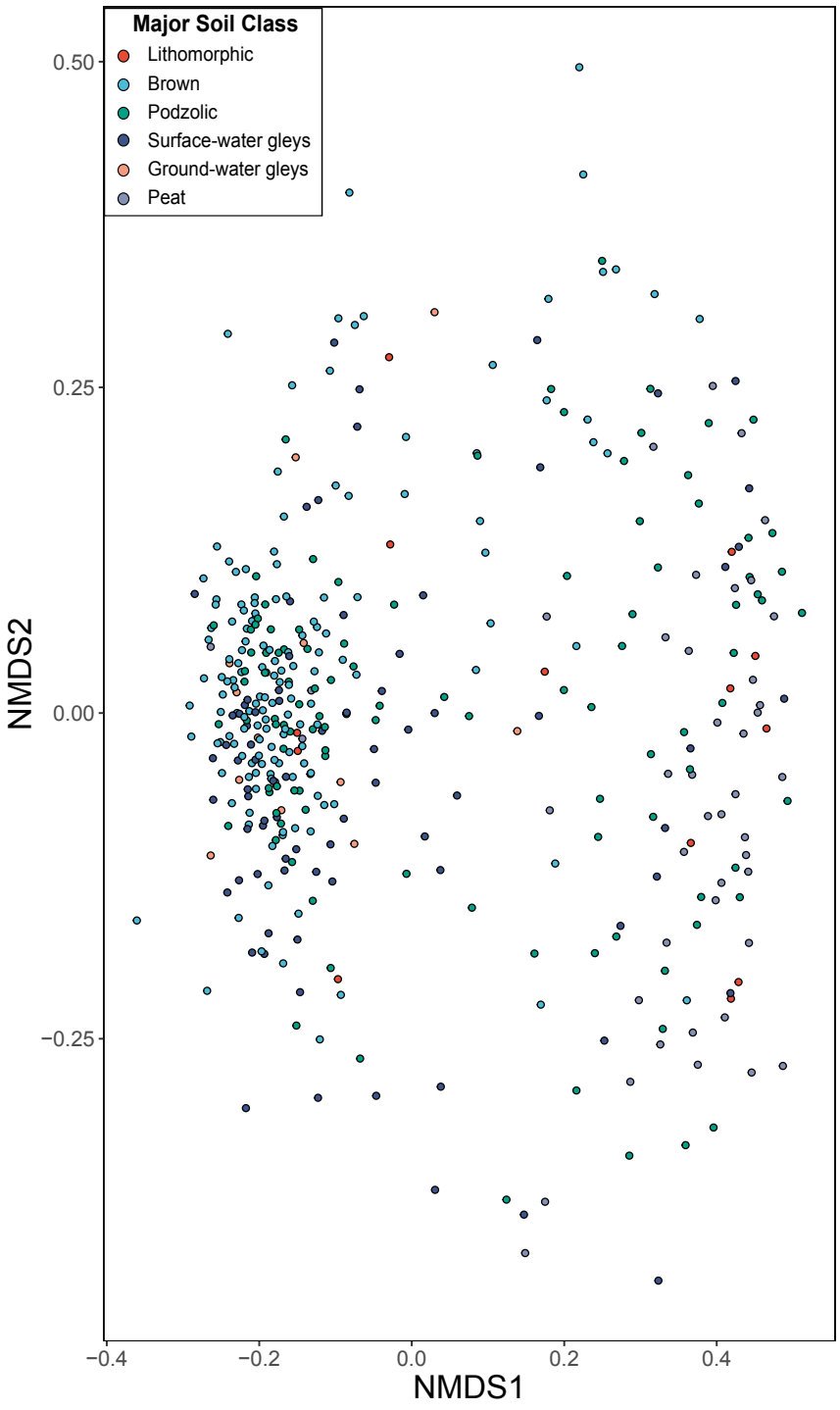


Fig. 6. Non-metric dimensional scaling ordinations of fungal community composition from ITS1 data across GMEP sites (stress = 0.13). Samples are coloured by soil type. Results of both PERMANOVA ($F_{5,407} = 4.44$, $p = 0.001$) and of testing dispersion of variances ($F_{5,407} = 9.72$, $p = 0.001$) were significant.

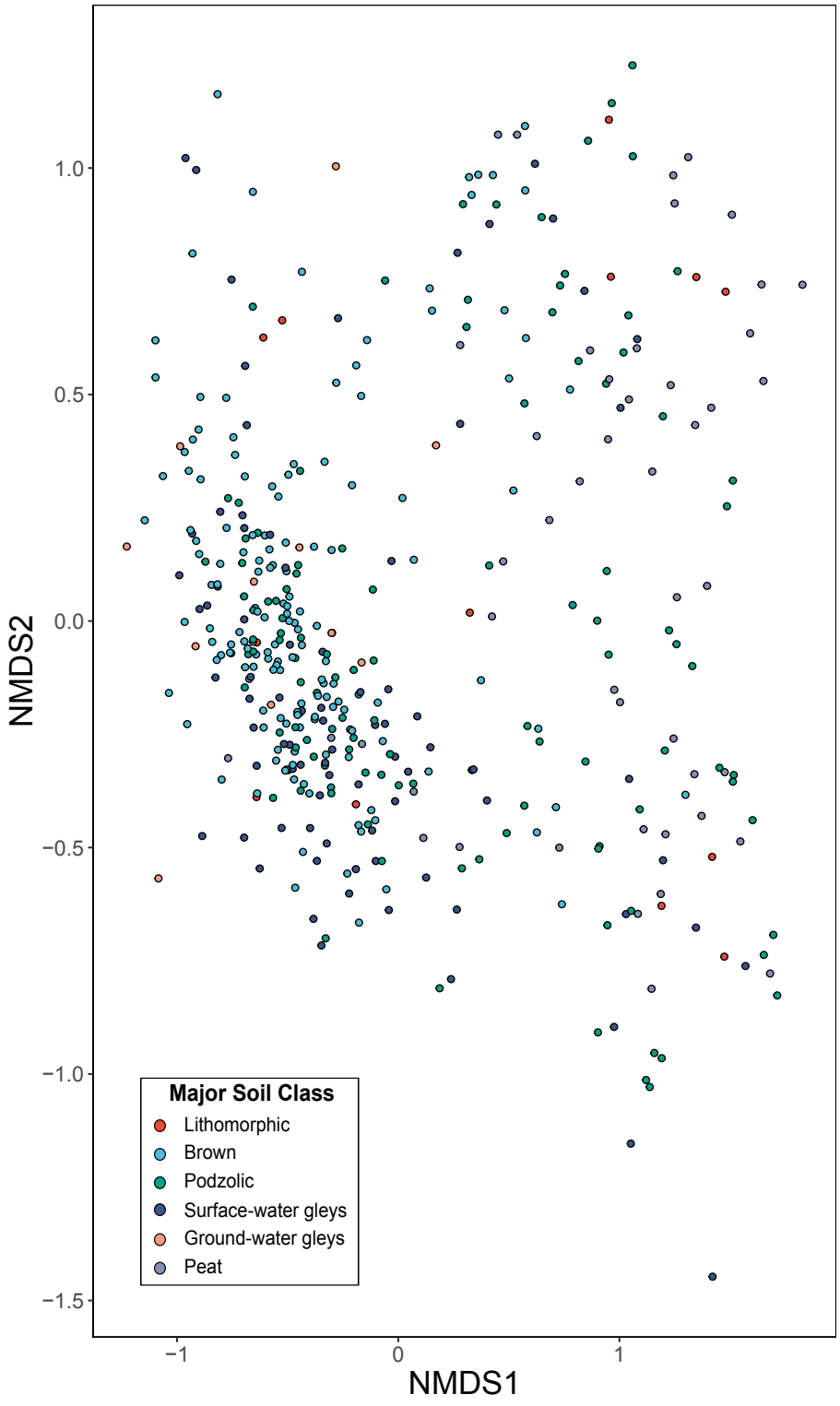


Fig. 7. Non-metric dimensional scaling ordinations of fungal community composition from 18S data across GMEP sites (stress = 0.11). Samples are coloured by soil type. Results of both PERMANOVA ($F_{5,416}$ = 6.0, p = 0.001) and of testing dispersion of variances ($F_{5,416}$ = 6.91, p = 0.001) were significant.

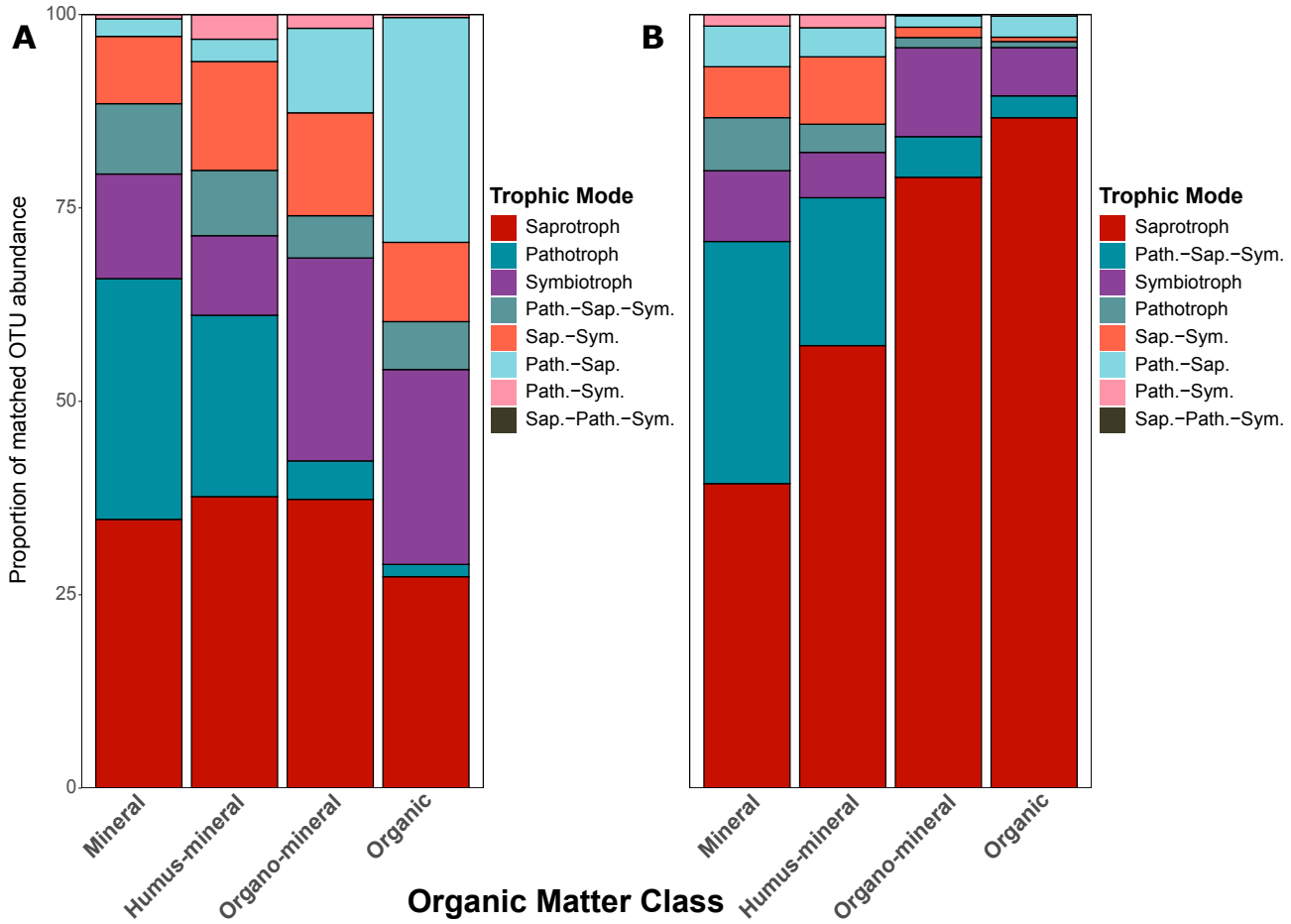


Fig. 8. Proportionate abundances of fungal OTUs matched to FUNGuild trophic groups for **A**) ITS1 and **B**) 18S data across organic matter classes. Organic matter classes are ordered by increasing percent organic matter. Abbreviations for multi-trophic mode groups are as follows: Path.-Sap. (Pathotroph-Saprotroph); Path.-Sap.-Sym. (Pathotroph-Saprotroph-Symbiotroph); Path.-Sym. (Pathotroph-Symbiotroph); Sap.-Path.-Sym (Saprotroph-Pathotroph-Symbiotroph); Sap.-Sym. (Saprotroph-Symbiotroph).

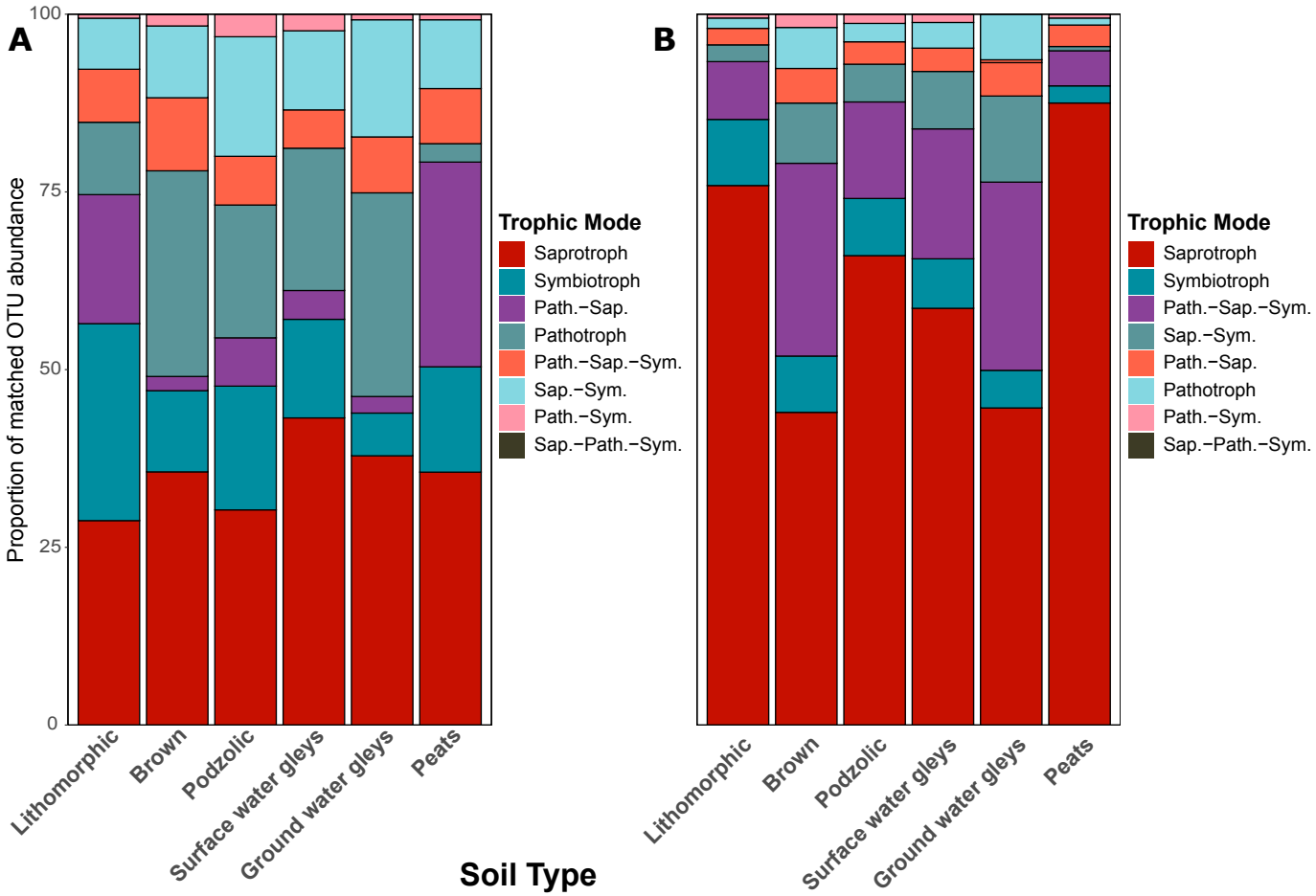


Fig. 9. Proportionate abundances of fungal OTUs matched to FUNGuild trophic groups for **A**) ITS1 and **B**) 18S data across soil types. Soil types are ordered by increasing approximate moisture content. Abbreviations for multi-trophic mode groups are as follows: Path.-Sap. (Pathotroph-Saprotroph); Path.-Sap.-Sym. (Pathotroph-Saprotroph-Symbiotroph); Path.-Sym. (Pathotroph-Symbiotroph); Sap.-Path.-Sym (Saprotroph-Pathotroph-Symbiotroph); Sap.-Sym. (Saprotroph-Symbiotroph).

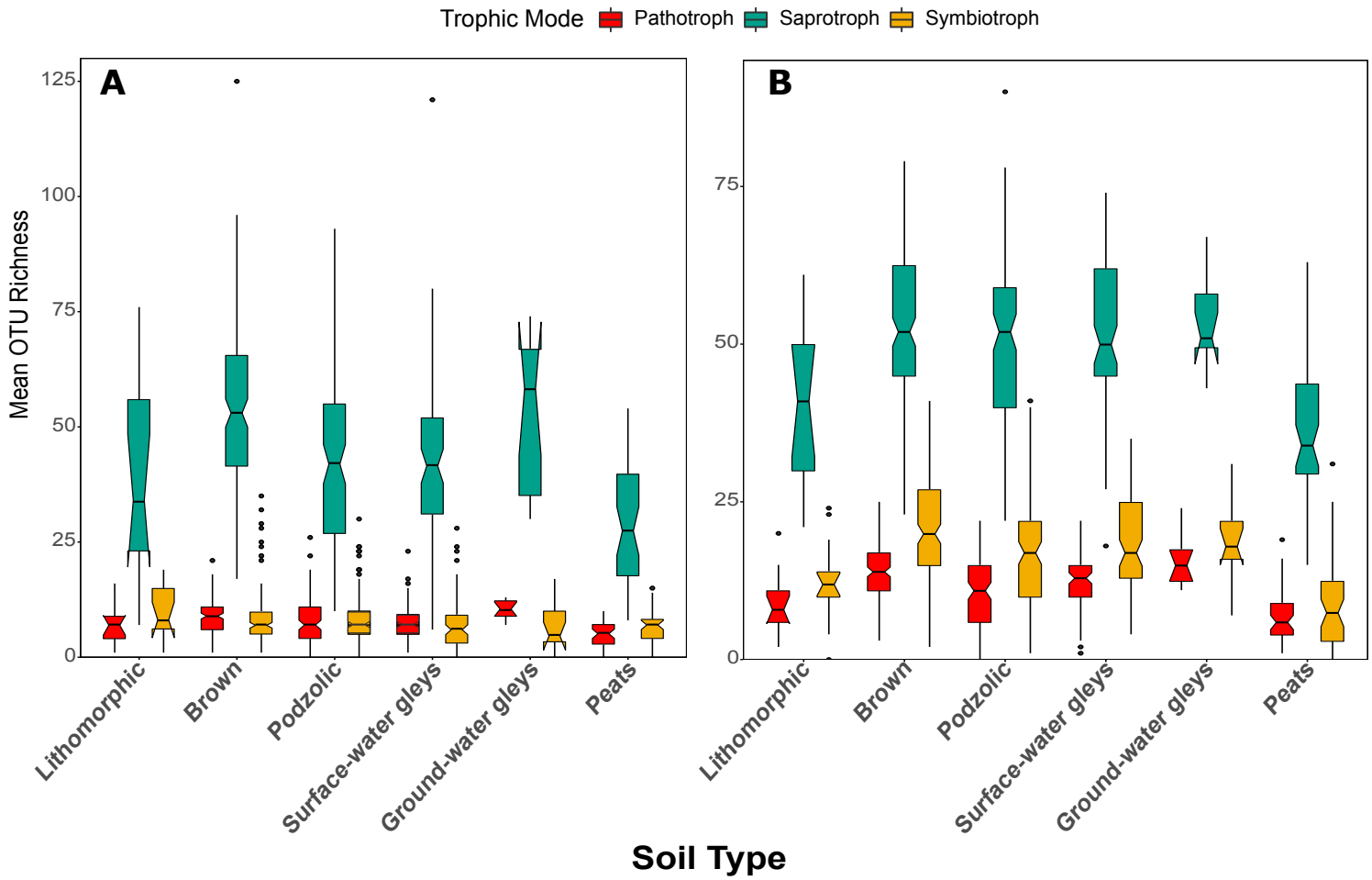


Fig. 10. Boxplots of richness of fungal OTUs matched to the pathotrophic, saprotroph, and symbiotroph trophic modes in FUNGuild for **A**) ITS1 and **B**) 18S datasets plotted against soil types. Soil types are listed in order of increasing approximate moisture content. Boxes cover the first and third quartiles and horizontal lines denote the median. Black dots represent outliers beyond the whiskers, which cover 1.5X the interquartile range.

Appendix 4

Supplementary Material for Chapter 6

This manuscript has been submitted to the *European Journal of Soil Science*

George, P.B.L.[1,2,*], Coelho, K.P.[3], Creer, S.[1], Lebron, I.[2], Emmett, B.A.[2], Robinson, D.A.[2], Jones, D.L.[1,4], 2019. *In Review*

Supplementary Tables

Table 1. Mean values (± SE) of soil physical and chemical variables of each Aggregate Vegetation Class (AVC). Differences between variables were determined by ANOVAs and Tukey's *post-hoc* tests. Results are as follows: total C ($F_{6, 427} = 89.13$ $p < 0.001$), total N ($F_{6, 427} = 61.03$, $p < 0.001$), C :N ratio ($F_{6, 427} = 94.41$, $p < 0.001$), total P ($F_{6, 424} = 7.1$, $p < 0.001$), organic matter ($F_{6, 428} = 107.02$, $p < 0.001$), sulphate concentration ($F_{6, 332} = 5.33$, $p < 0.001$), pH ($F_{6, 428} = 69.56$, $p < 0.001$), moisture content ($F_{6, 425} = 93.78$, $p < 0.001$), elevation ($F_{6, 429} = 78.42$, $p < 0.001$), and mean annual precipitation ($F_{6, 429} = 72.6$, $p < 0.001$). Means followed by the same letter are not significantly different from each other.

Environmental variable	Crops/weeds	Fertile grassland	Infertile grassland	Lowland wood	Upland wood	Moorland grass-mosaic	Heath/bog
Total C (%)	3.87 (± 0.83)d	4.75 (± 0.2)d	5.85 (± 0.33)d	5.78 (± 1.07)d	9.7 (± 2.25)c	12.2 (± 2.07)b	23.57 (± 1.88)a
Total N (%)	0.32 (± 0.05)d	0.45 (± 0.02)d	0.49 (± 0.02)d	0.4 (± 0.06)d	0.58 (± 0.1)c	0.83 (± 0.11)b	1.05 (± 0.09)a
C :N ratio	11.4 (± 0.81)cd	10.5 (± 0.13)d	11.6 (± 0.27)cd	13.9 (± 0.75)bc	15.9 (± 0.7)b	14.4 (± 0.42)b	20.7 (± 0.94)a
Organic matter (%)	7.5 (± 1.62)d	9.4 (± 0.34)d	11.3 (± 0.55)d	10.71 (± 1.7)d	18.8 (± 4.16)c	23.0 (± 3.72)b	39.23 (± 3.6)a
Sulphates (mg kg^{-1})	124.6 (± 20.08)a	68.8 (± 6.75)b	57.7 (± 3.06) c	96.3 (± 14.04)ab	63.1 (± 8.09)abc	49.4 (± 4.22)c	55.2 (± 7.75)c
pH ($CaCl_2$)	4.73 (± 0.26)b	5.2 (± 0.08)a	4.73 (± 0.05)b	4.31 (± 0.26)b	3.57 (± 0.1)cd	3.85 (± 0.09)c	3.84 (± 0.1)d
Moisture content (g g^{-1})	0.27 (± 0.08)d	0.41 (± 0.02)d	0.52 (± 0.03)d	0.44 (± 0.09)d	2.11 (± 0.76)c	3.25 (± 0.53)b	6.86 (± 0.74)a
Elevation (m)	88.7 (± 47.7)cd	109.3 (± 8.6)d	167.3 (± 8.7)c	119.1 (± 16.3)cd	297.8 (± 20.6)b	406.6 (± 19.2)a	380.6 (± 19.7)a
Mean annual precipitation (mL)	968.44 (± 69.01)c	1078.19 (± 24.71)c	1177.05 (± 18.91)c	1100.12 (± 52.28)c	1405.33 (± 65.35)b	2027.23 (± 74.39)a	1771.2 (± 58.19)a

Note: total C, total N, organic matter, and sulphate were all $\log_{10}$ transformed for ANOVA.

Supplementary Figures

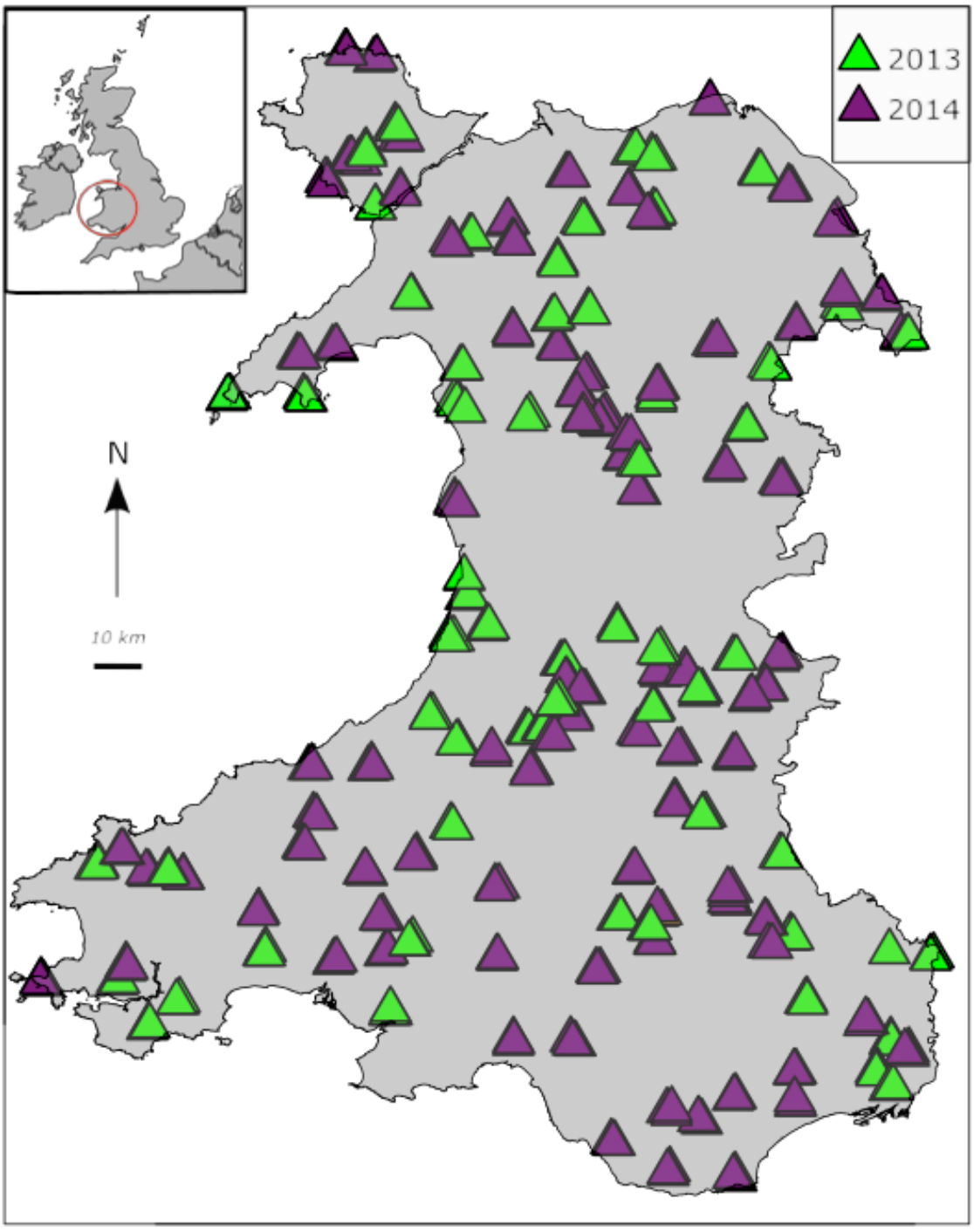


Fig. 1. Map of GMEP sample locations. To protect landowner anonymity, each triangle gives an approximate location of every 1 km^2 plot from which samples were taken

. **References**

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K., Fierer, N., Peña, A.G., Goodrich, J.K., Gordon, J.I., Huttley, G.A., Kelley, S.T., Knights, D., Koenig, J.E., Ley, R.E., Lozupone, C.A., McDonald, D., Muegge, B.D., Pirrung, M., Reeder, J., Sevinksky, J.R., Turnbaugh, P.J., Walters, W.A., Widmann, J., Yatsunenko, T., Zaneveld, J., Knight, R. 2010. QIIME allows analysis of high-throughput community sequencing data. Nature Methods 7, doi: 10.1038/nmeth.f.303

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Lozupone, C.A., Turnbaugh, P.J., Fierer, N., Knight, R., 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. Proceedings of the National Academy of Sciences 108 (Suppl. 1), 4516-4522.

DeSantis, T.Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E.L., Keller, K., Huber, T., Dalevi, D., Hu, P., Andersen, G.L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. Applied and Environmental Microbiology 72, 5069-5072.

Edgar, R.C., 2010. Search and clustering orders of magnitude faster than BLAST. Bionformatics 26, 2460-2461.

George, P.B.L., Lallias, D., Creer, S., Seaton, F.M., Kenny, J.G., Eccles, R.M., Griffiths, R.I., Lebron, I., Emmett, B.A., Robinson, D.A., Jones, D.L. 2019. Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. Nature Communications 10, 1107.

Martin, M., 2011. CUTADAPT removes adapter sequences from high-throughput sequencing reads. EMBnet Journal 17, doi: 10.14806/ej.17.1.200

Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F. 2016. VSEARCH: a versatile open source tool for metagenomics. PeerJ 4, e2584.

Sagova-Mareckova, M., Cermak, L., Novotna, J., Plhackova, K., Forstova, J., Kopecky, J., 2008. Innovative methods for soil DNA purification tested in soils with widely differing characteristics. Applied and Environmental Microbiology 74, 2902-2907.

Wang, Q., Garrity, G.M., Tiedje, J.M., Cole, J.R., 2007. Naïve Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. Applied and Environmental Microbiology 73, 5261-5267.

Appendix 5

Supplementary Material for Chapter 7

This manuscript is in preparation to be submitted to *Bioinformatics*

George, P.B.L.[1,2,*], Fidler, D.B.[1], Van Nostrand, J.D.[3], Atkinson, J.A.[4], Mooney, S.J.[4], Creer, S.[1], Griffiths, R.I..[2], McDonald, J.E.[1], Robinson, D.A.[2], Jones, D.L.[1,5], 2019. *In Preparation*.

Appendix 5.1 Complete bioindicator results

We used LDA with LEfSe to identify taxa indicative of vegetated or fallow soils in 1 and 10 year-old treatments. For prokaryotes, there were more indicator taxa in 10 year-old soils than 1 year-old soils (Fig. 7.5a, c). In the 1 year-old dataset there were 11 indicator taxa, of which, 5 were characteristic of vegetated and 6 of fallow soils (Fig. 7.5a). The genera *Rhodoplanes* and *Flavobacterium* along with the prospective family-level group EB1017 were the best bacterial indicator taxa of vegetated 1 year-old sites. Whereas, *Nitrosotalea devanaterra* and *Rhodanobacter* were the best indicators for 1 year-old fallow sites (Fig. 7.5a). There were considerable more indicator taxa in the 10 year-old dataset based on default (LDA score > 2.5) parameters (Appendix 5 Fig. 4); the most important (LDA score > 4) are shown in Fig. 7.5c. Nakamurellaceae, *Nitrospira*, and *Solirubrobacter* were the best indicators of 10 year-old vegetated soils. *Methylosinus* was the strongest indicator of 10 year-old fallow sites; interestingly *N. devanterra* was also a strong indicator in this data set (Fig. 7.5c).

Identification of the differentially abundant OTUs through DESeq2 revealed some congruence with indicator taxa from LDA analyses. For example, *N. devanterra* (OTU_1863), and two *Rhodanobacter* OTUs (OTU_865, OTU_870) are highlighted as being differentially abundant in 1 year-old fallow soils (Fig. 7.5b). Additionally, an OTU identified as *Streptomyces* (OTU_932) and *S. radiopugnans* (OTU_504), was more abundant in fallow

sites, supporting the identification of *S. radiopugnans* as an indicator taxon from LDA analysis (Fig. 7.5a,b). A full list of differentially abundant OTUs from 1 year-old soils can be found in Appendix 5 Table 2.

In 10 year-old soils, differentially abundant OTUs identified as RB40 (OTU_1686) and *Solirubrobacter* (OTU_1472) matched vegetation marker taxa from LDA analyses (Fig. 4c,d). *Methylosinus* (OTU_894) and *N. devanaterra* (OTU_1863) also appeared in the differential abundant fallow OTUs, matching the LDA data (Fig. 7.5c,d). There were also several fallow OTUs identified as *Bacillus* (OTU_1328, OTU_1318) supporting the findings of *B. foraminis* and *B. coahuilensis* as bioindicators (Fig. 7.5c,d; Appendix 5 Fig. 4). Additionally, Candidatus Koribacter (OTU_1635, OTU_693), and *Paenbacillus* also appear as differentially abundant in fallow soils (Fig. 7.5d; Appendix 5 Table 3), which support the LDA results highlighting the importance of *Paenbacillus* and Koribacteraceae in these soils (Fig. 7.5c; Appendix 5 Fig. 4). This is also true for *Nocardioides* (OTU_46) and *Nakamurella* (OTU_669) that appear as differentially abundant OTUs in vegetated soils, and their respective families as indicator taxa in LDA data (Fig. 7.4c; Appendix 7 Fig. 4).

Some differentially abundant OTUs from 1 year-old sites belonged to indicator taxa from 10 year-old sites. The prospective Acidobacteria family RB40 was identified as an important indicator taxon (Appendix 5 Fig. 4) and appeared as a differentially abundant OTU (OTU_1830) from 1 year-old vegetated sites (Fig. 7.5b), along with closely related OTUs (OTU_1443, OTU_1762). Similarly, the uncultured archaeal group SAGMA_X is differentially abundant OTU (OTU_1859) in the 1 year-old and as an indicator taxon in 10 year-old fallow soils (Fig. 7.5b; Appendix 5 Fig. 4).

Unexpectedly, other differentially abundant OTUs matched to indicator taxa from opposing treatments. For example, *Rhodoplanes* appears as an indicator of vegetated soil in LDA data from 1 year-old soils (Fig. 7.5a) but OTUs identified as *Rhodoplanes* in fallow soils from both the 1 year-old (OTU_1627) and 10 year-old (1633) datasets (Fig. 7.5b, d). This was also the case for *Pilimelia* (OTU_1524) and the opposite trend is apparent for OTU_48, identified as Nocardioidaceae (Fig. 7.5, Appendix 5 Fig. 4).

Using the same criteria as for prokaryotes, bioindicators were identified for fungi (Fig. 7.6) and protists (Fig. 7.7). Notably, there were much fewer fungal indicator taxa and differentially abundant OTUs for both 1 year-old and 10 year-old sites (Fig. 7.6). Indeed only 4 indicator taxa (2 vegetated and 2 fallow) from 1 year-old sites and 9 indicator taxa (3 vegetated and 6 fallow) from 10 year-old taxa were identified from LDA data (Fig. 7.6a,c). In 1 year-old data, Orbiliomycetes and *Coprinopsis brunneofibrillosa* were indicative of vegetated soils and *Cotylidia undulate* and *Mucor heimalis* were indicative of fallow soils (Fig. 7.6a). Orbiliomycetes was also the strongest indicator of vegetated soils in 10 year-old soils, along with *Psathyrella globosivelata* and *Drechsiera* sp., *Onygenales* and *C. candidolanata* were the strongest indicators of fallow (Fig. 7.6c).

The differentially abundant fungal OTUs in vegetated soils did not match the LDA-identified indicator taxa. Rather, these included 3 OTUs identified as Stephanosporaceae (OTU_301, OTU_272, OTU_246), 2 different *Mycena* species (OTU_46, OTU_560), 1 *Conocybe* (OTU_216), and an unnamed Ascomycota (i.e. OTU_243). There were no differentially abundant OTUs for 1 year-old fallow sites (Chapter 7 Fig. 6b; Appendix 5 Table 4). The only match between differentially abundant OTUs and taxa from LDA data in the 10 year-old sites was OTU_221, identified as Orbiliomycetes, from vegetated soils (Chapter 7

Fig. 6c, d). Other differentially abundant OTUs included taxa present in 10 year-old vegetated sites such as Stephanosporaceae (OTU_272), *M. valida* (OTU_560), and *C. fuscimarginata* (OTU_39) (Fig. 7.6d; Appendix 5 Table 5). OTUs indicative of fallow in 10 year-old soils were identified as *Trichomerium foliicola* (OTU_358), *Plenodomus biglobosus* (OTU_337), Schizoporaceae (OTU_325), and Agaricomycetes (i.e. OTU_62) of uncertain affinity (Fig. 7.6d; Appendix 5 Table 5).

In the protistan dataset, as with the bacterial data, there were a large number of indicators and so only taxa with an LDA score greater than 3.5 are presented in Chapter 7 Fig 7a and c. For the 1 year-old data, LDA revealed 20 indicator taxa (9 vegetated and 11 fallow) under default parameters (Appendix 5 Fig. 5). The genera *Spongomonas* and *Pseudoplatyphyra* (Fig. 7.7a) were the strongest protist indicators of vegetated samples in 1 year-old soils, though a number of other genera were also identified (Appendix 5 Fig. 5). Protistan groups indicative of fallow soils included the families Vampyrellidae, Thaumatomonadidae, and Trinematidae (Fig. 7.6a) as well as ambiguously identified cercozoans and Oomycetes (Appendix 5 Fig. 5). In the 10 year-old soils, there were 60 taxa indicative of vegetated soils and 29 taxa indicative of fallow soils (Fig. 7.7c; Appendix 5 Fig. 6). *Chloroidium*, *Spumella*, *Woronia* were the strongest indicators of vegetated soils (Fig. 7.7c). *Heteromita*, and *Metopion* were the genera most associated with fallow soils, along with the families and Trebouxiophyceae and Cercocmonadidae, and Stramenopiles from the MAST_12C group (Fig. 7.7c).

Differentially abundant protistan OTUs in both 1 year-old and 10 year-old soils showed similar trends to bacterial OTUs, with some supporting LDA scores in like treatments within age class and across age class, as well as incongruence across age classes. Congruence

was found between differentially abundant OTUs such as *Spongomonas* (OTU_1120), *Hypotrichia* (OTU_2052), and *Paulinella* (OTU_5676) with vegetated indicator taxa from 1 year-old vegetated soils (Fig. 7.7a,b; Appendix 5 Table 6). Similarly, differentially abundant OTUs belonging to Vampyrellidae (OTU_1530) and MAST_12C group Stramenopiles appear in the 1 year-old fallow soils (Fig. 7.7b). *Heteromita* (OTU_9307) and members of the order Euglyphida (OTU_2179, OTU_2212, OTU_5394) were differentially abundant in 10 year-old fallow soils (Fig. 7.7d; Appendix 5 Table 7). Unclassified OTUs belonging to the alveolate order Eugregarinorida were differentially abundant in the vegetated 10 year-old soils (Fig. 7.7d) and identified as an indicator taxa by LDA (Appendix 5 Fig. 6).

Incongruent matches between treatments and age classes included a number of OTUs belonging to Thecofilosea appearing as differentially abundant in vegetated and fallow treatments of both age classes (Fig. 7.7b,d; Appendix 5 Table 6; Appendix 5 Table 7 despite this group also being labelled an indicator taxa of vegetated soils in the 10 year-old soils only (Appendix 5 Fig. 6). Also, vegetated soils from both age classes were associated with differentially abundant OTUs belonging to Oomycetes, including potential representatives of *Phytophthera infestans* (OTU_412), a major plant pest (Fig. 7.7b,d; Appendix 5 Table 6; Appendix 5 Table 7).

Supplementary Tables

Table 1. Soil porosity from X-ray μCT analysis conducted at whole column and aggregate scale for vegetated vs. 10 year fallow only. Mean values (± SE) are presented for both treatments. Significant differences indicated by *** ($p<0.001$).

		Vegetated	**Fallow**
Column scale	Mean pore size (mm^2)	1.47 (± 0.01) ***	1.30 (± 0.01)
	Porosity (%)	16.09 (± 3.51)	6.18 (± 1.55)
	Total Pore Area (mm^2)	437745.55 (± 95445.62)	167986.71 (± 42062.95)
	PSD Ratio ($D_{10:60}$)	109.76 (± 36.41)	57.72 (± 12.79)

	Euler number	27413.0 (± 40231.7)	88963.0 (± 14565.0)
Aggregate scale	Mean pore size (mm^2)	0.05 (± 0.00) ***	0.01 (± 0.00)
	Porosity (%)	10.62 (± 1.14)	8.58 (± 1.05)
	Total Pore Area (mm^2)	206.37 (± 22.24)	166.90 (± 20.49)
	PSD Ratio ($D_{10:60}$)	7.17 (± 1.10)	5.31 (± 0.72)
	Euler number	25852.5 (± 12911.0)	-9440 (± 12124.6)

Table 2. Prokaryote bioindicators of fallow and vegetated treatments from 1 year-old soils identified by DESeq2 analyses.

OTU Identity	Treatment	Taxa
OTU_1628	Fallow	*Sporosarcina*
OTU_48	Fallow	Nocardioidaceae
OTU_1863	Fallow	*Nitrosotalea devanaterra*
OTU_865	Fallow	*Rhodanobacter*
OTU_870	Fallow	*Rhodanobacter*
OTU_1627	Fallow	*Rhodanobacter*
OTU_932	Fallow	*Streptomyces*
OTU_504	Fallow	*Streptomyces radiopugnans*
OTU_1859	Fallow	SAGMA-X
OTU_1856	Vegetated	Candidatus Solibacter
OTU_26	Vegetated	*Flavobacterium*
OTU_1830	Vegetated	RB40
OTU_1161	Vegetated	*Methylibium*
OTU_1762	Vegetated	RB40
OTU_1443	Vegetated	RB40

Table 3. Prokaryote bioindicators of fallow and vegetated treatments from 10 year-old soils identified by DESeq2 analyses.

OTU Identity	Treatment	Taxa	OTU Identity	Treatment	Taxa
OTU_913	Fallow	Sinobacteraceae	**OTU_603**	Vegetated	DA101
OTU_1484	Fallow	*Ammoniphilus*	**OTU_28**	Vegetated	*Adhaeribacter*
OTU_472	Fallow	Koribacteraceae	**OTU_27**	Vegetated	Chitinophagaceae
OTU_870	Fallow	*Rhodanobacter*	**OTU_864**	Vegetated	*Mycobacterium*
OTU_647	Fallow	Candidatus Solibacter	**OTU_1167**	Vegetated	*Steroidobacter*
OTU_514	Fallow	*Bacillus*	**OTU_397**	Vegetated	*Polaromonas*
OTU_795	Fallow	Intrasporangiaceae	**OTU_313**	Vegetated	Cytophagaceae
OTU_140	Fallow	Koribacteraceae	**OTU_1398**	Vegetated	Syntrophobacteraceae

OTU_1521	Fallow	Acidobacteriaceae	**OTU_1849**	Vegetated	*Limnohabitans*
OTU_1024	Fallow	DA101	**OTU_388**	Vegetated	Intrasporangiaceae
OTU_1366	Fallow	*Dokdonella*	**OTU_1000**	Vegetated	Solirubrobacteraceae
OTU_1791	Fallow	*Kaistobacter*	**OTU_16**	Vegetated	Gaiellaceae
OTU_1533	Fallow	Bradyrhizobiaceae	**OTU_1566**	Vegetated	Chitinophagaceae
OTU_471	Fallow	Planococcaceae	**OTU_1019**	Vegetated	*Sphingomonas*
OTU_1737	Fallow	Sphingomonadaceae	**OTU_62**	Vegetated	Gaiellaceae
OTU_1692	Fallow	Chitinophagaceae	**OTU_151**	Vegetated	PRR-10
OTU_320	Fallow	Ellin515	**OTU_68**	Vegetated	Solirubrobacteraceae
OTU_1714	Fallow	*Sporosarcina*	**OTU_1178**	Vegetated	*Rhodoplanes*
OTU_1627	Fallow	*Rhodoplanes*	**OTU_95**	Vegetated	*Luteolibacter*
OTU_1859	Fallow	SAGMA-X	**OTU_387**	Vegetated	DA101
OTU_634	Fallow	Ellin515	**OTU_72**	Vegetated	Gaiellaceae
OTU_1593	Fallow	0319-6A21	**OTU_533**	Vegetated	*Streptomyces*
OTU_925	Fallow	*Bacillus*	**OTU_60**	Vegetated	Rhizobiaceae
OTU_1255	Fallow	Koribacteraceae	**OTU_97**	Vegetated	Chitinophagaceae
OTU_348	Fallow	*Acinetobacter rhizosphaerae*	**OTU_1773**	Vegetated	Nocardioidaceae
OTU_231	Fallow	*Bacillus*	**OTU_492**	Vegetated	Ellin6075
OTU_1704	Fallow	DA101	**OTU_1770**	Vegetated	*Pedomicrobium*
OTU_805	Fallow	Rhodospirillaceae	**OTU_775**	Vegetated	DA101
OTU_702	Fallow	*Sporosarcina ginsengi*	**OTU_1792**	Vegetated	Ellin515
OTU_1844	Fallow	*Bacillus foraminis*	**OTU_1808**	Vegetated	*Geobacter*
OTU_1223	Fallow	*Hyphomicrobium*	**OTU_1384**	Vegetated	Gaiellaceae
OTU_1274	Fallow	Koribacteraceae	**OTU_1110**	Vegetated	*Pseudomonas*
OTU_498	Fallow	*Bacillus*	**OTU_973**	Vegetated	*Rhodoplanes*
OTU_1524	Fallow	*Pilimelia*	**OTU_1565**	Vegetated	Agrobacterium
OTU_1633	Fallow	*Rhodoplanes*	**OTU_222**	Vegetated	*Sphingomonas*
OTU_478	Fallow	*Bacillus cereus*	**OTU_687**	Vegetated	Ellin517
OTU_1077	Fallow	*Paenibacillus*	**OTU_583**	Vegetated	Cytophagaceae
OTU_1549	Fallow	Rhodospirillaceae	**OTU_278**	Vegetated	Rhodospirillaceae
OTU_476	Fallow	Koribacteraceae	**OTU_272**	Vegetated	*Rubrivivax*
OTU_1668	Fallow	Koribacteraceae	**OTU_1331**	Vegetated	*Rhodoplanes*
OTU_1120	Fallow	*Lysinibacillus massiliensis*	**OTU_1465**	Vegetated	Ellin515
OTU_1012	Fallow	*Bacillus*	**OTU_176**	Vegetated	Ellin6075
OTU_1809	Fallow	Micromonosporaceae	**OTU_407**	Vegetated	Gaiellaceae
OTU_714	Fallow	EB1003	**OTU_577**	Vegetated	Solirubrobacteraceae
OTU_1591	Fallow	*Bacillus*	**OTU_213**	Vegetated	Phyllobacteriaceae
OTU_750	Fallow	*Novosphingobium*	**OTU_1093**	Vegetated	*Rhizobium*
OTU_667	Fallow	Koribacteraceae	**OTU_83**	Vegetated	*Luteolibacter*
OTU_1218	Fallow	Sinobacteraceae	**OTU_436**	Vegetated	DA101
OTU_1315	Fallow	Koribacteraceae	**OTU_161**	Vegetated	Xanthomonadaceae
OTU_756	Fallow	Rhodospirillaceae	**OTU_1114**	Vegetated	*Phycicoccus*

OTU_905	Fallow	Koribacteraceae	**OTU_288**	Vegetated	Nocardioidaceae
OTU_690	Fallow	Ellin515	**OTU_1153**	Vegetated	Candidatus Solibacter
OTU_1774	Fallow	*Kaistobacter*	**OTU_837**	Vegetated	Sinobacteraceae
OTU_737	Fallow	Pirellulaceae	**OTU_1827**	Vegetated	EB1017
OTU_871	Fallow	Rhodoplanes	**OTU_564**	Vegetated	*Mesorhizobium*
OTU_602	Fallow	DA101	**OTU_1719**	Vegetated	*Methylibium*
OTU_1034	Fallow	Bradyrhizobiaceae	**OTU_427**	Vegetated	Nocardioidaceae
OTU_465	Fallow	Planococcaceae	**OTU_1103**	Vegetated	*Salinibacterium*
OTU_865	Fallow	*Rhodanobacter*	**OTU_1443**	Vegetated	RB40
OTU_1700	Fallow	*Rhodoplanes*	**OTU_303**	Vegetated	EB1017
OTU_1448	Fallow	*Ammoniphilus*	**OTU_332**	Vegetated	EB1017
OTU_491	Fallow	*Bacillus*	**OTU_588**	Vegetated	EB1017
OTU_932	Fallow	*Streptomyces*	**OTU_349**	Vegetated	Nocardioidaceae
OTU_911	Fallow	*Bacillus*	**OTU_1854**	Vegetated	RB40
OTU_1763	Fallow	*Paenisporosarcina*	**OTU_253**	Vegetated	*Pedobacter*
OTU_1697	Fallow	*Paenisporosarcina*	**OTU_13**	Vegetated	DA101
OTU_710	Fallow	*Rhodoplanes*	**OTU_82**	Vegetated	Chitinophagaceae
OTU_1388	Fallow	Sphingomonadaceae	**OTU_553**	Vegetated	*Kribbella*
OTU_1741	Fallow	Syntrophobacteraceae	**OTU_789**	Vegetated	DA101
OTU_641	Fallow	Koribacteraceae	**OTU_165**	Vegetated	Candidatus Solibacter
OTU_915	Fallow	Koribacteraceae	**OTU_653**	Vegetated	C111
OTU_788	Fallow	Gaiellaceae	**OTU_640**	Vegetated	DA101
OTU_454	Fallow	*Bacillus coahuilensis*	**OTU_760**	Vegetated	*Variovorax paradoxus*
OTU_1552	Fallow	*Nitrospira*	**OTU_1681**	Vegetated	C111
OTU_1357	Fallow	*Rhodoplanes*	**OTU_1185**	Vegetated	Rhodospirillaceae
OTU_894	Fallow	*Methylosinus*	**OTU_437**	Vegetated	Pirellulaceae
OTU_1273	Fallow	Planococcaceae	**OTU_368**	Vegetated	*Microlunatus*
OTU_1511	Fallow	Candidatus Koribacter	**OTU_600**	Vegetated	DA101
OTU_1075	Fallow	*Bacillus*	**OTU_1564**	Vegetated	Cytophagaceae
OTU_1300	Fallow	*Bacillus*	**OTU_1361**	Vegetated	Bradyrhizobiaceae
OTU_1628	Fallow	*Sporosarcina*	**OTU_1625**	Vegetated	RB40
OTU_819	Fallow	Candidatus Solibacter	**OTU_941**	Vegetated	*Pseudomonas umsongensis*
OTU_606	Fallow	*Bacillus*	**OTU_26**	Vegetated	*Flavobacterium*
OTU_42	Fallow	*Sporosarcina*	**OTU_43**	Vegetated	EB1017
OTU_1108	Fallow	Acidobacteriaceae	**OTU_100**	Vegetated	C111
OTU_672	Fallow	*Bacillus*	**OTU_1064**	Vegetated	auto67_4W
OTU_693	Fallow	Candidatus Koribacter	**OTU_54**	Vegetated	*Mycobacterium*
OTU_744	Fallow	Koribacteraceae	**OTU_705**	Vegetated	Ellin506
OTU_1374	Fallow	Candidatus Koribacter	**OTU_1837**	Vegetated	Pirellulaceae
OTU_604	Fallow	Candidatus Koribacter	**OTU_1124**	Vegetated	Comamonadaceae
OTU_840	Fallow	Candidatus Koribacter	**OTU_1686**	Vegetated	RB40
OTU_896	Fallow	Sinobacteraceae	**OTU_1161**	Vegetated	*Methylibium*

OTU_108	Fallow	Candidatus Solibacter	**OTU_269**	Vegetated	Gemmataceae
OTU_1483	Fallow	Acetobacteraceae	**OTU_61**	Vegetated	HTCC2089
OTU_770	Fallow	*Rhodoplanes*	**OTU_953**	Vegetated	Micrococcaceae
OTU_1863	Fallow	*Nitrosotalea devanaterra*	**OTU_295**	Vegetated	*Mesorhizobium*
OTU_1718	Fallow	*Sporosarcina*	**OTU_276**	Vegetated	*Janibacter*
OTU_1318	Fallow	*Bacillus*	**OTU_101**	Vegetated	C111
OTU_1635	Fallow	Candidatus Koribacter	**OTU_884**	Vegetated	*Agrobacterium*
OTU_1794	Fallow	*Paenisporosarcina*	**OTU_339**	Vegetated	Comamonadaceae
OTU_1328	Fallow	Bacillus	**OTU_601**	Vegetated	Ellin6075
OTU_48	Vegetated	Nocardioidaceae	**OTU_377**	Vegetated	Xanthomonadaceae
OTU_1276	Vegetated	Comamonadaceae	**OTU_103**	Vegetated	*Nocardioides*
OTU_666	Vegetated	Ellin515	**OTU_367**	Vegetated	Gaiellaceae
OTU_50	Vegetated	Patulibacteraceae	**OTU_46**	Vegetated	Nocardioides
OTU_36	Vegetated	*Nocardioides*	**OTU_93**	Vegetated	Solirubrobacteraceae
OTU_522	Vegetated	Phenylobacterium	**OTU_281**	Vegetated	Chitinophagaceae
OTU_229	Vegetated	Streptomycetaceae	**OTU_1728**	Vegetated	*Rhodoplanes*
OTU_798	Vegetated	*Pedomicrobium*	**OTU_1404**	Vegetated	*Rubrivivax*
OTU_1830	Vegetated	RB40	**OTU_1206**	Vegetated	auto67_4W
OTU_944	Vegetated	*Streptomyces*	**OTU_1044**	Vegetated	*Janibacter*
OTU_957	Vegetated	*Arthrobacter*	**OTU_1472**	Vegetated	Solirubrobacter
OTU_704	Vegetated	Chitinophagaceae	**OTU_400**	Vegetated	Sinobacteraceae
OTU_1419	Vegetated	*Rhodoplanes*	**OTU_964**	Vegetated	C111
OTU_787	Vegetated	Comamonadaceae	**OTU_86**	Vegetated	Nocardioidaceae
OTU_428	Vegetated	*Mesorhizobium*	**OTU_263**	Vegetated	Gaiellaceae
OTU_1426	Vegetated	Comamonadaceae	**OTU_319**	Vegetated	*Nitrospira*
OTU_1528	Vegetated	*Rhodoplanes*	**OTU_669**	Vegetated	Nakamurellaceae
OTU_1762	Vegetated	RB40			

Table 4. Fungi bioindicators of fallow and vegetated treatments from 1 year-old soils identified by DESeq2 analyses

OTU Identity	Treatment	Taxa
OTU_225	Vegetated	Ascomycota
OTU_560	Vegetated	*Mycena valida*
OTU_243	Vegetated	Ascomycota
OTU_267	Vegetated	Ascomycota
OTU_46	Vegetated	*Mycena flavoalba*
OTU_246	Vegetated	Stephanosporaceae
OTU_272	Vegetated	Stephanosporaceae
OTU_301	Vegetated	Stephanosporaceae
OTU_216	Vegetated	*Conocybe*

Table 5. Fungi bioindicators of fallow and vegetated treatments from 10 year-old soils identified by DESeq2 analyses.

OTU Identity	Treatment	Taxa
OTU_358	Fallow	*Trichomerium foliicola*
OTU_337	Fallow	*Plenodomus biglobosus*
OTU_325	Fallow	Schizoporaceae
OTU_335	Fallow	Agaricales
OTU_62	Fallow	Agaricomycetes
OTU_336	Vegetated	*Mucor*
OTU_128	Vegetated	*Volucrispora graminea*
OTU_299	Vegetated	Ascomycota
OTU_326	Vegetated	Agaricomycetes
OTU_560	Vegetated	*Mycena valida*
OTU_187	Vegetated	*Stropharia aeruginosa*
OTU_142	Vegetated	*Gymnopus pinophilus*
OTU_73	Vegetated	*Asterodon ferruginosus*
OTU_322	Vegetated	*Psathyrella globosivelata*
OTU_76	Vegetated	*Rugosomyces persicolor*
OTU_223	Vegetated	*Conocybe pallidospora*
OTU_361	Vegetated	Orbiliomycetes
OTU_101	Vegetated	Orbiliomycetes
OTU_243	Vegetated	Ascomycota
OTU_13	Vegetated	*Cristinia helvetica*
OTU_214	Vegetated	*Conocybe*
OTU_93	Vegetated	*Drechslera*
OTU_221	Vegetated	Orbiliomycetes
OTU_241	Vegetated	Gloniaceae
OTU_625	Vegetated	Agaricomycetes
OTU_272	Vegetated	Stephanosporaceae
OTU_237	Vegetated	Basidiomycota
OTU_327	Vegetated	*Coprinopsis brunneofibrillosa*
OTU_254	Vegetated	Unknown fungi
OTU_39	Vegetated	*Conocybe fuscimarginata*
OTU_267	Vegetated	Ascomycota

Table 6. Protist bioindicators of fallow and vegetated treatments from 1 year-old soils identified by DESeq2 analyses

OTU Identity	Treatment	Taxa	OTU Identity	Treatment	Taxa
OTU_3227	Fallow	P34.6 (Stramenopiles)	**OTU_5704**	Vegetated	Elev-18S-1089 (Apicomplexa)
OTU_4152	Fallow	*Monosiga ovata*	**OTU_4463**	Vegetated	Peritrichia
OTU_2463	Fallow	*Monocystis*	**OTU_5525**	Vegetated	Novel Clade Gran-6 (Cercozoa)
OTU_1065	Fallow	Chrysophyceae	**OTU_419**	Vegetated	*Phytophthora infestans* (T30-4)
OTU_3219	Fallow	*Gymnophrys*	**OTU_2185**	Vegetated	*Cercomonas*
OTU_7192	Fallow	Oomycetes	**OTU_3406**	Vegetated	*Rhogostoma*
OTU_7144	Fallow	Thecofilosea	**OTU_5109**	Vegetated	Colpodida
OTU_2126	Fallow	P34.6 (Stramenopiles)	**OTU_1120**	Vegetated	*Spongomonas*
OTU_95	Fallow	*Rhogostoma*	**OTU_420**	Vegetated	*Phytophthora infestans* (T30-4)
OTU_2201	Fallow	Thecofilosea	**OTU_2589**	Vegetated	Glissomonadida
OTU_3487	Fallow	Thaumatomonadidae	**OTU_412**	Vegetated	*Phytophthora infestans* (T30-4)
OTU_4185	Fallow	Cercozoa (sp. DDB-2008c)	**OTU_1187**	Vegetated	Oomycetes
OTU_1373	Fallow	*Micronuclearia podoventralis*	**OTU_1894**	Vegetated	Trebouxiophyceae
OTU_3376	Fallow	Thecofilosea	**OTU_5676**	Vegetated	*Paulinella*
OTU_4190	Fallow	*Gymnophrys*	**OTU_8939**	Vegetated	E-A1 (Thecofilosea)
OTU_4330	Fallow	Haptoria	**OTU_1121**	Vegetated	*Pseudoplatyophyra*
OTU_6262	Fallow	Thecofilosea	**OTU_1746**	Vegetated	Chloroplastida
OTU_3752	Fallow	*Gymnophrys*	**OTU_230**	Vegetated	E-A1 (Thecofilosea)
OTU_875	Fallow	Oomycetes	**OTU_4836**	Vegetated	Haptoria
OTU_2065	Fallow	*Eocercomonas tribula*	**OTU_4430**	Vegetated	Bicosoecida
OTU_1453	Fallow	Novel Clade Gran-5 (Cercozoa)	**OTU_2652**	Vegetated	*Chromera*
OTU_1530	Fallow	Vampyrellidae	**OTU_889**	Vegetated	*Cercomonas*
OTU_3085	Fallow	Bicosoecida gen. 1 sp. EK-2010a	**OTU_4237**	Vegetated	Discicristoidea
OTU_4408	Fallow	*Gymnophrys*	**OTU_2825**	Vegetated	*Cercomonas fastiga*
OTU_4743	Fallow	*Gymnophrys*	**OTU_5113**	Vegetated	MAST-12C (Stramenopiles)
OTU_3448	Fallow	Vampyrellidae	**OTU_1610**	Vegetated	*Peregrinia clavideferens*
OTU_1050	Fallow	*Monosiga ovata*	**OTU_3646**	Vegetated	Thecofilosea
OTU_3742	Fallow	*Heteromita*	**OTU_7006**	Vegetated	LEMD098 (Apicomplexa)
OTU_4057	Fallow	Novel Clade Gran-4 (Cercozoa)	**OTU_3959**	Vegetated	*Salpingoeca*
OTU_4149	Fallow	*Bodomorpha*	**OTU_4889**	Vegetated	Oomycetes
OTU_1449	Fallow	Thecofilosea	**OTU_4556**	Vegetated	*Cercomonas*
OTU_2275	Fallow	Thaumatomonadidae	**OTU_1965**	Vegetated	*Sterkiella*
OTU_2719	Fallow	*Pseudodifflugia*	**OTU_2179**	Vegetated	*Stenophora robusta*
OTU_2610	Fallow	Novel Clade Gran-4 (Cercozoa)	**OTU_2212**	Vegetated	13-1.8 (Cercozoa)
OTU_1978	Fallow	*Heteromita*	**OTU_1517**	Vegetated	*Cercomonas*
OTU_2347	Fallow	Cercomonadidae	**OTU_5200**	Vegetated	*Heteromita*
OTU_4859	Fallow	Cercozoa	**OTU_4825**	Vegetated	*Lecythium*
OTU_2149	Fallow	MAST-12C (Stramenopiles)	**OTU_2538**	Vegetated	*Colpoda*
OTU_2052	Vegetated	*Hypotrichia*	**OTU_5302**	Vegetated	Thecofilosea
OTU_4802	Vegetated	*Arcuospathidium namibiense tristicha*	**OTU_1630**	Vegetated	JBNA46 (Stramenopiles)

Table 7. Protist bioindicators of fallow and vegetated treatments from 10 year-old soils identified by DESeq2 analyses.

OTU Identity	Treatment	Taxa	OTU Identity	Treatment	Taxa
OTU_2910	Fallow	MAST-12C (Stramenopiles)	**OTU_2677**	Vegetated	*Cercomonas*
OTU_3290	Fallow	*Gymnophrys* (sp. DDB-2008c)	**OTU_4361**	Vegetated	*Woronina*
OTU_2164	Fallow	Cercomonadidae	**OTU_1497**	Vegetated	*Aplanochytrium*
OTU_874	Fallow	Trebouxiophyceae	**OTU_4920**	Vegetated	*Sorodiplophrys*
OTU_2748	Fallow	*Cercomonas*	**OTU_9188**	Vegetated	*Nudifila*
OTU_3460	Fallow	*Developayella elegans*	**OTU_1138**	Vegetated	*Cercomonas*
OTU_2531	Fallow	*Nudifila* (*Incertae sedis*)	**OTU_1671**	Vegetated	Bacillariophyceae
OTU_2495	Fallow	*Cercomonas*	**OTU_1791**	Vegetated	*Holosticha*
OTU_3210	Fallow	*Euglypha*	**OTU_5903**	Vegetated	Glissomonadida
OTU_8741	Fallow	*Trachelius*	**OTU_4828**	Vegetated	*Cercomonas fastiga*
OTU_2098	Fallow	Thecofilosea	**OTU_1997**	Vegetated	*Woronina*
OTU_4855	Fallow	Haptoria	**OTU_6059**	Vegetated	*Gymnophrys*
OTU_2213	Fallow	13-1.8 (Cercozoa)	**OTU_4334**	Vegetated	*Massisteria*
OTU_2417	Fallow	*Gymnophrys*	**OTU_1039**	Vegetated	*Colpodida*
OTU_1978	Fallow	*Heteromita*	**OTU_2228**	Vegetated	*Cercomonas*
OTU_2254	Fallow	13-1.8 (Cercozoa)	**OTU_4598**	Vegetated	*Copromyxa* (sp. PKD 2011)
OTU_2347	Fallow	Cercomonadidae	**OTU_707**	Vegetated	*Schizoplasmodiopsis* (sp. F3)
OTU_3315	Fallow	Bicosoecida gen. 1 sp. EK-2010a	**OTU_2202**	Vegetated	*Cercomonas*
OTU_5394	Fallow	*Cercomonas*	**OTU_932**	Vegetated	Ulvophyceae
OTU_4273	Fallow	*Cercomonas*	**OTU_4693**	Vegetated	Vampyrellidae
OTU_2152	Fallow	Eocercomonas sp. HFCC 908	**OTU_1452**	Vegetated	*Gymnophrys*
OTU_266	Fallow	Oomycetes	**OTU_9024**	Vegetated	*Cercomonas*
OTU_3227	Fallow	P34.6 (Stramenopiles)	**OTU_1956**	Vegetated	Chloroplastida
OTU_4176	Fallow	H1-10 (Centrohelida)	**OTU_1043**	Vegetated	Pseudoperkinsidae
OTU_1503	Fallow	*Sorodiplophrys*	**OTU_2097**	Vegetated	*Monocystis*
OTU_4819	Fallow	*Cercomonas*	**OTU_2299**	Vegetated	*Vorticella*
OTU_6494	Fallow	*Gymnophrys*	**OTU_3300**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_2263	Fallow	*Heteromita*	**OTU_4251**	Vegetated	*Cercomonas*
OTU_4933	Fallow	*Heteromita*	**OTU_5323**	Vegetated	*Parietochloris*
OTU_2344	Fallow	*Heteromita*	**OTU_5855**	Vegetated	Thecofilosea
OTU_4881	Fallow	Cyrtolophosidida	**OTU_563**	Vegetated	*Woronina*
OTU_3757	Fallow	*Gymnophrys*	**OTU_117**	Vegetated	*Dictyochloropsis*
OTU_226	Fallow	Haptoria	**OTU_2652**	Vegetated	*Chromera*
OTU_5110	Fallow	*Bodomorpha*	**OTU_3055**	Vegetated	*Maryna*
OTU_4968	Fallow	Bicosoecida	**OTU_951**	Vegetated	*Chloromonas vernalis*
OTU_2126	Fallow	P34.6 (Stramenopiles)	**OTU_4870**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_7144	Fallow	Thecofilosea	**OTU_1512**	Vegetated	Phytomyxea
OTU_250	Fallow	Trebouxiophyceae	**OTU_5020**	Vegetated	Thaumatomonadidae
OTU_4755	Fallow	Vampyrellidae	**OTU_5281**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_4253	Fallow	*Cercomonas*	**OTU_159**	Vegetated	Salpingoecidae
OTU_863	Fallow	MAST-12C (Stramenopiles)	**OTU_919**	Vegetated	*Chloromonas augustae*

OTU_8255	Fallow	P34.6 (Stramenopiles)	**OTU_3479**	Vegetated	*Maryna*
OTU_1943	Fallow	Opisthokonta	**OTU_4375**	Vegetated	Thecofilosea
OTU_2957	Fallow	*Cercomonas*	**OTU_5063**	Vegetated	*Heteromita*
OTU_2841	Fallow	*Gymnophrys*	**OTU_4927**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_9307	Fallow	*Heteromita*	**OTU_5054**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_6485	Fallow	*Heteromita* (sp. A10p145at5)	**OTU_3845**	Vegetated	*Cercomonas*
OTU_4958	Fallow	Vampyrellidae	**OTU_5100**	Vegetated	*Heteromita*
OTU_2678	Fallow	Glissomonadida	**OTU_846**	Vegetated	*Capsaspora*
OTU_3850	Fallow	Amb-18S-1124 (Glissomonadida)	**OTU_5805**	Vegetated	Cercozoa (sp. ATCC 50530)
OTU_4997	Fallow	*Allantion*	**OTU_4286**	Vegetated	13-1.8 (Cercozoa)
OTU_2992	Fallow	*Sandona mutans*	**OTU_1407**	Vegetated	*Hartmannella cantabrigiensis*
OTU_1491	Fallow	*Hypotrichia*	**OTU_1265**	Vegetated	*Dictyostelium*
OTU_712	Fallow	Cercomonadidae	**OTU_3261**	Vegetated	Pseudoperkinsidae
OTU_128	Fallow	Bicosoeca	**OTU_5075**	Vegetated	Bicosoecida
OTU_3851	Fallow	H1-10 (Centrohelida)	**OTU_3124**	Vegetated	*Bodomorpha*
OTU_2201	Fallow	Thecofilosea	**OTU_2301**	Vegetated	*Nudifila*
OTU_2870	Fallow	*Cercomonas*	**OTU_842**	Vegetated	*Vermamoeba*
OTU_4073	Fallow	*Cercomonas*	**OTU_3446**	Vegetated	*Heteromita*
OTU_1282	Fallow	Chlorophyta	**OTU_2311**	Vegetated	*Bryometopus*
OTU_3117	Fallow	Euamoebida	**OTU_6620**	Vegetated	*Ceromonas fastiga*
OTU_3631	Fallow	CCW10 (Cercozoa)	**OTU_2249**	Vegetated	Novel Clade 3 (Cercozoa)
OTU_2370	Fallow	*Colopodida*	**OTU_1725**	Vegetated	Amb-18S-504 (Chlorophyta)
OTU_3614	Fallow	Eocercomonas (sp. HFCC 908)	**OTU_857**	Vegetated	*Colpodida*
OTU_5430	Fallow	Amb-18S-1124 (Glissomonadida)	**OTU_1344**	Vegetated	*Didymium*
OTU_872	Fallow	*Cercomonas*	**OTU_4679**	Vegetated	Oomycetes
OTU_4798	Fallow	*Metopion*	**OTU_5588**	Vegetated	Cercomonadida
OTU_2375	Fallow	Eocercomonas (sp. HFCC 908)	**OTU_2719**	Vegetated	*Pseudodiffugia*
OTU_928	Fallow	Chrysophyceae	**OTU_922**	Vegetated	Litostomatea
OTU_4329	Fallow	Discicristoidea	**OTU_334**	Vegetated	*Sellaphora minima*
OTU_223	Fallow	Eocercomonas (sp. HFCC 908)	**OTU_2384**	Vegetated	Eugregarinorida
OTU_5288	Fallow	*Cercomonas fastiga*	**OTU_2343**	Vegetated	*Cercomonas*
OTU_4926	Fallow	*Gymnophrys*	**OTU_2983**	Vegetated	H1-10 (Centrohelida)
OTU_2237	Fallow	Oxytrichidae	**OTU_1075**	Vegetated	*Prasiola furfuracea*
OTU_5025	Fallow	*Heteromita*	**OTU_7373**	Vegetated	*Colpodida*
OTU_3752	Fallow	*Gymnophrys*	**OTU_2040**	Vegetated	Trebouxiophyceae
OTU_4866	Fallow	*Gymnophrys* (DDB-2008c)	**OTU_2386**	Vegetated	*Poterioochromonas*
OTU_2630	Fallow	*Heteromita*	**OTU_745**	Vegetated	Chlorophyceae
OTU_3084	Fallow	Eocercomonas (sp. HFCC 908)	**OTU_9183**	Vegetated	E-A1 (Thecofilosea)
OTU_4333	Fallow	*Cavernomonas*	**OTU_215**	Vegetated	Elev-18S-1089 (Apicomplexa)
OTU_2275	Fallow	Thaumatomonadidae	**OTU_5662**	Vegetated	*Cercomonas*
OTU_4152	Fallow	*Monosiga ovata*	**OTU_3619**	Vegetated	*Metopion*
OTU_2100	Fallow	Cercomonadidae	**OTU_4581**	Vegetated	*Colpodida*
OTU_3503	Fallow	*Cercomonas*	**OTU_209**	Vegetated	*Apusomonas*
OTU_4067	Fallow	*Peregrinia clavideferens*	**OTU_2466**	Vegetated	LG21-05 (Chrysophyceae)

OTU_8463	Fallow	Thecofilosea	**OTU_1895**	Vegetated	*Aphanomyces*
OTU_21	Vegetated	Amb-18S-431 (Thecofilosea)	**OTU_7501**	Vegetated	Thecofilosea
OTU_1776	Vegetated	Elev-18S-1089 (Eugregarinorida)	**OTU_2061**	Vegetated	*Cercomonas fastiga*
OTU_2447	Vegetated	*Cercomonas fastiga*	**OTU_1878**	Vegetated	*Colpodida*
OTU_4385	Vegetated	Chlorophyceae	**OTU_4995**	Vegetated	Oomycetes
OTU_412	Vegetated	*Phytophthora infestans* T30-4	**OTU_290**	Vegetated	*Cryptodifflugia*
OTU_4164	Vegetated	*Cercomonas fastiga*	**OTU_1136**	Vegetated	*Platyophrya*
OTU_897	Vegetated	Oxytrichidae	**OTU_1610**	Vegetated	*Peregrinia clavideferens*
OTU_2260	Vegetated	Chlorophyceae	**OTU_6615**	Vegetated	Cercozoa (sp. ATCC 50530)
OTU_1156	Vegetated	Trebouxiophyceae	**OTU_2419**	Vegetated	Haptoria
OTU_1354	Vegetated	*Chloroidium*	**OTU_3317**	Vegetated	*Cercomonas*
OTU_4330	Vegetated	Haptoria	**OTU_1884**	Vegetated	Trebouxiophyceae
OTU_2186	Vegetated	*Cercomonas fastiga*	**OTU_2816**	Vegetated	Acanthocystidae
OTU_2716	Vegetated	Haptoria	**OTU_6883**	Vegetated	*Heteromita*
OTU_4889	Vegetated	Oomycetes	**OTU_274**	Vegetated	*Hantzchia*
OTU_2274	Vegetated	Glissomonadida	**OTU_7479**	Vegetated	Haptoria
OTU_3963	Vegetated	LEMD098 (Apicomplexa)	**OTU_3381**	Vegetated	*Colpodida*
OTU_7988	Vegetated	LEMD098 (Apicomplexa)	**OTU_5089**	Vegetated	*Heteromita*
OTU_1218	Vegetated	Mischococcales	**OTU_3817**	Vegetated	Cercozoa (sp. ATCC 50530)
OTU_419	Vegetated	*Phytophthora infestans* T30-4	**OTU_4974**	Vegetated	*Ciliophrys infusionum*
OTU_4802	Vegetated	*Arcuospathidium namibiense tristicha*	**OTU_2338**	Vegetated	*Allanition*
OTU_8323	Vegetated	Pseudoperkinsidae	**OTU_5687**	Vegetated	Novel Clade Gran-6 (Cercozoa)
OTU_1568	Vegetated	Trebouxiophyceae	**OTU_4314**	Vegetated	*Cercomonas*
OTU_5956	Vegetated	*Monocystis*	**OTU_24**	Vegetated	*Flamella*
OTU_4237	Vegetated	Discicristoidea	**OTU_4367**	Vegetated	*Echinamoeba thermarum*
OTU_2185	Vegetated	*Cercomonas*	**OTU_4989**	Vegetated	*Colpodida*
OTU_1854	Vegetated	*Paraphysomonas*	**OTU_4545**	Vegetated	BOLA868 (Euamoebida)
OTU_754	Vegetated	*Gymnodinium*	**OTU_3025**	Vegetated	*Cercomonas*
OTU_3134	Vegetated	Vampyrellidae	**OTU_5044**	Vegetated	*Gymnophrys*
OTU_1785	Vegetated	Thecofilosea	**OTU_2304**	Vegetated	13-1.8 (Cercozoa)
OTU_3596	Vegetated	Haptoria	**OTU_5452**	Vegetated	Novel Clade Gran-6 (Cercozoa)
OTU_4896	Vegetated	*Allantion*	**OTU_2259**	Vegetated	*Cercomonas*
OTU_2182	Vegetated	Euglyphida	**OTU_4849**	Vegetated	Chrysophyceae
OTU_5287	Vegetated	*Metopion*	**OTU_5920**	Vegetated	*Eocercomonas tribula*
OTU_91	Vegetated	Perkinsidae	**OTU_1761**	Vegetated	Thecofilosea
OTU_1510	Vegetated	*Hypotrichia*	**OTU_6702**	Vegetated	Perkinsidae
OTU_2059	Vegetated	*Cercomonas ambigua*	**OTU_3073**	Vegetated	Opisthokonta
OTU_46	Vegetated	Haptoria	**OTU_4280**	Vegetated	*Nudifila*
OTU_3406	Vegetated	*Rhogostoma*	**OTU_875**	Vegetated	Oomycetes
OTU_649	Vegetated	*Hypotrichia*	**OTU_3646**	Vegetated	Thecofilosea
OTU_5057	Vegetated	Trebouxiophyceae	**OTU_2243**	Vegetated	Cercozoa
OTU_858	Vegetated	Oomycetes	**OTU_1501**	Vegetated	Chrysophyceae
OTU_2326	Vegetated	Phytomyxea	**OTU_6479**	Vegetated	Vampyrella
OTU_1263	Vegetated	Haptoria	**OTU_3501**	Vegetated	Bacillariophyceae
OTU_1950	Vegetated	Oxytrichidae	**OTU_4342**	Vegetated	Novel Clade Gran-3

					(Cercozoa)
OTU_2663	Vegetated	Pseudoperkinsidae	**OTU_405**	Vegetated	*Chloroidium saccharophilum*
OTU_2380	Vegetated	*Cardiostomatella*	**OTU_703**	Vegetated	*Ischnamoeba* sp. FN352
OTU_5193	Vegetated	*Paraphysomonas*	**OTU_5201**	Vegetated	SCM37C52 (Alveolata)
OTU_154	Vegetated	*Bryometopus*	**OTU_230**	Vegetated	E-A1 (Thecofilosea)
OTU_2771	Vegetated	Thecofilosea	**OTU_933**	Vegetated	Chlorophyta
OTU_4912	Vegetated	*Cercomonas*	**OTU_1962**	Vegetated	*Amastigomonas* sp. IVY8c
OTU_3910	Vegetated	Bacillariophyceae	**OTU_2547**	Vegetated	*Hartmanella*
OTU_2613	Vegetated	Thecofilosea	**OTU_5498**	Vegetated	*Diadesmis gallica*
OTU_3956	Vegetated	*Homalogastra setosa*	**OTU_930**	Vegetated	*Halteria*
OTU_496	Vegetated	Jakobida	**OTU_3578**	Vegetated	Pseudoperkinsidae
OTU_8023	Vegetated	*Colpodida*	**OTU_2916**	Vegetated	Euamoebida
OTU_2252	Vegetated	*Heteromita* (sp. B128)	**OTU_1618**	Vegetated	*Strichococcus jenerensis*
OTU_4628	Vegetated	Perkinsidae (A31)	**OTU_1778**	Vegetated	Chlorophyceae
OTU_3637	Vegetated	Trebouxiophyceae	**OTU_4300**	Vegetated	Amb-18S-1124 (Glissomonadida)
OTU_5222	Vegetated	Colpodea	**OTU_6653**	Vegetated	*Cercomonas*
OTU_2463	Vegetated	*Monocystis*	**OTU_5504**	Vegetated	*Heteromita*
OTU_403	Vegetated	Trebouxiophyceae	**OTU_3109**	Vegetated	E-A1 (Thecofilosea)
OUT_5354	Vegetated	*Bodomorpha*			

Supplementary Figures

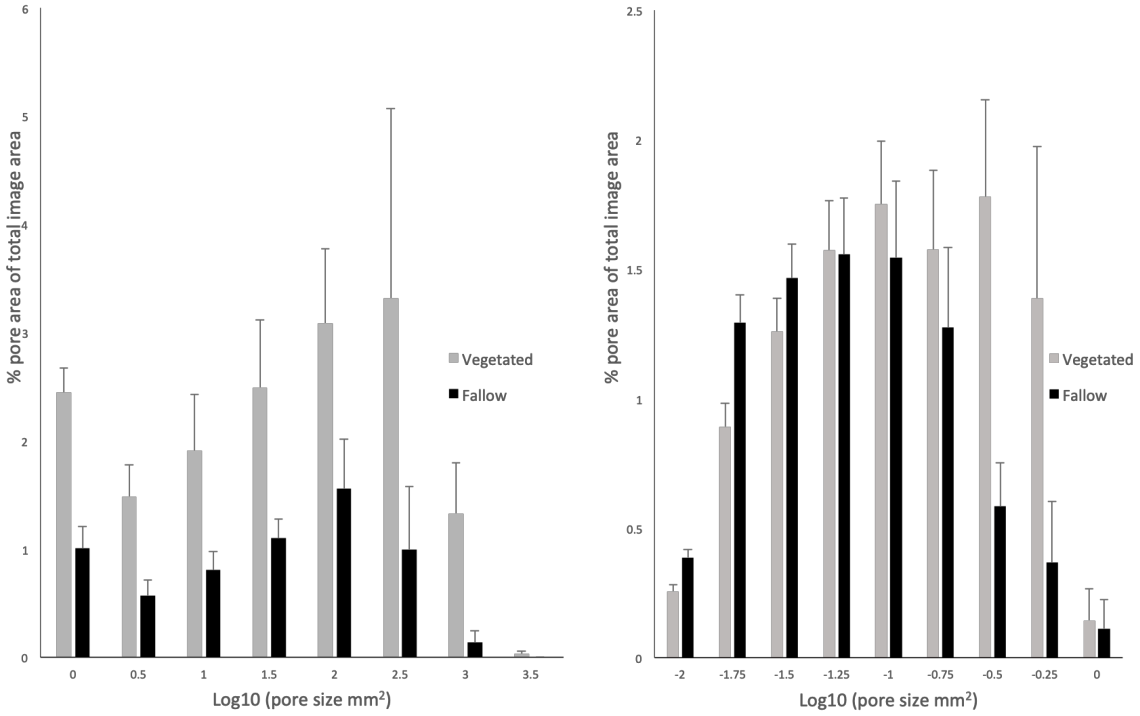


Fig. 1. Mean pore size distribution collected using x-ray μCT on whole column data (A) and aggregate data (B). Error bars = standard error of mean.

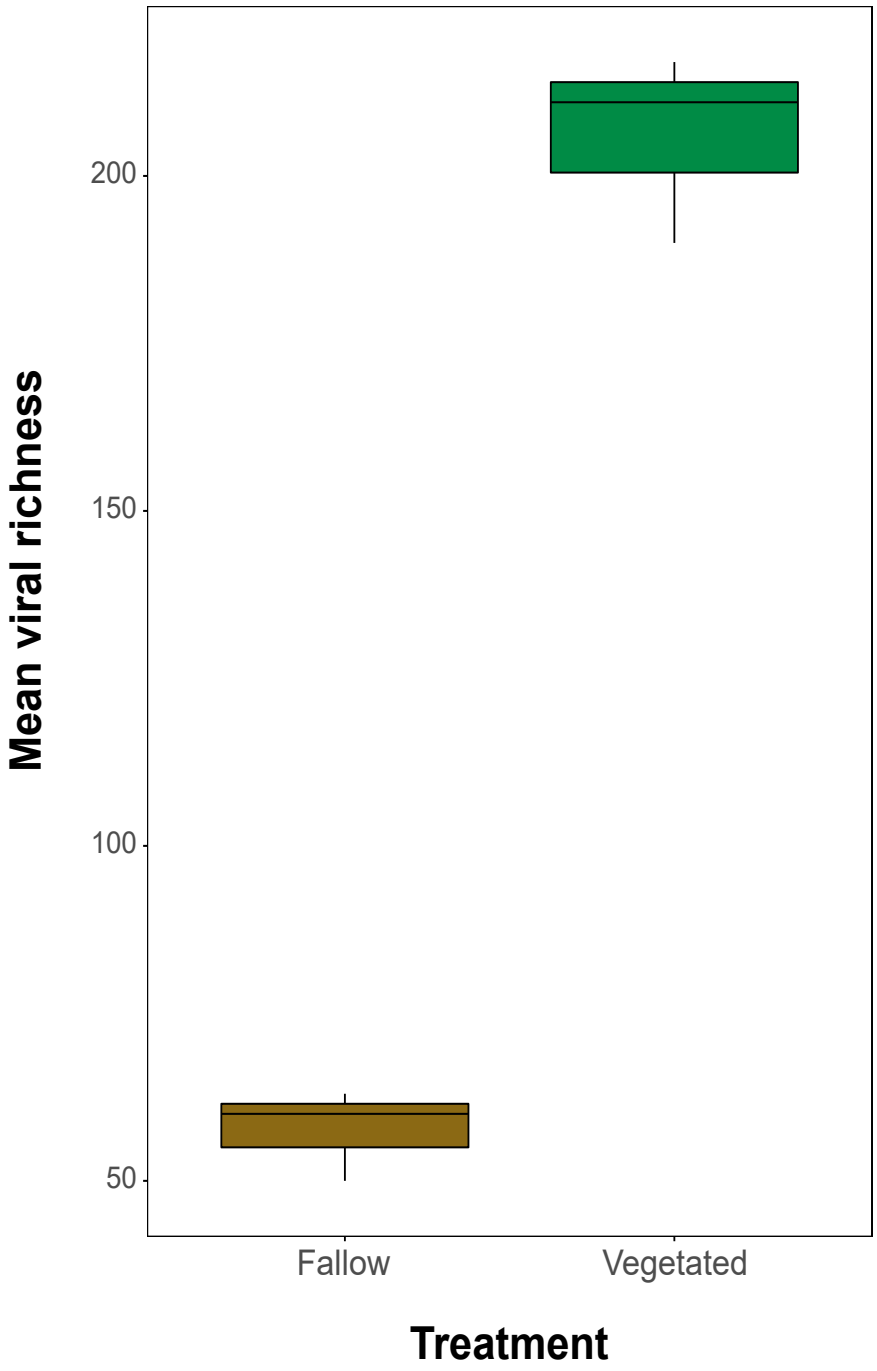


Fig. 2. Viral gene richness derived from Geochip data for 10 year old covered and vegetated soils.

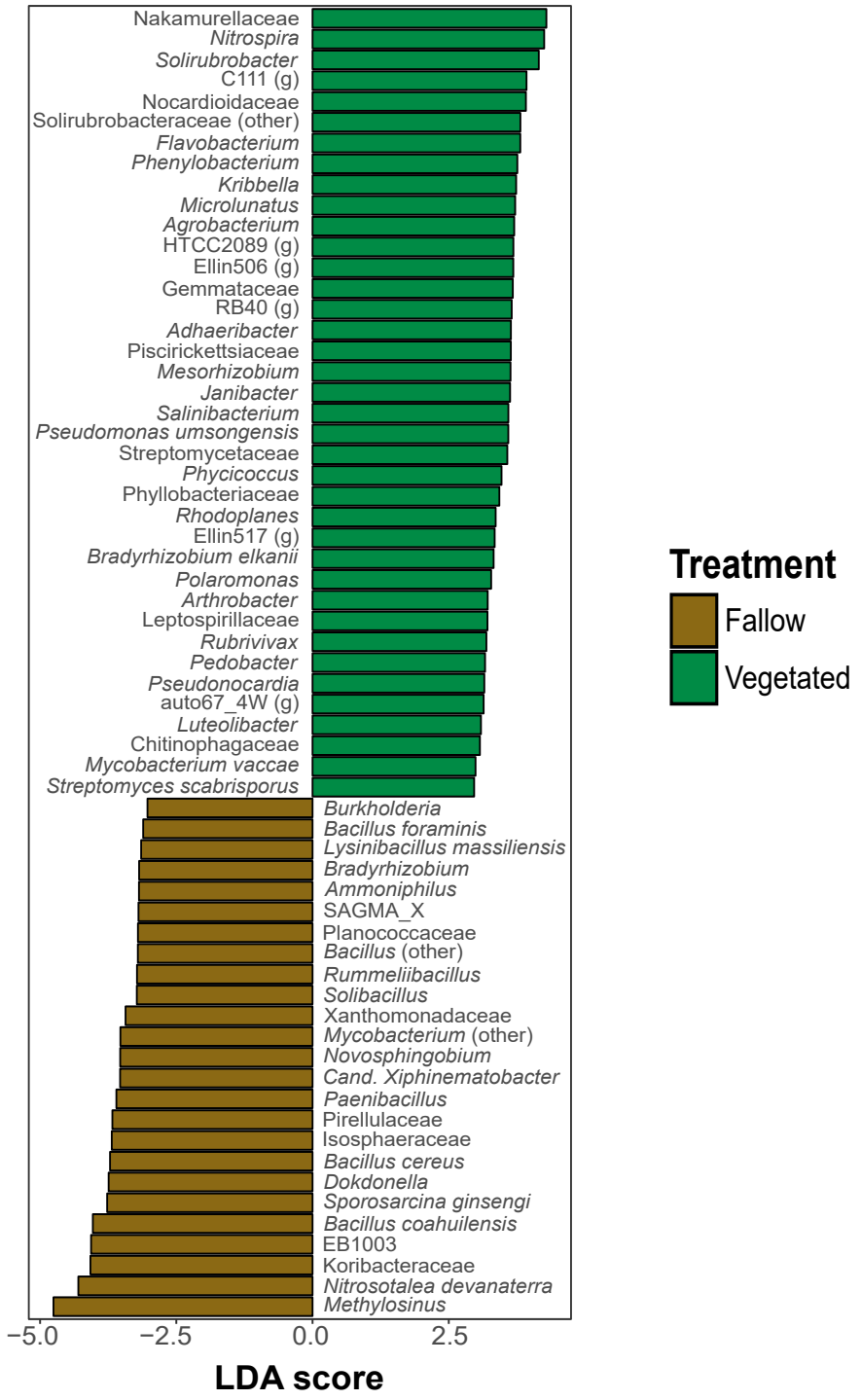


Fig. 3. Complete differential abundance of prokaryotic indicator taxa of covered and vegetated soils identified using linear discriminant analyses (LDA > 2.5) for 10 year-old soils.

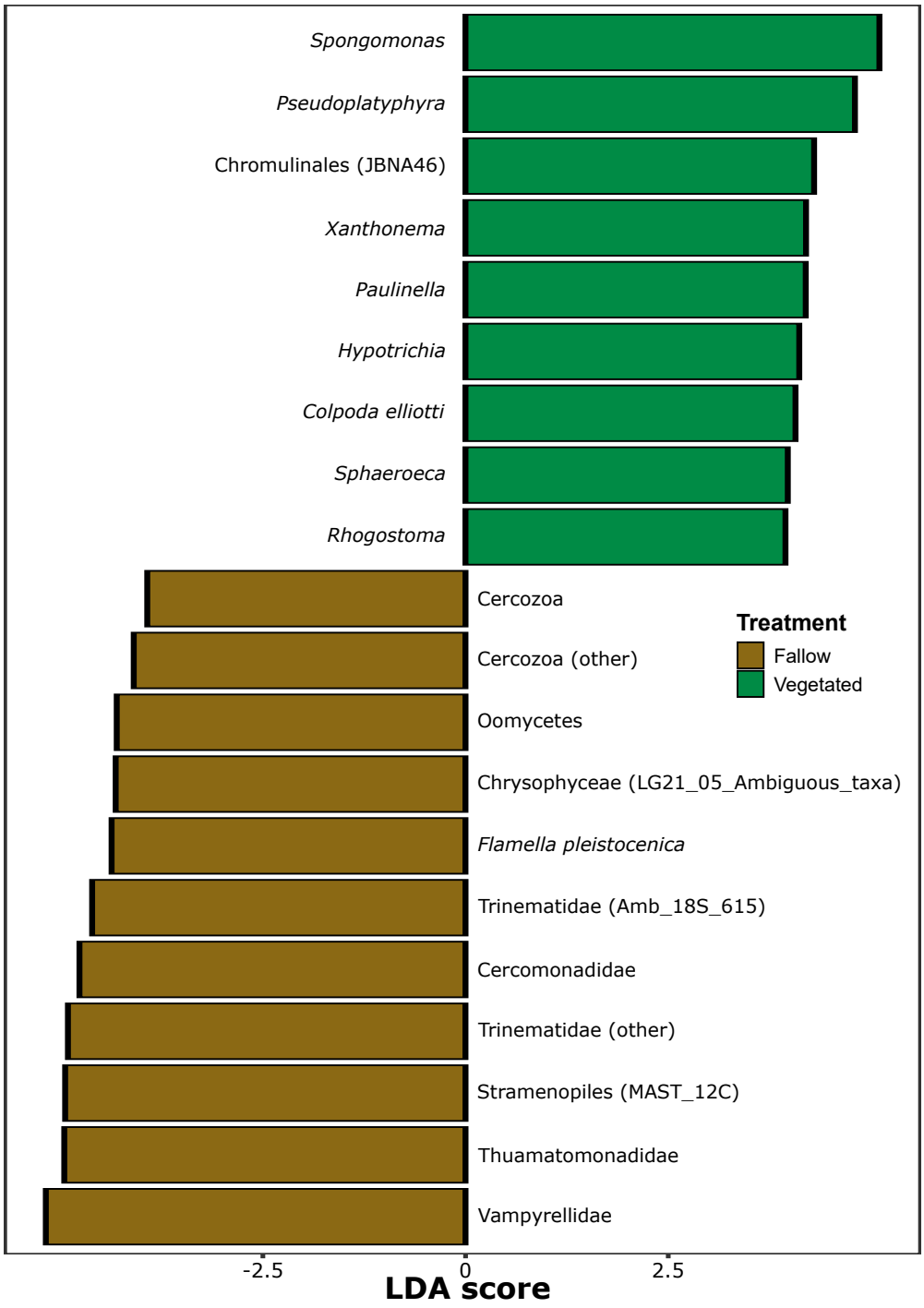


Fig. 4. Complete differential abundance of protistan indicator taxa of covered and vegetated soils identified using linear discriminant analyses (LDA > 2.5) for 1 year-old soils.

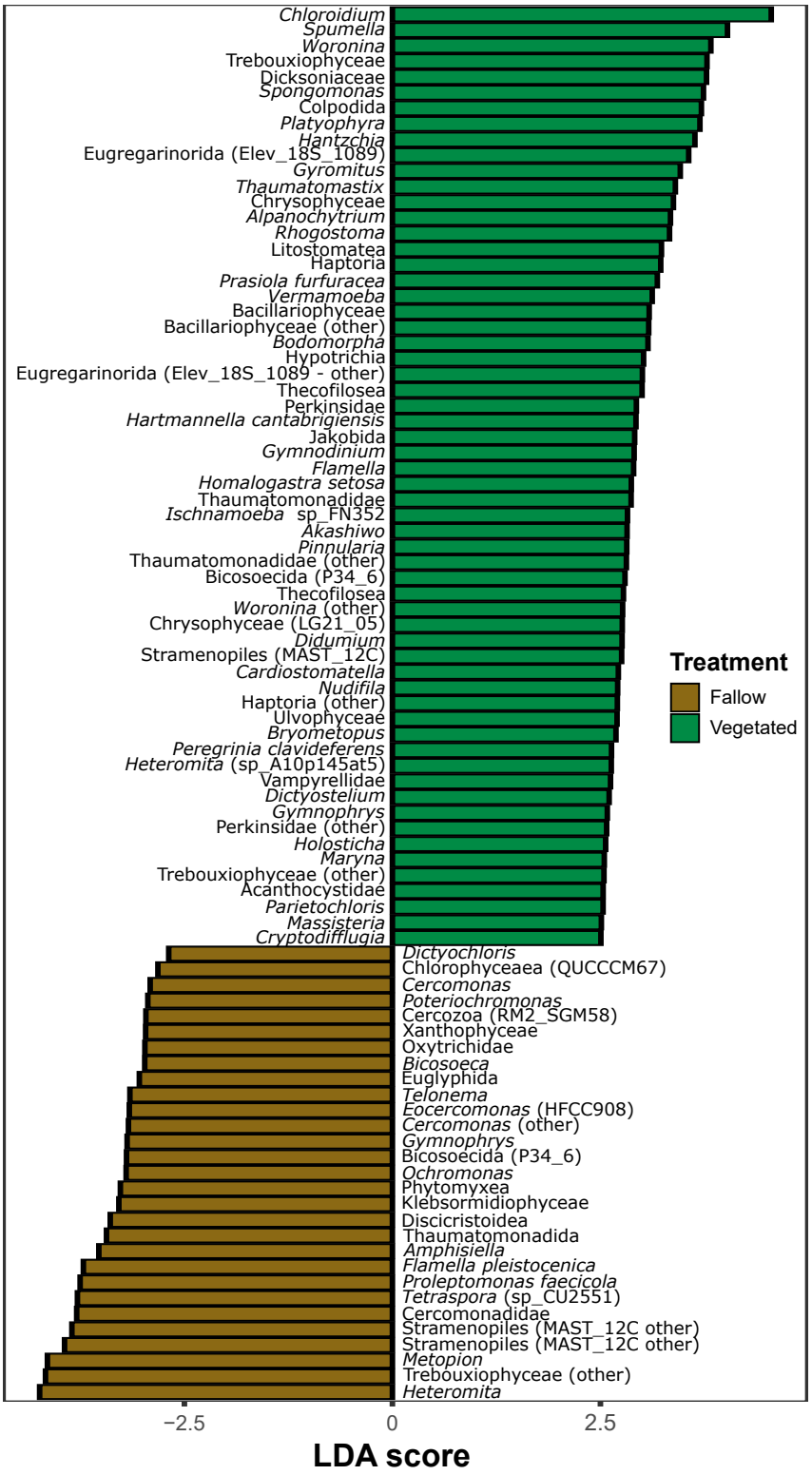


Fig. 5. Complete differential abundance of protistan indicator taxa of covered and vegetated soils identified using linear discriminant analyses (LDA > 2.5) for 10 year-old soils.

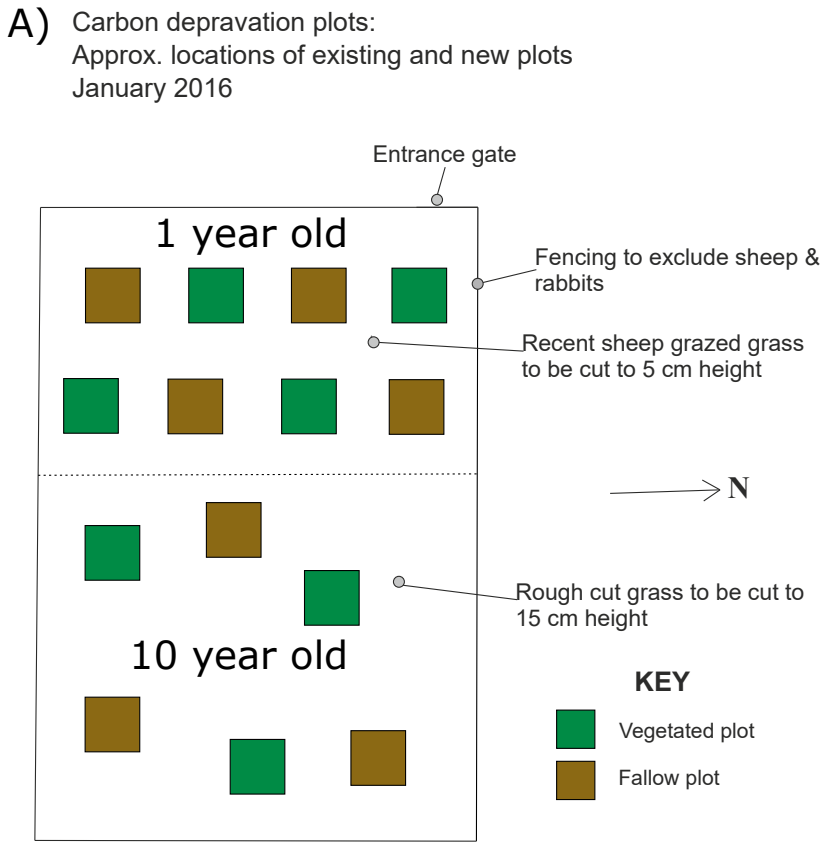


Fig. 6 A) Schematic map of fallow and vegetated plot layout; **B**) a detailed description of the fallow plot design.

Appendix 6

Co-authored paper that uses data generated as part of this thesis published in *Science of the Total Environment*

Plant and soil communities are associated with the response of soil water repellency to environmental stress

Fiona M. Seaton[1,2*], Davey L. Jones[2,3], Simon Creer[2], Paul B.L. George[2,1], Simon M. Smart[4], Inma Lebron[1], Gaynor Barrett[1], Bridget A. Emmett[1] and David A. Robinson[1]

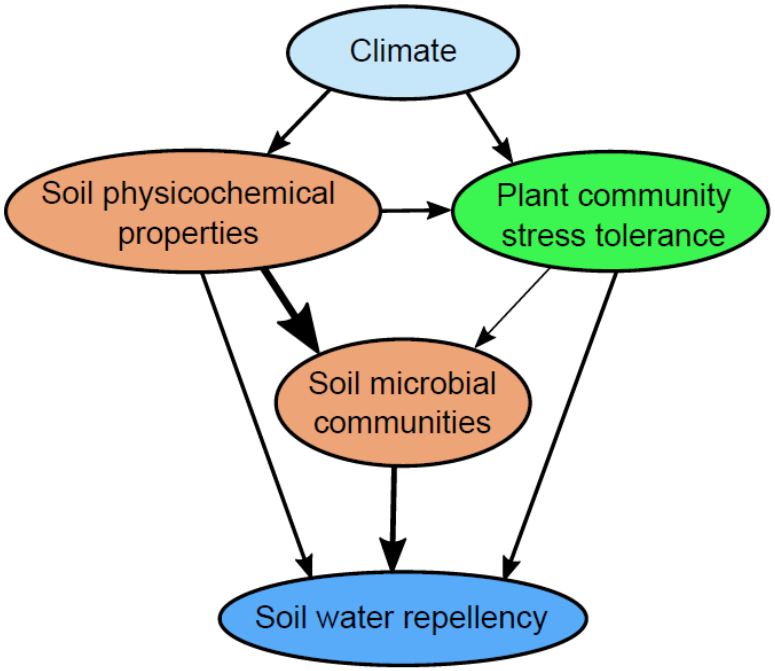
[1] School of Natural Sciences, Bangor University, Bangor, Gwynedd, LL57 2UW, UK

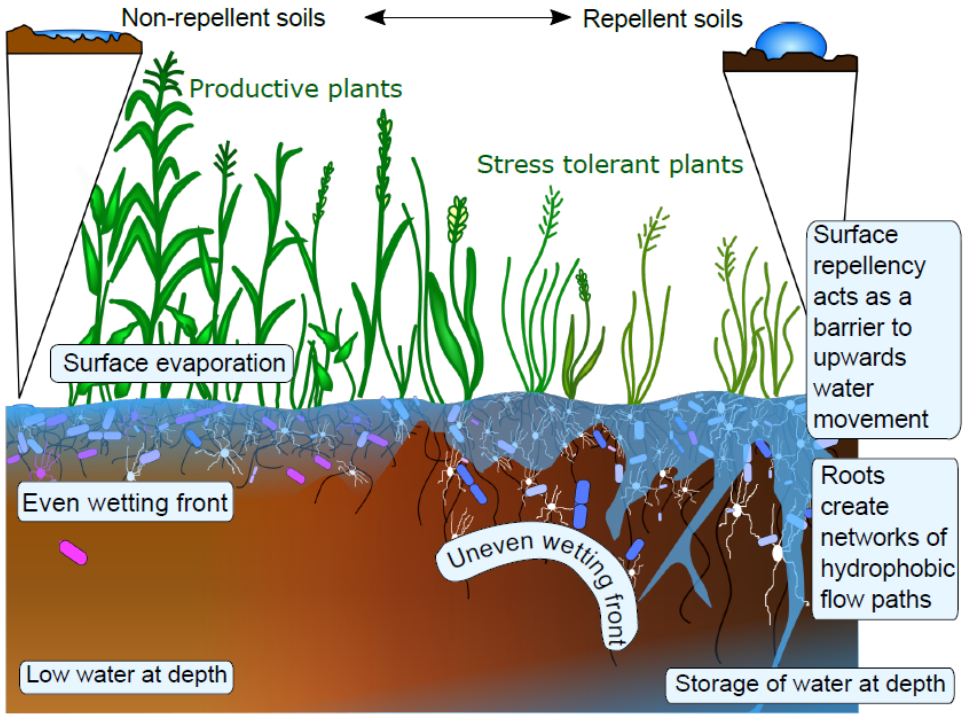
[2] Centre for Ecology & Hydrology, Bangor, Gwynedd, LL57 2UW, UK

[3] SoilsWest, UWA School of Agriculture and Environment, The University of Western Australia, Perth WA 6009, Australia

[4] Centre for Ecology & Hydrology, Lancaster, Bailrigg, Lancaster, LA1 4AP, UK

Study concept and design: Seaton, Robinson. Laboratory analyses: Lebron, Barrett, Seaton. Bioinformatics analyses: George. Analysis and interpretation of data: Seaton, Smart, Jones, Robinson, Creer. Drafting of the manuscript: Seaton. Critical revision of the manuscript for important intellectual content: Robinson, Jones, Creer, Smart. Statistical analysis: Seaton. Obtained funding: Emmett. Study supervision: Robinson, Emmett.


Climate
Soil physicochemical properties
Plant community stress tolerance
Soil microbial communities
Soil water repellency


Non-repellent soils
Repellent soils
Productive plants
Stress tolerant plants
Surface repellency acts as a barrier to upwards water movement
Surface evaporation
Even wetting front
Uneven wetting front
Roots create networks of hydrophobic flow paths
Low water at depth
Storage of water at depth

Abstract

A warming climate and expected changes in average and extreme rainfall emphasise the importance of understanding how the land surface routes and stores surface water. The availability and movement of water within an ecosystem is a fundamental control on biological and geophysical activity, and influences many climatic feedbacks. A key phenomenon influencing water infiltration into the land surface is soil hydrophobicity, or water repellency. Despite repellency dictating the speed, volume and pattern of water infiltration, there is still major uncertainty over whether this critical hydrological process is biologically or physicochemically controlled. Here we show that soil water repellency is likely driven by changes in the plant and soil microbial communities in response to environmental stressors. We carried out a field survey in the summers of 2013 to 2016 in a variety of temperate habitats ranging across arable, grassland, forest and bog sites. We found that moderate to extreme repellency occurs in 68% of soils at a national scale in temperate ecosystems, with 92% showing some repellency. Taking a systems approach, we show that a wetter climate and low nutrient availability alter plant, bacterial and fungal community structure, which in turn are associated with increased soil water repellency across a large-scale gradient of soil, vegetation and land-use. The stress tolerance of the plant community and associated changes in soil microbial communities were more closely linked to changes in repellency than soil physicochemical properties. Our results indicate that there are consistent responses to diverse ecosystem stresses that will impact plant and microbial community composition, soil properties, and hydrological behaviour. We suggest that the ability of a biological community to induce such hydrological responses will influence the resilience of the whole ecosystem to

environmental stress. This highlights the crucial role of above-belowground interactions in mediating climatic feedbacks and dictating ecosystem health.

1. Introduction

The frequency and intensity of extreme climatic events is predicted to increase over the next century and beyond (IPCC, 2014). Soil moisture has been shown to have major implications for carbon storage and related climatic feedbacks (Green et al., 2019), therefore it is more important than ever to understand how the flow of water interacts with ecosystem health and the mechanisms controlling water fluxes at the land-atmosphere interface. There are still many uncertainties surrounding how water, soil, and vegetation will respond to the escalation of climatic stress in addition to prevailing land use stresses. Resilience to change varies between ecosystems, yet in most cases resilience and recovery only occur within limits and are less likely under multiple stressors (Côté, Darling, & Brown, 2016). Biological communities shift in response to stress, and soil physicochemical properties change in tandem, creating an overall ecosystem response (van der Putten et al., 2013). Further, the ecosystem response to one stressor, such as drought, may change the response to another, such as flood. Many habitat stressor responses and feedbacks are as yet unknown but are globally important if we are to model and predict impacts helping to mitigate ecosystem damage (Robinson et al., 2019).

Soil water repellency fundamentally changes the way water infiltrates and moves through the soil. A water repellent (hydrophobic) soil is defined by the behaviour of liquid on the soil surface, with repellent soils causing water drops to bead and resist capillary absorption. Previous seminal work on water repellency has emphasised its impact on hydrological processes through increasing surface runoff and soil erodibility,

predominantly in fire driven systems (S.H. Doerr, Shakesby, & Walsh, 2000; Goebel, Bachmann, Reichstein, Janssens, & Guggenberger, 2011). To date, it is often negative impacts of repellency associated with crop production, flood risk, water quality and biogeochemical cycling that have been the focus of the literature (Dekker & Ritsema, 1994; S.H. Doerr et al., 2000). However, an emerging body of work provides evidence for the ecological role of repellency in promoting the resilience of plant communities and soil carbon stock to wildfire and drought stress in various ecosystems (Kettridge et al., 2014; Robinson, Lebron, Ryel, & Jones, 2010; Zeppenfeld et al., 2017). Water repellency has been shown to induce unsaturated preferential flow of water into the soil rather than piston flow in many soils (Dekker & Ritsema, 1994; Rye & Smettem, 2017). Of the 17 ecosystem service categories identified by Costanza *et al.* (1997), twelve benefit from preferential flow and three are affected detrimentally (Clothier, Green, & Deurer, 2008).

Water repellency induces increased runoff if the soil has no macropores and unsaturated preferential flow of water into the soil, rather than piston flow, in the presence of macropores (Dekker & Ritsema, 1994). The partitioning between preferential flow and surface run-off will depend on a number of factors in addition to the degree of repellency, e.g. texture, macropore density the topography of the area and the spatial pattern of repellency, which is often highly spatially heterogeneous (Bodí et al., 2013; S.H. Doerr et al., 2000). With preferential flow, water penetrates deeper into the soil profile by following roots or other macropores generating fingered flow, while with piston flow it penetrates evenly down the soil profile (Bogner, Gaul, Kolb, Schmiedinger, & Huwe, 2010). In an ecosystem where the spatial pattern of plants can adjust to the heterogeneity of infiltration due to repellency, preferential flow can be an advantage. For

example, preferential flow can result in greater storage of water at depth (Rye & Smettem, 2018) which can increase a plant's resilience to drought stress and give an advantage to deep-rooting plants over shallow-rooting plants in drought stressed environments (De Boeck & Verbeeck, 2011; Zeppenfeld et al., 2017). Whereas, in agricultural production systems where the pattern of plants is predetermined and there are limited macropores for the development of preferential flow paths soil moisture spatial heterogeneity and dry spots results in yield loss.

Water repellency is considered to be created by the amount, nature and configuration of soil organic material (Doerr et al., 2000; Mao et al., 2019), yet there is still uncertainty over the origins of the hydrophobic compounds in global soils (Mao, Nierop, Rietkerk, Sinninghe Damsté, & Dekker, 2016; Schaumann et al., 2007; Spohn & Rillig, 2012). Until now, potential mechanisms for inducing water repellency have not been tested at realistic scales, hampering the emergence of a coherent theory across habitat types for the development and persistence of water repellency. In this work we analysed soil repellency across a wide range of habitats (Fig. 1) within a temperate oceanic climate. This wide range of biota within a limited climatic range enabled us to evaluate the relative role of biotic influence on repellency versus soil physicochemical influences, without confounding effects of climate. We characterised the plant community and soil physicochemical properties within 1326 sites, including 425 sites in which the belowground communities were measured, allowing an in-depth look at how the whole ecosystem shifts in tandem with soil hydrological shifts. Given the emerging evidence discussed we hypothesise that:

1) Soil water repellency depends on habitat, particularly showing greater persistence in those habitats that experience environmental stress such as drought and high acidity.
2) Persistence of repellency depends on the microbial community composition, as microbes can adapt to water stress by either becoming repellent or producing repellent compounds to aid water conservation.

We test these hypotheses through the following objectives: (i) measure repellency across habitat types and determine its prevalence; (ii) test the relationship between soil, plant and microbial communities and the persistence of soil repellency; and (iii) explore whether our pre-identified physicochemical and biological variables predict the changes in repellency across land use.

2. Methods

2.1 Field sampling design

We used data collected as part of the Glastir Monitoring and Evaluation Programme (GMEP) field measurement program in Wales, a sampling domain of ~2,000,000 ha comprising varied land use and topography and situated on the oceanic Atlantic seaboard of NW Europe (Emmett & the GMEP team, 2017). There were 300 individual 1 km squares randomly selected from within land classification strata and each included 5 vegetation plots (Fig. 1, Fig. S1). The sites were selected to be representative of the range of habitat types across Wales; consequently, different grassland habitats were sampled extensively, complemented by substantial numbers of woodland and wetland sites (Table S1). Sampling occurred over a five month period across each of the summers of 2013 to 2016, each square was only surveyed once over the four years with different squares being surveyed each year. Every plot had a vegetation survey performed

for a 200 m^2 square and where possible soil samples taken at the south corner of an inner 2m square (Fig. S1). A soil core for physicochemical analysis was taken with a plastic corer of 5 cm diameter down to 15 cm depth. The squares from the first two years of the survey had soil samples for microbiology taken from three randomly selected plots within the square. Soil samples for microbiology were taken using a gouge auger at 5 points around the physicochemical soil core location down to 15 cm, and then bulking together the samples. The surveyors assigned each plot to a habitat according to the Joint Nature Conservation Committee criteria (Jackson, 2000). The main habitats included in this study were: arable; improved grassland; neutral grassland; acid grassland; broadleaved woodland; coniferous woodland; dwarf shrub heath; fen, marsh and swamp; bog; and bracken.

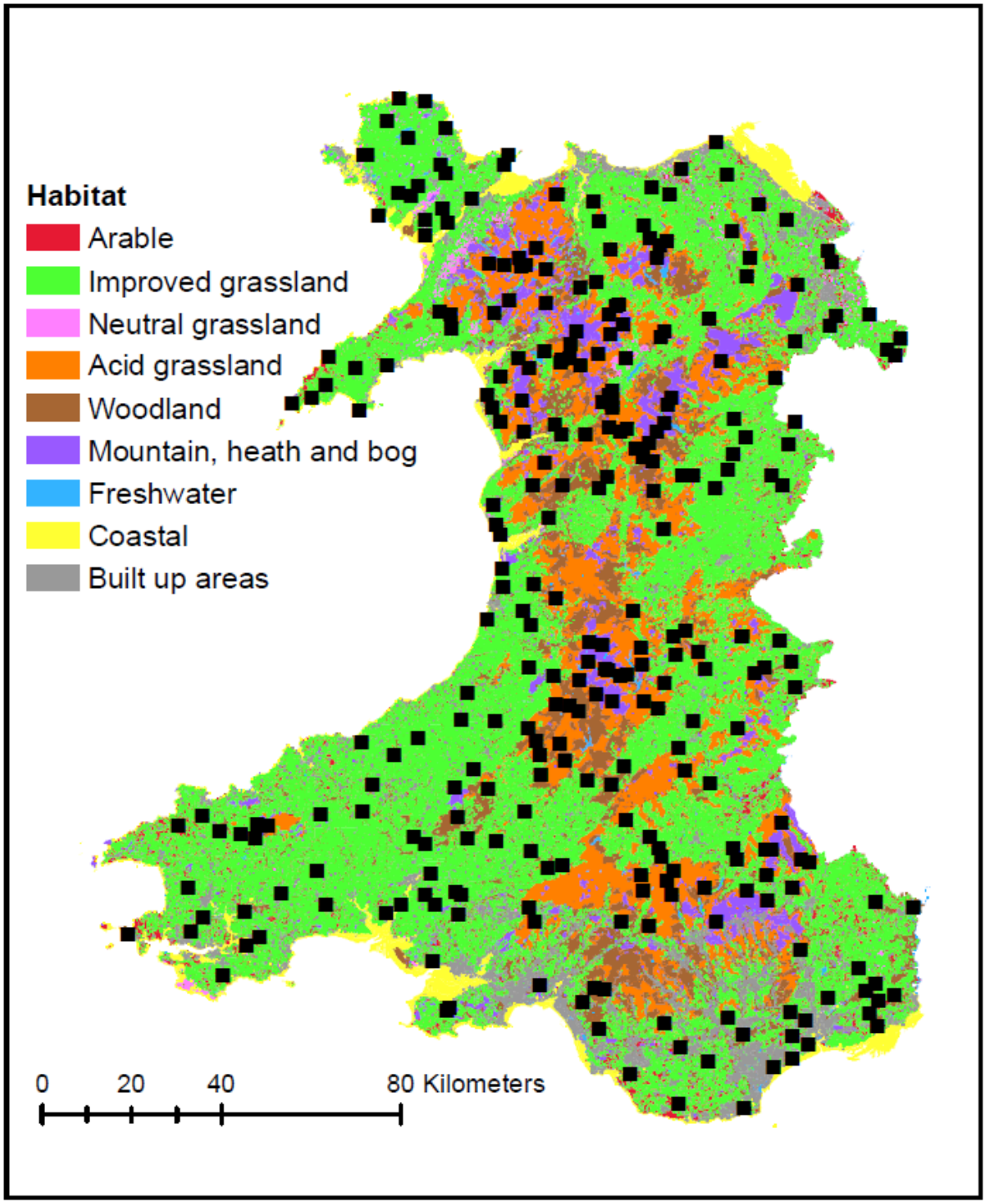


Figure 1: A map of the survey square locations and the range of habitats included in the survey. The white circles represent approximate survey square locations. The habitats shown are aggregated from the categories within the Land Cover Map 2015. These aggregated habitat classes were not obtained using the same methods as the field survey assignment so care must be taken in linking the results.

Elevation data was taken from NEXTmap based on the GPS coordinates of the plots. Precipitation is the Standardised Annual Average Rainfall for 1961-1990 calculated on a 1 km grid. Drought is a measure of the annual average number of dry spell events, defined as 14 day events with less than 2 mm rainfall per day, over the previous 30 years to sample collection and calculated on 5 km grid square basis. All precipitation and drought data came from the Met Office © Crown copyright 2017. The Land Cover Map 2015 was used to represent the range of habitats across Wales (Rowland et al., 2017).

2.2 Soil physicochemical laboratory analyses

Analysis of soil variables was undertaken using the methods of the Countryside Survey (Emmett et al., 2008). Soil pH was measured by suspending 10 g of fresh soil in 0.01 M $CaCl_2$ in a 1:2.5 (weight/volume) soil suspension (Avery & Bascomb, 1974). The pH used was measured in $CaCl_2$ instead of deionised water as the $CaCl_2$ solution has similar ionic strength to the soil solution in fertilised temperate soils and thus the pH is more representative of field conditions (Schofield & Taylor, 1955).

The surface 2 cm of the air-dry core was removed intact for water repellency measurement using the water drop penetration time method on the soil surface (Stefan H. Doerr, 1998) in the laboratory between 50-60% relative humidity. Six 1 ml droplets of deionised water were dropped on top of the soil surface from a height of 1 cm using a pipette. The absorption of the water droplets was recorded using video recording equipment, enabling measurement of the WDPT at a precision of 1s. This surface section of the soil was recombined with the rest of the core for further processing. The complete soil samples had particles greater than 2 mm size removed and the remaining fine earth

fraction ground by a deagglomerator (Pulverisette 8). Soil carbon of the fine earth fraction of the soil was measured by oxidative combustion followed by thermal conductivity detection using an Elementar Vario EL analyser. The soil water content was calculated as the volumetric percentage of the fine earth fraction of the soil, taking into account the volume of particles >2 mm removed.

2.3 Biological community data

2.3.1 Plant community analysis

Multiple indices of plant community properties were calculated, including both those based on Ellenberg indicator values (Hill, Preston, & Roy, 2004) and those based on Grime's CSR theory. Grime's CSR theory states that species can be categorised into competitors, stress tolerators and ruderals (Grime, 1977; Hodgson, Grime, Hunt, & Thompson, 1995). For these indices the score assigned to each plant species was taken and then a mean score per plot calculated based on species identity. Within this analysis we used Ellenberg fertility and Grime's stress tolerance.

2.3.2 Microbial community analysis

DNA was extracted using a mechanical lysis and homogenisation in triplicate from 0.25 g of soil per sample using PowerLyzer PowerSoil DNA Isolation Kits (MO-BIO) after pre-treatment with 750 μl of 1 M $CaCO_3$ (Sagova-Mareckova et al., 2008). Amplicon libraries were created using primers for the 16S (bacteria) and ITS1 (fungi) regions of the rRNA marker gene using a two-round PCR. The primer combinations used for the first round were 515F/806R (V4 16S) for 16S libraries (Caporaso et al., 2011; Walters et al., 2011) and ITS5/5.8S_fungi (ITS1) for ITS1 libraries (Epp et al., 2012). For a full description of the methods used see George et al. (2019). Amplicon libraries of

2013 samples were constructed at Bangor University. Library preparation for 2014 samples and Illumina sequencing for both years were conducted at the Liverpool Centre for Genome Research. Sequences with limited sample metadata have been uploaded to The European Nucleotide Archive with the following primary accession codes: PRJEB27883 (16S) and PRJEB28028 (ITS1).

All bioinformatics were performed on the Supercomputing Wales system. Illumina adapters were trimmed from sequences using Cutadapt (Martin, 2011). The sequences were then de-multiplexed, filtered, quality-checked, and clustered using a combination of USEARCH v. 7.0 (Edgar, 2010) and VSEARCH v. 2.3.2 (Rognes, Flouri, Nichols, Quince, & Mahé, 2016) programmes. Sequences with a maximum error greater than 1 and > 200 basepairs were removed following the merging of forward and reverse reads for all sequences. Operational taxonomic units (OTUs) were clustered using open reference methodology as described in George et al., (2019). Filtered sequences were matched first against either the GreenGenes v. 13_8 (DeSantis et al., 2006) or UNITE v. 7.2 (Kõljalg et al., 2013) databases,. Ten per cent of sequences that failed to match were clustered *de novo* and used as a new reference database for failed sequences. Sequences that failed to match with the *de novo* database were subsequently clustered *de novo*. Chimeric sequences were removed. Taxonomy was assigned to OTUs using QIIME (Caporaso et al., 2010) with RDP methodology (Q. Wang, Garrity, Tiedje, & Cole, 2007) from the GreenGenes database v. 13_8 and UNITE database v. 7.2 for the 16S and ITS1 data, respectively. Singletons and OTUs appearing in only 1 sample were removed from OTU tables following taxonomic assignment. All non-bacterial and non-fungal OTUs were removed from each OTU table.

To account for variation in read depth across samples, fungal data was rarefied to 1750 reads and bacterial data was rarefied to 18800 reads using the vegan package (Oksanen et al., 2018; Weiss et al., 2017). Rarefaction was repeated 100 times for fungi and 50 times for bacteria and the rounded mean used for all analyses. Fungal OTUs were also assigned to trophic mode using FUNGuild (Nguyen et al., 2016). In total 53.2% of the OTUs were assigned to a trophic mode, 82.9% of those assignations being rated probable or highly probable. The FUNGuild data was rarefied to 1500 read depth 100 times and the mean value across the repetitions used to calculate the proportions of OTUs identified to be solely pathotrophic, symbiotrophic or saprotrophic. Due to the low proportion of solely pathotrophic fungi within our samples only the symbiotrophic and saprotrophic proportions were used in the statistical analysis.

2.4 Statistical analysis

All statistical analysis was undertaken in R (R Core Team, 2018), and were performed on the natural logarithm of the median WDPT. The WDPT was categorised into the WDPT ratings of Doerr et al. (2006). Fig. 2 was created using the ggplot2 package (Wickham, 2009). Non-metric multidimensional scaling of the OTUs was performed using the vegan package (Oksanen et al., 2018) using Sørensen community composition distances.

Structural equation modelling (SEM) was used to evaluate the factors influencing water repellency in our dataset. This approach involves proposing a causative model, taking into account direct and indirect pathways, then fitting to the data and critically evaluating the proposed causative model. A set of climate, soil and plant variables were selected based on previous work constructing hypothesised relationships consistent with

mechanisms that could drive repellency. These variables were built into a piecewise structural equation model (SEM) (Shipley, 2000) using Bayesian multilevel models (Bürkner, 2017; Clough, 2012), and evaluated using Shipley's test of d-separation (Shipley, 2009, 2013). Further details on the SEM approach and parameter selection are contained within the supplementary information.

3. Results

3.1 Soil water repellency at the national scale

Overall, we found that 92% of the soils showed at least slight water repellency with 32% showing severe to extreme water repellency (Table S1). We found that water repellency was strongly associated with soil carbon, water content and the composition of the plant and soil microbial communities at a site (Fig. 2). Soil carbon had the largest impact upon water repellency in both the model across the full dataset (Fig. 3b, Table S2) and the model with microbial data (Fig. 2b, Table S3).

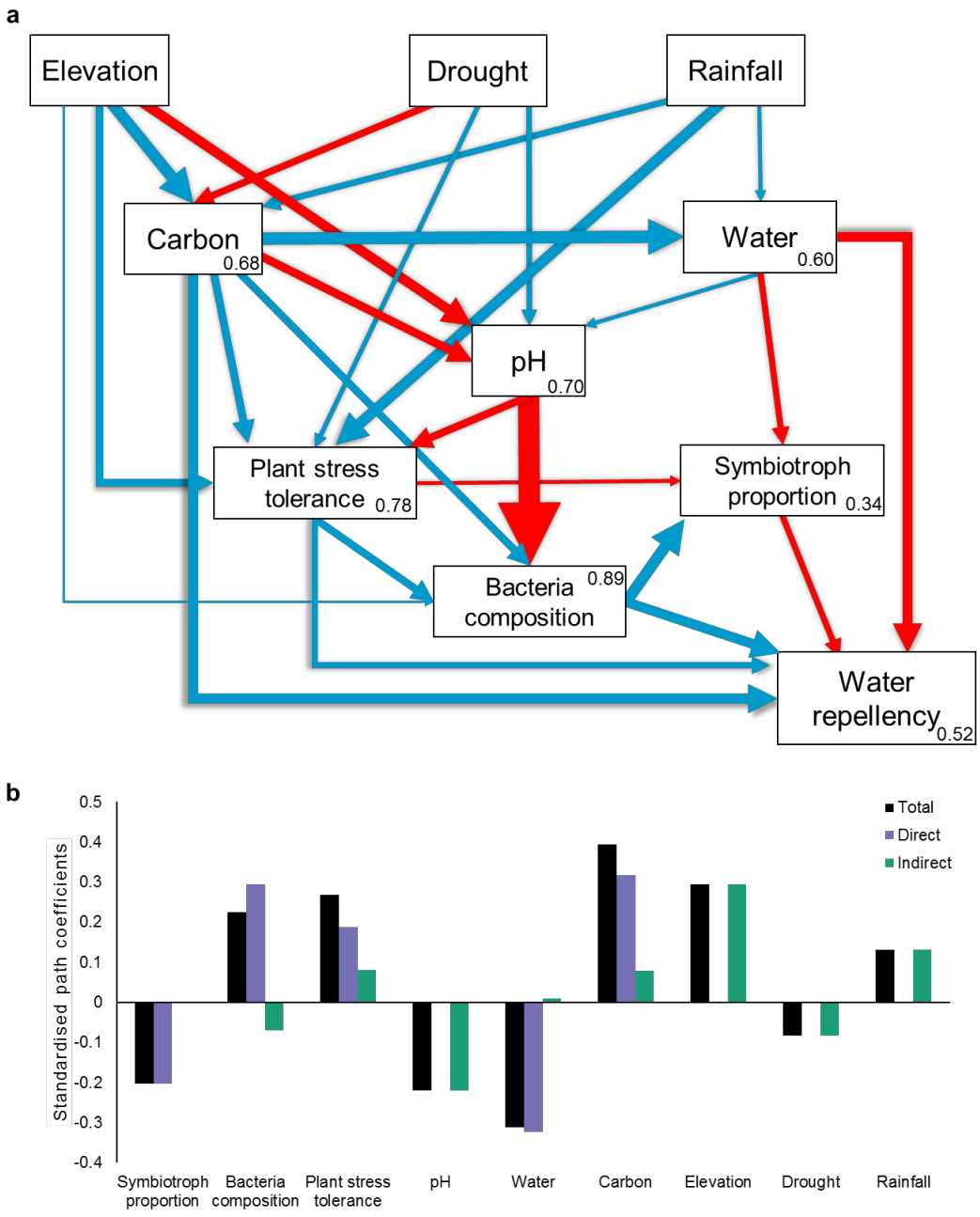


Figure 2: Structural equation modelling reveals soil water repellency is strongly influenced by biological community composition. a The width of the arrow joining two boxes is proportional to the strength of the relationship, i.e. the parameter estimate. Positive relationships are represented by a blue arrow, negative by red and endogenous variables feature the proportion of variance within the variable explained by the model, the conditional R^2 value, in the corner of the box. The model fitted the data well (C = 20.22, p = 0.68, n = 425) and all other SEMs tested had a ΔAICc score > 2. The full output from the model is in Supplementary Table 2. b The total, direct and indirect effect of each predictor on soil water repellency as estimated from the model parameters.

3.2 Biological influences on water repellency

Plant stress tolerance strongly impacted water repellency, having a direct impact that was over 50% higher than the effects of soil pH, soil water or climatic variables across the entire dataset (Fig. 3, Table S2). Although precipitation and drought were negatively correlated, both significantly increased the Grime stress tolerance score of a site. The stress score as a representative of the plant community was responsive to multiple forms of climatic stress as well as pH stress. A stress tolerant plant community at a site was associated with more repellent soils. The stress tolerance of the plant community impacts repellency directly and indirectly through differences in the soil microbial communities.

Both bacterial and fungal community composition explained significant residual variance in soil water repellency once changes in soil carbon, pH and water content were accounted for ($p < 0.001$), indicating a direct link between the soil microbial communities and water repellency. Soil water repellency decreased with increasing proportions of symbiotrophic fungi (Fig. 2), the majority of which were ectomycorrhizal in this dataset (61%). Bacterial composition had a particularly high direct impact upon repellency (93% of the impact of soil carbon, the source of hydrophobic material; Fig. 2b, Table S3).

3.3 Mediation of climate and pH stress

Within our model the impacts of environmental stressors on repellency were completely mediated by changes in the biological communities at a site. Within the model without microbial data there are direct links between precipitation, drought and repellency (Fig. 3) however these were not present in the model with microbial data (Fig.

2). Water repellency does increase considerably with elevation, and alters with changing rainfall regime, yet this was entirely mediated by changes in soil properties and the biological community (Fig. 2b). We also found no further association between soil pH and water repellency once changes in the soil bacterial community composition were accounted for.

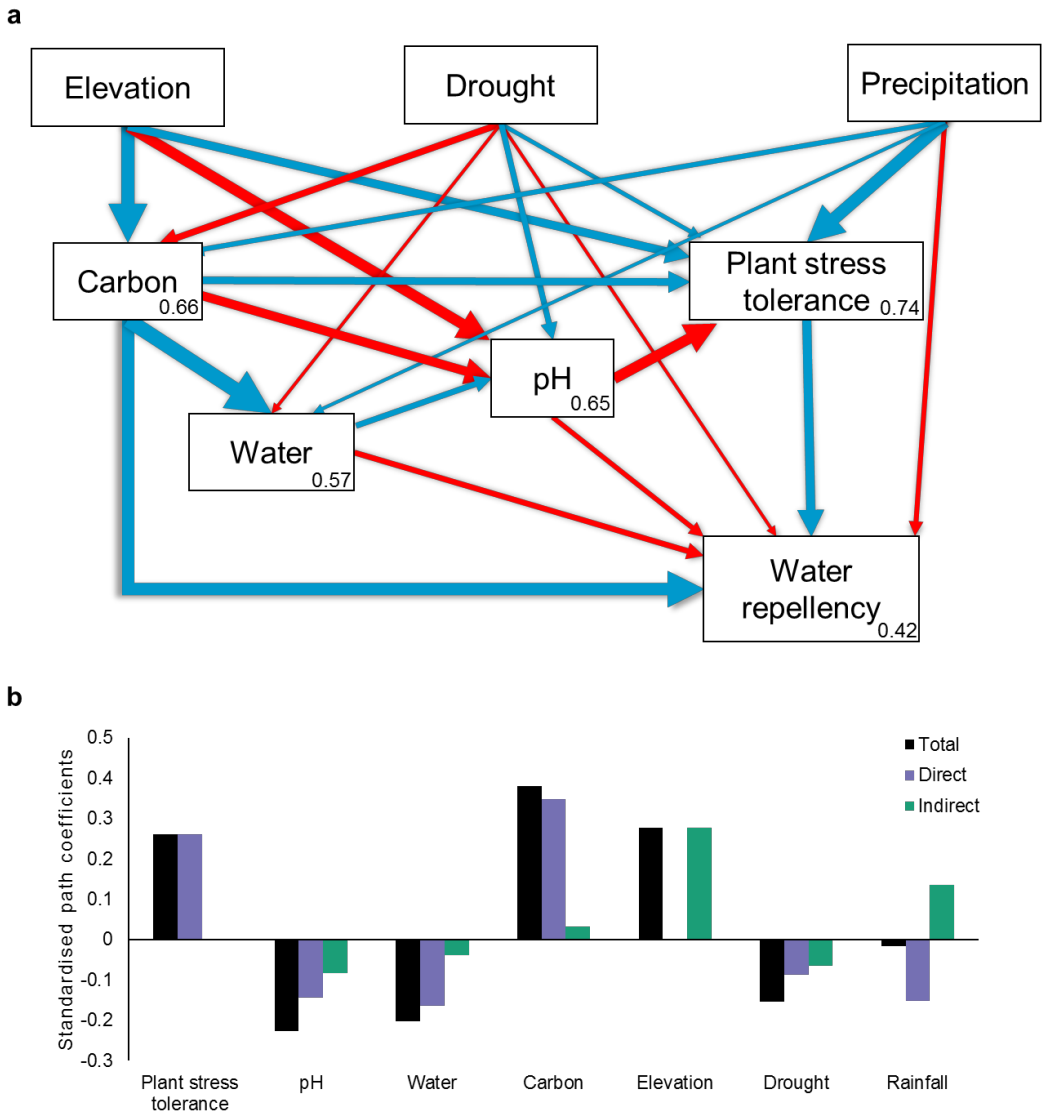


Figure 3: Structural equation modelling reveals the drivers of soil water repellency across the entire dataset. a The width of the arrow joining two boxes is proportional to the strength of the relationship, i.e. the parameter estimate. Positive relationships are represented by a blue arrow, negative by red and endogenous variables feature the proportion of variance within the variable explained by the model, the conditional R^2 value, in the corner of the box. The model fitted the data well (C=8.40, p=0.40, n=1326), and all other SEMs tested had ΔAICc > 2. The full output from the model is in Supplementary Table 3. b The total, direct and indirect effect of each predictor on soil water repellency is depicted as estimated from the model parameters.

3.4 Influence of land use on soil water repellency

Repellency varied across the different habitat types in our study, with higher repellency in low productivity habitats such as acid grassland and bog compared to high productivity habitats such as improved grassland. Repellency was highly variable within most habitat types, particularly in broadleaved woodlands and fens (Fig. 4). Arable systems had significantly lower water repellency than all other habitat types (Fig. 4, Table S1). The low water repellency of arable systems persisted after accounting for their higher pH and lower soil carbon content (ANOVA on impact of habitat on residuals for whole dataset: $F_{9,1295} = 7.394$; $p < 0.0001$; Table S4) and different microbial communities (ANOVA on impact of habitat on residuals: $F_{2,380}=2.458$; $p = 0.01$; Table S5). Arable habitats were the only habitats that were still different from other habitats after accounting for soil physicochemical and biotic variables (Table S5).

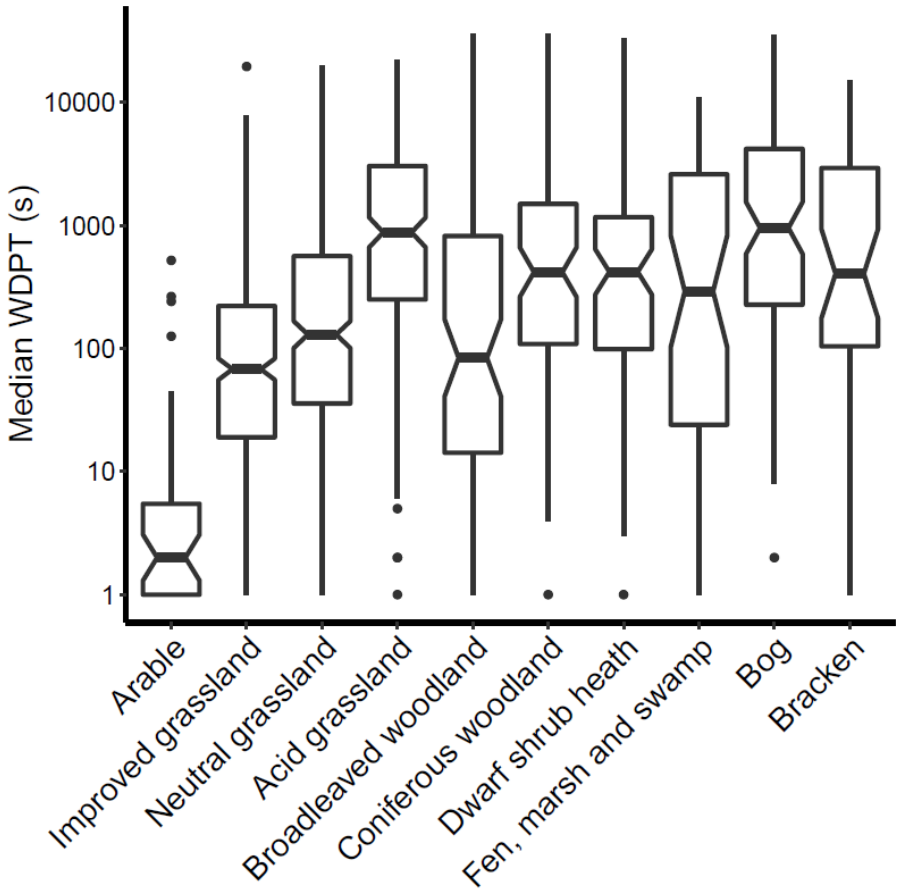


Figure 4.4: Arable systems show lower water repellency than all other habitat types. Water repellency increases with decreasing fertility of grassland (improved to neutral to acid grassland). The non-overlap of notches indicates that their medians are approximately significantly different at a 95% confidence level. Other habitats had lower sample sizes, overlapping notches and it is more difficult to draw strong conclusions.

4. Discussion

4.1 Biological influences on soil water repellency

We found that repellency is higher in ecosystems with greater soil carbon, higher plant stress tolerance and associated changes in soil pH and microbial communities (Fig. 5). The strong influence of soil carbon on water repellency is consistent with previous work (Hermansen et al., 2019; Mao, Nierop, Dekker, Dekker, & Chen, 2019; J.-T. Wang et al., 2016), but the association between plant community stress tolerance, microbial composition and repellency is novel. Our results provide evidence supporting literature conjecture that the ability to induce water repellency could confer a competitive

advantage to plants within stressful environments (Robinson et al., 2010; Verboom & Pate, 2006). Multiple types of environmental stressors, including both climatic and acidity related stressors, have been found to be related to repellency. Surface water repellency can divert water deeper into the soil profile through inducing preferential flow of water and preventing water movement upwards by creating a evaporative barrier layer at the soil surface providing dual protection from evaporation (S. H. Doerr et al., 2006; Rye & Smettem, 2017). In semi-arid ecosystems the pattern of soil moisture in relation to trees suggests that the trees respond to drought by inducing water repellency to promote water flux down their root systems into deeper soil layers (Robinson et al., 2010; Verboom & Pate, 2006). Rhizosphere hydrophobicity has been found in modelling exercises to give a competitive advantage for plant growth due to greater acquisition of water and mitigating the impacts of drought stress (Kroener, Zarebanadkouki, Bittelli, & Carminati, 2016; Zeppenfeld et al., 2017).

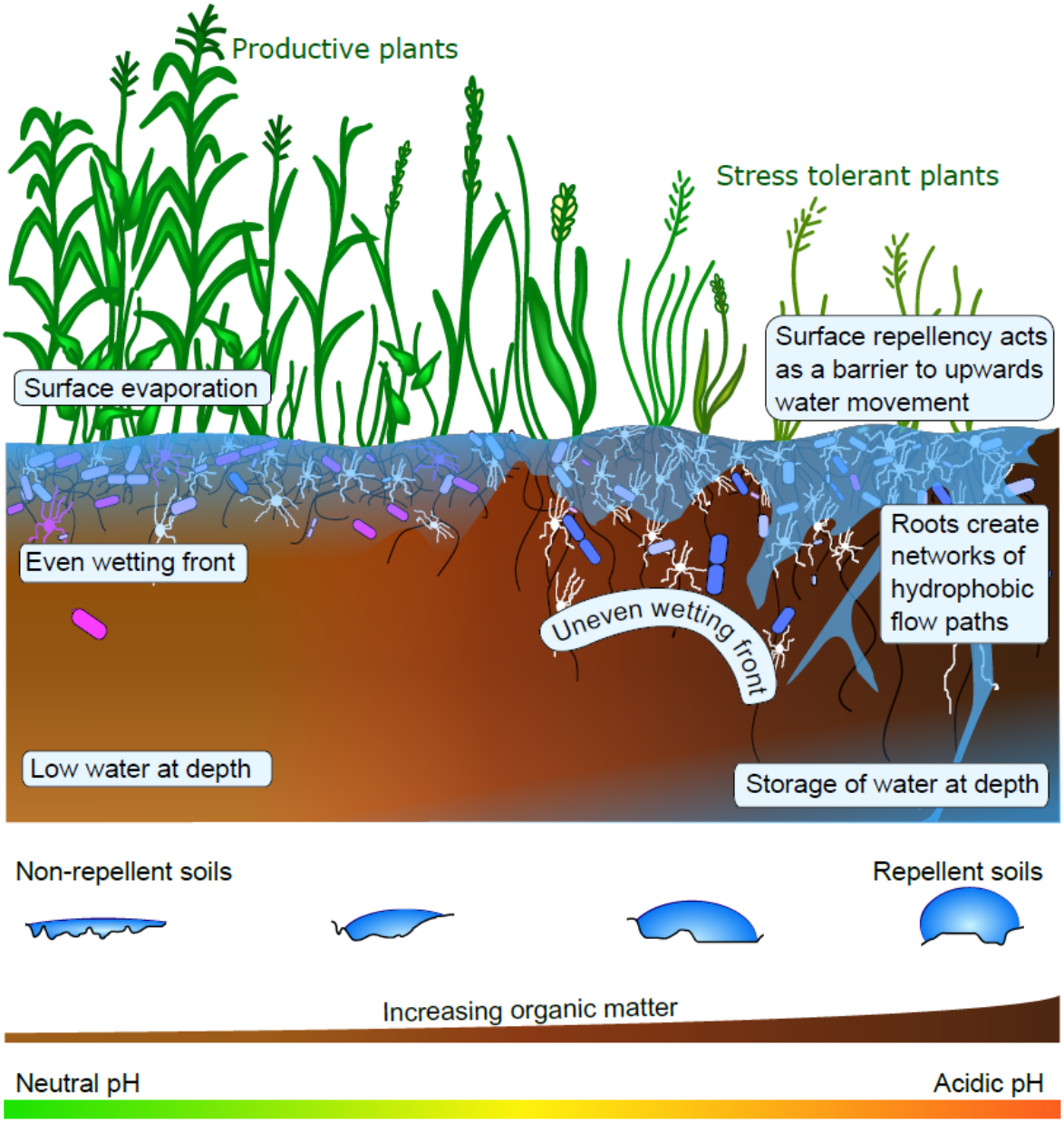


Figure 5: A representation of the change in repellency across an environmental stress gradient and its impact upon water fluxes in the soil when dry. Upon the left of the diagram we have a plant community that is adapted to be competitive in low-stress environments, highly productive with a non-repellent soil. Water infiltrates the soil in a piston flow manner. On the right we have a stress tolerant plant community with a repellent soil that alters water infiltration to follow preferential flow paths. This results in greater water next to plant roots and stored at depth within the soil.

We know from different parts of the literature that plant exudates (Svenningsson, Sundin, & Liljenberg, 1990), fungal mats (Spohn & Rillig, 2012), and bacterial communities (Achtenhagen, Goebel, Miltner, Woche, & Kästner, 2015) can all respond

to stress by producing water repellent compounds. For the microbial community the production of water repellent compounds can be an important survival mechanism both in dry and saturated systems. For example, Unestam (1991) argued that the lipoid, hydrophobic fungal surface protected both the fungus and tree roots against desiccation during drought periods. Furthermore, he observed that the mycorrhizal roots withstood a drier soil environment in rhizoscopes than did the hydrophilic non-mycorrhizal roots. Another advantage is that hydrophobic mycorrhizal hyphae may translocate water more efficiently, being less susceptible to water loss (Duddridge, Malibari, & Read, 1980; Read, Francis, & Finlay, 1985). In saturated conditions, Unestam (1991) argued that the fungal mats, particularly the complex hydrophobic structures, such as the mantle, cords, and patches, could produce air pockets. As obligate aerobes, saturation for extended periods would cause death, so the air pockets could provide a lifeline.

Bacteria have been found to produce extremely water repellent biofilms (Epstein, Pokroy, Seminara, & Aizenberg, 2011) (Epstein et al., 2011). One aspect of this repellency is that it prevents the penetration of antimicrobials into the biofilm. This has been exploited in crop protection where the biofilm development can shield roots from waterborne pathogens. Moreover, it has been argued that both hydrophobic bacterial cell walls and bacterial biofilms protect bacteria from desiccation or bursting in response to cycles of drying and rapid rewetting (Achtenhagen et al., 2015). Water stress was shown to activate a number of processes in microorganisms, (Morales, Parlange, & Steenhuis, 2010; Schimel, Balser, & Wallenstein, 2007). Hence our proposal that the development of water repellency is an ecosystem response to a stressful environment, as a means of protection for microbes and better resource allocation with plants. Our results, covering

climatic stress, soil physicochemical properties, plant and soil microbial communities together, support the development of such an ecological theory.

4.2 Persistence of repellency

Microbial communities are quicker to respond to change than plants and our results indicate that repellency could be induced by microbes on short timescales in response to environmental stressors. There is still much uncertainty over the persistence of repellency over time and space (Bodí et al., 2013; Leighton-Boyce, Doerr, Shakesby, & Walsh, 2007; K. Müller et al., 2014; Rye & Smettem, 2015). Our study analysed the air-dry repellency of the soil, which can be interpreted as the ability of the sample to become repellent upon drying and thus would be less variable over time than repellency of the fresh soil surface. The different ways in which repellency is created and maintained may be a critical factor in determining how long repellency will persist. Some studies have found that hydrophobicity can originate from plant material, both litter and root exudates, which clearly indicates a potential for long term maintenance of repellency by plants (Cesarano, Incerti, & Bonanomi, 2016; Hallett et al., 2009; Mao et al., 2016; Naveed et al., 2018). Microbial communities are more changeable than plants yet could still result in the long term ability to induce repellency. Microbes both create and destroy repellent compounds, and changes in the composition of the community help determine water repellency.

4.3 Evaluating the directionality of links and mediation in SEM

Within our analysis we assumed that soil repellency was caused by changes in the microbial community, rather than the reverse. We consider that repellency is caused by hydrophobic compounds within the soil (Hermansen et al., 2019; Mainwaring et al.,

2013; Mao et al., 2019), however, it is feasible that the physical configuration of soil components could play some role, which remains largely unexplored (Benard et al., 2018). It is these hydrophobic factors that we consider to be altered by biotic communities. It is possible that the hydrophobic compounds within the soil could be altering the microbial communities through changing the suitability of the environment (Barnard, Osborne, & Firestone, 2013; Or, Smets, Wraith, Dechesne, & Friedman, 2007; G. Wang & Or, 2013). However microbial communities are both the source of, and mediator of, the breakdown of hydrophobic compounds (Achtenhagen et al., 2015; Chau, Goh, Vujanovic, & Si, 2012; Li et al., 2018; Schaumann et al., 2007). There is likely a feedback mechanism whereby, as the physical environment is altered by the production or degradation of hydrophobic compounds, this then forces changes in microbial communities which are adapted to different situations. We believe that the shorter feedback is in the direction of microbes to repellency, and it is this we have included in our model.

We have found complete mediation of climatic and some physicochemical stressors on repellency. Thus once we know the biotic community composition we do not need to know the wider environmental conditions to be able to predict repellency. In particular, the complete mediation of pH related influences on repellency by the microbial community is of interest. This suggests that the change in water repellency with pH found in many observational studies (Lebron, Robinson, Oatham, & Wuddivira, 2012; Mirbabaei, Shahrestani, Zolfaghari, & Abkenar, 2013; Zavala, García-moreno, Gordillo-rivero, Jordán, & Mataix-solera, 2014) is not likely to be due to chemical modification of particles, which has been found to alter water repellency in pH

modification experiments (Amer, Schaumann, & Diehl, 2017; Diehl, 2013). The complete mediation of climatic stressors upon repellency suggests that the influence of climate on soil surface water content will be strongly impacted by the biological community at a site, with implications for earth system modelling (Goebel et al., 2011; Green et al., 2019). The infiltration of water into the soil in these systems is driven by biological factors, not physicochemical, and will therefore change as biological communities are placed under increasing stress.

4.4.4 Influence of land use on soil water repellency

The differing land uses within our study had differing repellency, however the impact of land use on repellency was in most cases explained by the variation in carbon, pH and biotic communities across the land use types. This supports the findings of Doerr et al. (2006), who also found a land cover dependency for soil water repellency in the United Kingdom. Repellency is known to have a strong role in the function of some land use types. For example, within some peatland systems extreme water repellency was created after fire, which lowered evaporation, allowed the maintenance of a high water table, and increased speed of ecosystem recovery compared to systems that did not become repellent after fire (Kettridge et al., 2014). With regard to stress it has been found that, in pasture systems a negative relationship between productivity and repellency has been found (K. Müller et al., 2014). This suggests that the competitive advantage found by the aforementioned modelling studies (Kroener et al., 2016; Zeppenfeld et al., 2017) are limited to locations that are undergoing stress and are potentially therefore less productive. Our results are consistent with this as stress resilient plant species are found in less productive sites.

There is however one habitat in which knowing the carbon, water and biotic community does not mean that you can predict repellency: arable. Arable systems have lower than predicted repellency even after taking into account soil physicochemical, above and belowground community composition. There is something qualitatively different about arable systems which results in lower repellency, perhaps due to the mechanical disturbance of the soil through tillage, which has been found to reduce water repellency and infiltration (Karin Müller et al., 2016; Roper, Ward, Keulen, & Hill, 2013). Water repellency is likely to be related to soil biophysical structure, the networks of roots, fungal hyphae and microbial biofilms that permeate the soil and follow, create and maintain preferential flow paths for water infiltration.

4.5 Water repellency and biological community response to stress

The concept of water repellency as an adaptive stress response suggests that the ability to induce water repellency promotes ecosystem resilience to drought and other stressors. Access to water stores has been shown to be crucial in determining carbon loss and plant resilience during drought (De Boeck & Verbeeck, 2011). We propose that water repellency indicates a healthy ecosystem response to stress, and the inability of tilled land to induce water repellency can be interpreted as an unhealthy lack of resilience. We have found that multiple different natural stressors: drought; high precipitation and low nutrient status acidic soils had a consistent relationship with our realistic large-scale gradient of soil water repellency. It is the biological communities which are more closely related to soil repellency than physicochemical factors, showing the importance of ecology in modifying hydrological processes through feedbacks that will help conserve water. The homogeneity of response indicates there are consistent

mechanisms induced by biological communities across ecosystem types to increase resilience. These mechanisms are those we should be interested in monitoring and influencing to understand, predict and mitigate ecosystem shifts in response to increasing stress from land use and climate change.

5. References

Achtenhagen, J., Goebel, M. O., Miltner, A., Woche, S. K., & Kästner, M. (2015). Bacterial impact on the wetting properties of soil minerals. *Biogeochemistry*, *122*(2–3), 269–280. doi: 10.1007/s10533-014-0040-9

Amer, A., Schaumann, G. E., & Diehl, D. (2017). The effect of pH modification on wetting kinetics of a naturally water-repellent coniferous forest soil. *European Journal of Soil Science*, *68*(3), 317–326. doi: 10.1111/ejss.12419

Avery, B. W., & Bascomb, C. L. (1974). *Soil Survey Laboratory Methods. Soil Survey Technical Monograph No. 6, Harpenden*.

Barnard, R. L., Osborne, C. A., & Firestone, M. K. (2013). Responses of soil bacterial and fungal communities to extreme desiccation and rewetting. *The ISME Journal*, *7*(11), 2229–2241. doi: 10.1038/ismej.2013.104

Benard, P., Zarebanadkouki, M., Hedwig, C., Holz, M., Ahmed, M. A., & Carminati, A. (2018). Pore-scale distribution of mucilage affecting water repellency in the rhizosphere. *Vadose Zone Journal*, *17*, 170013. doi: 10.2136/vzj2017.01.0013

Bodí, M. B., Muñoz-Santa, I., Armero, C., Doerr, S. H., Mataix-Solera, J., & Cerdà, A. (2013). Spatial and temporal variations of water repellency and probability of its occurrence in calcareous Mediterranean rangeland soils affected by fires. *Catena*, *108*, 14–25. doi: 10.1016/j.catena.2012.04.002

Bogner, C., Gaul, D., Kolb, A., Schmiedinger, I., & Huwe, B. (2010). Investigating flow mechanisms in a forest soil by mixed-effects modelling. *European Journal of Soil Science*, *61*(6), 1079–1090. doi: 10.1111/j.1365-2389.2010.01300.x

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*(1), 1–28. doi: 10.18637/jss.v080.i01

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., … Knight, R. (2010). QIIME allows analysis of high-throughput community sequencing data. *Nature Methods*, *7*(5), 335–336. doi: 10.1038/nmeth.f.303

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., … Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences*, *108*(Suppl 1), 4516–4522. doi: 10.1073/pnas.1000080107

Cesarano, G., Incerti, G., & Bonanomi, G. (2016). The influence of plant litter on soil water repellency: insight from 13C NMR spectroscopy. *PloS One*, *11*(3), e0152565. doi: 10.1371/journal.pone.0152565

Chau, H. W., Goh, Y. K., Vujanovic, V., & Si, B. C. (2012). Wetting properties of fungi mycelium alter soil infiltration and soil water repellency in a γ-sterilized wettable and repellent soil. *Fungal Biology*, *116*(12), 1212–1218. doi: 10.1016/j.funbio.2012.10.004

Clothier, B. E., Green, S. R., & Deurer, M. (2008). Preferential flow and transport in soil: progress and prognosis. *European Journal of Soil Science*, *59*(1), 2–13. doi: 10.1111/j.1365-2389.2007.00991.x

Clough, Y. (2012). A generalized approach to modeling and estimating indirect effects in ecology. *Ecology*, *93*(8), 1809–1815. doi: 10.1890/11-1899.1

Costanza, R., D'Arge, R., de Groot, R., Farber, S., Grasso, M., Hannon, B., … van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature*, *387*(6630), 253–260. doi: 10.1038/387253a0

Côté, I. M., Darling, E. S., & Brown, C. J. (2016). Interactions among ecosystem stressors and their importance in conservation. *Proceedings. Biological Sciences*, *283*(1824). doi: 10.1098/rspb.2015.2592

De Boeck, H. J., & Verbeeck, H. (2011). Drought-associated changes in climate and their relevance for ecosystem experiments and models. *Biogeosciences*, *8*(5), 1121–1130. doi: 10.5194/bg-8-1121-2011

Dekker, L. W., & Ritsema, C. J. (1994). How water moves in a water repellent sandy soil: 1. Potential and actual water repellency. *Water Resources Research*, *30*(9), 2507–2517. doi: 10.1029/94WR00749

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

Diehl, D. (2013). Soil water repellency: dynamics of heterogeneous surfaces. *Colloids and Surfaces A: Physicochemical and Engineering Aspects*, *432*, 8–18. doi: 10.1016/j.colsurfa.2013.05.011

Doerr, S. H., Shakesby, R. a., Dekker, L. W., & Ritsema, C. J. (2006). Occurrence, prediction and hydrological effects of water repellency amongst major soil and land-use types in a humid temperate climate. *European Journal of Soil Science*, *57*(5), 741–754. doi: 10.1111/j.1365-2389.2006.00818.x

Doerr, S.H., Shakesby, R. A., & Walsh, R. P. D. (2000). Soil water repellency: its causes, characteristics and hydro-geomorphological significance. *Earth-Science Reviews*, *51*(1–4), 33–65. doi: 10.1016/S0012-8252(00)00011-8

Doerr, Stefan H. (1998). On standardizing the 'water drop penetration time' and the 'molarity of an ethanol droplet' techniques to classify soil hydrophobicity: a case study using medium textured soils. *Earth Surface Processes and Landforms*, *23*(7), 663–668. doi: 10.1002/(SICI)1096-9837(199807)23:7<663::AID-ESP909>3.0.CO;2-6

Duddridge, J. A., Malibari, A., & Read, D. J. (1980). Structure and function of mycorrhizal rhizomorphs with special reference to their role in water transport. *Nature*, *287*(5785), 834–836. doi: 10.1038/287834a0

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

Emmett, B. A., Frogbrook, Z. L., Chamberlain, P. M., Griffiths, R., Pickup, R., Poskitt, J., … Wood, C. M. (2008). *Countryside Survey technical report no. 03/07*.

Emmett, B. A., & the GMEP team. (2017). *Glastir Monitoring & Evaluation Programme. Final report to Welsh Government*.

Epp, L. S., Boessenkool, S., Bellemain, E. P., Haile, J., Esposito, A., Riaz, T., … Brochmann, C. (2012). New environmental metabarcodes for analysing soil DNA: Potential for studying past and present ecosystems. *Molecular Ecology*, *21*(8), 1821–1833. doi: 10.1111/j.1365-294X.2012.05537.x

Epstein, A. K., Pokroy, B., Seminara, A., & Aizenberg, J. (2011). Bacterial biofilm shows persistent resistance to liquid wetting and gas penetration. *Proceedings of the National Academy of Sciences*, *108*(3), 995–1000. doi: 10.1073/pnas.1011033108

George, P. B. L., Lallias, D., Creer, S., Seaton, F. M., Kenny, J. G., Eccles, R. M., … Jones, D. L. (2019). Divergent national-scale trends of microbial and animal biodiversity revealed across diverse temperate soil ecosystems. *Nature Communications*, *10*(1), 1107. doi: 10.1038/s41467-019-09031-1

Goebel, M., Bachmann, J., Reichstein, M., Janssens, I. A., & Guggenberger, G. (2011). Soil water repellency and its implications for organic matter decomposition - is there a link to extreme climatic events? *Global Change Biology*, *17*(8), 2640–2656. doi: 10.1111/j.1365-2486.2011.02414.x

Green, J. K., Seneviratne, S. I., Berg, A. M., Findell, K. L., Hagemann, S., Lawrence, D. M., & Gentine, P. (2019). Large influence of soil moisture on long-term terrestrial carbon uptake. *Nature*, *565*(7740), 476–479. doi: 10.1038/s41586-018-0848-x

Grime, J. P. (1977). Evidence for the existence of three primary strategies in plants and its relevance to ecological and evolutionary theory. *The American Naturalist*, *111*(982), 1169–1194.

Hallett, P. D., Feeney, D. S., Bengough, A. G., Rillig, M. C., Scrimgeour, C. M., & Young, I. M. (2009). Disentangling the impact of AM fungi versus roots on soil structure and water transport. *Plant and Soil*. doi: 10.1007/s11104-008-9717-y

Hermansen, C., Moldrup, P., Müller, K., Jensen, P. W., van den Dijssel, C., Jeyakumar, P., & de Jonge, L. W. (2019). Organic carbon content controls the severity of water repellency and the critical moisture level across New Zealand pasture soils. *Geoderma*, *338*, 281–290. doi: 10.1016/J.GEODERMA.2018.12.007

Hill, M. O., Preston, C. D., & Roy, D. B. (2004). *PLANTATT - attributes of British and Irish plants: status, size, life history, geography and habitats*. Centre for Ecology & Hydrology.

Hodgson, J. G., Grime, J. P., Hunt, R., & Thompson, K. (1995). *The Electronic Comparative Plant Ecology*. Springer.

IPCC. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, R.K. Pachauri and L.A. Meyer (eds.)]*.

Jackson, D. L. (2000). *Guidance on the interpretation of the Biodiversity Broad Habitat Classification (terrestrial and freshwater types): Definitions and the relationships with other classifications, JNCC Report 307*.

Kettridge, N., Humphrey, R. E., Smith, J. E., Lukenbach, M. C., Devito, K. J., Petrone, R. M., & Waddington, J. M. (2014). Burned and unburned peat water repellency: implications for peatland evaporation following wildfire. *Journal of Hydrology*, *513*, 335–341. doi: 10.1016/j.jhydrol.2014.03.019

Kõljalg, U., Nilsson, R. H., Abarenkov, K., Tedersoo, L., Taylor, A. F. S., Bahram, M., … Larsson, K.-H. (2013). Towards a unified paradigm for sequence-based identification of fungi. *Molecular Ecology*, *22*(21), 5271–5277. doi: 10.1111/mec.12481

Kroener, E., Zarebanadkouki, M., Bittelli, M., & Carminati, A. (2016). Simulation of root water uptake under consideration of nonequilibrium dynamics in the rhizosphere. *Water Resources Research*, *52*(8), 5755–5770. doi: 10.1002/2015WR018579

Lebron, I., Robinson, D. A., Oatham, M., & Wuddivira, M. N. (2012). Soil water repellency and pH soil change under tropical pine plantations compared with native tropical forest. *Journal of Hydrology*, *414–415*, 194–200. doi: 10.1016/j.jhydrol.2011.10.031

Leighton-Boyce, G., Doerr, S. H., Shakesby, R. A., & Walsh, R. P. D. (2007). Quantifying the impact of soil water repellency on overland flow generation and erosion: a new approach using rainfall simulation and wetting agent on in situ soil G. *Hydrological Processes*, *21*, 2337–2345. doi: 10.1002/hyp.6744

Li, B., Pales, A. R., Clifford, H. M., Kupis, S., Hennessy, S., Liang, W.-Z., … Darnault, C. J. G. (2018). Preferential flow in the vadose zone and interface dynamics: impact of microbial exudates. *Journal of Hydrology*, *558*, 72–89. doi: 10.1016/J.JHYDROL.2017.12.065

Mainwaring, K., Hallin, I. L., Douglas, P., Doerr, S. H., & Morley, C. P. (2013). The role of naturally occurring organic compounds in causing soil water repellency. *European Journal of Soil Science*, *64*(5). doi: 10.1111/ejss.12078

Mao, J., Nierop, K. G. J., Dekker, S. C., Dekker, L. W., & Chen, B. (2019). Understanding the mechanisms of soil water repellency from nanoscale to ecosystem scale: a review. *Journal of Soils and Sediments*, *19*(1), 171–185. doi: 10.1007/s11368-018-2195-9

Mao, J., Nierop, K. G. J., Rietkerk, M., Sinninghe Damsté, J. S., & Dekker, S. C. (2016). The influence of vegetation on soil water repellency-markers and soil hydrophobicity. *Science of The Total Environment*, *566*, 608–620. doi: 10.1016/j.scitotenv.2016.05.077

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10. doi: 10.14806/ej.17.1.200

Mirbabaei, S. M., Shahrestani, M. S., Zolfaghari, A., & Abkenar, K. T. (2013). Relationship between soil water repellency and some of soil properties in northern Iran. *Catena*, *108*, 26–34. doi: 10.1016/j.catena.2013.02.013

Morales, V. L., Parlange, J. Y., & Steenhuis, T. S. (2010). Are preferential flow paths perpetuated by microbial activity in the soil matrix? A review. *Journal of Hydrology*, *393*(1–2), 29–36. doi: 10.1016/j.jhydrol.2009.12.048

Müller, K., Deurer, M., Jeyakumar, P., Mason, K., van den Dijssel, C., Green, S., & Clothier, B. (2014). Temporal dynamics of soil water repellency and its impact on pasture productivity. *Agricultural Water Management*, *143*, 82–92. doi: 10.1016/j.agwat.2014.06.013

Müller, Karin, Carrick, S., Meenken, E., Clemens, G., Hunter, D., Rhodes, P., & Thomas, S. (2016). Is subcritical water repellency an issue for efficient irrigation in arable soils? *Soil and Tillage Research*, *161*, 53–62. doi: 10.1016/j.still.2016.03.010

Naveed, M., Brown, L. K., Raffan, A. C., George, T. S., Bengough, A. G., Roose, T., … Hallett, P. D. (2018). Rhizosphere-scale quantification of hydraulic and mechanical properties of soil impacted by root and seed exudates. *Vadose Zone Journal*, *17*(1), 0. doi: 10.2136/vzj2017.04.0083

Nguyen, N. H., Song, Z., Bates, S. T., Branco, S., Tedersoo, L., Menke, J., … Kennedy, P. G. (2016). FUNGuild: an open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology*, *20*, 241–248. doi: 10.1016/j.funeco.2015.06.006

Oksanen, J., Blanchet, G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., … Wagner, H. (2018). *vegan: Community Ecology Package*.

Or, D., Smets, B. F., Wraith, J. M., Dechesne, A., & Friedman, S. P. (2007). Physical constraints affecting bacterial habitats and activity in unsaturated porous media - a review. *Advances in Water Resources*, *30*(6–7), 1505–1527. doi: 10.1016/j.advwatres.2006.05.025

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Read, D. J., Francis, R., & Finlay, R. D. (1985). Mycorrhizal mycelia and nutrient cycling in plant communities. In A. Fitter, D. Atkinson, D. J. Read, & M. Usher (Eds.), *Ecological interactions in soil* (pp. 193–218). Oxford, UK: Blackwell.

Robinson, D. A., Hopmans, J. W., Filipovic, V., van der Ploeg, M., Lebron, I., Jones, S. B., … Tuller, M. (2019). Global environmental changes impact soil hydraulic functions through biophysical feedbacks. *Global Change Biology*, *25*(6), 1895–1904. doi: 10.1111/gcb.14626

Robinson, D. A., Lebron, I., Ryel, R. J., & Jones, S. B. (2010). Soil water repellency: a method of soil moisture sequestration in pinyon–juniper woodland. *Soil Science Society of America Journal*, *74*(2), 624. doi: 10.2136/sssaj2009.0208

Rognes, T., Flouri, T., Nichols, B., Quince, C., & Mahé, F. (2016). VSEARCH: a versatile open source tool for metagenomics. *PeerJ*, *4*, e2584. doi: 10.7717/peerj.2584

Roper, M. M., Ward, P. R., Keulen, A. F., & Hill, J. R. (2013). Under no-tillage and stubble retention, soil water content and crop growth are poorly related to soil water repellency. *Soil and Tillage Research*, *126*, 143–150. doi: 10.1016/j.still.2012.09.006

Rowland, C. S., Morton, R. D., Carrasco, L., McShane, G., O'Neil, A. W., & Wood, C. M. (2017). *Land Cover Map 2015 (1km dominant aggregate class, GB)*. doi: 10.5285/711c8dc1-0f4e-42ad-a703-8b5d19c92247

Rye, C. F., & Smettem, K. R. J. (2015). Seasonal and interannual variability of the effective flow cross-sectional area in a water-repellent soil. *Vadose Zone Journal*, *14*(3). doi: 10.2136/vzj2014.10.0141

Rye, C. F., & Smettem, K. R. J. (2017). The effect of water repellent soil surface layers on preferential flow and bare soil evaporation. *Geoderma*, *289*, 142–149. doi: 10.1016/j.geoderma.2016.11.032

Rye, C. F., & Smettem, K. R. J. (2018). Seasonal variation of subsurface flow pathway spread under a water repellent surface layer. *Geoderma*, *327*, 1–12. doi: 10.1016/J.GEODERMA.2018.04.008

Sagova-Mareckova, M., Cermak, L., Novotna, J., Plhackova, K., Forstova, J., & Kopecky, J. (2008). Innovative methods for soil DNA purification tested in soils with widely differing characteristics. *Applied and Environmental Microbiology*, *74*(9), 2902–2907. doi: 10.1128/AEM.02161-07

Schaumann, G. E., Braun, B., Kirchner, D., Rotard, W., Szewzyk, U., & Grohmann, E. (2007). Influence of biofilms on the water repellency of urban soil samples. *Hydrological Processes*, *21*(17), 2276–2284. doi: 10.1002/hyp.6746

Schimel, J., Balser, T. C., & Wallenstein, M. (2007). Microbial stress response physiology and its implications for ecosystem function. *Ecology*, *88*(6), 1386–1394. doi: 10.1890/06-0219

Schofield, R. K., & Taylor, A. W. (1955). The measurement of soil pH. *Soil Science Society of America Journal*, *19*(2), 164–167. doi: 10.2136/sssaj1955.03615995001900020013x

Shipley, B. (2000). A new inferential test for path models based on directed acyclic graphs. *Structural Equation Modeling*, *7*(2), 206–218. doi: 10.1207/S15328007SEM0702_4

Shipley, B. (2009). Confirmatory path analysis in a generalized multilevel context. *Ecology*, *90*(2), 363–368. doi: 10.1890/08-1034.1

Shipley, B. (2013). The AIC model selection method applied to path analytic models compared using a d-separation test. *Ecology*, *94*(3), 560–564. doi: 10.1890/12-0976.1

Spohn, M., & Rillig, M. C. (2012). Temperature- and moisture-dependent soil water repellency induced by the basidiomycete Agaricus bisporus. *Pedobiologia*, *55*(1), 59–61. doi: 10.1016/j.pedobi.2011.10.006

Svenningsson, H., Sundin, P., & Liljenberg, C. (1990). Lipids, carbohydrates and amino acids exuded from the axenic roots of rape seedlings exposed to water-deficit stress. *Plant, Cell and Environment*, *13*, 155–162. doi: 10.1111/j.1365-3040.1990.tb01287.x

Unestam, T. (1991). Water repellency, mat formation, and leaf-stimulated growth of some ectomycorrhizal fungi. *Mycorrhiza*, *1*(1), 13–20. doi: 10.1007/BF00205897

van der Putten, W. H., Bardgett, R. D., Bever, J. D., Bezemer, T. M., Casper, B. B., Fukami, T., … Wardle, D. A. (2013). Plant-soil feedbacks: the past, the present and future challenges. *Journal of Ecology*, *101*(2), 265–276. doi: 10.1111/1365-2745.12054

Verboom, W. H., & Pate, J. S. (2006). Bioengineering of soil profiles in semiarid ecosystems: the "phytotarium" concept. A review. In *Plant and Soil* (Vol. 289). doi: 10.1007/s11104-006-9073-8

Walters, W. A., Caporaso, J. G., Lauber, C. L., Berg-Lyons, D., Fierer, N., & Knight, R. (2011). PrimerProspector: de novo design and taxonomic analysis of barcoded polymerase chain reaction primers. *Bioinformatics (Oxford, England)*, *27*(8), 1159–1161. doi: 10.1093/bioinformatics/btr087

Wang, G., & Or, D. (2013). Hydration dynamics promote bacterial coexistence on rough surfaces. *The ISME Journal*, *7*(2), 395–404. doi: 10.1038/ismej.2012.115

Wang, J.-T., Zheng, Y.-M., Hu, H.-W., Li, J., Zhang, L.-M., Chen, B.-D., … He, J.-Z. (2016). Coupling of soil prokaryotic diversity and plant diversity across latitudinal forest ecosystems. *Scientific Reports*, *6*, 19561. doi: 10.1038/srep19561

Wang, Q., Garrity, G. M., Tiedje, J. M., & Cole, J. R. (2007). Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied and Environmental Microbiology*, *73*(16), 5261–5267. doi: 10.1128/AEM.00062-07

Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., … Knight, R. (2017). Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome*, *5*(1), 27. doi: 10.1186/s40168-017-0237-y

Wickham, H. (2009). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.

Zavala, L. M., García-moreno, J., Gordillo-rivero, Á. J., Jordán, A., & Mataix-solera, J. (2014). Natural soil water repellency in different types of Mediterranean woodlands. *Geoderma*, *226–227*, 170–178. doi: 10.1016/j.geoderma.2014.02.009

Zeppenfeld, T., Balkenhol, N., Kóvacs, K., Carminati, A., Bennett, M., & Green, S. (2017). Rhizosphere hydrophobicity: a positive trait in the competition for water. *Plos One*, *12*(7), e0182188. doi: 10.1371/journal.pone.0182188