SPATIAL SCALING OF SOIL MICROBES UNDER DIFFERENT LAND USES

SERENA KATHERINE THOMSON
BSc (Hons), MSc

A thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences, University of Warwick

March 2015

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank my supervisors Dr Gary D. Bending, Dr. Christopher van der Gast and Dr David Bass for their support and guidance throughout this journey. Their patience and nurturing has been instrumental in the completion of this work. I am also grateful to the Natural Environment Research Council (NERC) for providing funding for this thesis.

I would also like to acknowledge the help and support of colleagues at the University of Warwick, particularly within Gary Bending's laboratory group including Lawrence Davies, Christopher Barnes, Caitlin Burns, Paul Hunter, Mark Day. A special thanks goes to Philip James, Christina Hazard and Leah Cuthbertson for their help with R programming, statistical analysis and all things scripts! Thank you too to Sally Hilton and Paul Gosling for willingly sharing their hard-earned samples with me. I do not underestimate the time and effort that went in to collecting that data. I hope I have done it justice! Thank you to other members of Life Sciences who I have not mentioned, who's friendly faces have made my time at the University of Warwick an enjoyable experience.

Finally, I would like to thank my husband Christopher for not only helping me with programming and running scripts, but for putting up with the tears and tantrums! Thank you also to my parents for their guidance and support.

I dedicate this work to my late father. My inspiration.

Declaration

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

Serena Katherine Thomson

List of Abbreviations

AES - agri-environment schemes

AOA - ammonia oxidising archaea

ANOSIM - Analysis of similarity

ANOVA - Analysis of variance

AMF - Arbuscular mycorrhizal fungi

BaYMV - Barley yellow mosaic virus

BSBV - Beet soil-borne virus

BLAST – Basic local alignment search tool

C - Carbon

DNA - Deoxyribonucleic acid

DGGE - Denaturing gel gradient electrophoresis

EMF – Ectomychorrizal fungi

GSP - Genome Sequencing Program

ITS – Internal transcribed spacer non-functional region of rRNA

K – Potassium

Mg - Magnesium

N – Nitrogen

NGS - Next generation sequencing

NMDS – Non metric multidimensional scaling

P – Phosphorus

PCR – Polymerase chain reaction

PGM - Personal Genome Machine

PLFA - Phospholipid-fatty acid

QIIME – Quantitative Insights into Microbial Ecology

OTU – Operational taxonomic unit

RNA - Ribonucleic acid

SBWMV – Soil-borne wheat mosaic virus

SMS - ‘single-molecule’ sequencing

TRFLP - Terminal restriction fragment length polymorphism

WYMV - wheat yellow mosaic virus

ZMW - Zero Mode Waveguide

Summary

There has been an increasing emphasis placed on understanding microbial biogeography, in order to enable the patterns and processes governing the spatial distribution of soil microbiota to be determined. Due to current food security issues, this is particularly important within agricultural systems given the fundamental role microorganisms play in the maintenance of crop health and productivity. With evidence in favour of both ubiquity and endemism, complicated by systems, scales and communities, there is a need to address the question of microbial biogeography within a single system.

A range of field experimental resources were used to investigate factors controlling the assembly of soil microbial communities. Microorganisms across all three domains of life demonstrated spatial scaling, in which there was no single universal driver. Land-use management was an important driver of eukaryote distribution, but also impacted the drivers of bacterial and eukaryote taxa groups under land-use practice. When considering microbial community structure, a pan microbial relationship between abundance and distribution was shown for the first time, across all microbial groups. Furthermore, partitioning microbial communities into common and rare groups provided information on the processes operating on the community and highlighted the importance of land-use management for shaping the structure of communities. Finally, a case study on plasmodiophorids increased current estimates of plasmodiophorid diversity in the soil. Also different communities were associated with the rhizosphere compared with the bulk soil, under different hosts. Plant development stage was also an important consideration acting on this previously understudied but highly significant group of protists to crop health.

CHAPTER I: INTRODUCTION AND REVIEW OF THE LITERATURE

1.1 Agricultural intensification in the context of food security

1.1.1 Human population pressure on crop production

The world's population is currently estimated at 7.2 billion and expected to rise to 9.5 billion by 2050 (Curtis and Halford, 2014). Currently our food requirements are being met through cash crops such as barley, maize, rice and wheat (Tilman, 1999). In particular, world production of wheat grain in 2010 has grown to 651 t, to become one of the most-produced crops in the world, exceeded by rice (672 t) and maize (844 t). Many Asian countries in particular have seen huge increases in wheat consumption. For example Indonesia experienced a 262-fold increase in wheat consumption between 1962 and 2012 and a 105-fold increase was seen in Bangladesh (Curtis and Halford, 2013). In addition to human consumption, grains are being produced for grazing livestock and fuels such as bioethanol production. In the US alone, 42% of the maize produced was used for bioethanol production in 2012 (Curtis and Halford, 2013). As a result of this increased requirement for crops, demand on resources is increasing. Whilst global grain production has doubled in the last 50 years, the allocation of arable land for agriculture has only increased by 9% (Godfray et al., 2010). Therefore emphasis has shifted to increasing yield in cash crops through agricultural intensification to obtain maximum productivity from the limited land resource.

1.1.2 *Impacts of agricultural intensification on soil microbes*

Agricultural intensification has been successful in increasing crop productivity per unit area, thereby helping meet global food demands (Bommarco et al.,

2013). Often crops are grown in monocultures or shortened crop rotations (Bennett et al., 2011). In addition, increased chemical input (as opposed to organic manure and natural pest control), physical disturbance, mechanisation and the use of high-yielding crop varieties have resulted in a number of negative local consequences. These include a reduction in aboveground biodiversity, increased soil erosion and lower soil fertility (Matson, 1997; Postma-Blaauw et al., 2010). The lack of aboveground diversity coupled with the associated decline in soil health can have a dramatic impact on total biodiversity including soil invertebrates and microorganisms (Lienhard et al., 2014; Lupwayi et al., 1998; Matson, 1997). Species richness of arbuscular mycorrhizal fungi (AMF) in particular, has been shown to decrease in high-input monocropping systems (Lienhard et al., 2014; Oehl et al., 2003). This is significant given their role in protecting plants from root pathogens, to which 10% of crops are lost (Strange and Scott, 2005) and the facilitation of immobile nutrients such as phosphorous for plant utilization. Consequently soil processes such as maintenance of soil structure, decomposition and nutrient cycling, which maintain plant health, can be severely affected (Matson, 1997; Strange and Scott, 2005).

1.1.3 Impact of microorganisms on plant communities

Microbes are responsible for the cycling of nutrients within the soil via decomposition and nitrogen mineralization (Berthrong et al., 2013). As such, they play a significant role in maintaining aboveground diversity (van der Heijden et al., 2008). For example mycorrhizal fungi form symbiotic associations with plant roots, supplying the plant with nutrients to aid plant growth via their extensive mycelial network. Both arbuscular (AMF) and ectomychorrizal (EMF)

fungi are involved in carbon cycling, which is fundamental to agroecosytem functioning (Leake et al., 2004). Specifically, they are responsible for increasing plant uptake of key nutrients in particular P, in addition to N and K, in exchange for carbohydrates supplied by the host. AM fungi also offer protection from pathogens by enhancing tolerance or increasing their resistance. Finally they also function in stabilising soil aggregates, resulting in a macroporous structure that is more resistant to erosion (Oehl et al., 2003).

Many bacterial species play a significant role in maintaining plant productivity through nitrogen fixation (van der Heijden et al., 2008), whilst others are involved in denitrifying and nitrifying processes (Horz et al., 2004). Rhizobacteria such as the *Sinorhizobium* for example, colonise plant roots and form symbioses with plants to fix nitrogen and promote growth through the production of hormones (Hayat et al., 2010). Free-living microbes are also known to indirectly influence plant productivity by affecting nutrient supply and the partitioning of resources. For example they can do this by nutrient mineralization whereby insoluble and soluble organic matter is broken down and converted into inorganic plant available forms. They have also been shown to affect plant nutrient availability through the weathering of soils via the exudation of organic acids, in addition to the solubilisation of a number of forms of precipitated P (van der Heijden et al., 2008).

1.2 An introduction to microbial biogeography

1.2.1 The importance of understanding spatial scaling of microbes in the soil

Understanding the spatial distribution of microbes in the soil provides essential information regarding the underlying mechanisms that regulate microbial

biodiversity. Given the role microorganisms play in biogeochemical cycling and maintaining ecosystem functioning, these are essential considerations if we are to effectively manage our land and reduce the impact of anthropogenic change such as agricultural intensification (Prosser et al., 2007). By assessing beta-diversity, which describes how community composition changes across the landscape, the processes acting on communities can be determined. For example, some communities may be dispersal limited, whilst others are shaped by environmental heterogeneity. This quantitative information is useful when informing policy makers and will facilitate in the assessment and management of microbes and their contribution to ecosystem processes such as soil erosion and soil health along with climate change (Prosser et al., 2007).

It is well known that macroorganisms exhibit spatial scaling, however whether these patterns were present in microorganisms remained largely unknown until the last ten years. It was previously assumed that all microorganisms were cosmopolitan in their environment, which stemmed from Baas-Becking's statement "everything is everywhere, but, the environment selects". This has since been referred to as the ubiquitous dispersal hypothesis, which facilitates random dispersal due to the small body-size below approximately 1 mm and vast population size associated with microorganisms. As a result microorganisms are not limited by dispersal but have the capacity to disperse over long distances passively, in addition to having low probabilities of local extinction (Bass et al., 2007; Fenchel, 2003; Green and Bohannan, 2006). Studies investigating the distribution of Atlantic deep sea protists such as *Rhynchomonas parvulus* have demonstrated cosmopolitan communities (Scheckenbach et al., 2005). Finlay et al., (2001) also demonstrated random

spatial distributions in soil across scales of 4 m^2, 10,000 m^2 and globally in testate amoebae and ciliates, that varied based on their abundance. Bacterial studies addressing all taxonomic levels provide evidence for a cosmopolitan distribution from classes of bacteria such as cyanobacteria and actinobacteria, to genus and species level. For example the marine species *Nitrosococcus oceani* strain was found to be distributed throughout the world's oceans (Ramette and Tiedje, 2007; Rejmánková et al., 2004; Ward and O'Mullan, 2002). Finally microbial eukaryotes have also demonstrated cosmopolitan distribution, for example the human pathogenic fungus *Aspergillus fumigatus*, was shown to be globally distributed (Pringle et al., 2005).

However there is now evidence to suggest that distinct biogeographical patterns exist in many microbial groups, where microorganisms are not randomly distributed but exhibit spatially predictable aggregate patterns such as microbial endemism (Green and Bohannan, 2006; Whitaker et al., 2003). Green et al., (2004) investigated the spatial distribution of ascomycete fungi for example. They found that ascomycete fungi demonstrated distance-decay relationships whereby communities became less similar to one another with increasing distance. Whilst the importance of environmental processes has been well documented in the literature (Table 1.2), a number of other studies have also demonstrated that distance-decay relationships exist in bacteria, archaea and other eukaryotes such as AM fungi (Feinstein and Blackwood, 2013; van der Gast et al., 2011a; Ranjard et al., 2013; Reche et al., 2005; Whitaker et al., 2003) (Table 1.1).

Table 1.1 A selection of studies demonstrating either a significant distance-decay relationship (indicated with "yes") or no significant distance-decay relationship (indicated with a "no").

Organism	Environment	Distance-Decay effect	Reference
Bacteria	Phyllosphere	Yes	Finkel et al., 2012
Bacteria	Soil	No	Sayer et al., 2013
Bacteria	Lake	Yes	Reche et al., 2005
AM fungi	Soil	Yes	Green et al., 2004
Ciliate Protist	Sea	No	Stock et al., 2013
AM fungi	Soil	Yes	van der Gast et al., 2011
Archea	Hot Springs	Yes	Whitaker et al., 2003
Bacteria	Soil	Yes	Ranjard et al., 2013

With literature demonstrating spatial scaling in a number of microorganisms, emphasis has now shifted towards determining the processes that shape beta diversity within communities.

1.2.2 Mechanisms: Niche and Neutral processes

When making predictions regarding the distribution and abundance of species in space and time, there are two main theories that are commonly referred to. The first is niche-based theory, which describes species-environment relationships. It stems from the competitive exclusion principle, which states that species that share the same niches (i.e. same traits or response to the environment) cannot coexist indefinitely and that species differ based on an unlimited number of niches (Lekberg et al., 2007). That is, species that are similar are expected to have different habitat preferences and resource requirements, which reduces the chances of competitive exclusion through trade-offs (Beck et al., 2015). Modern niche theory addresses a number of factors that may influence the spatial distribution of a community, which are not limited to consumable resources. These include spatial and temporal heterogeneity (both intrinsic and extrinsic), space and interactions between other species (Chase, 2005).

A complementary and far more simplistic theory is neutral theory, which states that similar species will have the same life history strategies with no competitive advantages. These species are said to be competitively identical and the overall abundance of a species or speciation rate is not a result of variation between a species' trait (McGill et al., 2007). Spatial processes such as dispersal limitation, the size of the metacommunity, along with demographic stochasticity including extinction, immigration and speciation, govern patterns of diversity, relative abundance and composition (Beck et al., 2015; Chase, 2005; Lekberg et al., 2007). Given the simplicity of the model, there are questions that cannot be addressed with neutral theory. For example, neutral theory does not recognise that species differ in their ecological traits, therefore predictions regarding the impact of habitat fragmentation or invasive species on community cannot be made. Similarly whilst stochastic changes in community composition can be predicted with neutral theory, it cannot predict directed (non-random) shifts in composition through time. For example changes in the environment such as increased nutrient concentration or temperature can favour a species, which, due to its traits, are able to respond optimally to the new conditions (Chase, 2005).

There is support for both theories in the literature. Many studies have shown that environmental variables correlate with the spatial distribution of microorganisms. For example pH is a well-known driver in the spatial scaling of bacteria, even at coarse levels of taxonomic resolution (such as phylum-level; Acidobacteria, Actinobacteria and Bacteroidetes). For example, Griffiths et al., (2011) found that Acidobacteria were more abundant in low pH soils, and in addition abundance was associated with factors such as climate, soil biochemistry and plant community composition (Fierer and Jackson, 2006;

Lauber et al., 2009; Rousk et al., 2010). Whilst the specific mechanisms responsible for these correlations have yet to be confirmed, general explanations for the strong correlation of bacteria with pH have been put forward. The first is that soil pH acts as an integrated variable providing an index of soil conditions. This is based on the fact that many other soil variables are either directly or indirectly linked to soil pH including moisture, nutrient availability and salinity (Lauber et al., 2009). The second is that microorganisms have an optimum range in which they grow. For example for many microorganisms, the intracellular pH is within 1 pH unit of neutral. Therefore deviations in extracellular pH are likely to cause the microorganism physiological stress and are likely to impact fitness (Fierer and Jackson 2006). For the reason mentioned, it is difficult to determine whether pH directly is responsible for differences in bacterial community composition, given that soil pH is correlated with a number of biotic and abiotic factors. For instance C and N cycling has been shown to be significantly affected by pH. Specifically a Rothamsted and Woburn based study by Kemmitt et al., (2006) showed that an increase in soil acidity resulted in reduced C and N cycling through a decline microbial activity. Similarly pH has been shown to correlate with crenarchaeal distribution (Bru et al., 2011). Furthermore a multitude of environmental parameters including ammonia concentration, organic matter, temperature and oxygen have been shown to play a role in shaping microbial communities including ammonia oxidising archaea (AOA). For example, whilst most AOA are found at temperatures ranging from 22 to 37 °C, some AOA (including extreme acidophilic and hyperthermophilic *Crenarchaeota* are able to tolerate temperatures between 82 and 97 °C and pH ranges between 2.5 and 7 (Zhalnina et al., 2012, Reigstad et al., 2008). Other

studies have demonstrated the importance of soil type, moisture and soil nutrient status in predicting the spatial distribution of AM fungi (Hazard et al., 2014; Lauber et al., 2008). The community composition of other eukaryotes such as soil protists (Alveolates and Rhizaria) has been shown to correlate with climatic conditions that regulate annual soil moisture availability (Bates et al., 2012) (Table 1.2).

The role of spatial processes in driving the distribution of microorganisms has also been recognised in the literature, although there are far fewer studies that directly measure geographical distance as a driver of microbial distribution (Martiny et al., 2006). Bacterial communities inhabiting the extreme environment of the phyllosphere of *Tamarix* leaves correlated with geographical distance as opposed to pH (Finkel et al., 2012). Papke et al., (2003) and Whitaker et al., (2003) investigated the importance of distance in the composition of *Synechococcus* and *Sulfolobus* respectively. They also concluded that geographical distance was the main driver of their distributions within a hot springs environment (Table 1.2).

Finally, studies have shown that a combination of the two theories (environmental factors in addition to dispersal) can act on communities. High beta diversity can be the result of a number of mechanisms that are unlikely to be mutually exclusive (Kivlin et al., 2011). Dumbrell et al., (2009) investigated the mechanisms driving the composition and structure of AM fungi. They concluded that both neutral and niche processes were acting on the community in the form of dispersal limitation and pH, although niche processes were the primary mechanism regulating the community. Similarly Kivlin et al., (2011) investigated the drivers of AM fungal distribution. They too found evidence for dispersal

limitation in addition to plant community type, soil moisture and temperature. This is also supported in bacterial studies that have demonstrated that both environmental processes and dispersal are involved in spatial scaling in soil communities and lake sediments (Ranjard et al., 2013; Xiong et al., 2012) (Table 1.2).

It has been proposed that the niche and neutral models represent two extremes of a continuum (Beck et al., 2015). Recent niche-based models have included dispersal limitation and incorporated large-scale processes such as speciation and extinction. Therefore the trend is for the integration of niche and neutral theories. What might be a more ecologically relevant question when addressing the mechanisms responsible for driving the spatial distribution of microorganisms, is how much of a role do niche and neutral based mechanisms play in shaping community assembly (Beck et al., 2015).

Table 1.2 Studies of the effects of distance, environment and both distance and environment on microbial composition. Scale is based on categories provided in Hanson et al., (2012); local = 0 - 100 km, regional = 101 - 5,000 km, intercontinental = > 5,000 km.

				Effect of:		
Organisms	Scale	Habitat	OTU	distance	environment	Reference
Bacteria	Regional	Soil	TRFLP	No	Yes	Griffiths et al., 2011
Bacteria	Local	Soil	Bar-coded pyrosequencing	No	Yes	Rousk et al., 2010
Bacteria	Intercontinental	Soil	TRFLP	No	Yes	Fierer and Jackson, 2006
Bacteria	Intercontinental	Soil	Bar-coded pyrosequencing	No	Yes	Lauber et al., 2009
AM fungi	Regional	Soil	TRFLP	No	Yes	Hazard et al., 2013
Archaea	Regional	Soil	qPCR	No	Yes	Bru et al., 2011
Protist	Intercontinental	Soil	Bar-coded pyrosequencing	No	Yes	Bates et al., 2012
Bacteria	Local	Phyllosphere	16S rRNA gene pyrosequencing	Yes	No	Finkel et al., 2012
Synecoccocus	Intercontinental	Hot Springs	16S/ITS sequence	Yes	No	Papke et al., 2003
Sulfolobus	Intercontinental	Hot Springs	MLS of isolates	Yes	No	Whitaker et al., 2003
Bacteria	Local	Lake	DGGE	Yes	No	Reche et al., 2005
AM fungi	Intercontinental	Soil	18S.28S sequence	Yes	Yes	Kivlin et al., 2011
Bacteria	Regional	Soil	ARISA	Yes	Yes	Ranjard et al., 2013
AM fungi	Local	Soil	TRFLP	Yes	Yes	Dumbrell et al., 2009
Bacteria	Intercontinental	Lake sediment		Yes	Yes	Xiong et al., 2012
Ascomycete	Local	Soil	ARISA	Yes	Yes	Green et al., 2004
AM Fungi	Regional	Soil	TRFLP	Yes	Yes	Gast et al., 2011
Bacteria	Regional	Lake	ARISA	Yes	Yes	Yannarell et al., 2005

1.2.3 Impact of habitat type on spatial scaling in the soil

The different ecological processes that act on communities, are likely to be influenced by a number of additional factors including habitat, community and spatial scale (Feinstein and Blackwood, 2013). For example, the type of habitat will have an impact on the mechanisms that drive the spatial distribution of a community. As a result there is a wealth of information examining microbial community composition across disparate habitats. Habitats including lakes, marine environments, soil, salt marshes and hot springs, have highlighted the importance of a range of variables that act on the microbial communities that inhabit them. For example, bacterial communities within lake environments were driven by geographical distance in addition to water clarity and pH gradients (Yannarell and Triplett, 2005). This was supported by a separate study, which also demonstrated the importance of geographical distance, where it was shown to be a significant predictor of bacterial communities in the mountain lakes of Sierra Nevada in Spain (Reche et al., 2005). The spatial predictors of bacterial communities in the soil have been extensively studied. As mentioned, the consensus in the literature emphasizes the importance of soil pH in driving the spatial scaling of bacteria and this has been supported by studies across the globe with evidence originating from North America, the United Kingdom and the Arctic (Chu et al., 2010; Lauber et al., 2009; Rousk et al., 2010). Within salt marshes, whilst bacteria communities demonstrated a distance-decay relationship, distribution correlated with environmental heterogeneity over spatial processes (Horner-Devine et al., 2004). A similar study however showed that salt marsh communities were driven by both the environment and

geographical distance, but that the influence of each depended on the scale of the study (Martiny et al., 2011).

1.2.4 Impact of scale on spatial scaling in the soil

The categorization of scales varies widely in the literature but can loosely be divided into continental (>5,000 km), regional (1 km - 5000 km) and local scale (<1 m -1 km). Martiny et al., (2011) demonstrated that the relative importance of environmental variables and geographical distance differed across three spatial scales (within marshes, across marshes and across continents). This highlights the importance of considering scale when investigating microbial biogeography. Geographical distance influenced *Nitrosomonadales*, an order of proteobacteria within marshes, whilst sediment moisture was the main predictor of communities within marshes, compared to water temperature and nitrate concentration, which were the primary predictors at the regional and continental scale (Martiny et al., 2011). The general consensus however in the literature (although there are exceptions to the rule) is that environmental variables are commonly identified as predictors of microbial distribution at local scales (Horner-Devine et al., 2004; Martiny et al., 2006). For example, in their study ranging up to 1 km, Zinger et al., (2011) demonstrated that beta diversity patterns of crenarchaea, fungi and bacteria in alpine soil were all influenced by plant community composition.

At regional scales, both the environment and geographical distance are expected to play a role in influencing the composition of communities. In some studies both factors have been identified within a single study. For example, ascomycete community turnover could be predicted by geographical distance in addition to environmental heterogeneity (Green et al., 2004). Typically however

the evidence for both factors playing a role has been derived from a multitude of studies. Cho and Tiedje, (2000) for example, demonstrated the influence of geographical distances between 5 m and 80 km in the community composition of fluorescent *Pseudomonas* strains in soil. Whereas other studies have shown environmental heterogeneity to be the main factor in determining the distribution of bacteria, crenarchaea and AM fungi (Bru et al., 2011; Hazard et al., 2013). At the global scale, distance has been shown to be the major predictor in the spatial scaling of microorganisms (Whitaker et al., 2003, Martiny et al., 2006).

1.2.5 Impact of community on spatial scaling in the soil

Few studies investigating the drivers of beta diversity have focussed on more than one taxon in a single study and as such there is limited understanding of the variation in the scaling of beta diversity across distinct taxa. This is an important consideration given that drivers of beta diversity are likely to reflect varying life history strategies and trait complexes such as body size and dispersal ability, that act on microbes (Barton et al., 2013). For example, soil protozoa are thought to be dispersed via percolating rainwater, in addition to burrowing invertebrates (Finlay et al., 2001). Whilst wind currents are reportedly the main mode of dispersal for spores of soil fungi, bacteria on the other hand have been shown to be transported using hyphal biofilms of fungi (van der Heijden et al., 2008, Kivlin et al., 2014). Of the few studies that have addressed more than one taxon, (Rousk et al., 2010) they investigated bacterial and fungal community composition and showed differences between the two domains. For example, pH was a strong predictor of bacterial community composition, whilst fungal community composition was only weakly related to pH demonstrating the

different life history strategies of the two groups. Other studies on fungi have shown the importance of geographical distance over the environment in shaping ascomycete communities (Green et al., 2004), compared with environmental factors in the distribution of AM fungi (van der Gast et al., 2011a). Differences in the drivers of microbial communities between and within microbial taxa are common in the literature but difficult to collate, for a number of reasons. For instance, a variety of methodologies are often employed to determine the importance of a set of predictors, including TRFLP (Terminal restriction fragment length polymorphism), DGGE (Denaturing Gradient Gel electrophoresis) and 454 pyrosequencing. Each of the studies sampled different locations, environments, suite of variables etc. This makes it difficult to definitively determine the main factors influencing microbial groups. For example, Bru et al., (2011) demonstrated the importance of pH in driving the distribution of crenarchaea communities, whilst Angel et al., (2010) showed that climate and vegetation cover explained the differences in archaeal community composition. Finally, Bates et al., (2011) conducted a global scale study of archaeal distribution and found C:N was the best predictor of archaea distribution.

1.2.6 Land use management and microbial distribution patterns

The processes responsible for explaining the variability in microbial communities have largely been studied within a natural environment. It is of fundamental importance to also consider the impact of land use management on the spatial scaling of microbial communities, given the influence of agricultural practice on soil biochemistry. Anthropogenic disturbance in the form of pesticide

and fertilizer application, crop management i.e. rotation and tillage practices, add selective pressures to the microbial community, which have been shown to result in changes in diversity and composition of microbial communities. For example intensively managed land is often associated with lower levels of diversity in AM fungi and bacteria (Lupwayi et al., 1998; Oehl et al., 2003). Similarly conventional farms are typically associated with lower levels of between-site microbial diversity when compared with organic farms in both AM fungi and bacteria (Verbruggen et al., 2010). Furthermore, studies have demonstrated that land use management impacts the biogeographical distribution of AM fungi, whereby AM fungal community composition and structure within conventional farms was significantly different to organic farms (van der Gast et al., 2011a; Hazard et al., 2013, 2014). However few studies have investigated whether microbial spatial scaling is affected by land use management, which can affect the physical and chemical properties of the soils and ultimately soil health. Tillage and physical disturbance such as soil compaction and crusting can result in a reduction in pore space and therefore soil organic carbon (Bronick and Lal, 2005). Additionally a change in plant species composition and the application of fertilisers can impact the chemical properties of the soil such as altering the C and N balance, which will impact soil microbial communities (Ye et al., 2009, Lauber et al., 2008, Jangid et al., 2008). It is therefore essential to understand the implications of anthropogenic change (in this case conversion of arable land to set aside) on the spatial distribution of microorganisms given their role in ecosystem services such as CO_2 respiration and decomposition, in addition to nitrogen cycling. This information will aid policy makers in the effective

management of the soil in order to improve ecosystem services and soil function (Martiny et al., 2006, Acosta-Martínez et al., 2008).

1.3 Cropping systems and microbial communities – with reference to Plasmodiophorids

1.3.1 Plasmodiophorid phylogeny and life cycles

In order to fully understand the role microbes play in the soil, it is necessary to consider the distribution of soil microorganisms across the landscape (van der Gast et al., 2011a). One particular group for which there is an incomplete understanding regarding spatial ecology, and in particular their diversity and distribution within the soil, is the 'phytomyxid' protist group. This is a fundamental issue given that of the 41 species known in the literature, the majority are classified as pathogens and as such are responsible for huge economic losses in the crop industry, examples of which are shown in Figure 1.1 (Neuhauser et al., 2011a).

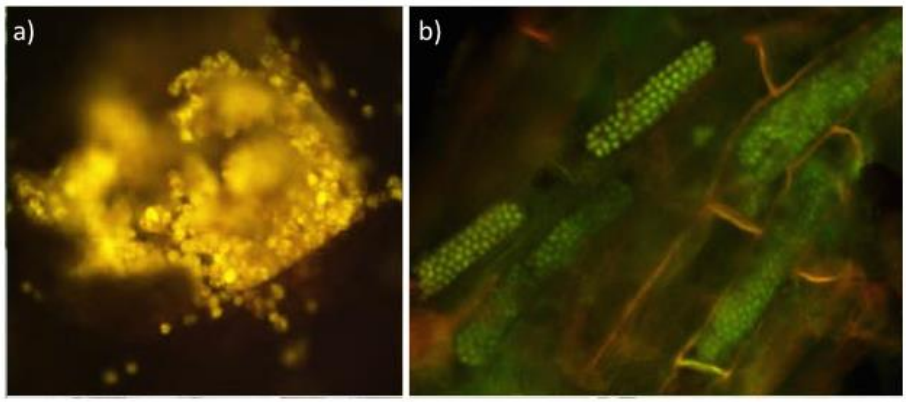


Figure 1.1 a) Yellow/orange autofluorescence of *Sorosphaera viticola* resting spores (cytosori) in grapevine roots; b) Green - acridine orange stained resting spores (cytosori) of *Polymxya graminis* in *Poa* spp (Images supplied by Martin Kirchmair and Sigrid Neuhauser).

Phytomyxids (also referred to as Phytomyxea) include two distinct orders: phagomyxida, known to parasitize brown algae and diatoms, and

plasmodiophorida, which are parasites of green plants and oomycetes (Neuhauser et al., 2011b). Phytomyxids occupy a range of ecosystems including soil, freshwater and marine environments. Along with their sister taxa Vampyrellidae, they group within Endomyxa, the subphylum of Cercozoa which belong to the protist supergroup Rhizaria (Braselton, 1995). Phytomyxids are obligate biotrophs with complex multiphasic life cycles consisting of two main developmental phases. These include two types of plasmodia (sporogenic and sporangial) and two types of zoospores (primary and secondary) (Braselton, 1995; Neuhauser et al., 2011b), which are the only part of the life cycle outside of the host. Their complex life cycle and small size (3-6 μm) make them difficult to identify morphologically and therefore there is limited understanding of their biodiversity (Neuhauser et al., 2014) (Figure 1.2).

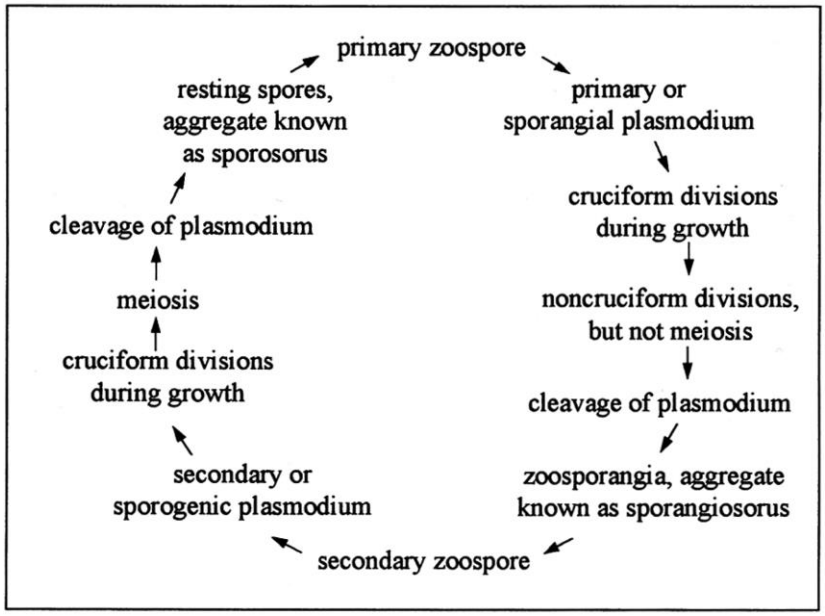


Figure 1.2: Generic life cycle of plasmodiophorids (taken from Braselton, 1995).

1.3.2 *The role of plasmodiophorids in the soil as protist pathogens*

Phytomyxids are likely to serve as important components of the soil food web. The walls of the resting spores of *Plasmodiophora brassicae* for example contain

25% chitin, 17.5% lipids and as much as 33% protein (Neuhauser et al., 2011b). These spores however can remain viable in the soil for many years whereby the spore inoculum has a half-life of 3.6 years (Wallenhammar, 1996). The spores of *Plasmodiophora brassicae* for example contain five spore walls composed of chitin and carbohydrates, proving the spore with excellent protection from degradation by extra-cellular enzymes (Dixon, 2014).

Plasmodiophorids are however most well known for their role as plant pathogens and viral vectors of major plant diseases in economically significant cash crops. It is well documented that plasmodiophorids pose a serious threat to cereal and grass production of many crop species (Dixon, 2009; Kanyuka et al., 2003).

Plasmodiophora brassicae infects *Brassica* species and crucifers across the world causing clubroot disease. It functions by disrupting the host growth regulator metabolism and in doing so forms galls on the roots, which serve as ideal environments for the pathogen (Dixon et al., 2014) (Figure 1.3). The galls inhibit water and nutrient transport to the plant thereby wilting leaves and stunting the plant's growth (Bulman et al., 2006; Hwang et al., 2012; Manzanares-Dauleux et al., 2001). Infestation is thought to be affected by soil type where clay soils have been shown to be associated with the highest levels of infestation when compared with other soil types such as silt, fine and coarse sand. The severity of clubroot infection is also thought to be related to soil moisture content and the physical condition of the soil (Wallenhammar et al., 1996). Interestingly however there are some studies in the literature that suggest that the severity of *Plasmodiophora brassicae* infection can be mitigated through

the presence of other plant species such as mint, chive, parsley and basil (Dixon et al., 2014).

Figure 1.3 1 - Clubroot symptoms in an infected brassicae root, 2 - uninfected roots of the same age plant, 3 - resting spores of Plasmodiophora brassicae (Images supplied by Martin Kirchmair and Sigrid Neuhauser).

Another gall forming protist is *Spongospora subterranea,* which causes powdery scab in potatoes (*Solanum tuberosum).* By replicating in the roots and stolon they produce pustules or lesions on the tubers (Gilchrist et al., 2011). These structures are home to spore aggregations (sporosori), which too can remain dormant in the soil for over 10 years. The infection not only results in unmarketable potatoes but also impacts root function, which is expected to reduce yield (Falloon, 2008). Gilchrist et al., (2011) demonstrated that potato height and foliar dry weight were reduced by 23 % and 32 % respectively, corresponding to a 30 % yield reduction, with just a 5 % coverage of galls on the root surface.

Unlike *Plasmodiophora brassicae, Polymyxa graminis* causes damage by acting as a vector for a suite of viruses infecting the root epidermal cells of wheat plants where the protist itself is thought to do very little damage. A large number of crops are affected, including barley, which is infected by barley yellow mosaic virus (BaYMV), and wheat, infected by wheat yellow mosaic virus (WYMV) and soil-borne wheat mosaic virus (SBWMV) (Tamada and Kondo, 2013, Table 1.3). Although symptoms vary depending on the crop and type of virus that is infected, the general symptoms include discoloration of leaves such as yellowing or mottling and subsequent stunted growth (Kanyuka et al., 2003; Ward and Adams, 1998) .

Other plasmodiophorids include *Polymyxa betae*, which is morphologically very similar to *Polymyxa graminis* and is a vector of the beet soil-borne virus (BSBV) in sugar beet. *Sorosphara viticola* is a parasite of grapevines, infecting the cortical tissue of vine roots, whilst *Woronina spp* is a parasite of the oomycete *Pythium*, which itself is a parasite. The group is therefore diverse and complex in their modes of transmission and infection and pose a serious threat to the health of a variety of crop species.

Table 1.3 Cereal viruses transmitted by *Polymyxa graminis*. *Formally accepted virus species appear in italics, and tentative species are in regular font (Taken from Kanyuka et al., 2003).

Virus*	Acronym	Genus	Natural hosts	Distribution
Rice stripe necrosis virus	RSNV	*Benyvirus (?)*	Rice	West Africa, South and Central America Europe
Barley mild mosaic virus	BaMMV	*Bymovirus*	Barley	Europe, Japan, China, Korea
Barley yellow mosaic virus	BaYMV	*Bymovirus*	Barley	Europe, Japan, China, Korea
Oat mosaic virus	OMV	*Bymovirus*	Oats	Europe, USA
Rice necrosis mosaic virus	RNMV	*Bymovirus*	Rice	Japan, India
Wheat spindle streak mosaic virus	WSSMV	*Bymovirus*	Wheat, rye, triticale	North America, Europe
Wheat yellow mosaic virus	WYMV	*Bymovirus*	Wheat	Japan, India
Chinese wheat mosiac virus	CWMV	*Furovirus*	Wheat	China
Oat golden stripe virus	OGSV	*Furovirus*	Oats	Europe, USA
Soil-borne cereal mosaic virus	SBCMV	*Furovirus*	Wheat, rye, tricale	Europe
Soil-borne wheat mosaic virus	SBWMV	*Furovirus*	Wheat, barley, rye, triticale	North America,?elsewhere
Sorghum chlorotic spot virus	SrCSV	*Furovirus*	Sorghum	USA
Peanut clump virus	PCV	*Pecluvirus*	Peanut, sorghum	India, West Africa
Aubian wheat mosaic virus	AWMV	?	Wheat	France, UK

1.3.3 Implications of plasmodiophorid infestation on crop production

The cost of phytomyxid infections although high, is difficult to gauge. Crop losses as a result of infestation vary widely in the literature and depend on the crop and pathogen. For example a survey described in Dixon (2009), assessed *Plasmodiophora brassicae* infections in multiple countries and concluded that infestation averaged at 11%. The average infection rate in countries such as Wales and Scotland were as high as 45 % and 48 % respectively compared with countries such as Canada, which were less than 1 % (Dixon 2009) (Table 1.4).

Table 1.4 Worldwide survey of infection rates of *Plasmodiophora brassicae* (taken from Dixon et al., 2009).

Country	Average % infection	Country	Average % infection
Australia	6	Japan	5
Canada	0.11	Netherlands	10
Czech republic	10	New Zealand	15
Denmark	5	Norway	12
England	6	Poland	4
Finland	4	Scotland	48
France	3	Sweden	1
Germany	8	USA	10
Ireland	17	Wales	45

Much of the literature on plasmodiophorids is directed towards methods of detection, their detrimental impact on crops and management solutions such as longer rotations (Dixon, 2009; Faggian and Strelkov, 2009; Falloon, 2008). As a result very few species have been studied and little to no emphasis has been placed on diversity of the group as a whole.

1.4 Microbial community profiling

1.4.1 Next generation sequencing technologies advances and applications in microbial ecology

It has been nearly 45 years since the introduction of Sanger sequencing and only 10 years since next generation sequencing (NGS) technologies were introduced to the field of biology, courtesy of Roche's 454. Before the advent of NGS technologies, molecular microbial community profiling was achieved through techniques such as terminal restriction fragment length polymorphism (TRFLP) and denaturing gradient gel electrophoresis (DGGE). TRFLP profiles microbial communities using fluorescently labelled primers to generate fluorescently-labelled terminal restriction fragments which can be separated out using electrophoresis (Applied Biosystems). DGGE on the other hand relies on different fragments having different denaturation profiles (based on nucleotide composition) to separate out DNA from different taxa (Muyzer and Smalla, 1998). Since the introduction of NGS however there has been an explosion of studies, increasing current estimates of microbial diversity from environmental samples, and understanding of microbial processes and patterns (Reeder and Knight, 2010). A number of marker genes are used for biodiversity analysis including 16S rRNA, which is typically used for bacterial identification, whilst 18S rRNA and the internal transcribed spacer (ITS) of the rDNA is used for eukaryote studies.

Current NGS technologies have the capability for genome sequencing (generation of sequence reads from fragmented libraries), RNAseq or transcriptome sequencing (generation of sequence reads from a pool of cDNA library fragments, which have been generated through the reverse transcription of

RNA molecules). Thirdly, sequence reads can be generated from a pool of PCR-amplified molecules (amplicon sequencing) (Shokralla et al., 2012). NGS technologies can be partitioned into two major categories, those that rely on PCR-based methods and those that utilise 'single-molecule' sequencing (SMS) technologies. PCR- based methods include 454 pyrosequencing by Roche 454 Genome sequencer, (Roche Applied Sciences, Basel, Switzerland), based on a sequencing-by-synthesis method which generates ~400,000 250 bp reads in GS FLX implementation and up to 800,000 400 bp reads with titanium reagents (Quince et al., 2011; Siqueira et al., 2012a). The second is HiSeq by Illumina (San Diego, CA, USA), which, is based on bridge PCR amplification and generates 100 bp reads and finally the SOLiD system (Applied Biosystems, Foster City, CA, USA), which uses sequencing-by-oligo ligation technology (as opposed to sequencing-by-synthesis used by Illumina and 454 platforms). The SOLiD system generates reads that are between 50 bp and 75 bp in length. The latter two technologies generate between 30 and 100 million reads (although Illumina offers the highest throughput per run at the best price (van Dijk et al., 2014). Finally Ion Torrent (formerly Life Technologies) released the Personal Genome Machine (PGM), which uses semiconductor technology to detect incorporated nucleotides (Siqueira et al., 2012a, van Dijk et al., 2014). SMS technologies include Heliscope (Helicos BioSciences Corp., Cambridge, MA, USA), which also incorporates a sequencing-by-synthesis approach and is capable of producing around 1 billion sequence reads. The second system is the SMRT (single-molecule real-time) system (Pacific Biosciences, Menlo Park, CA, USA) (Shokralla et al., 2012). This is another real-time fluorescence-based SMS platform and is capable of incorporating ten or more nucleotides every second in

several thousand parallel ZMWs (Zero Mode Waveguide), a type of nano-structure used for DNA polymerization (Shokralla et al., 2012) (Table 1.5).

Table 1.5 Next-generation sequencing technologies (Taken from Siqueira et al., 2012a)

Platform	Library preparation	Chemistry	Read Length	Bases per run	Run time
Roche 454 GS FLX Tit	Emulsion PCR	Pyrosequencing	400	500 Mb	10 h
Illumina/solexa	Bridge PCR	Reversible terminators	100	18-35 Gb	4-9 d
SOLiD	Emulsion PCR	Sequencing by ligation	50	30-50 Gb	7-14 d
Helicos	Single molecule	Reversible terminators	32	37 Gb	8 days
Sanger	PCR and cloning	Dye terminators	800	800 bp	3 hours

1.4.2 The advantages and disadvantages of NGS technologies

The advancement of NGS brought with it several benefits. For example there is no longer the requirement to clone DNA fragments, which produced hundreds of sequences. Instead the 'cell-free' system has enabled the production of hundreds of thousands (454) to millions of reads (Illumina and SOLiD) in a single run, thereby substantially increasing throughput (Mardis, 2008; Metzker, 2010). Clone-based sequencing can also involve multiple steps in sample preparation compared with high-throughput techniques, which are comparatively more streamlined, therefore saving time. Given that sequencing no longer requires electrophoresis, but is run in parallel, the cost per base is much lower than traditional sequencing methods (and is continuing to decrease in price) (Schuster, 2007). For example the cost to sequence the human genome has dropped over the past 10 years from over $10 million at the turn of the century to around $5,000 in 2014 (Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) (Figure 1.4).

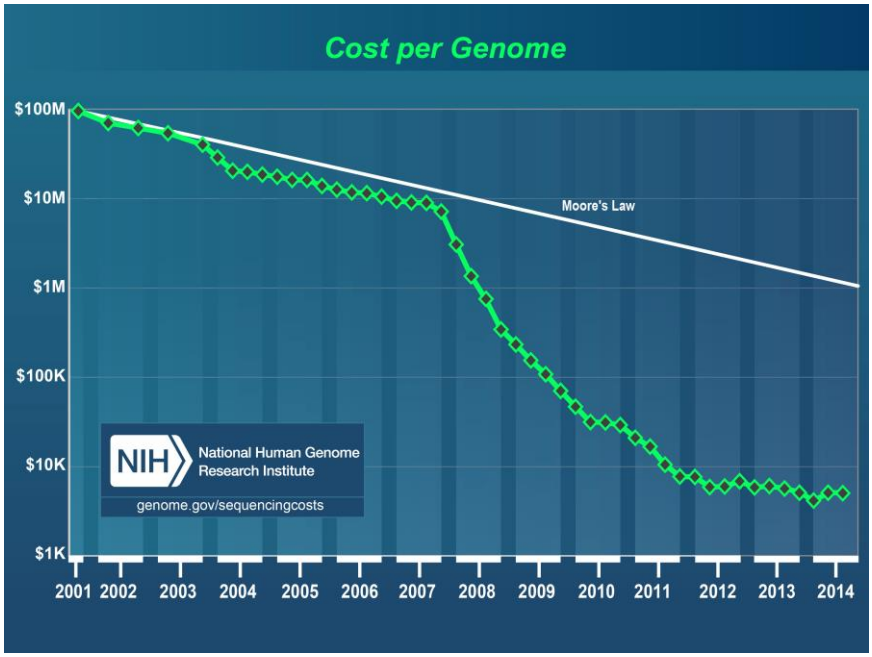


Figure 1.4 Decline in costs associated with sequencing the Human Genome from 2001 to 2014 (Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

Whilst there are many obvious benefits associated with NGS, there are also disadvantages associated with the techniques that must be considered. Short read length (~250 bp or less) makes reliable taxonomic assignments at species level difficult and are therefore more reliable at phyla, class, order, family and genera (Metzker 2010). Secondly the run time for NGS technologies are typically longer than traditional Sanger sequencing (between 8 hours and 10 days) depending on the platform and whether single or paired-end reads are generated (Mardis, 2008). Given the quantity of data generated using high-throughput techniques, diversity estimates using 454 sequencing can often be over inflated through spurious OTUs. This has lead to studies making claims about the 'rare biosphere' that are likely to be an over estimate (Quince et al., 2011; Sogin et al., 2006). It is therefore important to obtain as true a representation of diversity as possible through the filtering of noise. Three main sources of error have been reported all of which can inflate diversity estimates

including sequencing error, where the presence of homopolymers (a continual run of the same nucleotide) often results in base insertion and deletion errors. Secondly PCR single base substitutions are associated with the amplification of the PCR product at the start of the process and finally chimeras in which sequences are composed of two or more true sequences (Quince et al., 2009). Whilst programmes exist to remove noise associated with homopolymers, which are a major source of error, these algorithms are based on read abundance information and can remove the 'rare' (less abundant) taxa, along with valuable biological signals (Bik et al., 2012).

1.4.3 Bioinformatics as a tool for data analysis

The generation of much larger datasets has required computational advances in order to handle the data produced in terms of management, transfer and storage. Given the likelihood of overestimating diversity and richness in microbial community analysis (due to the abundance of low-quality read), there is a need for noise detection algorithms (Quince et al., 2009). Therefore computational algorithms have been developed to process and analyse the large amount of sequencing and PCR-based errors that can be generated with NGS. Popular pipelines include QIIME and Mothur, which have been developed to aid in the processing of raw sequences from 'clean up' to interpretation of large datasets. This has enabled the characterization of patterns in whole microbial communities based on marker gene surveys (Caporaso et al., 2011).

QIIME in particular encompasses a multitude of tools to remove noise, analyse data and visualize results. Denoiser and AmpliconNoise are tools, which cluster raw flowgrams (patterns of light intensities) accounting for errors

associated with miscalls when homopolymers are present. The drawback to these programs is that they are computationally expensive, requiring more computing power than is typically available on an average office computer and take a long time to run. ACACIA however is a tool for homopolymer error-correction whereby peak memory usage is ~14x lower than that used by AmpliconNoise (although 1-4x higher than Denoiser). It is also ~500x faster than Denoiser and over 2,000 faster than AmpliconNoise. It is however less sensitive to the correction of substitution errors as it only deals with homopolymer errors but is a valid alternative to AmpliconNoise and Denoiser as it maintains sensitivity (Bragg et al., 2012).

PCR based errors can be addressed using ChimeraSlayer and Uchime, which are designed to detect chimeric sequences (generated when incomplete extension occurs during PCR to produce two new fragments which act as primers in the next round of PCR) (Quince et al., 2011). It is thought that between 1 % and 45 % of a sample can be made up of chimeric sequences (Haas et al., 2011). To address these issues, ChimeraSlayer was developed by Haas et al., (2011) which was superseded by Perseus (Quince et al., 2011). UChime (developed by Edgar et al., 2011) was reported to be comparable to Perseus, balancing specificity and sensitivity, however a more accurate option currently exits; USEARCH (all three algorithms are reviewed in Schloss et al., (2011)). UPARSE is a separate pipeline (also developed by Edgar) that encompasses a number of steps involving removal of barcodes and primers, OTU clustering in addition to chimera filtering which greater accuracy than previous tools (Edgar, 2013). Software tools are being developed which require less computationally

expensive algorithms meaning larger datasets can be handled keeping up with the rapid advancement of sequencing technologies.

1.4.4 The future of next generation sequencing

Fierce competition now exists between sequencing manufacturers such as Roche, Illumina and Life Technologies to advance sequencing technologies further. Improvements include reducing the running costs, increasing sequence output per run and read length and reducing base-calling errors (Mardis 2011). At the time of writing, state of the art technology included third generation technologies such as SMRT, which are moving in this direction by addressing the drawbacks associated with traditional second-generation technologies such as Illumina. These advancements include having lower error rates with higher accuracy. Speed is also a relevant consideration with technologies moving to produce entire genomes in under a day at a low cost (Schadt et al., 2010).

High-throughput sequencing has advanced the field of microbiology and enabled the high-resolution analysis of the ecological and functional roles of microbial communities. Whilst whole genome analyses such as metagenomics and metatranscriptomics are popular, considerable value is still associated with 16S, 18S, 23S and ITS rRNA based studies. They provide insights into the ecological characteristics of whole microbial communities and allow the exploration of previously undiscovered biodiversity providing an insight into whole ecosystems. A continued challenge however is the processing of growing datasets and storage associated with the technology, which will need to be refined and made accessible to biologists (Bik et al., 2012, Caporaso et al., 2011).

1.5 AIMS AND OBJECTIVES

The overall aim of this work was to investigate microbial structure and composition under land use management, with reference to plasmodiophorids as a case study. The thesis has been divided into five sections in total, with three self-contained experimental chapters (Chapters II – IV), each with their own defined aims, methods, results and discussion sections, in addition to a general introduction (Chapter I) and a general discussion (Chapter V).

Chapter II aimed to determine the drivers of spatial scaling in Bacteria, Archaea and Eukaryote communities under contrasting land use management. The objectives considered were:

(i) To determine if there is a universal driver of microbial spatial scaling in different taxa groupings of microorganisms within and between the different land uses.
(ii) To determine whether land use management has a discernible effect on α and β-diversity for any of the microbial groupings.

Chapter III aimed to determine the impact of land use management on the commonness and rarity of microbes. The objectives considered were:

(i) To determine whether coherent metacommunities exists across all three domains of life with an agricultural soil system, as indicated by a universal positive relationship between taxa abundance and distribution?

(ii) To investigate whether the community structure of multiple microbial groups can be partitioned into common and rare members.
(iii) To determine the impact of land use management on the common and rare communities.
(iv) To understand the impact of land use management on microbial community structure in terms of dominance and evenness.

Chapter IV aimed to investigate novel plasmodiophorid diversity and dynamics in agricultural soil and assess the factors influencing their community composition in relation to yield decline. The objectives considered were:

(i) To test newly-designed PCR primers for the specific amplification of plasmodiophorid 18S rDNA and determine the performance of these primers in the characterization of plasmodiophorid communities.
(ii) To determine the key factors that shape plasmodiophorid communities within an agricultural system.
(iii) To identify the common and rare taxa within the community and their seasonal and spatial dynamics.
(iv) To investigate whether there is a relationship between dynamics of plasmodiophorid communities and crop yield.

CHAPTER II: COMPARISON OF THE DRIVERS DETERMINING THE SPATIAL SCALING OF SOIL BACTERIA, ARCHAEA AND EUKARYOTE COMMUNITIES UNDER CONTRASTING LAND USES

2.1. Introduction

Rapid global population growth is driving an unprecedented demand for food production, and the resulting agricultural intensification is a leading cause of wide scale habitat loss and decline in global biodiversity (Chivian and Bernstein, 2008; Pywell *et al.*, 2012). Consequently, sustainable intensification is needed to allow more food to be produced from the same area of agricultural land while reducing environmental impacts (Royal Society, 2009; Godfray *et al.*, 2010). There has been a strong drive for sustainable agricultural management, balancing agricultural production and conservation of biodiversity, across various parts of the world (Rands *et al.*, 2010) and particularly in Europe through agri-environment schemes (AES). Set-aside was a European Union AES introduced in 1992 to ameliorate the unintended consequence of agricultural intensification, with field margins or strips of agricultural land taken out of production and left fallow ('set-aside') for a varying number of years. The scheme has resulted in reported positive impacts on above ground macro-diversity including birds, pollinators and plants (Pywell *et al.*, 2012; Baker *et al.*, 2012). However, the impact of this land use change for belowground microbial biodiversity has yet to be elucidated.

Intensive land management can compromise soil fertility and quality, which can have significant effects on crop health (Matson, 1997). For example

shortened crop rotations and crops grown in monocultures can impact disease and pathogen build up which can result in yield decline (Choudhary *et al.*, 2011; Hilton *et al.*, 2013). Of the 14 crops which constitute the majority of human food resources, all are susceptible to infection by bacteria, fungi, protozoa, and viruses as a result of agricultural intensification (Strange and Scott, 2005). Conversely, other microorganisms provide a multitude of benefits to plants through symbioses that result in increased nutrient availability, provision of disease resistance or enhancement of stress tolerance. Arbuscular mycorrhizal fungi (AMF) for example, form symbioses with over 80% of plant species, including most crop species, in which photosynthetic assimilates are traded for nutrients acquired by the AMF from the soil (Morrissey *et al.*, 2004). Given the important roles soil microorganisms perform in determining crop productivity and maintenance of aboveground biodiversity there is an urgent need to understand the spatial distribution and scaling of soil microbial biodiversity and distribution patterns.

From a fundamental perspective, it is now widely accepted that many microorganisms are not ubiquitous in their spatial distribution and can demonstrate significant turnover in β-diversity at different spatial scales (Hanson *et al.*, 2012; Martiny *et al.*, 2006). Patterns of microbial community composition and diversity are shaped by the processes of dispersal, selection, drift, and evolution, also referred to as speciation, mutation, or diversification (Vellend, 2010; Hanson *et al.*, 2012; Nemergut *et al.*, 2011 respectively). The impact of these processes on microbial communities depends on factors such as scale, habitat type and microbial group. Studies assessing multiple scales have provided evidence that at intercontinental scales, dispersal (represented by geographical

distance effects) influences bacterial distributions (Cho and Tiedje, 2000) whereas at local scales, selection (represented by environment effects) can be the main predictor of spatial scaling (Horner-Devine *et al.*, 2004). At the regional scale both environment and geographical distance can be important in determining distribution patterns. For example Ranjard *et al.* (2013) showed that bacterial turnover was affected by both environment and geographical distance.

Patterns of microbial community distribution can also vary between habitats. Lake based studies have shown that bacterial β-diversity can be driven by dispersal or dispersal and selection (Reche *et al.*, 2005; Yannarell and Triplett, 2005). In hot spring ecosystems, bacterial (Papke *et al.*, 2003) and archaeal (Whitaker *et al.*, 2003) distributions were driven by geographical distance only. However, bacterial communities in salt marsh environments were driven by environmental factors (Horner-Devine *et al.*, 2004) or environment and geographical distance (Martiny *et al.*, 2011). Soil based studies on bacteria have emphasised the importance of the environment, particularly pH, in driving microbial distributions (Griffiths *et al.*, 2011; Fierer and Jackson, 2006; Lauber *et al.*, 2008). Distribution of archaea in soil tend to be driven by environmental factors such as pH (Bates *et al.*, 2011) and C:N (Bru *et al.*, 2011). There is however strong evidence for the influence of both environment (soil temperature, soil type) and geographical distance on fungal distributions, (Dumbrell *et al.*, 2009; Green *et al.*, 2004; Kivlin *et al.*, 2011), while for soil protists, spatial scaling has been shown to be largely dependent on environmental factors such as soil moisture (Bates *et al.*, 2012).

Agricultural systems provide an important platform to investigate microbial sensitivity to anthropogenic changes (Herold *et al.*, 2014; Verbruggen

et al., 2010). Studies have shown that land use management (often conventional versus organic, tillage regimes and variation in grassland management methods) can impact bacterial (Lupwayi *et al.*, 1998; Jangid *et al.*, 2008) and fungal diversity and relative abundance (Sayer *et al.*, 2013; van der Gast *et al.*, 2011a). Other studies have shown that the variables associated with certain management practices are more significant in determining microbial distribution patterns than the land use type itself (Hazard *et al.*, 2013; Lauber *et al.*, 2008).

Our fundamental understanding of microbial biodiversity and distribution has benefitted greatly over the last decade with an ever-increasing number of studies addressing microbial spatial distribution at different scales, across a diverse array of habitat types, and usually focused on particular microbial groupings in each instance. This diversity of studies and conclusions could act as an unintended confounding factor, with evidence for geographical, environmental drivers (either environment in general or different specific environmental factors), or both, being generalised from non-comparable experimental systems. Indeed, the task of determining whether there are consistent drivers of microbial diversity and distribution patterns is further confounded when independent studies of the same habitat type and microbial groupings report conflicting evidence for the underlying drivers of the spatial distribution patterns observed (Hanson et al., 2012). However, contrasting drivers of microbial distribution patterns between studies could reflect very different life histories and dispersal mechanisms for distinct microbial groups. Studies of multiple microbial groupings in single study systems are needed to address these issues.

In the current study we investigated drivers of spatial scaling at the

community level for multiple microbial groupings, covering all three domains of life, sampled from multiple farm sites. These consisted of paired arable fields and AES set-aside margins, with a maximum separation of 390 km, on a range of soil types and latitudes in the UK. The study aimed to determine the impact of land use change, environment and geographical distance on the spatial scaling of diverse microorganism groups in soil. Furthermore, since set-aside locations ranged between 6-17 years from conversion, the influence of time since land use change, on microbial distribution patterns, could be investigated. We hypothesised that (1) there is no universal driver of microbial spatial scaling in different taxa groupings of microorganisms within and between the different land uses and (2) that land use management will not have discernible effects on α and β-diversity for any of the microbial groupings.

2.2. Materials and Methods

Experimental design, sampling, DNA extraction was carried out by Paul Gosling, Molecular lab work was carried out by Steffen Jost, Hanna Hartikainen, Shazia Mahamdallie and Michelle Gardner. Sequencing was funded by an FWF grant (Austrian Science Fund), awarded to Jens Boenigk and carried out by Eurofins (Germany).

2.2.1. Location

17 sites were selected (from 15 locations) across England, with a maximum separation of 390 km. Sites were located within the farms of a range of organisations, including agricultural colleges, research institutes, ADAS, RSBP and The Game Conservancy Trust, in addition to commercial farms. The sites

were on clay, clay loam, silt, sandy loam and loam soil types. Set-aside locations occurred either within cropped arable fields or were adjacent to cropped arable fields, and ranged between six and 17 years in age since conversion from arable fields (Figure 2.1).

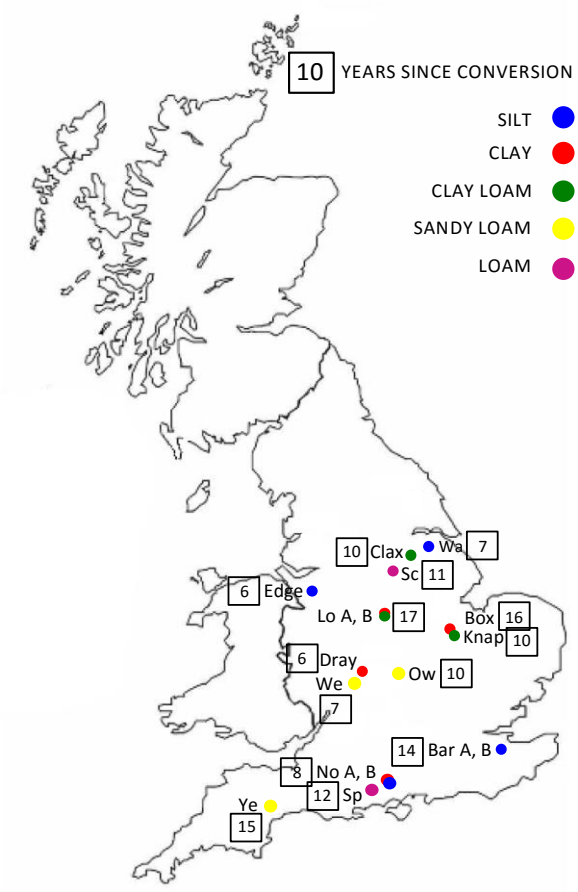


Figure 2.1 Map showing 15 site locations that were sampled across England including soil type and age since conversion from set-aside (shown in box). Bar A, B, Barfrestone A and Barfrestone B; Bo, Boxworth; Clax, Claxby Moor; Dray, Drayton; Edge, Edgmond; Knap, Knapwell; Lo A, B, Loddington A and Loddington B; No A, B, Northington A and Northington B; Ow, Old Warden A; Sc, Scartho; Sp, Sparsholt; Wa, Waddingham; We, Wellesbourne; Ye, Yettington A. In depth site details are provided in Table 2.1)

2.2.2. Sampling

At each field location, five 10 x 10 metre plots within both the set-aside and the arable area were assigned. For between-field comparisons, plots were established

in the centre of the adjacent arable and set-aside fields, while for within-field comparisons a 10 m partition separated the set-aside and arable plots. Using a 5 cm diameter auger, 20 x 0-10 cm depth soil cores were taken from within each plot between March and April 2008. The soil cores from within each plot were pooled for 454 pyrosequencing and for analysis of soil pH, total carbon and nitrogen, total and available phosphorous (Olsen P), mineral N, potassium, magnesium and rainfall; as detailed in Gosling (2014). Time since conversion from arable to set-aside was detailed at each location (Table 2.1).

Table 2.1 Soil characteristics, land use treatment and site details. Site names with a '1' denote within-field comparisons, site names with a '2' denote between-field comparisons.

Site	Land Use	pH	NO3 (ug/g)	K (ug/g)	Mg (ug/g)	Olsen P (ug/g)	Total P (ug/g)	%C	%N	Rainfall (mm)
[1]Barfrestone A	Arable	7.5	86.8	274.4	41.5	61.5	1241.5	3.6	0.5	720
	Set-aside	7.7	18.0	211.6	56.4	22.8	882.1	3.7	0.5	720
[1]Barfrestone B	Arable	7.4	189.2	343.6	60.9	101.2	1141.7	3.9	0.6	720
	Set-aside	7.8	12.3	255.6	71.8	31.1	703.6	3.8	0.5	720
[2]Boxworth	Arable	7.3	62.2	190.3	92.3	25.9	629.3	2.2	0.2	550
	Set-aside	6.6	14.5	292.1	170.6	37.6	700.1	3.1	0.3	550
[1]Claxby Moor	Arable	6.7	31.3	141.2	66.0	18.4	288.1	1.3	0.1	530
	Set-aside	7.1	4.9	92.3	72.4	11.2	284.5	1.7	0.1	530
[2]Drayton	Arable	7.6	4.4	293.7	102.4	18.7	620.7	2.6	0.3	625
	Set-aside	6.7	8.7	411.8	313.0	26.1	562.5	3.5	0.3	625
[1]Edgmond B	Arable	6.2	7.6	115.7	126.4	19.4	357.9	1.1	0.1	655
	Set-aside	5.9	6.6	127.9	169.6	15.0	357.8	1.4	0.1	655
[1]Knapwell	Arable	6.2	19.3	321.4	114.2	46.6	834.7	4.1	0.5	550
	Set-aside	7.1	19.4	372.1	98.2	37.9	721.0	4.5	0.4	550
[1]Loddington A	Arable	7.0	66.3	236.5	132.0	20.3	844.5	3.7	0.5	660
	Set-aside	7.3	25.6	341.5	169.8	17.9	811.5	4.8	0.6	660
[1]Loddington B	Arable	5.9	99.3	208.5	196.1	33.7	1269.5	2.9	0.4	660
	Set-aside	5.7	18.7	334.3	238.3	44.6	1236.1	3.2	0.4	660
[1]Northington A	Arable	7.8	26.8	265.6	41.7	31.9	1216.0	3.0	0.3	800
	Set-aside	8.0	13.3	216.4	36.4	31.4	615.0	2.8	0.3	800
[1]Northington B	Arable	7.8	20.8	229.4	68.9	27.2	1116.3	2.7	0.4	800
	Set-aside	7.8	16.8	430.2	60.7	51.5	1097.0	3.4	0.5	800
[2]Old Warden A	Arable	7.1	20.4	250.9	99.3	51.4	717.8	2.3	0.2	584
	Set-aside	7.2	27.5	274.4	78.6	67.1	783.6	2.6	0.2	584
[2]Scartho	Arable	5.5	59.7	199.1	157.9	21.3	488.8	1.9	0.2	565
	Set-aside	5.4	7.0	178.1	103.1	18.2	433.8	2.2	0.2	565
[1]Sparsholt	Arable	7.6	89.3	190.4	54.0	72.5	1167.4	3.8	0.4	800
	Set-aside	7.8	29.1	332.7	56.6	58.5	1177.6	3.6	0.5	800
[2]Waddingham	Arable	7.8	30.0	120.3	44.8	27.1	575.8	2.7	0.3	533
	Set-aside	7.9	7.1	238.3	40.5	35.9	644.2	2.5	0.3	533
[1]Wellesbourne	Arable	6.7	3.1	160.0	67.1	59.5	688.7	0.8	0.1	625
	Set-aside	6.4	4.6	197.2	91.6	50.4	612.4	1.3	0.1	625
[1]Yettington A	Arable	5.9	48.1	207.0	57.3	120.8	616.2	1.2	0.1	972
	Set-aside	5.4	28.1	226.9	65.2	111.6	695.9	1.6	0.2	972

2.2.3. Pyrosequencing

DNA was extracted from each plot and then pooled for analysis using the Qiagen DNeasy Blood and Soil kit (Qiagen, Hilden, Germany) following the manufacturer's guidelines. PCR amplifications of DNA were conducted using four different primer sets, which variously targeted the V4 and V9 region of the 18S and 16S rRNA gene. The primer set 1087Fmod and 1492r (Roesch *et al.*, 2007) were used to amplify bacterial DNA. Primers 1391F (Lane, 1991; Stoeck *et al.*, 2010) and Euk B (Medlin *et al.*, 1998; Stoeck *et al.*, 2010) were used to amplify a broad spectrum of eukaryotes. A semi-nested PCR approach was used for Cercozoa. The first round primers were 25F and 1256R; the second round used the same reverse primer with 3NDf (Bass *et al.*, 2009). Nested PCR was also used Nematodes using primers Nem_18S_F and Nem_18S_R (Floyd et al., 2005).

The final concentrations in all of the PCR reactions were: 1 µl of DNA template 20 µl PCR reaction with 0.4 units of Phusion polymerase, primers at 0.25 µM final concentration, and dNTPs at 0.2 mM final concentration, including 4 µl Phusion buffer and 12.2 µl water. The PCR conditions consisted of an initial denaturation at 94 °C for 4 min and 35 cycles of: 30 s at 95 °C, annealing for 60 sec at 53 °C, elongation for 2 min at 72 °C followed by a final extension step of 10 min at 72 °C for broad bacterial primers and nematode primers; 30 s at 95 °C, annealing for 60 sec at 60 °C, elongation for 2 min at 72 °C followed by a final extension step of 10 min at 72 °C for broad eukaryote primers; 30 s at 95 °C, annealing for 60 sec at 70 °C, elongation for 2 min at 72 °C followed by a final extension step of 10 min at 72 °C for broad cercozoan primers. Pyrosequencing was carried out using the 454 Genome Sequencer FLX System (454 Life Science

Branford, CT, USA).

2.2.4. Sequence processing

Raw sequences were processed within the Quantitative Insights Into Microbial Ecology (QIIME) v. 1.7.0 pipeline based on the default settings (Caporaso *et al.*, 2010a). Modifications to the default settings included setting the maximum number of barcode errors and maximum number of ambiguous bases to zero (default is 1.5 and 6 respectively) in order to increase the stringency of the filtering process. The minimum sequence length was reduced to 100 bp to take into account the shorter eukaryote reads (default is 200bp). Sequences were denoised using 'Denoiser' (Reeder and Knight, 2010) and clustered into phylotypes at 97% similarity using Uclust (Edgar, 2010). Sequence read lengths varied from an average of 160 bp for eukaryotes, 350 bp for Cercozoa and 400 bp for bacteria and Nematodes. The 16S rRNA gene sequences were aligned to the Greengenes Core reference alignment, Feb. 4 2011 version (DeSantis *et al.*, 2006). The 18S rRNA gene sequences were aligned to the SILVA database, release 108 (http://www.arb-silva.de/), both of which utilised PyNAST (Caporaso *et al.*, 2010b). ChimeraSlayer (Haas *et al.*, 2011) was used to remove chimeric sequences based on reference alignments. Taxonomy was assigned using BLAST (Altschul *et al.*, 1990) against Greengenes February 4th 2011 version (McDonald *et al.*, 2011) and SILVA taxonomy and reference databases (version 108). Finally OTU (operational taxonomic unit) tables were constructed. After removal of "no blast hits" (based on a maximum BLAST e-value of 0.001) and non-target sequences, the total number of OTUs (determined at 97% similarity) and the total number of sequences per sample, (n= 34, all sites), across all sites for each taxon were calculated (Table 2.2).

Table 2.2 Total number of sequences and operational taxonomic units (OTUs) for each taxonomic group, across all 34 samples. NB. Individual taxa group totals will not add up to the total for the bacteria domain, as not all taxa groups were included from the sequencing data.

Domain	Taxa	Total Sequences	Total OTUs
Bacteria	Bacteria (Total)	158886	17176
	Acidobacteria	41440	3688
	Actinobacteria	54570	1605
	Alphaproteobacteria	11929	1482
	Bacteroidetes	13328	1248
	Betaproteobacteria	18054	1289
	Deltaproteobacteria	10586	1881
	Firmicutes	18039	1432
	Gammaproteobacteria	10212	1649
	Proteobacteria	50069	6308
	Verrucmicrobia	11452	1010
Archaea	Archaea	5123	138
	Crenarchaea	3255	22
	Euryarchaea	1868	116
Eukyarotes	Ascomycota	41352	722
	Basidiomycota	35063	565
	Glomeromycota	1528	86
	Rhizaria	29295	607
	Cercozoa	3188	335
	Stramenopiles	10115	318
	Nematoda	66885	1524

From the four primer sets, 21 microbial taxa groupings were chosen for statistical analysis. Actinobacteria, Acidobacteria, Proteobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Verrucmicrobia, Firmicutes, Bacteroidetes and total bacteria were taken from the bacterial primers. From the eukaryote primers, Basidiomycota, Ascomycota, Glomeromycota, Stramenopiles, Rhizaria, Archaea, Crenarchaea and Euryarchaea were chosen. Finally, Nematode and Cercozoa sequences were derived from the Nematode and Cercozoa specific primers. The mean number of sequences per sample ranged from 1967 sequences

for Nematodes to an average of 392 sequences per sample for Bacteroidetes (Table 2.3).

Table 2.3 Each taxa group and corresponding primer set

Taxon	Primer Set
Bacteria	Bacteria
Acidobacteria	Bacteria
Actinobacteria	Bacteria
Alphaproteobacteria	Bacteria
Bacteroidetes	Bacteria
Betaproteobacteria	Bacteria
Deltaproteobacteria	Bacteria
Firmicutes	Bacteria
Gammaproteobacteria	Bacteria
Proteobacteria	Bacteria
Verrucomicrobia	Bacteria
Archaea	General Eukaryote
Crenarchaea	General Eukaryote
Euryarchaea	General Eukaryote
Ascomycota	General Eukaryote
Basidiomycota	General Eukaryote
Glomeromycota	General Eukaryote
Rhizaria	General Eukaryote
Stramenopile	General Eukaryote
Nematode	Nematode
Cercozoa	Cercozoa

2.2.5. Statistical analyses of the data

Distance-decay relationships were used to assess community similarity with increasing distance and to provide information on β-diversity at the regional scale (van der Gast *et al.*, 2011a). The rate of decay in similarity (*d*) was calculated based on the Bray-Curtis index of dissimilarity (1- dissimilarity converted the data to similarities for visual purposes) and a power law equation fitted to the curves (as detailed in Magurran, 2003). Distance-decay relationships were investigated for each microbial grouping, across each land use practice, and with both land uses combined.

Mantels and partial Mantels (Mantel, 1967) were performed to correlate β-diversity with soil parameters and distance. For each taxa grouping, dissimilarity matrices (based on the Bray-Curtis index) were constructed using the Vegan package and function 'vegdist' in R v2.15.2 (R Development Core Team, 2010). A distance matrix based on pairwise Euclidan distances was generated from the grid coordinates and environmental matrices were produced using the Euclidian and Gower index to account for categorical variables. For instance for soil type, a zero was used when the values were the same and a one when they were different. The matrices were produced using the R package 'Cluster' and function 'daisy'. Relatedness between these matrices was determined using Pearson's correlation coefficient (r), based on 999 permutations, (P = 0.05). The preliminary Mantel results were used to inform partial Mantels for each of the microbial groupings. This enabled the correlation between species composition and the significant environmental variables to be determined whilst accounting for geographical distance, and vice versa. The data were partitioned into arable and set-aside treatments and the analyses re-run to remove any treatment effects.

Additionally, a complementary direct ordination approach, canonical correspondence analysis (CCA) and principal coordinates of neighbour matrices (PCNM), which uses proportional OTU abundance to divide β-diversity into spatial and environmental components was performed (Borcard and Legendre, 2002). PCNM produces spatial predictors in the form of eigenvectors, derived from positive eigenvalues (Borcard *et al.*, 2004) and allows for the detection and quantification of spatial patterns over multiple scales (Sweetman *et al.*, 2010). PCNM analysis was performed using the 'PCNM' function from the PCNM

package in R (R development Core Team, Vienna, Austria). PCNM vectors were retained from positive eigenvalues and used in a stepwise procedure (based on 999 permutations; $\alpha < 0.05$), which incorporated the 'Forward Select' function within CANOCO for Windows v5.03 (Ter Braak, 1988). Variation in community composition was determined using significant environmental and spatial variables in a canonical correspondence analysis (CCA). Only when both PCNM and environmental variables significantly correlated with the species data (incorporating distance as a covariable) were partial CCAs performed. In the case where only environmental variables were significant, no partial CCAs were run. Where a significant linear trend occurred between distance and species data, the 'var-part' function was implemented in CANOCO, which detrended the data and partitioned the variance accordingly. The CCA method applied to the data was derived from Hazard *et al.*, (2013) and is shown in more detail in Table 2.4.

Table 2.4 Calculations used to derive values for overall Canonical Correspondence analysis (CCA), using methods detailed in Hazard *et al., (*2012)

% of Variation	Calculation
Environment	Environment-PCNM (sum all canonical eigenvalues pCCA)/total inertia * 100
PCNM	% Environmental variance - % Environmental only
Environment + PCNM	PCNM% - % overlap
% Environmental Variance	Inertia Environmental (sum all canonical eigenvalues, environmental CCA) / total inertia*100
PCNM%	Amount explained by PCNM/total inertia*100
Amount explained by PCNM	Total inertia - sum all eigenvalues (pCCA)

Meta-analyses were used to summarise the *Hedges' d* effect size of each treatment for each taxa grouping (Borenstein *et al.*, 2009). The effect size calculates the standardized mean difference between treatments whilst taking into account the sample variance and the standard error of the mean (Rogers *et*

al., 2013a). Samples were rarefied, whereby each dataset was randomly subsampled to the lowest sequence number per sample, using a custom R script (Gihring *et al.*, 2012; Rogers *et al.*, 2013a) (Figure 2.2). The script calculated three complementary diversity indices and their associated standard deviations (based on 1000 iterations) to characterise the microbial communities. These included richness (*S**), reciprocal of Simpson's index (*1/D*) and Shannon-Weiner index (*H*). *S** calculates the total 'species' in a community (Pommier *et al.*, 2010), *1/D* considers 'species' count, evenness and sampling effort and *H* considers abundance and evenness (McCaig *et al.*, 1999; Zhou *et al.*, 2002; Koid *et al.*, 2012). Differences in the indices were computed using the method of Solow (1993).

```
setwd("~/Desktop/Spatial_analyses/testing_site_removal/Diversity_again/")
data1<-read.csv("Verruc_Dall_rerun.csv")

library(vegan)

s_num<- 3666#number you want to resample i.e. what is the lowest number of sequense in you samples


# run the resamp function
resamp<-function(x){
 sample_x<-sample(x,s_num,replace=TRUE)#check if you want replacement or not
 sample_y<-table(sample_x)
 samp_richness<-length(unique(sample_x)) # to give species richness
 D<-diversity(sample_y,index="invsimpson") # function in vegan to give diversity indices
 H<-diversity(sample_y,index="shannon") # can do simpsons and shannons
 out<-c(samp_richness, D, H)  # this sets the outputs so it gives them in the correct order and format
 names(out)<-c("Richness", "invSimpson", "Shannon")
 return(out)
}

# make a matrix for the output, the number of columns can be edited for different datasets
diversity<-matrix(nrow=3, ncol=15)
colnames(diversity)<-c("Mean_Richness", "Mean_invSimpson", "Mean_Shannon","Med_Richness",
                "Med_invSimpson", "Med_Shannon","2.5_Richness","97.5_Richness",
                "2.5_invSimpson", "97.5_Simpson","2.5_Shannon","97.5_Shannon",
                "Rich_SD","invSimp_SD","Shan_SD")

## Now run the loop for all the samples in the dataset
#Median is used as it is not so sensative to skews in the data
#The data can be checked for skews


for (i in (seq(from=1,to=4,by=2))){
 x<-rep(na.omit(data1[,i]),na.omit(data1[,i+1]))
 rep1000<-replicate(1000, resamp(x)) # the you run the resamp function 1000 times
 MeanDiv<-rowMeans(rep1000) # this gives you the mean of each diversity measure
 Median_rich<-median(rep1000[1,])
 Median_D<-median(rep1000[2,])
 Median_H<-median(rep1000[3,])
 Rich_95<-quantile(rep1000[1,],c(.025,0.975))
 simp_95<-quantile(rep1000[2,],c(0.025,0.975))
 shan_95<-quantile(rep1000[3,],c(0.025,0.975))
 RichStDEV<-sqrt(var(rep1000[1,]))
 simpStDEV<-sqrt(var(rep1000[2,]))
 shanStDEV<-sqrt(var(rep1000[3,]))
 diversity[i,]<-as.numeric(c(MeanDiv,Median_rich,Median_D, Median_H,
                  Rich_95, simp_95, shan_95, RichStDEV, simpStDEV, shanStDEV)) # these are the Column names in the output
}

diversity2<-diversity[-2,] ## remove the NA rows
diversity2

write.csv(diversity2,"VerrucDall_rerun.csv")
```

Figure 2.2 Custom R script designed to re-sample the data based on a minimum sequence number, based on 1000 iterations. Species richness S*, Shannon-Weiner index (*H*) and reciprocal of Simpson's index (*1/D*) were calculated for each site/sample.

2.3. Results

2.3.1. Distance decay relationship

Distance-decay analysis showed that there was a strong negative relationship between community similarity and geographic distance in a range of taxa across all domains (Table 2.5 and Figure 2.3). Significant relationships were shown in bacteria as a whole ($P = 0.03$, $d = -0.01$), Actinobacteria ($P = 0.04$, $d = -0.02$), Archaea as a whole ($P = <0.001$, $d = -0.06$) and Crenarchaea ($P = <0.001$, $d = -0.06$) (Table 2.5). A significant distance-decay effect was also demonstrated for

three of the six eukaryote taxa groupings including Glomeromycota ($P = 0.01$, $d = -0.08$), Basidiomycota ($P = 0.00$, $d = -0.03$), and Stramenopiles ($P = 0.01$, $d = -0.02$).

Table 2.5 Distance-decay summary statistics. N is the number of samples, n is the number of pairwise comparisons. The power law regression statistics include the rate of decay in similarity (d), the coefficient of determination (r^2) and significance (P). The asterisk denotes a significant distance-decay relationship (P=<0.05).

	Taxa	N	n	d	r^2 (%)	P
Bacteria	Bacteria Combined	30	435	-0.01	1.6	0.03*
	Set-aside	15	105	-0.05	10.6	0.00*
	Arable	15	105	0.03	2.2	0.19
	Acidobacteria Combined	30	435	-0.04	0.1	0.58
	Set-aside	15	105	-0.04	7.0	0.02*
	Arable	15	105	0.01	0.1	0.81
	Actinobacteria Combined	30	435	-0.02	1.3	0.04*
	Set-aside	15	105	-0.05	7.9	0.01*
	Arable	15	105	-0.02	0.3	0.63
	Alphaproteobacteria Combined	30	435	-0.01	0.8	0.10
	Set-aside	15	105	-0.04	8.5	0.01*
	Arable	15	105	-0.04	3.2	0.12
	Bacteroidetes Combined	30	435	-0.62	0.6	0.19
	Set-aside	15	105	-0.03	7.9	0.01*
	Arable	15	105	-0.02	0.9	0.42
	Betaproteobacteria Combined	30	435	-0.56	0.5	0.22
	Set-aside	15	105	-0.04	5.2	0.05*
	Arable	15	105	-0.01	0.1	0.75
	Deltaproteobacteria Combined	30	435	-0.52	0.4	0.24
	Set-aside	15	105	-0.06	13.7	0.00*
	Arable	15	105	-0.07	7.0	0.02*
	Firmicutes Combined	30	435	-0.38	0.2	0.45
	Set-aside	15	105	-0.01	0.2	0.69
	Arable	15	105	0.01	0.1	0.81
	Gammaproteobacteria Combined	30	435	-0.01	0.8	0.11
	Set-aside	15	105	-0.03	7.4	0.02*
	Arable	15	105	-0.02	0.8	0.43
	Proteobacteria Combined	30	435	-0.01	0.9	0.09
	Set-aside	15	105	-0.04	9.4	0.01*
	Arable	15	105	-0.03	1.7	0.25
	Verrucomicrobia Combined	30	435	-0.02	0.8	0.11
	Set-aside	15	105	-0.07	4.8	0.06
	Arable	15	105	-0.02	0.2	0.73
Archaea	Archaea Combined	32	496	-0.06	5.8	0.00*
	Set-aside	15	105	-0.05	0.6	0.49
	Arable	17	136	-0.10	3.3	0.07
	Euryarchaea Combined	25	300	-0.02	0.2	0.51

	Set-aside	13	78	-0.05	0.2	0.78
	Arable	12	66	-0.08	4.3	0.12
	Crenarchaea Combined	31	465	-0.06	6.0	<0.001*
	Set-aside	14	91	0.01	0.1	0.72
	Arable	17	136	-0.07	2.3	0.13
Fungi	Glomeromycota Combined	31	465	-0.08	1.9	0.01*
	Set-aside	17	136	-0.12	6.4	0.01*
	Arable	14	91	-0.32	4.2	0.10
	Ascomycota Combined	34	561	-0.01	0.2	0.41
	Set-aside	17	136	0.01	0.1	0.75
	Arable	17	136	0.00	0.0	0.95
	Basidiomycota Combined	34	561	-0.03	2.4	0.00*
	Set-aside	17	136	-0.06	4.1	0.05*
	Arable	17	136	-0.02	0.3	0.61
Protists	Rhizaria Combined	34	561	-0.01	0.6	1.07
	Set-aside	17	136	-0.01	0.2	0.67
	Arable	17	136	-0.01	0.3	0.62
	Stramenopiles Combined	34	561	-0.02	1.8	0.01*
	Set-aside	17	136	-0.02	0.7	0.41
	Arable	17	136	-0.02	0.9	0.35
	Cercozoa Combined	34	561	-0.01	0.6	0.11
	Set-aside	17	136	0.00	0.0	1.00
	Arable	17	136	-0.02	0.4	0.54
Animalia	Nematodes Combined	33	528	-0.17	0.3	0.31
	Set-aside	16	120	-0.10	3.8	0.08
	Arable	17	136	-0.03	0.2	0.69

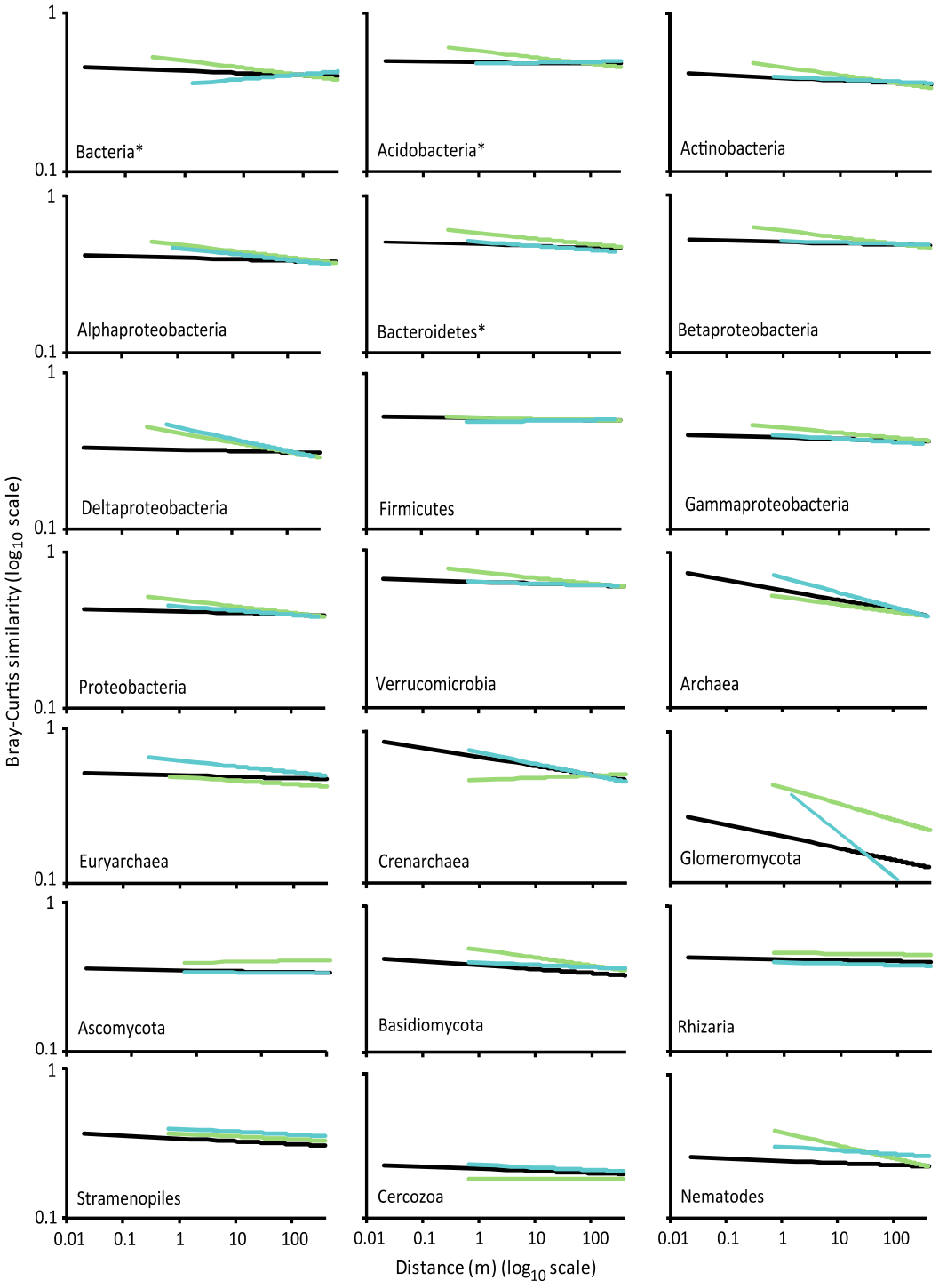


Figure 2.3 Distance decay plots for each taxa grouping showing community dissimilarity (shown as 1-Bray-Curtis index for visual purposes) across combined (solid black line), set-aside (green line) and arable (blue line) land management. Pairwise dissimilarities are plotted as a function of distance between the sites. T-distribution method (Fowler *et al.*, 1998) was used to calculate the slope of the distance-decay relationship for the two management practices. Significant differences between set-aside and arable land practice are denoted with an asterisk. Summary statistics are listed in Table 2.5.

Distance-decay analyses showed that under set-aside management, all bacterial taxa groupings showed a significant distance-decay relationship, with the exception of Firmicutes (P = 0.69, *d* = -0.01) and Verrucomicrobia (P = 0.06, *d* = -0.07). For eukaryotes, two of the three fungal taxa groupings showed significant distance-decay relationships under set-aside including Glomeromycota (P = 0.01, *d* = -0.12) and Basidiomycota (P = 0.05, *d* = -0.06), in which community similarity decreased with increasing distance (Table 2.5). Under arable land use management however, significant distance-decay effects were only present in Deltaproteobacteria (P = 0.02, *d* = -0.07). There was no significant relationship between community similarity and geographic distance for any other eukaryote or prokaryote group. Therefore land use management appeared to have a major impact on the composition of the communities when considering distance. Using the t-distribution method (Fowler *et al.,* 1998), the slopes of the distance-decay relationships for the two management practices (set-aside and arable), were found to be significantly different in the bacterial domain only (Figure 2.3). There was no significant difference between the slopes of the distance-decay relationships under arable and set-aside management for any Eukaryote or Archaea taxa grouping. The slopes of the distance-decay relationship were significantly different under arable and set-aside management in total bacteria, (arable; *d* = 0.03; set-aside; *d* = -0.05; t = 3.19; d.f. = 152; P = <0.001), Acidobacteria (arable; *d* = 0.01; set-aside; *d* = -0.04; t = 2.14; d.f. = 152; P = 0.03) and Bacteroidetes (arable; *d* = -0.02; set-aside; *d* = -0.03; t = 1.97; d.f. = 152; P = 0.05) (Table 2.5 for *d* values, Table 2.6 for T-statistic values).

Table 2.6: T-distribution test statistics comparing the regression slopes for the arable and set-aside land use, based on the method detailed in Fowler *et al.*, (1998).

	P	D.F	T statistic
Bacteria	0.00	152.00	3.19
Acidobacteria	0.03	152.00	2.14
Actinobacteria	0.06	152.00	1.88
Alphaproteobacteria	0.20	152.00	1.29
Bacteroidetes	0.05	152.00	1.97
Betaproteobacteria	0.15	152.00	1.45
Deltaproteobacteria	0.08	152.00	1.75
Firmicutes	0.64	152.00	0.47
Gammaproteobacteria	0.07	152.00	1.83
Proteobacteria	0.10	152.00	1.66
Verrucomicrobia	0.16	152.00	1.43
Archaea	0.30	175.00	1.03
Euryarchaea	0.18	100.00	1.36
Crenarchaea	0.22	198.00	1.23
Glomeromycota	0.07	161.00	1.85
Ascomycota	0.77	198.00	0.29
Basidiomycota	0.26	198.00	1.13
Rhizaria	0.91	198.00	0.12
Stramenopiles	0.87	198.00	0.17
Cercozoa	0.65	198.00	0.46
Nematodes	0.21	182.00	1.26

2.3.2. Environment and geographical distance effects

When assessing the combined data (both set-aside and arable land use practice), we found that environment was the key driver of all prokaryote and eukaryote community distribution (Table 2.7).

Table 2.7 Partial Mantel summary statistics for the correlation between taxa structure and environment and geographical distance. The partial Mantel statistic r(SG.E) estimates the correlation between S (community similarity) and geographical distance (G), whilst controlling for the effects of environment (E). Conversely r(SE.G) estimates the correlation between S and E whilst controlling for G. P determines whether the regression coefficients were significantly different from zero (based on 999 permutations). Significant correlations are indicated with an asterisk.

		r(SG.E)		r(SE.G)	
		r	p	r	p
Bacteria	Bacteria Combined	0.00	0.49	0.27	0.00*
	Set aside	0.28	0.029*	0.02	0.47
	Arable	-0.05	0.56	0.22	0.01*
	Acidobacteria Combined	0.11	0.12	0.25	0.00*
	Set aside	0.18	0.14	0.67	0.00*
	Arable	-	-	-	-
	Actinobacteria Combined	0.03	0.33	0.24	0.00*
	Set aside	0.15	0.17	0.16	0.13
	Arable	0.14	0.19	0.28	0.01*
	Alphaproteobacteria Combined	0.04	0.31	0.27	0.00*
	Set aside	0.08	0.27	0.16	0.11
	Arable	-	-	-	-
	Bacteroidetes Combined	0.10	0.16	0.12	0.01*
	Set aside	0.28	0.036*	0.00	0.48
	Arable	-0.04	0.55	0.16	0.05*
	Betaproteobacteria Combined	-0.01	0.50	0.26	0.00*
	Set aside	0.05	0.38	0.15	0.15
	Arable	-	-	-	-
	Deltaproteobacteria Combined	0.07	0.12	0.22	0.00*
	Set aside	0.09	0.25	0.48	0.00*
	Arable	-0.02	0.52	0.17	0.05*
	Firmicutes Combined	-0.08	0.79	0.11	0.02*
	Set aside	-0.05	0.56	0.19	0.05*
	Arable	-	-	-	-
	Gammaproteobacteria Combined	0.01	0.47	0.31	0.00*
	Set aside	0.16	0.15	0.10	0.23
	Arable	-0.12	0.79	0.43	0.00*
	Proteobacteria Combined	-0.01	0.55	0.24	0.00*
	Set aside	0.12	0.22	0.17	0.12
	Arable	-	-	-	-
	Verrucomicrobia Combined	0.05	0.26	0.18	0.00*
	Set aside	0.22	0.07	0.50	0.00*
	Arable	0.04	0.41	0.21	0.02*
Archaea	Archaea Combined	-0.09	0.89	0.52	0.00*
	Set aside	0.10	0.21	0.44	0.00*
	Arable	0.05	0.32	0.53	0.00*
	Euryarchaea Combined	-0.05	0.68	0.48	0.00*
	Set aside	0.04	0.36	0.67	0.01*

	Arable	-0.18	0.86	0.28	0.03*
	Crenarchaea Combined	-0.06	0.88	0.53	0.00*
	Set aside	0.15	0.09	0.46	0.00*
	Arable	0.06	0.23	0.27	0.01*
Fungi	Glomeromycota Combined	0.12	0.08	0.21	0.00*
	Set aside	0.13	0.17	0.17	0.11
	Arable	0.08	0.30	0.06	0.28
	Ascomycota Combined	-0.08	0.90	0.43	0.00*
	Set aside	-0.13	0.87	0.38	0.00*
	Arable	-0.12	0.90	0.42	0.00*
	Basidiomycota Combined	0.07	0.15	0.42	0.00*
	Set aside	0.17	0.04*	0.01	0.43
	Arable	-0.02	0.55	0.19	0.06
Protists	Rhizaria Combined	0.11	0.11	0.20	0.00*
	Set aside	-	-	-	-
	Arable	0.09	0.21	0.26	0.02*
	Stramenopiles Combined	0.95	0.10	0.56	0.00*
	Set aside	0.02	0.42	0.52	0.00*
	Arable	0.07	0.30	0.31	0.00*
	Cercozoa Combined	0.13	0.07	0.06	0.02*
	Set aside	-	-	-	-
	Arable	-	-	-	-
Animalia	Nematodes Combined	0.15	0.04*	0.18	0.00*
	Set aside	0.24	0.04*		
	Arable	-	-	-	-

With the exception of Nematodes, geographical distance did not correlate with the β-diversity of any taxa (when accounting for environment) (Table 2.8). The spatial scaling of Actinobacteria, Acidobacteria, Verrucomicrobia, Proteobacteria, Alphaproteobacteria and Betaproteobacteria β-diversity ($P = 0.003$, $r = 0.25$; $P = 0.01$, $r = 0.23$; $P = 0.01$, $r = 0.17$; $P = 0.001$, $r = 0.28$; $P = 0.001$, $r = 0.27$; $P = 0.002$, $r = 0.26$, respectively) was driven solely by pH. Bacteroidetes, Deltaproteobacteria and total bacteria β-diversity however, significantly correlated with soil type only ($P = 0.01$, $r=0.12$; $P = 0.001$, $r = 0.18$, $P = 0.001$, $r = 0.20$ respectively). Variation in Firmicute β-diversity remained unexplained (Table 2.8).

Table 2.8 Partial Mantel analyses for the association between taxa community composition with geographical distance and environmental variables under combined conditions. Significant P values are denoted with an asterisk, after accounting for Bonferroni Correction ranging between 0.003 and 0.03, (r) estimates the correlation between each variable.

		Actinobacteria		Acidobacteria		Bacteroidetes		Verrucomicrobia		Firmicutes		Proteobacteria		Alphaproteobacteria		Betaproteobacteria		Deltaproteobacteria		Gammaproteobacteria		Bacteria	
Parameter	Controlling for	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance	0.25	0.003*	0.23	0.01*			0.17	0.01*			0.28	0.001*	0.27	0.001*	0.26	0.002*	0.11	0.06	0.20	0.01	0.16	0.02
C	Distance																					0.14	0.07
N	Distance																						
N03	Distance																						
Mg	Distance																						
K	Distance																						
OP	Distance																			0.26	0.01		
TP	Distance																						
Rainfall	Distance											0.07	0.23							0.19	0.03	0.06	0.26
Soil Type	Distance	0.12	0.02			0.12	0.018*	0.11	0.03	0.11	0.02	0.10	0.03					0.18	0.001*	0.11	0.01	0.20	0.001*
Land Use	Distance	0.01	0.26	0.03	0.14	0.01	0.26	0.06	0.09	-0.01	0.53	0.03	0.17	0.03	0.18	0.02	0.22	0.00	0.44	0.04	0.14		
Distance	pH	0.11	0.12	-0.02	0.58			0.04	0.28			0.05	0.25	0.04	0.31	-0.01	0.50	0.09	0.16	0.17	0.03	0.11	0.11
Distance	C																					0.14	0.06
Distance	N																					0.14	0.07
Distance	N03																					0.16	0.03
Distance	Mg																					0.16	0.05
Distance	K																					0.17	0.04
Distance	OP																			0.08	0.15	0.14	0.04
Distance	TP																					0.16	0.05
Distance	Rainfall											0.04	0.28							0.03	0.40	0.07	0.22
Distance	Soil Type	0.17	0.04			0.10	0.16	0.08	0.15	-0.08	0.79	0.13	0.07					0.11	0.12	0.22	0.003*	0.14	0.06
Distance	Land Use	0.20	0.02	0.06	0.28	0.09	0.21	0.09	0.12	-0.10	0.84	0.14	0.06	0.11	0.09	0.08	0.17	0.13	0.10	0.26	0.002*		

		Glomeromycota		Ascomycota		Basidiomycota		Stramenopiles		Rhizaria		Cercozoa		Nematodea		Archaea		Euryarchaea		Crenarchaea	
		r	P	r	P			r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance	0.29	0.001*	0.21	0.001*	0.28	0.001*	0.33	0.001*	0.1	0.08					0.62	0.001*	0.28	0.01	0.53	0.001*
C	Distance			0.14	0.01	0.24	0.002*									0.23	0.004*			0.31	0.001*
N	Distance	0.08	0.14	0.14	0.01	0.29	0.001*			0.11	0.05					0.19	0.002*			0.33	0.001*
N03	Distance	0.29	0.02	0.15	0.03																
Mg	Distance			0.18	0.01											0.36	0.001*	0.49	0.002*	0.36	0.001*
K	Distance																				
OP	Distance	0.22	0.02																		
TP	Distance					0.09	0.08									0.12	0.04				
Rainfall	Distance	0.01	0.45																	0.01	0.42
Soil Type	Distance			0.13	0.01	0.11	0.03									0.18	0.01	0.21	0.01	0.16	0.00
Land Use	Distance	0.25	0.001*	0.29	0.001*	0.238	0.002*	0.46	0.002*	0.14	0.01*	0.06	0.03*	0.18	0.004*	-0.02	0.69				
Distance	pH	0.12	0.07	-0.05	0.79	0.08	0.1	0.03	0.32	0.11	0.08					0.02	0.38	-0.07	0.77	0.04	0.22
Distance	C			-0.01	0.58	0.13	0.02									0.12	0.06			0.12	0.03
Distance	N	0.08	0.13	-0.02	0.62	0.1	0.04			0.12	0.09					0.11	0.05			0.11	0.04
Distance	N03	0.19	0.02	-0.01	0.55															0.17	0.00
Distance	Mg			0.02	0.39											0.18	0.01	0.03	0.38	0.15	0.01
Distance	K																			0.17	0.01
Distance	OP	0.1	0.09																	0.17	0.01
Distance	TP					0.14	0.02									0.12	0.05			0.12	0.02
Distance	Rainfall	0.15	0.06																	0.12	0.02
Distance	Soil Type			0	0.5	0.15	0.01									0.15	0.01	0.00	0.46	0.17	0.01
Distance	Land Use	0.28	0.002*	0.03	0.34	0.18	0.01	0.17	0.01*	0.15	0.05	0.13	0.07	0.15	0.04	0.19	0.01				

Total Archaea β-diversity correlated with pH (P = 0.001, r = 0.62), C (P = 0.004, r = 0.23), N (P = 0.002, r = 0.19) and Mg (P = 0.001, r = 0.36). Similarly, Euryarchaea β-diversity correlated with Mg (P = 0.002, r = 0.49). These environmental drivers were also significant in the β-diversity of Crenarchaea whereby pH, (P = 0.001, r = 0.53), C (P = 0.001, r = 0.31), N (P = 0.001, r = 0.33) and Mg (P = 0.001, r = 0.36) were significant (Table 2.8). For four of the seven eukaryote taxa, pH was also a significant driver of microbial distribution including Glomeromycota (P = 0.001, r = 0.29), Ascomycota (P = 0.001, r = 0.21), Basidiomycota (P = 0.001, r = 0.28) and Stramenopiles (P = 0.001, r = 0.33). Crucially, land use management significantly correlated with all seven eukaryote taxa groups investigated but was not shown to influence the spatial scaling of any prokaryote group (Table 2.8). CCA showed that environment significantly influenced three bacterial taxa groupings, all fungal taxa groupings, Stramenopiles and Euryarchaea. Geographical distance significantly influenced Verrucomicrobia, Cercozoa, Nematodes, Archaea and Crenarchaea. Although in most cases, the amount of variation explained by both environment and geographical distance was similar. Seven of the eleven bacteria taxa groupings along with Rhizaria remained unexplained by the model (Table 2.9).

Table 2.9 Canonical Correspondence Analysis for each taxa grouping. Values shown represent percentage contribution of PCNM (geographical distance), environmental variables, both PCNM and the environment combined. Undetermined is the amount of variation explained by the model.

	BACTERIA										
% of Variation	Actinobacteria	Acidobacteria	Bacteroidetes	Verrucomicrobia	Firmicutes	Proteobacteria	Alphaproteobacteria	Betaproteobacteria	Deltaproteobacteria	Gammaproteobacteria	Bacteria
Environment	-	3.60%	-	3.68%	7.65%	-	-	-	3.76%	-	-
PCNM	-	3.56%	-	12.58%	3.64%	-	-	-	3.50%	-	-
Environment + PCNM	-	0.64%	-	1.06%	0.56%	-	-	-	0.50%	-	-
Undetermined	100.00%	92.20%	100.00%	82.68%	88.15%	100.00%	100.00%	100.00%	92.24%	100.00%	100.00%

	FUNGI				PROTISTS		ANIMALIA	ARCHAEA		
% of Variation	Glomeromycota	Ascomycota	Basidiomycota	Stramenopiles	Rhizaria	Cercozoa	Nematodes	Archaea	Euryarchaea	Crenarchaea
Environment	17.22%	38.00%	32.00%	14.03%	-	6.69%	15.06%	13.94%	12.63%	19.36%
PCNM	16.25%	6.00%	20.88%	-	-	7.55%	17.53%	27.27%	9.95%	20.56%
Environment + PCNM	6.30%	4.00%	23.20%	-	-	0.89%	2.08%	15.42%	2.91%	20.43%
Undetermined	60.23%	56.00%	23.92%	85.97%	100.00%	85.76%	65.33%	43.37%	74.51%	39.65%

ANOVA results (as detailed in Defra (2011) Final report for IF0138) showed significant differences between set-aside and arable sites for all of the soil characteristics (with the exception of total N). For example NO_3 was 32.46 μg g^{-1} in arable soil compared to 12.09 μg g^{-1} in set-aside soil (significant $P < 0.001$) and pH was higher under set-aside conditions with a value of 7.05 compared to 7.0 under arable conditions (significant $P < 0.05$). Environment was therefore shown to be a significant driver of microbial distribution under both arable and set-aside conditions. Under set-aside conditions, environment correlated with taxa from all three domains, whilst geographical distance correlated with taxa groupings belonging to two of the three domains. Environment correlated with bacteria including Acidobacteria (P = <0.001, r = 0.67), Deltaproteobacteria (P = <0.001, r = 0.48), Firmicutes (P = 0.05, r = 0.19) and Verrucomicrobia (P = <0.001, r = 0.50). Environment also correlated with all Archaea taxa including total Archaea (P = <0.001, r = 0.44), Euryarchaea (P = 0.01, r = 0.67) and Crenarchaea (P = <0.001, r = 0.46) (Table 2.7). Within the eukaryotes, environment correlated with ascomycete fungi (P = <0.001, r = 0.38) and Stramenopiles (P = <0.001, r = 0.52). Geographic distance correlated with two eukaryote taxa, Basidiomycota (P = 0.04, r = 0.17) and Nematodes (P = 0.04, r = 0.24) and two bacterial taxa groupings, total bacteria (P = 0.029, r = 0.28) and Bacteroidetes (P = 0.036, r = 0.28). When investigating the relative importance of individual environmental factors, partial Mantels revealed that pH was a key predictor of distribution in all bacterial taxa groupings within the bacteria domain (excluding Firmicutes and Deltaproteobacteria), all of the fungal groupings including Glomeromycota (P = 0.001, r = 0.54), Ascomycota (P = 0.001, r = 0.38) and Basidiomycota (P = 0.002, r = 0.50) and the protist taxa Stramenopiles (P = 0.001, r = 0.52) in the

eukaryotes. pH also correlated with total Archaea ($P = 0.001$, $r = 0.71$) and Crenarchaea ($P = 0.001$, $r = 0.60$) in addition to Mg in Euryarchaea ($P = 0.009$, $r = 0.67$) and Crenarchaea ($P = 0.003$, $r = 0.44$), (Table 2.10). There was no correlation between the time since conversion from arable farms to set-aside and β-diversity for any taxa (data not shown).

Table 2.10: Partial Mantels analyses for the association between taxa community composition and both geographical distance and environmental variables under set-aside conditions. Significant P values are denoted with an asterisk, after accounting for Bonferroni Correction ranging between 0.003 and 0.03

Parameter	Controlled for	Actinobacteria		Acidobacteria		Bacteroidetes		Verrucomicrobia		Firmicutes		Proteobacteria		Alphaproteobacteria		Betaproteobacteria		Deltaproteobacteria		Gammaproteobacteria		Bacteria	
		r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance	0.62	0.001*	0.66	0.001*	0.51	0.001*	0.31	0.01*			0.48	0.003*	0.48	0.001*	0.40	0.007*	0.23	0.03	0.41	0.001*	0.51	0.001*
C	Distance	0.24	0.05																			0.18	0.09
N	Distance	0.32	0.01			0.10	0.18	0.25	0.02	0.19	0.05	0.12	0.11									0.26	0.02
N03	Distance																	0.20	0.05				
Mg	Distance																						
K	Distance																						
OP	Distance																	0.26	0.06	0.25	0.04		
TP	Distance																						
Rainfall	Distance	0.20	0.11			0.04	0.39					0.16	0.14	0.15	0.14	0.14	0.18			0.09	0.30	0.01	0.47
Soil Type	Distance																	0.23	0.02				
Distance	pH	0.13	0.18	0.19	0.14	0.22	0.09	0.09	0.25			0.13	0.16	0.09	0.24	0.05	0.34	0.31	0.03	0.18	0.09	0.23	0.04
Distance	C	0.34	0.02																			0.38	0.004
Distance	N	0.34	0.01			0.38	0.01	0.21	0.12	-0.05	0.56	0.30	0.02									0.38	0.001*
Distance	N03																	0.37	0.01			0.38	0.01
Distance	Mg																					0.39	0.002*
Distance	K																					0.41	0.002*
Distance	OP																	0.25	0.04	0.18	0.11	0.43	0.001*
Distance	TP																					0.40	0.001*
Distance	Rainfall	0.12	0.24			0.26	0.06					0.11	0.23	0.09	0.27	0.05	0.36			0.18	0.15	0.28	0.01
Distance	Soil Type																	0.40	0.006*			0.39	0.002*
Distance	age	0.34	0.02	0.39	0.02	0.38	0.01	0.21	0.12	-0.04	0.54	0.30	0.01	0.26	0.03	0.20	0.07	0.40	0.01	0.32	0.01	0.39	0.003*

Parameter	Controlled for	Glomeromycota		Ascomycota		Basidiomycota		Rhizaria		Stramenopiles		Cercozoa		Nematodea		Archaea		Euryarchaea		Crenarchaea	
		r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance	0.54	0.001*	0.38	0.001*	0.50	0.002*			0.52	0.001*					0.71	0.001*			0.60	0.001*
C	Distance					0.36	0.001*									0.26	0.04			0.35	0.01
N	Distance					0.41	0.002*									0.24	0.04			0.36	0.08
N03	Distance																			0.20	0.07
Mg	Distance															0.35	0.03	0.67	0.009*	0.44	0.003*
K	Distance																				
OP	Distance	0.30	0.04																		
TP	Distance																				
Rainfall	Distance	0.19	0.08			0.00	0.51														
Soil Type	Distance																				
Land Use	Distance																				
Distance	pH	0.20	0.07	-0.13	0.87	0.10	0.15			0.02	0.42			0.23	0.03	-0.08	0.74			0.00	0.45
Distance	C					0.22	0.006*							0.23	0.05	0.08	0.24			0.11	0.19
Distance	N					0.21	0.02							0.24	0.04	0.06	0.31			0.10	0.17
Distance	N03													0.28	0.02					0.10	0.21
Distance	Mg													0.23	0.04	0.41	0.13	0.04	0.36	0.17	0.08
Distance	K													0.24	0.05						
Distance	OP	0.19	0.06											0.29	0.005*						
Distance	TP													0.24	0.05						
Distance	Rainfall	0.12	0.19			0.17	0.03							0.22	0.05						
Distance	Soil Type													0.23	0.04						
Distance	Land Use																				
Distance	Age	0.33	0.01			0.24	0.02														

Under arable land use management, environment was the key driver of β-diversity in both prokaryotes and eukaryotes including six of the 11 bacterial taxa, all Archaea and three of the seven eukaryote groups. Geographic distance was not a significant predictor of β-diversity for any of the taxa groupings. Partial Mantels revealed that the distribution of four of the 11 bacterial taxa groupings (Actinobacteria, Verrucomicrobia, Gammaproteobacteria and total bacteria), correlated with soil type only. Variation in all other bacterial taxa groupings remained unexplained. pH was a significant driver of two of the three fungal groups including Glomeromycota ($P = 0.006$, $r = 0.30$), Ascomycota ($P = 0.01$, $r = 0.26$), Stramenopiles ($P = 0.003$, $r = 0.29$) and two of the three Archaea taxa groupings including total Archaea ($P = 0.001$, $r = 0.59$) and Crenarchaea ($P = 0.001$, $r = 0.53$). Mg was a significant driver of Basidiomycota ($P = 0.004$, $r = 0.34$) and Archaea ($P = 0.01$, $r = 0.31$) (Table 2.11).

Table 2.11: Partial Mantels analyses for the association between taxa community composition and both geographical distance and environmental variables under arable conditions. Significant P values are denoted with an asterisk, after accounting for Bonferroni Correction ranging between 0.003 and 0.03.

Parameter	Controlling for	Actinobacteria		Acidobacteria		Bacteroidetes		Verrucomicrobia		Firmicutes		Proteobacteria		Alphaproteobacteria		Betaproteobacteria		Deltaproteobacteria		Gammaproteobacteria		Bacteria	
		r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance																						
C	Distance																						
N	Distance																						
N03	Distance																						
Mg	Distance																						
K	Distance																						
OP	Distance																			0.33	0.02		
TP	Distance																						
Rainfall	Distance																			0.29	0.03		
Soil Type	Distance	0.28	0.008*			0.16	0.05	0.21	0.02*									0.17	0.05	0.28	0.004*	0.2216	0.01*
Land Use	Distance																						
Distance	pH																						
Distance	C																						
Distance	N																						
Distance	N03																						
Distance	Mg																						
Distance	K																						
Distance	OP																			-0.07	0.68		
Distance	TP																						
Distance	Rainfall																			-0.14	0.8		
Distance	Soil Type	0.14	0.19			-0.04	0.55	0.04	0.41									-0.02	0.52	0.14	0.16	-0.05	0.56

Parameter	Controlling for	Glomeromycota		Ascomycota		Basidiomycota		Rhizaria		Stramenopiles		Cercozoa		Nematode		Archaea		Euryarchaea		Crenarchaea	
		r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P	r	P
pH	Distance	0.3	0.006*	0.26	0.01*	0.22	0.04	0.13	0.12	0.29	0.003*					0.59	0.001*	0.244	0.044	0.526	0.001*
C	Distance					0.17	0.05									0.18	0.05			0.27	0.01
N	Distance	0.21	0.05			0.001	0.47	0.16	0.07							0.15	0.06			0.30	0.01
N03	Distance	0.24	0.05																		
Mg	Distance			0.2	0.03	0.34	0.004*									0.31	0.01*	0.40	0.01	0.25	0.02
K	Distance																				
OP	Distance																				
TP	Distance					0.23	0.02	0.31	0.006*												
Rainfall	Distance																				
Soil Type	Distance			0.33	0.004*					0.19	0.03					0.22	0.03	0.37	0.01		
Distance	pH	0.004	0.48	-0.13	0.92	0.01	0.44	0.15	0.12	0.04	0.38					0.04	0.35	-0.06	0.65	0.00	0.47
Distance	C					0.02	0.42									0.11	0.16			0.05	0.25
Distance	N	0.03	0.42			0.03	0.38	0.16	0.14							0.10	0.17			0.04	0.31
Distance	N03	0.04	0.39																		
Distance	Mg			-0.09	0.84	0.05	0.33									0.13	0.12	0.04	0.37	0.07	0.21
Distance	K																				
Distance	OP																				
Distance	TP					-0.06	0.69	0.07	0.24												
Distance	Rainfall																				
Distance	Soil Type			-0.07	0.77					0.1	0.21					0.15	0.12	0.04	0.42		

2.3.3. Impact of land use management on alpha diversity

Meta-analysis showed that land use management (i.e. conversion from arable to set-aside land) resulted in a significant increase in alpha diversity of five of the seven eukaryote groups, including all fungal taxa, Rhizaria and Nematodes (Figure 2.4). In contrast land management had no significant effect on the alpha diversity of any prokaryote, with the exception of Acidobacteria, which showed an increase in alpha diversity in set-aside relative to conventional arable management, as confirmed by two of the three indices (1/D and H' diversity indices).

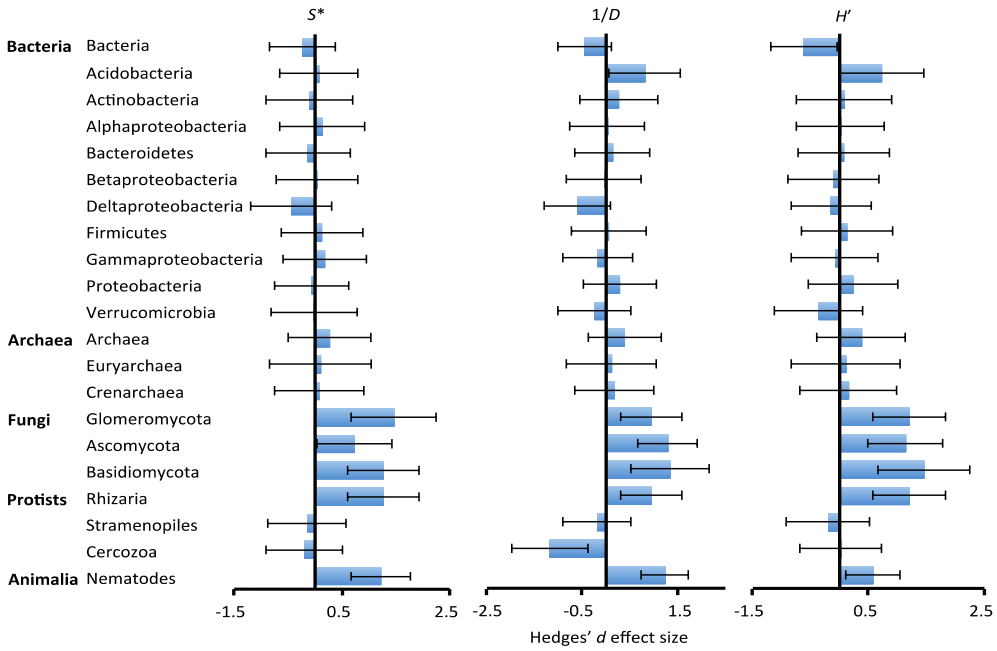


Figure 2.4 Meta-analysis of local level differences in diversity by land use change as measured by Hedges' *d* effect size. S – Richness; 1-D – Reciprocal of Simpsons Diversity Index; H – Shannons Diversity Index. Columns represent effect size and error bars (n=1000) represent s.e (vi) of the effect size. (The s.e bars that pass through zero represent no significant effect of conversion from arable to set-aside on diversity. Positive s.e bars represent an increase in alpha diversity following conversion).

2.4. Discussion

We show here for the first time that taxa groupings from all three domains demonstrate spatial scaling across the landscape (as indicated by significant distance-decay relationships). When demonstrating the biogeographical patterns of microorganisms, literature has relied heavily on bacteria as a model (King *et al.*, 2010; Bell, 2010; Fierer and Jackson, 2006) and to a lesser extent AM fungi (Kivlin *et al.,* 2011, van der Gast *et al.,* 2011a, Green et al., 2004). However this study has highlighted a biogeographical distribution not only in bacteria and AM fungi but also in taxa that are under represented in the literature such as Basidiomycota (Feinstein and Blackwood, 2013), Stramenopiles (protists) (Bates *et al.,* 2012) and certain cercomonad Protists (Bass *et al.*, 2007).

When assessing the combined data we found that environment was the key driver of all prokaryote and eukaryote community distributions. With the exception of nematodes, geographical distance did not correlate with the β-diversity of any group. When determining which environmental factors were responsible for these correlations, the environmental drivers of microbial distribution differed significantly between domains. Distribution of the majority of bacteria taxa groupings correlated with either pH or soil type or both. Variation in Archaea β-diversity correlated with pH, C, N and Mg. Crucially, land use management significantly correlated with all eukaryote taxa groupings investigated but was not shown to influence the spatial scaling of any prokaryote taxa groupings. The fundamental differences in environmental drivers of prokaryote β-diversity and eukaryote β-diversity at the landscape scale, highlighted here, could reflect their diverse life history traits including reproduction and growth, in addition to varying dispersal mechanisms (van der

Gast *et al.,* 2011a). Competitive fitness as a life history strategy is also widely variable between domains (Prosser *et al.*, 2007).

Bacteria exhibit narrow intracellular pH ranges required for optimal growth and are likely to be susceptible to minor pH changes in the environment, given that they are single-celled organisms. As a result the distribution of bacteria is often shown to correlate with pH (Fierer and Jackson, 2006; Ramette and Tiedje, 2006; Lauber *et al.,* 2008; Griffiths *et al*., 2011; Rousk et al., 2010). Very few studies have made direct comparisons between multiple soil types and microbial biogeography. The importance of soil type in determining the distribution of bacteria is unsurprising given the close association between edaphic variables and different soil types. It is interesting to note that Firmicutes were the only taxon that did not correlate with either environmental variables or geographical distance. Very few studies have investigated spatial scaling in Firmicutes and as a result they are substantially under represented in the literature. It is possible that variables not investigated in this study including physical disturbance such as tillage, may have more of an influence on Firmicutes than other bacterial taxa groupings. Whilst this difference to other bacterial taxa may be linked to their predominant dispersal mechanism, as spore formers and subsequent ability to persist in unfavourable conditions, further work is required to understand why Firmicutes are so different to other bacterial taxa (Shade *et al.*, 2013).

Archaea distribution was influenced by pH, C and N. However the strong influence of Mg shown in all Archaea taxa groupings investigated has not been shown before. The importance of C:N ratio, along with pH and to a lesser extent moisture content, have been highlighted in the literature as important drivers of

Archaea biogeography. However spatial scaling in soil Archaea is under represented in the literature and of the soil-based studies that have addressed this, Mg is very rarely measured (Wessén *et al.*, 2010, 2011; Bates et al., 2011). Pasternak *et al.*, (2013) measured Mg in relation to Archaea distribution and found no correlation in their semi-arid/arid sites. They recorded Mg in the range of 0.49 – 85.5 mg/kg whereas the current study recorded Mg levels in the range of 28.4 – 317.2 μg/g suggesting correlations occur at higher Mg ranges.

Eukaryotes on the other hand are known to be driven by an array of abiotic and biotic factors such as above ground (plant) community composition (de Vries *et al.*, 2012), C and N levels (Lauber *et al.,* 2008) and soil moisture (Kivlin *et al.,* 2011; Klamer and Hedlund, 2004). Dumbrell *et al.,* 2009 showed that pH played an important role in the distribution of AM fungi, linking pH levels to rates of decomposition and phosphate availability. pH is also thought to play a crucial part in shaping plant community composition, therefore may indirectly and directly influence AM fungi (Newman and Reddell, 1987). Our results not only reinforce the importance of pH in driving the distribution of AM fungi, but concomitantly its importance in determining the distribution of ascomycete and basidiomycete fungi. Another key finding is the importance of land use as a driver of β-diversity of every eukaryote group. Many fungal, Nematode and protist genera are obligate biotrophs of higher organisms, particularly plants, and this may limit their dispersal capabilities (Neuhauser *et al.,* 2011a; Neuhauser *et al.*, 2014; Duplessis *et al.*, 2011; Bouwmeester *et al.*, 2007). Therefore the paucity of above-ground diversity associated with arable conditions may contribute to the differences in β-diversity seen in below-ground microbial groups. Similarly, the distribution of predatory Nematodes and protists

may be tied to that of specific organisms at lower trophic levels on which they feed (Yeates *et al.*, 1993).

Few studies have assessed how land use management impacts microbial distance-decay relationships. Van der Gast *et al*., (2011a) found that intensive agriculture reduced turnover rates of arbuscular fungi across the landscape, relative to organic management. We however found that turnover rates were reduced in just three of the bacterial taxa groupings on conversion from arable to set-aside. We did reveal however that eukaryote alpha diversity was impacted by conversion from arable to set-aside. Alpha diversity of eukaryotes could be negatively impacted by fertiliser inputs, tillage and crop rotation, typical of conventional management, which result in selection of similar communities, reducing diversity (van der Gast et al., 2011a).

This study also showed that land use management impacted the drivers of microbial spatial scaling. When investigating the relative importance of individual environmental factors, different environmental variables were important in determining β-diversity depending on the type of land use management. For example bacteria β-diversity correlated with soil type under arable conditions whilst variation in the rest of the bacterial taxa remained unexplained. Under set-aside conditions however, pH was a key predictor of distribution in most of the bacterial taxa. In contrast to this, land use management did not influence Archaea spatial distribution, with the same environmental variables significant in both set-aside and arable treatments. Within the eukaryotes, pH was the main driver of spatial distribution within all of the fungal groups and the Stramenopiles under both arable and set-aside conditions (with the exception of Basidiomycota under arable conditions). Different

environmental variables influenced the β-diversity of Basidiomycota fungi, with Mg a significant driver of Basidiomycota distribution under arable conditions compared with pH, C, N and geographic distance under set-aside conditions (mentioned above). Only minor differences in the drivers of beta diversity occurred within the other eukaryote taxa under different land use conditions.

Studies have investigated the impact of land use on bacterial and fungal alpha diversity and abundance. Many have demonstrated a negative impact of intensive land use management on soil microbiota (van der Gast *et al.*, 2011a; Roesch *et al.*, 2007; Acosta-Martínez *et al.*, 2008). Of the few studies that have investigated whether land use management itself is a driver of spatial distribution of microorganism, they conclude that edaphic properties were more influential than land use type (Lauber *et al.*, 2008; Kuramae *et al.*, 2012; Hazard *et al.*, 2012). However, they did not explore how agricultural management impacts the drivers of β-diversity. Archaea and most protists for example were shown to be driven by the same edaphic properties irrespective of land use management condition. This study demonstrated that land use management is an important driver in the distribution of eukaryotes only and the drivers of bacterial and certain eukaryote distributions such as basidiomycota, are impacted under varying land use management. In this case, further research is needed to explore the relationship between the taxa grouping and the drivers for example bacteria and soil type and basidiomycota and Mg.

No study has simultaneously examined the biogeographical patterns of all three domains across more than one land-use type in a single agricultural system. We conclude that there is no single universal driver of β-diversity across the landscape. Major differences in the drivers of spatial scaling occurred between

and within domains and at the phylum level. The landscape distribution of eukaryotic taxa was determined by environment and land use, while for bacterial taxa, environment was the key driver. Furthermore the environmental variables associated with community composition differed between prokaryotic and eukaryotic taxa groupings depending on land use management. Further research is needed to identify the processes that influence the spatial distributions of each of the domains at a landscape scale.

CHAPTER III: IMPACT OF LAND USE MANAGEMENT ON THE COMMONNESS AND RARITY OF MICROBES

3.1 Introduction

Understanding the patterns of biological diversity have long been of interest to ecologists as they provide an insight into the fundamental processes that shape communities including gene flow, population dynamics and how organisms are likely to respond to change and stress such as climate change and disturbance (Sagarin et al., 2006, Preston et al., 1948, Ulrich 2008). A universal concept in community ecology that dates back to Darwin (1859) describes variation in the proportional abundances of taxa in a community as a universal pattern of commonness and rarity:

> In no environment…are all species equally common. Instead, it is universally the case that some are very abundant, others only moderately common, and the remainder-often the majority-rare. This pattern is repeated across taxonomic groups", (Magurran 2004).

Investigating the commonness and rarity within local communities has been used as a basis to explain the excess of rare species of macrofauna such as fish and insects (Magurran and Henderson, 2003, Ulrich and Zalewski, 2006). However the advent of faster, cheaper NGS technology has provided a greater understanding of microbial biodiversity. With the increase in sampling depth, rare taxa have been the focus of microbial ecology particularly in marine bacterial communities (Campbell et al., 2011, Sjostedt et al., 2012, Pedros-Alios, 2012, Galand et al., 2009). Furthermore a limited number of studies have

investigated species abundance distributions (SADs) in soil bacteria and fungi, where the common and rare members of communities have been identified and potential processes acting on their distributions discussed (Unterseher et al., 2011, Pereira e Silva et al., 2013).

Empirical patterns of SADs have been quantified since the 20th century, by fitting known distributions to data in order to evaluate assemblages (Magurran et al., 2011, Ulrich and Ollik, 2004). A number of models have since been proposed to describe distributions including the broken stick model, which predicts extremely even abundances of organisms and the geometric series, which predicts extremely uneven abundances of organisms (Tokeshi, 1993). However the log-series distribution suggested by Fisher et al. (1943) was one of the earliest attempts to use a mathematical model to describe the relationship between the number of species and the number of individuals in those species in a natural community. Its function was to highlight the processes determining biological diversity of the species in question. The model predicts that a few species will be abundant whilst a large proportion of the community will be rare species. Consequently the model is likely to apply to situations in which one or a few factors dominate the ecology of an assemblage (Magurran 2004). It was typically used to described species poor communities and/or those under perturbation.

Preston (1948) was the first to describe stochastic variation in relative abundance using the log-normal distribution, which has become the most frequently used distribution to describe SADs, based on his work on birds and moths. The model expresses abundance on a log scale and for a long time was the default model to explain the majority of large, unperturbed communities,

where their distribution is the product of the effect of a number of random influences on large heterogeneous assemblages (Bardgett 2005). The model makes no assumptions about demographic, ecological or evolutionary mechanisms that are acting on bacterial community structure, (Dunbar et al., 2002). The log-normal distribution is therefore highly applicable and important to biological communities, making it popular in ecology.

Models that describe species abundance such as the log-normal and log-series require just two critical parameters; an estimate of the population size of the most abundant species and an estimate of the total species richness in the community in order to identify commonness and rarity in a meta-population context. These models have been typically applied to local communities and are concerned with the abundance of constituent species at a particular locality or habitat. It has been proposed that completely sampled animal and plant communities will follow a log-normal SAD, whereas under-sampled communities fail to fit log-normal distributions. Furthermore log-normal distributions are commonly used as a null model when describing bacterial species abundance (Dunbar et al., 2002, Bardgett 2005). Microbial studies investigating structure within local communities have been limited to single studies of bacteria and fungi (Galand et al., 2009, Unterseher et al., 2011). However no study has looked at community structure in terms of commonness and rarity across multiple domains within a single system.

Furthermore the quantification of species distribution patterns has also been expanded to consider metacommunities in terms of spatial (site occupancy) and temporal (persistence) distributions (Dunbar et al., 2002, van der Gast et al., 2011a). Metacommunities can be defined as large-scale regional assemblages of

individuals that share similar trophic characteristics, which exists as a series of local communities. These local communities are thought to be linked by dispersal of potentially interacting species (Prosser 2007). Studies have found that it is biologically intuitive to fit different models to different parts of the SADs (Magurran et al., 2011). For example, Magurran and Henderson, (2003) investigated the SADs of fish communities from Hinckley Point, Bristol Channel, UK. They found that the community was best explained as two overlapping SADs, whereby the common species were fitted with a log-normal distribution and the rare species with a log-series distribution. This has also been shown in microbial studies including clinical bacterial studies of cystic fibrosis patients (van der Gast et al., 2011b).

One method used in the characterization of metacommunities, relies on fitting the data to a Poisson distribution (Magurran and Henderson, 2003, van der Gast et al., 2011a, Rogers et al., 2013a). By decomposing the overall distribution using an index of dispersion based on the ratio of variance to mean abundance, the community could be fitted to a Poisson distribution. 2.5 % and 97.5 % confidence limits of the χ^2 distribution were used to visualize non-random (common) and random (rare) dispersal through space, (Krebs, 1999, van der Gast et al., 2011a). These confidence limits can result in anomalies, whereby a highly persistent taxon will be classified as rare. It might, therefore be present in every sample, but in low abundance. Similarly, a taxon may be classified as common that has low persistence, where it may have a large abundance in one or two samples. Common and rare taxa are said to have very similar variance-mean ratios, which may make distinguishing between the two groups problematic. Furthermore patterns of dispersal and spatial distribution could be missed using

the Poisson distribution. Other studies have opted to use persistence or occupancy thresholds, based either on the number of samples a taxa is present in or the number of time points in which it was recorded. For example studies have used persistence whereby taxa occurring in 50 % or more of samples are classified as common, and less than 50 %, rare (Dolan et al., 2009, Unterseher et al., 2011).

Metacommunity analysis therefore enables the investigation of the variation in local communities through space to be taken into account by considering persistence in addition to abundance and be visualised using a distribution abundance relationship (DAR). DARs allow the exploration of coherent metacommunities in which rational immigration and dispersal across local communities is present. The species frequency distribution is said to be unimodal whereby rare taxa make up the majority of the community and are restricted in their distribution, whilst the common taxa (of which there are fewer) are widely distributed. Within a metacommunity, it is expected that a positive relationship exists between distribution and abundance with a continuum of species that are widespread and abundant and others that are endemic and rare. This relationship reflects the stochastic processes of local colonization and extinction. It predicts that the distribution and abundance of taxa are likely to fluctuate at random resulting in shifts in the core and rare members of a community and been shown extensively for macroorganisms (Guo et al., 2000, Hanski 1982, Gaston et al., 2000, Magurran and Henderson 2003). This has also been explored in pathogenic communities of bacteria within human hosts, which have provided a crucial insight into the behaviour of common taxa (often disease

causing) for example in the intestine and respiratory tract (Tap et al., 2009, Claesson et al., 2011, van der Gast et al., 2011b).

Common microorganisms are typically more widespread, abundant and functionally active in their environment, whilst the highly diverse rare taxa however may also collectively impact global biogeochemical processes. It is also thought that they act as a seed bank in that the rare organisms may be adapted to certain conditions in different places or in different seasons for example and therefore may thrive by dispersing (Fuhrman 2009). The seed bank may apply to the whole organism or just its genes. For example genes from a rare individual may be transferred to another creating a recombinant that is better adapted to a habitat. The seed bank therefore is valuable to the community in terms of coping with anthropogenic change or disturbance. Although taxa may be rare when sampled in a local community (or a single point in space or time), a change in environmental conditions could result in the rare taxa flourishing and becoming common and therefore functionally active. This is particularly significant when investigating the impact of land-use change on the microbial metacommunity (Fuhrman 2009). However the impact of disturbance in the form of land-use management on the common and rare members of a community, for example conversion from arable to set-aside (arable land left fallow), have yet to be elucidated for microbial communities. No study has addressed the persistence of common and rare microbial taxa in the soil in response to human impact, which will provide a detailed insight into how microbial communities respond to change.

This study aimed to compare the impact of disturbed (arable), and non-disturbed (the agri-environment scheme set-aside) environments, on the

community structure of multiple microbial metacommunities. Using 454 high-throughput pyrosequencing, we sampled from multiple farm sites (paired arable fields and AES set-aside margins) with a maximum separation of 390 km, on a range of soil types and latitudes in the UK**,** in order to (i) determine whether coherent metacommunities exist across all three domains of life with an agricultural soil system, as indicated by a universal positive relationship between taxa abundance and distribution (ii) investigate whether the community structure of multiple microbial groups can be partitioned into common and rare members (iii) determine the impact of land use management on the common and rare communities (iv) understand the impact of land use management on microbial community structure in terms of dominance and evenness. It is expected that under arable conditions, given that above-ground diversity is very low and inputs such as fertilisers, herbicides create a selective pressure, the community is predetermined to follow a dominant structure. Set-aside on the other hand has substantial above-ground diversity and therefore is likely to result in a more even community structure (Ager et al., 2010).

3.2. Materials and Methods

3.2.1. Sampling, site characterization, isolation and extraction of soil DNA, PCR amplification, pyrosequencing of 18S rRNA and 16S rRNA genes and sequence processing

Analysis was based on 16S rRNA and 18S rRNA gene pyrosequencing data generated from soil samples collected from 34 samples at 17 farm sites, (paired

arable fields and AES set-aside margins) with a maximum separation of 390km. Detailed methods including site information, isolation and extraction of soil DNA, PCR amplification, pyrosequencing of 18S rRNA and 16S rRNA genes and sequence processing are described in Chapter 2. From the four primer sets, 21 microbial taxa groupings were chosen for statistical analysis. Actinobacteria, Acidobacteria, Proteobacteria, Alphaproteobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria, Verrucmicrobia, Firmicutes, Bacteroidetes and total bacteria were taken from the bacterial primers. From the eukaryote primers, Basidiomycota, Ascomycota, Glomeromycota, Stramenopiles, Rhizaria, Archaea, Crenarchaea and Euryarchaea were chosen. Finally, Nematode and Cercozoa sequences were derived from the Nematode and Cercozoa specific primers. The mean number of sequences per sample ranged from 1967 sequences for Nematodes to an average of 392 sequences per sample for Bacteroidetes.

3.2.2. Statistical analysis

Using proportional abundances, species abundance distributions (SAD) were fitted to the data (Magurran and Henderson, 2003; Preston, 1960). The number of sites occupied (distribution/persistence) and the percentage mean relative abundance of each taxonomic group, under each land use management condition was calculated. Persistence totalled the number of samples for which each OTU was observed and the mean proportional abundance was taken for each OTU across all samples. Proportional mean relative abundance ($\log_{10}$ scale) was plotted on a scatter graph and a simple linear regression fitted to the graph to

determine whether the relationship between abundance and distribution was significant.

Taxa within each metacommunity were partitioned into common and rare groups for both arable and set-aside treatments using a 75% cut-off frequency based on persistence. OTUs that occurred in 75% or more of the samples were persistent (in addition to being most abundant) and classified as common. OTUs that occurred in less than 75% of samples were occasional (in addition to being typically less abundant) and classified as rare (as used in Unterseher et al., 2011). Statistical significance of differences between land use condition (set-aside and arable) for the whole, common and rare metacommunities were determined using ANOSIM (one-way analysis of similarity) (PRIMER, version 6, Primer-E). ANOSIM was used to calculate the level of dissimilarity between samples (global R), where an R value of 1 indicated that the populations from the land use treatment being tested were completely dissimilar to each other. An R value of zero suggested the populations were completely random. An associated significance level (P), based on 999 permutations was also calculated.

To determine which OTUs (either common or rare groups) made the largest contribution to the differences between land use treatment, similarity percentage analyses (SIMPER) were performed. PRIMER software (v6) was used to calculate similarity percentages (SIMPER) to determine taxa contributions to the average Bray-Curtis dissimilarity between set-aside and arable communities. SIMPER calculates the overall percentage contribution that each OTU makes to the average dissimilarity between the two sets of samples (Clarke and Gorley, 2006). To determine how much of the abundance the

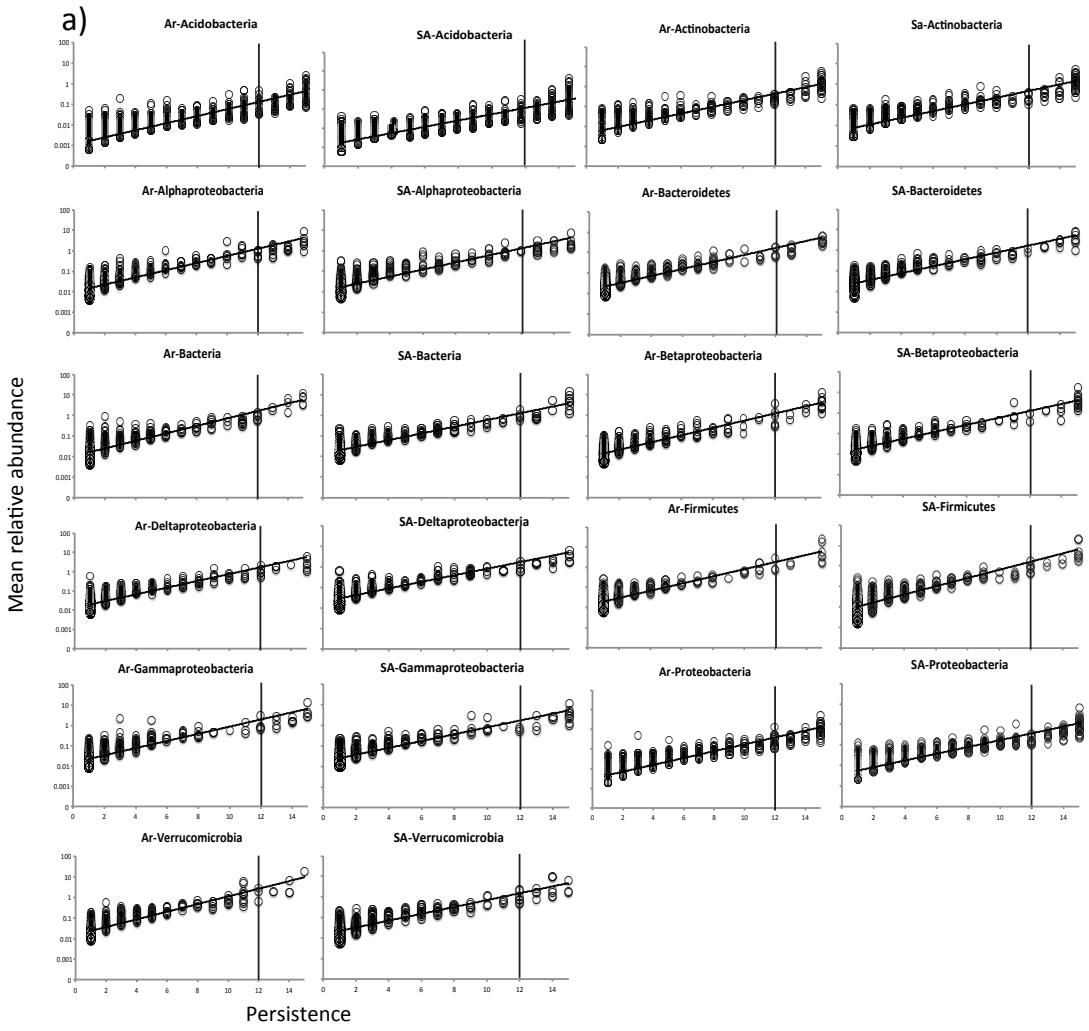
common and rare groups accounted for, the total relative contribution (%) for both groups was calculated within each taxa group.

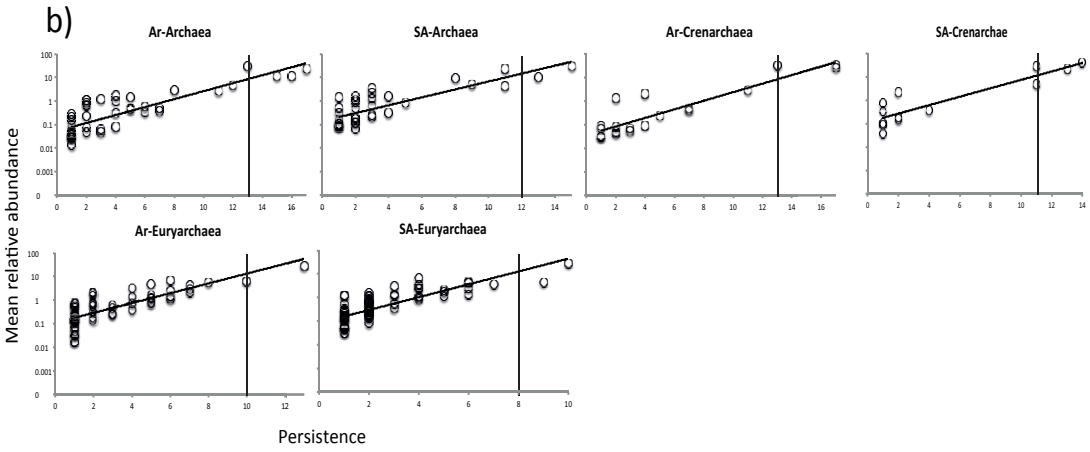
Rank abundance plots were used to determine differences in microbial community structure. Rank-abundance plots were derived from the relative abundance of each taxon, which was standardized to percentages (for each sample). For each plot, rank order was listed on the x axis, with relative abundance (log_{10}) transformed on the y axis. A simple linear regression was fitted to each model to give a slope value (b) using the equation $log_{10}y = a + bx$, (a is the intercept). The slope value (b) was then be used as a descriptive statistic indicating change in the structure of a community defined as evenness and dominance. A steeper slope would indicate a more dominant community structure. Linear regression, coefficients of determination (r^2) and significance (P) were calculated using Minitab software (version 16.1.0, Minitab, University Park, PA). A linear regression test using the t-distribution method was used to compare slopes, detailed in Fowler (1998). Comparisons were made between all common and all rare taxa present in set-aside and arable conditions for each taxa group.

3.3. Results

3.3.1. Relationship between taxa abundance and distribution across all three domains

To visualize the species abundance distributions (SAD) within the bacterial, archaeal and eukaryote metacommunities, percentage mean relative abundance of OTUs and local community occupancy was plotted for each taxa group (Figures 3.1a-3.1c).


a)
Ar-Acidobacteria
SA-Acidobacteria
Ar-Actinobacteria
Sa-Actinobacteria
Ar-Alphaproteobacteria
SA-Alphaproteobacteria
Ar-Bacteroidetes
SA-Bacteroidetes
Ar-Bacteria
SA-Bacteria
Ar-Betaproteobacteria
SA-Betaproteobacteria
Ar-Deltaproteobacteria
SA-Deltaproteobacteria
Ar-Firmicutes
SA-Firmicutes
Ar-Gammaproteobacteria
SA-Gammaproteobacteria
Ar-Proteobacteria
SA-Proteobacteria
Ar-Verrucomicrobia
SA-Verrucomicrobia
Mean relative abundance
Persistence


b)
Ar-Archaea
SA-Archaea
Ar-Crenarchaea
SA-Crenarchae
Ar-Euryarchaea
SA-Euryarchaea
Mean relative abundance
Persistence

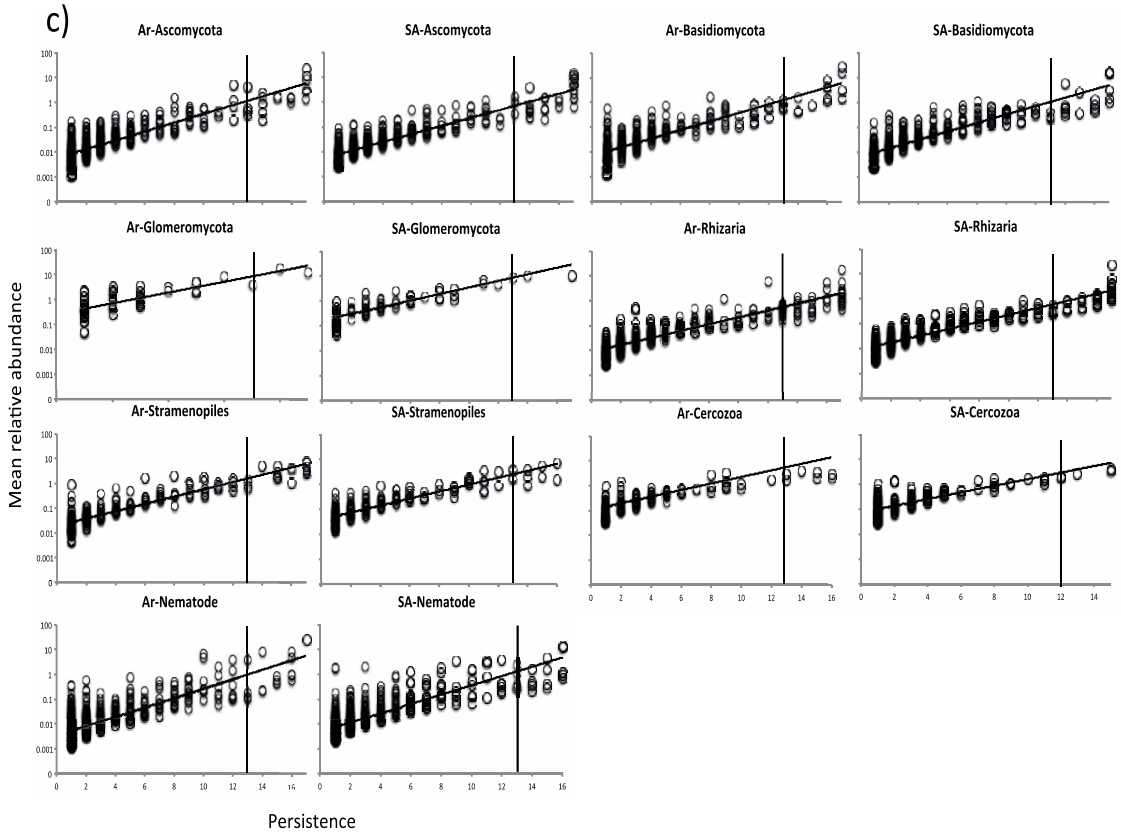


Figure 3.1 Distribution of (a) bacterial taxa groups; (b) archaea taxa groups, and; (c) eukaryote taxa groups across land use management (SA; set-aside, A; arable). Given is occupancy (number of samples for which each taxa was observed), plotted against percentage mean relative abundance (log_{10} scale) across all samples. All relationships are significant at P=<0.001. The vertical line represents a 75% cut-off whereby OTUs occurring in 75% or more of the samples (to the right of the line) are classified as common and those that did not are classified as rare.

There was a significant positive relationship between percentage mean relative abundance and occupancy for all 21 taxa groups encompassing all three domains for both arable and set-aside land use management (P = <0.001) (Table 3.1).

Table 3.1 Logistic regression analysis to determine relationship between percentage mean relative abundance and persistence.

		Arable		Set Aside	
		r^2	P	r^2	P
Bacteria	Bacteria	69.7	<0.001	73.9	<0.001
	Acidobacteria	64.4	<0.001	75.7	<0.001
	Actinobacteria	69.9	<0.001	76.9	<0.001
	Alphaproteobacteria	72.5	<0.001	70.1	<0.001
	Bacteroidetes	67.6	<0.001	74.9	<0.001
	Betaproteobacteria	70.6	<0.001	73.5	<0.001

	Deltaproteobacteria	68.6	<0.001	70.5	<0.001
	Firmicutes	64.3	<0.001	54.5	<0.001
	Gammaproteobacteria	64.2	<0.001	68.2	<0.001
	Proteobacteria	69.5	<0.001	71.7	<0.001
	Verrucomicrobia	71.1	<0.001	72.1	<0.001
Archaea	Archaea	72.4	<0.001	69.2	<0.001
	Euryarchaea	63.5	<0.001	70.0	<0.001
	Crenarchaea	73.0	<0.001	78.1	<0.001
Eukaryotes	Glomeromycota	59.5	<0.001	77.1	<0.001
	Ascomycota	70.9	<0.001	79.1	<0.001
	Basidiomycota	71.6	<0.001	78.4	<0.001
	Rhizaria	80.5	<0.001	83.3	<0.001
	Stramenopiles	79.6	<0.001	78.5	<0.001
	Cercozoa	74.4	<0.001	57.1	<0.001
	Nematodes	62.7	<0.001	66.5	<0.001

All microbial taxa groups could be partitioned into common and rare communities. Under both set-aside and arable land use management, there were fewer common OTUs than rare OTUs within every taxa group. Within the bacterial taxa group (excluding total bacteria), between 17-75 common OTUs and 593-3334 rare were present under set-aside, compared to 9-83 common OTUs and 638-3959 rare OTUs under arable conditions. Within archaea, 2-4 common OTUs and 9-82 rare OTUs were present in set-aside conditions, compared to 2-3 common OTUs and 17-73 rare OTUs under arable (excluding overall archaea). Finally within the eukaryotes, 3-45 common OTUs and 74-936 rare OTUs occurred within the set-aside land use management, compared with 3-49 common OTUs and 49-946 rare OTUs within arable land use management. This suggests that the rare OTUs accounted for the majority of the diversity within each taxa group (Table 3.2).

Furthermore, whether the common or rare OTUs collectively accounted for the majority of the abundance in the metacommunity varied between and within domains across both arable and set-aside conditions. Common OTUs

within the bacterial domain, accounted for an average of 38.12 % of the total relative abundance compared with 51.27 % for the rare OTUs. In all of the bacterial taxa (with two exceptions), the common OTUs contributed less within the whole community than the rare taxa (acidobacteria common 42.75 %, rare 46.7 %; actinobacteria common 36.83 %, rare 53.03; alphaproteobacteria common 31.2 %, rare 58.67 %; bacteroidetes common 40.48 %, rare 49.4 %; deltaproteobacteria common 23.2 %, rare 66.89 %; firmicutes common 43.36 %, rare 46.78 %; gammaproteobacteria common 25.93 %, rare 63.86 %; proteobacteria common 35.7 %, rare 57.22 %; verrucomicrobia common 44.89 %, rare 45.19 %). Within total bacteria, the common OTUs contributed 40.78 % to the total mean relative abundance compared with 40.73 % within the rare OTUs and common OTUs within the betaproteobacteria taxa group contributed 54.23 % to the mean total relative abundance compared with 35.78 % within the rare OTUs.

Within the archaea domain common OTUs accounted for an average 58.62 % and rare OTUs 48.20 % of the total relative abundance. Common OTUs contributed less within the whole community than the rare taxa within Euryarchaea only (common 22.21 %, rare 68.2 %). In both total archaea and crenarchaea, the common OTUs contributed more to the mean relative total abundance than the rare OTUs (total archaea common 62.62 %, rare 28.19 %; crenarchaea common 91.04 %, rare 0 %, based on an approximate 90 % cut-off). Percentage contribution was calculated as the mean contribution divided by the mean dissimilarity across samples, with cumulative percentage contribution set at 90 %. The rare OTUs within crenarchaea contributed very little to the mean

dissimilarity, accounting for the last 10 % of the cumulative abundance between 90 % and 100 %, (the rare OTUs were not shown in the table).

Eukaryote common OTUs accounted for an average 48.90 % and rare OTUs 41.22 % of the total relative abundance. Within three of the seven eukaryotes, the common OTUs contributed less than the rare OTUs within the whole community (glomeromycota common 35.48 %, rare 54.98 %; cercozoa common 16.25 %, rare 73.81 %; nematodes common 40.86 %, rare 49.2 %). In the remaining four eukaryote taxa groups, common taxa contributed more within the whole community than the rare taxa (ascomycota common 72.59 %, rare 17.47 %; basidiomycota common 63.54 %, rare 26.51 %; rhizaria common 61.82 %, rare 28.16 % (Table 3.2).

Table 3.2. Total number of common and rare OTUs within set-aside and arable fields, including the % contribution for both common and rare groups across both land use management practices combined. An approximate 90% cumulative percentage contribution cut-off was used (SA – Set aside, A – Arable).

		Common		Rare		Common Contribution in terms of abundance (%)	Rare Contribution in terms of abundance (%)	Up to 90% cumulative contribution
Domain	Taxa	SA OTUs	SA OTUs	SA OTUs	A OTUs			
Bacteria	Bacteria	243	230	9540	10549	40.78	40.43	81.21
	Acidobacteria	60	64	2003	2296	42.75	46.7	89.45
	Actinobacteria	28	32	883	975	36.83	53.03	89.86
	Alphaproteobacteria	17	21	802	927	31.2	58.67	89.87
	Bacteroidetes	24	17	666	810	40.48	49.4	89.88
	Betaproteobacteria	21	23	716	807	54.23	35.78	90.01
	Deltaproteobacteria	19	20	931	1236	23.2	66.89	90.09
	Firmicutes	17	12	966	675	43.36	46.78	90.14
	Gammaproteobacteria	18	19	882	984	25.93	63.86	89.79
	Proteobacteria	75	83	3334	3959	35.7	57.22	92.92
	Verrucomicrobia	22	9	593	638	44.89	45.19	90.08
Archaea	Archaea	2	4	35	44	62.62	28.19	90.81
	Euryarchaea	2	2	82	73	22.21	68.2	90.41
	Crenarchaea	4	3	9	17	91.04	0	91.04
Eukaryotes	Glomeromycota	3	3	74	49	35.48	54.98	90.46
	Ascomycota	32	21	506	456	72.59	17.47	90.06
	Basidiommycota	22	17	449	341	63.54	26.51	90.05
	Rhizaria	45	49	438	397	61.82	28.16	89.98
	Stramenopiles	12	21	212	217	51.77	38.39	90.16
	Cercozoa	5	7	238	219	16.25	73.81	90.06
	Nematodes	22	17	936	946	40.86	49.2	90.06

3.3.2. Determining the impact of land use management on spatial distribution at the metacommunity level

ANOSIMs were run to determine whether land use management impacted community composition within each domain. The results showed that

when comparing community composition within the whole community (combined common and rare OTUs) significant differences in the community composition between land use management were evident in all seven eukaryote taxa groups (glomeromycota r = 0.25, P = 0.001, ascomycota r = 0.32, P = 0.001, basidiomycota r = 0.24, P = 0.001, rhizaria r = 1.00, P = 0.001, stramenopiles r = 0.51, P = 0.001, cercozoa r = 0.06, P = 0.03, nematodes r = 0.19, P = 0.002 (Table 3.3). There were no significant differences in the community composition of either bacteria or archaea taxa between land use management practices. However when assessing the common and rare OTUs separately within each land use management condition, the ANOSIMS showed that both the common and rare communities (with the exception of common Crenarchaea and rare Euryarchaea) were significantly different under set-aside fields compared with arable fields (Table 3.3).

Table 3.3 ANOSIM summary statistics for set-aside and arable practice within whole, common and rare communities. Significant differences (P = ≤ 0.05) are denoted with an asterisk

		Whole		Common		Rare	
	Whole	r	P	r	P	r	P
Bacteria	Bacteria	0.06	0.13	0.29	0.001*	0.33	0.001*
	Acidobacteria	0.04	0.16	0.17	0.01*	0.24	0.01*
	Actinobacteria	-0.004	0.39	0.32	0.001*	0.23	0.01*
	Alphaproteobacteria	0.03	0.20	0.21	0.005*	0.16	0.01*
	Bacteroidetes	0.03	0.15	0.17	0.002*	0.22	0.001*
	Betaproteobacteria	0.02	0.23	0.11	0.03*	0.29	0.001*
	Deltaproteobacteria	-0.01	0.57	0.35	0.001*	0.12	0.02*
	Firmicutes	0.00	0.44	0.11	0.02*	0.33	0.001*
	Gammaproteobacteria	0.04	0.16	0.17	0.002*	0.14	0.01*
	Proteobacteria	0.04	0.17	0.23	0.01*	0.20	0.004*
	Verrucomicroba	0.05	0.13	0.55	0.001*	0.44	0.001*
Archaea	Archaea	-0.05	0.95	0.43	0.001*	0.46	0.001*
	Euryarchaea	-0.04	0.76	0.12	0.02*	0.04	0.22
	Crenarchaea	-0.04	0.79	-0.01	0.50	0.19	0.002*
Eukaryotes	Glomeromycota	0.25	0.001*	0.59	0.001*	0.19	0.001*
	Ascomycota	0.32	0.001*	0.43	0.001*	0.50	0.001*
	Basidiomycota	0.24	0.002*	0.31	0.003*	0.32	0.001*
	Rhizaria	1.00	0.001*	0.99	0.001*	0.98	0.001*
	Stramenopiles	0.51	0.001*	0.69	0.001*	0.44	0.001*
	Cercozoa	0.06	0.03*	0.59	0.001*	0.22	0.001*
	Nematodes	0.19	0.002*	0.59	0.001*	0.66	0.001*

SIMPER analyses were run to determine how the communities differed between set-aside and arable conditions. The SIMPER analyses show that within every bacterial taxa group, every OTU that contributed ≥1 % to the differences in land use management, were present in both set-aside and arable samples. However within each bacterial taxa group a number of OTUs (represented by <1 % abundance contribution to the difference in land use management) occurred in either set-aside or arable samples as opposed to both habitat types.

Within the bacterial domain, community shifts in common and rare taxa occurred under land use management (when looking at taxa with an abundance ≥1 % contribution). Within Actinobacteria, MC47, *Solirubroacterales* and *Streptomyces* were classified as common under set-aside and rare under arable, along with *Rhizobiales* in Alphaproteobacteria *and Bacillaceae*, *Planococcaceae* and *Paenibacillaceae* in firmicutes. *Spartobacteriales* (OTU ID 20621, 18817) *and Verrucomicrobiales* (OTU ID 2220, 16082) within Verrucomicrobia were also classified as common under set-aside and rare under arable. Within Alphaproteobacteria *Rhodospiralles* was classified as common under arable and rare under set-aside, in addition to *Burkholderiales i*n betaproteobacteria and *Burkolderiales* within proteobacteria (Appendix I; Tables 3.4-3.24).

All OTUs within total archaea and crenarchaea (with one exception in total archaea) occurred in both set-aside and arable samples. Within archaea however community shifts in common and rare taxa were evident. In total archaea, *Nitrososphaera* and an uncultured crenarchaea (OTU ID 7293) were classified as a common OTU under arable conditions but rare under set-aside. In crenarchaea, an uncultured crenarchaea (OTU ID 591) was classified as a common OTU under set-aside and rare under arable. Although the majority of

Euryarchaea taxa occurred in both land use management practices, uncultured Euryarchaea (OTU ID 8957) and *Methanobacteria* (OTU ID 1) occurred in just arable or set-aside conditions respectively (when considering only OTUs that contributed >1 % to the difference in land use management). Other taxa were specific to each habitat type but were <1 % abundant. Minor shifts in the common taxa occurred within Euryarchaea, for example an uncultured Euryarchaea (OTU ID 9289) was classified as common under arable conditions and rare under set-aside whilst another uncultured Euryarchaea (OTU ID 4198) was common under set-aside and rare under arable (Appendix I; Tables 3.4-3.24).

With regards to the eukaryotes, of the OTUs that contributed >1 % abundance towards the difference in land use management, all OTUs within Ascomycota, Basidiomycota, Rhizaria and Stramenopiles were present in both set-aside and arable practice. OTUs that were present in either land use practice occurred within Cercozoa, whereby Cercozoa sp. COHH1 (OTU ID 463) was present only in set-aside sites, whilst an uncultured Cercozoa (OTU ID 160) was present only in arable sites. Within Glomeromycota, *Glomus* (OTU ID 8750) and unclassified *Glomeraceae* (OTU ID 9610) occurred in arable sites only, whereas *Glomus* (OTU ID 5538) was only present within set-aside sites. Within Nematodes, *Aphelenchoid*es, *Paratrichodorus* and *Diplolaimella* were present in set-aside sites only.

Within the eukaryote taxa groups, certain OTUs were classified as common under arable and rare under set-aside including an uncultured ascomycota (OTU ID 2956) within the ascomycota, *Agaricomycetes* (OTU ID 4734) and *Atractiellales* (OTU ID 9106) within basidiomycota and uncultured

cercozoa (OTU ID 183 and 326) and *Cercomonas* (OTU ID 330), within cercozoa. *Ambiospora* (OTU ID 10839) and *Glomus* (OTU ID 5199) within glomeromycota were also classified as common under arable and rare under set-aside, including *Sorosphaera* (OTU ID 7671) within Rhizaria and unclassified *Aphanomyces* (OTU ID 6049), unclassified *Xanthophyceae* (OTU ID 3360), *Phytophthora* (OTU ID 6846), unclassified *Saprolegnia* (OTU ID 6278), *Sellaphora* (OTU ID 5826) and unclassified *Xanthophyceae* (OTU ID 941) within stramenopiles.

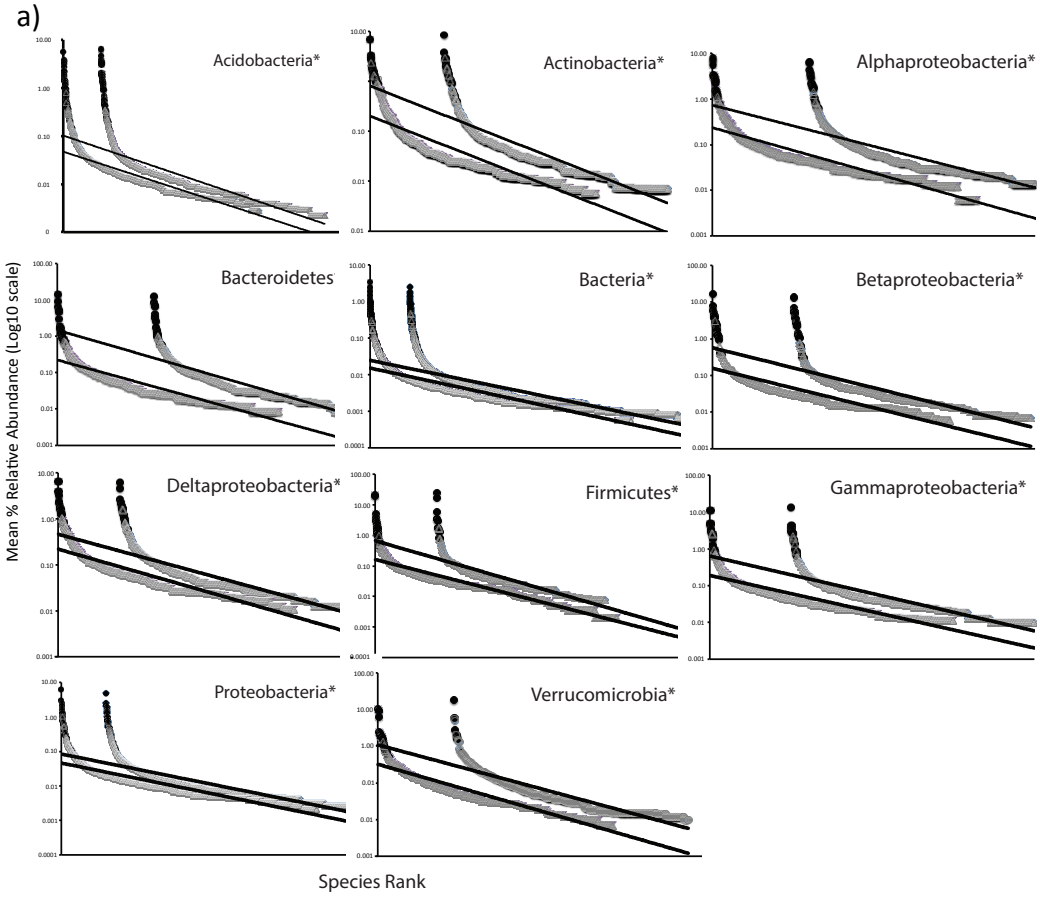
Other OTUs were classified as common under set-aside conditions and rare under arable conditions including *Paraphaeosphaeria* (OTU ID 8758), *Capnobotryella* (OTU ID 10190), *Tubeufiacaea* (OTU ID 9126) and *Neobulgaria* (OTU ID 6833) within ascomycota, uncultured cercozoa (OTU ID 372) within cercozoa, *Glomus* (OTU ID 2849 and 7284) within Glomeromycota. Within Rhizaria, *Cercomonadida* environmental samples (OTU ID 2428, and 2520) were also classified as common under set-aside conditions and rare under arable conditions, along with *Rhizidiomyces* (OTU ID 771) in stramenopiles.

Within the nematode taxa group, it is important to note that although the same common taxa were present in each land use management, differences within the common communities were shown. This is because the results have highlighted common OTUs with a relative abundance contribution >1 %. Many of the common OTUs were <1 % abundant but still highly persistent (hence classified as common). It is shifts within these taxa that are likely to be driving some of the differences within the land use management.

3.3.3. Impact of land use management on community structure in terms of dominance and evenness

Given that the common and rare communities were shown to be significantly different under land use management, rank abundance plots were used to visualize how the metacommunity structure changed for both common and rare groups under set-aside and arable management (Figures 3.2a-3.2c). The slope values (b) were used as a descriptive statistic of evenness. A value of zero indicates perfect evenness (Ager et al., 2010). When assessing combined metacommunity structure (common and rare together), the t-distribution test, (Fowler et al., 1998), demonstrated that the slopes were significantly different (P = <0.001 - 0.003) in all bacterial taxa groups, with the exception of Firmicutes. A change in metacommunity structure was observed in arable and set-aside samples in total bacteria (arable = b -1.154, set-aside = b -1.146), acidobacteria (arable = b -1.175, set-aside = b -1.165), actinobacteria (arable = b -1.332, set-aside = b -1.320), alphaproteobacteria (arable = b -1.143, set-aside = b -1.119), bacteroidetes (arable = b -1.322, set-aside = b 1-.262), betaproteobacteria (arable = b -1.246, set-aside = b -1.233), deltaproteobacteria (arable = b -1.117, set-aside = b -1.084), gammaproteobacteria (arable = b -1.132, set-aside = b -1.111), proteobacteria (arable = b -1.136, set-aside = b -1.125) and verrucomicrobia (arable = b -1.278, set-aside = b -1.319) (Table 3.3). A more gentle slope under set-aside treatment indicated an increase in community evenness following conversion from arable to set-aside (Table 3.3). An exception to this included Verrucomicrobia, whereby (although the slopes were significantly different under each land use management), conversion from arable to set-aside resulted in an increase in community dominance.

Within all eukaryote taxa groups, the t-distribution test, (Fowler et al., 1998), demonstrated that the slopes were significantly different (P = <0.001). A change in metacommunity structure was observed between arable and set-aside samples in all eukaryote taxa groups including glomeromycota (arable = b -1.311, set-aside = b -1.275), ascomycota (arable = b -1.701, set-aside = b -1.616), basidiomycota (arable = b -1.758, set-aside = b -1.645), rhizaria (arable = b -1.665, set-aside = b -1.593), stramenopiles (arable = b -1.701, set-aside = b -1.437) and cercozoa (arable = b -1.137, set-aside = b -1.068). There were no significant differences between arable and set-aside samples in any of the archaea taxa groups.



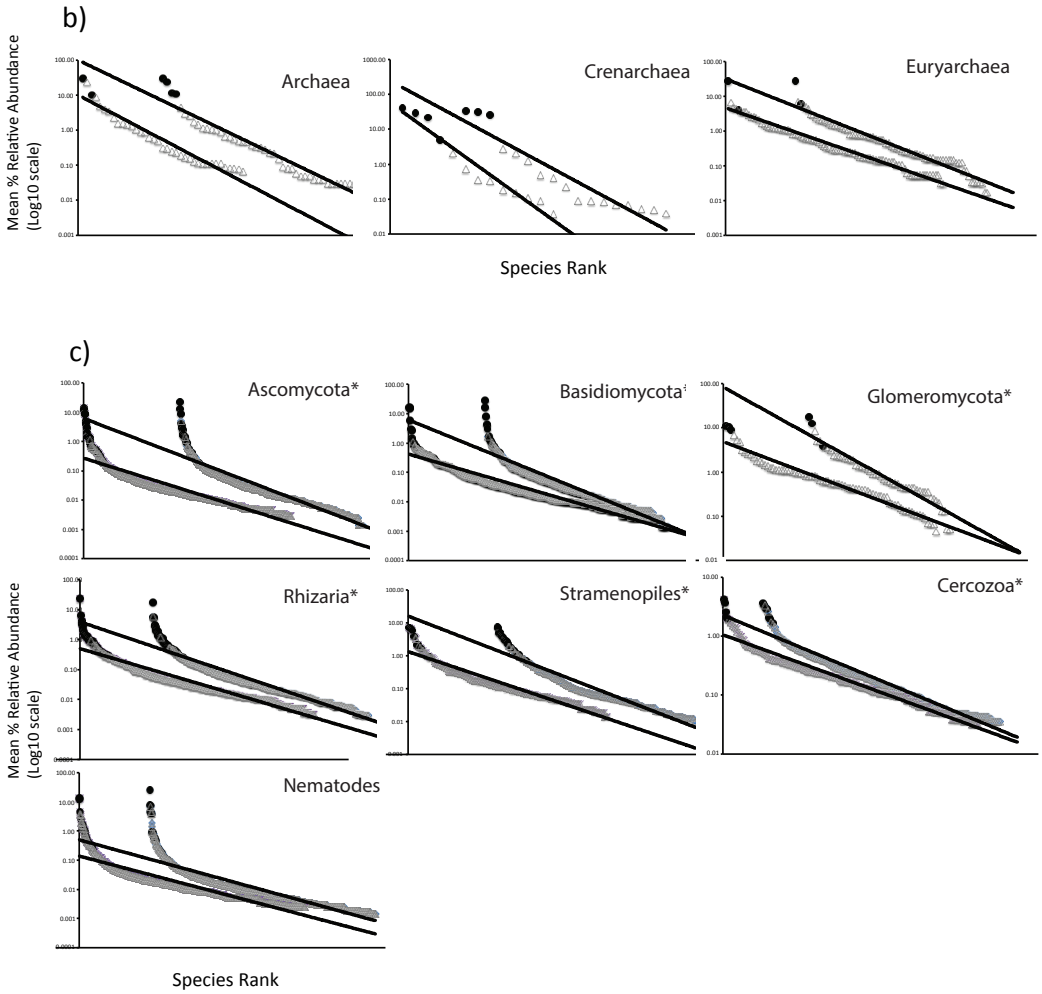


Figure 3.2a-c: Rank abundance distributions of (a) bacterial taxa groups; (b) archaea taxa groups, and; (c) eukaryote taxa groups, whereby the left hand slope represents the arable community and right hand slope represents the set-aside community. The common taxa are shown in black closed circles and the rare taxa in open grey triangles. The asterisk denotes a significant difference between slopes, determined by the t-distribution based on a linear regression test (Fowler, 1998).

Community structure was significantly different under land use management within the common metacommunity in six of the eleven bacteria taxa in which the slopes were significantly different (P = <0.001 – 0.05). A change in the metacommunity structure was observed between arable and set-aside samples in bacteria (arable = b -0.91, set-aside = b -0.91), Alphaproteobacteria (arable = b -0.81, set-aside = b -0.80), Bacteroidetes (arable = b -1.19, set-aside = b -1.10), Gammaproteobacteria (arable = b -0.92, set-aside

= b -1.01), Proteobacteria (arable = b -0.88, set-aside = b -0.89) and Verrucomicrobia (arable = b -1.25, set-aside = b -1.06).

A significant difference in the slopes was shown in one of the three Archaea taxa (total Archaea (P = 0.02)) and a change in metacommunity structure was observed under arable (b = -0.16) and set-aside (b = -0.46).

A significant difference in the slopes was demonstrated in four of the seven eukaryotes including Nematodes, Cercozoa, Stramenopiles and Ascomycota (P = <0.001 – P = 0.05). A change in metacommunity structure was observed between arable and set-aside in nematodes (arable = b -1.98, set-aside = b -1.42), cercozoa (arable = b -0.40, set-aside = b -0.52), stramenopiles (arable = b -0.80, set-aside = b -0.78) and ascomycota (arable = b -1.40, set-aside = b -1.36).

In all taxa with the exception of Gammaproteobacteria, Proteobacteria, Archaea and Cercozoa, conversion from arable to set-aside conditions resulted in an even community structure (as indicated by a more gentle slope). In these taxa, conversion of arable to set-aside conditions resulted in a more dominant community structure (Table 3.25).

When assessing the rare taxa under each land use management, the t-distribution test demonstrated that the slopes were significantly different (P = <0.001 – 0.02) in all taxa across all domains. Within the bacteria domain, a change in community structure was observed in the arable metacommunity when compared with the set-aside metacommunity, in total bacteria (arable = b -0.96, set-aside = b -0.94), acidobacteria (arable = b -0.92, set-aside = b -0.91), actinobacteria (arable = b -1.12, set-aside = b -1.11), alphaproteobacteria (arable = b -0.96, set-aside = b -1.01), bacteroidetes (arable = b -1.14, set-aside = b -

0.99), betaproteobacteria (arable = b -1.06, set-aside = b -0.95), deltaproteobacteria (arable = b -0.99, set-aside = b -0.94), firmicutes (arable = b -1.00, set-aside = b -1.19), gammaproteobacteria (arable = b -0.97, set-aside = b -0.94), proteobacteria (arable = b -0.95, set-aside = b -0.93) and verrucomicrobia (arable = b -1.17, set-aside = b -1.07).

Within the archaea domain, a change in community structure was observed in the arable metacommunity when compared with the set-aside metacommunity, in total archaea (arable = b -1.73, set-aside = b -1.76), crenarchaea (arable = b -1.85, set-aside = b -1.63) and euryarchaea (arable = b -1.42, set-aside = b -1.42). Within the eukaryote domain, a change in community structure was observed in the arable metacommunity when compared with the set-aside metacommunity, in glomeromycota (arable = b -1.15, set-aside = b -1.13), ascomycota (arable = b -1.40, set-aside = b -1.23), basidiomycota (arable = b -1.15, set-aside = b -1.40), rhizaria (arable = b -1.27, set-aside = b -1.20), stramenopiles (arable = b -1.37, set-aside = b -1.20), cercozoa (arable = b -1.03, set-aside = b -0.98) and nematodes (arable = b -1.40, set-aside = b -1.36).

In every taxa group within the combined metacommunity, common and rare groups, a gradual slope was observed under set-aside conditions indicating an increase in community evenness following conversion from arable to set-aside with the exception of Firmicute rare, total Archaea common and rare and Basidiomycota common and rare where conversion from arable to set-aside resulted in a dominant community (Table 3.25).

Table 3.25 T-distribution test statistics comparing slopes within arable and set-aside for whole, common and rare OTUs.

		Whole			Common			Rare		
Domain	Taxa	b	P	T	b	P	T	b	P	T
Bacteria	Bacteria Set Aside	-1.146	<0.001	10.9	-0.91	0.01	2.46	-0.94	<0.001	11.64
	Bacteria Arable	-1.154			-0.91			-0.96		
	Acidobacteria Set aside	-1.165	<0.001	6.05	-1.01	0.06	1.87	-0.91	<0.001	7.96
	Acidobacteria Arable	-1.175			-1.04			-0.92		
	Actinobacteria Set aside	-1.320	0.003	3.01	-0.62	0.38	0.89	-1.11	0.00	3.24
	Actinobacteria Arable	-1.332			-0.73			-1.12		
	Alphaproteobacteria Set aside	-1.119	<0.001	9.59	-0.80	0.05	2.06	-1.01	<0.001	13.49
	Alphaproteobacteria Arable	-1.143			-0.81			-0.96		
	Bacteroidetes Set aside	-1.262	<0.001	4.07	-1.10	<0.001	6.02	-0.99	0.02	2.32
	Bacteroidetes Arable	-1.322			-1.19			-1.14		
	Betaproteobacteria Set aside	-1.233	0.002	3.12	-0.96	0.98	0.02	-0.95	<0.001	4.38
	Betaproteobacteria Arable	-1.246			-1.08			-1.06		
	Deltaproteobacteria Set aside	-1.084	<0.001	11.86	-0.71	0.93	0.09	-0.94	<0.001	13.66
	Deltaproteobacteria Arable	-1.117			-0.73			-0.99		
	Firmicutes Set aside	-1.380	0.08	1.73	-1.35	0.25	1.17	-1.19	<0.001	13.89
	Firmicutes Arable	-1.188			-1.44			-1.00		
	Gammaproteobacteria Set aside	-1.111	0.001	3.22	-1.01	0.05	2.02	-0.94	<0.001	3.81
	Gammaproteobacteria Arable	-1.132			-0.92			-0.97		
	Proteobacteria Set aside	-1.125	<0.001	13.01	-0.89	0.01	2.80	-0.93	<0.001	16.81
	Proteobacteria Arable	-1.136			-0.88			-0.95		
	Verrucmicrobia Set aside	-1.319	<0.001	4.622	-1.06	0.01	2.88	-1.07	0.01	2.74
	Verrucmicrobia Arable	-1.278			-1.25			-1.17		
Archaea	Archaea Set aside	-1.955	0.44	0.77	-0.46	0.02	7.97	-1.76	0.01	2.58
	Archaea Arable	-2.215			-0.16			-1.73		
	Euryarchaea Set aside	-1.523	0.46	0.74	-	-	-	-1.42	<0.01	4.58
	Euryarchaea Arable	-1.545			-			-1.42		
	Crenarchaea Set aside	-2.996	0.26	1.15	-1.31	0.08	2.63	-1.63	0.01	3.09
	Crenarchaea Arable	-2.887			-0.24			-1.85		
Eukaryote	Glomeromycota Set aside	-1.275	<0.001	3.52	-0.14	0.11	2.77	-1.13	<0.001	11.41
	Glomeromycota Arable	-1.311			-1.29			-1.15		
	Ascomycota Set aside	-1.616	<0.001	8.44	-1.36		4.62	-1.23	<0.001	15.00
	Ascomycota Arable	-1.701			-1.40	<0.001		-1.40		
	Basidiomycota Set aside	-1.645	<0.001	15.62	-1.62	0.19	1.33	-1.40	<0.001	22.19
	Basidiomycota Arable	-1.758			-1.52			-1.15		
	Rhizaria Set aside	-1.593	<0.001	7.26	-1.03	0.61	0.52	-1.20	<0.001	10.63
	Rhizaria Arable	-1.665			-1.04			-1.27		
	Stramenopiles Set aside	-1.437	<0.001	5.14	-0.78	<0.001	4.20	-1.20	<0.001	4.20
	Stramenopiles Arable	-1.701			-0.80			-1.37		
	Cercozoa Set aside	-1.068	<0.001	6.56	-0.52	0.05	2.28	-0.98	<0.001	7.77
	Cercozoa Arable	-1.137			-0.40			-1.03		
	Nematoda Set aside	-1.500	0.09	1.69	-1.42	<0.001	5.43	-1.36	0.01	2.45
	Nematoda Arable	-1.542			-1.98			-1.40		

3.4. Discussion

This is the first study to highlight a universal pan microbial relationship. Positive species-abundance relationships have been shown in a range of macrofauna and plants and individual microbial taxa such as bacteria (Guo et al., 2000; Rogers et al., 2013a). The current study shows a significant positive relationship in multiple microbial groups, across all domains of life. It demonstrates a coherent community structure whereby a regional community exists that is comprised of

trophically similar individuals in local communities linked by rationale dispersal of interacting species between or into those local communities in the same habitat, at the landscape scale (Rogers et al., 2013a). Our data support the moderate endemicity model across all taxa groups, whereby some OTUs were widely distributed across the landscape and more locally abundant, whilst other taxa were restricted in their distribution and typically less abundant (Foissner 2008). This provides evidence for a universal relationship of emigration and distribution across the different taxa regardless of land use management, body size (not measured) or life history strategy. Many hypotheses exist to explain the mechanisms underlying the species abundance distribution. For example species adapted to a broader set of niches will be able to inhabit more habitat types. Alternatively local abundance is achieved through increased dispersal success and a reduced extinction rate. It has been proposed that distribution abundance relationships are a result of the interaction between processes involving local resources and extinction/dispersal mechanisms and that the two are not mutually exclusive (Guo et al., 2000; Werner et al., 2014). Positive distribution abundance relationships (like those shown in Figure 3.1) have also been demonstrated in macrofauna and clinical microbial studies (Guo et al., 2000, van der Gast et al., 2011a). Some studies however failed to find positive distribution abundance relationships. For example Barberan et al., (2011) who investigated network analyses in soil bacteria and archaea from a range of ecosystems across continents, did not find this trend in their study. This is likely to be due to the large scales at which their study was conducted extended across different habitats (Antarctica, North and South America), which does not allow for rationale immigration between communities. Furthermore a clinical study investigating

ascitic microbiota also found no relationship between distribution and bacterial abundance. This is likely to be due to the lack of common immigration route into local communities with the movement of bacteria from more than one area of the body, resulting in high species diversity regardless of sample occupancy (Rogers et al., 2013b).

Every microbial taxa group could be partitioned into common and rare groups. By using a bimodal model to partition the species abundance distributions, an indication of the underlying causal processes resulting in commonness and rarity can be inferred including immigration, extinction, abiotic and biotic factors. Random dispersal is said to be the predominant mechanism shaping spatial scaling of rare taxa, whilst biological mechanisms including species interactions such as competition drive common taxa (Magurran and Henderson 2003). Given that every taxa group could be partitioned into the two groups, it suggests that within every taxa group across microbial domains, communities are driven by a combination of both biological factors and neutral (dispersal) mechanisms at this spatial scale. This was supported in Chapter 2 which demonstrated that both environmental and geographical distance were important predictors of the spatial scaling of microorganisms in the soil. However within every taxa group across all three domains, there were a greater number of rare taxa than common taxa in each community, regardless of land use management. This suggests that the rare communities were driving the diversity within each community but also provides information regarding the dynamics of each community. The higher number of rare taxa within each community suggests that a large proportion of the community is contributing little to ecosystem processes and is likely to be limited by dispersal (Siqueira et al.,

2012b). Although the common taxa were relatively less abundant in bacteria and some eukaryotes (more abundant in archaea and some eukaryotes), these species are likely to account for the majority of the ecosystem functioning and community interactions and be governed by abiotic and biotic factors (Hol et al., 2010).

When assessing the whole community (common and rare combined) in each taxa group, significant differences between land use management were only evident in the eukaryote community. This supports the findings in Chapter 2, which showed that land use management was a significant driver of eukaryote spatial distribution. However by partitioning the data into common and rare groups, differences in common and rare communities within the Archaea and Bacteria domains were also evident under different land use management practices. The data showed that in each taxa group, the majority of common OTUs that contributed >1 % towards the differences in land use management, were present in both set-aside and arable sites. These common OTUs could therefore be classified as habitat generalists, which tend to be more widespread and permanent in the community and unaffected by land use management. Within every taxa group however, habitat specialists were present that occurred in either set-aside or arable conditions. These taxa typically contributed <1% to the difference in land use management and were classified as rare taxa. They are therefore likely to be transient within the community, have higher emigration and extinction rates (Magurran, 2007). Land use management is therefore predominantly impacting the rare community. This approach has enabled predictions to be made based on anthropogenic disturbance and indicated that land use management impacted the rare members of each community across all

three domains (depending on their fitness).

In all taxa groups, when comparing between land use management practices, the majority of taxa (that contributed >1 % to the differences in land use management) were consistently rare and others remained relatively common. However conversion from arable to set-aside practice did result in shifts in the common and rare taxa within each taxa group. For example certain OTUs were classified as common under arable and rare under set-aside and others common under set-aside and rare under arable. These changes could be due to stochastic colonization and extinction brought about by some kind of change. Rare taxa are likely to act as a seed bank (first proposed by Pedrós-Alió, 2006) and therefore a change in their environment (in this case brought about by land-use change) could result in shifts of the common and rare taxa within the community (Fuhrman, 2009). For instance certain rare taxa could be better adapted to the new conditions and therefore thrive under the changed environment and become common. Other common taxa may be less adapted to the new conditions and become rare, as demonstrated by the current study. Whilst significant differences in soil properties have been shown between land use management (Chapter 2), the majority of changes in community structure and composition are within the rare community. It is therefore important to consider that Bray Curtis distance methods like those used in the current study are more influenced by the dominant taxa, so shifts in common and rare taxa under set-aside and arable may be masked by both the variability in soil properties or the type of statistical test applied.

Communities typically unaffected by disturbance or perturbation are characterised as having evenly distributed communities and tend to be replaced

by more dominant communities when disturbed (Ager et al., 2010). In the current study, the anthropogenic mediated selection pressure associated with arable farming including fertilizer and pesticide application and physical disturbance such as tillage represented the perturbed environment. This study showed that conversion from a perturbed environment (arable) to a more stable environment (in this case set-aside) resulted in a more even community structure in the majority of taxa. This change in structure applied to both common and rare groups within the metacommunity. Exceptions included total archaea, basidiomycota and firmicutes where conversion from arable to set-aside resulted in a dominant community structure. As this pattern was shown in a single taxa group from each domain, further research would be recommended to determine whether this is a sampling artefact.

The current study has shown a universal, pan microbial relationship between abundance and distribution. By assessing each community according to its common and rare groups, in addition to traditional structure parameters such as dominance and evenness, the importance of land use management in shaping microbial community structure has been highlighted. Although the distance-decay approach has previously been very useful in understanding the spatial distribution of taxa (see Chapter 2), determining species area distributions (SADs) and portioning the data into common and rare groups, provides a more intimate understanding of distribution of taxa within metacommunities. These models provide a platform to discuss the mechanisms and processes shaping these communities and how microbial community structure is likely to be impacted by disturbance.

CHAPTER IV: REVEALING NOVEL PLASMODIOPHORID DIVERSITY AND DYNAMICS IN AGRICULTURAL SOIL

4.1. Introduction

Plasmodiophorida is a protistan order of parasites/symbionts of green plants and oomycetes. Plasmodiophorids include the causative agents and vectors of disease leading to economically significant crop losses (Neuhauser et al, 2014). Together with the order Phagomyxida, they make up the class Phytomyxea. Along with their sister taxon Vampyrellidae, Phytomyxea group within Endomyxa, a subphylum of Cercozoa, which is part of the super group Rhizaria (Archibald and Keeling, 2004; Bass et al., 2009; Cavalier-Smith and Chao, 2003).

Some plasmodiophorids pose a serious threat to crop production given their notorious role as disease agents (Kanyuka et al., 2003, Dixon, 2009). One of the most commercially important of these is *Plasmodiophora brassicae,* which causes club root disease in cruciferous plants such as oilseed rape (OSR). Clubroot has been shown to result in average crop losses of 10-15% on a global scale (Hwang et al., 2012, Dixon et al., 2009).

Other plasmodiophorids act as vectors of soil-borne viruses that cause plant diseases. For example *Polymyxa betae* transmits beet necrotic yellow vein virus (BNYVV), a major disease of sugarbeet, which results in substantial sugar yield loss (Desoignies et al., 2014; Hassanzadeh Davarani et al., 2014; Lemaire et al., 1988). Similarly *Polymyxa graminis* infects a multitude of crops and can transmit a number of viruses, for example soil-borne wheat mosaic virus (SBWMV), which is considered as one of the most important diseases of winter

wheat in central and Eastern USA, and soil-borne cereal mosaic virus (SBCMV) (Kanyuka et al., 2003; Ward et al., 2004).

To date studies on plasmodiophorids have aimed to understand their phylogeny, improve methods of detection and determine the mechanisms of infection in species responsible for crop losses (Burki et al., 2010; Edgcomb et al., 2011; Faggian and Strelkov, 2009). The complexity of the plasmodiophorid's six-stage life cycle, coupled with their small size (3-6 μm), makes them difficult to study. As a result, studies have targeted a select few plasmodiophorids, leaving the group as a whole grossly understudied, and as a consequence there is little understanding of the biodiversity of the group at a community level.

A recent molecular survey investigated plasmodiophorid biodiversity by focussing on host associations between plants and soil plasmodiophorids within rhizosphere and root-free soil. 81 potentially new OTUs were discovered, significantly adding to the 41 known phytomyxid (combined plasmodiophorid and phagomyxid) lineages (Neuhauser et al., 2014). This suggests that many lineages remain uncharacterised, the biological function of which is unknown.

Given the importance of plasmodiophorids as crop disease agents and viral disease vectors, understanding the diversity and distribution of the group within agricultural systems is an important goal in order to develop sustainable disease management approaches. A novel insight into the structure of plasmodiophorid communities and the factors modulating this can be deduced by partitioning the community into common and rare members.

By using sample occupancy (number of samples containing at least one OTU) in addition to abundance, common members of the plasmodiophorid communities can be identified for the first time, defined as those that are locally

abundant and frequent in the community. Rare members however are likely to be present in low abundance and exist only occasionally in the community (Magurran and Henderson et al., 2003, Hanski 1982). This is particularly relevant given that so little is known about the ecological processes acting on plasmodiophorid communities including immigration and extinction, and competition and niche partitioning in plasmodiophorids, all of which are currently unknown (Ulrich and Zalewski, 2006).

The factors that shape plasmodiophorid communities within habitats also remain to be resolved. Plant roots are zones of intense microbial activity, with rhizodeposition selecting a distinct microbial community, the biomass of which can be orders of magnitude higher than in the bulk soil (Berg and Smalla, 2009; Bonkowski, 2004). Most understanding of these factors, shaping the composition and function of the rhizosphere microbiome comes from bacterial and fungal communities, with very limited understanding of most microbial eukaryote groups. Plasmodiophorids form strong associations with their hosts, which are often green plants, but in some instances, they can also infect other parasites including heterotrophic stramenopiles, e.g. *Woronina pythii,* which infects the oomycete pathogen *Pythium* spp.

Plasmodiophorids are obligate biotrophs, requiring their hosts for the completion of their life cycle and to enable them to successfully reproduce (Neuhauser et al., 2014). Consequently plasmodiophorid community composition is likely to vary between the rhizosphere and bulk soil due to shifts in the composition of available plant and microbial hosts. Current understanding of such interactions are limited, although there is evidence for rhizosphere preferences in some plasmodiophorids. For example *Polymyxa graminis* has

been reported in both rhizosphere and bulk soil, whereas *Polymyxa beta* and *Sorosphaera viticola* tend to be localised within the rhizosphere (Neuhauser *et al.,* 2014).

Rhizosphere microbial communities can vary between plant species (Bossio et al., 1998; Garbeva et al., 2004; Wieland et al., 2001). It has recently been shown that some species of plasmodiophorid are associated with a greater variety of hosts than previously thought. Examples include *Spongospora subterranea,* which is thought to infect a multitude of hosts within *Poaceae, Brassicaceae, Leguminosae* and *Geraniacaea* as opposed to being associated with a limited number of host plants. *Polymyxa graminis* also has a wide host range including cereal crops such as barley, wheat, *Poa* spp., pearl millet, *Arabidopsis*; *Polymyxa betae,* which was considered to be a specialist pathogen of sugarbeet (Desoignies et al., 2014), has been found to also infect wheat (Smith et al., 2013).

Rhizosphere communities are well known to vary according to plant growth stage, (Houlden et al., 2008; Lupwayi et al., 1998; Smalla et al., 2001). These differences may reflect changes in root physiology and particularly the quality and quantity of rhizodeposits with young plants associated with the highest amount of exudation and the largest rhizosphere microbial populations (Chaparro et al., 2014; Garbeva et al., 2014; Gomes et al., 2001; Houlden et al., 2008). There is some evidence that plasmodiophorids may show preferences for young plants. For example, Hwang et al. (2012) found that disease severity of *Plasmodiophora brassicae* declined as plants aged (Hwang et al., 2012).

There is also evidence that agricultural rotations can change the dynamics of microbial communities inhabiting the rhizosphere, resulting in significant

shifts in community composition. Continuous cropping and growing crops in monoculture can result in a decrease in diversity of bacterial and fungal rhizosphere communities (Lupwayi et al., 1998), and in particular promote the abundance of deleterious microbes, leading to 'yield decline', which is associated with significant economic losses (Bennett et al., 2012). To date studies have focussed on understanding the role of bacteria and fungi in the development of yield decline, and despite the importance of plasmodiophorids as rhizosphere biota, their involvement has not been extensively studied (Bennett et al., 2012; Hilton et al., 2013). Continuous shortened rotations or crops grown in monoculture can increase the abundance of specific plasmodiophorid pathogens. For example *Plasmodiophorida brassicae* infestations in oilseed rape are managed with longer rotation breaks. This is because the pathogen's spores are able to remain dormant in the soil for many years, with an estimated half-life of almost 4 years (Dixon, 2014; Gossen et al., 2013; Wallenhammar, 1996). The extent to which shifts in plasmodiophorid community composition as a result of shortened crop rotations or growth of crops in monoculture contribute to yield decline remains unclear, since studies to date have considered only the contributions of bacterial and fungal communities (Garbeva et al., 2004, Bennett et al., 2011, Hilton et al., 2013).

To summarise, plasmodiophorids are important rhizosphere-acting plant obligate biotrophs in which diversity is largely uncharacterized, and the factors, controlling community composition, are unclear. With their potential to cause significant health impacts on a variety of crops there is a need to understand which factors determine community diversity and composition, in order to devise approaches to manage and reduce their economic impacts.

In this study an ecologically well-defined and agriculturally relevant scenario was used to determine the drivers of plasmodiophorid communities. The aims were to i) test newly-designed PCR primers for the specific amplification of plasmodiophorid 18S rDNA and determine the performance of these primers in the characterization of plasmodiophorid communities ii) determine the key factors that shape plasmodiophorid communities within an agricultural system iii) identify the common and rare taxa within the community and their seasonal and spatial dynamics; and iv) investigate whether there is a relationship between dynamics of plasmodiophorid communities and crop yield. Roche 454 pyrosequencing was used to investigate the factors shaping plasmodiophorid community composition. This method reduces taxonomic biases associated with physical screening for individual known pathogens and circumvents sampling limitations associated with detecting cryptic and poorly known microbes thereby enabling the detection of rare species and increasing accuracy of diversity estimates.

4.2. Materials and Methods

DNA samples for this investigation were provided by Dr Sally Hilton from the experiments described in Hilton et al., (2013).

4.2.1. Experimental design and sampling

To determine plasmodiophorid community differences between the rhizosphere and bulk soil under different plant hosts, a total of 160 samples from rhizosphere and bulk soil of oil seed rape OSR (cv. Winner) and winter wheat (cv. Brompton), taken over a two-year period from a field trial in East Anglia,

UK (52° 33' N and 1° 2' E) were analysed. To assess the impact of rotation and plant growth stage on plasmodiophorid communities, 40 samples were collected from five OSR and wheat rotations (Table 4.1) during the fourth year of the trial in June 2007, and a further 40 samples were taken in the fifth year of the field trial in November 2007, March 2008 and June 2008. The soil sampled was a sandy clay loam with a pH of 6.6 and available P, K, Mg and SO_4 of 32.4, 111, 28, 30.6 mg kg^{-1} respectively.

Table 4.1 Cropping history of rotations in years 1-4 and year 5. Rhizosphere and bulk soil samples were collected in June 2007 of the fourth year and in November 2007, March 2008 and June 2008 of the 5th year. Samples from alternate OSR and OSR – 2-year- gap in year 5 were collected from different plots to those in year 4 (O = OSR, W = wheat).

	Year of trial	
Rotation	**1-4**	**5**
Continuous OSR	OOOO	OOOO
Continuous Wheat	OWWW	OWWW
Alternate OSR	WOWO	OWOWO
OSR - 2 Year Gap	OWWO	WOWWO
Virgin OSR (Years 1-4) Wheat after OSR (Year 5)	WWWO	WWWOW

As part of this trial, the field was ploughed and pressed each season before establishment. Drilling occurred at the beginning of September for OSR, mid September for the first winter wheat, followed by mid-October for subsequent wheat (Hilton et al., 2013). Local commercial best practice was adhered to with regards to the application of fertiliser; 200 kg ha^{-1} N and 30 kg ha^{-1} S.

In June 2007 (pre-harvest), samples were collected from continuous OSR, continuous wheat, alternate OSR (last rotation was OSR), OSR with a 2 year gap (last rotation was OSR) and virgin OSR. November 2007 (seedling stage), March 2008 (stem extension) and June 2008 (pre-harvest) samples were collected in the

5th year of the trial and consisted of the same five rotations, which were sampled in June 2007.

Each rotation (within each time point) was replicated four times within randomised 24 m x 6 m plots. Additionally each plot was sub-divided into three equal sub-plots, where the central sub-plot was used for yield data. Bulk soil and rhizosphere samples were collected from each of the sub-plots to give 6 pooled plant or bulk soil samples. 5 mm sections of fine root were extracted, shaken free of loose soil and 0.5 g of material, which included closely adhering soil, was retained as the rhizosphere sample. A 3 mm sieve was used to obtain 0.5 g of soil for the bulk soil samples. The DNA was extracted using a FastDNA® SPIN kit for soil (MP Biomedicals LLC, UK) following the manufacturer's guidelines for all steps, except to use a Mini Beadbeater-8 cell disrupter for a 3 minute period in place of a FastPrep® machine) (Biospec products, Inc., USA). 10 µL of the original DNA was diluted with 40 µL of sterile water to give a 1:5 diluted stock solution.

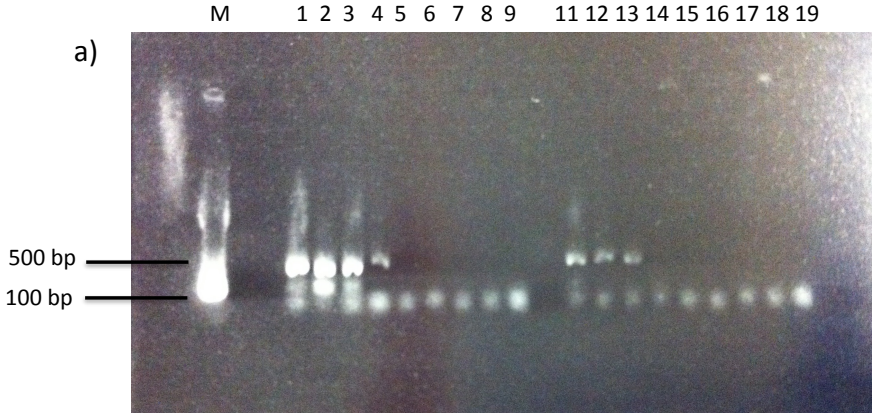
4.2.2.Primer selection and PCR design and testing

A reference pan-eukaryote alignment (Glücksman et al., 2010) was used to design the new plasmodiophorid-specific primer pair 1301f (5'-GATTGAAGCTCTTTCTTGATCACTTC-3') and 1801gram (5'-ACGGAAACCTTGTTACGACTTC-3), which amplify the V6-V9 region of the 18S rRNA gene (18S rDNA).

An appropriate annealing temperature needed to be determined for the 1301f and 1801gram primer pair in order to obtain a balance between specificity, phylogenetic coverage across the known diversity of plasmodiophorids, and the number of positive reactions across the samples. The optimum temperature for

amplification was determined with a PCR gradient based on the following temperatures; 61.3 °C, 62.3 °C, 63.5 °C, 64.8 °C, 66.0 °C, 66.8 °C, 67.4 °C, 67.9 °C and 68.0 °C. 12 µl DNA (3μl x 4 replicates per sample) was taken from four of the OSR/wheat samples from both soil and rhizosphere at random along with a positive control (pure *Plasmodiophora brassicae* DNA from Chinese cabbage, in Kematen, Tirol, Austria) with the aim of determining at which temperature the primers amplified best.

The PCR reaction (50 μl) comprised of 47 μl MegaMix (Microzone Limited, UK), 1 μl eDNA (at 1:5 dilution from original extraction), 1 μl of forward primer (10 μM) and 1 μl of reverse primer (10 μM). The thermocycling protocol included denaturation at 95 °C for 5 minutes, then 35 cycles of 95 °C for 30 seconds, annealing temperature (gradient detailed above) for 40 seconds, 72 °C for 2 minutes, then the final extension was 72 °C for 7 minutes and indefinite hold at 4 °C. The PCR products were run on an agarose gel at 80 V for 30 minutes, stained with GelRed™ and viewed under UV light to assess the bands. The PCR reactions were most successful (indicated with a single, clear band appearing at around 500 bp) at both of the lower temperatures (63.5 °C, 64.8 °C) for all samples including the control (Figure 4.1).



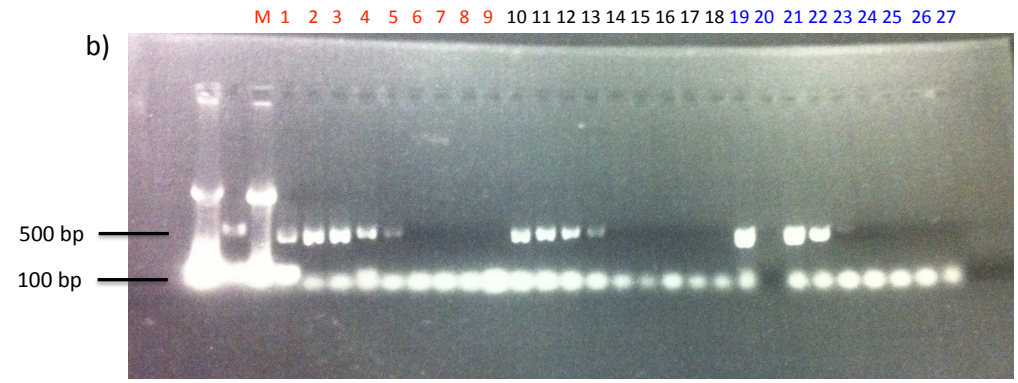


Figure 4.1 Gel 1 (left): Ladder (M), followed by Continuous Rhizosphere (OSR), lanes 1-9, then Continuous Bulk (OSR), lanes 11-19. Gel 2 (right): Ladder (M), followed by continuous rhizosphere (wheat) lanes 1-9 in red, continuous bulk (wheat) lanes 10 – 18 in black and a positive control, lanes 19 – 27 in blue. There were no gaps in between samples (with 9 temperatures tested for each). The temperatures run from left to right include 61.3, 62.3, 63.5, 64.8, 66.0, 66.8, 67.4, 67.9, 68.0 °C. Samples were run with GeneScan 1200™ LIZ ladder (Applied Biosystems, UK).

4.2.3. 454 Amplicon pyrosequencing to determine plasmodiophorid diversity within bulk soil and rhizosphere of OSR and wheat

160 separately MID-tagged amplicon libraries were 454-sequenced across two GS Junior plates (Micropathology Ltd, Coventry, UK). 1 μl of the extracted DNA at a final reaction concentration of 0.2 μM was used with MyTaq HS 5x mastermix from Bioline. All 160 samples were run at 62 °C (given initial low amplicon yields at 63.5 °C, 64.8 °C obtained by Micropathology). The cycling parameters for the first round were 35 cycles of PCR with 95 °C for 2 min, 95 °C for 20s, 62 °C for 20 s and 72 °C for 20 s. Following amplification, 1 μl of first round PCR was added to a second round PCR using the same reagents, 30 cycles of PCR with 95 °C for 2 min, 95 °C for 20 s, 55 °C for 20 s and 72 °C for 20 s. This enabled fusion primer sequences to be added and used a universal tail based on M13. Sample cleanup was performed using AMPure XP beads at a ratio of 0.6:1 and samples were equimolar pooled following quantitation using a Shimadzu MultiNA. The pool was then diluted to 1 molecule per μl. The sequencing protocol was performed entirely according to the manufacturers

protocol with no deviations (libL emPCR kit) (Roche 454 Sequencing system software manual, v 2.5p1). Amplicon processing was applied to the raw data by the onboard software (Micropathology, Coventry, UK). Pyrosequencing yielded approximately 120 000 sequences per plate.

4.2.4. Bioinformatic processing of 454 sequence data

QIIME 1.8.0 software (Caporaso et al., 2010) was used to filter the raw sequence files according to a quality score of 25, sequence length between 200 and 1000 bp, zero primer mismatches, up to six homopolymers, zero ambiguous bases and a maximum of 1.5 barcode errors. The fasta files were de-multiplexed and partitioned based on sample identifiers. The trimmed sequences were then incorporated into the UParse pipeline which uses a greedy OTU clustering algorithm based on sequence abundances (Edgar, 2013) to remove singletons (OTUs with less than 2 sequences across all samples). The amplicon reads were clustered into OTUs (at 97% sequence similarity) resulting in 24 OTUs, which were reintegrated into QIIME for taxonomic assignment to a custom phytomyxid database using BLAST (provided by Sigrid Neuhauser).

4.2.5. Statistical analysis

The 24 OTUs were aligned with as many plasmodiophorid sequences as possible from Neuhauser et al. (2014); i.e. those that completely overlapped with the 3' SSU rDNA fragment amplified in the present study. The alignment was created using the 1-ins-I algorithm in MAFFT (Katoh et al., 2002) and then refined by eye. Maximum Likelihood phylogenetic analyses was performed using RAxML

BlackBox (GTR model with CAT approximation); bootstrap values were mapped onto the tree with the highest likelihood value. RAxML analyses (Stamatakis, 2006) were carried out via the Cipres Science Gateway Portal (Miller et al., 2010).

Using the phylogenetic tree, potential outliers were identified and cross-referenced using BLAST (Altschul et al., 1990) on NCBI's nr/nt database in Genbank, showing that three OTUs were from plants and were therefore removed from further analysis. To avoid over-estimating biodiversity, OTU classification was manually checked by aligning each representative OTU sequence against the reference alignment, based on the phylogenetic placement of sequences to known lineages. OTUs that were different from a sequence in the reference database by three or more nucleotide positions in two or more variable regions of the amplicon were considered distinct lineages; those more similar to reference sequences were considered to belong to the reference lineage.

4.2.5.1. Influence of rhizosphere and bulk soil of OSR and wheat, rotation and plant growth stage on plasmodiophorid community composition

To analyse the compositional differences between the plasmodiophorid communities under the different experimental variables, the ordination technique non-metric multidimensional scaling (NMDS) was applied to the data. These variables were rhizosphere and bulk soil, crop (i.e. OSR and wheat), rotation and growth stage. Based on untransformed proportional abundance data, patterns in community composition could be visualised using the Bray-Curtis distance matrix. NMDS was used as it makes no assumptions regarding species distributions, (i.e. it does not require Gaussian distribution), which makes it a

robust choice for the detection of ecological patterns (Hillstrom and Lindroth, 2008; Nekola, 2003). ANOSIM *R* and *P* values were generated using the Bray-Curtis measure of similarity. *R* scales from +1 to -1 were used where +1 indicates that all the most similar samples are within the same groups. *R* equals 0 occurs if the high and low similarities are perfectly mixed and bear no relationship to the group. A value of -1 indicates that the most similar samples are all outside of the groups (Carson et al., 2007). An associated significance level (P), based on 999 permutations was also given (Hilton et al., 2013). To determine which OTUs made the largest relative contribution (%) to the observed assemblage differences among material type, host, rotation and plant growth stage, similarity percentage analyses (SIMPER) were performed. SIMPER analysis compares average abundances and determines the percentage contribution that each OTU makes to the average dissimilarity between groups of interest. It then lists them in order of importance in discriminating the two sample sets (Axelsson et al., 2011). All non-parametric multivariate analyses including NMDS, ANOSIM and SIMPER analysis were conducted using Primer-E statistical software (PRIMER, version 6, Primer-E, Plymouth Marine Laboratory, UK).

Each OTU was colour-coded based on its mean relative abundance. OTUs highlighted in red indicate that the OTU had an abundance between 51 % and 100 %, orange between 26-50 %, yellow between 11-25 %, green between 2-10 % and blue between 0-1 % abundance. In order to visualize in which treatments (rhizosphere OSR and wheat, bulk soil OSR and wheat) OTUs were highly abundant (compared to those with a lower abundance), species abundance distributions (SAD) were fitted to the data (Magurran and

Henderson, 2003; Preston, 1960). The number of sites occupied (distribution/persistence) and the percentage mean relative abundance of each taxonomic group was calculated in each treatment and plotted on scatter graphs as detailed in Chapter 3. To further understand the impact of different variables on the plasmodiophorid communities, OTUs were categorised as common or rare, using mean relative abundance in addition to persistence (frequency of occurrence of each OTU across all samples within a group). The data were sorted by persistence and the upper quartile persistence values were calculated. OTUs that occurred in the 75 % upper quartile of the dataset (i.e. occurred in 75 % or more of the sites) were classified as common, whereas those that did not, were classified as rare.

4.2.5.2. Influence of rhizosphere and bulk soil of OSR and wheat, rotation and plant growth stage on plasmodiophorid community structure

Structure (defined as dominance and evenness) was determined by plotting the rank-abundance distributions (based on the relative mean abundance). The slope value (determined from a simple linear regression mode) provided an indication of evenness or dominance, enabling community structure to be assessed. The t-distribution method was then used to determine whether the regression line slopes was significantly affected by experimental variables, as described in Fowler et al., (1998).

4.3. Results

4.3.1. Novel plasmodiophorid diversity and phylogenetic analyses

Excluding plant OTUs (representing 0.26% of the total sequences), a total of 196,196 processed sequences, with an average read length of 404.3 bp were obtained across all 160 samples, (1226 sequences per sample on average). Based on clustering with UParse and cross-referencing against an alignment and phylogenetic tree, 24 OTUs (average of 6539.9 reads assigned to each OTU) were identified across all samples. A rarefaction curve was plotted and showed that diversity plateaued at around 100 sequences (Figure 4.2 a and b).

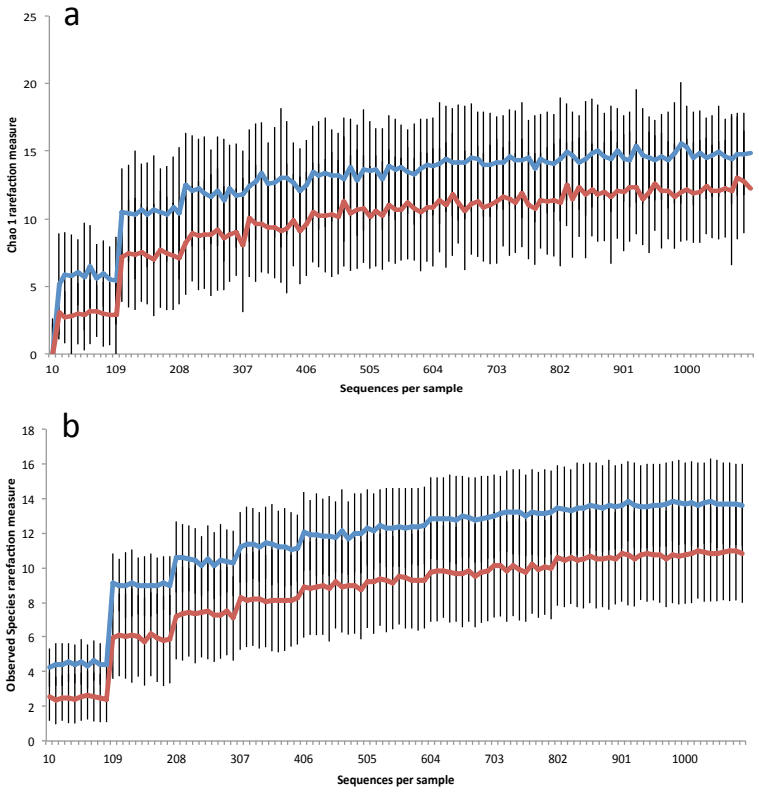


Figure 4.2 a and b: alpha diversity estimates Chao1 (a) and Observed Species (b). Blue line represents the bulk soil and the red line represents the rhizosphere

Four of the OTU sequences were classified as characterised plasmodiophorids, based on the criteria described in the Methods: *Polymyxa graminis* (OTU 2), *Polymyxa beta* (OTU 11), *Spongospora subterranea nasturtii*

(OTU 1) and *Woronina spp.* (OTU 5). A further three OTUs were identical or very similar to plasmodiophorid lineages from environmental sequences; FN690466_clone_1D9_Plasmo (OTU 22), EF024345_clone_Elev_18s_763_Plasmo (OTU 12), and EU910610_clone_D20_Plasmo (OTU 6). The remaining 15 OTUs were previously unknown, potentially representing novel taxa. OTU 4 grouped in the same clade as *Plasmodiophora brassicae,* to which it appears to be a close relative.

Across all samples, the most abundant lineage was OTU 1, which had an identical 18S rDNA sequence to *Spongospora subterranea nasturtii* in the fragment sequenced. OTU 1 accounted for 43% of the total sequences across all data (Fig. 2). The second most abundant was OTU 23 (a novel lineage most closely related to *Sorosphaera veronicae*) representing 20.8% of the total sequences and the third was OTU 2 (identical 18S rDNA sequence to *Polymyxa graminis)*, representing 11.6% of the total sequences. 98.11% of the total number of sequences across all samples were derived from eight OTUs (Figure 4.3).

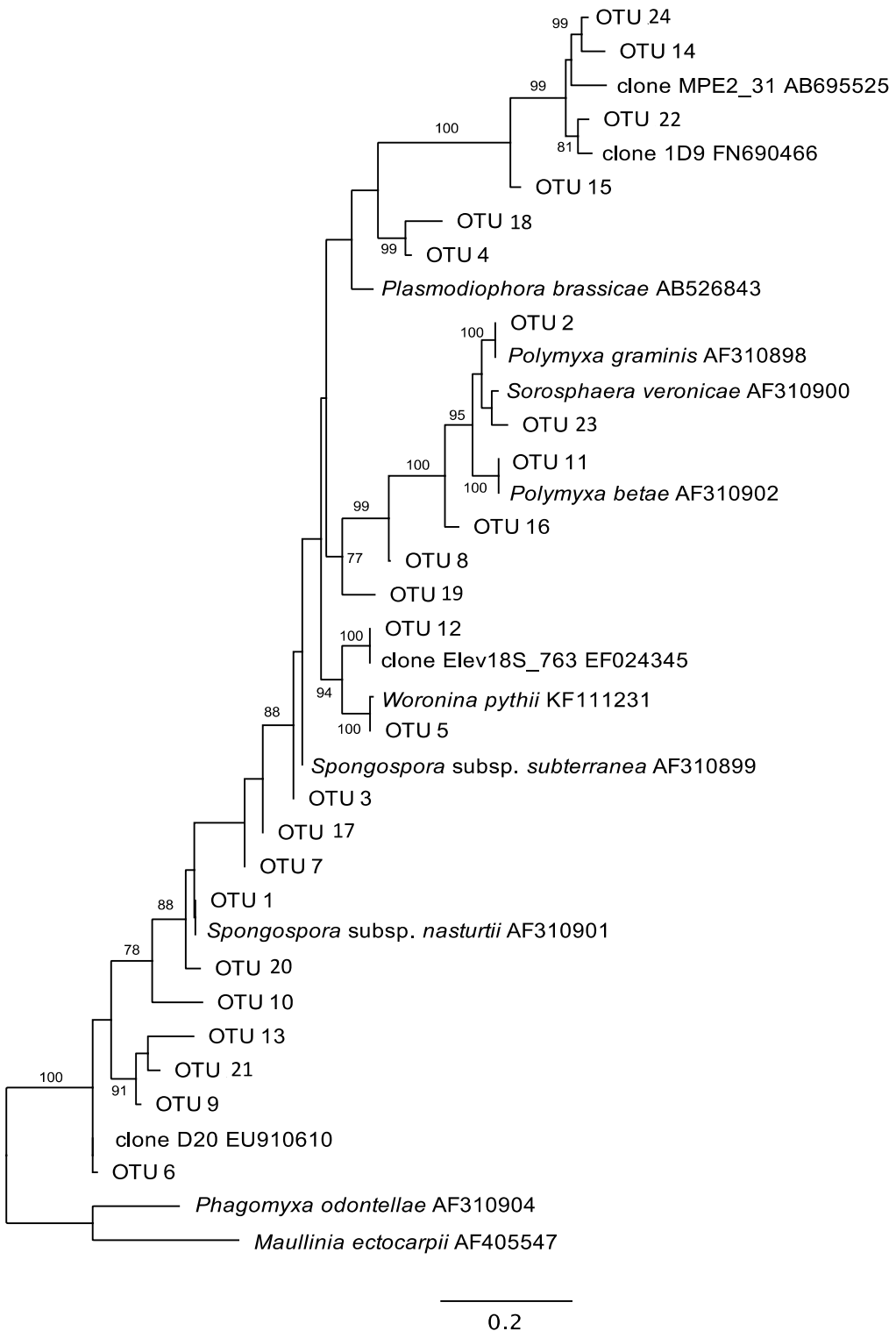


Figure 4.3: Phylogenetic tree of 24 OTUs from the rhizosphere and bulk soil, complemented with sequences from Neuhauser *et al*., (2011) from the alignment and phagomyxid outgroup. The reference alignment was analyzed by RAxML BlackBox (GTR model with CAT approximation); bootstrap values were mapped onto the tree with the highest likelihood value. RAxML analyses (Stamatakis, 2006) were carried out via the Cipres Science Gateway Portal (Miller et al., 2010).

4.3.2. The influence of the rhizosphere and bulk soil on plasmodiophorid communities

The relative distribution of plasmodiophorid communities was compared within the rhizosphere and bulk soil. Plasmodiophorid assemblages in the rhizosphere were found to be significantly different from those in bulk soil in all seasons (June 2007; P = 0.001, r = 0.30; November 2007; P = 0.23, r = 0.002; March 2008; P = 0.34, r = 0.001; June 2008; P = 0.001, r = 0.49) (Figure 4.4, Table 4.2).

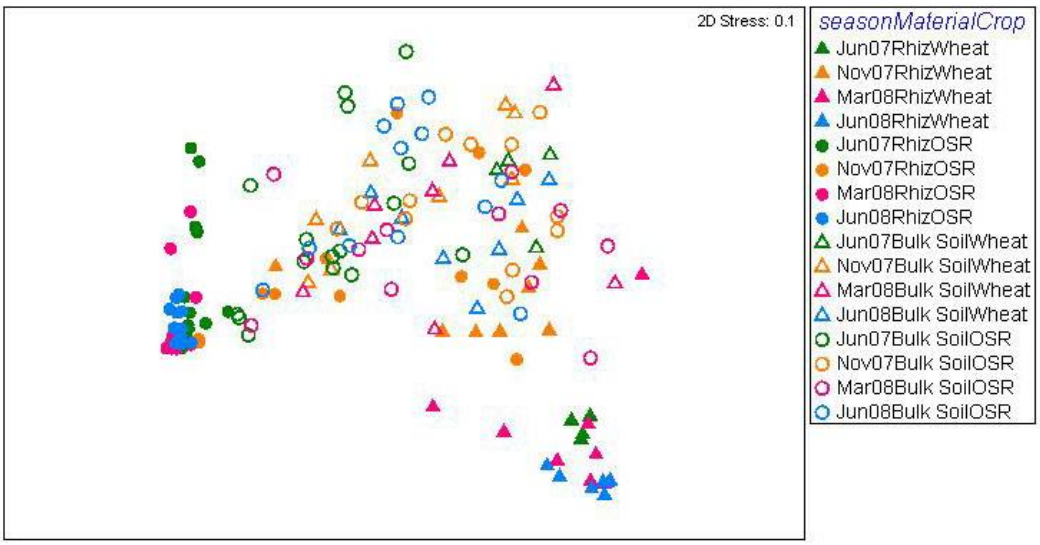


Figure 4.4: Non-metric multi-dimensional scaling plot showing *differences* in community composition in the rhizosphere (Rhiz) and bulk soil (Bulk Soil) of OSR and wheat hosts at different plant growth stages; June 2007 (Jun07), November 2007 (Nov07), March 2008 (Mar08) and June 2008 (Jun08).

Table 4.2: ANOSIM statistics comparing the significance of differences in plasmodiophorid communities within bulk soil and rhizosphere OSR and wheat host plants over time.

Treatments compared	Jun-07		Nov-07		Mar-08		Jun-08	
	R	P	R	P	R	P	R	P
Rhizosphere vs Bulk Soil	0.30	0.001*	0.23	0.002*	0.34	0.001*	0.49	0.001*
Rhizosphere: OSR vs Wheat	1.00	0.001*	0.14	0.07	1.00	0.001*	1.00	0.001*
Bulk Soil: OSR vs Wheat	0.53	0.001*	0.08	0.14	-0.07	0.79	-0.11	0.92

Of the 24 OTUs detected, 21 were found in both rhizosphere and bulk soil samples. However OTUs 16, 18 and 19 were only found in the rhizosphere.

To visualize the species abundance distributions (SAD) within the plasmodiophorid metacommunities, percentage mean relative abundance of OTUs and local community occupancy was plotted. There was a significant positive relationship between percentage mean relative abundance and occupancy for all treatments (rhizosphere OSR and wheat, bulk soil OSR and wheat) under each growth stage encompassing all three domains for both arable and set-aside land use management (P = 0.002 - <0.001) (Figure 4. 5).

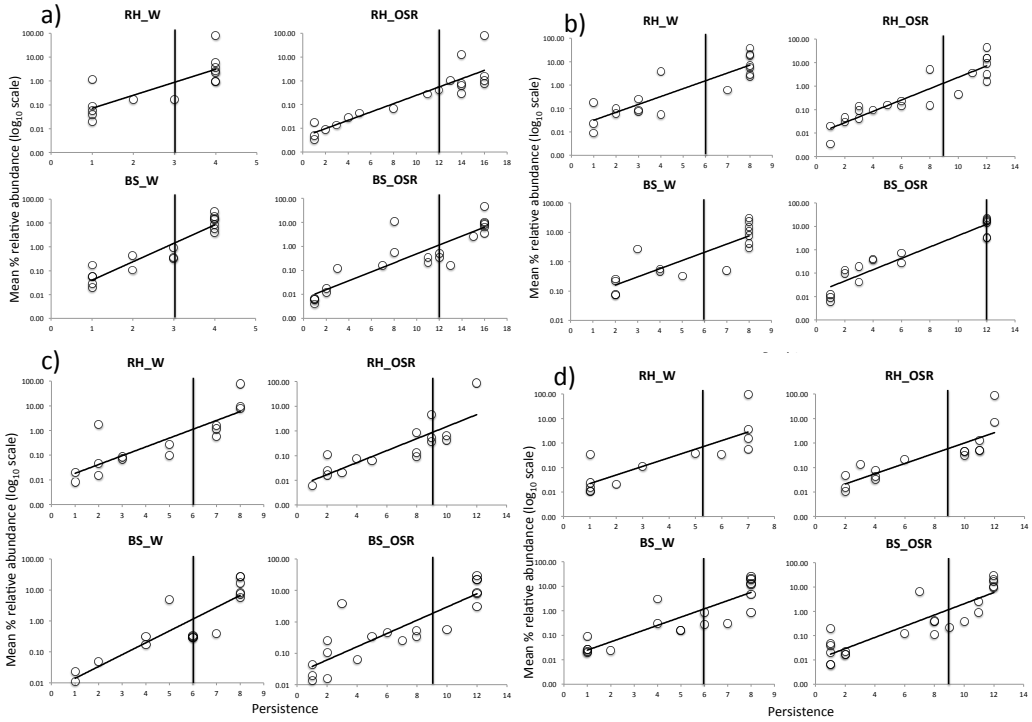


Figure 4.5: Distribution of plasmodiophorids across different treatments (RH_W – rhizosphere wheat; RH_OSR – rhizosphere OSR; BS_W – bulk soil wheat; BS_OSR – bulk soil OSR), under different growth development stages – June 2007 (a), November 2007 (b), March 2008 (c) and June 2008 (d). Given is occupancy (number of samples for which each taxa was observed), plotted against percentage mean relative abundance (log_{10} scale) across all samples. All relationships are significant at P=<0.001. The vertical line represents a 75% cut-off whereby OTUs occurring in 75% or more of the samples (to the right of the line) are classified as common and those that did not are classified as rare.

Across all plant growth stages, OTUs 1, 2, 4, 5, 6 and 23 were classified as common (see Methods) in both rhizosphere and bulk soil in either or both plant type (Figure 4.6). OTU 20 was classified as common in rhizosphere and bulk soil in June 2007 and November 2007 only. OTUs 11, 12, 13, 14,15, 16, 17, 18, 19, 21 and 22 were consistently classified as rare under both rhizosphere and bulk soil conditions. No OTU was consistently classified as rare in rhizosphere and common in bulk soil or vice versa.

The same common taxa were found in both rhizosphere and bulk soil, however the significant differences in the structure of communities were due to differences in their relative abundance (Figure 4.6). Whilst SIMPER analyses showed that OTUs 1, 2, 4, 5, 3 and 23 contributed most strongly to assemblage differences between rhizosphere and bulk sample, the rhizosphere was dominated by OTUs 1 and 23. OTU 1 and OTU 23 had a higher mean relative abundance in the rhizosphere in June 2007 (OTU 1 64.55 %, OTU 23 16.93 %), November 2007 (OTU 1 34.79 %, OTU 23 24.70 %), March 2008 (OTU 1 56.00 %, OTU 23 30.74 %) and June 2008 (OTU 1 57.27 %, OTU 23 34.61 %) (Table 4.3). However, bulk soil communities were more even in structure with OTUs 1, 2, 4, 5, 23 each contributing to the differences between the rhizosphere and bulk soil across all plant growth stages (Table 4.3). OTU 2, 4 and 5 had a greater mean relative abundance in bulk soil compared to rhizosphere. OTU 2 (*P.graminis*) had a mean relative abundance of 20.15 % in bulk soil compared to 0.52 % in rhizosphere in June 2008, 23.03 % in bulk soil and 16.59 % in rhizosphere in November 2007, 23.98 % in bulk soil, 4.08 % in rhizosphere in March 2008 and 13.78 % in bulk soil, 1.82 % in rhizosphere in June 2007 (Table 4.3). Similarly OTU 4 (novel lineage closely related to *P.brassicae)* had a

relative abundance of 11.84 % in bulk soil compared to 0.41 % in rhizosphere in June 2008, 16.57 % in bulk soil and 3.13 % in rhizosphere in November 2007, 8.41 % in bulk soil, 0.69 % in rhizosphere in March 2008 and 11.67 % in bulk soil, 0.96 % in rhizosphere in June 2007. Finally OTU 5 (*Woronina spp*.,) had a relative abundance of 10.30 % in bulk soil compared to 0.33 % in rhizosphere in June 2008, 13.14 % in bulk soil and 8.28 % in rhizosphere in November 2007, 7.86 % in bulk soil, 0.52 % in rhizosphere in March 2008 and 7.99 % in bulk soil, 1.56 % in rhizosphere in June 2007 (Table 4.3). Figure 4.6 shows that OTU 6 (EU910610_clone_D20_Plasmo) was also more abundant in bulk soil compared with rhizosphere (with one exception where it was marginally more abundant in the rhizosphere of OSR plants in November 2007). Furthermore OTUs 8, 9, 10, 11 and 20 were also more abundant in bulk soil than rhizosphere irrespective of host plant, but in very low relative abundances.

Jun-07	Treatment			
OTU ID	RH_W	RH_OSR	BS_W	BS_OSR
1	2.62*	80.04*	8.72*	45.32*
2	6.02*	0.77*	30.36*	9.63*
3	1.20	12.97*	0.06	11.02
4	2.26*	0.63*	16.13*	10.55*
5	3.6*	1.01*	13.92*	6.51*
6	0.99*	0.28	5.82*	2.52*
7	0.00	0.41	0.00	0.16*
8	0.17	0.04	0.44	0.36
9	0.02	0.03	0.37	0.53
10	0.06	0.07	0.33	0.35
11	0.17	0.02	0.92	0.21
12	0.09	0.00	0.18	0.01
13	0.00	0.00	0.02	0.01
14	0.00	0.01	0.00	0.00
15	0.00	0.00	0.06	0.02
16	0.00	0.00	0.00	0.00
17	0.04	1.07*	0.03	0.54
18	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.01
20	0.91*	1.60*	4.12*	3.64*
21	0.00	0.01	0.00	0.12
22	0.00	0.01	0.00	0.00
23	81.69*	0.74*	18.43*	8.32*
24	0.00	0.29*	0.11	0.16

Nov-07	Treatment			
OTU ID	RH_W	RH_OSR	BS_W	BS_OSR
1	20.90*	44.13*	30.26*	19.51*
2	17.93*	15.70*	23.49*	22.78*
3	3.96	5.18	2.68	0.41
4	5.47*	1.57*	14.95*	17.71*
5	6.81*	9.27*	11.07*	14.56*
6	2.90*	3.73*	4.25*	3.21*
7	0.10	0.04	0.07	0.01
8	0.09	0.23	0.58	0.28
9	0.08	0.05	0.47	0.36
10	0.62*	0.44*	0.50*	0.75
11	0.01	0.15	0.33	0.14
12	0.18	0.00	0.22	0.00
13	0.06	0.14	0.00	0.00
14	0.00	0.03	0.00	0.01
15	0.02	0.02	0.00	0.10
16	0.00	0.00	0.00	0.00
17	0.25	0.16	0.25	0.00
18	0.00	0.00	0.00	0.00
19	0.00	0.00	0.00	0.00
20	2.35*	3.12*	2.86*	3.30*
21	0.00	0.09	0.00	0.20
22	0.00	0.10	0.00	0.00
23	38.22*	15.71*	7.94*	16.64*
24	0.06	0.15	0.07	0.04

Mar-08	Treatment			
OTU ID	RH_W	RH_OSR	BS_W	BS_OSR
1	8.18*	87.88*	27.20*	29.61*
2	9.57*	0.43*	27.44*	21.67*
3	1.79	4.40	4.80	3.89
4	1.63*	0.06	8.29*	8.49*
5	1.16*	0.10	7.49*	8.10*
6	0.55*	0.52	5.90*	3.14*
7	0.02	0.82	0.32	0.25
8	0.26	0.02	0.40*	0.46
9	0.07	0.00	0.29	0.56*
10	0.08	0.02	0.33	0.34
11	0.01	0.00	0.30	0.34
12	0.00	0.00	0.01	0.00
13	0.00	0.01	0.00	0.10
14	0.00	0.07	0.00	0.02
15	0.01	0.00	0.05	0.01
16	0.00	0.00	0.00	0.00
17	0.00	0.11	0.00	0.04
18	0.00	0.00	0.00	0.02
19	0.00	0.00	0.00	0.00
20	0.10	0.13	0.32	0.54
21	0.00	0.00	0.02	0.06
22	0.00	0.00	0.00	0.00
23	76.29*	0.38	16.68*	22.11*
24	0.05	0.65*	0.17	0.26

Jun-08	Treatment			
OTU ID	RH_W	RH_OSR	BS_W	BS_OSR
1	3.49*	88.64*	24.23*	29.35*
2	0.57*	0.49*	18.94*	20.96*
3	0.35	7.32*	3.03	6.25
4	0.35*	0.44*	13.44*	10.77*
5	0.38	0.30*	11.68*	9.38*
6	1.47*	0.21	4.80*	2.58*
7	0.01	1.29*	0.30	0.12
8	0.01	0.05	0.87*	0.38
9	0.02	0.01	0.84	0.91*
10	0.02	0.08	0.31*	0.22
11	0.00	0.00	0.16	0.39*
12	0.00	0.00	0.03	0.02
13	0.00	0.00	0.02	0.02
14	0.00	0.14	0.00	0.05
15	0.00	0.00	0.00	0.01
16	0.00	0.00	0.00	0.19
17	0.00	0.04	0.03	0.02
18	0.00	0.00	0.09	0.00
19	0.00	0.00	0.00	0.04
20	0.02	0.03	0.28	0.40
21	0.00	0.00	0.02	0.01
22	0.00	0.00	0.00	0.00
23	93.20*	0.43*	20.75*	17.81*
24	0.11	0.51*	0.16	0.11

Figure 4.6 Persistence overlaid with relative abundance for June 2007, November 2007, March 2008 and June 2008 for each treatment; RH_W (rhizosphere wheat), RH_OSR (rhizosphere OSR), BS_W (bulk soil wheat), BS_OSR (bulk soil OSR). OTUs highlighted in red indicate that the OTU had a relative abundance between 51% and 100%, orange between 26-50%, yellow between 11-25%, green between 2-10% and blue between 0-1% abundance. OTUs with an asterisk were classified as common based on the OTU occupying more than 75% of the sites.

Table 4.3 Similarity percentage analysis (SIMPER) of OTUs most responsible for differentiating plasmodiophorid communities within the rhizosphere and bulk soil

Season	OTU	% Contribution to difference	Cumulative contribution to difference (%)	Mean relative abundance in Rhizosphere	Mean relative abundance in Bulk Soil
Jun-08	1	33.58	33.58	57.27	27.30
	23	27.49	61.07	34.61	18.46
	2	13.89	74.97	0.52	20.15
	4	8.06	83.03	0.41	11.84
	5	7.03	90.06	0.33	10.30
Nov-07	1	28.09	28.09	34.79	23.84
	23	20.88	48.98	24.70	13.14
	2	14.92	63.90	16.59	23.03
	4	14.31	78.21	3.13	16.57
	5	7.36	85.56	8.28	13.14
	3	5.61	91.17	4.70	1.32
Mar-08	1	34.62	34.62	56.00	28.65
	23	25.67	60.59	30.74	19.93
	2	17.63	78.21	4.08	23.98
	4	5.88	84.10	0.69	8.41
	5	5.66	89.76	0.52	7.86
	3	4.68	94.44	3.36	4.26
Jun-07	1	36.76	36.76	64.55	38.00
	23	18.63	55.39	16.93	10.35
	3	12.24	67.63	10.62	8.83
	2	10.43	78.06	1.82	13.78
	4	9.10	87.16	0.96	11.67
	5	5.56	92.72	1.56	7.99

4.3.3. Influence of OSR and wheat on plasmodiophorid communities within the rhizosphere and bulk soil

Within the rhizosphere, crop type significantly impacted plasmodiophorid community composition. Different communities assembled according to the type of crop in all plant growth stages, with the exception of the November 2007 samples ($P = 0.001$, $r = 1$, based on 999 permutations) (Figs 4.4; 4.7 – 4.10 Table 4.2).

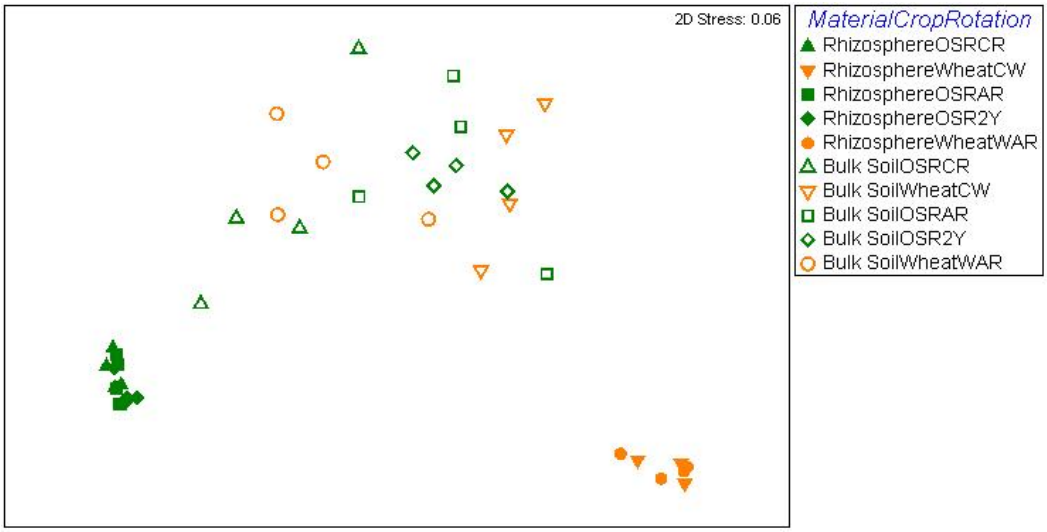


Figure 4.7 Non-metric multi-dimensional scaling plot showing rhizosphere (solid) and bulk soil (open) 18S rDNA communities obtained from different rotations of OSR (green) and wheat (orange) for June 2008.

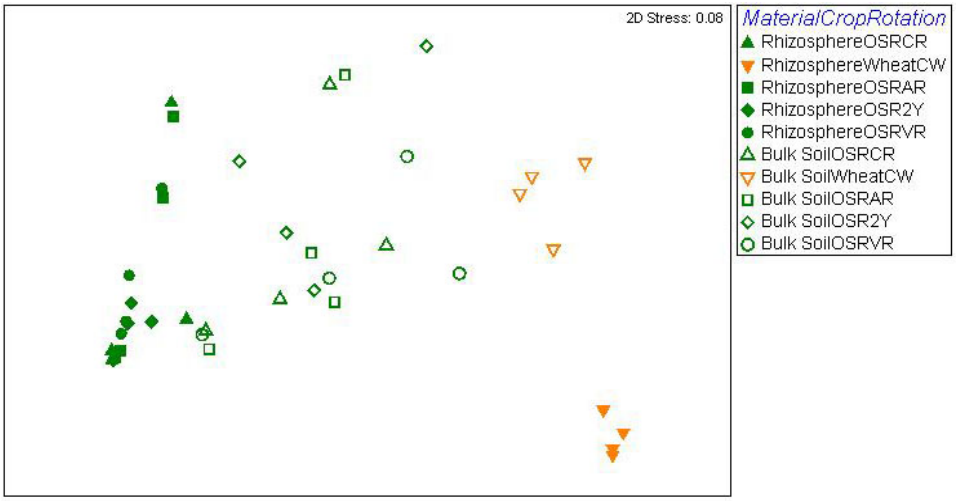


Figure 4.8 Non-metric multi-dimensional scaling plot showing rhizosphere (solid) and bulk soil (open) 18S rDNA communities obtained from different rotations of OSR (green) and wheat (orange) for June 2007.

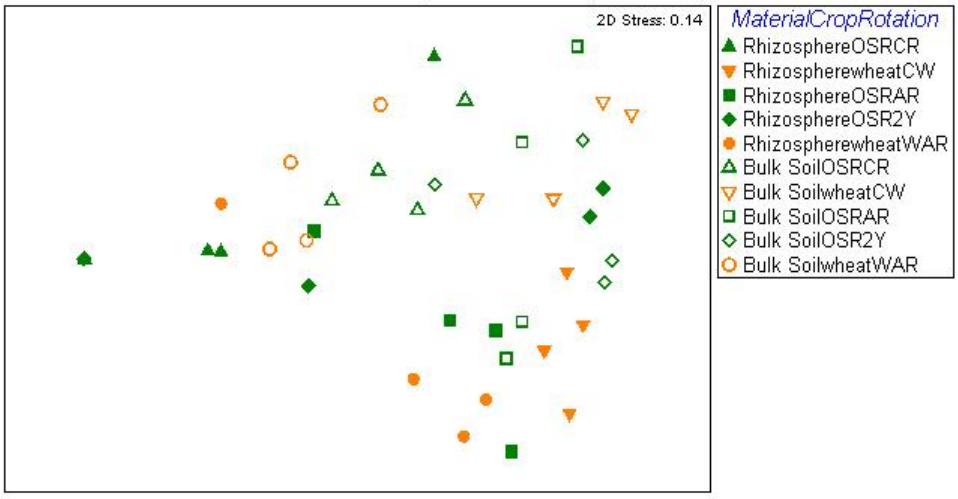


Figure 4.9 Non-metric multi-dimensional scaling plot showing rhizosphere (solid) and bulk soil (open) 18S rDNA communities obtained from different rotations of OSR (green) and wheat (orange) for November 2007.

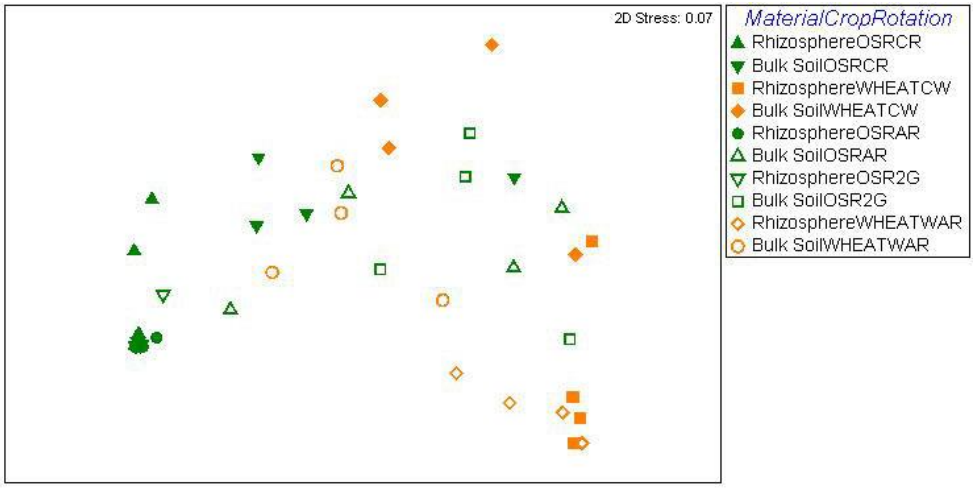


Figure 4.10 Non-metric multi-dimensional scaling plot showing rhizosphere (solid) and bulk soil (open) 18S rDNA communities obtained from different rotations of OSR (green) and wheat (orange) for March 2008.

Plasmodiophorid communities in November 2007 were different to all the other time points but were the same in wheat and OSR rhizosphere. Within the rhizosphere, the community composition was similar in each of the time points following the seedling stage in November. Furthermore although there were consistent differences between host plants (with the exception of Nov 2007), i.e. wheat and OSR, there were no community differences between June 2007 and June 2008 for either plant. Given that these communities were sampled a year apart and from different plots, this indicates that consistent, predictable communities had developed within the rhizosphere of both OSR and wheat. Within the bulk soil however, crop type only significantly impacted plasmodiophorid community distribution in June 2007 ($P = 0.001$, $r = 0.53$), Table 4.2, Figure 4.6.

OSR rhizosphere communities were dominated by OTU 1 (*Spongospora subterranea nasturtii*) and categorised within the 51 % - 100 % abundance range (orange in Figure 4.5). Specifically OTU 1 accounted for 88.64 % of the 454

reads across OSR rhizosphere samples in June 2008, 87.88 % in March 2008 and 80.04 % in June 2007 (Figure 4.6). OTU 1 was shown by SIMPER analysis to account for 44.86 % of the difference between OSR and wheat rhizosphere samples in June 2008, 44.94 % in March 2008 and 41.91 % in June 2007. All other OSR rhizosphere OTUs were represented by <1 % of reads (with the exception of OTU 3 in June 2007, which had a mean relative abundance of 12.97 % in the OSR rhizosphere) (Table 4.4). The plasmodiophorid community within the OSR bulk soil differed markedly to that in the OSR rhizosphere. The most highly represented lineage in OSR bulk soil was OTU 1 and therefore was classified in the 26 % - 50 % abundance range (orange in Figure 4.6), specifically with an average abundance of 45.32 %. The average abundance of other OTUs in the OSR bulk soil was around 10 % (compared with < 1 % in the rhizosphere). OTUs 3 (11.02 %), 4 (10.55 %), 2 (9.63 %), 23 (8.32 %) and 5 (6.51 %) accounted for 91.46 % of the difference between OSR wheat within the bulk soil (Table 4.5). Furthermore in June 2007, June 2008 and March 2008, in which crop type significantly impacted plasmodiophorid community composition, OTU 24 and 3 were classified as a common OTU within OSR rhizosphere but not in OSR bulk soil. Conversely OTU 6 was classified as a common OTU in June 2007, June 2008 and March 2008 within OSR bulk soil but not within OSR rhizosphere (Figure 4.6).

Table 4.4 Similarity percentage analysis (SIMPER) of OTUs most responsible for differentiating plasmodiophorid communities within the rhizosphere of OSR and wheat

Season	OTU	% Contribution to difference	Cumulative contribution to difference (%)	Mean relative abundance in OSR	Mean relative abundance in Wheat
Jun-08	23	48.87	48.87	0.43	93.20
	1	44.86	93.73	88.64	3.49
Nov-07	1	33.42	33.42	44.07	20.88
	23	29.12	62.53	15.73	38.16
	2	12.86	75.39	15.72	17.90
	3	7.82	83.21	5.19	3.96
	5	4.93	88.15	9.27	6.80
	4	4.27	92.42	1.57	5.48
Mar-08	1	44.94	44.94	87.88	8.18
	23	43.55	88.49	0.38	76.29
	2	5.27	93.77	0.43	9.57
Jun-07	23	43.82	43.82	0.74	81.69
	1	41.91	85.72	80.04	2.62
	3	6.81	92.53	12.97	1.20

Table 4.5 Similarity percentage analysis (SIMPER) of OTUs most responsible for differentiating plasmodiophorid communities within the bulk soil of OSR and wheat, in June 2007. There was no significant difference between OSR and wheat in June 2008, March 2008 and November 2007.

OTU	% Contribution to difference	Cumulative contribution to difference (%)	Mean relative abundance in bulk soil OSR	Mean relative abundance in bulk soil wheat
1	33.79	33.79	45.32	8.72
2	19.83	53.62	9.63	30.36
23	10.63	64.25	8.32	18.43
4	10.16	74.41	11.02	0.06
5	9.78	84.2	10.55	16.13
6	7.26	91.46	6.51	13.92

In contrast, wheat rhizosphere communities were dominated by OTU 23, a relative of *Sorosphaera veronicae* and categorised in the 51 % - 100 % (red in Figure 4.6), specifically 93.20 % of reads in June 2008; 76.29 % in March 2008; 81.69 % in June 2007), which accounted for 48.87 % of the difference between wheat rhizosphere and bulk soil in June 2008, 43.55 % in March 2008 and 43.82 % in June 2007 (Table 4.4). OTU 1 was represented at a much lower level (between 2 % - 10 % coloured green in Figure 4.6), specifically 3.49 % in June 2008, 2.62 % in June 2007 and 8.18 % in March 2008) (Figure 4.5; Table 4.4). Within wheat bulk soil communities (in June 2007), the most highly represented

lineage in was OTU 2 (classified orange between 26 % - 50 % in Figure 4.6) with a mean relative abundance of 30.36 %, followed by OTU 23 (average of abundance 18.43 %), OTU 5 (average of abundance 16.13 %) and OTU 6 (average of abundance 13.92 %) (Table 4.5). Across all significant time points, there were no consistent differences in the OTUs classified as common under wheat rhizosphere and wheat bulk soil.

4.3.4. Influence of rotation on plasmodiophorid communities within the rhizosphere and bulk soil

In most cases, bulk soil and rhizosphere plasmodiophorid assemblages were not significantly affected by crop rotation. Within the rhizosphere, rotation only significantly impacted community composition between continuous-wheat and wheat-after-OSR in November 2007 ($P = 0.03$, $r = 0.37$) (Table 4.6). OTU 23 (*Sorosphaera veronicae*) dominated continuous-wheat rotations, with an average relative abundance of 40.85 %. This was followed by OTU 2 (identical 18S rDNA sequence to *Polymyxa graminis),* which had an average relative abundance of 26.64 %. This compared to wheat-after-OSR rotations in which OTU 23 (*Sorosphaera veronicae*) and OTU 1 (*Spongospora subterranea nasturtii)* had similar average relative abundances (35.47 % for OTU 23 and 31.89 % for OTU 1).

Table 4.6 ANOSIM statistics on the significant comparisons of plasmodiophorid communities under OSR and wheat rotations within bulk soil and rhizosphere, by plant growth stages

	Rhizosphere		Bulk Soil	
	Continuous Wheat vs Wheat after OSR	Continuous Wheat vs Wheat after OSR	Continuous OSR vs OSR 2 Year Gap	Continuous OSR vs Alternate OSR
Jun-07	-	-	-	
Nov-07	R = 0.37, P = 0.03	R = 0.38, P = 0.03		R = 0.40, P = 0.03
Mar-08	-	R = 0.33, P = 0.03	-	
Jun-08	-	-	R = 0.63, P = 0.03	

In bulk soil, significant differences in community composition occurred between continuous-wheat and wheat-after-OSR in November 2007 (P = 0.03, r = 0.38) and March 2008 (P = 0.03, r = 0.33) (Table 4.6). In both time points, continuous-wheat rotations were characterised by a dominance of OTU 2 (*P.graminis)* (average relative abundance of 38.13 % in November 2007 and 42.6 % in March 2008). In wheat-after-OSR rotations on the other hand, OTU 1 (*Spongospora subterranea nasturtii)* was most dominant (47.47 % in November 2007 and 41.63 % in March 2008) (Table 4.7).

Table 4.7 Significant simper analyses for wheat rotations under rhizosphere and bulk soil for each season

Material/Crop	Season	OTU	% Contribution to difference	Cumulative contribution to difference (%)	Mean relative abundance under Continuous Wheat	Mean relative abundance under Wheat after OSR
Rhizosphere	Nov-07	1	27.97	27.97	10.38	31.89
		2	23.26	51.23	26.64	9.17
		24	22.4	73.63	40.85	35.47
		3	10.18	83.8	0.25	7.67
		4	4.7	88.5	6.79	4.16
		5	4.05	92.55	7.56	6.04
Bulk Soil	Nov-07	1	35.26	35.26	12.89	47.47
		2	30.09	65.35	38.13	8.62
		4	6.84	72.19	16.55	13.26
		5	6.75	78.94	11.47	10.61
		24	5.63	84.57	8.47	7.42
		3	4.93	89.5	0.52	4.84
		6	4.35	93.85	5.46	3.02
	Mar-08	2	30.65	30.65	42.6	12.28
		1	29.17	59.82	12.77	41.63
		24	14.2	74.02	15.66	17.7
		3	7.38	81.4	1.53	8.07
		5	5.84	87.24	8.66	6.31
		6	5.16	92.41	7.77	4.03

In bulk soil OSR samples, differences occurred between continuous-OSR and OSR-2-year-gap in June 2008 (P = 0.03, r = 0.63) and between continuous-OSR and alternate-OSR in November 2007 (P = 0.03, r = 0.40) (Table 4.6). At

both time points, OTU 1 (*Spongospora s. nasturtii)* was the most abundant with an average relative abundance under continuous-OSR of 34.49 % in November 2007 and 48.13 % in June 2008. This compared with alternate-OSR whereby OTUs 4 (relative abundance of 24.14 %), OTU 23 (*Sorosphaera veronicae)* (relative abundance of 23.95 %) and OTU 2 (*P.graminis*) (relative abundance of 18.73 %) were most abundant within November 2007. In June 2008, OTU 23 (*Sorosphaera veronicae)* (relative abundance of 24.06 %), OTU 2 (*P.graminis)* (average relative abundance of 23.59 %) and OTU 1 (*Spongospora s. nasturtii) (*average relative abundance of 19.05 %) were most abundant within OSR-2-year-gap (Table 4.8).

Table 4.8 Significant simper analyses for OSR rotations under bulk soil for each season

		OTU	% Contribution to difference	Cumulative contribution to difference (%)	Mean relative abundance under Continuous OSR	Mean relative abundance under Alternate OSR
Bulk Soil	Nov-07	1	29.35	29.35	34.49	11.63
OSR		24	23.33	52.68	6.8	23.95
		4	18.25	70.93	15.53	24.14
		5	10.32	81.25	17.13	10.59
		2	6.41	87.66	17.72	18.73
		6	4.41	92.07	1.45	4.64
					Continuous OSR	OSR 2-Year Gap
Bulk Soil	Jun-08	1	35.86	35.86	48.13	19.05
OSR		24	21.62	57.48	5.51	24.66
		3	15.67	73.45	10.47	8.05
		2	12.92	86.36	12.72	23.59
		4	5.39	91.75	11.79	12.58

Crop yield data were available for June 2007 and June 2008. Although there were no significant differences in plasmodiophorid composition in the rhizosphere between these two time points, significant differences in plasmodiophorid communities were shown in June 2008 bulk soil samples between continuous OSR and OSR with a two-year-gap. The June 2008 yield data (Hilton et al., 2013) for rotation was overlaid with the average relative abundance of the three most abundant OTUs. The data showed lower yields associated with continuous OSR (yield 2.97 t/ha) and OSR with a two-year-gap

(3.53 t/ha), compared with 11.27 t/ha under continuous wheat, 12.64 t/ha under wheat after OSR and 11.28 t/ha under alternate OSR. Whilst the relative abundance of OTU 1 (*Spongospora s. nasturtii)* was highest under continuous OSR (48.13 %), it was not significantly different to the higher yielding rotations, other than continuous wheat (P = 0.03, r = 0. 68) and OSR with a two-year-gap (mentioned). Furthermore given that OTU abundance of all three OTUs under OSR with a two-year-gap was not consistently higher than the other rotations associated with higher yields, it is unlikely that plasmodiophorid abundance was the sole influence of OSR yield in the bulk soil (Figure 4.11).

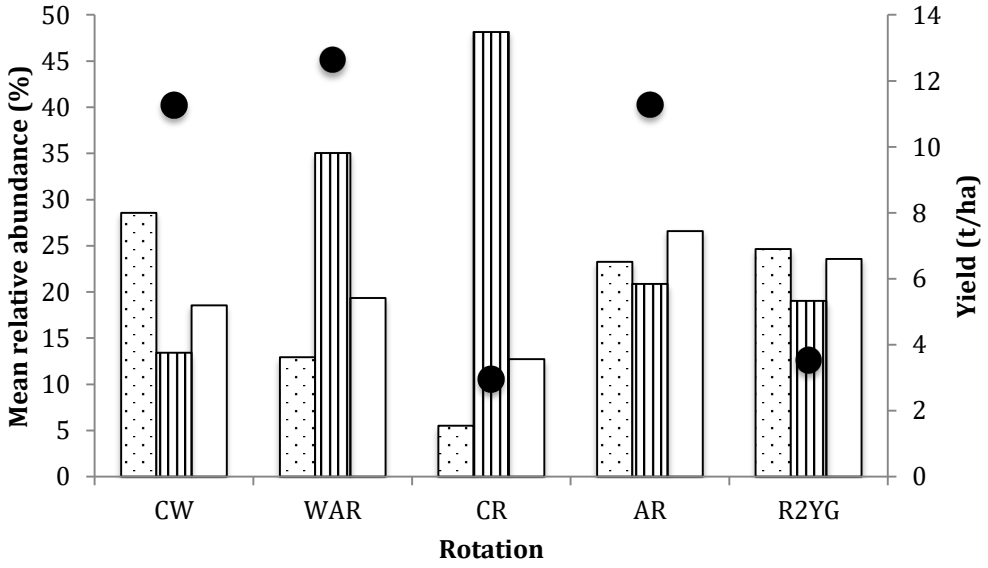


Figure 4.11 Graph showing mean relative abundance (%) of OTU 1 (vertical lines), OTU 2 (no fill) and OTU 23 (spots), coupled with yield (t/ha) for June 2008 across each rotation in the bulk soil (black circles). CW – Continuous wheat, WAR – wheat after OSR, CR – continuous OSR, AR – alternate OSR and R2YG – OSR with a 2-year gap.

4.4. Discussion

4.4.1. Novel plasmodiophorid diversity within the soil

This study provides evidence that plasmodiophorid diversity in the soil has been substantially underestimated. In just the single site used in the current study, 15 potentially new lineages were identified, proving almost a 50 % increase on the current estimate of ca. 40 sequence types (Neuhauser et al., 2011) Additional work is required to determine whether these lineages are distinct through morphological and parasitological analyses. Greater plasmodiophorid diversity in soil systems has also recently been highlighted by Neuhauser et al., (2014), whereby 11 distinct lineages related to *Woronina pythii* (OTU 5) alone were discovered. Before Neuhauser et al., (2014) and the present study, biodiversity investigations of plasmodiophorids were conducted using morphology and host screening and therefore do not provide sequence data that can be included in my analyses.

4.4.2. Influence of OSR and wheat on plasmodiophorid communities within the rhizosphere and bulk soil

The study by Neuhauser et al., (2014) investigated the host ranges of plasmodiophorids and in doing so discovered novel plasmodiophorid diversity. However the factors shaping plasmodiophorid communities have not previously been investigated. I have shown that plasmodiophorid communities within the rhizosphere were significantly different to those in bulk soil for both OSR and wheat crops, and were significantly different between the rhizosphere of these two crop species. Studies investigating bacterial and fungal populations have demonstrated an increase in the relative abundance of some taxa in the

rhizosphere, resulting in a dominant community structure (Bonkowski 2004). This compares to a more even structure in the bulk soil, where a greater number of taxa are represented by a lower abundance (Gomes et al., 2001). This is the first time this has also been shown for plasmodiophorid communities. Differences between the rhizosphere and bulk soil were based on shifts in the relative abundance of *S.subterranea nasturtii* (OTU 1), *Polymyxa graminis* and *Sorosphaera veronicae* (OTU 2 and 23) within the rhizosphere, compared with an more even community structure within the bulk soil, which consisted of *P. graminis*, a *P.brassicae* relative (OTU 4), *Woronina spp* (OTU 5) and *S.subterranea nasturtii* (OTU 1). This is likely due to the close association between plasmodiophorids and the plant rhizosphere, mediated by, for example nutrient availability, of which root exudates including amino acids, sugars and organic acids are a source (Berg et al., 2006).

Structural and functional diversity of bacteria within the rhizosphere is known to be influenced by a number of abiotic and biotic factors including plant growth stage, grazers, pesticide application, soil type and plant developmental stage (Garbeva et el., 2004). Plant species is thought to be a major determinant of microbial communities within the rhizosphere (Berg and Smalla 2009). This gives rise to a plant-specific rhizosphere effect, which has been shown in bacterial and fungal communities (Berg and Smalla, 2009; Kent and Triplett, 2002). Here we show for the first time a plant species effect on plasmodiophorid diversity in the rhizosphere. Wheat plants were dominated by *Sorosphaera veronicae* (OTU 23), whilst OSR was dominated by *Spongospora nasturtii* (OTU 1). *Spongospora nasturtii* (OTU 1) is thought to exhibit high host specificity, only infecting *Brassicaea*. Whilst studies have reported that

Spongospora nasturtii causes crook root disease in watercress, in which growth is stunted, very little has been published regarding its association with other plant species, even within the *Brassicales* (Claxton et al., 1996). Given that it was present in such high abundance within the rhizosphere, further research is needed to determine the potential impact of *S.nasturtii* (OTU 1) on OSR. OTU 23 was highly dominant in wheat. It is closely related to *Sorosphaera veronicae* typically known to only be associated with Plantaginaceae, specifically *Veronica* species (Neuhauser et al., 2014). This may however be an artefact of under sampling or biased sampling given that the rest of the clade that OTU belongs to or is associated with grasses. OTU 23 has a clearly distinct SSU sequence from S*orosphaera veronicae* and is therefore likely to be a different sequence type. Given that other minor SSU differences in plasmodiophorids can be associated with strong biological differences, it is certainly feasible that OTU 23, be associated with grass crops such as wheat. Further research is needed therefore to determine how closely related OTU 23 is to *Sorosphaera veronicae* and the host range of this potential new sequence type.

4.4.3. Understanding the drivers of plasmodiophorid communities through common and rare partitioning

The current study shows a significant positive relationship in plasmodiophorids for all treatments. Therefore as demonstrated in Chapter 3 and in a number of other studies, commonness and rarity of (in this case) plasmodiophorids was related to their permanence (Magurran and Henderson, 2003, van der Gast et al., 2011a, Rogers et al., 2011a). By partitioning the plasmodiophorid OTUs into common and rare groups based on sample occupancy, the importance of

dispersal and abiotic and biotic factors can be determined. This provides a novel insight into the drivers of plasmodiophorid beta-diversity given that common and rare species are expected to be biologically distinct and as a result have different ecological requirements (Magurran and Henderson 2003, Ulrich and Zalewski, 2006). Common taxa tend to be widespread in the community and locally abundant. Their distribution is thought to be shaped by species interactions such as competition (Ulrich and Zalewski 2006). Rare taxa on the other hand, are known to be occasional in the community, have low abundance and restricted distributions. (Magurran and Henderson 2007). Given the higher abundance and widespread nature of the common taxa, there is a need to identify the novel common taxa (i.e. OTU 6, 20 and 4) in order to determine their impact on plant hosts. In the current study, there was no detection of *P.brassicae* and consequently no symptoms of clubroot. OTU 4 however, which we show to be related to *P.brassicae,* was detected. OTU 4 was present in 6.5 % of all samples and was most abundant in the bulk soil but found in both OSR and wheat samples. Given that its relative *P.brassicae* has recently been shown to infect a number of plant hosts including *Poacaea* in addition to *Brassicaea* (Neuhauser et al., 2014), its host range is much greater than previously thought. OTU 4 could therefore also pose a potential threat to crops, but there is the possibility that it has no detrimental impact on crops. In addition OTU 5 (*Woronina*) and OTU 2 (*Polymyxa graminis*) were classified as common members of the community and found in greater abundance in bulk soil. Their distribution pattern therefore is likely to be governed by biological factors such as competition. They could also be more susceptible to plant defence chemicals released by the plants during different stages of its development. This is especially interesting since *Woronina*

spp is a parasite of the oomycete *Pythium* i.e. a hyperparasite (parasite of a parasite) in plants. This provides an exciting and novel insight into the complex interactions amongst plasmodiophorid taxa given that the literature to date is focussed on phylogeny and host preferences of *Woronina* spp*,* rather than the factors driving its distribution in the environment (Gleason et al., 2014; Neuhauser et al., 2014).

4.4.4. Influence of plant developmental stage on plasmodiophorid communities

Plant developmental stage is known to impact rhizosphere microbial communities. For example studies on bacterial communities in the rhizosphere have shown that distinct communities are associated with young plants compared with other stages of the plant's development, whilst other studies have shown distinct fungal communities associated with the senescence stage of a plant's development (Duineveld et al., 2001; Hannula et al., 2012). Distinct plasmodiophorid communities associated with different plant development stages were evident in my current study. The rhizosphere communities associated with the seedling stage (November 2007), were distinct from those at all other time points. The difference in plasmodiophorid community during the earliest phase of development may be associated with changes in the composition of rhizodeposits at different plant growth stages. For example early stages of a plant's development are associated with enhanced levels of plant-defense proteins (Chaparro et al., 2014). This difference could also be related to other abiotic and biotic factors including the increased complexity in above ground community composition that follows colonization and an increase in microbial competition as the community becomes more established. The plasmodiophorid

community then appeared to become fixed by March whereby the same plasmodiophorid community was present in subsequent time points. This is further supported by the fact that the community composition was the same in June samples of 2007 and 2008, despite plants coming from different plots. This demonstrates that the community is likely to be predictable across years and plant growth stages.

4.4.5. Influence of rotation on plasmodiophorid communities within the rhizosphere and bulk soil

When addressing the impact of rotation, Hilton et al., (2007) found that the enrichment of fungal pathogens (*Olpidium brassicae* and *Pyrenochaeta lycopersici*) within continuous OSR rotations (compared with other rotations) was associated with a 25% reduction in crop yield. In the current study, no significant effect of rotation on the plasmodiophorid community was found in the rhizosphere. Whilst a yield decline in the continuous OSR rotation was experienced in both June 2007 and June 2008, plasmodiophorid communities showed no significant difference in relative abundance under continuous OSR relative to the other rotations. This suggests that the plasmodiophorids detected in this study are not a main contributor to yield decline in OSR. However within the bulk soil, the increase in the relative abundance of OTU 1 (*Spongospora s. nasturtii)* in continuous OSR in June 2008 was associated with a decline in yield. These differences in the bulk soil could potentially be contributing to yield decline, however further research would be needed to test this. For example a measure of absolute abundance or biomass would be required to determine whether there are significantly more *Spongospora s. nasturtii* in the continuously

grown OSR rotation compared with the other rotations. Finally it would be interesting to determine whether plasmodiophorid occurrence if positively or negatively correlated with other taxa, including pathogens in both the rhizosphere and bulk soil.

In conclusion, this study has increased our current estimates of plasmodiophorid diversity in the soil and provided an insight into the factors that impact plasmodiophorid community composition and structure. Rhizosphere communities were associated with distinct communities relative to the bulk soil. This might be expected given that plasmodiophorids require a primary host plant to complete their life cycle and reproduce. Furthermore different host plants were shown to be associated with different plasmodiophorid communities, suggesting that rhizosphere environment is fundamental to determining plasmodiophorid community composition. Further research could investigate whether this is also shown in other plant species. Different crop plants from the same location could be investigated to determine whether plant species have differing rhizosphere effects on plasmodiophorid diversity, selecting for different species from the seed bank.

The importance of plant growth stage in determining plasmodiophorid community has also been demonstrated, communities associated with the early plant development stage being distinct from all other time points. Furthermore, we show that the same community was present in plants collected from June in successive years, and in different plots, demonstrating that plasmodiophorid community composition appears to be predictable.

Whilst the data showed no impact of rotation on plasmodiophorids in the rhizosphere, differences within the bulk soil were seen in certain wheat and OSR

rotations. In bulk soil OSR plants, a decline in yield coincided with an increase in the relative abundance of *Spongospora subterranea nasturtii.* This warrants further investigation to determine the role of this plasmodiophorid (if any) in yield decline.

A better understanding of plasmodiophorid diversity may be essential for the ecological and economic management of crop species. Furthermore by understanding the factors that impact abundance and distribution within the whole community, protists can be managed more effectively with the aim of reducing impacts such as yield decline and ultimately increasing crop productivity.

CHAPTER V: GENERAL DISCUSSION

5.1. General overview of findings

This work investigated the spatial scaling, community composition and community structure of microorganisms in the soil, with a detailed look at a particularly understudied group of protists. It has demonstrated that in a single agricultural system, microbes across all three domains of life demonstrate spatial scaling at the landscape scale. Microbial distribution patterns differ between taxa groups and major differences are demonstrated in the drivers of spatial scaling between and within domains, with no single universal driver of microbial spatial scaling in soil. Land-use management is also an important driver in the distribution of eukaryotes, whilst the drivers of bacterial and particular eukaryote distributions are impacted under varying land-use management. This could suggest that the community composition of certain taxa groups may function as biological indicators of soil health. When investigating the structure within

microbial communities, a universal pan microbial relationship between abundance and distribution is demonstrated for the first time. Every microbial taxa group could be partitioned into common and rare groups, which further elucidated the importance of land-use management on microbial community structure.

5.1.1 Understanding the drivers of spatial scaling and the impact of land use management

This work investigated the distribution of twenty-one microbial groups across every domain of life from bacteria and archaea to a range of eukaryotes such as fungi, protists and nematodes across England. The influence of environmental parameters and geographical distance were explored, in addition to the effects of land use on microbial beta diversity, in the form of conversion from set-aside to arable field practice.

Prokaryotes and eukaryotes differed with regards to the factors that influenced their distribution. Environmental factors and land use were important predictors in the distribution of every eukaryote taxa. This compared with bacterial and archaeal taxa, in which only environmental parameters, namely pH and soil type in bacteria and pH, C, N and Mg in Archaea were important. Of the few studies that have investigated whether land use management itself is a significant driver of the spatial distribution of microorganisms, they concluded that edaphic properties were more influential than land use type (Lauber *et al.*, 2008; Kuramae *et al.*, 2012; Hazard *et al.*, 2012). However, the impact of agricultural management on the drivers of beta diversity had not been explored in the literature. This study showed that the environmental variables associated with

community composition differed depending on land use management. For example, bacteria beta diversity correlated with soil type under arable conditions in a few taxa groups, however under set-aside conditions, pH was a key predictor of distribution in most of the bacterial taxa. Within the eukaryotes, pH was the main driver of spatial distribution within all of the fungal groups and the Stramenopiles under both arable and set-aside conditions (with the exception of Basidiomycota under arable conditions). Different environmental variables however influenced the beta diversity of Basidiomycota fungi, with Mg a significant driver of Basidiomycota distribution under arable conditions compared with pH, C, N and geographic distance under set-aside conditions. Only minor differences in the drivers of beta diversity occurred within the other eukaryote taxa groups under different land use conditions. Similarly land use management did not influence Archaea spatial distribution.

5.1.2 Understanding the effect of land use management in the context of the commonness and rarity of microorganisms

By assessing the species abundance distributions of each microbial taxa group using sample occupancy, key information regarding the processes that shape microbial communities could be hypothesized. For example, the universal positive relationship between taxa abundance and distribution highlighted, showed that each microbial community consisted of members that were widely dispersed and locally abundant and others which were restricted in their distribution and less abundant. Each coherent community represents a continuum of taxa that are driven by different processes. For example highly abundant and therefore common taxa are likely to be driven by niche processes whilst those,

which are rare (represented by few individuals) will be limited by dispersal (Magurran and Henderson 2003, Ulrich and Zalewski 2006). This supported the work in Chapter 2 which showed that both environmental and geographical distance were important predictors of the spatial scaling of microorganisms in the soil.

By partitioning each microbial taxa group into common and rare members, land use management was also shown to impact the community structure of bacteria, archaea and eukaryote taxa groups. Furthermore, under set-aside or arable practice, particular OTUs from each domain were present in one or the other land use condition demonstrating that land use management particularly affected the rare members of the community. Shifts occurred however in both common and rare members under each land use management practice. This suggests that land use management impacts members of the community differently where certain taxa within each domain are sensitive to disturbance in their environment, whilst others are more robust. Similarly, disturbance impacted the dominance and evenness of the community whereby conversion from arable to set-aside conditions resulted in more even community structures for the majority for taxa groups investigated.

5.1.3 Novel insights into the diversity of plasmodiophorids and factors affecting their community composition

The case study on plasmodiophorids, renowned in the literature for their detrimental impacts on crop health, also highlighted that the community could be partitioned into common and rare members, which provided a novel insight into the drivers of plasmodiophorid beta diversity. Taxa including *Spongospora*

subterranean nasturtii and *Sorosphaera veronicae* were identified as being common in the environment and are likely to be driven by abiotic and biotic factors for example.

A number of additional factors were explored using an alternative agricultural system in plasmodiophorid communities including the impact of growth development stage, plant host, human interference in the form of rotation and soil type (rhizosphere and bulk soil). The investigation showed that not only is the diversity of the group substantially underestimated, but that major differences in community composition occur between the rhizosphere and bulk soil, different host plants (wheat and OSR) and at different growth stages. Plasmodiophorid communities were affected by plant growth stage in the same way as many bacteria and fungal taxa have been shown to be (Duineveld et al., 2001, Hannula et al., 2012). The rhizosphere communities associated with the seedling stage were distinct to all other time points which may highlight the importance of colonisation of above ground community composition, increases in belowground competition as the environment establishes or changes in the rhizosphere exudates associated with early stage plants. Differences occurred in the communities depending on whether they were found in the bulk soil or rhizosphere and specific to certain plant hosts.

This work has therefore highlighted the importance of a number of different abiotic and biotic factors, including dispersal on the distribution of bacteria, archaea and eukaryotes, in particular protists in the soil. By assessing community composition in addition to aspects of structure, an insight into the mechanisms that shape microbial diversity have been explored including

immigration and extinction, speciation and competition (Prosser et al., 2007, Martiny et al., 2006).

5.2 Future directions

5.2.1. The advances and challenges of community profiling technologies

This work was based on amplicon surveys using 454 pyrosequencing, which provides the excellent sampling depth to characterize complex microbial communities at an affordable cost and good processing times (Tedersoo et al., 2010). Pyrosequencing-based studies that profile diversity within microbial communities have been largely based on ITS, 16S, 18S, 23S rRNA amplicons of environmental samples. However the length of the amplicon is still a concern with gene fragments limited to between 100 bp and 400 bp (Tedersoo et al., 2010). PCR-based errors and sequencing errors associated with these molecular approaches including base-call errors and homopolymers, which are known to inflate diversity estimates, particularly with regards to the rare biosphere are also a concern (Kunin et al., 2010, Huse et al., 2007).

A shift towards metagenomic and metatranscriptomic approaches are increasingly being applied to environmental data to identify functional and ecological roles of microorganisms. High-throughput shotgun sequencing technologies provide unbiased insights into all molecular aspects of a community allowing the simultaneous investigation of genes and their functions (Ulrich et al., 2008, Tinge et al., 2005, Raes and Bork 2008). High-throughput techniques are facilitating the monitoring of genes (metagenomics), transcript and protein levels (meta-transcriptomics and meta-proteomics) and metabolites (metabolomics). Metagenomics for example allow the discovery of novel genes

in the environment and large genomic inserts containing phylogenetic "anchors" that can enable direct links to be made to a microbial taxon. Metagenomics based on DNA however does not provide information on the functional role of genes in the environment, which therefore calls for the use of meta-transcriptomics (Ulrich et al., 2008).

Improvements in the profiling of communities brings with it challenges. Larger datasets require super-computers to collate the information into registries and store the vast amounts of data produced. There is also a lack of information regarding the genetics, biochemistry and physiology of model organisms, whereby large proportions of genes cannot be annotated, which currently limits the use of metagenomics in microbial diversity studies for example. In addition, each sequencing centre currently uses different annotation pipelines to process the sequences, which could potentially result in different results, so there is a need for standardisation of the processing of sequencing (Gilbert et al., 2011).

5.2.2. Recognising the impact of land use management on microbial communities and the effect on soil ecosystem processes

The current work has shown that microorganisms demonstrate spatial scaling and are driven by a number of different factors within the current agricultural system, with conversion from arable to set aside being an important condition to consider at the landscape scale. Given the evidence that microorganisms are distributed non-randomly in the soil, studies focus on areas of high microbial activity such as the rhizosphere when addressing soil processes. The varying correlation of different drivers with prokaryotes and eukaryotes could also provide an initial insight into the effective management of potential indicator species within this

type of land use management for example. Knowing how human interference can impact processes specific to certain domains or even specific taxa groups is essential in the prediction and management of the soil for future policy makers. For instance the impact of land-use change on soil microbial distribution patterns and community structure is fundamental to understanding crop health given the significant influence microorganisms have on crops, and their sensitivity to anthropogenic change (Jangid et al., 2008, van der Gast et al., 2011a). Bacteria in the soil for example influence ecosystem processes that contribute to the provision of essential ecosystem services through carbon and nutrient cycling. They are therefore recognised as having an essential role in the maintenance of soil fertility and ultimately productivity (Sayer et al., 2013, Barrios, 2007). A change (through management) in microbial community structure and composition may therefore have important consequences for soil functionality and must be managed effectively. Studies have shown that shifts in microbial community composition associated with conversion from conventional tillage to no-tillage can impact soil function by altering nutrient cycling processes (Barrios 2007).

To expand our understanding of the impact of anthropogenic disturbance on microbial communities, given its importance highlighted in this study, alternative agricultural systems should be explored. These could include conventional versus organic management in which fertilizer and tillage are addressed, along with investigating the impact of climate change and pollutant effects. Furthermore, additional abiotic and biotic factors could be investigated when assessing the drivers of microbial distribution in the soil. For example, Mg

is rarely addressed in the literature and highlighted in this study as an important predictor of Archaea distribution at larger ranges.

5.2.3. Understanding the role of plasmodiophorids in the soil and impact on crop yield

With regards to plasmodiophorids specifically, further research is needed to identify the novel lineages recognized in the current study. There is also a need to explore the taxa highlighted as common and determine functionality within the soil, given that, of the known plasmodiophorid taxa, many have a detrimental role in the health of crop plants. Finally, whilst this study showed no direct link of plasmodiophorids to yield decline in OSR, the associated increase in relative abundance of *Spongospora subterranea nasturtii* with continuous OSR in bulk soil warrants further investigation through targeting yield experiments.

References

Acosta-Martínez, V., Dowd, S., Sun, Y. and Allen, V. (2008). Tag-encoded pyrosequencing analysis of bacterial diversity in a single soil type as affected by management and land use. *Soil Biol. Biochem.* **40**, 2762–2770.

Ager, D., Evans, S., Li, H., Lilley, A. K. and van der Gast, C. J. (2010). Anthropogenic disturbance affects the structure of bacterial communities. *Environ. Microbiol.* ***12***, 670–678.

Altekar, G., Dwarkadas, S., Huelsenbeck, J.P. and Ronquist, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics.* ***20*****,** 407–415.

Altschul, SF., Gish, W., Miller, W., Myers. E. W. and Lipman, D. J. (1990). Basic Local Alignment Search Tool. *J. Mol. Biol.* **215**, 403–410.

Archibald, J. M. and Keeling, P. J. (2004). Actin and ubiquitin protein sequences support a cercozoan/foraminiferan ancestry for the plasmodiophorid plant pathogens. *J. Eukaryot. Microbiol.* ***51*****,** 113–118.

Axelsson, E. P., Hjältén, J., LeRoy, C. J., Whitham, T. G., Julkunen-Tiitto, R. and Wennström, A. (2011). Leaf litter from insect-resistant transgenic trees causes changes in aquatic insect community composition: Leaf litter from Bt trees and aquatic insects. *J. Appl. Ecol*. ***48***, 1472–1479.

Baker, D. J., Freeman, S.N., Grice, P.V. and Siriwardena, G. M. (2012). Landscape-scale responses of birds to agri-environment management: a test of the English Environmental Stewardship scheme: *Evaluating the effects of ES for farmland birds*. *J. Appl. Ecol.* **49**, 871–882.

Bardgett, R. (2005). The Biology of Soil: A community and ecosystem approach (Biology of Habitats Series). Oxford: Oxford University Press.

Barton, P. S., Cunningham, S. A., Manning, A. D., Gibb, H., Lindenmayer, D. B. and Didham, R. K. (2013). The spatial scaling of beta diversity: Spatial scaling of beta diversity. *Glob. Ecol. Biogeogr.* ***22*****,** 639–647.

Bass, D., Chao, E. E-Y., Nikolaev, S., Yabuki, A., Ishida, K., Berney, C., Pakzad, U., Wylezich, C. and Cavalier-Smith, T. (2009). Phylogeny of Novel Naked Filose and Reticulose Cercozoa: Granofilosea cl. n. and Proteomyxidea Revised. *Protist.* **160**, 75–109.

Bass, D., Richards, T.A., Matthai, L., Marsh, V. and Cavalier-Smith, T. (2007). DNA evidence for global dispersal and probable endemicity of protozoa. *BMC. Evol. Biol.* **7,** 162.

Bates, S.T., Berg-Lyons, D., Caporaso, J.G., Walters, W.A., Knight, R. and Fierer, N. (2011). Examining the global distribution of dominant archaeal populations in soil. *ISME J.* **5**, 908–917.

Bates, S. T., Clemente, J. C., Flores, G. E., Walters, W. A., Parfrey, L. W., Knight, R. and Fierer, N. (2012). Global biogeography of highly diverse protistan communities in soil. *ISME J.* **7**, 652–659.

Beck, S., Powell, J. R., Drigo, B., Cairney, J. W. G. and Anderson, I. C. (2015). The role of stochasticity differs in the assembly of soil- and root-associated fungal communities. *Soil. Biol. Biochem.* ***80*****,** 18–25.

Bell, T. (2010). Experimental tests of the bacterial distance–decay relationship. *ISME J.* **4**, 1357–1365.

Bennett, A. J., Bending, G. D., Chandler, D., Hilton, S. and Mills, P. (2012). Meeting the demand for crop production: the challenge of yield decline in crops grown in short rotations. *Biol. Rev*. ***87*****,** 52–71.

Berg, G., Opelt, K., Zachow, C., Lottmann, J., Götz, M., Costa, R. and Smalla, K. (2006). The rhizosphere effect on bacteria antagonistic towards the pathogenic fungus Verticillium differs depending on plant species and site: Rhizosphere-associated antagonistic bacteria. *FEMS Micro. Ecol.* ***56*****,** 250–261.

Berg, G. and Smalla, K. (2009). Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Micro. Ecol.* ***68***, 1–13.

Berthrong, S. T., Buckley, D. H. and Drinkwater, L. E. (2013). Agricultural Management and Labile Carbon Additions Affect Soil Microbial Community Structure and Interact with Carbon and Nitrogen Cycling. *Microb. Ecol*. ***66***, 158–170.

Bik, H. M., Sung, W., De Ley, P., Baldwin, J. G., Sharma, J., Rocha-Olivares, A. and Thomas, W. K. (2012). Metagenetic community analysis of microbial eukaryotes illuminates biogeographic patterns in deep-sea and shallow water sediments: Metagenetic Community Analysis. *Mol. Ecol.* ***21***, 1048–1059.

Bommarco, R., Kleijn, D. and Potts, S. G. (2013). Ecological intensification: harnessing ecosystem services for food security. *Trends Ecol. Evol.* ***28*****,** 230–238.

Bonkowski, M. (2004). Protozoa and plant growth: the microbial loop in soil revisited. *New Phytol.* ***162*****,** 617–631

Borcard, D. and Legendre, P. (2002). All-scale spatial analysis of ecological data by means of principal coordinates of neighbour matrices. *Ecol Model.* **153**, 51–68.

Borcard, D., Legendre, P., Avois-Jacquet, C. and Tuomisto, H. (2004). Dissecting the spatial structure of ecological data at multiple scales. *Ecology.* **85**, 1826–1832.

Borenstein. M,. Hedges, L. V., Higgins, J. P. T. and Rothstein, H. R. (2009). Introduction to Meta-Analysis. Wiley; 1 edition.

Bossio, D. A., Scow, K. M., Gunapala, N. and Graham, K. J. (1998). Determinants of soil microbial communities: effects of agricultural management, season, and soil type on phospholipid fatty acid profiles. *Microb. Ecol.* ***36*****,** 1–12.

Bouwmeester, H. J., Roux, C., Lopez-Raez, J.A. and Bécard, G. (2007). Rhizosphere communication of plants, parasitic plants and AM fungi. *Trends Plant Sci.* **12**, 224–230.

Bragg, L., Stone, G., Imelfort, M., Hugenholtz, P. and Tyson, G.W. (2012). Fast, accurate error-correction of amplicon pyrosequences using Acacia. *Nat. Method.* ***9*****,** 425–426.

Braselton, J. (1995). Current Status of the Plasmodiophorids. *Crit. Rev. Microbiol.* **21,** 263–275.

Bronick, C.J. and Lal, R. (2005) Soil structure and management: a review. *Geoderma*. **124,** 3-22.

Bru. D., Ramette, A., Saby, N. P. A., Dequiedt, S., Ranjard, L., Jolivet, C., Arrouays, D. and Philippot, L. (2011). Determinants of the distribution of nitrogen-cycling microbial communities at the landscape scale. *ISME J.* **5**, 532–542.

Bulman, S., Siemens, J., Ridgway, H. J., Eady, C. and Conner, A. J. (2006). Identification of genes from the obligate intracellular plant pathogen, *Plasmodiophora brassicae*. *FEMS Micro. Lett.* ***264*****,** 198–204.

Burki, F., Kudryavtsev, A., Matz, M. V., Aglyamova, G. V., Bulman, S., Fiers, M., Keeling, P.J. and Pawlowski, J. (2010). Evolution of Rhizaria: new insights from phylogenomic analysis of uncultivated protists. *BMC Evol. Biol.* ***10*****,** 377.

Campbell, B.J., Yu, L., Heidelberg, J.F. and Kirchman, D.L. (2011). Activity of abundant and rare bacteria in a coastal ocean. *Proc. Natl. Acad. Sci.* ***108*****,** 12776–12781.

Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Pea, 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, N., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunenko, T., Zaneveld, J. and Knight, R. (2010a). QIIME allows analysis of high-throughput community sequencing data. *Nat Methods.* **7**, 335–336.

Caporaso, J. G., Bittinger, K., Bushman, F. D., DeSantis, T. Z., Andersen, G. L. and Knight, R. (2010b). PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics.* **26**, 266–267.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N. and Knight, R. (2011). Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proc. Natl. Acad. Sci.* ***108*****,** 4516–4522.

Carson, J.K., Rooney, D., Gleeson, D.B. and Clipson, N. (2007). Altering the mineral composition of soil causes a shift in microbial community structure. *FEMS Micro. Ecol.* ***61*****,** 414–423.

Cavalier-Smith, T. and Chao, E.E.-Y. (2003). Molecular Phylogeny of Centrohelid Heliozoa, a Novel Lineage of Bikont Eukaryotes That Arose by Ciliary Loss. *J. Mol. Evol.* ***56*****,** 387–396.

Chaparro, J.M., Badri, D.V. and Vivanco, J.M. (2014). Rhizosphere microbiome assemblage is affected by plant development. *ISME J.* ***8*****,** 790–803.

Chase, J.M. (2005). Towards a really unified theory for metacommunities. *Funct. Ecol*. ***19*****,** 182–186.

Chivian, E. and Bernstein, A. (2008). Sustaining Life: How Human Health Depends on Biodiversity. Oxford University Press.

Cho, J-C. and Tiedje, J. M. (2000). Biogeography and Degree of Endemicity of Fluorescent Pseudomonas Strains in Soil. *Appl Environ Microbiol* **66**:5448–5456.

Choudhary, D. K., Sharma, K. P and Gaur, R. K. (2011). Biotechnological perspectives of microbes in agro-ecosystems. *Biotechnol Lett.* **33**, 1905–1910.

Chu, H., Fierer, N., Lauber, C.L., Caporaso, J.G., Knight, R. and Grogan, P. (2010). Soil bacterial diversity in the Arctic is not fundamentally different from that found in other biomes: Bacterial biogeography in arctic soils. *Environ. Microbiol.* ***12*****,** 2998–3006.

Claesson, M.J., Cusack, S., O'Sullivan, O., Greene-Diniz, R., de Weerd, H., Flannery, E., Marchesi, J.R., Falush, D., Dinan, T., Fitzgerald, G., Stanton, C., van Sinderen, D., O'Connor, M., Harnedy, N., O'Connor, K., Henry, C., O'Mahony, D., Fitzgerald, A., Shanahan, F., Twonmey, C., Hill, C., Ross, R, P. and O'Toole, P. W. (2011). Composition, variability, and temporal stability of the intestinal microbiota of the elderly. *Proc. Natl. Acad. Sci.* ***108*****,** 4586–4591.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* ***18*****,** 117–143.

Clarke, K. R. and Gorley, R. N., 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E, Plymouth.

Claxton, J. R., Potter, U. J., Blakesley, D. and Clarkson, J.M. (1996). An ultrastructural study of the interaction between Spongospora subterranea f. sp. nasturtii and watercress roots. *Mycol. Res.* ***100*****,** 1431–1439.

Curtis, T. and Halford, N. G. (2014). Food security: the challenge of increasing wheat yield and the importance of not compromising food safety: Wheat yield and food safety. *Ann. Appl. Biol*. ***164*****,** 354–372.

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favoured races in the struggle for life*. London: John Murray.

DeSantis, T. Z., Hugenholtz, P., Larsen, N., Rojas, M., Brodie, E. L., Keller, K., Huber, T., Dalevi, D., Hu, P and Andersen, G. L. (2006). Greengenes, a Chimera-Checked 16S rRNA Gene Database and Workbench Compatible with ARB. *Appl Environ Microbiol.* **72**, 5069–5072.

Desoignies, N., Carbonell, J., Moreau, J. S., Conesa, A., Dopazo, J. and Legrève, A. (2014). Molecular interactions between sugar beet and *Polymyxa betae* during its life cycle: Molecular interactions between sugar beet and *Polymyxa betae*. *Ann. Appl. Biol.* ***164*****,** 244–256.

Van Dijk, E. L., Auger, H., Jaszczyszyn, Y. and Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends Genet.* ***30*****,** 418–426.

Dixon, G. R. (2009). The Occurrence and Economic Impact of Plasmodiophora brassicae and Clubroot Disease. *J. Plant Growth Regul.* ***28*****,** 194–202.

Dixon, G.R. (2014). Clubroot (*Plasmodiophora brassicae* Woronin) – an agricultural and biological challenge worldwide. *Can. J. Plant Pathol*. ***36*****,** 5–18.

Dolan, J. R., Ritchie, M. E., Tunin-Ley, A. and Pizay, M. D. (2009). Dynamics of core and occasional species in the marine plankton: tintinnid ciliates in the north-west Mediterranean Sea. *J. Biogeogr.* ***36*****,** 887–895.

Duineveld, B.M., Kowalchuk, G.A., Keijzer, A., van Elsas, J.D. and van Veen, J.A. (2001). Analysis of Bacterial Communities in the Rhizosphere of Chrysanthemum via Denaturing Gradient Gel Electrophoresis of PCR-Amplified 16S rRNA as Well as DNA Fragments Coding for 16S rRNA. *Appl. Environ. Microbiol.* ***67*****,** 172–178.

Dumbrell, A. J., Nelson, M., Helgason, T., Dytham, C. and Fitter, A. H. (2009). Relative roles of niche and neutral processes in structuring a soil microbial community. *ISME J.* **4**, 337–345.

Dunbar, J., Barns, S.M., Ticknor, L.O. and Kuske, C.R. (2002). Empirical and Theoretical Bacterial Diversity in Four Arizona Soils. *Appl. Environ. Microbiol.* ***68*****,** 3035–3045.

Duplessis, S., Cuomo, C. A., Lin, Y-C., Aerts, A., Tisserant, E., Veneault-Fourrey, C., Joly, D. J., Hacquard, S., Amselem, J., Cantarel, B. L., Chiu, R., Coutinho, P. M., Feau, N., Field, M., Frey, P., Gelhaye, E., Goldberg, J., Grabherr, M. G., Kodira, C. D., Kohler, A., Kües, U., Lindquist, E. A., Lucas, S. M., Mago, R., Mauceli, E., Morin, E., Murat, C., Pangilinan, J. L., Park, R., Pearon, M., Quesneville, H., Rouhier, N., Sakthikumar, S., Salamov, A. A., Schmutz, J., Selles, B., Shapiro, H., Tanguay, P., Tuskan, G. A., Henrissat, B., Ven de Peer, Y., Rouze, P., Ellis, J. G., Dodds, P. N., Schein, J. E., Zhong, S., Hamelin, R. C., Grigoriev, I. V., Szabo, L. J and Marton F. (2011). Obligate biotrophy features unraveled by the genomic analysis of rust fungi. *Proc Natl Acad Sci*. **108**, 9166–9171.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics.* **26**, 2460–2461.

Edgar, R. C., Haas, B. J., Clemente, J. C., Quince, C. and Knight, R. (2011). UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* ***27*****,** 2194–2200.

Edgar, R. C. (2013). UPARSE: highly accurate OTU sequences from microbial amplicon reads. *Nat. Methods*. ***10*****,** 996–998.

Edgcomb, V., Orsi, W., Bunge, J., Jeon, S., Christen, R., Leslin, C., Holder, M., Taylor, G. T., Suarez, P., Varela, R. and Epstein, S. (2011). Protistan microbial observatory in the Cariaco Basin, Caribbean. I. Pyrosequencing vs Sanger insights into species richness. *ISME J*. ***5*****,** 1344–1356.

Faggian, R. and Strelkov, S. E. (2009). Detection and Measurement of Plasmodiophora brassicae. *J. Plant Growth Regul.* ***28*****,** 282–288.

Falloon, R. E. (2008). Control of Powdery Scab of Potato: Towards Integrated Disease Management. *Am. J. Potato Res.* ***85*****,** 253–260.

Feinstein, L. M. and Blackwood, C. B. (2013). The spatial scaling of saprotrophic fungal beta diversity in decomposing leaves. *Mol Ecol.* **22**, 1171–1184.

Fenchel, T. (2003). Biogeography for Bacteria. *Science.* ***301*****,** 925–926.

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

Finkel, O. M., Burch, A. Y., Elad, T., Huse, S. M., Lindow, S. E., Post, A. F. and Belkin, S. (2012). Distance-Decay Relationships Partially Determine Diversity Patterns of Phyllosphere Bacteria on Tamrix Trees across the Sonoran Desert. *Appl. Environ. Microbiol.* ***78*****,** 6187–6193.

Finlay, B., Esteban, G., Clarke, K. and Siqueira, J. F. (2001). Biodiversity of Terrestrial Protozoa Appears Homogeneous across Local and Global Spatial Scales. *Protist.* ***152*****,** 355–366.

Fisher, R., Corbet, S. and Williams, C.. (1943). The Relation Between the Number of Species and the Number of Individuals in an Random Sample of an Animal Population. *J. Anim. Ecol.* ***12*****,** 42–58.

Floyd, R.M., Rogers, A.D., Lambshead, P.J.D. and Smith, C.R. (2005). Nematode-specific PCR primers for the 18S small subunit rRNA gene. *Mol. Ecol. Notes*. ***5*****,** 611–612.

Foissner, W. (2008). Protist diversity and distribution: some basic considerations. *Biodivers. Conserv.* ***17*****,** 235–242.

Fowler, J., Cohen, L.,and Jarvis, P. (1998). Practical Statistics for Field Biology (Chichester; New York: Wiley).

Fuhrman, J. A. (2009). Microbial community structure and its functional implications. *Nature*. ***459*****,** 193–199.

Galand, P. E., Casamayor, E. O., Kirchman, D. L. and Lovejoy, C. (2009). Ecology of the rare microbial biosphere of the Arctic Ocean. *Proc. Natl. Acad. Sci.* ***106*****,** 22427–22432.

Garbeva, P., Hordijk, C., Gerards, S. and de Boer, W. (2014). Volatile-mediated interactions between phylogenetically different soil bacteria. *Front. Microbiol.* ***5*****.** 1-9.1

Van der Gast, C. J., Gosling, P., Tiwari, B. and Bending, G. D. (2011a). Spatial scaling of arbuscular mycorrhizal fungal diversity is affected by farming practice: Spatial scaling of arbuscular mycorrhizal fungi. *Environ. Microbiol.* **13**, 241–249.

Van der Gast, C. J., Walker, A.W., Stressmann, F. A., Rogers, G. B., Scott, P., Daniels, T. W., Carroll, M. P., Parkhill, J. and Bruce, K. D. (2011b). Partitioning core and satellite taxa from within cystic fibrosis lung bacterial communities. *ISME J.* ***5*****,** 780–791.

Gaston, K. J., Blackburn, T. M., Greenwood, J. J., Gregory, R. D., Quinn, R. M. and Lawton, J. H. (2000). Abundance–occupancy relationships. *J. Appl. Ecol.* ***37*****,** 39–59.

Gihring, T. M., Green, S. J. and Schadt, C. W. (2012). Massively parallel rRNA gene sequencing exacerbates the potential for biased community diversity comparisons due to variable library sizes: Pyrosequencing exacerbates sample size bias. *Environ. Microbiol.* **14**, 285–290.

Gilchrist, E., Soler, J., Merz, U. and Reynaldi, S. (2011). Powdery scab effect on the potato Solanum tuberosum ssp. andigena growth and yield. *Trop. Plant Pathol.* ***36*****,** 350–355.

Gilbert, J. A., Meyer, F. and Bailey, M. J. (2011). The future of microbial metagenomics (or is ignorance bliss?). *ISME J.* ***5*****,** 777.

Gleason, F. H., Lilje, O., Marano, A.V., Sime-Ngando, T., Sullivan, B. K., Kirchmair, M. and Neuhauser, S. (2014). Ecological functions of zoosporic hyperparasites. *Front. Microbiol.* ***5*****.** 1-10

Glücksman, E., Bell, T., Griffiths, R.I. and Bass, D. (2010). Closely related protist strains have different grazing impacts on natural bacterial communities: Protist grazing of bacterial communities. *Environ. Microbiol.* ***12***, 3105–3113.

Godfray, H. C J., Beddington, J. R., Crute, I. R., Haddad, L., Lawrence, D., Muir, J. F., Pretty, J., Robinson, S., Thomas, S. M and Toulmin, C. (2010). Food Security: The Challenge of Feeding 9 Billion People. *Science.* **327**, 812–818.

Gomes, N. C. M., Heuer, H., Schönfeld, J., Costa, R., Mendonca-Hagler, L. and Smalla, K. (2001). Bacterial diversity of the rhizosphere of maize (Zea mays) grown in tropical soil studied by temperature gradient gel electrophoresis. *Plant Soil*. ***232*****,** 167–180.

Gosling, P., Proctor, M., Jones, J. and Bending, G. D. (2014b). Distribution and diversity of Paraglomus spp. in tilled agricultural soils. *Mycorrhiza.* ***24*****,** 1–11.

Gossen, B. D., Mcdonald, M. R., Hwang, S. F., Strelkov, S. E. and Peng, G. (2013). A comparison of clubroot development and management on canola and Brassica vegetables. *Can. J. Plant Pathol*. ***35*****,** 175–191.

Green, J. L., Holmes, A. J., Westoby, M., Oliver, I., Briscoe, D., Dangerfield, M., Gillings, M. and Beattie, A. J. (2004). Spatial Scaling Of Microbial Eukaryote Diversity. *Nature.* **432**, 747–750.

Green, J. and Bohannan, B.J.M. (2006). Spatial scaling of microbial biodiversity. *Trends Ecol. Evol.* ***21*****,** 501–507.

Griffiths, R. I., Thomson, B, C., James, P., Bell, T., Bailey, M. and Whiteley, A. S. (2011). The bacterial biogeography of British soils: Mapping soil bacteria. *Environ. Microbiol.* **13**, 1642–1654.

Guo, Q., Brown, J.H. and Valone, & T.J. (2000). Abundance and distribution of desert annuals: are spatial and temporal patterns related? *J. Ecol.* ***88*****,** 551–560.

Haas, B. J., Gevers, D., Earl, A. M., Feldgarden, M., Ward, D. V., Giannoukos G., Ciulla, D., Tabbaa, D., Highlander, S. K., Sodergren, E., Methé, B. and DeSantis, T. Z.(2011). Chimeric 16S rRNA sequence formation and detection in Sanger and 454-pyrosequenced PCR amplicons. *Genome Res.* **21**, 494–504.

Hannula, S.E., de Boer, W. and van Veen, J. (2012). A 3-Year Study Reveals That Plant Growth Stage, Season and Field Site Affect Soil Fungal Communities while Cultivar and GM-Trait Have Minor Effects. *PLoS ONE.* ***7*****,** e33819.

Hanski, I. (1982). Dynamics of Regional Distribution: The Core and Satellite Species Hypothesis. *Oikos.* ***38*****,** 210–221.

Hanson, C. A., Fuhrman, J.A., Horner-Devine, M. C. nad Martiny, J. B. H. (2012). Beyond biogeographic patterns: processes shaping the microbial landscape. *Nat. Rev. Microbiol.* **14**, 497-506.

Hassanzadeh Davarani, F., Rezaee, S., Mahmoudi, S.B., Norouzi, P., Safarnejad, M.R. and Safarpour, H. (2014). Detection and molecular characterization of Polymyxa betae, transmitting agent of sugar beet rhizomania disease, in Iran. *Span. J. Agric. Res.* ***12***, 787.

Hayat, R., Ali, S., Amara, U., Khalid, R. and Ahmed, I. (2010). Soil beneficial bacteria and their role in plant growth promotion: a review. *Ann. Microbiol.* ***60*****,** 579–598.

Hazard, C., Gosling, P., van der Gast, C. J., Mitchell, D. T., Doohan, F. M. and Bending, G. D. (2013). The role of local environment and geographical distance in determining community composition of arbuscular mycorrhizal fungi at the landscape scale. *ISME J.* **7**, 498–508.

Hazard, C., Gosling, P., Mitchell, D. T., Doohan, F. M. and Bending, G. D. (2014). Diversity of fungi associated with hair roots of ericaceous plants is affected by land use. *FEMS Microbiol. Ecol.* ***87*****,** 586–600.

Hillstrom, M.L. and Lindroth, R. (2008). Elevated atmospheric carbon dioxide and ozone alter forest insect abundance and community composition. Insect *Conserv. Divers.* ***1*****,** 233–241.

Van der Heijden, M. G. A., Bardgett, R. D. and van Straalen, N. M. (2008). The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecol. Lett.* ***11*****,** 296–310.

Herold, N., Schöning, I., Gutknecht, J., Alt, F., Boch, S., Müller, J., Oelmann, Y., Socher, S. A., Wilcke, W., Wubet, T. and Schrumpf, M. (2014). Soil property and management effects on grassland microbial communities across a latitudinal gradient in Germany. *Appl Soil Eco.l* **73**, 41–50.

Hilton, S., Bennett, A. J., Keane, G., Bending, G. D., Chandler, D., Stobart, R. and Mills, P. (2013). Impact of Shortened Crop Rotation of Oilseed Rape on Soil and Rhizosphere Microbial Diversity in Relation to Yield Decline Van Overbeek, L (ed). *PLoS ONE.* **8**, e59859.

Hol, W. H. G., de Boer, W., Termorshuizen, A. J., Meyer, K. M., Schneider, J. H. M., van Dam, N. M., van Veen, J. A.,and van der Putten, W. H. (2010). Reduction of rare soil microbes modifies plant-herbivore interactions. *Ecol. Lett.* ***13,*** *292–301.*

Horner-Devine, M. C., Carney, K. M. and Bohannan, B. J, M. (2004). An ecological perspective on bacterial biodiversity. *Proc. R. Soc. B. Biol. Sci.* **271**, 113–122.

Horz, H. P., Barbrook, A., Field, C. B. and Bohannan, B. J. (2004). Ammonia-oxidizing bacteria respond to multifactorial global change. *Proc. Natl. Acad. Sci. U. S. A.* ***101*****,** 15136–15141.

Houlden, A., Timms-Wilson, T. M., Day, M. J. and Bailey, M. J. (2008). Influence of plant developmental stage on microbial community structure and activity in the rhizosphere of three field crops: Plant and growth stage effects on microbial populations. *FEMS Microbiol. Ecol.* ***65*****,** 193–201.

Huber, L., Eisenbeis, G., Ruhl, E.H., Pagay, V. and Kirchmair, M. (2007). Distribution and host range of the grapevine plasmodiophorid Sorosphaera viticola. *VITIS.* ***46***, 23-25

Hwang, S. F., Strelkov, S. E., Feng, J., Gossen, B. D. and Howard, R.J. (2012). Plasmodiophora brassicae: a review of an emerging pathogen of the Canadian

canola (Brassica napus) crop: Progress on canola clubroot research. *Mol. Plant Pathol.* ***13*****,** 105–113.

Jangid, K., Williams, M. A., Franzluebbers, A. J., Sanderlin, J. S., Reeves, J. H., Jenkins, M. B. Endale, D. M., Coleman, D. C. and Whitman, W. B. (2008). Relative impacts of land-use, management intensity and fertilization upon soil microbial community structure in agricultural systems. *Soil. Biol. Biochem.* **40**, 2843–2853.

Kanyuka, K., Ward, E. and Adams, M. J. (2003). Polymyxa graminis and the cereal viruses it transmits: a research challenge. *Mol. Plant Pathol.* ***4*****,** 393–406.

Katoh, K., Misawa, K., Kuma, K. and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* ***30***, 3059–3066.

Kemmitt, S. J., Wright, D., Goulding, K. W. T. and Jones, D. L. (2006). pH regulation of carbon and nitrogen dynamics in two agricultural soils. *Soil Bio and Biochem.* **38**, 898-911.

Kent, A.D. and Triplett, E. W. (2002). Microbial Communities and Their Interactions in Soil and Rhizosphere Ecosystems. *Annu. Rev. Microbiol.* ***56*****,** 211–236.

King, A. J., Freeman, K. R., McCormick, K. F., Lynch, R. C., Lozupone, C., Knight, R., and Scmidt, S. K. (2010). Biogeography and habitat modelling of high-alpine bacteria. *Nat Commun.* **1**, 1–6.

Kivlin, S. N., Hawkes, C. V. andTreseder, K. K. (2011). Global diversity and distribution of arbuscular mycorrhizal fungi. *Soil. Biol. Biochem.* **43**, 2294–2303.

Klamer, M. and Hedlund, K. (2004). Fungal diversity in set-aide agricultural soil investigated using terminal-restriction fragment length polymorphism. *Soil. Biol. Biochem.* **36**, 983–988.

Koid, A., Nelson, W. C., Mraz, A. and Heidelberg, K. B. (2012). Comparative Analysis of Eukaryotic Marine Microbial Assemblages from 18S rRNA Gene and Gene Transcript Clone Libraries by Using Different Methods of Extraction. *Appl. Environ. Microbiol.* **78**, 3958–3965.

Krebs, C. J. (1999). Ecological Methodology (Menlo Park, Calif: Benjamin Cummings).

Kunin, V., Engelbrektson, A., Ochman, H. and Hugenholtz, P. (2010). Wrinkles in the rare biosphere: pyrosequencing errors can lead to artificial inflation of diversity estimates. *Environ. Microbiol.* ***12*****,** 118–123.

Kuramae, E. E., Yergeau, E., Wong, L. C., Pijl, A.S., Veen, J. A. and Kowalchuk, G. A. (2012). Soil characteristics more strongly influence soil bacterial communities than land-use type. *FEMS. Micro. Ecol.* **79**, 12–24.

Lane D. (1991). Nucleic Acid Techniques in Bacterial Systematics. 1 edition. Wiley: Chichester ; New York.

Lauber, C. L., Strickland, M. S., Bradford, M. A. and Fierer, N. (2008). The influence of soil properties on the structure of bacterial and fungal communities across land-use types. *Soil. Biol. Biochem.* **40**, 2407–2415.

Lauber, C. L., Hamady, M., Knight, R., and 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.

Leake, J., Johnson, D., Donnelly, D., Muckle, G., Boddy, L. and Read, D. (2004). Networks of power and influence: the role of mycorrhizal mycelium in controlling plant communities and agroecosystem functioning. *Can. J. Bot.* ***82*****,** 1016–1045.

Lemaire, O., Merdinoglu, D., Valentin, P., Putz, C., Ziegler-Graff, V., Guilley, H., Jonard, G. and Richards, K. (1988). Effect of beet necrotic yellow ven virus RNA composition on transmission by Polymyxa betae. *Virology*. ***162*****,** 232–235.

Lekberg, Y., Koide, R. T., Rohr, J. R., Aldrich-Wolfe, L. and Morton, J. B. (2007). Role of niche restrictions and dispersal in the composition of arbuscular mycorrhizal fungal communities. *J. Ecol*. ***95*****,** 95–105.

Lienhard, P., Terrat, S., Prévost-Bouré, N. C., Nowak, V., Régnier, T., Sayphoummie, S., Panyasiri, K., Tivet, F., Mathieu, O., Levêque, J., Maron, P. and Ranjard, L. (2014). Pyrosequencing evidences the impact of cropping on soil bacterial and fungal diversity in Laos tropical grassland. *Agron. Sustain. Dev.* ***34*****,** 525–533.

Lupwayi, N. Z., Rice, W. A and Clayton, G. W. (1998). Soil microbial diversity and community structure under wheat as influenced by tillage and crop rotation. *Soil. Biol. Biochem.* **30**, 1733–1741.

Magurran, A. E. and Henderson, P. A. (2003). Explaining the excess of rare species in natural species abundance distributions. *Nature*. ***422*****,** 714–716.

Magurran, A. E. (2004). Measuring Biological Diversity. Blackwell Publishing: Oxford, UK. 256p.

Magurran, A. E. (2007). Species abundance distributions over time. *Ecol. Lett.* **10,** 347–354.

Magurran, A. E., Khachonpisitsak, S. and Ahmad, A.B. (2011). Biological diversity of fish communities: pattern and process. *J. Fish Biol.* ***79*****,** 1393–1412.

Mantel N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Res.* **27**, 209–220.

Manzanares-Dauleux, M. J., Divaret, I., Baron, F. and Thomas, G. (2001). Assessment of biological and molecular variability between and within field isolates of Plasmodiophora brassicae. *Plant Pathol.* ***50*****,** 165–173.

Mardis, E. R. (2008). Next-Generation DNA Sequencing Methods. *Annu. Rev. Genomics Hum. Genet.* ***9*****,** 387–402.

Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., Horner-Devine, M. C., Kane, M., Adams Krumins, J., Kuske, C. R., Morin, P. J., Naeem, S., Øvreds, L., Reysenbach, A., Smith, V. H. and Staley, T. (2006). Microbial biogeography: putting microorganisms on the map. *Nat. Rev. Microbiol.* **4**, 102–112.

Martiny, J. B. H., Eisen, J. A., Penn K., Allison, S. D. and Horner-Devine, M. C. (2011). Drivers of bacterial -diversity depend on spatial scale. *Proc. Natl. Acad. Sci.* **108**, 7850–7854.

Matson, P. A. (1997). Agricultural Intensification and Ecosystem Properties. *Science,* **277**, 504–509.

McCaig, A. E., Glover, L. A. and Prosser, J. I. (1999). Molecular analysis of bacterial community structure and diversity in unimproved and improved upland grass pastures. *Appl. Environ. Microbiol.* **65**, 1721–1730.

McDonald, D., Price, M. N., Goodrich, J., Nawrocki, E. P., DeSantis, T. Z., Probst. A,. Andersen, G. L., Knight, R. and Hugenholtz, P. (2011). An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. *ISME J.* **6**, 610–618.

McGill, B. J., Etienne, R.S., Gray, J. S., Alonso, D., Anderson, M. J., Benecha, H. K., Dornelas, M., Enquist, B. J., Green, J. L., He, F., Hurlbert, A. H., Magurran, A. E., Marquet, P. A., Maurer, B. A., Ostling, A., Soykan, C. U., Ugland, K. I. and White, E. P. (2007). Species abundance distributions: moving beyond single prediction theories to integration within an ecological framework. *Ecol. Lett.* ***10*****,** 995–1015.

Medlin, L., Elwood, H. J., Stickel, S. and Sogin, M. L. (1998). The Characterization Of Enzymatically Amplified Eukaryotic 16S-like rRNA-coding Regions. *Gene.* **71**, 491–499.

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

Miller, M. A., Pfeiffer, W. and Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In Gateway Computing Environments Workshop (GCE), 2010, pp. 1–8.

Morrissey, J. P., Dow, J. M., Mark, G. L. and O'Gara, F. (2004). Are microbes at the root of a solution to world food production? *EMBO. Rep.* **5**, 922–926.

Muyzer, G.,and Smalla, K. (1998). Application of denaturing gradient gel electrophoresis (DGGE) and temperature gradient gel electrophoresis (TGGE) in microbial ecology. *Antonie Van Leeuwenhoek.* ***73*****,** 127–141.

Nekola, J. C. (2003). Large-scale terrestrial gastropod community composition patterns in the Great Lakes region of North America. *Divers. Distrib*. ***9*****,** 55–71.

Nemergut, D. R., Costello, E. K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K., Fierer, N., Townsend, A. R., Cleveland, C. C., Stanish, L. and Knight, R. (2011). Global patterns in the biogeography of bacterial taxa: Global bacterial biogeography. *Environ. Microbiol.* **13**, 135–144.

Neuhauser, S., Kirchmair, M. and Gleason, F. H. (2011a). The ecological potentials of Phytomyxea ("plasmodiophorids") in aquatic food webs. *Hydrobiologia*. **659,** 23–35.

Neuhauser, S., Kirchmair, M. and Gleason, F. H. (2011b). Ecological roles of the parasitic phytomyxids (plasmodiophorids) in marine ecosystems - a review. *Mar. Freshw. Res*. **62**, 365.

Neuhauser, S., Martin, K., Simon, B. and David, B. (2014). Cross-kingdom host shifts of phytomyxid parasites. *BMC. Evol. Biol.* **14**, 33.

Newman, E. I. and Reddell, P. (1987). The distribution of mycorrhizas among families of vascular plants. *New Phytol.* **106**, 745–751.

Oehl, F., Sieverding, E., Ineichen, K., Mader, P., Boller, T. and Wiemken, A. (2003). Impact of Land Use Intensity on the Species Diversity of Arbuscular Mycorrhizal Fungi in Agroecosystems of Central Europe. *Appl. Environ. Microbiol.* ***69,*** 2816–2824.

Papke, R. T., Ramsing, N. B., Bateson, M. M. and Ward, D. M. (2003). Geographical isolation in hot spring cyanobacteria. *Environ. Microbiol.* **5**, 650–659.

Pasternak, Z., Al-Ashhab, A., Gatica, J., Gafny, R., Avraham, S., Minz, D., Gillor, O. and Jurkevitch, E. (2013). Spatial and Temporal Biogeography of Soil Microbial Communities in Arid and Semiarid Regions López-García, P (ed). *PLoS ONE.* **8**, e69705.

Pedrós-Alió, C. (2006). Marine microbial diversity: can it be determined? *Trends Microbiol.* **14,** 257–263.

Pedrós-Alió, C. (2012). The Rare Bacterial Biosphere. *Annu. Rev. Mar. Sci*. **4,** 449–466.

Pereira e Silva, M. C., Schloter-Hai, B., Schloter, M., van Elsas, J. D. and Salles, J. F. (2013). Temporal Dynamics of Abundance and Composition of Nitrogen-Fixing Communities across Agricultural Soils. *PLoS ONE.* ***8,*** e74500.

Pommier, T., Neal, P., Gasol, J., Coll, M., Acinas, S. and Pedrós-Alió, C. (2010). Spatial patterns of bacterial richness and evenness in the NW Mediterranean Sea explored by pyrosequencing of the 16S rRNA. *Aquat. Microb. Ecol.* **61**, 221–233.

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

Preston, F. (1948). The Commonness and Rarity of Species. *Ecology*. **29,** 254–283.

Preston, F. (1960). Time and Space and the Variation of Species. *Ecology*. **41,** 611–627.

Pringle, A., Baker, D. M., Platt, J. L., Wares, J. P., Latge, J. P. and Taylor, J. W. (2005). Cryptic speciation in the cosmopolitan and clonal human pathogenic fungus Aspergillus fumigatus. *Evolution*. **59,** 1886–1899.

Prosser, J. I., Bohannan, B. J., Curtis, T. P., Ellis, R. J, Firestone, M. K., Freckleton, R. P., Green, J. L., Green, L. E., Killham, K., Lennon, J. L., Osborn, M., Solan, M., van der Gast, C. J. and Young, J. P. W. (2007). The role of ecological theory in microbial ecology. *Nat. Rev. Microbiol.* **5**, 384–392.

Pywell, R.F., Heard, M. S., Bradbury, R. B., Hinsley, S., Nowakowski, M., Walker, K. J. and Bullock , J. M. (2012). Wildlife-friendly farming benefits rare birds, bees and plants. *Biol. Lett.* **8**, 772–775.

Quince, C., Lanzén, A., Curtis, T. P., Davenport, R. J., Hall, N., Head, I. M., Read, L. F. and Sloan, W. T. (2009). Accurate determination of microbial diversity from 454 pyrosequencing data. *Nat. Methods*. ***6*,** 639–641.

Quince, C., Lanzen, A., Davenport, R. J. and Turnbaugh, P. J. (2011). Removing noise from pyrosequenced amplicons. *BMC Bioinformatics.* ***12*,** 38.

Raes, J. and Bork, P. (2008). Molecular eco-systems biology: towards an understanding of community function. *Nat. Rev. Microbiol.* ***6*,** 693–699.

Ramette, A and Tiedje, J. M. (2006). Biogeography: An Emerging Cornerstone for Understanding Prokaryotic Diversity, Ecology, and Evolution. *Microb. Ecol.* **53**, 197–207.

Rands, M. R. W., Adams, W. M., Bennun, L., Butchart, S. H. M., Clements, A., Coomes, D., Entwistle, A., Hodge, I., Kapos, V., Scharlemann, J. P. W., Sutherland, W. J. and Vira, B. (2010). Biodiversity Conservation: Challenges Beyond 2010. *Science.* **329**, 1298–1303.

Ranjard, L., Dequiedt, S., Chemidlin Prévost-Bouré, N., Thioulouse, J., Saby, N. P. A., Lelievre, M., Maron, P. A., Morin, F. E. R., Jolivet, C., Arrouays, D. and Lemanceau, P. (2013). Turnover of soil bacterial diversity driven by wide-scale environmental heterogeneity. *Nat. Commun.* **4**, 1434.

Reche, I., Pulido-Villena, E., Morales-Baquero, R. and Casamayor, E. O. (2005). Does ecosystem size determine aquatic bacterial richness? *Ecology.* **86**, 1715–1722.

Reeder, J. and Knight, R. (2010). Rapidly denoising pyrosequencing amplicon reads by exploiting rank-abundance distributions. *Nat. Methods.* **7**, 668–669.

Reigstad, L. J., Richter, A., Daims, H., Urich, T., Schwark, L. and Schleper, C. (2008) Nitrification in terrestrial hot springs of Iceland and Kamchatka: Nitrification in hot springs. *FEMS Microbiol Ecol* **64**:167–174.

Rejmánková, E., Komárek, J. and Komárková, J. (2004). Cyanobacteria—a neglected component of biodiversity: patterns of species diversity in inland marshes of northern Belize (Central America). *Divers. Distrib*. ***10***, 189–199.

Roesch, L. F. W., Fulthorpe, R. R., Riva, A., Casella, G., Hadwin, A. K. M, Kent, A. D, Daroub, S. H., Camargo, A. A. O., Farmerie, W. G. and Triplett, E. W. (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J*. **1,** 283- 290.

Rogers, G. B., Cuthbertson, L., Hoffman, L. R., Wing, P. A., Pope, C., Hooftman, D. A., Lilley, A. K., Oliver, A., Carroll, M. P., Bruce, K. D. and van der Gast, C. J. (2013a). Reducing bias in bacterial community analysis of lower respiratory infections. *ISME J.* **7**, 697–706.

Rogers, G. B., van der Gast, C. J., Bruce, K.D., Marsh, P., Collins, J. E., Sutton, J. and Wright, M. (2013b). Ascitic Microbiota Composition Is Correlated with Clinical Severity in Cirrhosis with Portal Hypertension. PLoS ONE. ***8*,** e74884.

Ronquist, F. and Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics. **19**,* 1572–1574.

Rousk, J., Bååth, E., Brookes, P. C., Lauber, C. L., Lozupone, C., Caporaso, J. G., Knight, R. and Fierer, N. (2010). Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J*. ***4*,** 1340–1351.

Royal Society (Great Britain). (2009). Reaping the benefits: science and the sustainable intensification of global agriculture. The Royal Society: London.

Sagarin, R. D., Gaines, S. D. and Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends Ecol. Evol.* **21**, 524–530.

Sayer, E. J., Wagner, M., Oliver, A. E., Pywell, R. F., James, P., Whiteley, A. S. and Heard, M. S. (2013). Grassland management influences spatial patterns of soil microbial communities. *Soil. Biol. Biochem.* **61**, 61–68.

Schadt, E. E ., Turner, S. and Kasarskis, A. (2010). A window into third-generation sequencing. *Hum. Mol. Genet. **19*,** R227–R240.

Scheckenbach, F., Wylezich, C., Weitere, M., Hausmann, K. and Arndt, H. (2005). Molecular identity of strains of heterotrophic flagellates isolated from surface waters and deep-sea sediments of the South Atlantic based on SSU rDNA. *Aquat. Microb. Ecol. **38*,** 239–247.

Schloss, P. D., Gevers, D. and Westcott, S. L. (2011). Reducing the Effects of PCR Amplification and Sequencing Artifacts on 16S rRNA-Based Studies. *PLoS ONE*. ***6*,** e27310.

Schuster, S. C. (2007). Next-generation sequencing transforms today's biology. *Nat. Methods*. ***5*****,** 16–18.

Shade, A., Caporaso, J. G., Handelsman, J., Knight, R. and Fierer, N. (2013). A meta-analysis of changes in bacterial and archaeal communities with time. *ISME J.* **7**, 1493–1506.

Shokralla, S., Spall, J. L., Gibson, J.F. and Hajibabaei, M. (2012). Next-generation sequencing technologies for environmental DNA research: Next-Generation Sequencing For Environmental DNA. *Mol. Ecol*. ***21*****,** 1794–1805.

Siqueira, J. F., Fouad, A. F. and Rôças, I. N. (2012a). Pyrosequencing as a tool for better understanding of human microbiomes. *J. Oral Microbiol.* ***4*****,** 1-15.

Siqueira, T., Bini, L.M., Roque, F. O., Marques Couceiro, S. R., Trivinho-Strixino, S. and Cottenie, K. (2012b). Common and rare species respond to similar niche processes in macroinvertebrate metacommunities. *Ecography*. ***35***, 183–192.

Sjostedt, J., Koch-Schmidt, P., Pontarp, M., Canback, B., Tunlid, A., Lundberg, P., Hagstrom, A. and Riemann, L. (2012). Recruitment of Members from the Rare Biosphere of Marine Bacterioplankton Communities after an Environmental Disturbance. *Appl. Environ. Microbiol.* ***78*****,** 1361–1369.

Smalla, K., Wieland, G., Buchner, A., Zock, A., Parzy, J., Kaiser, S., Roskot, N., Heuer, H. and Berg, G. (2001). Bulk and Rhizosphere Soil Bacterial Communities Studied by Denaturing Gradient Gel Electrophoresis: Plant-Dependent Enrichment and Seasonal Shifts Revealed. *Appl. Environ. Microbiol*. ***67*****,** 4742–4751.

Smith, M. J., Adams, M. J. and Ward, E. (2013). Ribosomal DNA analyses reveal greater sequence variation in *Polymyxa* species than previously thought and indicate the possibility of new ribotype-host-virus associations: Ribotype-host-virus associations in *Polymyxa* spp. *Environ. Microbiol. Rep.* ***5*****,** 143–150.

Sogin, M. L., Morrison, H. G., Huber, J. A., Welch, D. M., Huse, S. M., Neal, P. R., Arrieta, J. M. and Herndl, G.J. (2006). Microbial diversity in the deep sea and the underexplored "rare biosphere." *Proc. Natl. Acad. Sci*. **103,** 12115–12120.

Solow, A R. (1993). A Simple Test for Change in Community Structure. *J Anim Ecol,* **62**, 191.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* **22**, 2688–2690.

Stoeck, T., Bass, D., Nebel, M., Christen, R., Jones, M. D. M., Breiner H, W. and Richards, T. A. (2010). Multiple marker parallel tag environmental DNA sequencing reveals a highly complex eukaryotic community in marine anoxic water. *Mol. Ecol.* **19**, 21–31.

Strange, R. N. and Scott, P.R., (2005). Plant Disease: A Threat to Global Food Security. *Annu. Rev. Phytopathol.* **43**, 83–116.

Sweetman, J. N., Rühland, K. M. and Smol, J. P. (2010). Environmental and spatial factors influencing the distribution of cladocerans in lakes across the central Canadian Arctic treeline region. *J. Limnol.* **69**, 76–87.

Tamada, T. and Kondo, H. (2013). Biological and genetic diversity of plasmodiophorid-transmitted viruses and their vectors. *J. Gen. Plant Pathol.* **79,** 307–320.

Tap, J., Mondot, S., Levenez, F., Pelletier, E., Caron, C., Furet, J.-P., Ugarte, E., Muñoz-Tamayo, R., Paslier, D.L.E., Nalin, R., Dore, J. and Lecierc, M. (2009). Towards the human intestinal microbiota phylogenetic core. *Environ. Microbiol.* **11**, 2574–2584.

Team, R. D. C. (2010). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org

Tedersoo, L., Nilsson, R.H., Abarenkov, K., Jairus, T., Sadam, A., Saar, I., Bahram, M., Bechem, E., Chuyong, G.,and Kõljalg, U. (2010). 454 Pyrosequencing and Sanger sequencing of tropical mycorrhizal fungi provide similar results but reveal substantial methodological biases. *New Phytol.* ***188*****,** 291–301.

Ter Braak, C. J. F. (1988). CANOCO-a FORTRAN Programme for Canonical Community Ordination by [partial][detrended][canonical] Correspondence Analysis, Principal Components Analysis and Redundancy Analysis (version 2.1). Agricultural Mathematics Group.

Tilman, D. (1999). Global environmental impacts of agricultural expansion: the need for sustainable and efficient practices. *Proc. Natl. Acad. Sci.* **96,** 5995–6000.

Tringe, S., von Mering, C., Kobayashi, A., Salamov, A., Chen, K., Podar, M., Short, J., Mathur, E., Detter, J., Bork, P., Hugenholtz, P. and Rubin, E. M. (2005). Comparative Metagenomics of Microbial Communities. *Science*. ***308*****,** 554–558.

Tokeshi, M. (1993). Species Abundance Pattern and Community Structure. *Adv. Ecol. Res.* **24,** 111–186.

Ulrich, W. (2008). Species Abundance Distributions in Space and Time. *Ecol. Quest.* **9,** 15–20.

Ulrich, W. and Ollik, M. (2004). Frequent and occasional species and the shape of relative-abundance distributions. *Divers. Distrib.* ***10*****,** 263–269.

Ulrich, W. and Zalewski, M. (2006). Abundance and co-occurrence patterns of core and satellite species of ground beetles on small lake islands. *Oikos.* ***114***, 338–348.

Ulrich, T., Lanzén, A., Qi, Ji., Huson, D. H., Schleper, C. and Schuster, S. C. (2008). Simultaneous Assessment of Soil Microbial Community Structure and Function Through Analysis of the Meta-transcriptome. *PLoS ONE.* **3,** e2527.

Unterseher, M., Jumpponen, A., Öpik, M., Tedersoo, L., Moora, M., Dormann, C.F. and Schnittler, M. (2011). Species abundance distributions and richness estimations in fungal metagenomics - lessons learned from community ecology: COMMUNITY ECOLOGY IN FUNGAL METAGENOMICS. *Mol. Ecol.* ***20*****,** 275–285.

Vellend, M. (2010). Conceptual synthesis in community ecology. *Q. Rev. Biol.* **85**, 183–206.

Verbruggen, E., Röling, W. F. M., Gamper, H. A., Kowalchuk, G. A., Verhoef, H. A. and van der Heijden, M. G. A. (2010). Positive effects of organic farming on below-ground mutualists: large-scale comparison of mycorrhizal fungal communities in agricultural soils. *New Phytol.* **186**, 968–979.

De Vries, F.T., Manning, P., Tallowin, J. R. B., Mortimer, S. R., Pilgrim, E. S., Harrison, K. A., Hobbs, P. L., Quirk, H., Shipley, B., Cornelissen, J. H. C., Kattge, J. and Bardgett, R. D. (2012). Abiotic drivers and plant traits explain landscape-scale patterns in soil microbial communities Johnson, N (ed). *Ecol. Lett.* **15**, 1230–1239.

Wallenhammar, A.C. (1996). Prevalence of Plasmodiophora brassicae in a spring oilseed rape growing area in central Sweden and factors influencing soil infestation levels. *Plant. Pathol.* ***45*****,** 710–719.

Ward, B. B. and O'Mullan, G. D. (2002). Worldwide Distribution of Nitrosococcus oceani, a Marine Ammonia-Oxidizing -Proteobacterium, Detected by PCR and Sequencing of 16S rRNA and amoA Genes. *Appl. Environ. Microbiol.* ***68*****,** 4153–4157.

Ward, E. and Adams, M. J. (1998). Analysis of ribosomal DNA sequences of Polymyxa species and related fungi and the development of genus- and species-specific PCR primers. *Mycol. Res.* ***102*****,** 965–974.

Ward, E., Kanyuka, K., Motteram, J., Kornyukhin, D. and Adams, M. J. (2004). The use of conventional and quantitative real-time PCR assays for Polymyxa graminis to examine host plant resistance, inoculum levels and intraspecific variation. *New Phytol.* ***165*****,** 875–885.

Werner, E. E., Davis, C. J., Skelly, D. K., Relyea, R. A., Benard, M. F. and McCauley, S. J. (2014). Cross-scale interactions and the distribution-abundance relationship. *PloS One.* **9,** e97387.

Wessén, E., Hallin, S. and Philippot, L. (2010). Differential responses of bacterial and archaeal groups at high taxonomical ranks to soil management. *Soil. Biol. Biochem.* **42**, 1759–1765.

Wessén, E., Söderström, M., Stenberg, M., Bru, D., Hellman, M., Welsh, A., Thomsen, F., Klemedtson, L., Philippot, L. and Hallin, S. (2011). Spatial distribution of ammonia-oxidizing bacteria and archaea across a 44-hectare farm related to ecosystem functioning. *ISME J.* **5**, 1213–1225.

Whitaker, R. J., Grogan, D. W. and Taylor, J. W. (2003). Geographic Barriers Isolate Endemic Populations of Hyperthermophilic Archaea. *Science.* **301**, 976–978.

Wieland, G., Neumann, R. and Backhaus, H. (2001). Variation of Microbial Communities in Soil, Rhizosphere, and Rhizoplane in Response to Crop Species, Soil Type, and Crop Development. *Appl. Environ. Microbiol.* **67,** 5849–5854.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J., Yang, Y., Yao, T., Knight, R. and Chu, H. (2012). Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau: Bacterial biogeography in alkaline sediments. *Environ. Microbiol.* **14,** 2457–2466.

Yannarell, A. C. and Triplett, E. W. (2005). Geographic and Environmental Sources of Variation in Lake Bacterial Community Composition. *Appl. Environ. Microbiol.* **71**, 227–239.

Ye, R., Wright, A. L., Inglett, K., Wang, Y., Ogram, A. V. and Reddy, K. R. (2009). Land-use effects on soil nutrient cycling and microbial community dynamincs in the everglades agricultural area, Florida. *Comm. in Soil Science and Plant Analysis.* **40,** 2725-2742.

Yeates, G. W., Bongers, T., De Goede, R. G. M., Freckman, D. W. and Georgieva, S. S. (1993). Feeding Habits in Soil Nematode Families and Genera--An Outline for Soil Ecologists. *J Nematol.* **25**, 315–331.

Zhalnina, K., de Quadros, P.D., Camargo, F.A.O. and Triplett, E.W. (2012). Drivers of archaeal ammonia-oxidizing communities in soil. *Front. Microbiol: **3**, 1-9.*

Zhou, J., Xia, B., Treves, D. S., Wu, L-Y., Marsh, T. L. and O'Neill, R. V. (2002). Spatial and Resource Factors Influencing High Microbial Diversity in Soil. *Appl. Environ. Microbiol.* **68**, 326–334.

Appendix I

Table 3.4 – SIMPER analysis for Acidobacteria. Columns shown in red represent the core taxa. Uncoloured OTUs represent satellite taxa. Only taxa with a % contribution greater than 1% are shown.

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
18878	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	4.57	3.57	1.24	1.45	2.41	2.41
11232	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	3.17	2.06	0.98	1.39	1.9	4.31
10763	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	3.98	3.23	0.92	1.41	1.79	6.1
11826	Bacteria;Acidbacteria;Chlracidbacteria;	1.74	2.16	0.91	1.09	1.76	7.87
18254	Bacteria;Acidbacteria;Chlracidbacteria;	2.02	2.7	0.85	1.36	1.64	9.51
23213	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	3.16	3.42	0.85	1.31	1.64	11.15
11067	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	6.39	5.42	0.81	1.44	1.57	12.72
15833	Bacteria;Acidbacteria;Iibactere;Iibacterale; libacteraceae;Candidatulibacter;	1.04	1.37	0.75	1.06	1.45	14.17
18057	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	3.69	3.7	0.71	1.36	1.37	15.54
10657	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	1.78	1.98	0.63	1.44	1.22	16.76
14485	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	2.09	2.3	0.61	1.46	1.18	17.94
10425	Bacteria;Acidbacteria;Acidbacteria;Acidbacteriale;	1.92	1.29	0.6	1.42	1.15	19.1
21550	Bacteria;Acidbacteria;Chlracidbacteria;	0.57	1.39	0.54	0.97	1.05	20.15

Table 3.5 – SIMPER analysis for Actinobacteria

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
8421	Bacteria;Actinobacteria;Actinobacteria;MC47	7.94	7.25	1.98	1.39	3.3	3.3
11338	Bateria;Actinobacteria;Actinobacteria;MC47	1.88	3.44	1.62	1.27	2.69	5.99
23962	Bacteria;Actinobacteria;Actinobacteria; Actinomycetales;Propionibacteriaceae	3.78	3.27	1.61	1.14	2.67	8.66
23758	Bateria;Actinobacteria;Actinobacteria;MC47	2.79	1.97	1.39	1.21	2.31	10.97
18389	Bacteria;Actinobacteria;Actinobacteria; Actinomycetales;Nocardioidaceae	1.48	2.82	0.99	0.86	1.65	12.62
582	Bateria;Actinobacteria;Actinobacteria;MC47	1.82	1.06	0.86	1.2	1.43	14.05
2825	Bateria;Actinobacteria;Actinobacteria;MC47	2.94	2.7	0.81	1.45	1.35	15.4
2851	Bateria;Actinobacteria;Actinobacteria;MC47	1.47	1.65	0.79	1.28	1.31	16.71
13256	Bacteria;Actinobacteria;Actinobacteria; Solirubrobacterales	1.41	1.01	0.77	0.8	1.28	17.99
3124	Bateria;Actinobacteria;Actinobacteria;MC47	1.94	1.54	0.76	1.35	1.26	19.25
5808	Bacteria;Actinobacteria;Actinobacteria; Rubrobacterales;Rubrobacteraceae;	1.11	0.66	0.71	0.76	1.18	20.43
18999	Bacteria;Actinobacteria;Actinobacteria; Actinomycetales;Microbacteriaceae;	1.96	1.41	0.69	1.31	1.15	21.58
8801	Bacteria;Actinobacteria;Actinobacteria;MC47	2.07	1.81	0.68	1.42	1.13	22.71
23695	Bacteria;Actinobacteria;Actinobacteria; Actinomycetales;Micrococcaceae;	1.3	1.59	0.66	1.15	1.09	23.8
2711	Bacteria;Actinobacteria;Actinobacteria; Actinomycetales;Intrasporangiaceae;	1	1.51	0.64	1.34	1.07	24.87
17670	Bateria;Actinobacteria;Actinobacteria;MC47	0.29	1.31	0.64	0.96	1.06	25.94
268	Actinomycetales;Streptomycetaceae; Streptomyces	0.71	1.34	0.61	0.61	1.02	26.96
23605	Bateria;Actinobacteria;Actinobacteria;MC47	0.44	0.98	0.6	0.68	1	27.95

Table 3.6 – SIMPER analysis for Alphaproteobacteria

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
20247	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Rhodoplane	4.6	7	2.68	1.19	4.36	4.36
18516	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Bradyrhizobium	6.38	5.53	1.56	0.96	2.54	6.9
3889	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae	6.17	7.3	1.44	1.5	2.34	9.25
18323	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Rhodoplanes	3.6	3.23	1.21	1.36	1.97	11.22
17011	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Pedomicrobium	2.88	3.07	1.02	1.38	1.67	12.89
647	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	1.87	0.61	0.97	1.06	1.58	14.47
6715	Bacteria;Proteobacteria;Alphaproteobacteria;Sphingomonadales;Sphingomonadaceae;Kaistobacter	0.83	2.47	0.96	0.77	1.57	16.03
4181	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	1.77	1.24	0.95	0.83	1.54	17.57
17304	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae;Rhodoplanes	3.71	3.38	0.91	1.31	1.48	19.05
2619	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Amaricoccus	2.06	1.27	0.87	1.21	1.42	20.47
5869	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Phyllobacteriaceae;Mesorhizobium	2.99	2.77	0.85	1.32	1.39	21.86
14403	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae	2.45	2.53	0.76	1.35	1.24	23.11
1043	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Bradyrhizobiaceae;Balneimonas	1.38	0.78	0.71	0.78	1.16	24.27
14999	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodospirillales;Rhodospirillaceae;Skermanella	1.19	0.79	0.7	0.83	1.14	25.41
7083	Bacteria;Proteobacteria;Alphaproteobacteria;Rhizobiales;Hyphomicrobiaceae	0.67	1.56	0.68	1.02	1.1	26.51
18304	Bacteria;Proteobacteria;Alphaproteobacteria;Rhodobacterales;Rhodobacteraceae;Amaricoccus	1.84	1.07	0.67	1.55	1.09	27.61

Table 3.7 – SIMPER analysis for Archaea

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
3235	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);Candidatus Nitrososphaera; Candidatus Nitrososphaera gargensis	31.65	23.41	14.8	1.3	27.54	27.54
287	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	24.56	29.6	7.79	1.03	14.49	42.02
7293	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	11.52	8.96	7.45	0.9	13.86	55.88
4838	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	11.13	10.16	3.62	1.03	6.73	62.61
1261	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	4.87	4.8	3.11	1.07	5.78	68.39
1446	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured archaeon	1.29	3.57	2.24	0.51	4.16	72.55
622	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured archaeon	2.68	4.14	2.17	0.98	4.05	76.6
8786	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	1.3	2.22	1.58	0.51	2.93	79.53
490	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	1.89	1.55	1.46	0.62	2.72	82.25
187	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Terrestrial Miscellaneous Gp (TMEG);uncultured euryarchaeote	1.09	1.3	1.09	0.49	2.03	84.28
4677	Archaea;Crenarchaeota;Soil Crenarchaeotic Group (SCG);uncultured crenarchaeote	1.11	1.07	1.02	0.45	1.9	86.18
1094	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Terrestrial Miscellaneous Gp (TMEG);uncultured archaeon	0.25	1.58	0.89	0.31	1.66	87.84
6557	Archaea;Euryarchaeota;Halobacteria;Halobacteriales;Deep Sea Hydrothermal Vent Gp 6 (DHVEG-6);marine metagenome	0.99	0.89	0.87	0.45	1.63	89.47
3553	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Benthic Group D and DHVEG-1;uncultured Thermoplasmatales archaeon	0	1.43	0.71	0.27	1.33	90.8

Table 3.8 – SIMPER analysis for Ascomycota

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
177	Eukaryota;Fungi;Dikarya;Ascomycota;mitosporic Ascomycota;Phoma;Phoma sp. Y3 EG-2010	22.18	11.42	8.09	1.1	12.56	12.56
10976	Eukaryota;Fungi;Dikarya;Ascomycota;environmental samples;uncultured Ascomycota	12.68	8.67	4.99	1.17	7.75	20.31
9381	Eukaryota;Fungi;Dikarya;Ascomycota;mitosporic Ascomycota;Tetracladium;Tetracladium setigerum	8.64	13.75	4.78	1.25	7.42	27.73
9201	Eukaryota;Fungi;Dikarya;Ascomycota;environmental samples;uncultured Ascomycota	0.3	8.74	4.23	0.86	6.57	34.3
8758	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Dothideomycetes;Pleosporomycetidae; Pleosporales;Montagnulaceae;Paraphaeosphaeria;Paraphaeosphaeria sp. N119	5.22	1.34	2.63	0.49	4.09	38.39
9270	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Leotiomycetes;Helotiales;Leotiaceae; Neobulgaria;Neobulgaria premnophila	0.45	4.63	2.09	1.72	3.25	41.64
2956	Eukaryota;Fungi;Dikarya;Ascomycota;environmental samples;uncultured Ascomycota	4.16	0.6	2.09	0.47	3.25	44.89
5118	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Pezizomycetes;Pezizales;Pyronemataceae; Rhodotarzetta;Rhodotarzetta rosea	0.94	3.97	1.96	0.75	3.04	47.93
3790	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Dothideomycetes;Pleosporomycetidae; Pleosporales;Montagnulaceae;Paraphaeosphaeria;Paraphaeosphaeria sp. N119	3.25	5.2	1.87	0.97	2.91	50.84
9771	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Leotiomycetes;Helotiales;Rutstroemiaceae; Lambertella;Lambertella corni-maris	2.73	3	1.27	1.12	1.97	52.81
9440	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Pezizomycetes;Pezizales;Pyronemataceae; Melastiza;Melastiza cornubiensis	2.44	0.45	1.15	0.64	1.79	54.6
10190	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Dothideomycetes;Dothideomycetidae; Capnodiales;Teratosphaeriaceae;mitosporic Teratosphaeriaceae;Capnobotryella;Capnobotryella sp. MA 3612	1.48	1.56	1.11	0.91	1.72	56.32
587	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Pezizomycetes;Pezizales;Pyronemataceae; Melastiza;Melastiza cornubiensis	1.85	1.37	1	0.62	1.55	57.87
9251	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Leotiomycetes;Helotiales;Helotiaceae; Chlorencoelia;Chlorencoelia torta	1.64	1.48	0.98	0.73	1.53	59.4
9126	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Dothideomycetes;Dothideomycetes Incertae sedis;Tubeufiaceae;Tubeufia;Tubeufia sp. CT101	0.27	2.03	0.93	0.91	1.44	60.84
10926	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Sordariomycetes;Sordariomycetidae; Coniochaetales;unclassified Coniochaetales;Coniochaetales sp. GMG_C4	1.54	1.35	0.74	1.16	1.15	61.99
6149	Eukaryota;Fungi;Dikarya;Ascomycota;mitosporic Ascomycota;Phoma;Phoma sp. Y3 EG-2010	1.6	0.51	0.71	1	1.1	63.09
6833	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Leotiomycetes;Helotiales;Leotiaceae; Neobulgaria;Neobulgaria premnophila	0.36	1.38	0.66	1.04	1.03	64.12
9753	Eukaryota;Fungi;Dikarya;Ascomycota;Pezizomycotina;Sordariomycetes;Xylariomycetidae; Xylariales;Xylariaceae;unclassified Xylariaceae;Xylariaceae sp. Y X-28	1.48	1.06	0.65	1.21	1.01	65.12

Table 3.9 – SIMPER analysis for Bacteroidetes

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
9844	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales Bacteria;Bacteroidetes;Sphingobacteria;	13.04	14.29	2.58	1.27	4.88	4.88
4655	Sphingobacteriales;Flexibacteraceae;Cytophaga Bacteria;Bacteroidetes;Sphingobacteria;	6.86	5.09	2.39	1.5	4.51	9.39
21078	Sphingobacteriales;Saprospiraceae	6.76	4.81	2.08	1.21	3.94	13.33
16965	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales	4.12	6.29	2.07	1.27	3.92	17.25
23151	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales	8.03	9.09	1.55	0.87	2.93	20.18
12024	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales	3.17	4.83	1.33	1.3	2.52	22.69
6785	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales	2.84	2.95	1.08	1.01	2.03	24.73
1157	Bacteria;Bacteroidetes;Sphingobacteria;Sphingobacteriales	2.39	1.67	0.83	1.22	1.58	26.31
3697	Bacteria;Bacteroidetes;Sphingobacteria; Sphingobacteriales;Flexibacteraceae;Cytophaga Bacteria;Bacteroidetes;Sphingobacteria;	1.86	1.77	0.74	1.55	1.4	27.71
21732	Sphingobacteriales;Flexibacteraceae;Cytophaga Bacteria;Bacteroidetes;Sphingobacteria;	1.48	1.64	0.65	1.11	1.23	28.94
7897	Sphingobacteriales;Flexibacteraceae;Cytophaga	1.01	0.19	0.57	0.39	1.08	30.02

Table 3.10 – SIMPER analysis for Bacteria

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
10650	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacilluslongiquaesitum	1.86	3.13	1.29	0.98	2.15	2.15
16230	Bacteria;Proteobacteria;Betaproteobacteria	1.77	1.91	0.82	1.18	1.37	3.52
16017	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus	1.35	2.34	0.78	1.07	1.3	4.82

Table 3.11 – SIMPER analysis for Basidiomycota

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
5953	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Tremellomycetes;Tremellales;Tremellaceae;environmentalsamples;uncultured Tremellaceae	28.14	16.44	9.66	1.42	15.55	15.55
7697	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Agaricomycetidae;Boletales;environmentalsamples;uncultured Boletaceae	3.02	15.99	7.11	1.17	11.44	26.99
5542	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Agaricomycetidae;Agaricales;Tricholomataceae;Cleistocybe;Cleistocybe	15.75	14.98	4.94	1.39	7.95	34.94
6298	Tremellales;Tremellaceae;Filobasidiella;Filobasidiella/Cryptococcus neoformanscryptococcus gattii VM276	8.03	5.8	4.23	1.13	6.82	41.76
4734	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Auriculariales;environmentalsamples;uncultured Auriculariaceae	4.23	0.65	1.99	0.79	3.2	44.96
1898	Eukaryota;Fungi;Dikarya;Basidiomycota;basidiomycete yeastsp. BG02-6-6-1-5	3.88	3.09	1.87	0.99	3.01	47.96
7589	Agaricomycetidae;Boletales;Tapinellineae;Tapinellaceae;Tapinella;Tapinella atrotomentosa	1.5	2.7	1.25	1.1	2.01	49.97
4635	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Agaricomycetidae;Agaricales;Pluteaceae;Volvariella;Volvariella gloiocephala	1.69	0.21	0.89	0.37	1.43	51.4
9106	Atractiellomycetes;Atractiellales;unclassified Atractiellales;Atractiellalessp. AH33906	1.65	0.03	0.82	0.77	1.31	52.72
2031	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Polyporales;Polyporaceae;Trichaptum;Trichaptum laricinum	0.74	1.11	0.81	0.52	1.31	54.03
7041	Eukaryota;Fungi;Dikarya;Basidiomycota;Agaricomycotina;Agaricomycetes;Auriculariales;environmentalsamples;uncultured Auriculariaceae	1.42	1.53	0.69	1.27	1.11	55.13

Table 3.12 – SIMPER analysis for Betaproteobacteria

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
16230	Bacteria;Proteobacteria;Betaproteobacteria	14.1	16.43	6.21	1.34	12.06	12.06
5583	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae	4.02	2.87	2.17	1.1	4.21	16.27
7917	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales	4.96	3.66	2.04	1.02	3.96	20.23
17852	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales	5.35	6.33	1.93	1.04	3.76	23.99
3740	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Alcaligenaceae;Achromobacter	2.28	5.02	1.87	0.97	3.63	27.62
8442	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Methylibium	7.32	7.53	1.58	1.37	3.07	30.69
2717	Bacteria;Proteobacteria;Betaproteobacteria;Rhodocyclales	3.43	2.67	1.53	0.96	2.96	33.65
2493	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	5.96	4.4	1.08	1.26	2.1	35.75
6322	Bacteria;Proteobacteria;Betaproteobacteria	2.46	2.72	1.05	1.22	2.04	37.79
20250	Bacteria;Proteobacteria;Betaproteobacteria;Nitrosomonadales;Nitrosomonadaceae;Nitrosospira;Nitrosospiramultiformis	1.03	2.57	1.01	1.28	1.97	39.76
9438	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Oxalobacteraceae	4.69	4	0.9	1.37	1.75	41.51
21934	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Comamonadaceae	2.73	1.86	0.82	1.15	1.59	43.1
14152	Bacteria;Proteobacteria;Betaproteobacteria	2.18	1.81	0.78	1.11	1.51	44.61
11291	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales	2.4	1.13	0.71	1.53	1.38	45.99
14214	Bacteria;Proteobacteria;Betaproteobacteria	1.54	1.31	0.69	1.14	1.33	47.33
16308	Bacteria;Proteobacteria;Betaproteobacteria;Burkholderiales;Methylibium	2.56	2.24	0.57	1.17	1.1	48.42
5742	Bacteria;Proteobacteria;Betaproteobacteria;Methylophilales	1.22	1.23	0.55	1.35	1.07	49.5
7315	Bacteria;Proteobacteria;Betaproteobacteria	1.39	1.6	0.51	1.29	1	50.49

Table 3.13 – SIMPER analysis for Cercozoa

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
188	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Cercomonadidae;environmental samples;uncultured cercomonad	3.28	1.78	2.32	0.87	2.95	2.95
118	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;environmental samples;Cercomonadida environmental sample	3.4	2.62	2.05	1.07	2.6	5.55
183	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	3.69	1.42	1.91	0.95	2.43	7.98
130	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.67	3.72	1.77	1.37	2.25	10.23
479	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	2.91	1.97	1.74	1.09	2.22	12.45
54	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Spongomonas;Spongomonas minima	2.96	4.2	1.74	1.14	2.21	14.66
19	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Cercomonadidae;Massisteria;Massisteria marina	2.69	0.5	1.72	0.47	2.19	16.84
330	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Cercomonadidae;Cercomonas;Cercomonas agilis	2.1	1.71	1.29	1.24	1.64	18.49
372	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.14	2.01	1.25	1.07	1.59	20.08
200	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.6	0.54	1.07	0.58	1.36	21.44
342	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	0.89	1.43	1.05	0.89	1.33	22.77
326	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.99	1.93	1.03	1.54	1.31	24.08
471	Eukaryota;Rhizaria;Cercozoa;Vampyrellidae;Platyreta;Platyreta germanica	1.72	1.07	1.01	1.01	1.29	25.37
283	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Cercomonadidae;Cercomonas;Cercomonas sp. SmallSA	0.48	1.65	0.97	1.18	1.24	26.6
160	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.44	0	0.96	0.35	1.22	27.82
463	Eukaryota;Rhizaria;Cercozoa;Cercozoa sp. COHH1	0	1.5	0.94	0.28	1.19	29.01
15	Eukaryota;Rhizaria;Cercozoa;Silicofilosea;Thaumatomonadida;Thaumatomastigidae;Thaumatomastix;Thaumatomastix sp. CC002-Boundary	0.63	1.17	0.87	0.9	1.1	30.12
51	Eukaryota;Rhizaria;Cercozoa;environmental samples;uncultured cercozoan	1.01	1	0.86	0.89	1.09	31.21
556	Eukaryota;Rhizaria;Cercozoa;Cercozoa sp. WA81p124at18LS	1.39	0.32	0.85	0.71	1.09	32.29
211	Eukaryota;Rhizaria;Cercozoa;Cercomonadida;Cercomonadidae;Cercomonas;Cercomonas sp. SmallSA	0.89	0.64	0.85	0.35	1.08	33.37

Table 3.14 – SIMPER analysis for Crenarchaea

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
3068	Archaea;Crenarchaeota;Soil Crenarchaeotic Group(SCG);Candidatus Nitrososphaera;Candidatus Nitrososphaera gargensis	36.43	28.79	16.1	1.29	37.01	37.01
269	Archaea;Crenarchaeota;Soil Crenarchaeotic Group(SCG);uncultured crenarchaeote	32.66	40.88	11.06	1.19	25.43	62.44
4596	Archaea;Crenarchaeota;Soil Crenarchaeotic Group(SCG);uncultured crenarchaeote	23.42	21.4	9.98	1.03	22.95	85.39
591	Archaea;Crenarchaeota;Soil Crenarchaeotic Group(SCG);uncultured archaeon	3.03	4.91	2.46	0.96	5.65	91.03

Table 3.15 – SIMPER analysis for Deltaproteobacteria

TaxaID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
14772	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae	4.59	3.88	1.6	1.26	2.3	2.3
18170	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae	5.8	6.4	1.57	1.27	2.26	4.56
520	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Polyangiaceae	2.25	2.42	0.97	0.98	1.39	5.95
628	Bacteria;Proteobacteria;Deltaproteobacteria;Desulfuromonadales;Geobacteraceae;Geobacter	1.67	1.07	0.94	0.9	1.36	7.31
12887	Bacteria;Proteobacteria;Deltaproteobacteria;Syntrophobacterales;Syntrophobacteraceae	2.32	2.28	0.92	1.35	1.32	8.63
16452	Bacteria;Proteobacteria;Deltaproteobacteria;Entotheonellales;Entotheonellaceae;Candidatus Entotheonella	1.73	0.85	0.89	0.99	1.28	9.91
13114	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Haliangiaceae	2.2	1.97	0.87	1.31	1.25	11.17
6626	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Haliangiaceae	1.52	1.69	0.87	0.84	1.25	12.41
17409	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales	1.88	1.61	0.86	1.36	1.24	13.65
6751	Bacteria;Proteobacteria;Deltaproteobacteria;Myxococcales;Haliangiaceae	1.86	1.77	0.84	1.14	1.2	14.86

Table 3.16 – SIMPER analysis for Euryarchaea

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
7773	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	29.88	25.69	9.85	1.31	14.13	14.13
4174	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	0.95	7.4	3.86	0.59	5.53	19.66
3554	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	5.68	5	3.75	1.07	5.37	25.03
9289	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	5.13	4.02	3.09	1.15	4.44	29.47
463	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	5.41	2.82	2.99	0.99	4.29	33.75
4198	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	5.06	3.33	2.54	1.06	3.64	37.4
175	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Terrestrial Miscellaneous Gp(TMEG);uncultured euryarchaeote	3.54	1.46	2.18	0.63	3.12	40.52
6515	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	2.5	3.66	2.11	1.05	3.02	43.54
9217	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	0.95	3.86	2.07	0.85	2.97	46.51
6867	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	1.26	3.21	2.03	0.51	2.91	49.42
6222	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	2.57	0.91	1.64	0.43	2.35	51.77
3349	Archaea;Euryarchaeota;Halobacteria;Halobacteriales;Deep Sea Hydrothermal Vent Gp 6(DHVEG-6);uncultured euryarchaeote	2.3	1.1	1.52	0.54	2.18	53.95
4455	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	0.88	2.81	1.43	0.74	2.05	55.99
1341	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	0.52	2.54	1.38	0.61	1.97	57.97
1039	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Terrestrial Miscellaneous Gp(TMEG);uncultured euryarchaeote	2.57	0.48	1.33	0.75	1.91	59.88
5133	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	1.69	1.89	1.23	1.02	1.76	61.64
7491	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Benthic Group D and DHVEG-1;uncultured eukaryote	1.49	1.4	1.19	0.73	1.7	63.34
8957	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	0	2.32	1.16	0.54	1.66	65
1	Archaea;Euryarchaeota;Methanobacteria;Methanobacteriales;Methanobacteriaceae;Methanobacterium;Methanobacterium formicicum	2.25	0	1.13	0.43	1.62	66.62
5915	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	2.06	1.19	1.08	0.91	1.55	68.17
7774	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	1.47	1.26	1.05	0.81	1.51	69.68
10077	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	1.8	0.15	0.95	0.37	1.36	71.04
6085	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	1.3	1.48	0.93	0.96	1.34	72.38
8762	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	1.36	0.91	0.92	0.68	1.32	73.69
2794	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	1.42	0.74	0.9	0.73	1.3	74.99
6109	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Benthic Group D and DHVEG-1;uncultured eukaryote	1.6	0.26	0.85	0.6	1.22	76.21
6346	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;Marine Group II;uncultured marine group II euryarchaeote	0	1.7	0.85	0.41	1.22	77.43
66	Archaea;Euryarchaeota;Thermoplasmata;Thermoplasmatales;CCA47;uncultured eukaryote	0.31	1.21	0.71	0.59	1.01	78.45

Table 3.17 – SIMPER analysis for Firmicutes

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
10650	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacilluslongiquaesitum	23.11	19.59	5.32	1.29	10.7	10.7
16017	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;	17.48	19.49	4.43	1.32	8.91	19.62
22605	Bacteria;Firmicutes;Bacilli;Bacillales;Planococcaceae;Sporosarcina;	2.9	3.62	1.34	1.34	2.7	22.32
8483	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;	3.6	2.63	1.27	1.18	2.57	24.89
9824	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillusfirmus	6.22	5.17	1.26	1.42	2.54	27.42
20699	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacillusfumarioli	1.72	1.89	1.01	1.2	2.04	29.46
17290	Bacteria;Firmicutes;Bacilli;Bacillales;Planococcaceae	1.37	2.28	0.88	0.8	1.77	31.24
4693	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus	3.53	3.05	0.87	1.18	1.74	32.98
9644	Bacteria;Firmicutes;Bacilli;Bacillales;Planococcaceae;Paenisporosarcina	1.72	1.26	0.78	1.16	1.57	34.55
21582	Bacteria;Firmicutes;Bacilli;Bacillales;Planococcaceae	1.73	1.24	0.74	1.4	1.49	36.03
5383	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus	1.45	1.88	0.62	1.29	1.24	37.27
18084	Bacteria;Firmicutes;Bacilli;Bacillales;Paenibacillaceae;Paenibacillus;Paenibacilluschondroitinus	0.79	1.1	0.61	0.91	1.23	38.51
3455	Bacteria;Firmicutes;Bacilli;Bacillales;Bacillaceae;Bacillus;Bacilluslongiquaesitum	1.77	0.97	0.58	1.3	1.18	39.68
9870	Bacteria;Firmicutes;Clostridia;Clostridiales;Peptococcaceae;Desulfosporosinus;	0.47	0.97	0.52	0.95	1.05	40.73

Table 3.18 – SIMPER analysis for Gammaproteobacteria

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
23245	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae	13.54	10.4	2.14	1.17	3.38	3.38
10827	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	1.96	2.88	1.89	0.86	2.98	6.36
14975	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	2.96	5.06	1.42	1.23	2.23	8.59
6484	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	2.07	3.22	1.37	1.34	2.16	10.74
	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae;						
2765	Steroidobacter	4.29	4.59	1.31	1.41	2.07	12.81
	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae;						
15796	Rhodanobacter	0.41	2.55	1.27	0.46	2	14.81
23709	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae	2.92	2.39	1.24	1.17	1.96	16.77
	Bacteria;Proteobacteria;Gammaproteobacteria;Enterobacteriales;Enterobacteriaceae;						
1428	Pantoea	2.24	0.11	1.16	0.29	1.82	18.6
1399	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	3.16	3.3	1.04	1.39	1.64	20.24
14783	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae	2.78	2.34	1.02	1.19	1.61	21.85
16324	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	1.52	2.75	0.97	1.07	1.53	23.38
552	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae	2.69	2.67	0.87	1.37	1.37	24.74
19751	Bacteria;Proteobacteria;Gammaproteobacteria;Chromatiales;Sinobacteraceae	1.41	0.98	0.77	0.81	1.21	25.95
153	Bacteria;Acidobacteria;Acidobacteria-5	1.82	1.77	0.72	1.32	1.13	27.08
10197	Bacteria;Proteobacteria;Gammaproteobacteria;Xanthomonadales;Xanthomonadaceae	1.61	2.03	0.7	1.33	1.1	28.19

Table 3.19 – SIMPER analysis for Glomeromycota

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Archaeosporales;Ambispioraceae;Ambispora;						
10839	Ambispora leptoticha	17.4	0.98	8.64	0.78	10.24	10.24
5199	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus geosporum	12.35	5.16	6.3	0.99	7.47	17.72
2849	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus tunicatum	8.47	8.55	5.91	0.94	7.01	24.73
7284	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus tunicatum	1.16	10.4	5.03	1.01	5.96	30.69
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Unclassified Glomeraceae;						
3084	Glomus mycorrhizal symbiont of Marchantia foliacea	3.81	9.8	4.05	1.21	4.8	35.49
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Paraglomerales;Paraglomeraceae;Paraglomus;						
7251	Paraglomus occultum	5.05	5.6	3.77	0.92	4.47	39.96
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Unclassified Glomeraceae;						
2582	Glomus mycorrhizal symbiont of Marchantia foliacea	2.27	7.34	3.71	1.03	4.4	44.36
8750	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus eburneum	5.06	0	2.53	0.47	3	47.36
8508	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus eburneum	3.09	2.41	2.25	0.72	2.67	50.03
10599	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus irregulare	2.38	2.15	2.11	0.5	2.5	52.54
6752	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus irregulare	3.4	0.69	1.95	0.37	2.31	54.85
3342	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus sp. INBRrmc	3.65	0.32	1.92	0.46	2.27	57.12
7627	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus tunicatum	1.46	3.03	1.82	0.78	2.16	59.29
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Unclassified Glomeraceae;						
9491	Glomus mycorrhizal symbiont of Marchantia foliacea	0.36	3.45	1.8	0.62	2.13	61.42
9974	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus tunicatum	1.05	2.97	1.67	0.85	1.98	63.39
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Unclassified Glomeraceae;						
6241	Glomus mycorrhizal symbiont of Marchantia foliacea	2.31	1.09	1.61	0.4	1.91	65.3
9369	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus tunicatum	1.93	1.5	1.5	0.61	1.78	67.08
2063	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus eburneum	2.09	1.26	1.44	0.51	1.71	68.79
5538	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus geosporum	0	2.65	1.33	0.34	1.57	70.36
6497	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus sp. 5014b24.	1.85	1.51	1.32	0.75	1.56	71.93
2622	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus eburneum	2.25	0.45	1.21	0.72	1.43	73.36
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Unclassified Glomeraceae;						
9610	Glomus mycorrhizal symbiont of Marchantia foliacea	2.04	0	1.02	0.28	1.21	74.57
5291	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus sp. 5014b24.	0.93	1.35	0.96	0.75	1.14	75.71
	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Diversisporales;Acaulosporaceae;Acaulospora;						
9913	Acaulospora scrobiculata	1.24	0.87	0.93	0.65	1.11	76.81
9083	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus irregulare	1.34	0.52	0.89	0.38	1.06	77.87
9495	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus sp. 5014b24.	0.92	0.99	0.86	0.49	1.02	78.9
6985	Eukaryota;Fungi;Glomeromycota;Glomeromycetes;Glomerales;Glomeraceae;Glomus;Glomus sp. W3347	0.65	1.18	0.86	0.41	1.02	79.92

Table 3.20 – SIMPER analysis for Nematode

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Cephaloboidea;						
2089	Cephalobidae; Cephalobus; Cephalobus sp. PS1143	24.6	11.95	9.77	1.19	13.36	13.36
	Eukaryota; Metazoa; Nematoda; Chromadorea; Araeolaimida; Plectoidea; Plectidae;						
310	unclassified Plectidae; Plectidae sp. PDL-2005	4.86	13.35	5.71	0.84	7.8	21.16
2187	Eukaryota; Metazoa; Nematoda; Nematoda environmental samples; uncultured nematoda	7.05	3.03	4.03	0.76	5.5	26.67
	Eukaryota; Metazoa; Nematoda; Enoplea; Dorylaimida; Dorylaimina; Dorylaimoidea;						
283	Qudsianematidae; Qudsianematinae; Eudorylaimus; Eudorylaimus carteri	7.68	4.43	3.48	1.2	4.76	31.43
	Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Cephaloboidea;						
526	Cephalobidae; Acrobeloides; Acrobeloides maximus	7.93	2.51	3.39	1.02	4.64	36.07
	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidoidea;						
872	Aphelenchoididae; Aphelenchoides; Aphelenchoides saprophilus	3.97	3.55	3.02	0.67	4.13	40.2
	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidoidea;						
1211	Aphelenchoididae; Aphelenchoides; Aphelenchoides blastophtorus	4.66	0.44	2.38	0.53	3.25	43.45
388	Eukaryota; Metazoa; Nematoda; Nematoda environmental samples; uncultured nematoda	2.1	3.61	2.09	0.84	2.86	46.32
1861	Eukaryota; Metazoa; Nematoda; Nematoda environmental samples; uncultured nematoda	3.8	1.53	2.04	0.85	2.79	49.1
1337	Eukaryota; Metazoa; Nematoda; Nematoda environmental samples; uncultured nematode	0.22	3.86	1.91	0.71	2.62	51.72
567	Eukaryota; Metazoa; Nematoda; Nematoda environmental samples; uncultured nematode	0.2	2.03	1.08	0.34	1.48	53.2
	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidoidea;						
1217	Aphelenchoididae; Aphelenchoides; Aphelenchoides blastophtorus	0	1.87	0.93	0.48	1.28	54.48
	Eukaryota; Metazoa; Nematoda; Enoplea; Triplonchida; Diphtherophorina;						
2020	Trichodoroidea; Trichodoridae; Paratrichodorus; Paratrichodorus macrostylus	0	1.8	0.9	0.26	1.23	55.71
	Eukaryota; Metazoa; Nematoda; Chromadorea; Monhysterida; Monhysteridae;						
1717	Diplolaimella; Diplolaimella dievengatensis	0	1.67	0.83	0.26	1.14	56.85
	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidoidea;						
2559	Aphelenchoididae; Aphelenchoides; Aphelenchoides saprophilus	0.11	1.62	0.81	0.69	1.11	57.96
	Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Aphelenchoidoidea;						
151	Aphelenchoididae; Laimaphelenchus; Laimaphelenchus heidelbergi	0.83	1.14	0.81	0.82	1.1	59.06
	Eukaryota; Metazoa; Nematoda; Enoplea; Enoplida; Oxystominoidea; Alaimidae;						
2224	Paramphidelus; Paramphidelus sp. IH-2004	0.95	1.59	0.74	1.1	1.02	60.08
	Eukaryota; Metazoa; Nematoda; Enoplea; Dorylaimida; Dorylaimina; Dorylaimoidea;						
2170	Qudsianematidae; Allodorylaimus; Allodorylaimus sp. PDL-2005	1.46	0.39	0.73	0.61	1	61.07

Table 3.21 – SIMPER analysis for Proteobacteria

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
16230	Bacteria; Proteobacteria; Betaproteobacteria	5.19	6.24	2.45	1.28	4.06	4.06
5583	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae	1.43	0.97	0.75	1.15	1.24	5.3
3740	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Oxalobacteraceae	0.82	1.94	0.74	0.88	1.23	6.53
7917	Bacteria; Proteobacteria; Betaproteobacteria; Rhodocyclales	1.79	1.29	0.72	0.95	1.2	7.73
17852	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Alcaligenaceae; Achromobacter	1.9	2.33	0.71	1.04	1.18	8.91
8442	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales; Methylibium	2.6	2.84	0.65	1.26	1.07	9.98
20247	Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales; Hyphomicrobiaceae; Rhodoplanes	1.04	1.54	0.61	1.14	1	10.98

Table 3.22 – SIMPER analysis for Rhizaria

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
1125	Eukaryota; Rhizaria; Cercozoa; Plasmodiophorida; Plasmodiophoridae; Polymyxa; Polymyxa graminis	17.67	22.68	7.2	1.59	13.29	13.29
	Eukaryota; Rhizaria; Cercozoa; Plasmodiophorida; Plasmodiophoridae; Spongospora; Spongospora						
7970	subterranea f. sp. subterranea	6.11	0.57	3.08	0.5	5.68	18.97
	Eukaryota; Rhizaria; Cercozoa; Plasmodiophorida; Plasmodiophoridae; Spongospora; Spongospora						
754	subterranea f. sp. nasturtii	5.39	6.4	2.91	0.7	5.37	24.34
	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae; environmental samples; uncultured						
262	cercomonad	5.53	4.21	1.47	0.92	2.72	27.06
6922	Eukaryota; Rhizaria; Cercozoa; Vampyrellidae; Platyreta; Platyreta germanica	2.52	3.85	1.39	1.13	2.56	29.62
8906	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; Cercomonadidae; Cercomonas; Cercomonas edax	2.92	1.32	1.37	0.76	2.52	32.15
4495	Eukaryota; Rhizaria; Cercozoa; environmental samples; uncultured cercozoan	3.01	2.84	0.97	1.23	1.8	33.94
	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; environmental samples; Cercomonadida environmental						
1862	sample	1.85	3.09	0.9	1.31	1.67	35.61
	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; environmental samples; Cercomonadida environmental						
2428	sample	0.38	1.73	0.71	1.3	1.31	36.92
	Eukaryota; Rhizaria; Cercozoa; Plasmodiophorida; Plasmodiophoridae; Sorosphaera; Sorosphaera						
7671	veronicae	1.46	0.45	0.68	0.76	1.26	38.19
4795	Eukaryota; Rhizaria; Cercozoa; Vampyrellidae; Platyreta; Platyreta germanica	1.18	0.29	0.67	0.54	1.23	39.42
10725	Eukaryota; Rhizaria; Cercozoa; environmental samples; uncultured cercozoan	1.9	1.37	0.66	1.09	1.22	40.64
1949	Eukaryota; Rhizaria; Cercozoa; Silicofilosea; Euglyphida; Trinematidae; Trinema; Trinema enchelys	2.23	1.61	0.66	1.2	1.21	41.85
	Eukaryota; Rhizaria; Cercozoa; Cercomonadida; environmental samples; Cercomonadida environmental						
2520	sample	0.16	1.37	0.63	0.79	1.17	43.02
8726	Eukaryota; Rhizaria; Cercozoa; Silicofilosea; Euglyphida; Euglyphidae; Euglypha; Euglypha f. ciliata	1.42	0.88	0.61	1.14	1.12	44.14
1266	Eukaryota; Rhizaria; Cercozoa; Cercozoa sp. TGS5	2.07	2.14	0.6	1.09	1.11	45.26
	Eukaryota; Rhizaria; Cercozoa; Silicofilosea; Thaumatomonadida; Thaumatomastigidae; environmental						
133	samples; Thaumatomastigidae environmental sample	1.25	1.33	0.59	1.25	1.08	46.34

Table 3.23 – SIMPER analysis for Stramenopiles

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
	Eukaryota;Stramenopiles;Bacillariophyta;Bacillariophyceae;Bacillariophycidae;Bacillariales;Bacillariaceae;						
7703	Nitzschia;Nitzschia sigma	7.11	1.66	3.06	0.65	4.45	4.45
	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;Tribonematales;Tribonemataceae;Tribonema;Tribonema						
1728	intermixum	7.51	3.82	2.92	1.63	4.25	8.71
7006	Eukaryota;Stramenopiles;Oomycetes;Pythiales;Pythiaceae;unclassified Pythiaceae;Pythiaceae sp. PHY1	0.74	5.79	2.65	0.91	3.85	12.56
	Eukaryota;Stramenopiles;Synurophyceae;Ochromonadales;Ochromonadaceae;environmental sample;						
10790	Ochromonadaceae environmental sample	3.62	6.88	2.52	1.14	3.66	16.22
	Eukaryota;Stramenopiles;Oomycetes;Saprolegniales;Saprolegniaceae;Aphanomyces;unclassified						
6049	Aphanomyces;Aphanomyces sp. APH2	4.91	0.77	2.32	0.86	3.38	19.6
3360	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;unclassified Xanthophyceae;Xanthophyceae sp. BX3	4.8	1.21	2.3	0.79	3.35	22.95
8098	Eukaryota;Stramenopiles;Oomycetes;Pythiales;Pythiaceae;unclassified Pythiaceae;Pythiaceae sp. PHY1	4.82	7.2	2.13	1.3	3.1	26.05
	Eukaryota;Stramenopiles;Chrysophyceae;Chromulinales;Chromulinaceae;Spumella;unclassified Spumella;						
8933	Spumella sp. GOT220	3.68	6.71	2.11	1.22	3.07	29.11
9310	Eukaryota;Stramenopiles;Oomycetes;Pythiales;Pythiaceae;unclassified Pythiaceae;Pythiaceae sp. PHY1	2.88	6.23	2.06	1.23	3	32.11
6846	Eukaryota;Stramenopiles;Oomycetes;Peronosporales;Phytophthora;Phytophthora infestans T30-4	1.76	3.26	1.78	0.47	2.6	34.71
	Eukaryota;Stramenopiles;Oomycetes;Saprolegniales;Saprolegniaceae;Saprolegnia;unclassified Saprolegnia;						
6883	Saprolegnia sp. SAP4	1.4	3.82	1.72	0.9	2.5	37.21
7766	Eukaryota;Stramenopiles;Stramenopile sp. MAST-12 KKTS_D3	0.8	3.6	1.72	1.1	2.5	39.71
	Eukaryota;Stramenopiles;Oomycetes;Saprolegniales;Saprolegniaceae;Saprolegnia;unclassified Saprolegnia;						
6278	Saprolegnia sp. SAP4	4	1	1.71	1.29	2.49	42.19
941	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;unclassified Xanthophyceae;Xanthophyceae sp. LA6L-5	2.99	0.97	1.28	1.01	1.86	44.06
2916	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;unclassified Xanthophyceae;Xanthophyceae sp. LA6L-5	1.55	1.99	1.26	0.83	1.84	45.89
3038	Eukaryota;Stramenopiles;Hyphochytriomycetes;Rhizidiomycetaceae;Rhizidiomyces;Rhizidiomyces apophysatus	3.09	0.92	1.22	0.88	1.77	47.66
771	Eukaryota;Stramenopiles;Hyphochytriomycetes;Rhizidiomycetaceae;Rhizidiomyces;Rhizidiomyces apophysatus	0.09	2.16	1.06	0.7	1.55	49.21
5661	Eukaryota;Stramenopiles;environmental samples;uncultured stramenopile	1.9	0.6	1.04	0.71	1.52	50.73
	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;Mischococcales;Pleurochloridaceae;Pleurochloris;						
6010	Pleurochloris meiringensis	2.68	2.16	1.03	1.22	1.5	52.23
	Eukaryota;Stramenopiles;Bacillariophyta;Bacillariophyceae;Bacillariophycidae;Naviculales;Naviculaceae;						
2327	Fistulifera;Fistulifera pelliculosa	2.25	1.81	1	1.16	1.45	53.68
10271	Eukaryota;Stramenopiles;Chrysophyceae;Hibberdiales;Hibberdiaceae;Hibberdia;Hibberdia magna	1.85	0.49	0.99	0.64	1.44	55.12
7587	Eukaryota;Stramenopiles;PX clade;Xanthophyceae;unclassified Xanthophyceae;Xanthophyceae sp. LA6L-5	1.65	0.53	0.96	0.67	1.39	56.51
	Eukaryota;Stramenopiles;Synurophyceae;Ochromonadales;Ochromonadaceae;environmental samples;						
3310	Ochromonadaceae environmental sample	0.68	1.8	0.93	0.86	1.35	57.86
	Eukaryota;Stramenopiles;Bacillariophyta;Bacillariophyceae;Bacillariophycidae;Naviculales;Sellaphoraceae;						
5826	Sellaphora;Sellaphora pupula	1.7	1.21	0.9	1.14	1.32	59.18
9191	Eukaryota;Stramenopiles;Eustigmatophyceae;Eustigmatales;Eustigmataceae;Vischeria;Vischeria punctata	0.58	1.41	0.73	0.89	1.06	60.23

Table 3.24 – SIMPER analysis for Verrucomicrobia

Taxa ID	Species	Arable Av.Abund	Set-aside Av.Abund	Av.Diss	Diss/SD	Contrib%	Cum.%
	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;						
8435	CandidatusXiphinematobacter	16.16	9.01	6.16	1.46	9.47	9.47
20621	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;MC18	6	10.52	4.38	1.41	6.73	16.2
18817	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;MC18	5.3	9	4.19	1.27	6.44	22.64
7559	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales	6.26	6.28	1.65	1.29	2.54	25.17
	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;						
14864	CandidatusXiphinematobacter	2.75	1.99	1.38	1.29	2.13	27.3
	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;						
11997	Verrucomicrobiasubdivision3	2.19	2.37	1.13	1.22	1.74	29.04
1650	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales	2.05	1.16	1.02	0.65	1.57	30.61
16672	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;MC18	1.93	2.32	1	1.39	1.54	32.15
	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;						
8973	Chthoniobacter	2.08	1.61	0.91	1.06	1.39	33.54
6432	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales	1.11	1.27	0.84	0.91	1.29	34.84
23335	Bacteria;Verrucomicrobia;Spartobacteria;Spartobacteriales;Spartobacteriaceae;MC18	1.49	2.03	0.84	1.08	1.29	36.12
	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;Verrucomicrobiaceae;						
10530	Luteolibacter	1.67	1.91	0.83	1.1	1.27	37.4
2220	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales	0.81	1.86	0.82	1.15	1.26	38.66
16082	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales	0.81	1.87	0.75	0.71	1.16	39.82
	Bacteria;Verrucomicrobia;Verrucomicrobiae;Verrucomicrobiales;						
12383	Verrucomicrobiasubdivision3	1.74	1.88	0.74	1.42	1.14	40.96