Characterisation of plant (*Brassica* spp.) and microbial rhizosphere functions

Christopher Charles Hale

A thesis submitted for the degree of Doctor of Philosophy

School of Life Sciences, University of Warwick

September 2017

Contents

List of Tables

List of Figures

Acknowledgements

I would like to thank Gary Bending for his guidance, supervision and general positive attitude over the course of my PhD. Graham Teakle and Hendrik Schäfer for their continued support. Paul Hunter, who helped me settle in to lab work in my first few years at Warwick. Guy Barker, who has helped me delve into the complicated world of plant genetics. Bex Southwell, for the good company during tea breaks. Emma Picot, who has nurtured my interest in bioinformatics and taught me to love R! I would like to give a special thanks to Sally Hilton who has guided me with optimism and encouragement at every stage of my PhD; in the lab, in the field and on the computer.

I am grateful to the Biotechnology and Biological Sciences Research Council (BBSRC) for providing the funding which gave me this wonderful opportunity and the Crop Improvement Research Club (CIRC), which has helped me view the impact of research on everyday life.

I would like to thank my parents for their unwavering belief and support throughout my life; before, during and hopefully after my PhD!

Finally, I would like to thank my partner Emily who has supported me through the highs and lows over the past few years and always sees the best in me.

Declaration

I declare that the work presented in this thesis was conducted by me under the direct supervision of Prof Gary Bending, Doctor Graham Teakle and Doctor Hendrik Schäfer, 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 another degree.

Christopher Charles Hale

List of Abbreviations

ANOSIM - Analysis of similarity
ANOVA - Analysis of variance
BLAST - Basic local alignment search tool
C – Carbon
DEFRA – Department for Environmental, Food and Rural Affairs
DNA - Deoxyribonucleic acid
EU - European Union
GSL - Glucosinolates
ITS - Internal transcribed spacer non-functional region of rRNA
K – Potassium
KEGG - Kyoto Encyclopedia of Genes and Genomes
MEGAN – Metagenome analyzer
N - Nitrogen
NGS - Next generation sequencing
NMDS - Non metric multidimensional scaling
P - Phosphorus
Pi – Inorganic phosphate
Po - Organic phosphate
PICRUSt - Phylogenetic Investigation of Communities by Reconstruction of Unobserved States
PCA - Principal Component Analysis
PCR - Polymerase chain reaction
QIIME - Quantitative Insights into Microbial Ecology
OReGIN – Oilseed Rape Genetic Improvement Network
OSR – Oilseed Rape
OTU - Operational taxonomic unit
RNA - Ribonucleic acid
SparCC Sparse Correlations for Compositional data
SSU - Small Subunit
STAMP - Statistical Analysis of taxonomic and functional profiles
VeGIN – Vegetable Genetic Improvement Network

Summary

The rhizosphere is defined as the area of soil surrounding plant roots, which is influenced by plant exudates. The rhizosphere hosts a diverse and dynamic microbiome, which is shaped by both plant and environmental factors. The plant-microbe and microbe-microbe functional interactions which occur in the rhizosphere can have significant impacts on plant growth. Developing understanding of the composition, functions and interactions of the rhizosphere microbiome and the factors which shape it, may prove valuable to improve agricultural sustainability.

The rhizosphere and bulk soil microbiomes of contrasting *Brassica napus* genotypes growing in the field under high and low N inputs were characterised using amplicon sequencing. Taxonomic identification, functional prediction tools and network analysis were used to gauge how nutrient availability and plant genotype influenced the microbiome. N availability was seen to have a greater influence on composition, function and connectivity of the microbiome than crop genotype, with varying effects on microbes from different Kingdoms.

Metatranscriptome analysis enables analysis of the functioning of the microbiome. The effectiveness of different methods for the separation of root and rhizosphere soil for metatranscriptome analysis was compared. Washing roots in water to separate roots and rhizosphere soil followed by freeze drying prior to RNA extraction was shown to be the best method to avoid distorting the metatranscriptome profile. Metatranscriptome analysis of field grown *B. napus* revealed increases in the rhizosphere relative to soil for protein metabolism functions, and the root compartment contained a high proportion of transcripts related to phage activity.

Plant rhizosphere functions were investigated using transcriptomic analysis of a diverse range of cultivated and wild *Brassica oleracea* plants. Uptake of PO_4 is a vital plant process but the identity of PO_4 transporters is unknown in *B. oleracea.* A number of putative *PHT1* PO_4 transporter genes were identified. Significant differences in expression of the putative PHT1 genes were found between cultivated and wild lines, which may inform future plant breeding strategies.

Chapter 1. Introduction

1.1 The importance of agriculture and future challenges

Plants play a vital role in feeding global human populations. Although diets vary across the world, on average 80% of our calorific intake comes from direct consumption of food crops such as grains, starchy roots, vegetable oils, vegetables and fruits (FAO, 2011). Additionally, the production of feed crops to allow for the keeping of livestock, contributes another 16% to global food needs (FAO, 2011). The human population has been rising at astonishing rates since the industrial revolution and is predicted to reach around 9.7 billion by the year 2050 (United Nations, 2015). Agricultural production will need to increase by an estimated 25-70% in order to feed this population (Hunter *et al*., 2017). Meanwhile the area of available arable land has decreased by 33% between the years of 1975 and 2015 (Cameron *et al*., 2015). Thus agriculture faces increasing demand for crops and decreasing space in which to grow them, presenting the need for greater efficiency of production systems.

A vital driver for increasing agricultural efficiency lies in gaining a better understanding of the plant-soil ecosystem. Modern agricultural practices involve vast inputs of water, chemical fertilisers and chemical protection products. For instance, in 2015 the global demand for nitrogen, phosphorus and potassium (NPK) fertilisers was 237,810,000 tonnes (FAO, 2015). These inputs have previously allowed for the increased productivity of agricultural systems. However they are associated with a number of environmental problems. Run-off of nutrients from agricultural systems can lead to devastating consequences to aquatic systems resulting from hypoxia and eutrophication (Alexander *et al*., 2008). Additionally, unlike N, which can be obtained from the air through the Haber-Bosch process (Haber, 1920) the amount of P available is finite. The phosphorus present in the majority of NPK fertilisers is mined from

rock phosphate. Rock phosphate reserves are limited, and some estimates predict that global rock phosphate supplies may run out by the year 2050 (Cordell *et al.*, 2009). The sustainable use of resources such as fertilisers is one of the main challenges for agriculture (Rockstrom *et al.*, 2017).

A major cause of agricultural losses are pests and diseases, which can account for potential crop yield losses of up to 50% (Oerke, 2005). The use of crop protection products (pesticides, herbicides, fungicides *etc.*) have reduced the level of loss due to these various pests and diseases to 29% (Oerke, 2005). However the use of these compounds may also have unintended and potentially adverse effects on the environment. The herbicide glyphosate has been shown to have the potential for long-term negative impacts on honeybee colony performance (Herbert *et al.*, 2014), and in a separate study was seen to alter the soil bacterial community (Newman *et al.*, 2016), which may have important and currently unknown effects on plant health.

Plant breeding has played a profound role in improving the productivity of agricultural systems. Since the domestication of plants some 11,000 years ago (Doebley *et al.*, 2006), selective breeding of plant varieties has driven a range of crop species suitable for modern cultivation. Subsequent selection and breeding has allowed for the development of highly efficient crop varieties such as the Nobel prize winning development of dwarf wheat by Norman Bourlag, which had much higher yield and disease resistance (Brown, 1970).

1.2 Oilseed rape (*Brassica napus* L.)

Oilseed rape (OSR) is an agricultural crop, which is grown globally for its seed oil, use as animal feed and also as a biofuel. By mass, it is the UK's third most produced crop, behind wheat and barley (Department of Environment, Food and Rural Affairs, 2012) and globally more than 35

million hectares of land is used for OSR growth. The vegetable oil produced from OSR accounts for 13.3% of global oil production and 65.3% of EU oilseed production (European_Commision, 2017). OSR has additional benefits as a break crop in rotation with wheat (Angus *et al.*, 1991) as it boosts disease resistance and yield due to differences in nutrient requirement, and it can act as a catch crop as it can assimilate N from the soil in late autumn and early winter.

OSR (*Brassica napus* L.) is part of the Brassicaceae family which also contains the model plant species *Arabidopsis thaliana.* Within the *Brassica* genus there are many wild species and six crop species. Three of these crop species are diploid; *Brassica rapa* (AA)*, Brassica nigra (BB) and Brassica oleracea (CC)* that have resulted from a polyploidism event between 7.9 and 14.6 million years ago (Lysak *et al.*, 2005).Three are amphidiploid species; *Brassica napus (AACC), Brassica juncea (AABB) and Brassica carinata (*BBCC) which have resulted from the hybridisation of the diploid *Brassica* species (Nagaharu, 1935) and collectively form the triangle of U (Figure 1.1).

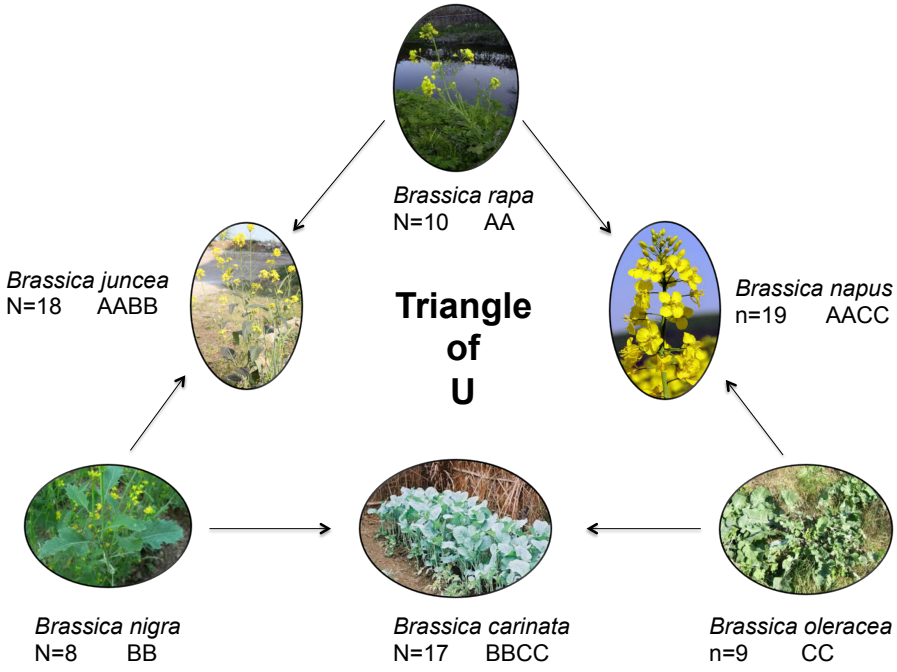


Figure 1.1 Triangle of U (Nagaharu, 1935), representing genetic relationships of the *Brassica* crop types, (n) is number of chromosomes and letter denotes genome classification.

1.3 Plant rhizosphere processes

The significance of roots to plant health cannot be understated. Roots provide structural roles in anchoring the plant to the ground, and act as the main way in which plants uptake water and most of their vital nutrients including N, P and K from the soil.

The rhizosphere is the zone of interaction between plant roots, the soil and the microbial world. The term was first coined by Lorenz Hiltner (Hiltner, 1904) and was used to describe the area of soil surrounding the roots that can be influenced by plant secretions. Plant roots release a number of different chemicals into the rhizosphere. This process is known as exudation. The composition and concentration of plant root exudates varies between plant species (Badri and Vivanco, 2009), growth stages (Chaparro *et al*., 2013), and can change under nutrient deficiency (Tawaraya *et al*., 2014). Root exudate chemicals can comprise a vast range of carbohydrates, amino acids, secondary metabolites, hormones and proteins (Huang *et al*., 2014). Mature plants secrete 5-10% of their total photosynthetically fixed carbon into the surrounding soil (Jones *et al*., 2004), and this figure is higher for growing plants (Bekku *et al*., 1997). This presents a vast input of carbon into the rhizosphere, and coupled with the sloughing of root cap cells and proton exudation from plant roots, represents significant impacts to the surrounding environmental conditions and ecosystem (Jones *et al*., 2009). The function of plant root exudates are varied. Plants have been shown to use root secretions to communicate with other plants, via mycorhhizas in order to warn of incoming pests (Babikova *et al*., 2013), to detect the identity of surrounding plant species and trigger behavioural changes (Semchenko *et al*., 2014), directly suppress pathogen growth (Raaijmakers *et al*., 2008), inhibit seed development of competing plant species (Singh *et al*., 2002) and prevent consumption by herbivores (Huber *et al*., 2016). Plant roots have also been shown to release enzymes such as acid phosphatase in order to increase nutrient uptake, (Tadano *et al*., 1993).

Root exudates also have indirect functional consequences for the plant due to influences on the microbial community. The exudation of certain chemicals has been seen to attract taxa to the rhizosphere in order to control the growth of pathogenic organisms. Secretions of malic acid from the roots of *Arabidopsis thaliana* was linked to significant increases in abundance of a plant-beneficial *Bacillius subtilus* strain (Rudrappa *et al*., 2008). When under attack from herbivores, increases in (E)-b-caryophyllene in maize roots was shown to attract nematode predators which acted as biological control (Rasmann *et al*., 2005). Plant root exudates also have a role in recruiting taxa with capabilities to increase plant nutrient access. Nitrogen, phosphorus and potassium are the three most common growth limiting factors for plants, which have limitations in uptake due to the specific redox state of these nutrients. Plants can only uptake phosphorus as HPO_4^{2-} or $H_2PO_4^-$, and most plants can only assimilate N as nitrates NO_3^- or ammonium NH_4^+. Some plant species have the ability to form symbioses with microbes in order to increase nutrient acquisition. Root exudates such flavones and flavonols have roles in initiating symbioses, including with rhizobia (nitrogen fixing bacteria), which are selectively recruited into root nodules of leguminous plants (Zhang *et al*., 2009). Similarly root exudation of C-glycosylflavonoid was shown to have a role in the initiation of mycorhhizal symbioses (Akiyama *et al*., 2001), which increase plant available P.

1.4 Rhizosphere microbial ecology

The soil is believed to harbour one of the most diverse microbial communities studied (Schloss and Handelsman, 2006). Plant root exudation enriches the rhizosphere with carbon which helps increases the biomass of microorganisms since carbon is the most common growth limiting factor for microbes (Alden *et al*., 2001). This rhizosphere microbiome plays important indirect roles for plant defence by controlling pathogenic microbial taxa *via* natural resource competition. It also plays an important role in nutrient transformation through activities of free living

nitrogen fixing bacteria (Hirsch and Mauchline, 2015), denitrifying bacteria and nitrifying bacteria which have active roles in redox transformations of N. Extracellular release of phytase and phosphate mobilizing enzymes by bacteria can increase the plant available P in the rhizosphere by converting organic P which is unavailable for plants, to inorganic P (Turner and Haygarth, 2005, Richardson *et al*., 2009).
The rhizosphere microbiome has been called a second genome available to plants (Berendsen *et al*., 2012), due to the substantial contribution that it plays for supporting plant growth. Different species of plant and even different genotypes of the same species have been shown to assemble different microbial communities from the same soil environment (Sasaki *et al*., 2013),(Zancarini *et al*., 2012).

Many factors have an influence of the assembly of rhizosphere microbial communities. As well as the nutrient status of the soil (Carvalhais *et al*., 2011), it has been shown that the growth stage of a plant affects the community composition (Micallef *et al*., 2009). However, it is not yet fully known to what extent plant species directly control their unique microbiomes, *A. thaliana* was shown to preferentially recruit Proteobacteria, Bacteroidetes and Actinobacteria in different soil types (Bulgarelli *et al*., 2012).

Rhizosphere and soil microbial communities are dynamic with vast numbers of ecological relationships and interactions. Symbioses are well studied between plants and bacteria (such as *Rhizobia*) and fungi (Arbuscular mycorrhizae) (van der Heijden *et al*., 2016). However, further symbioses between microbial taxa are also present in the rhizosphere, such as the endosymbiosism between Burkholderia bacteria and Rhizopus fungi (Partida-Martinez and Hertweck, 2005), in which the bacterial partner produces a toxin (rhizoxin) which increases fungal pathogenicity to plant roots and subsequently increases available nutrients to the fungus. There are numerous antagonisms within soil and rhizosphere microbial

communities through the production of antibiotic compounds. For example, surfactin and iturin production in *Bacillus subtillis,* which inhibits the growth of the *Pythium ultimum* fungus (Kinsella *et al*., 2009). Competition for nutrients and space is a constant factor which regulates microbial communities, and drives the high diversity in the soil (Hibbing *et al*., 2010). Furthermore, predations such as the grazing of bacteria by rhizosphere protists drives an intricate feedback loop between plants, bacteria and protists which regulates nutrient cycling in the rhizosphere (Bonkowski and Clarholm, 2012). Understanding how various factors influence these interactions and the assembly of rhizosphere microbial communities is the first step towards manipulation the rhizosphere microbiome to promote beneficial functions, thereby improving the efficiency and sustainability of agriculture.

1.5 Analysing microbial communities

The abundance and composition of soil microbial communities is highly variable, and is driven by a number of factors such as soil type, pH (Lauber *et al*., 2009) and temperature (Oliverio *et al*., 2017). Estimates of soil microbial populations were collated from molecular studies by Fierer and show that typically soil microbial communities are comprised of bacteria and fungi at highest biomass (~100-100000 μg biomass carbon per gram of soil), followed by the archaea (~10-100 μg biomass carbon per gram of soil) and protists and viruses (~0.001-10 μg biomass carbon per gram of soil) (Fierer, 2017). As mentioned these communities display massive variation across samples, however typical soil taxa include; Acidobacteria, Verrumicrobia, Bacteriodetes, Proteobacteria, Planctomycetes and Actinobacteria, Basidiomycota, Ascomycota, Zygomycota, Glomeromycota, Chytridiomycota,Crenarchaeota, Euryarchaeota, Parvarchaeota, Rhizaria, Alveolata, Stramenopiles Amoebozoa and Archaeplastida (Fierer, 2017); (Crowther et al., 2014). Rhizosphere microbial communities represent diverse and active populations (Berendsen *et al*., 2012) which are distinct from the

surrounding soil and are again determined by a vast range of abiotic factors and the species of plant (Berg and Smalla, 2009).
Before the development of molecular methods for microbial ecology, culture based methods were used to profile microbes from the environment. This led to improvements in understanding the functional capabilities of a large number of microbes. However only a fraction of the total microbial diversity has been cultured thus far, an estimated 1% (Hugenholtz, 2002). Even as culture collections continue to grow, there is a limit to the number of microorganisms that can be brought into culture due to complex dependencies and thus there is a scarcity of certain taxa (Stewart, 2012). The ability to simultaneously identify and quantify the diverse microbial taxa from an environment is not possible using culture based techniques.

Molecular based sequencing methods have revolutionised microbial ecology. Microbes exist in dynamic communities with a large range of ecological interactions. The ability to determine the abundance of organisms relative to each other, and to discern the taxonomic identity of the microbes present in an environment can reveal a detailed picture of the microbial ecosystem. Amplicon sequencing is the most common method used for the identification and relative quantification of microbiomes. The sequencing of specific regions of DNA, amplified by PCR from environmental DNA samples can be used to provide taxonomic and functional characterisation of communities. One of the biggest technological advances in the study of microbial communities was the ability to use the 16S rRNA gene for bacterial identification, and the 18S rRNA gene for eukaryotic identification (Woese *et al.*, 1990). This enabled the identification of previously unculturable or inaccessible microoganisms. The ribosomal RNA genes are used for molecular classification of organisms due to highly conserved regions of the gene to which primers can bind, and hypervariable regions which allow for unique determination of identity (Van de Peer *et al.*, 1996). Taxonomic classification using rRNA

genes is facilitated by the use of Operational Taxonomic Units (OTUs) based on sequence similarity. One of the benefits of the use of rRNA genes and OTU clustering is that estimations of taxonomic identity can be made, without the need to define bacterial species (Schmidt *et al*., 2014). One of the drawbacks of using rRNA genes for identification is that some species may share more than 97% similarity of rRNA genes which can lead to OTUs representing multiple species, for example *Bacillus globisporus* and *Bacillus psychrophilus* were seen to exhibit more than 99.5% sequence simialtity of the 16S rRNA gene which would lead to these distinct species being labeled as one OTU (Fox *et al*., 1992). Also, prokaryotic taxa may have between one and fifteen copies of the 16S rRNA gene (Lee *et al*., 2009), which can distort diversity estimates. For example, in a forest soil study, the relative abundance of Acidobacteria was artificially increased and the relative abundance of Firmicutes was decreased due to differences in 16S rRNA copy number (Vetrovsky and Baldrian, 2013). Due to the need to amplify the rRNA genes (amplicons) by PCR, biases can be introduced due to preferential binding of primers to the DNA of certain taxa, giving erroneous estimates of the relative abundance. Choice of primers was shown to be one of the main causes of errors in Illumina DNA amplicon sequencing (Schirmer *et al*., 2015).
High throughput targeted amplicon sequencing of the rRNA genes is a popular tool for the determination of microbial communities across environments. However, there are drawbacks of using targeted amplicon DNA sequencing to profile the microbial community. This method does not target the active community. The central dogma of molecular biology, is that DNA → RNA→ Protein (Crick, 1970). DNA represents the genomic potential of the organisms present, whereas RNA represents the expression of genes by living organisms. One potential issue of using DNA to identify and quantify the microbial community is that DNA has been shown to remain present in the soil environment for years as ‘relic’ DNA (Carini *et al*., 2016). The relic DNA has been shown to account for 40% of the total prokaryotic and fungal communities and as a consequence, has

the potential to inflate the diversity of the microbiome and shift the microbial community composition. The use of RNA instead of DNA to analyse microbial communities has allowed for the taxonomic identification of the active community (Thomas and Cebron, 2016).

1.6 Whole community identification

The ultimate aim within the study of microbial ecology is to simultaneously determine the identify and function of all members of microbial communities. Advances in sequencing technology have enabled the study of metagenomes and metatranscriptomes in the soil environment. Metagenomic studies seek to identify the complete genetic potential of the microbial community while metatranscriptomes can reveal the complete functional output of a microbial community. This rapidly advancing field of study currently relies on fragmentation of the total DNA or RNA in the environment, and the allocation of these stretches of sequence to reference databases based on percentage similarity. RNA-Seq approaches allow for the analysis of the metatranscriptome without prior selection of target genes, such as in RNA microarrays (Wang *et al.*, 2009). Metatranscriptomics approaches have identified differences in the microbial composition and function of soil and rhizosphere communities from a variety of plants (Turner *et al.*, 2013). Whilst these techniques have opened up new possibilities for linking the composition of microbial communities to function, there is a key limitation: The databases to which sequences are aligned do not (by any means) contain the genomes of all living taxa. Recent estimates have estimated there may be over one trillion microbial species on Earth (Locey and Lennon, 2016). Furthermore, the cost of ‘meta-omics’ approaches are much higher than targeted amplicon sequencing and significant computational power is required to cope with the volume of data produced.

1.7 The plant root transcriptome

High throughput sequencing methods have also revealed vast amounts of information regarding plant transcriptional activity. As mentioned previously, plant roots are a hotspot for various processes including nutrient uptake and secretion of chemicals as root exudates. These processes are coordinated through expression of genes in the root. Understanding how plants react to different environmental conditions such as drought (Janiak *et al.*, 2016) and nutrient limitation (Curci *et al.*, 2017) helps develop the understanding of the impacts that the environment can have on the plant, and the subsequent influence the plant will have on it's environment. By comparing the transcriptional fingerprint of plants with known functional differences, RNA-Seq transcriptomes can help identify the underlying causes of these changes and help identify targets for breeding or genetic modification. For example, Singh and co-workers (Singh *et al.*, 2016) were able to use rice cultivars to discover genes that control phenotypes such as lateral root development. At a larger scale, the root transcriptome of sweet potato (Ponniah *et al.*, 2017) and spinach (Xu *et al.*, 2015) plants have been compared with those of their wild ancestors in order to detect differences in expression pattern. This sort of approach leads to real improvements in understanding how plant systems develop over time and can reveal practical information about plant responses to biotic and abiotic stresses which can reveal potential breeding targets and ultimately help improve agricultural efficiency.

1.8 Aims and objectives

The overall aims of this thesis were to develop approaches to investigate plant and microbe functions in the rhizosphere. The work presented here has been divided into five chapters. Chapter 1 presents a general introduction into the research area. Chapters 2 - 4 are experimental chapters with defined structure (Introduction, Materials and Methods, Results, Discussion), and Chapter 5 provides a general discussion of the work.

The aim of Chapter 2 was to determine the relative abundance and potential functions of the microbial communities in the soil and in the rhizosphere of field grown *B. napus* and identify what impact N fertilisation and plant genotype had on the composition and function of these bacterial, protist and fungal communities.

The aim of Chapter 3 was to evaluate methods for extraction and preservation of RNA from the soil root and rhizosphere of field grown *B. napus* and investigate differences in the active microbial communities and functions of these compartments.

The aim of Chapter 4 was to investigate functional processes from the side of the plant, and identify potential phosphate transport genes in *B. oleracea* using similarity to *Arabidopsis thaliana* sequences. Further aims were to identify localised expression in the root and leaf tissues from a variety of wild and cultivated *Brassica* plants and finally detect differences in expression patterns between the wild species and crop types.

Chapter 2. The effect of nitrogen fertilisation and plant genotype on the diversity, composition and function of microbial communities in the rhizosphere and bulk soil of field grown oilseed rape.

2.1 Introduction

2.1.1 The rhizosphere

The microbial communities that inhabit the rhizosphere (plant roots and close-surrounding soil) can have substantial effects on plant health. Many rhizosphere microbes can influence plant health directly, acting as pathogens (Hilton *et al*., 2013, Raaijmakers *et al*., 2008) or as beneficial symbiotic partners (Bonfante and Genre, 2010). Others influence plant growth indirectly by increasing plant access to nutrients, for example through the release of nutrients from organic substrates (Jorquera *et al*., 2008), or by suppressing the growth of plant pathogens (Mendes *et al*., 2011).

Microbial communities inhabiting the rhizosphere are made up of a combination of prokaryotes, protists (single celled eukaryotes) and fungi (Philippot *et al*., 2013). The drivers which control the assembly and function of these communities are most well studied for bacteria (Bell *et al*., 2015) and fungi (Paungfoo-Lonhienne *et al*., 2015), while protists have received limited focus at the community level, despite their importance (Wilkinson *et al*., 2012). This reflects difficulties in creating cultures, the lack of suitable conserved primers for targeting protist small subunit rRNA (Adl *et al*., 2014), and the scarcity of reference sequences and poor annotation in rRNA databases (Geisen *et al*., 2015). However, protist root pathogens, such as the *Brassica napus* root pathogen *Plasmodiophora brassicae*, can have direct detrimental effects on plant health *via* infection of the root and formation of galls (clubroot) (Hwang *et al*., 2012). Protists can also indirectly affect plant health, for example, Amoebozoa can act as grazers of bacteria, causing shifts in bacterial community composition and

function in the rhizosphere (Rosenberg *et al*., 2009), with consequences for plant growth (Krome *et al*., 2009).

Current evidence suggests that rhizosphere microbial communities are comprised of specialists and generalists (Pandit *et al*., 2009). Generalist species generally have the potential to occur over a wide range of niches and usually have a wide range of potential metabolic processes, however generalist organism, such as *Desulfobulbus* are usually outcompeted in the microbial world by specialists (Carbonero *et al*., 2014). Specialist microbes, such as the symbiotic fungi (Bonfante and Genre, 2010) and rhizobia (Kamboj *et al*., 2008), show changes in abundance across ecological niches, such as root hairs or primary roots (Saleem *et al*., 2015). The host selection can be specific, for example rhizobia are only able to colonise plants from the *Fabaceae* genus. Specialist interactions, can even be specific to bacterial strain (Andrews and Andrews, 2017) and have impacts on plant growth that lead to significant yield increases (Peoples and Craswell, 1992). Specialist microbes are of particular interest in regards to the assembly of rhizosphere microbial communities as the interaction of a host plant can be unique to plant species, or even genotype (Muller *et al*., 2015), and the interaction can have a significant impact on growth.

Evidence is emerging that some free-living bacteria may also be rhizosphere specialists (Farrar *et al*., 2014). Proteobacteria and Actinobacteria have been shown to be root colonisers in *Arabidopsis thaliana* (Lundberg *et al*., 2012), and additionally the Actinobacteria have been implicated as a phylum which is directly recruited into active Arabidopsis plant roots in response to metabolic signals (Bulgarelli *et al*., 2012).

2.1.2 Drivers of microbiome assembly

Although there is an innate microbial community associated with the seed, the microbiome of the rhizosphere is primarily assembled from the

surrounding soil (Philippot *et al*., 2013). Assembly and organisation of the crop rhizosphere microbiome is the result of interactions between a myriad of factors (Edwards *et al*., 2015), which can broadly be categorized into; environment, crop genotype, and management practices (Busby *et al*., 2017).

Environmental factors, such as pH, moisture and salinity have an important role in shaping the microbial community structure of both the soil and the rhizosphere (Marschner *et al*., 2004, Griffiths *et al*., 2011). Geographical distance has also been shown to contribute to the selection of the rhizosphere microbial community for a number of plant species (Pereira e Silva *et al*., 2012). The species of plant itself (Berg and Smalla, 2009) and growth stage (Sugiyama *et al*., 2014b) also has an effect on the composition of the rhizosphere microbiome. Similarly, genotypes within a species can show differences in composition of rhizosphere microbial communities (Bulgarelli *et al*., 2015). However, an understanding of the relative importance and interactions of these factors under field conditions is still being developed. In a field grown sugarcane crop, plant genotype was found to have no significant effects on rhizosphere community composition (Yeoh *et al*., 2016). In agricultural systems, management practices, such as the application of fertiliser can influence soil bacterial community composition (Ramirez et al., 2010) (Fierer et al., 2012), and can affect rhizosphere composition (Gosling *et al*., 2013), although it is unclear whether this reflects direct effects, or indirect impacts via effects on the plant. Additionally, the responses of rhizosphere biota to plant and environmental drivers of assembly can be different. For example bacterial communities in the rhizosphere of soybean responded to N availability (Sugiyama *et al*., 2014b) while fungi did not (Sugiyama *et al*., 2014a).

Community composition is also shaped by the interactions between taxa. The highly diverse and abundant communities of the rhizosphere and soil allow a variety of ecological interactions, such as predation, parasitism, symbioses and direct competition (Kent and Triplett, 2002). Network analysis is an emerging approach to interpret these interactions in the

rhizosphere. In essence, networks display positive and negative correlations in the relative abundances of microbial taxa. Connectivity, which is a measure of the number of these interactions, has been shown to be greater in the rhizosphere than in soil under controlled conditions in the rhizosphere of *Avena fatua (*Shi *et al., 2016)*, and *Jacobaea vulgaris* (Yan *et al*., 2016). However, until recently, rhizosphere network studies have been mainly conducted in greenhouse (pot) conditions. Field based network analyses seem to suggest the connectivity of the microbial community in the soil may be greater than the rhizosphere (Fan *et al*., 2017). The extent to which these characteristics scale from greenhouse to the field environment are unclear.

2.1.3 Functional capabilities of microbial communities

Analyses of rhizosphere community composition typically focus on community shifts, the functional consequences of which have received much less consideration. Through amplicon sequencing of DNA regions that are both conserved and highly variable, it is possible to identify a vast number of different rhizosphere taxa. However, attempting to identify the functional capabilities or activity of communities requires more complex sequencing efforts, such as metagenomics (Ranjan *et al*., 2016) or metatranscriptomics (Bashiardes *et al*., 2016), which come with complications, such as gaining sufficient coverage and increased cost (Ogura, 2014). However, tools have been developed that utilise data from rRNA amplicon sequencing to estimate function. Phylogenetic Investigation of Communities by Reconstruction of Unobserved States (PICRUSt) (Langille *et al*., 2013) uses 16S rRNA genes in order to predict a likely metagenome using an ancestral-state reconstruction algorithm. Although PICRUSt predicted microbiomes have limitations, such as a dependency on sequenced genomes, and possible biases based on sample type, they can offer a valuable insight into the functional characteristics of a bacterial community, which have been found to correlate with full shotgun sequencing efforts (Zarraonaindia *et al*., 2015).

FUNGuild (Nguyen *et al*., 2016) places ITS sequences of fungi into functional guilds, based on literature searches, which are based on the most likely lifestyle of the taxonomic group to which the sequence aligns.

2.1.4 Aims

The aims of this chapter were: 1.) To compare the bacteria, fungal and protist communities of the rhizosphere and bulk soil of *Brassica napus* growing in the field, and determine the relative importance of N fertilisation and the genotype of the plant and microbial community composition; 2.) To determine whether the shifts in community composition associated with genotype and N fertilisation could have functional significance, using *in silico* predictions; and 3.) To identify interactions between taxa using network analysis and determine whether N treatment or genotype had an affect on the composition and characteristics of microbial networks.

2.2 Materials and Methods

2.2.1 Location, experimental design and sampling procedure

Rhizosphere and soil samples were obtained from a field site at Rothamsted Research, Hertfordshire, UK (51.803896°N -0.362505°E). The site was managed as a part of the Oilseed RapE Genetic Improvement Network (OREGIN) project. Ten different *B. napus* genotypes were grown in 10 m x 3 m plots. Each genotype was grown in four plots, two of which received a standard N fertiliser rate of 210 kg/ha of and two of which received no additional N. This allowed comparisons of community composition between; compartment (rhizosphere and soil), N levels in both the soil and rhizosphere and the genotype of the rhizosphere samples.

Soil and rhizosphere samples were collected from plots of ten selected genotypes in April 2014, during the pod development stage. The

genotypes were from a variety of geographic locations, crop types and oilseed breeding traits, which were selected to encompass genetic variability available within oilseed rape and its wider breeding material. The selected lines were; Canard, Couve nabica, Darmor, Ningyou 7, Rocket, Tapidor, Temple, Victor, York and Yudal (Table 2.1).

Table 2.1 List of *Brassica napus* genotypes used in this study, with additional origin, crop type and trait information. (GSL = glucosinolates)

Variety	Country of origin	Crop type	Traits
Canard	Great Britain	Winter forage rape	High GSL, high erucic acid
Couve nabica	Portugal	Couve nabica	High erucic acid
Darmor	France	Winter OSR	Low GSL, low erucic acid
Ningyou 7	China	Winter OSR	High GSL, high erucic acid
Rocket x Lizard DH line	Great Britain	Winter OSR	Low erucic acid
Tapidor DH	France	Winter OSR	Low GSL, low erucic acid
Temple	Great Britain	Winter OSR	Low GSL, low erucic acid
Victor	Sweden	Winter OSR	High GSL, high erucic acid
York	Great Britain	Swede	High GSL, high erucic acid
Yudal	South Korea	Spring OSR	High GSL, high erucic acid

Sampling of the rhizosphere was conducted by removing four plants from the soil, shaking to remove any loose adhering soil and cutting the root system at the root-stem junction. Sampling of the bulk soil was conducted by removal of the top 2 cm of soil and collecting 10 cm soil from an area adjacent to the plant. Lateral roots with the adhering soil were removed from the taproot, cut to 5 mm length and pooled. The soil was sieved (<4 mm). DNA was subsequently extracted from 0.5 g of rhizosphere and soil samples using the FastDNA® SPIN Kit for Soil, according to the manufacturers instructions (MP Biomedicals, Santa Ana, USA). DNA was

quantified using the Qubit 2.0 Fluorometer (Life Technologies, USA) with the broad range DNA assay and diluted to 5 ng/μl.

2.2.2 Determination of microbial community diversity and structure

PCR amplification was performed on the extracted DNA using 515f and 806r primers to amplify bacterial sequences, Euk_1391f and Euk_Br primers to amplify protist sequences and ITS3 and ITS4 primers for fungi (Table 2.2). The primer sets were modified at the 5′ end with adaptors, (TCG TCG GCA GCG TCA GAT GTG TAT AAG AGA CAG) – forward adaptor and (GTC TCG TGG GCT CGG AGA TGT GTA TAA GAG ACA G) – reverse adaptor (Illumina).

Table 2.2 Primer pairs used to amplify target DNA to identify bacterial, protist and fungal communities

Amplicon	Forward primer	Reverse primer	Reference
16S (V4 SSU rRNA)	**515f** 5' GTG CCA GCM GCC GCG GTA A 3'	**806r** 5' GAC TAC VSG GGT ATC TAA T 3'	(Caporaso *et al*., 2012)
18S (V9 SSU rRNA)	**Euk_1391f** 5' GTA CAC ACC GCC CGT C 3'	**Euk_Br** 5' TGA TCC TTC TGC AGG TTC ACC TAC 3')	(Caporaso et al., 2012, Amaral-Zettler et al., 2009)
ITS (ITS2 region)	**ITS3** 5' GCA TCG ATG AAG AAC GCA GC 3'	**ITS4** 5' TCC TCC GCT TAT TGA TAT GC 3'	(White *et al*., 1990).

The final volumes in all of the PCR reactions were: 3μl of DNA template (5 ng/μl), 1.25μl of forward primer (10 μM), 1.25μl of reverse primer (10 μM), 12.5μl of Q5 readymix (New England Biolabs, Hitchin UK), made up to a total volume of 25μl with water. The PCR was conducted using an initial denaturation at 98 °C for 30 seconds followed by 25 cycles (16S) or 30

cycles (ITS/18S) of: 10s denaturing at 98 °C, 15s annealing at 57 °C, 20s elongation at 72 °C followed by a final elongation step of 5 min at 72 °C in a Multigene Optimax thermal cycler (Labnet, USA). PCR products were purified using the AMPure XP beads (Beckman Coulter, Germany) according to the manufacturer's instruction. For each amplicon, library preparation was concluded by performing a PCR using the v1 and v2 Nextera XT Index kit (Illumina, USA). PCR was conducted using 3μl DNA (4 nM), 13μl Q5® Hot Start High-Fidelity 2X Master Mix (New England Biolabs), 2.5μl of each Nextera index primer (5 μM), made up to 26 μl in water. The PCR was conducted using an initial denaturation at 95 °C for 3 minutes followed by 8 cycles of: 20s denaturing at 98 °C, 15s annealing at 55 °C, 15s elongation at 72°C followed by a final elongation step of 5 min at 72 °C. The PCR product was purified as before. The DNA concentration was measured using the Qubit 2.0 Fluorometer (Life Technologies, USA) and diluted to 4 nM. The samples were pooled and 300bp paired end sequencing was performed on an Illumina MiSeq system (Illumina, USA).

2.2.3 Bioinformatics analysis

The raw sequences were automatically demultiplexed by the Illumina MiSeq. Trimmomatic v0.35 (Bolger *et al*., 2014), was used to remove low-quality bases. Paired-ends reads were assembled by aligning the forward and reverse reads, primers were trimmed and quality filtering was conducted with the setting (–fastq_maxee 0.5) using USEARCH (Edgar, 2010).

Unique sequences were sorted by abundance and singletons in the data set were discarded. Sequences were clustered to OTUs at 97% minimum identity threshold and chimeras were removed. Taxonomy was assigned using Quantitative Insights into Microbial Ecology (QIIME 1.8) (Caporaso *et al*., 2010) using the Greengenes reference database for 16S (McDonald *et al*., 2012), the Silva database for 18S (Quast *et al*., 2013)) and the ITS Unite database for ITS (Kõljalg *et al*., 2013), using the UCLUST algorithm

for 16S and 18S rRNA amplicons and the blast algorithm for the fungal sequences. An OTU table consisting of the OTU counts for each sample and the taxonomic classification was constructed with make_otu_table.py.

2.2.4 Statistical analysis

Alpha diversity (the diversity within samples) and Beta diversity (diversity between samples) were calculated using the Phyloseq package (McMurdie and Holmes, 2013) in R (R Core Team, 2015). The OTU table was normalised using a single random subsampling at the lowest depth library. The statistical significance of the differences in alpha diversity were conducted using a Kruskall Wallis test. The pairwise comparisons were conducted using Dunn tests and P values were corrected using Benjamini-Hochberg multiple corrections. These statistical analyses were conducted in R using the Rcompanion (Mangiafico, 2015) plyr (Wickham, 2011) and FSA (Ogle, 2017) packages. Analyses of PICRUSt generated metagenomes was conducted using Statistical Analysis of Metagenomic Profiles (STAMP) (Parks *et al*., 2014). Significant differences between relative abundance of taxa were generated using Kruskall-Wallace test and P values were corrected using the Benjamini-Hochberg multiple correction method. Non-metric multidimensional scaling plots were generated using ordinations from Bray-Curtis similarity and were visualised using the R packages ggplot2 (Wickham, 2009) and vegan (Oksanen *et al*., 2017). Primer 6 (PRIMER, version 6, Primer-E, Plymouth Marine Laboratory, UK) was used to conduct beta diversity analysis of the OTU tables, grouped at all taxonomic levels. A Bray-Curtis similarity matrix was created using the OTU tables. Following this differences between groups were statistically analysed using a one-way analysis of similarity (ANOSIM), which calculates dissimilarity between samples and a significance level. In order to determine which OTU (or phylum) were contributing to the differences observed in the ANOSIM, a similarity percentages (SIMPER) test was performed.

2.2.5 Functional analysis

PICRUSt was used to generate predicted metagenomes of the 16S community. The analysis was conducted using the Huttenhower labs online tool (available at https://huttenhower.sph.harvard.edu/galaxy) using the default settings. N cycle genes were identified from the work of Zhu and coworkers (Zhu *et al*., 2016); N-fixation (nifD, nifH), Urea hydrolysis (ureC), Assimilatory nitrate reduction (nasA, nirA), Nitrification (pmoA-amoA, hao) a,d Denitrification (norB, nosZ). The contribution of OTUs to the frequency of these genes in the predicted metagenome were calculated using contributions_to_metagenome.py and dividing the sum of the "countcontributionedbyOTU" for each sample by the number of normalised reads per sample. Significant differences in relative abundance of OTUs containing N cycle genes between rhizosphere and soil were calculated using a Wilcoxon ranked sum test. Differences under N treatment were calculated using a Dunn test. FUNGuild analysis provides a "guild" to assign to OTUs based on published studies. These analyses were conducted using the online guilds application (available at http://www.stbates.org/guilds/app.php). Significant differences were identified between relative abundance of guilds using Dunn tests and P values were corrected using Benjamini-Hochberg multiple corrections in R.

2.2.6 Network analysis

In order to investigate correlations of OTU abundances in samples, a network analysis was conducted. Networks were constructed using OTU tables pre-filtered for the top 10% most abundant OTUs. SparCC (Friedman and Alm, 2012) is a tool which estimates correlations between samples in compositional data sets, and computes P-values representing the degree of correlation. SparCC was conducted on the OTU tables with 20 iterations and the Pseudo P-values were assessed on 100 bootstrap replicates, and corrected for multiple comparisons using the Benjamini Hochberg procedure. Networks were subsequently generated in Gephi (Bastian *et al*., 2009) using Spearmans correlations of at least 0.25 and

Benjamini Hochberg P<0.05. Nodes were coloured by class and edges by correlation, the network was structured using the Fruchterman-Reingold algorithm.

2.3 Results

A total of 25,360,302 sequencing reads were returned and after filtering and paring of reads, there were 19,605,313 contigs that were assigned taxonomy (Table 2.3). Filtering was conducted in order to remove OTUs from the 16S table annotated as chloroplast, eukaryota, archea and mitochondria. Fungal, plant and metazoan sequences were filtered from the 18S rRNA OTU table.

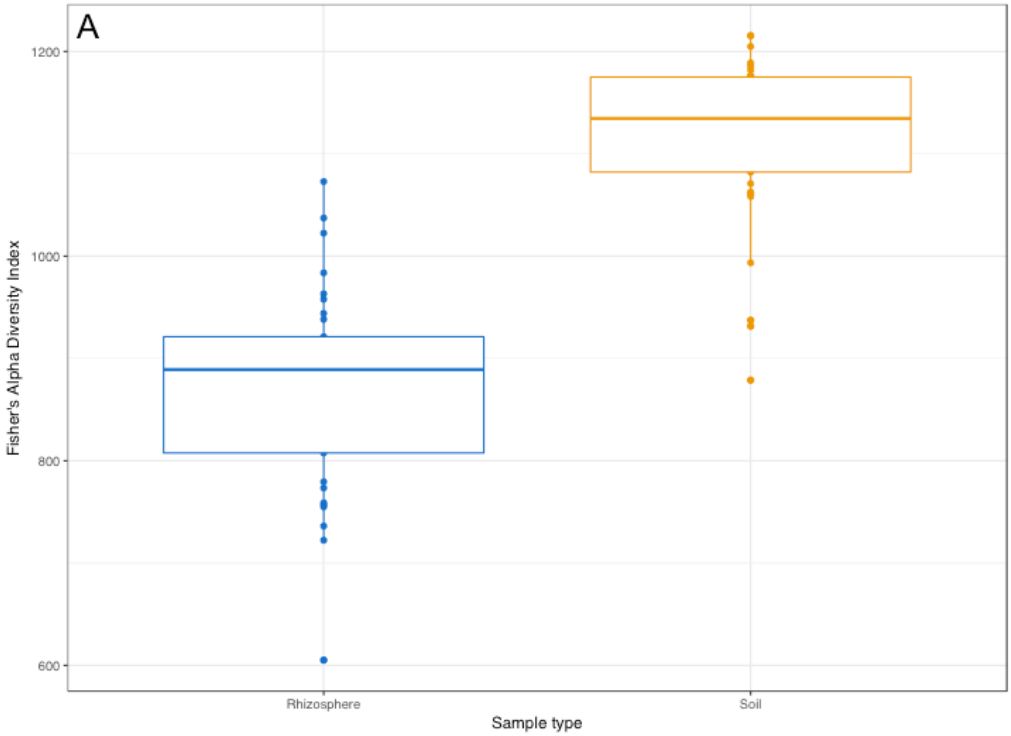
Table 2.3 Number of sequences assigned to each amplicon of study before and after filtering steps

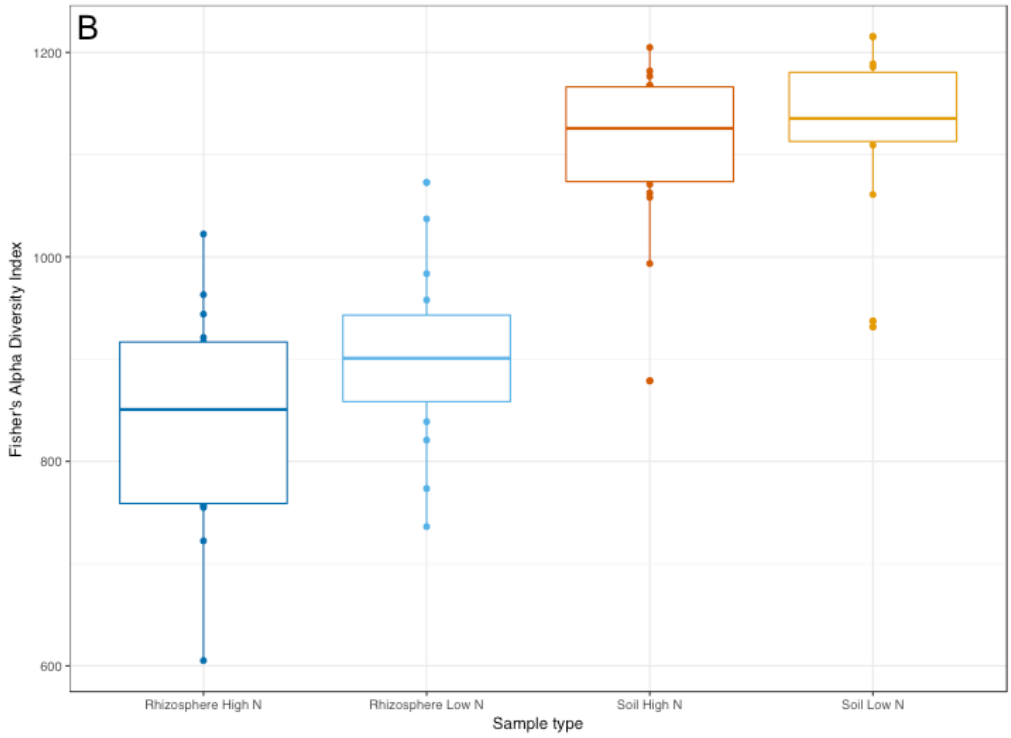
	Pre-filter		Post filter	
	Number of sequences	**Number of OTUs**	**Number of sequences**	**Number of OTUs**
16S	6,007,398	12,384	5,277,617	10,634
18S	5,064,818	11,297	164,238	1,311
ITS	2,003,810	1,916		

Data was denoised by removing any OTUs with less than 33 total sequences across the samples (at least one sequence in half of the samples). Datasets were normalised to 7,538 bacterial sequences per sample from 5,258 OTUs, 358 protist sequences per sample from 529 OTUs and 2,479 fungal sequences per sample from 1,187 OTUs.

2.3.1 Bacterial community composition

Bacterial Fisher's alpha diversity was significantly (P<0.001) greater in soil samples than in rhizosphere samples (Figure 2.1A), however N had no significant effect on bacterial alpha diversity in the rhizosphere (P=0.5) or the soil (P=0.7) (Figure 2.1B). Genotype had no significant impact on rhizosphere bacterial alpha diversity (Figure 2.1C).


A
1200
1000
800
600
Fisher's Alpha Diversity Index
Rhizosphere
Soil
Sample type


B
1200
1000
800
600
Fisher's Alpha Diversity Index
Rhizosphere High N
Rhizosphere Low N
Soil High N
Soil Low N
Sample type

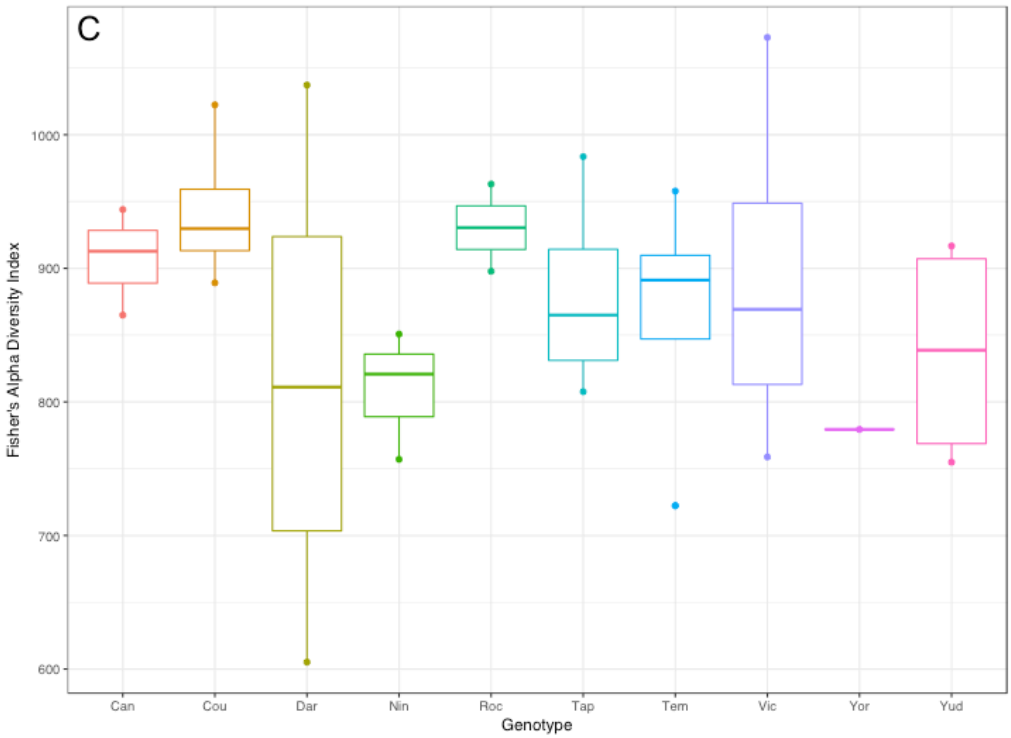


Figure 2.1 Box and whisker plot showing Fisher's alpha diversity of the bacterial communities in A.) the rhizosphere (blue) and soil (orange) samples. B.) The rhizosphere high N (dark blue), rhizosphere low N (light blue), soil high N (dark orange) and soil low N (light orange) samples. C.) The rhizosphere for each of the genotypes. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range.

The differences between bacterial communities were visualised using a MDS plot generated from a Bray-Curtis similarity matrix (Figure 2.2). ANOSIM showed that the bacterial communities of the soil and rhizosphere were significantly different from each other (R= 0.93), (P<0.001). There were no significant differences between the bacterial communities from high or low N soil (R=0.01), (P=0.3), however the rhizosphere high N and rhizosphere low N bacterial community compositions were found to be significantly different (R=0.22), (P=0.001). Genotype had no significant effect on the bacterial rhizosphere communities (R=0), (P=0.5).

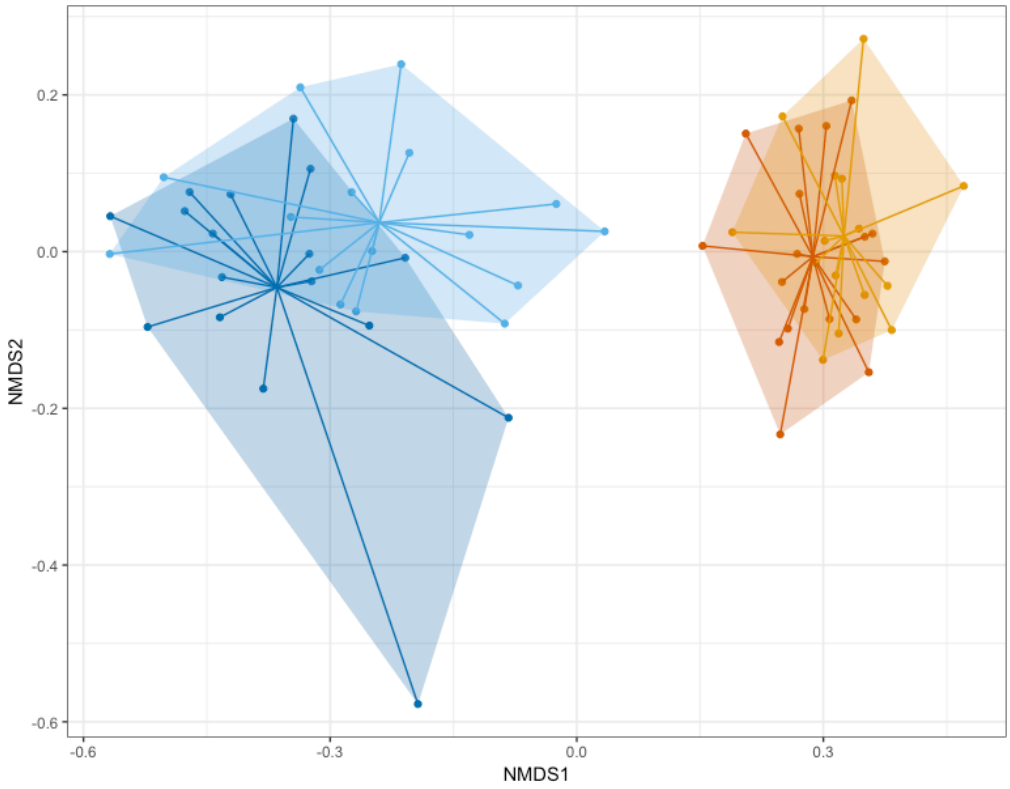


Figure 2.2 Non-metric multi dimensional scaling plot displaying the Bray-Curtis similarity of bacterial OTUs. Rhizosphere samples are indicated in blue coloured circles, soil samples are indicated by orange coloured circles. The level of nitrogen fertilisation is indicated by shade. Rhizosphere high N samples are in dark blue, rhizosphere low N samples are displayed in light blue. Soil high N samples are displayed in dark orange, soil low N samples are displayed in light orange. Samples from each treatment are connected to a group centroid inside coloured segments

The rhizosphere bacterial samples contained a significantly ($P<0.05$) higher relative abundance of OTUs belonging to the Actinobacteria, Bacteroidetes and Proteobacteria phyla, indicated by soil to rhizosphere enrichments of 3.3%, 7.8% and 10.6% respectively (Figure 2.3). Conversely the relative abundance of OTUs belonging to the Acidobacteria, Chloroflexi, Firmicutes, Gemmatimonadetes, Nitrospirae, Planctomycetes and Verrumicrobia phyla were significantly ($P<0.05$) enriched in the soil samples, indicated by enrichments of 11.0%, 1.3%, 1.4%, 1.6%, 0.6%, 2.4%, 2.5%, in the soil samples relative to the rhizosphere samples respectively.

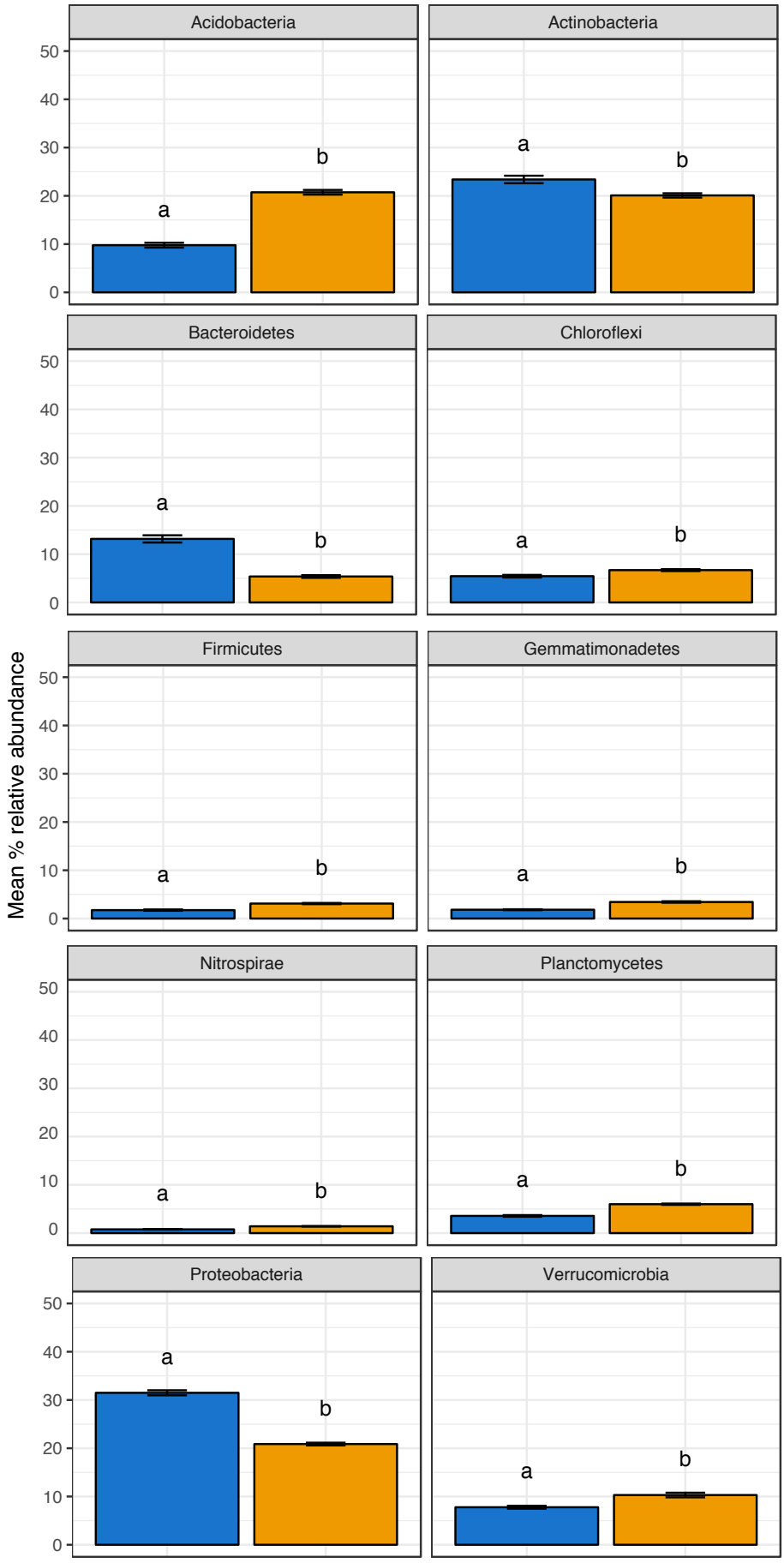
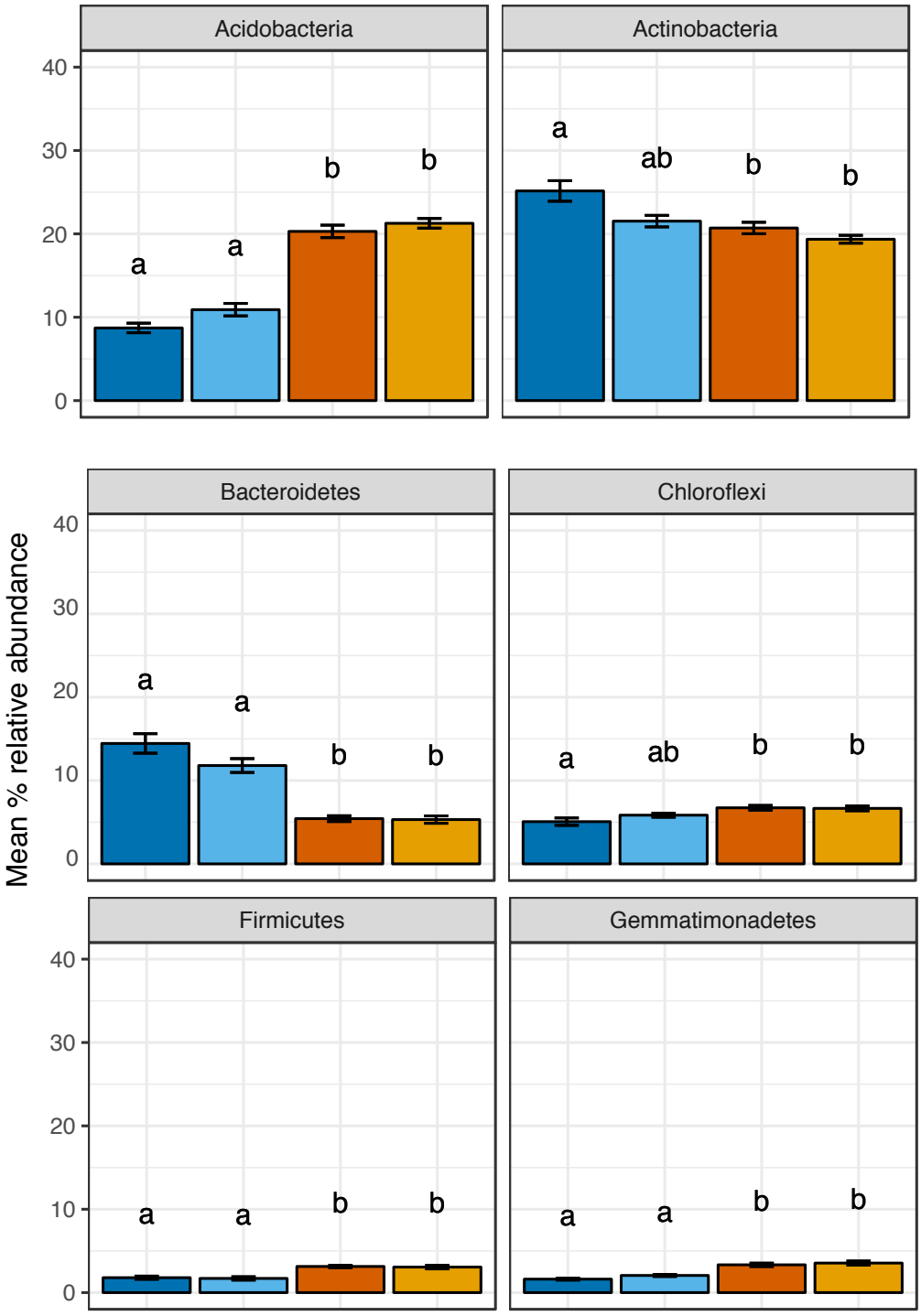


Figure 2.3 Facetted bar plots displaying mean relative abundance of the bacterial phyla in the rhizosphere (blue) and soil (orange) samples. Error bars display standard error of the mean. Groups with a different letter denote significant difference (P<0.05) based on a Kruskal Wallace test

There was no significant difference in relative abundance of bacterial phyla under the influence of N in the rhizosphere or the soil (Figure 2.4). No significant effect of genotype on the composition of bacterial phyla was identified.



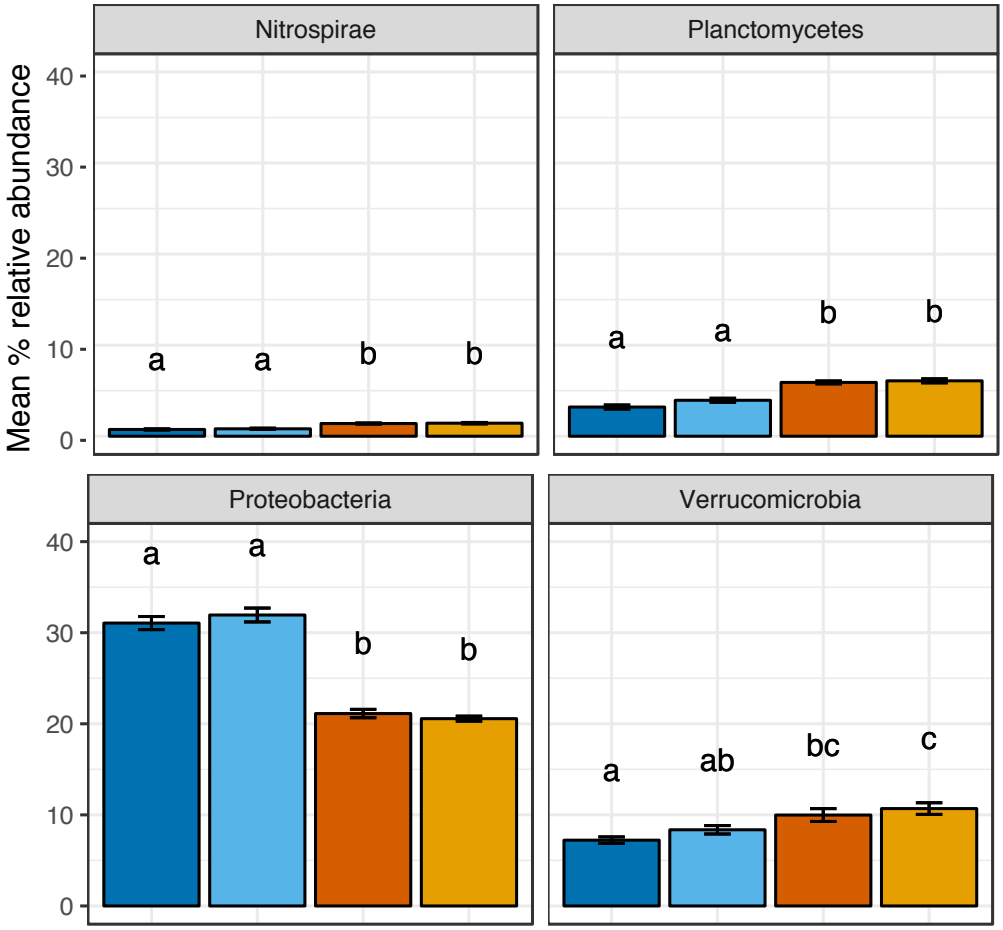


Figure 2.4 Facetted bar plots displaying mean relative abundance of the bacterial phyla. Averages of rhizosphere high N samples displayed in dark blue, rhizosphere low N samples displayed in light blue, soil high N samples displayed in dark orange and soil low N samples displayed in light orange. Error bars display standard error of the mean. Groups with a different letter denote significant difference (P<0.05) based on Dunn's test.

A Similarity percentages test (SIMPER) was performed in order to identify the OTUs that contributed to differences between compartments and treatments which have significant differences indicated by ANOSIM. Each OTU is assigned a percentage value of dissimilarity between the groups. The bacterial OTUs that showed the highest contribution to the dissimilarity (Over 1% of total dissimilarity) between the soil and the rhizosphere were *Flavobacterium* (2 OTUs), *Pedobacter,* Oxalobacteraceae*, Promicromonospora*, and all of these were enriched in the rhizosphere. Conversely, a *DA101* OTU was enriched in the soil (Table 2.4).

Table 2.4 Bacterial OTUs identified by SIMPER analysis that contribute to over 1% of differences between rhizosphere and soil samples

OTU ID	Highest level taxonomic information	Rhizosphere mean relative abundance (%)	Soil mean relative abundance (%)	Contribution to difference (%)
8	*Flavobacterium*	1.86	0.29	1.5
9	*Pedobacter*	1.51	0.03	1.4
4	*DA101*	0.8	2.13	1.3
6	*Oxalobacteraceae*	1.73	0.52	1.17
19	*Promicromonospora*	1.23	0.02	1.14
6770	*Flavobacterium*	1.25	0.06	1.12

Using SIMPER it was revealed that of the OTUs that contribute most to the difference between the N treatments in the rhizosphere, were all enrichments in the high N sample groups (Table 2.5). *Pedobacter*, *Promicromonospora*, Micrococcaceae, Oxalobacteriacea, and two *Flavobacterium* OTUs were enriched in the high N rhizosphere, by 1.47%, 0.89%, 0.95%, 0.40%, 0.34% and 0.45% respectively (Table 2.5).

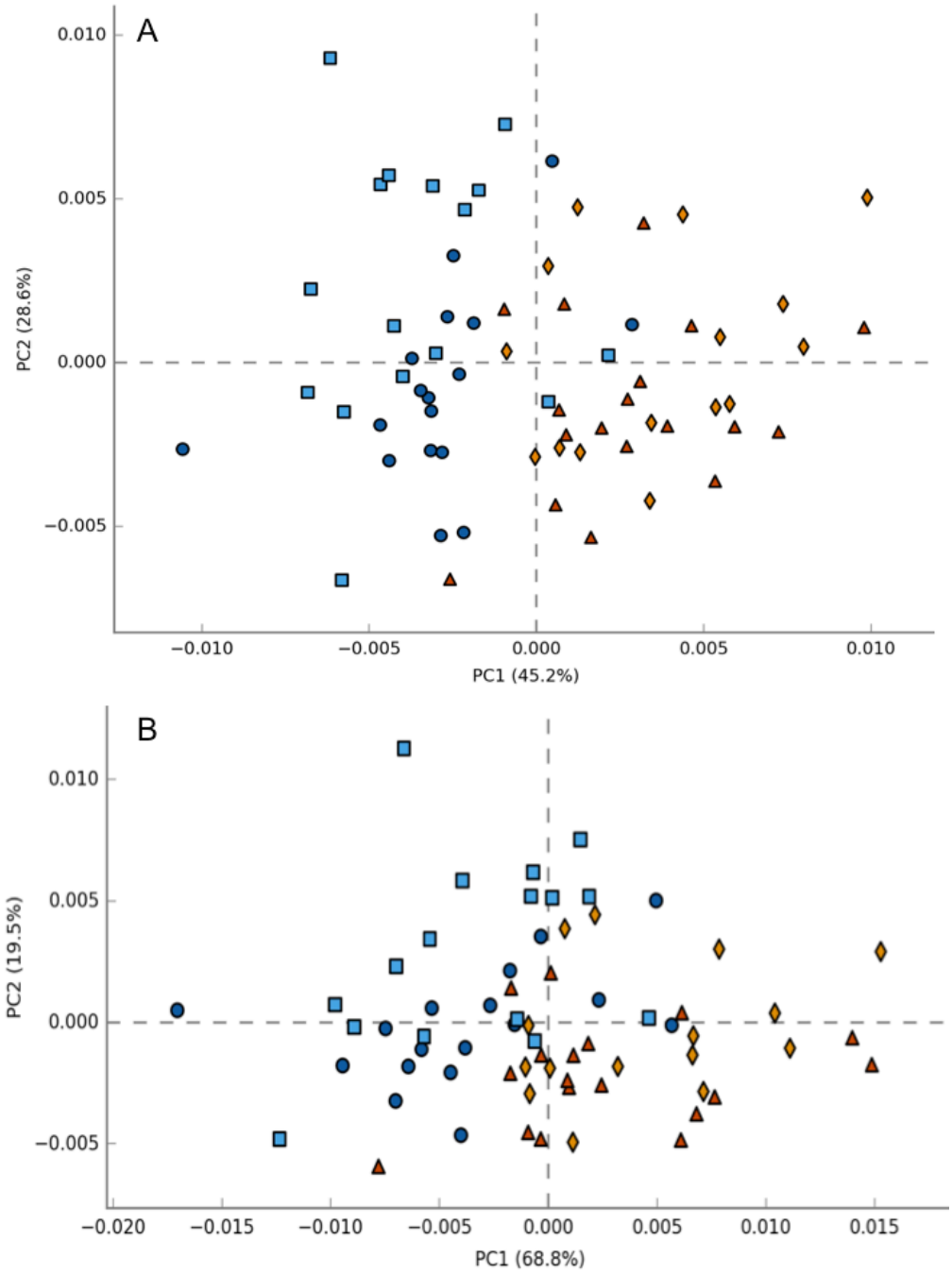
Table 2.5 Bacterial OTUs identified by SIMPER analysis that contribute to the top 10% of differences between the rhizosphere samples under contrasting N treatment, and the corresponding relative abundance of theses OTUs in the soil samples.

OTU ID	Highest level taxonomic information	Rhizosphere high N relative mean abundance (%)	Rhizosphere low N mean relative abundance (%)	Contribution to difference (%)
9	*Pedobacter*	2.22	0.75	2.23
19	*Promicromonospora*	1.66	0.77	1.73
2	*Micrococcaceae*	2.91	1.96	1.29
8	*Flavobacterium*	2.03	1.69	1.11
6770	*Flavobacterium*	1.47	1.02	1.03
6	*Oxalobacteraceae*	1.92	1.52	1.02
12023	*Flavobacterium*	1.15	1.03	0.88
8491	*Pseudomonas*	0.74	0.46	0.64

OTU ID	Highest level taxonomic information	Soil high N relative mean abundance (%)	Soil low N mean relative abundance (%)	Contribution to difference (%)
9	*Pedobacter*	0.04	0.01	0.05
19	*Promicromonospora*	0.02	0.02	0.03
2	*Micrococcaceae*	1.82	1.46	0.87
8	*Flavobacterium*	0.23	0.06	0.44
6770	*Flavobacterium*	0.08	0.04	0.09
6	*Oxalobacteraceae*	0.57	0.47	0.39
12023	*Flavobacterium*	0.31	0.24	0.39
8491	*Pseudomonas*	0.23	0.06	0.27

PICRUSt analysis was conducted in order to predict a metagenome. This allowed detection of possible functional differences between bacterial communities from various compartments. The predicted metagenomes were analysed using the KEGG classification, which is divided into a hierarchical 3-layer structure; Level 1 consists of broad classifications, such as ‘Metabolism’ and levels 2 and 3 have increasingly narrower classifications of metabolic pathways such as ‘Metabolism of Cofactors and Vitamins’ at level 2 and ‘Lipoic acid metabolism’ at level 3. There was no significant effect of plant genotype on the predicted metagenomes at any KEGG level ($P>0.05$). The predicted metagenomes of the soil and

rhizosphere sample groups were significantly (P<0.05) different at all three KEGG levels, however N did not significantly impact the overall predicted metagenome of the rhizosphere or soil (P>0.05) (Figure 2.5).



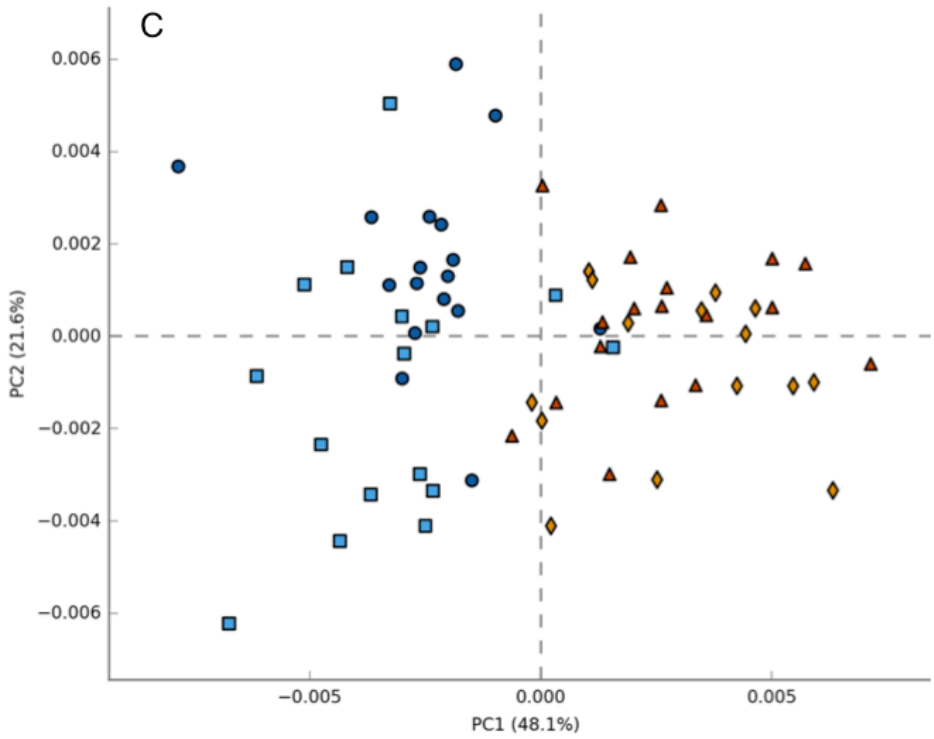


Figure 2.5 KEGG PCA plots displaying differences in PICRUST generated metagenome information at the KEGG level 1 (a) KEGG level 2 (b) and KEGG level 3(c) for PCA 1 and 2. Rhizosphere high N samples are displayed as dark blue circles, rhizosphere low N samples are displayed as light blue squares. Soil high N samples are displayed as dark orange triangles and soil low N samples are displayed as light orange diamonds.

At the KEGG level 1 there were significantly (P<0.05) greater predicted relative abundances of ‘Cellular processes’, ‘Poorly characterized’ and ‘Cellular Processes and Signalling’ in the rhizosphere compared to the soil (Figure 8). Conversely, there were significantly (P<0.05) greater relative abundances of predicted ‘Metabolism’, ‘Genetic Information Processing’ and ‘Environmental Information’ pathways in the soil than the rhizosphere (Figure 2.6).

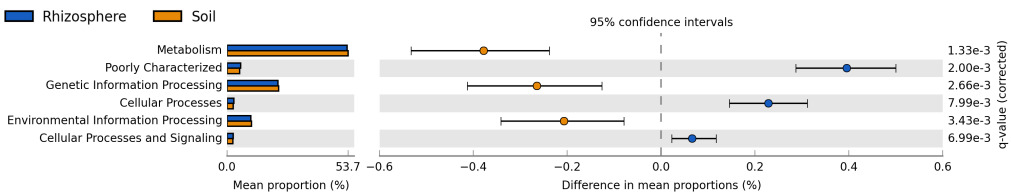


Figure 2.6 Mean relative abundance of the significantly different KEGG level 1 processes in the soil (orange) and rhizosphere (blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

At KEGG level 2, 37 gene groups were predicted to be significantly (P<0.05) different between the predicted soil and rhizosphere metagenomes (Figure 2.7). The largest differences in relative abundance were a greater prevalence in the rhizosphere of ‘Function unknown’, and ‘Cell motility’. Whereas in the soil the largest shifts by percentage relative abundance were ‘Replication and repair’, ‘Membrane transport’ and ‘Carbohydrate metabolism’.

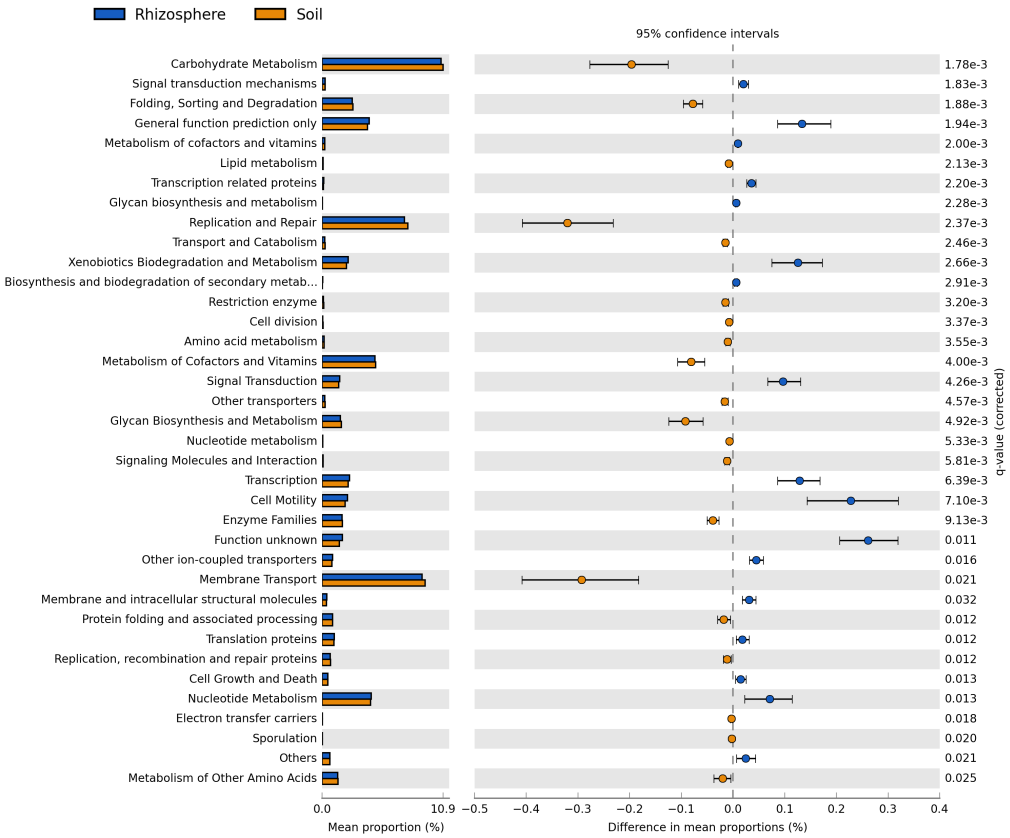
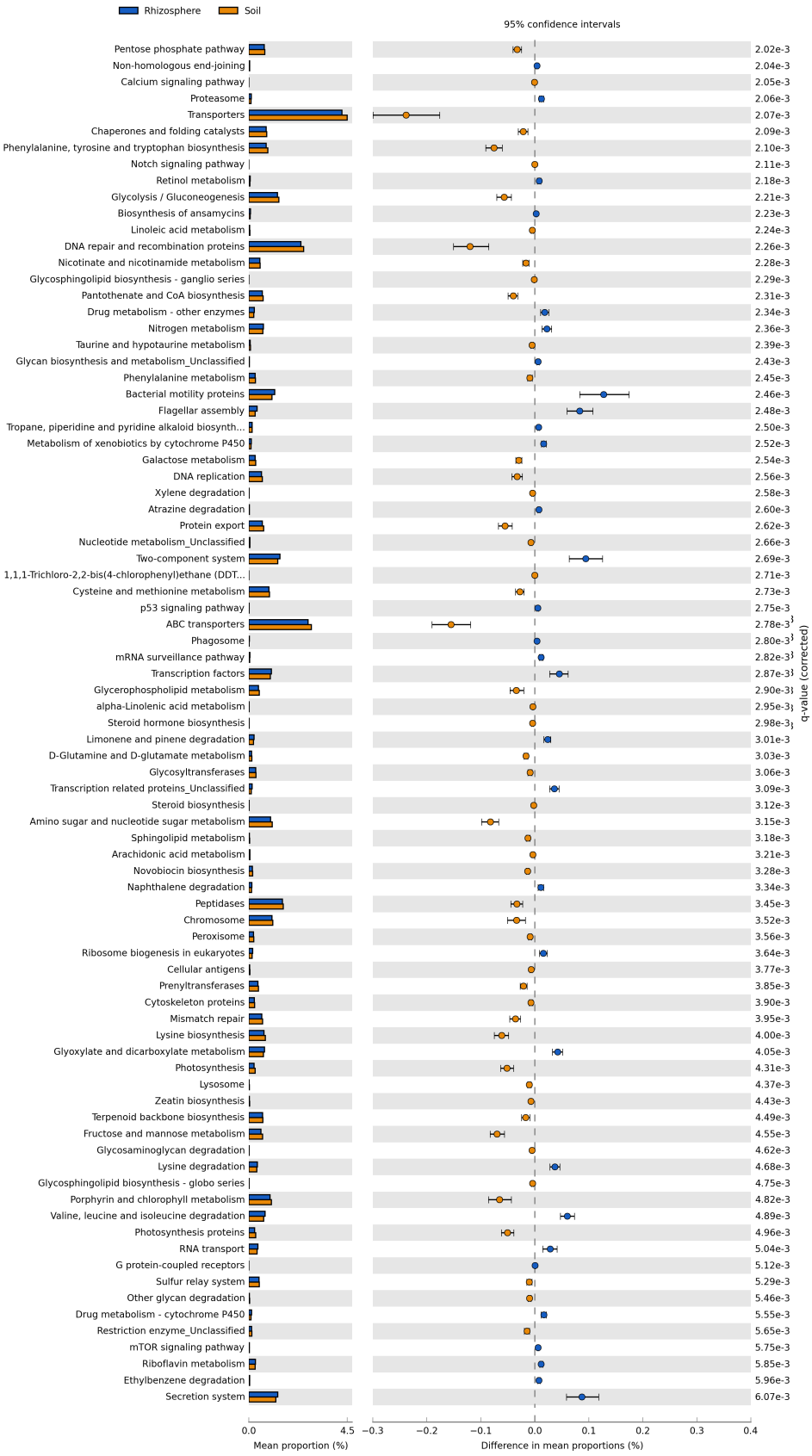


Figure 2.7 Mean relative abundance of the significantly different KEGG level 2 processes in the soil (orange) and rhizosphere (blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

At KEGG Level 3 the relative abundance of 165 predicted processes were significantly (P<0.05) different between rhizosphere and soil (Figure 2.8). The largest shifts by percentage relative abundance were increases in the rhizosphere of ‘Function Unknown’, ‘General function predicted only’ and ‘Bacterial motility proteins’. In the soil the processes that were increased by the largest relative abundance were ‘Transporters’ and ‘ABC transporters’.


Rhizosphere
Soil
95% confidence intervals
Pentose phosphate pathway 2.02e-3
Non-homologous end-joining 2.04e-3
Calcium signaling pathway 2.05e-3
Proteasome 2.06e-3
Transporters 2.07e-3
Chaperones and folding catalysts 2.09e-3
Phenylalanine, tyrosine and tryptophan biosynthesis 2.10e-3
Notch signaling pathway 2.11e-3
Retinol metabolism 2.18e-3
Glycolysis / Gluconeogenesis 2.21e-3
Biosynthesis of ansamycins 2.23e-3
Linoleic acid metabolism 2.24e-3
DNA repair and recombination proteins 2.26e-3
Nicotinate and nicotinamide metabolism 2.28e-3
Glycosphingolipid biosynthesis - ganglio series 2.29e-3
Pantothenate and CoA biosynthesis 2.31e-3
Drug metabolism - other enzymes 2.34e-3
Nitrogen metabolism 2.36e-3
Taurine and hypotaurine metabolism 2.39e-3
Glycan biosynthesis and metabolism_Unclassified 2.43e-3
Phenylalanine metabolism 2.45e-3
Bacterial motility proteins 2.46e-3
Flagellar assembly 2.48e-3
Tropane, piperidine and pyridine alkaloid biosynth... 2.50e-3
Metabolism of xenobiotics by cytochrome P450 2.52e-3
Galactose metabolism 2.54e-3
DNA replication 2.56e-3
Xylene degradation 2.58e-3
Atrazine degradation 2.60e-3
Protein export 2.62e-3
Nucleotide metabolism_Unclassified 2.66e-3
Two-component system 2.69e-3
1,1,1-Trichloro-2,2-bis(4-chlorophenyl)ethane (DDT... 2.71e-3
Cysteine and methionine metabolism 2.73e-3
p53 signaling pathway 2.75e-3
ABC transporters 2.78e-3
Phagosome 2.80e-3
mRNA surveillance pathway 2.82e-3
Transcription factors 2.87e-3
Glycerophospholipid metabolism 2.90e-3
alpha-Linolenic acid metabolism 2.95e-3
Steroid hormone biosynthesis 2.98e-3
Limonene and pinene degradation 3.01e-3
D-Glutamine and D-glutamate metabolism 3.03e-3
Glycosyltransferases 3.06e-3
Transcription related proteins_Unclassified 3.09e-3
Steroid biosynthesis 3.12e-3
Amino sugar and nucleotide sugar metabolism 3.15e-3
Sphingolipid metabolism 3.18e-3
Arachidonic acid metabolism 3.21e-3
Novobiocin biosynthesis 3.28e-3
Naphthalene degradation 3.34e-3
Peptidases 3.45e-3
Chromosome 3.52e-3
Peroxisome 3.56e-3
Ribosome biogenesis in eukaryotes 3.64e-3
Cellular antigens 3.77e-3
Prenyltransferases 3.85e-3
Cytoskeleton proteins 3.90e-3
Mismatch repair 3.95e-3
Lysine biosynthesis 4.00e-3
Glyoxylate and dicarboxylate metabolism 4.05e-3
Photosynthesis 4.31e-3
Lysosome 4.37e-3
Zeatin biosynthesis 4.43e-3
Terpenoid backbone biosynthesis 4.49e-3
Fructose and mannose metabolism 4.55e-3
Glycosaminoglycan degradation 4.62e-3
Lysine degradation 4.68e-3
Glycosphingolipid biosynthesis - globo series 4.75e-3
Porphyrin and chlorophyll metabolism 4.82e-3
Valine, leucine and isoleucine degradation 4.89e-3
Photosynthesis proteins 4.96e-3
RNA transport 5.04e-3
G protein-coupled receptors 5.12e-3
Sulfur relay system 5.29e-3
Other glycan degradation 5.46e-3
Drug metabolism - cytochrome P450 5.55e-3
Restriction enzyme_Unclassified 5.65e-3
mTOR signaling pathway 5.75e-3
Riboflavin metabolism 5.85e-3
Ethylbenzene degradation 5.96e-3
Secretion system 6.07e-3
q-value (corrected)
0.0
4.5
Mean proportion (%)
−0.3
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4
Difference in mean proportions (%)

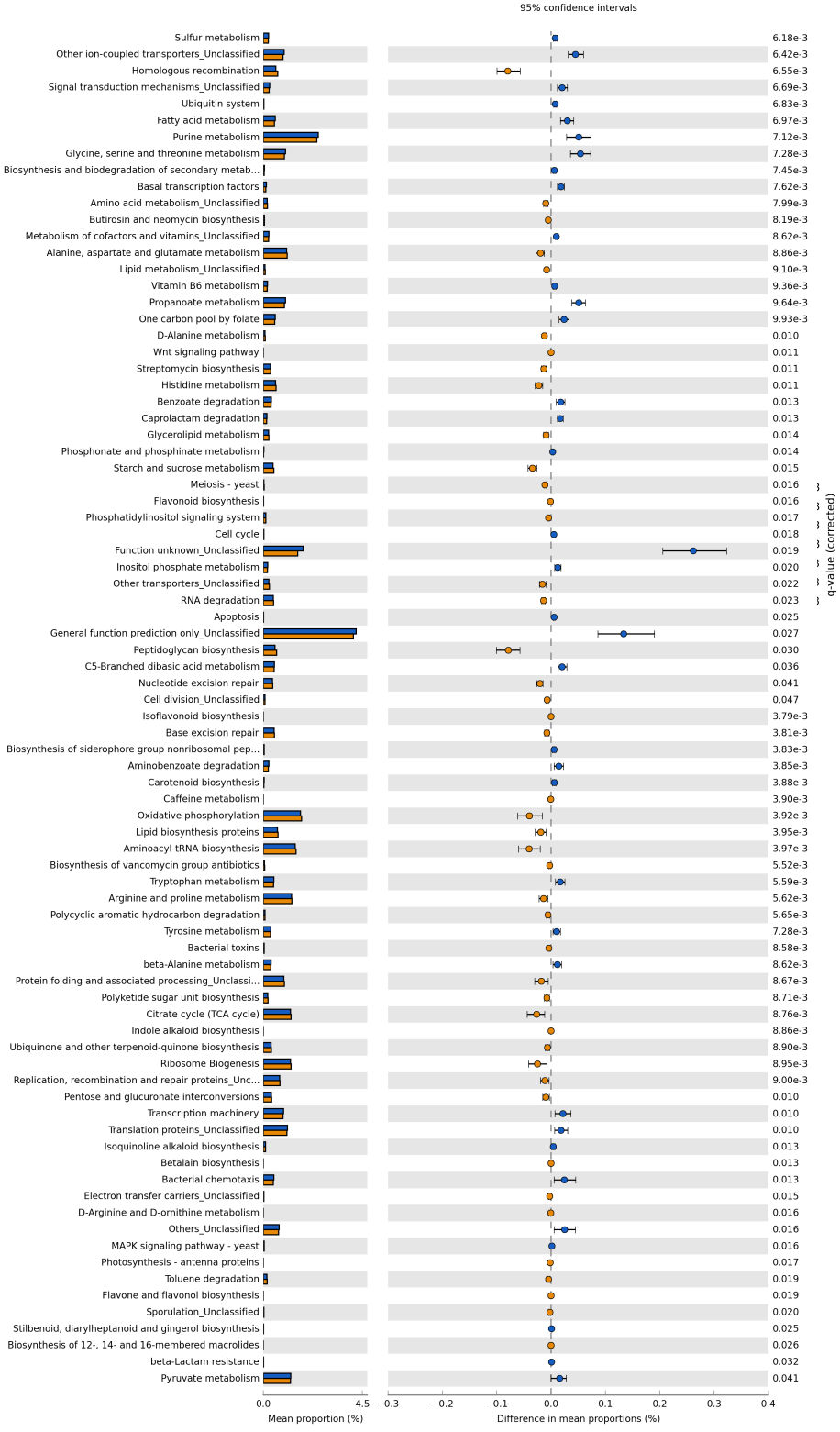
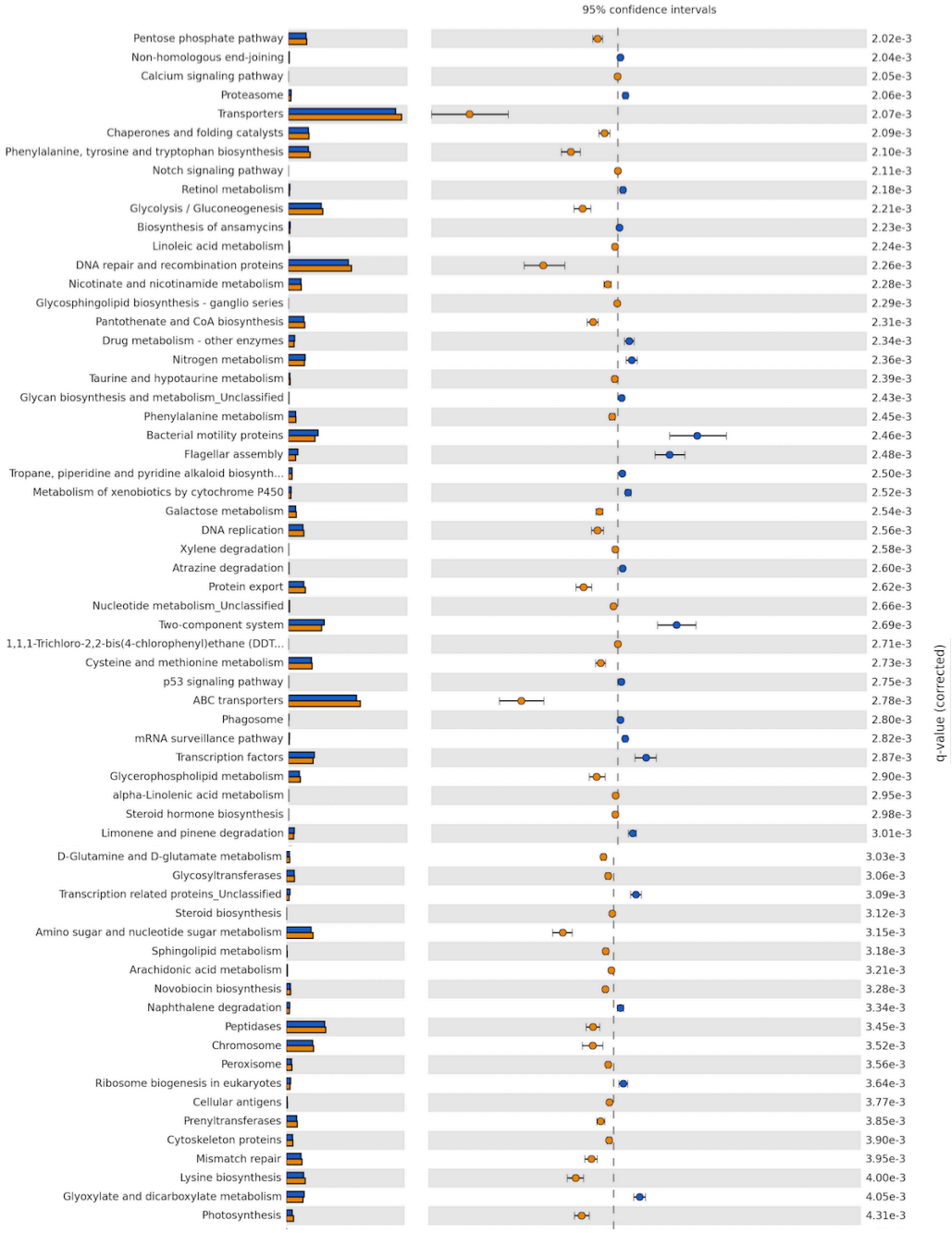
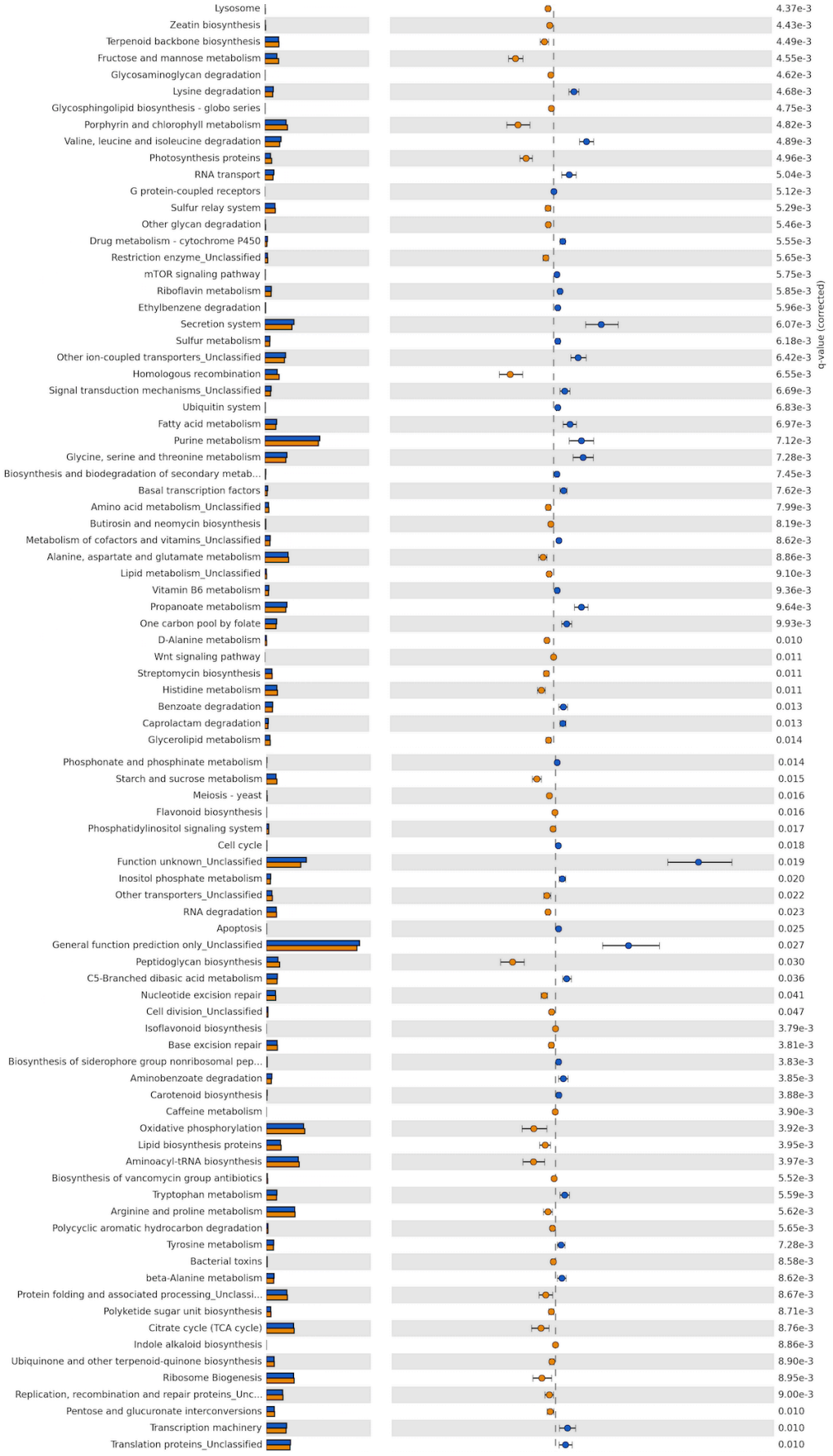


Figure 2.8 Mean relative abundance of the significantly different KEGG level 3 processes in the soil (orange) and rhizosphere (blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

At KEGG Level 3 the relative abundance of 165 predicted processes were significantly (P<0.05) different between rhizosphere and soil (Figure 2.9). The largest shifts by percentage relative abundance were increases in the rhizosphere of ‘Function Unknown’, ‘General function predicted only’ and ‘Bacterial motility proteins’. In the soil the processes that were increased by the largest relative abundance were ‘Transporters’ and ‘ABC transporters’.




Lysosome 4.37e-3
Zeatin biosynthesis 4.43e-3
Terpenoid backbone biosynthesis 4.49e-3
Fructose and mannose metabolism 4.55e-3
Glycosaminoglycan degradation 4.62e-3
Lysine degradation 4.68e-3
Glycosphingolipid biosynthesis - globo series 4.75e-3
Porphyrin and chlorophyll metabolism 4.82e-3
Valine, leucine and isoleucine degradation 4.89e-3
Photosynthesis proteins 4.96e-3
RNA transport 5.04e-3
G protein-coupled receptors 5.12e-3
Sulfur relay system 5.29e-3
Other glycan degradation 5.46e-3
Drug metabolism - cytochrome P450 5.55e-3
Restriction enzyme_Unclassified 5.65e-3
mTOR signaling pathway 5.75e-3
Riboflavin metabolism 5.85e-3
Ethylbenzene degradation 5.96e-3
Secretion system 6.07e-3
Sulfur metabolism 6.18e-3
Other ion-coupled transporters_Unclassified 6.42e-3
Homologous recombination 6.55e-3
Signal transduction mechanisms_Unclassified 6.69e-3
Ubiquitin system 6.83e-3
Fatty acid metabolism 6.97e-3
Purine metabolism 7.12e-3
Glycine, serine and threonine metabolism 7.28e-3
Biosynthesis and biodegradation of secondary metab... 7.45e-3
Basal transcription factors 7.62e-3
Amino acid metabolism_Unclassified 7.99e-3
Butirosin and neomycin biosynthesis 8.19e-3
Metabolism of cofactors and vitamins_Unclassified 8.62e-3
Alanine, aspartate and glutamate metabolism 8.86e-3
Lipid metabolism_Unclassified 9.10e-3
Vitamin B6 metabolism 9.36e-3
Propanoate metabolism 9.64e-3
One carbon pool by folate 9.93e-3
D-Alanine metabolism 0.010
Wnt signaling pathway 0.011
Streptomycin biosynthesis 0.011
Histidine metabolism 0.011
Benzoate degradation 0.013
Caprolactam degradation 0.013
Glycerolipid metabolism 0.014
Phosphonate and phosphinate metabolism 0.014
Starch and sucrose metabolism 0.015
Meiosis - yeast 0.016
Flavonoid biosynthesis 0.016
Phosphatidylinositol signaling system 0.017
Cell cycle 0.018
Function unknown_Unclassified 0.019
Inositol phosphate metabolism 0.020
Other transporters_Unclassified 0.022
RNA degradation 0.023
Apoptosis 0.025
General function prediction only_Unclassified 0.027
Peptidoglycan biosynthesis 0.030
C5-Branched dibasic acid metabolism 0.036
Nucleotide excision repair 0.041
Cell division_Unclassified 0.047
Isoflavonoid biosynthesis 3.79e-3
Base excision repair 3.81e-3
Biosynthesis of siderophore group nonribosomal pep... 3.83e-3
Aminobenzoate degradation 3.85e-3
Carotenoid biosynthesis 3.88e-3
Caffeine metabolism 3.90e-3
Oxidative phosphorylation 3.92e-3
Lipid biosynthesis proteins 3.95e-3
Aminoacyl-tRNA biosynthesis 3.97e-3
Biosynthesis of vancomycin group antibiotics 5.52e-3
Tryptophan metabolism 5.59e-3
Arginine and proline metabolism 5.62e-3
Polycyclic aromatic hydrocarbon degradation 5.65e-3
Tyrosine metabolism 7.28e-3
Bacterial toxins 8.58e-3
beta-Alanine metabolism 8.62e-3
Protein folding and associated processing_Unclassi... 8.67e-3
Polyketide sugar unit biosynthesis 8.71e-3
Citrate cycle (TCA cycle) 8.76e-3
Indole alkaloid biosynthesis 8.86e-3
Ubiquinone and other terpenoid-quinone biosynthesis 8.90e-3
Ribosome Biogenesis 8.95e-3
Replication, recombination and repair proteins_Unc... 9.00e-3
Pentose and glucuronate interconversions 0.010
Transcription machinery 0.010
Translation proteins_Unclassified 0.010
q-value (corrected)

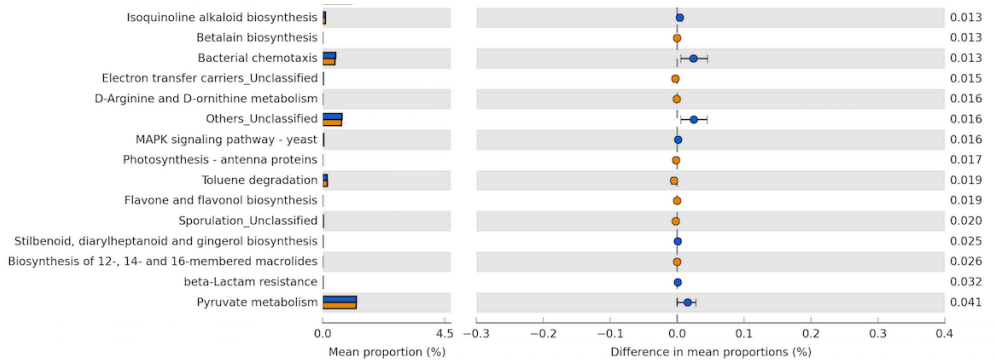


Figure 2.9 Mean relative abundance of the significantly different KEGG level 3 processes in the soil (orange) and rhizosphere (blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

N treatment did not significantly alter any of the KEGG level 1 processes in the soil, but in the rhizosphere there were significant (P<0.05) increases under high N of ‘Environmental Information Processing’ and an increase in the low N of ‘Poorly Characterised processes’ (Figure 2.10).

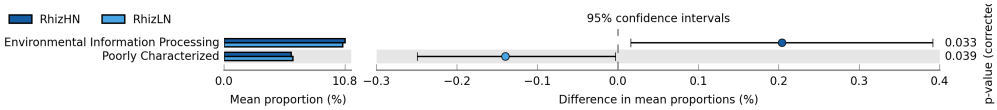


Figure 2.10 Mean relative abundance of the significantly different KEGG level 1 processes in the rhizosphere high N (dark blue) and low N (light blue), Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

At KEGG level 2, N treatment did not significantly alter any processes in the soil, however in the rhizosphere increases of relative abundance under high N in ‘Cell growth and death’, ‘Membrane and intracellular structural molecules’, ‘Glycan biosynthesis and metabolism’, ‘Metabolism of other amino acids’ and ‘Lipid metabolism’ were identified (Figure 2.11). In the predicted metagenomes the low N rhizosphere, there were significant increases of ‘General function and prediction only’, ’Carbohydrate metabolism’, ‘Signalling molecules and interaction’, ‘Nucleotide metabolism’, ‘Replication, recombination and repair proteins’, ‘Sporulation’ and ‘Enzyme families’.

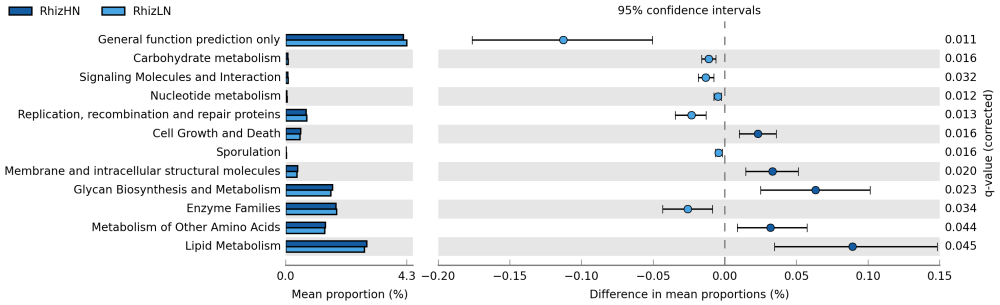


Figure 2.11 Mean relative abundance of the significantly different KEGG level 2 processes in the rhizosphere high N (dark blue) and low N (light blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

At KEGG level 3, N addition significantly (P<0.05) increased the relative abundance in the rhizosphere of ‘Glycine, serine and threonine metabolism’, ‘Oxidative phosphorylation’ and ‘Bacterial secretion system’ (Figure 2.12). The processes predicted to be at higher relative abundance in the low N rhizosphere were ‘‘Methane metabolism’, ‘General function prediction only’ and ‘Amino sugar and nucleotide sugar metabolism’.

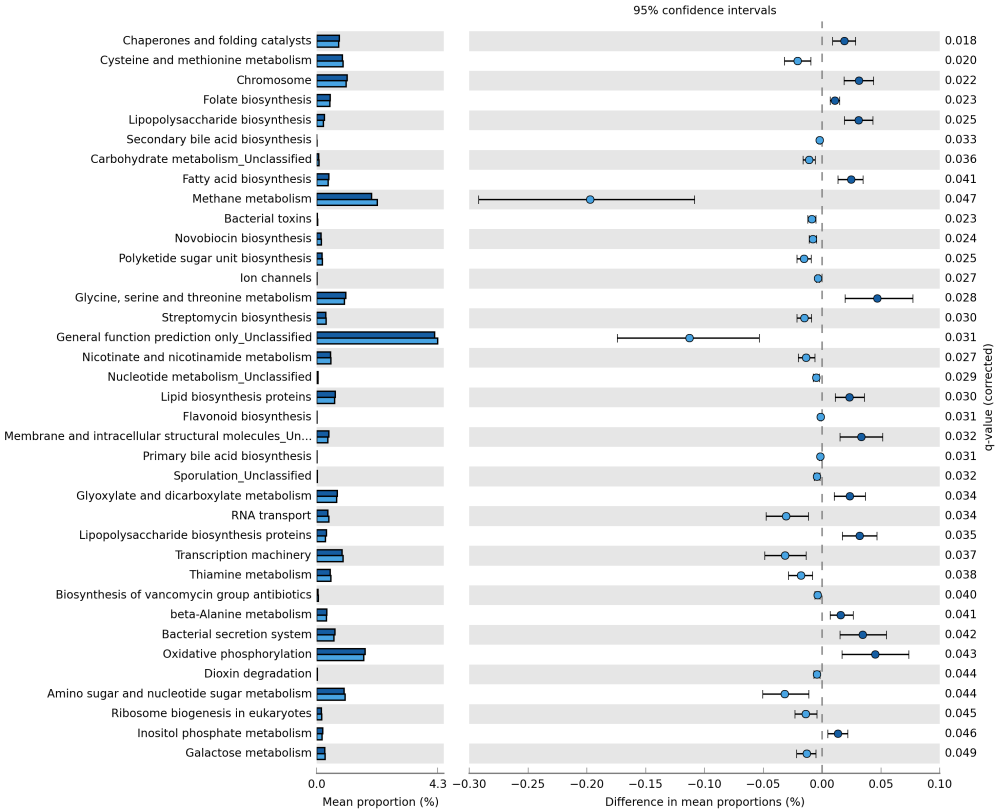


Figure 2.12 Mean relative abundance of the significantly different KEGG level 3 processes in the rhizosphere high N (dark blue) and low N (light blue). Shown with the difference in mean proportions of the processes and the corrected P value calculated by Whites non-parametric t-test.

In the rhizosphere the predicted metagenome had a significant (P<0.001) increase in the OTUs containing *ureC* gene by 0.4% relative to soil. In the soil the predicted metagenome had a significant (P<0.001) increase of OTUs containing *nirK* and *nirA* genes, relative to rhizosphere, both by 0.4% (Table 2.6).

Table 2.6 The mean relative abundance of OTUs predicted to contain specific nitrogen cycle genes in the rhizosphere and soil. Also shown is significance of the difference between rhizosphere and soil samples.

KEGG ID	Gene	Rhizosphere mean % relative abundance of OTUs	Soil mean % relative abundance of OTUs	Significance (Wilcoxon P value)
K04561	*norB*	1.44	1.59	0.1
K01428	*ureC*	1.73	1.30	<0.001
K02588	*nifH*	1.49	1.54	0.3
K02586	*nifD*	1.53	1.50	0.6
K00376	*nosZ*	1.54	1.49	0.4
K00372	*nasA*	1.48	1.55	0.3
K00368	*nirK*	1.30	1.73	<0.001
K00366	*nirA*	1.32	1.71	<0.001

Furthermore, the predicted metagenome of the low N treatment had a significant (P<0.05) increase of *nifH* relative to the high N treatment, the opposite of the effect for *ureC* (P<0.05) (Table 2.7).

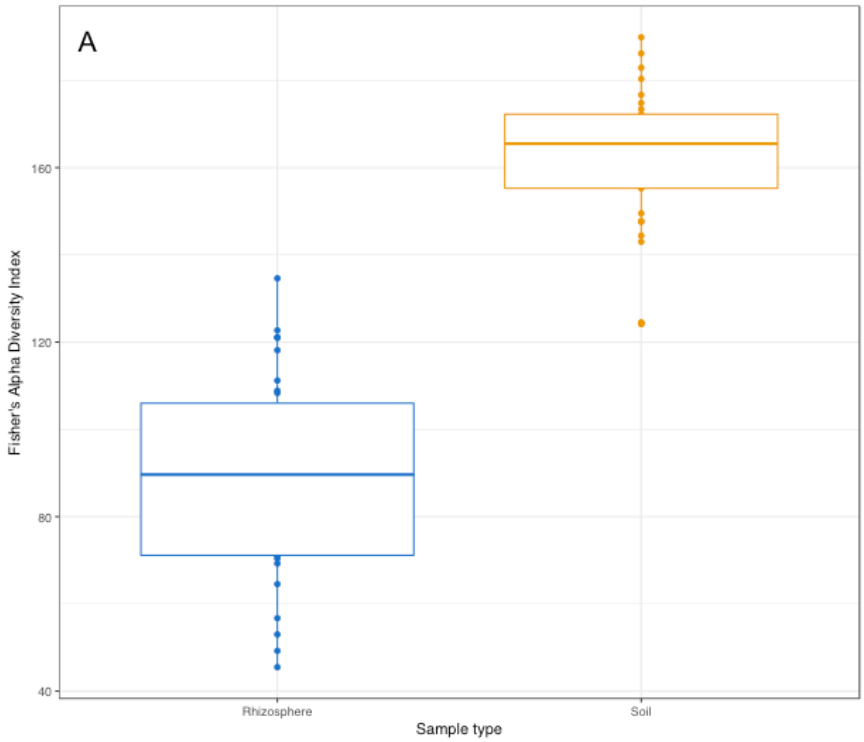
Table 2.7 The mean relative abundance of OTUs predicted to contain specific N cycle genes in the rhizosphere and soil under nitrogen treatment

KEGG ID	Gene	Rhizosphere high N mean % relative abundance of OTUs	Rhizosphere low N mean % relative abundance of OTUs	P value
K04561	*norB*	1.44	1.45	1.00
K01428	*ureC*	1.82	1.63	0.22
K02588	*nifH*	1.35	1.63	**0.03**
K02586	*nifD*	1.54	1.53	1.00
K00376	*nosZ*	1.56	1.53	0.99
K00372	*nasA*	1.44	1.53	1.00
K00368	*nirK*	1.35	1.24	0.73
K00366	*nirA*	1.29	1.34	0.57

KEGG ID	Gene	Soil high N mean % relative abundance of OTUs	Soil low N mean % relative abundance of OTUs	P value
K04561	*norB*	1.61	1.57	0.88
K01428	*ureC*	1.43	1.15	**0.04**
K02588	*nifH*	1.55	1.54	0.73
K02586	*nifD*	1.54	1.44	1.00
K00376	*nosZ*	1.60	1.36	0.83
K00372	*nasA*	1.58	1.51	0.82
K00368	*nirK*	1.79	1.67	0.83
K00366	*nirA*	1.78	1.64	0.67

2.3.2 Protist community composition

Protist Fisher's alpha diversity was significantly (P<0.001) greater in the soil than in the rhizosphere (Figure 2.13A), however nitrogen had no significant effect on protist alpha diversity in the rhizosphere (P=0.148) or the soil (P=0.451) (Figure 2.13B). Plant genotype had no significant impact on rhizosphere protist alpha diversity (Figure 2.13C).



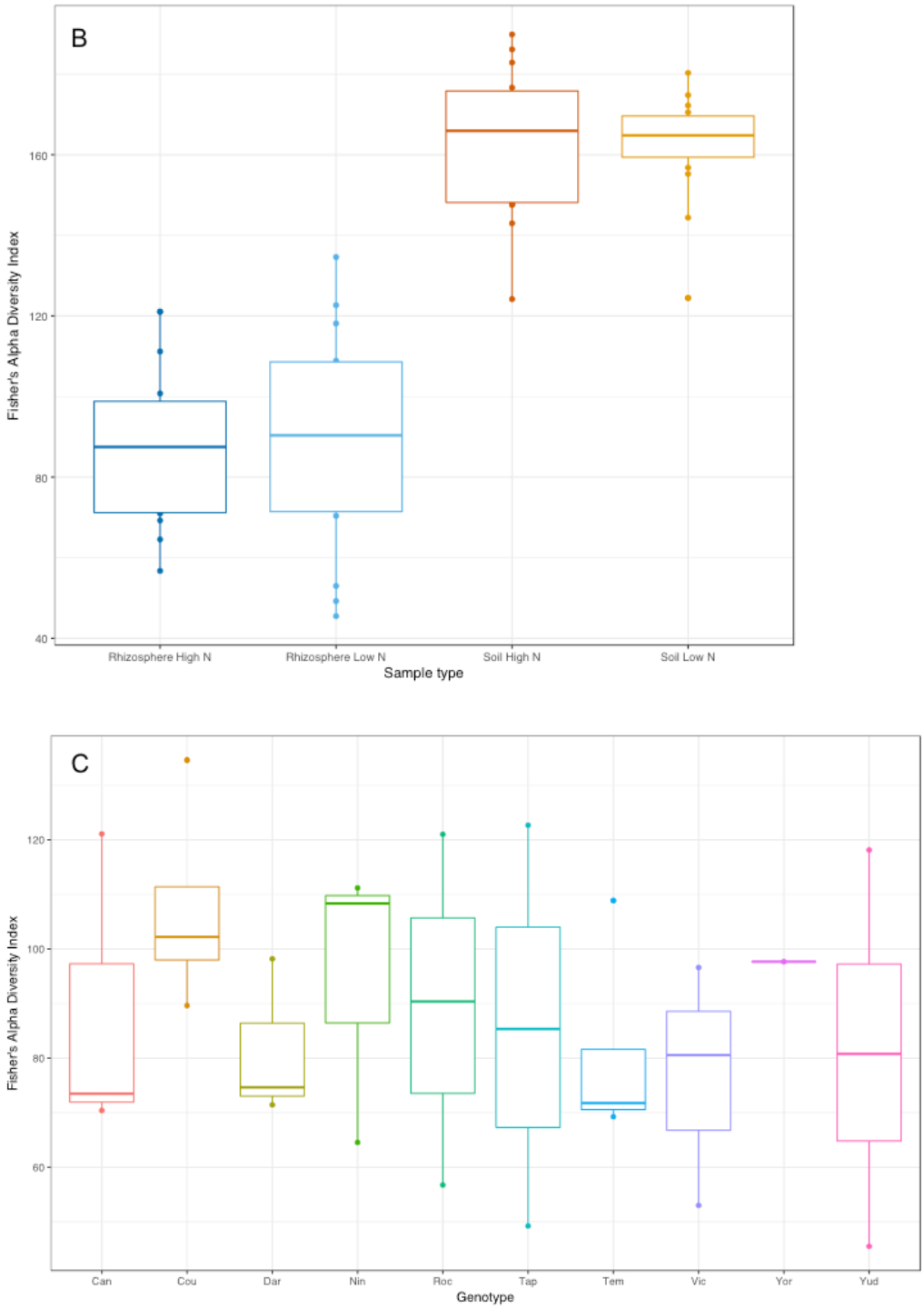


Figure 2.13 Box and whisker plot showing Fisher's alpha diversity of the protist communities in the A.) Rhizosphere (blue) and soil (orange) samples. B.) Rhizosphere high N (dark blue), rhizosphere low N (light blue), soil High N (dark orange) and soil low N(light orange) samples. C.) The rhizosphere for each of the genotypes. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range.

Differences in protist community composition between compartment and N treatment were visualised using MDS (Figure 2.14). ANOSIM showed that the protist communities of the soil and rhizosphere were significantly different from each other (R= 0.93), (P <0.001). The soil protist community composition was significantly different (R=0.146) (P<0.05) between the soil samples with and without nitrogen treatment, but N fertilisation did not affect the rhizosphere protist community (P=0.37). Plant genotype had no significant effect on the protist communities (P=0.06)

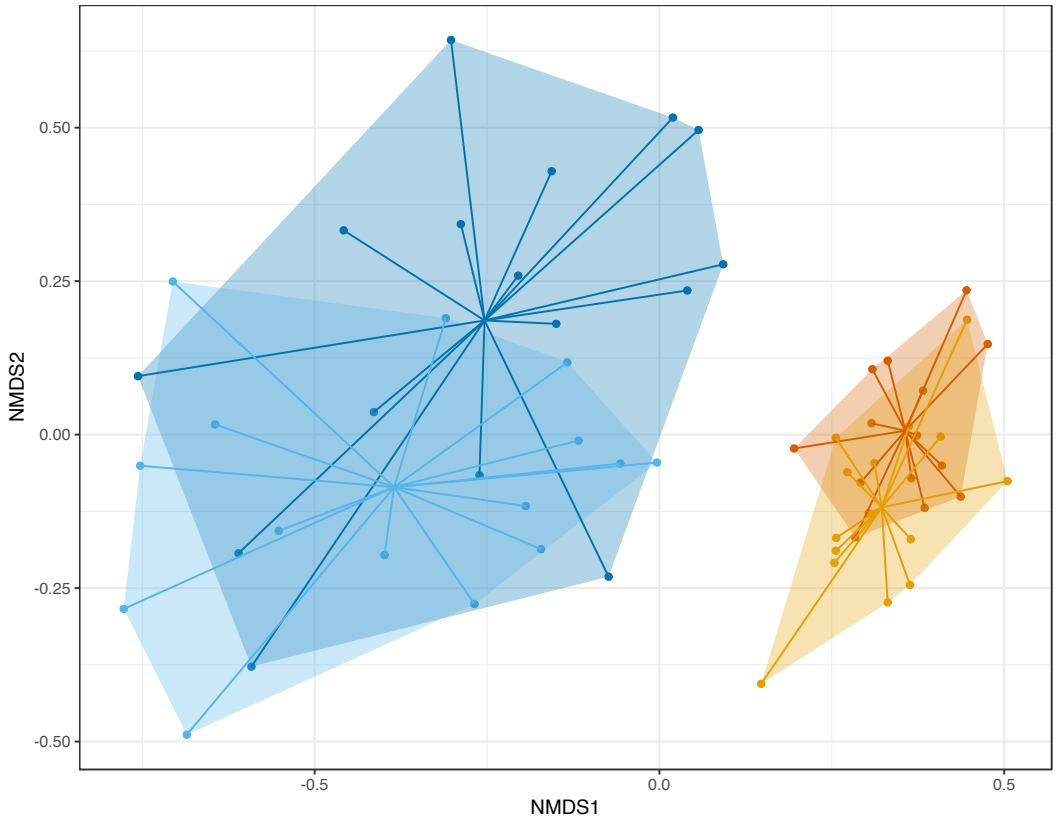


Figure 2.14 Non-metric multi-dimensional scaling plot displaying the Bray-Curtis similarity of protist OTUs. Rhizosphere samples are shown in blue, soil samples are shown in orange. The level of nitrogen fertilisation is indicated by shade. Rhizosphere high N samples are in dark blue, rhizosphere low N samples are displayed in light blue. Soil high N samples are displayed in dark orange, soil low N samples are displayed in light orange. Samples from each treatment are connected to a group centroid inside coloured segments

There was a significant (P<0.05) enrichment of 18.85% for the SAR (Stramenopiles, Alveolata and Rhizaria) in the rhizosphere of OSR (Figure 2.15). Conversely the Amoebozoa, Archaeplastida and Excavata were significantly (P<0.05) enriched in the soil compared to the rhizosphere, with mean increases in relative abundance of 6.02%, 11.13% and 0.7% respectively.

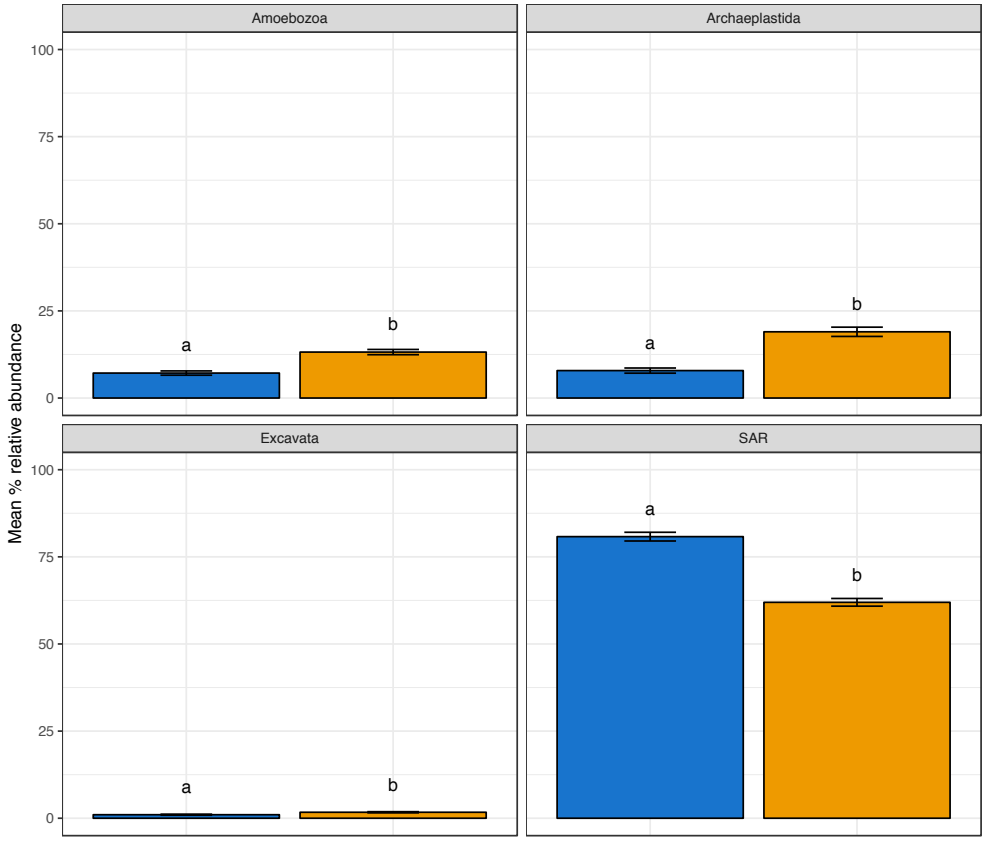


Figure 2.15 Facetted bar plots displaying mean relative abundance of abundant (greater that 1%) protist high level taxonomic groups in the rhizosphere (blue) and soil (orange) samples. Error bars display standard error of the mean. Letters denote significant difference (P<0.05) based on a Kruskal Wallace test

Nitrogen treatment had no significant effect on the relative abundance of any of the phyla in either the soil or rhizosphere (Figure 2.16).

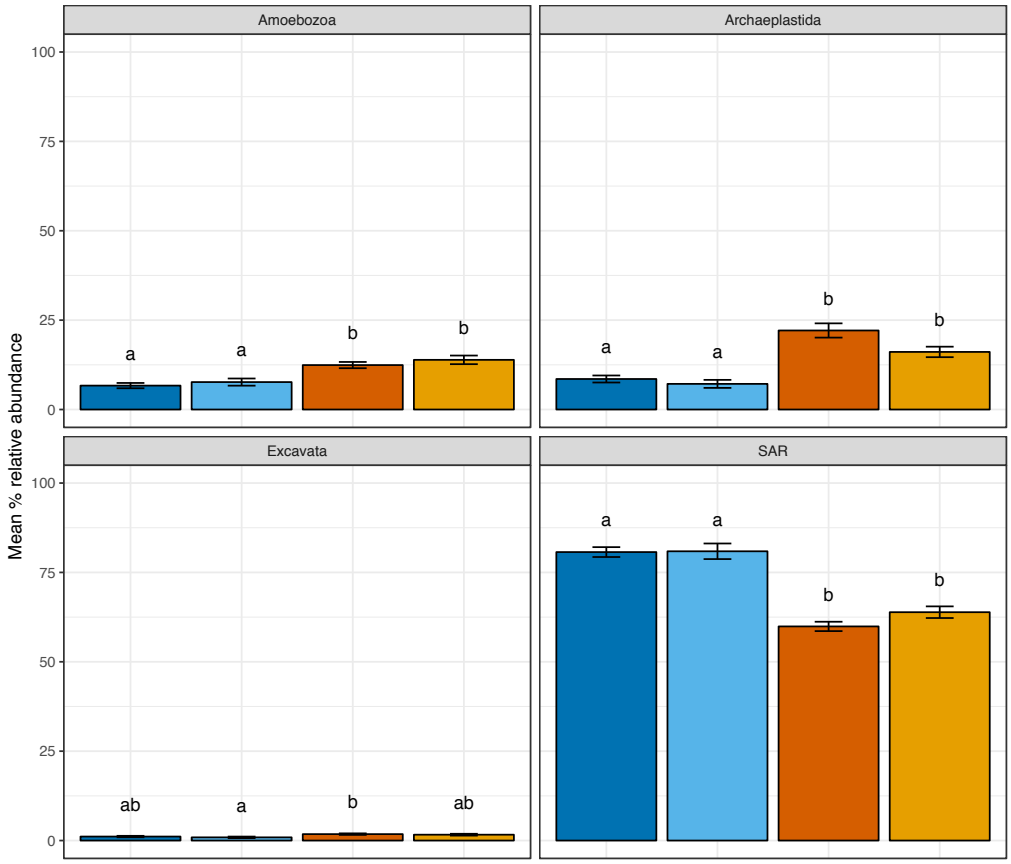


Figure 2.16 Facetted bar plots displaying mean relative abundance of the protist phyla in the rhizosphere high N samples displayed in dark blue, rhizosphere low N samples displayed in light blue, soil high N samples displayed in dark orange and soil low N samples displayed in light orange. Error bars display standard error of the mean. Letters denote significant difference (P<0.05) based on a Kruskal Wallace test, where groups that share a letter have no significant difference.

The protist OTUs which contributed to at least 2% of the difference between the rhizosphere and soil, as identified by a SIMPER, were *Spongospora subterranea f. sp. subterranea* and *Phytophthora infestans T30-4* which were enriched in the rhizosphere samples by 23.7% and 3.9% respectively (Table 2.8), and conversely, an *Ophiocytum* OTU was which had a greater relative abundance in the soil samples by 4.3%.

Table 2.8 Protist OTUs identified by SIMPER analysis, which contribute to at least 2% of differences between rhizosphere and soil samples.

OTU ID	Highest level taxonomic information	Rhizosphere mean relative abundance (%)	Soil mean relative abundance (%)	Contribution to difference (%)
28	*Spongospora subterranean f. sp. subterranea*	24	0.27	15.03
37	*Ophiocytum*	2.07	6.36	2.93
255	*Phytophthora infestans T30-4*	3.98	0.1	2.48

N treatment did not significantly shift the rhizosphere protist community composition. In the soil the OTUs identified by SIMPER as contributing to at least 2% of difference under N treatment were an *Ophiocytum* OTU and Pyrenosporomycetes OTU which had higher relative abundance in the low N samples by 0.8% and 4% respectively (Table 2.9).

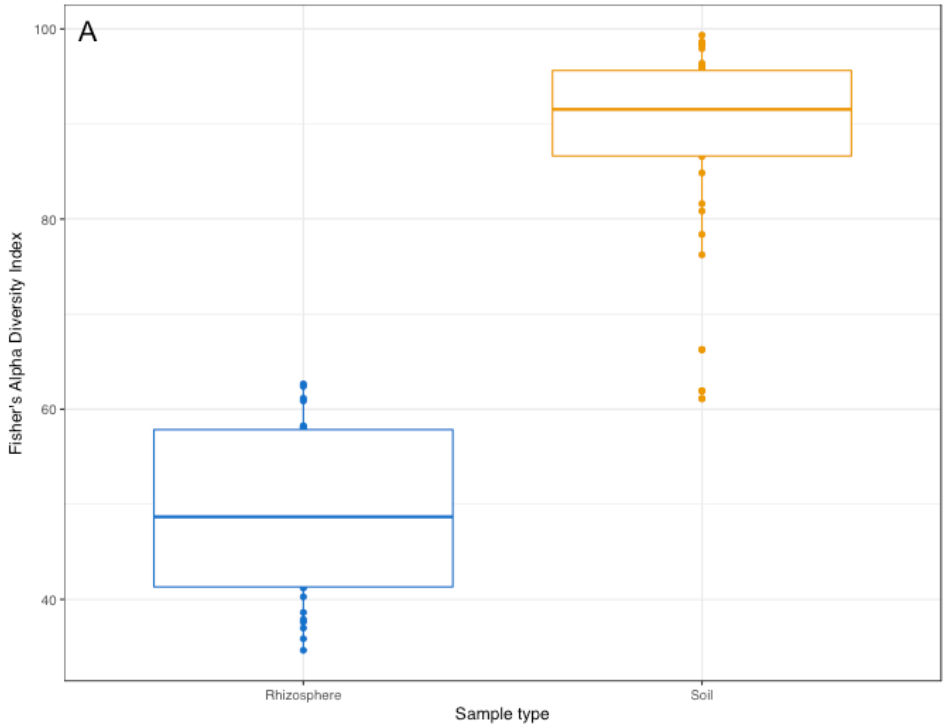
Table 2.9 Table of protist OTUs identified by SIMPER analysis that contribute to at least 2% of differences between the soil samples under contrasting N treatment, with the corresponding relative abundance of the rhizosphere samples shown for reference.

OTU ID	Highest level taxonomic information	Soil high N % relative abundance	Soil low N mean % relative abundance	Contribution to difference (%)
37	*Ophiocytium*	5.96	6.74	3.41
117	Peronosporomycetes	0.08	4.06	3.26

OTU ID	Highest level taxonomic information	Rhizo-sphere high N % relative abundance	Rhizo-sphere low N mean % relative abundance	Contribution to difference (%)
37	*Ophiocytium*	5.96	6.74	3.41
117	Peronosporomycetes	0.08	4.06	3.26

2.3.3 Fungal community composition and function

Fungal Fisher's alpha diversity was significantly (P<0.01) greater in soil than in the rhizosphere (Figure 2.17A), however N had no significant effect on fungal alpha diversity in the rhizosphere (P=0.44) or the soil (P=0.74) (Figure 2.17B). Plant Genotype had no significant impact on rhizosphere fungal alpha diversity (P=0.6) (Figure 2.17C).



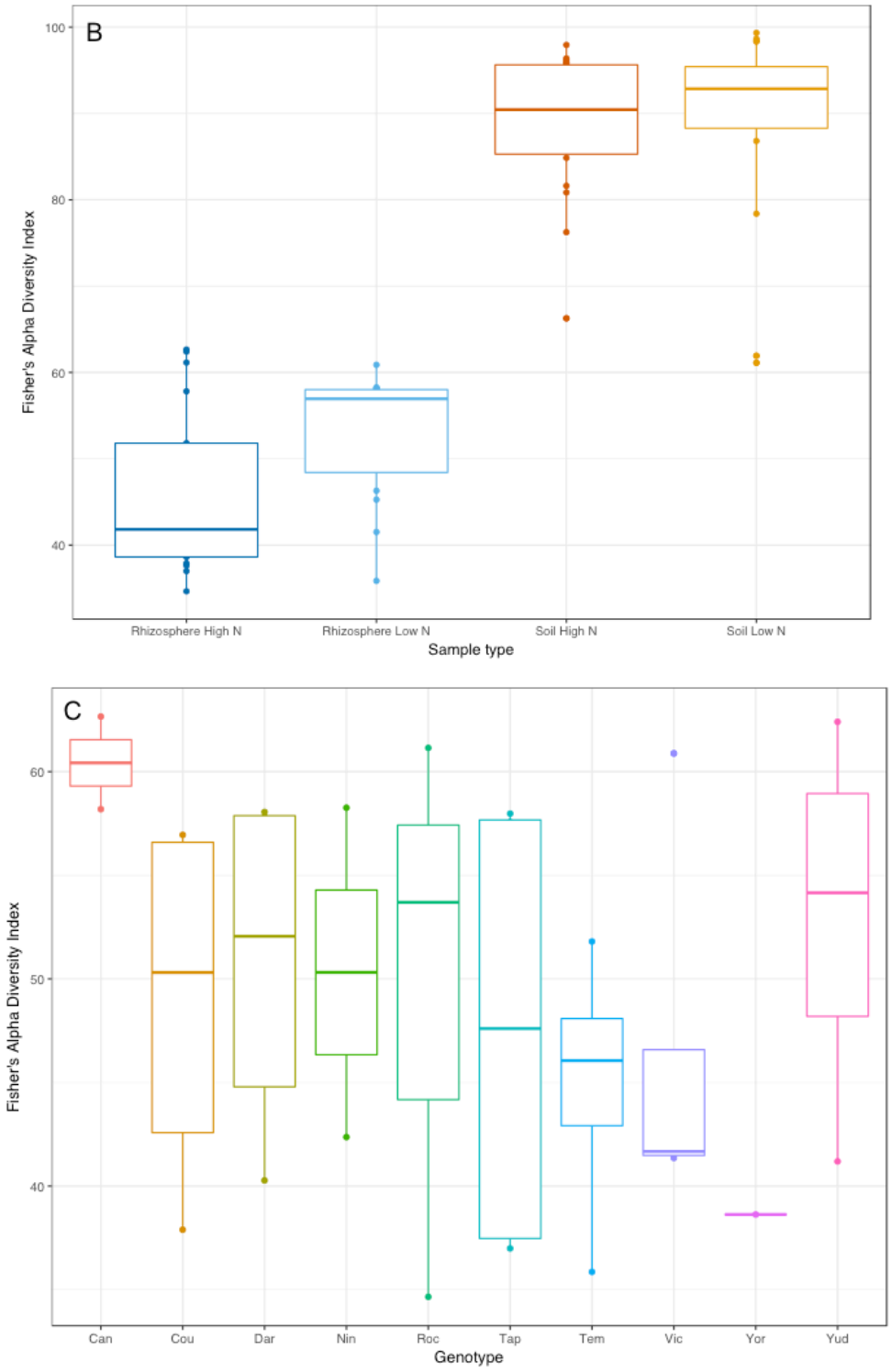


Figure 2.17 Box and whisker plot showing Fisher's alpha diversity of the fungal communities in A.) The rhizosphere (blue) and soil (orange) samples. B.) Rhizosphere high N (dark blue) , rhizosphere low N (light blue), soil High N (dark orange) and soil low N(light orange) samples. C.) The rhizosphere for each of the genotypes. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range.

The differences in fungal community composition were visualised using an NMDS plot (Figure 23). ANOSIM revealed that fungal community composition of the soil and rhizosphere were significantly different from each other (R= 0.5), (P<0.001). Furthermore, N fertilisation significantly altered both the soil (R=0.11) (P<0.026) and rhizosphere (R= 0.11) (P<0.010) fungal community composition. Genotype had no significant effect on fungal community composition (P=0.09) (Figure 2.18).

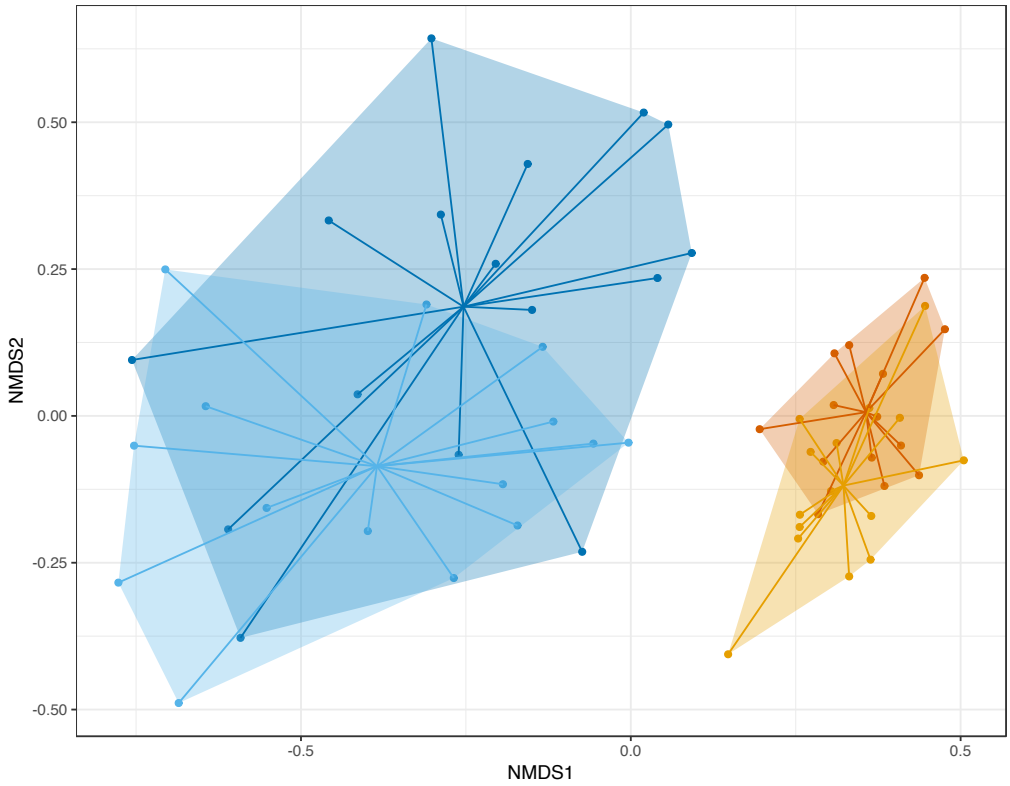


Figure 2.18 Multi-dimensional scaling plot displaying the Bray-Curtis similarity of bacterial OTUs. Rhizosphere samples represented by blue circles, soil samples are represented as orange circles. The level of nitrogen fertilisation is indicated by shade. Rhizosphere high N samples are in dark blue, rhizosphere low N samples are displayed in light blue. Soil high N samples are displayed in dark orange, soil low N samples are displayed in light orange. Samples from each treatment are connected to a group centroid inside coloured segments

The relative abundances of the Ascomycota and Chytridiomycota in the rhizosphere were significantly (P<0.05) greater than in the soil by 8.4% and 2.8% respectively, and the relative abundance of Basidiomycota and Zygomycota were significantly (P<0.05) greater in the soil by 8.9% and 2.0% respectively (Figure 2.19).

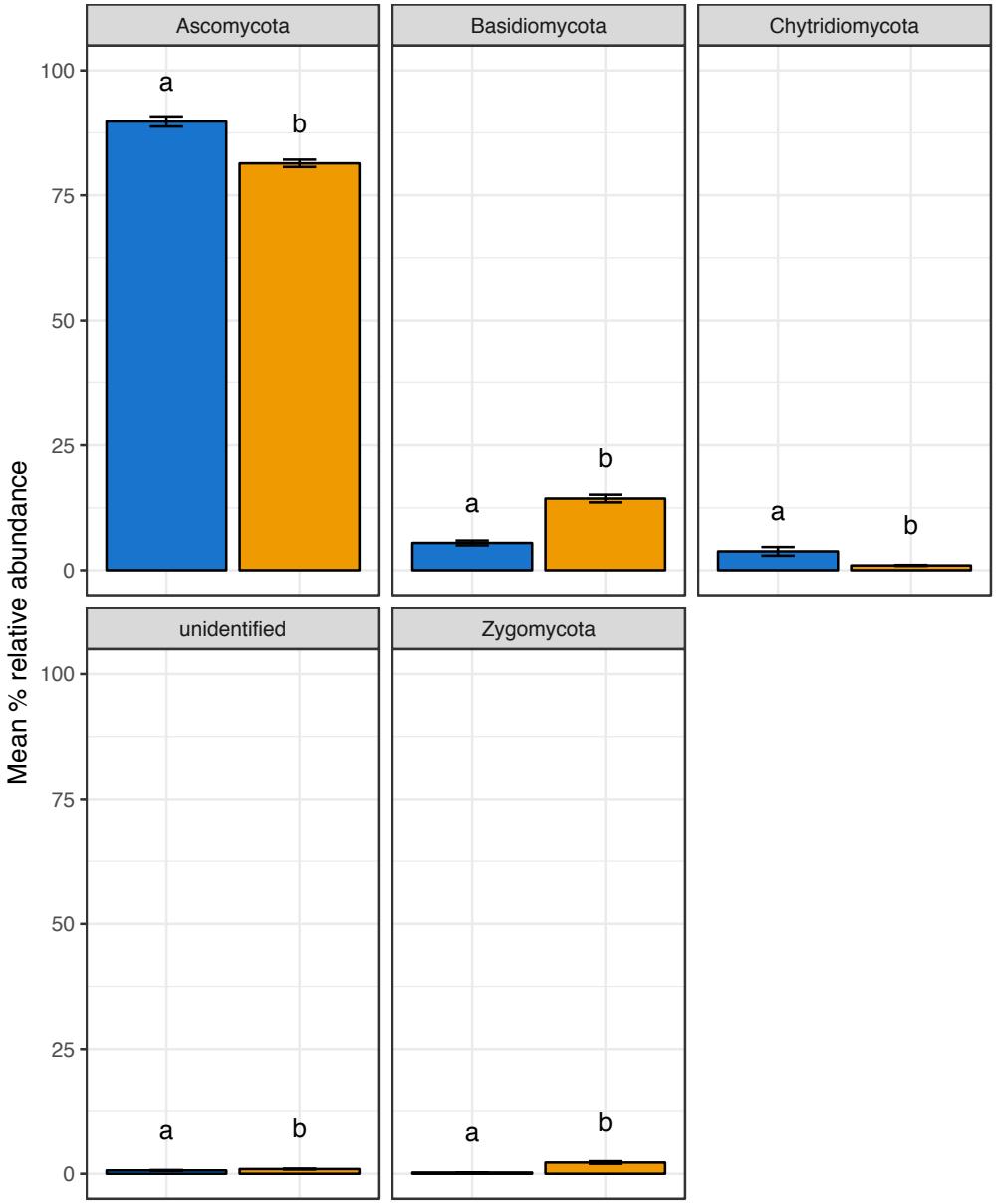


Figure 2.19 Facetted bar plots displaying mean relative abundance of the fungal phyla in the rhizosphere (blue) and soil (orange) samples. Error bars display standard error of the mean. Letters denote significant difference (P<0.05) based on a Kruskal Wallace test, where groups that share a letter have no significant difference.

N did not significantly affect the relative abundance of any phyla in the rhizosphere or soil (Figure 2.20).

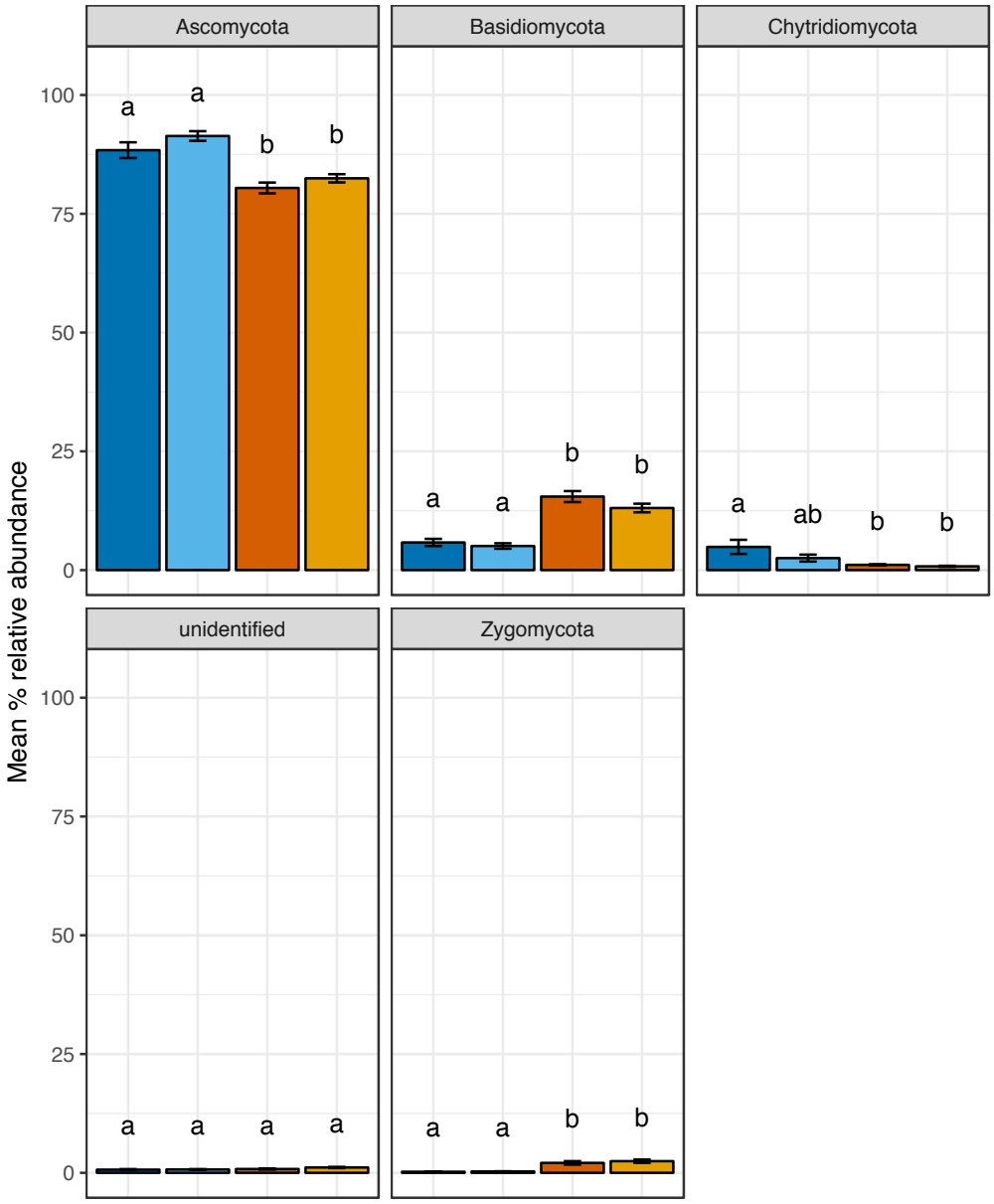


Figure 2.20 Facetted bar plots displaying mean relative abundance of the fungal phyla in the rhizosphere under high N (dark blue), rhizosphere low N (light blue), soil high N (dark orange) and soil low N samples. Error bars display standard error of the mean. Letters denote significant difference (P<0.05) based on a Kruskal Wallace test, where groups that share a letter have no significant difference.

The Fungal OTUs that contributed to at least 2% difference between the rhizosphere and soil were identified by SIMPER. The relative abundance of Leptosphaeria maculans, *Davidiella* tassiana, *Pyrenopeziza brassicae*, Leotiomycetes, *Olpidium brassicae* and *Chalastospora ellipsoidea* OTUs were enriched in the rhizosphere by 9.7%, 6.4%, 2.3%, 3.4%, 3.4% and 0.7% respectively (Table 2.10). Conversely Sordariomycetes and *Exophiala equina* OTUs had higher relative abundance in the soil by 3.1% and 2.4% respectively.

Table 2.10 Fungal OTUs identified by SIMPER analysis that contribute to the top 10% of differences between rhizosphere and soil samples.

OTU ID	Highest level taxonomic information	Rhizosphere mean relative abundance (%)	Soil mean relative abundance (%)	Contribution to difference (%)
2	*Leptosphaeria maculans*	15.6	5.9	9.9
3	*Davidiella tassiana*	14.6	8.2	8.2
6	*Pyrenopeziza brassicae*	4.6	2.3	3.4
23	*Leotiomycetes*	3.9	0.5	2.8
42	*Olpidium brassicae*	3.5	0.1	2.8
24	*Chalastospora ellipsoidea*	2.5	1.8	2.0
12	*Exophiala equina*	1.4	4.6	2.6
4	*Sordariomycetes*	1.9	4.3	2.4

In the rhizosphere, the fungal OTUs identified by SIMPER that contributed to at least 3% of the difference between the N treatments were *Davidiella tassiana, Leptosphaeria maculans, Pyrenopeziza brassicae,* Leotiomycetes and *Olpidium brassicae* OTUs, which were enriched in the high N treatment by 9.3% *2.5*% *1*% *1.2*% and *2*% respectively (Table 2.11).

Table 2.11 Fungal OTUs identified by SIMPER analysis that contribute to the top 10% of differences between the rhizosphere samples under contrasting N treatment, and the corresponding relative abundance of the soil samples.

OTU ID	Highest level taxonomic information	Rhizosphere high N relative mean abundance (%)	Rhizosphere low N mean relative abundance (%)	Contribution to difference (%)
3	*Davidiella tassiana*	18.9	9.6	11.0
2	*Leptosphaeria maculans*	16.7	14.3	10.8
6	*Pyrenopeziza brassicae*	5.1	4.1	5.0
23	*Leotiomycetes*	4.5	3.2	3.8
42	*Olpidium brassicae*	4.4	2.4	3.8

OTU ID	Highest level taxonomic information	Soil high N relative mean abundance (%)	Soil low N mean relative abundance (%)	Contribution to difference (%)
3	*Davidiella tassiana*	8.9	7.4	7.3
2	*Leptosphaeria maculans*	7.0	4.7	6.3
6	*Pyrenopeziza brassicae*	2.8	1.8	2.7
23	*Leotiomycetes*	0.6	0.4	0.6
42	*Olpidium brassicae*	0.1	0.0	0.1

In the soil, the fungal OTUs identified by SIMPER that contributed to at least 2% of the difference between the N treatments were *Davidiella tassiana, Leptosphaeria maculans, and Pyrenopeziza brassicae* OTUs, which were enriched in the high N treatment by *1.5*% *2.3*% and *1*% respectively (Table 2.12). Conversely, in the low N soil the relative abundance of Sordariomycetes*, Chalastospora ellipsoidea, Crytpococcus podzolicus* and *Exophiala equina* OTUs were greater by 0.4%, 0.8%, 0.9% and 0.5% respectively*.*

Table 2.12 Table of fungal OTUs identified by SIMPER analysis that contribute to the top 10% of differences between the soil samples under contrasting N treatment, and the corresponding relative abundance of the rhizosphere samples.

OTU ID	Highest level taxonomic information	Soil high N relative mean abundance (%)	Soil low N mean relative abundance (%)	Contribution to difference (%)
3	*Davidiella tassiana*	8.9	7.4	7.3
2	*Leptosphaeria maculans*	7.0	4.7	6.3
4	Sordariomycetes	4.1	4.5	3.0
6	*Pyrenopeziza brassicae*	2.8	1.8	2.7
24	*Chalatospora ellipoidea*	1.4	2.2	2.4
11	*Cryptococcus podzolicus*	1.2	2.1	2.1
12	*Exophalia equina*	4.3	4.8	2.0

OTU ID	Highest level taxonomic information	Rhizosphere high N relative mean abundance (%)	Rhizosphere low N mean relative abundance (%)	Contribution to difference (%)
3	*Davidiella tassiana*	18.9	9.6	11.0
2	*Leptosphaeria maculans*	16.7	14.3	10.8
4	Sordariomycetes	1.4	2.4	1.2
6	*Pyrenopeziza brassicae*	2.7	4.1	5.0
24	*Chalatospora ellipoidea*	2.6	2.3	2.3
11	*Cryptococcus podzolicus*	0.2	0.4	0.4
12	*Exophalia equina*	1.1	1.7	0.9

Using the FUNGuild program, a fungal guild was assigned to 49% of fungal OTUs. ANOSIM analysis revealed that rhizosphere and soil guild composition were significantly ($P<0.05$) different. In the rhizosphere there was a significant ($P<0.05$) increase of 15.2% in relative abundance of the Pathotroph guild (Figure 2.21). In the soil increases in the relative abundance of Pathotroph-Saprotroph, Pathotroph–Symbiotroph, Saprotroph-Symbiotroph and Symbiotroph guilds were identified by 4.6%, 0.4%, 0.9% and 0.07% respectively.

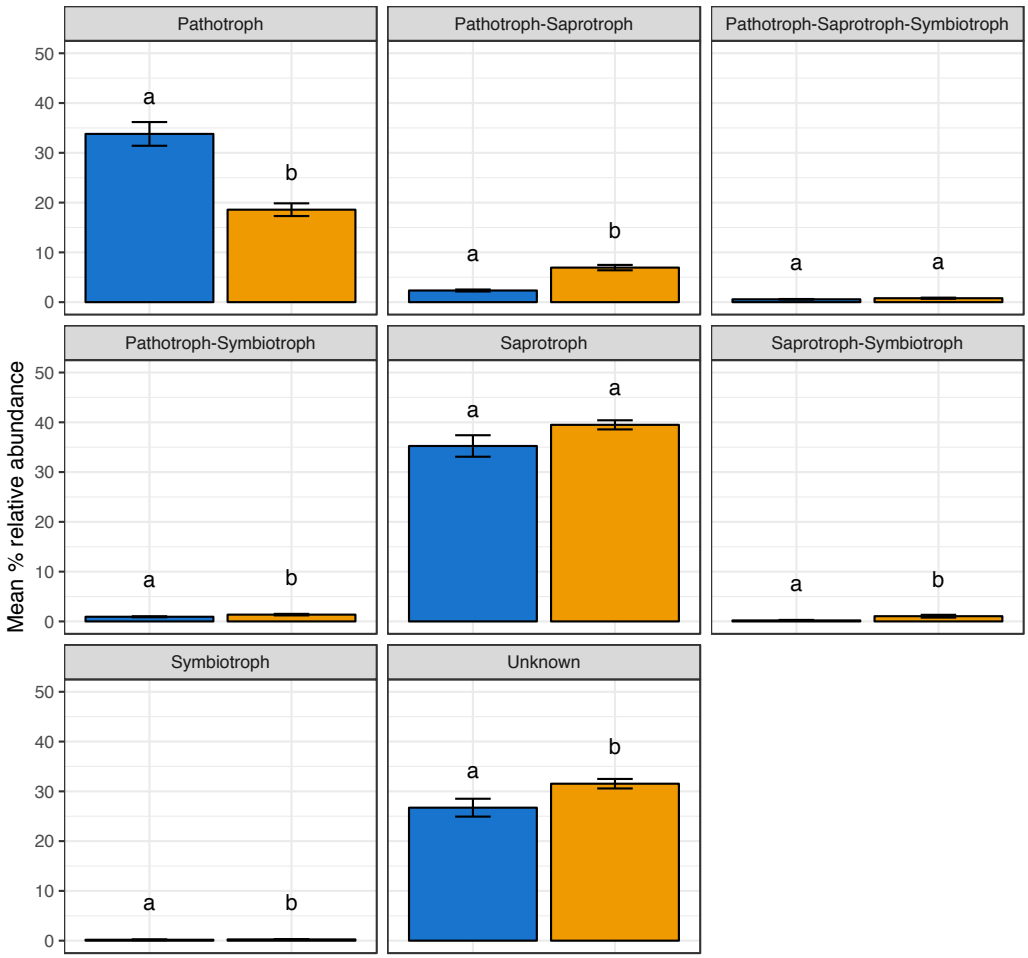


Figure 2.21 Facetted bar plots displaying mean relative abundance of fungal guilds in the rhizosphere (blue) and soil (orange) samples. Error bars display standard error of the mean. Letters denote significant difference (P<0.05) based on a Kruskal Wallace test, where groups that share a letter have no significant difference.

N treatment (Figure 2.22) had no significant effect on guild composition in the soil or rhizosphere. Furthermore, genotype had no effect on guild composition in the rhizosphere.

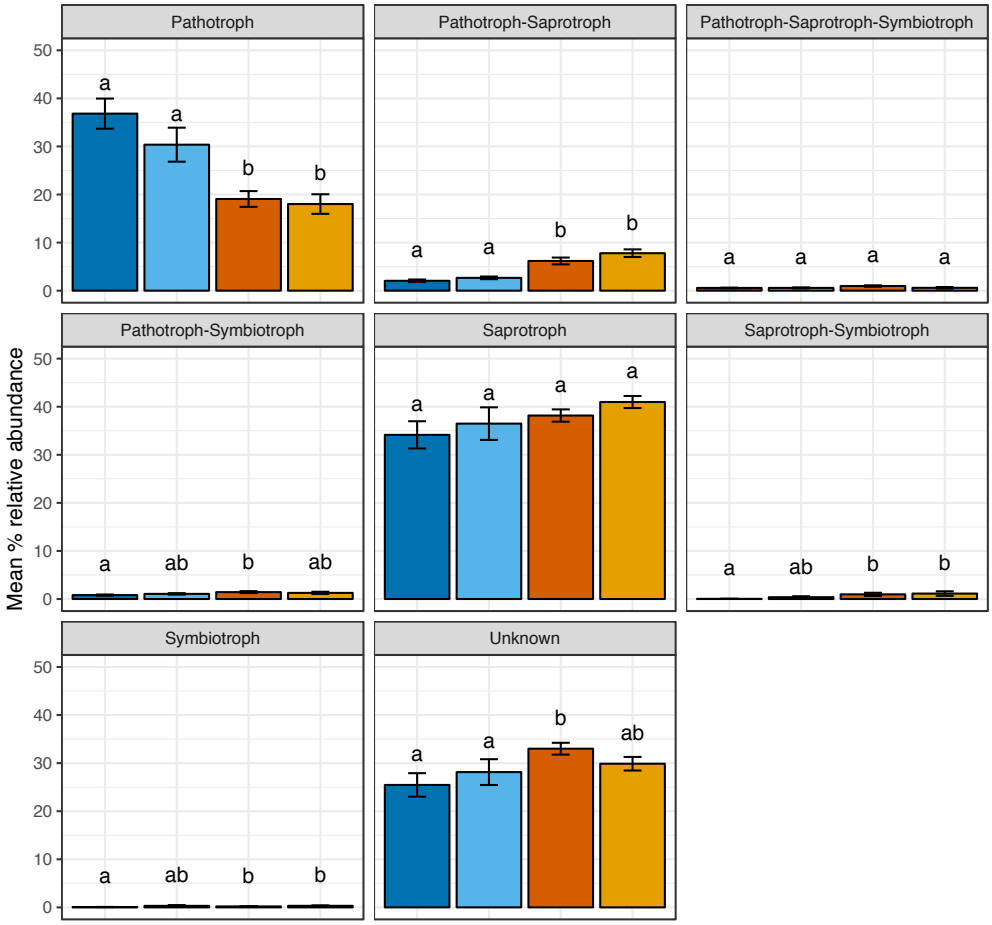
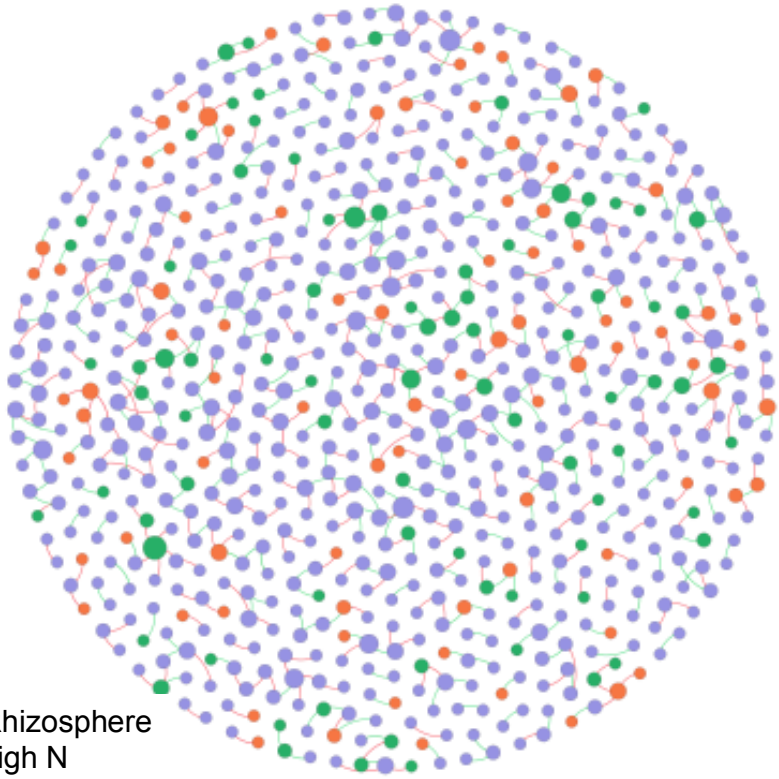


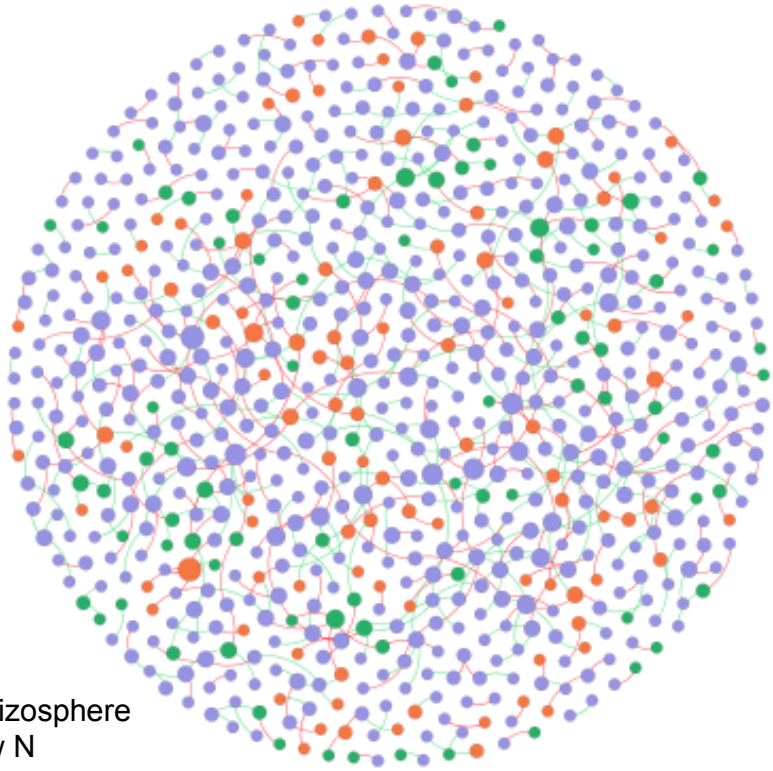
Figure 2.22 Facetted bar plots displaying mean relative abundance of the fungal guild composition of rhizosphere high N samples displayed in dark blue, rhizosphere low N samples displayed in light blue, soil high N samples displayed in dark orange and soil low N samples displayed in light orange. Error bars display standard error of the mean. Groups sharing a letter are not significantly different (P<0.05) based on a Kruskal Wallace test.

2.3.4 Network analysis

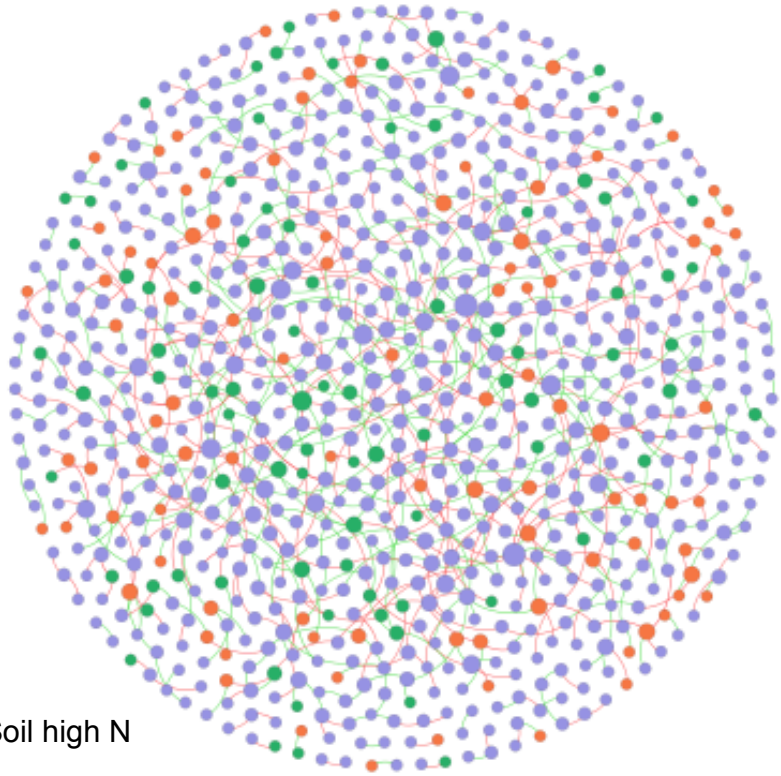
Network analysis revealed the potential interactions between different taxa. Network diagrams highlight any positive or negative correlation of the relative abundance of OTUs. The points on the network are referred to as nodes, and the connecting lines between the nodes are referred to as edges. Edges represent a significant correlation, either positive or negative between an OTU. The networks have been displayed as averages for each of the rhizosphere high N, rhizosphere low N, soil high N and soil low N groups (Figure 2.23).



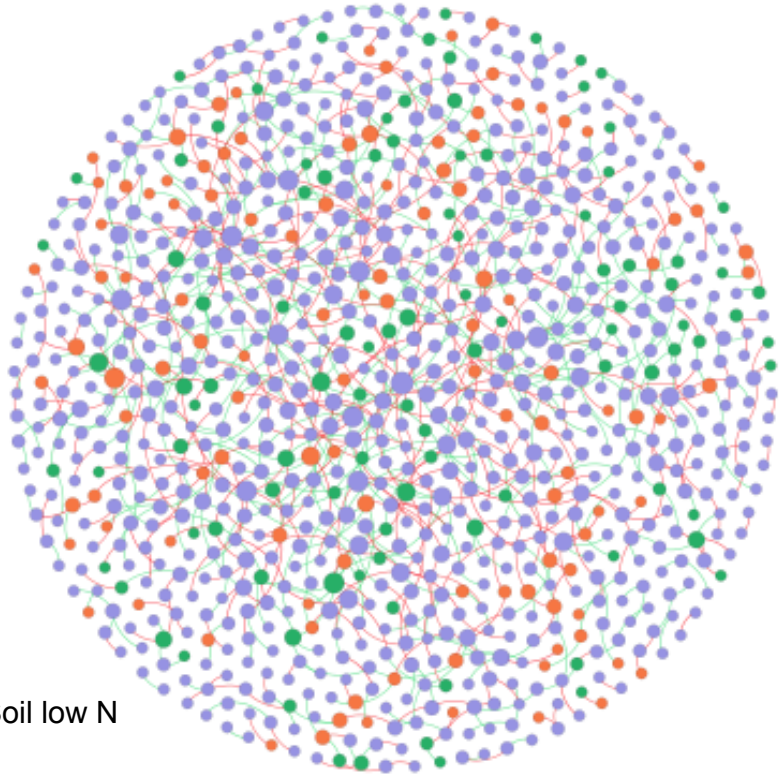
Rhizosphere
high N



Rhizosphere
low N



Soil high N



Soil low N

Figure 2.23 Network diagrams for soil and rhizosphere under N regimes. Bacteria are represented by purple nodes, protist OTUs are represented by green nodes and fungi are represented by orange nodes. Positive interactions are represented by green edges and negative interactions are indicated by red edges.

Network statistics indicate a higher level of connectivity in the soil microbial community than the rhizosphere. Soil networks contained more edges (864) and higher average degree (edges per node) (1.9) than the rhizosphere samples (Table 2.13), which contained an average of 583 and an average degree of 1.5). The nature of the interactions was similar, the rhizosphere networks contained 49.1% positive interactions compared to 49.3% positive interactions in the soil networks.

Table 2.13 Network information. Nodes represent OTUs from combined datasets. Edges represent the significant interactions between nodes. Average degree represents the mean number of edges per node. Modularity represents the density of edges inside modules compared to links between modules.

Group	Nodes	Edges	Average degree	Modularity
Rhizosphere high N	725	558	1.5	0.978
Rhizosphere low N	752	608	1.6	0.969
Soil high N	827	806	1.7	0.894
Soil low N	904	922	2.0	0.877
Rhizosphere	739	583	1.5	1.0
Soil	866	864	1.9	0.9

Group	Negative edges	Positive edges	Percentage negative	Percentage positive
Rhizosphere high N	285	273	51.1	48.9
Rhizosphere low N	308	300	50.7	49.3
Soil high N	415	391	51.5	48.5
Soil low N	460	462	49.9	50.1
Rhizosphere	286.5	296.5	49.1	50.9
Soil	426.5	437.5	49.3	50.7

N application decreased network connectivity in a similar manner in both the rhizosphere and soil. In the soil, the high N networks contained 116 fewer edges than the low N treatment, and had an average degree of 1.7 compared an average degree of 2.0 in the soil low N networks. In the rhizosphere networks the low N treatment had an average degree of 1.6 compared to the high N treatment, which had an average degree of 1.5, and, the high N treatment rhizosphere network contained 50 less edges than the low N treatment network. The percentage of negative interactions in both the soil and the rhizosphere were greater by 0.6% in the soil and 0.4% in the rhizosphere under higher N conditions.

2.4 Discussion

Significant differences were identified in bacterial, protist and fungal community composition between the rhizosphere and soil compartments. N treatment had distinct effects on each group, with assembly of bacterial communities affected in the rhizosphere only, protists in the soil only, and fungi affected in both the soil and rhizosphere. Network analysis provided indications that microbial community connectivity in both the soil and rhizosphere was affected by N treatment. For both fungi and bacteria, there was evidence for function differences between the rhizosphere and bulk soil, and for bacteria, predicted metagenomes indicated differences between the N treatments. However N application had no effect on distribution of fungal guilds in the soil or rhizosphere. Significantly, plant genotype had no effect on any aspect of microbial composition or function.

2.4.1 Plant genotype and microbial community

Although the microbiome of *Arabidopsis thaliana*, a model plant from the Brassicaceae, has been studied in depth (Lundberg *et al*., 2012), the rhizosphere microbiome of *Brassica napus*, has received little attention. These results show similarity between the *A. thaliana* and *B. napus* bacterial microbiomes, with Actinobacteria, Acidobacteria, Proteobacteria and Bacteroidetes dominating the rhizosphere bacterial communities.

There are no fungal or protist studies on the rhizosphere microbiome of *Arabidopsis thaliana* at the time of writing. The SIMPER analyses indicated that top OTUs enriched in the rhizosphere were *Flavobacterium* (16S, OTU 8) *Spongospora subterranea* (18S, OTU 28) and *Leptosphaerica maculans* (ITS, OTU 2). The *Flavobacterium* genus contains 140 known species that are mainly commensal, but can have both pathogenic or positive impacts on plants (Soltani *et al*., 2010). *Flavobacterium* are often recorded in higher abundance in the rhizosphere than the soil for plant species (Kolton *et al*., 2016), and have been shown to have the potential have a role in plant growth promotion (Umamaheswari *et al*., 2013). *Spongospora subterranea* (powdery scab) is known to be able to cause pathogenicity to potatoes (George *et al*., 2004). *B. napus* has the potential to act as a trapping plant for *Spongospora subterranea,* as it is colonised but slows the life cycle of this plasmodiophorid (Qu and Christ, 2006). *Leptosphaeria maculans* (stem canker) is an Ascomyceete pathogen of *Brassica* crops (Rouxel and Balesdent, 2005). *Pyrenopeziza brassicae,* (Light leaf spot) is a foliar pathogen that is not believed to have any life cycle below ground (Boys *et al*., 2007), potentially there is an unknown role of this pathogen in the rhizosphere of *B. napus*.

Plant genotype has frequently been described as having the potential to influence the assembly of rhizosphere microbial communities (Bulgarelli 2015, Micallef 2009). Genotype is thought to alter rhizosphere microbiome composition due to differences in concentration and composition of rhizodeposits that will influence the assembly of microbial communities (Bulgarelli *et al*., 2013).

No effect of plant genotype on the diversity of the rhizosphere microbial communities was identified. Additionally, no overall effect of genotype on the microbial community composition, or function in the rhizosphere was identified. This suggests that in field systems, exposure to environmental variables dilutes the effect that an individual plant can have on the assembly of its rhizosphere microbiome. This is supported by the work of

Edwards and colleagues (Edwards *et al*., 2015), who revealed that while rice genotype had a significant role in assembly of the rhizosphere microbiome under greenhouse conditions, this effect was lost when plants were grown in the field (Edwards *et al*., 2015). Similarly no effect of the genotype of *Boechera stricta* (Drummond's Rockcress, Brassicaceae family) on rhizosphere bacterial composition was found under field conditions (Wagner *et al*., 2016).

2.4.2 Effect of Nitrogen level on microbial communities

Overall, N significantly altered the rhizosphere bacterial communities, but not the soil bacterial communities. In the case of fungi, rhizosphere and soil microbial communities were significantly changed by N. Surprisingly, the protist communities were significantly altered by N in the soil, but not in the rhizosphere.

Availability of soil N has previously been shown to alter bacterial (Ramirez *et al*., 2010), protist (Krashevska *et al*, 2012) and fungal (Paungfoo-Lonhienne *et al*., 2015) community compositions. In the rhizosphere however, changes in N availability can influence community composition via direct effects, and indirectly, *via* the responses of the plant to N. The composition and concentration of plant root exudates can be determined by the physiological state of the plant (Chaparro *et al*., 2013), and increasing N fertilisation has been seen to positively correlate with the rhizodeposition of sugars, sugar alcohols and phenolics, and altered bacterial community composition (Zhu *et al*., 2016).

Whilst no direct measurements of rhizosphere exudation profiles were taken, the OTUs which responded to N in the rhizosphere, but not in the bulk soil, such as *Pedobacter*, *Promicromonospora* and *Flavobacterium* may indicate an indirect response to N of rhizosphere specialists, via the host plant. In contrast, OTUs that responded to N treatment in both the soil

and rhizosphere, such as Micrococcaceae, *Davidiella tassiana*, and *Leptospharea maculans* (ITS, OTU 2), could be responding directly to N.

Whilst individual taxa responded to N, the overall response of the bacterial community to N was only significant in the rhizosphere. In contrast, the fungal community significantly shifted in both the rhizosphere and soil. The response of the bacterial community to N level in the rhizosphere and not the soil implies that a factor other than the available N is driving the change in the microbial community.

The lower level of available N may have caused a shift in exudation profile in *Brassica napus* and may have contributed to significant shifts in bacterial community composition in the rhizosphere. Both plant health and nutrient availability can impact the concentration and composition of plant root exudates. Under N stress plants have been seen to increase exudation of glycerol, ribitol, fructose, and malate (Carvalhais *et al*., 2011). This in turn has the potential to influence the microbial community composition in the rhizosphere, for example malate and citrate concentrations have been shown to increase recruitment of a growth promoting bacteria in the rhizosphere (Tawaraya *et al*., 2014).The fungal community shifts in response to N were apparent in both in the soil and the rhizosphere. Although the shift in plant exudation patterns may have also impacted the influenced the fungal community, the N level changes alone were sufficient to drive a change in fungal community composition.

The protist data did not conform to either of these models, since the protist community composition was altered by change in N in the soil but not the rhizosphere. A possible cause for this could be the uneven level of sequencing and filtering required for amplicon sequencing of the protist community, resulted in a low diversity and low evenness. For future work, it would be beneficial to use a transcriptomic approach. The sequencing of total RNA would allow for the direct, quantitative study of microbial communities uniformly without amplification bias using total rRNA. Transcriptomic approaches would also incorporate the use of mRNA to

characterise the genuine functional fingerprint of microbial communities, rather than using functional predictions.

2.4.3 The impact of compartment and nitrogen on microbiome connectivity

Network analysis revealed that the soil samples had higher connectivity, indicated by the increased number of edges and average degree, than the rhizosphere samples. Increased connectivity represents an increase of interactions between members of the microbiome. The interactions can be positive, such as in some form of symbiotic or commensal relationship, or negative if taxa are competing for the same resources or directly preying on other taxa.

This is in contrast to existing rhizosphere network studies which have found that rhizosphere networks were more complex than soil networks (Shi *et al*., 2016) and (Yan *et al*., 2016). It has been hypothesized that the rhizosphere offers increased stability for interactions to occur, and a more heterogeneous environment. However the previous studies were conducted in sieved soil and conducted in a pot experiment. This causes substantial disruption of the soil structure and therefore microbial communities. Under such circumstances the rhizosphere may present a more stable environment, as communities will reassemble in response to root exudation. However, in undisturbed soil systems, the growth of the root presents a disruption to existing soil networks and this could be reflected a decrease in network complexity, relative to the soil, as was found in the current study. Network analysis of field grown rhizosphere and soil communities has revealed more complex networks in the soil than the rhizosphere (Mendes *et al*., 2014). Mendes and colleagues speculate that the increased diversity of the soil offers more opportunities for interactions than the less diverse rhizosphere. This would match these findings, as the diversity of the rhizosphere communities were all significantly less than the bulk soil samples.

Low N networks had more edges and higher average degree than the high N networks, implying an increase in connectivity. This could be due to an increase in competitiveness due to the decreased nutrient availability, or coordination in order to utilise limited resources, his trend of increased complexity correlating to increased N levels has been previously identified within soil bacteria (Ma *et al*., 2016).

The percentage of the total interactions was an average of 49.2% positive and 50.8% negative across all networks. This, again is in contrast to (Shi *et al*., 2013) who report predominantly (>80%) positive interactions in the rhizosphere albeit using a different algorithm, and (Yan *et al*., 2016) who reported a higher percentage of positive interactions in all rhizosphere and soil networks, using the same algorithm as used in this study. The lower percentage of positive interactions in this study could be as a result of the plant specific interactions of the microbial community, or the difference in community composition and interactions of different soil types, as well as changes in community assembly and connectivity induced by soil processing in the previous studies.

Overall network analysis is a new field. There are many tools and approaches for generating networks, such as SparCC (Friedman and Alm, 2012), MENAP (Deng *et al*., 2012) and CONET (Faust *et al*., 2012). Methods have not been standardised, and the impact that choice of software used to generate networks has on network structure is unclear. Additionally, the procedures, in which OTU tables are manipulated prior to network analysis, are not included, varied and often unclear.

2.4.4 The impact of compartment, nitrogen and plant genotype on predicted function of bacterial communities and fungal guilds.

Increased relative abundance of gene categories related to movement and motility were identified in the rhizosphere community, relative to the bulk soil. In particular, significant increase of ‘Cell motility’ and ‘Flagellar

assembly' in the rhizosphere. This corresponds with rhizosphere functional motility assays analyses conducted by (Czaban *et al*., 2007). Taxa that are able to colonise the rhizosphere from the soil, have increased motility compared to those from the bulk soil, as the ability to move quickly to colonise the plant root may offer a selective advantage (Ofek-Lalzar *et al*., 2014). Therefore it may be possible that other highly different categories identified by PICRUSt analysis might also be relevant. A significant increase in "Xenobiotics biodegradation and Metabolism" was identified in the rhizosphere, which could indicate the functional response of the bacterial community to the increased competition driven by plant exudation in the rhizosphere. Environments with increased competition and diversity may harbour more antimicrobial resistance (AMR) genes. Recent work has identified increased abundance of genes related to competition in the rhizosphere (Yergeau *et al*., 2014). No influence of genotype was identified on the predicted rhizosphere metagenomes.

Overall, N treatment had no significant effect on the predicted metagenome of the soil. However, in the rhizosphere, the relative abundance of fourteen KEGG level 3 processes were significantly greater in the predicted metagenomes of the high N treatment, and twenty-three were significantly greater in the low N treatment. Of the groups enriched under low N, five have the potential to function in biocontrol; Bacterial toxins (Jousset *et al*., 2009), Novobiocin synthesis (Flinspach. 2014), Polyketide sugar biosynthesis, streptomycin biosynthesis and Vancomycin biosynthesis (summarized in (Thiele-Bruhn, 2003). Antimicrobial activity, indicated by an increase of streptomycin producing genes has found to be higher in the rhizosphere than the soil, possibly reflecting increased competition following bacterial utilisation of exudates (Chaparro *et al*., 2014). However, the interaction of N with these processes is unknown. Under low N conditions the potential for antibiotic production may be greater, suggesting increased competition in the rhizosphere low N availability environment. This could have significance for approaches to identify novel antimicrobial compounds.

Another KEGG category of potential interest is the high increase in methane metabolism in the low N rhizosphere (Figure 2.12). N has been shown to be a limiting factor for methanotrophic microbes (Irvine *et al*., 2012). However of the OTUs enriched in the low N rhizosphere, the top two most abundant taxa were a *Methylybium* OTU *and a Methylotenera mobil*is OTU (Supplementary Table 1). *M. mobilis* has been identified as an obligate methylamine –utilising bacteria (Kalyuzhnaya *et al*., 2006). It is possible that the local depletion of plant available ammonium and nitrate drives enrichment of microbial populations capable of accessing N from organic compounds, such as methylamine.

The predicted relative abundance of N cycle genes were analysed, as outlined by Zhu and coworkers (Zhu *et al*., 2016). In the rhizosphere, a significant increase of the N fixation gene *nifH* under high N treatment was identified, but the other genes were not significantly affected by N treatment. It is possible the level of N fertilisation was not sufficient to drive a large-scale shift in N cycle genes. These findings also may indicate an indirect effect of the N on the rhizosphere bacterial community, via changes in plant health and exudation.

PICRUSt predicted metagenomes have been directly compared to soil metagenomes with a significant Spearman's correlation $P<0.001$, $R=0.81$ (Langille *et al*., 2013). However, the accuracy of the predicted metagenomes are limited by number of genome-sequenced organisms in the environment of study. Although PICRUSt is able to predict the metagenome organisms without whole genome sequencing using it's ancestral state algorithm, a number of closely related genome sequences are required in order to make accurate predictions. The soil microbiome, as previously discussed, comprised of a vast number of different taxa with no genome sequence data, so there is still a large proportion of the microbiome that may not be identified using this technique.

The fungal community of the rhizosphere contained a higher relative abundance of pathotrophic fungi than the soil. Fungal communities can

broadly be categorised as pathotrophic, saprotrophic or symbiotic, or a combination thereof (Tedersoo *et al*., 2014). No significant effect of this level of N fertilisation was identified on the guild composition of the fungi. The impact of N on fungal communities have been studied in sugar cane (Paungfoo-Lonhienne *et al*., 2015), and N treatment was seen to increase the level of pathotrophic fungi in the rhizosphere. The N levels used in this study were lower than those used by (Paungfoo-Lonhienne *et al*., 2015) which may account for this difference.

2.4.5 Conclusions

Assembly of the rhizosphere microbiome is of importance to modern agriculture, due to the impact that the composition of the microbial community can have on yield. Research into the drivers of rhizosphere community assembly can help identify conditions, management practices, environmental changes or crop phenotypes that can lead to a “beneficial” microbiome, which has the ability to feedback into crop yield.
In the present study, the rhizosphere microbiome of *Brassica napus* from an agricultural field in the UK was characterised. The most abundant bacterial phyla in the rhizosphere were similar to those observed in *A.thaliana* (Bulgarelli *et al*., 2012) with, Proteobacteria and Actinobacteria most represented in the rhizosphere, however in this study a more abundant Bacteroidetes community was present. The OTUs that were observed to have the highest relative abundance were *Flavobacterium*, *Pedobacter*, *Promicromonospora* and Micrococcacea which may have beneficial effects on plant health. Resolution of species and functional studies would be required in order to investigate further. Interestingly the relative abundance of a *Promicromonospora* OTU was enriched under high N in the rhizosphere, whereas the relative abundance of this OTU in the soil was unchanged. This pattern of enrichment may indicate a role of plant exudates in driving the abundance of this OTU rather than N level. Specialist beneficial microbes respond to changes in plant exudation profiles. For example, increases in exudation of a benzoxazinoid from

maize roots was seen to recruit a beneficial *Pseudomonas putida* strain KT2440 (Neal *et al*., 2012).

A possible specialist association may be present between this *Promicromonospora* OTU and *B. napus. Promicromonospora sp.* SE188 was seen to produce plant-growth promoting gibberlilins and have a high phosphate solubilisation potential which increased plant biomass (Kang *et al*., 2012). The fungal rhizosphere microbiome detected here contained mainly Ascomycetes such as *Leptospharea maculans*. The ITS3 and ITS4 primers used here did not amplify *Olpidium brassicae* to the same extent as has been seen in the rhizosphere of OSR previously (Hilton *et al*., 2013,). A potential novel life cycle role for *Pyrenopeziza brassicae,* (Light leaf spot) was seen as it is not expected in the rhizosphere. The protist OTUs most enriched in the rhizosphere were *Spongospora subterranea* (powdery scab) and *Phytophthora infestans* (late blight) which both have pathogenic potential. N level had a greater influence on community composition than the wide range of genotypes used as host plants.

Genotype of plant was not observed as having a significant effect on microbiome assembly. The diluting effect of the environmental conditions in a field setting may be the main factor for this lack of effect. A possible direction for future studies would be a comparative microbiome study of pot and field grown *B. napus* in order to determine the impact these conditions had on microbial community composition, function and connectivity. Bacterial communities responded to N in the rhizosphere but not the soil and may have been responding to an indirect impact of N fertilisation, such as plant exudates effects. This contrasts to fungal communities which shifted community composition in both the soil and rhizosphere under N. The rhizosphere bacterial community in the lower N treatment group had a predicted increase in antimicrobial compound producing genes, which may offer a new area of study for novel antibiotics.

Chapter 3. Optimisation of soil RNA preservation techniques for metatranscriptome analysis and comparison of the active communities of root, rhizosphere and bulk soil

3.1 Introduction

3.1.1 Dynamics of rhizosphere microbial communities

Microbial communities can display extraordinary diversity in natural environments. Especially so in the soil, where thousands different species of bacteria may be found in just one gram of soil (Roesch *et al*., 2007). On top of bacterial diversity, the soil also hosts large numbers of archaea, protists and fungi (Fierer, 2017). Microbial communities are recruited into the rhizosphere from soil, which supports distinct microbial communities to the soil as seen in Chapter 2. The microbial communities of the root and rhizosphere soil can have significant effects on plant health and productivity. The rhizosphere microbiome contains many taxa, which are potentially pathogenic to plants, which in an agricultural context can lead to significant losses of yield (Hilton *et al*., 2013). The abundance of pathogens species can be controlled by a competition for resources with other biota in the rhizosphere, predation by taxa from higher trophic levels (Jousset *et al*., 2009) and production of antimicrobial compounds (Haas and Keel, 2003). Microbes in the rhizosphere also perform biogeochemical cycling roles that benefit the plant, such as mycorrhizal fungi which provide plants access to soil nutrients such as P (Bonfante and Genre, 2010) and *Rhizobia,* which can fix N in root nodules (Andrews and Andrews, 2017).

As human populations grow, the efficiency of agriculture must be improved in order to cope with the increased demand for food. Developing an understanding of the dynamics of soil and rhizosphere microbial

communities is an important step towards manipulating rhizosphere processes to benefit humans.

3.1.2 Environmental microbiology methodologies

Historically, the study of microorganisms typically involved the isolation of an organism from an environment and assessing its functional capabilities using culture based methods and assays (Sutra *et al*., 2000). However, only a small percentage of the total microbial diversity has been successfully cultured (Rappé and Giovannoni, 2003). Whilst culture based methods are still vital for the understanding of individual microorganisms, over the past two decades molecular approaches have provided opportunities to simultaneously characterise the composition and functions of the diverse microbial communities typical of environmental compartments.

PCR amplicon sequencing has greatly extended the scope of rhizosphere microbial studies, with the search terms “Rhizosphere amplicon sequencing” identifying 881 of papers from the last 5 years on the NCBI website (NCBI, U.S. National Library of Medicine, 8600 Rockville Pike, Bethesda MD, 20894 USA). Although this obviously does not represent a definitive collection of all amplicon based rhizosphere studies, it demonstrates the publication of this type of study at a rate of one every other day. Through these DNA amplicon approaches, rhizosphere microbiomes have been developed for a growing number of plant species.

As described in detail in Chapter 1, a problem with using DNA to characterise the microbial community is that DNA can remain present in the soil environment for years as ‘relic’ DNA (Carini *et al*., 2016).
The use of RNA allows for the taxonomic identification of the active rhizosphere microbial community, as the half life for RNA is short (for example ~6.8 minutes for *E. coli* mRNA (Selinger *et al*., 2003) and

metatranscriptomics, which is the study of the total RNA in a given sample at a single time point, can reveal the function of these communities.

3.1.3 Rhizosphere metatranscriptomics

A study by Turner and colleagues marked the first rhizosphere metatranscriptome analysis (Turner *et al*., 2013), identifying a number of differences in the rhizosphere metatranscriptomes from a variety of plant species, and detecting increases in processes suspected to be beneficial in the rhizosphere such as cellulose degradation and methylotrophy. Later studies have discovered growth stage dependent changes in rhizosphere metatranscriptome profile in Arabidopsis (Chaparro *et al*., 2014), responses of rhizosphere metatranscriptomes to glyphosate treatment in the rhizosphere of plants with or without glyphosate tolerance (Newman *et al*., 2016) and the functional response of a rhizosphere microbiome following colonisation by a bacterial plant pathogen (Zhang *et al*., 2017). However these studies all took different approaches to rhizosphere soil sampling. Chaparro and colleagues did not separate root and rhizosphere soil (Chaparro *et al*., 2014). Turner and colleagues shook root samples to remove and collect rhizosphere soil (Turner *et al*., 2013). Newman and colleagues used soil cores from a rhizobox (40 × 20 × 2 cm container (Newman *et al*., 2016), where all soil in the cores was defined as rhizosphere soil. Zhang and colleagues scraped roots with brush pencils to collect rhizosphere soil (Zhang *et al*., 2017).

The microbial communities of the root (endorhizosphere) and rhizosphere soil (ectorhizosphere), have been shown to be significantly distinct from each other in many plant species (Bulgarelli *et al*., 2013). One of the aims of this chapter was to find a method of completely separating the rhizosphere soil from the root tissue, so that the metatranscriptomes in each compartment could be studied independently. Snap freezing root with rhizosphere attached would be the ideal preservation technique but

subsequent separation of the rhizosphere from the root would involve thawing, which would degrade RNA prior to extraction.

The majority (95-99%) of RNA present in soil samples can be ribosomal RNA (Mettel *et al*., 2010). Whilst this can be used for taxonomic profiling of the microbiome, analysis of the messenger RNA is required to determine the functional characteristics of the microbiome. Ribosomal RNA is often depleted prior to metatranscriptome sequencing. However the cost of these depletion techniques can limit the number of samples which can subsequently be sequenced. In order to develop cheap, high throughput methods for metatranscriptome analysis, it would be beneficial to avoid a ribodepletion step.

Due to the short half life of RNA (~5 minutes) to degradation from chemicals and enzyme activity, it is typically recommended that samples for RNA extraction should either be stored at −80°C or maintained in an RNA preservation solution such as LifeGuard™ Soil Preservation Solution (MO BIO Laboratories, USA). The effectiveness of Lifeguard solution to preserve the functional fingerprint of a microbiome has yet to be determined. If it does not have a significant impact on the taxonomic and functional fingerprint of the rhizosphere and root microbiomes it would provide a useful tool for separation of rhizosphere soil from the root whilst maintaining RNA integrity. However the cost of the Lifeguard solution may prohibit its use in large scale experiments. If samples could be frozen and subsequently thawed in Lifeguard then it would allow for preservation of a large number of samples followed by exploratory amplicon sequencing work in order to identify samples of interest for focussed metatranscriptomics. However it is unknown whether Lifeguard is effective at preserving RNA in samples that have been thawed in the solution.

Cost effective methodologies for separating the rhizosphere soil from the root without altering the transcriptional fingerprint of the respective communities would allow comparative analysis of the microbial interactions and functions within these compartments. A potential way to

separate the closely bound rhizosphere soil from the root is to wash the roots in water and flash freeze the rhizosphere wash and roots separately in liquid nitrogen. This allows a short time period between removal of the plant from the ground and preservation of RNA. In the lab these samples can be freeze-dried to remove all water and RNA can be extracted. This Freeze-drying method may allow for separation of rhizosphere soil from root without significant degradation of RNA.

3.1.4 Aims

The aims of this study were to 1.) Determine the extent to which different RNA preservation techniques influence the active microbial community and transcriptional fingerprint the soil. 2.) Determine differences in the active microbial composition and functional fingerprint between root, rhizosphere soil and bulk soil compartments 3.) Determine whether samples that had not undergone a ribosomal rRNA depletion would still yield sufficient mRNA to give an indication of the function of the microbiome

3.2 Materials and Methods

3.2.1 Experimental design.

The experiment was designed in order to be able to compare the effectiveness of different approaches for preserving soil for subsequent extraction of RNA. The four techniques were; 1.) Freeze-Dried, 2.) Frozen 3.) Lifeguard and 4.) Thawed (frozen then thawed in Lifeguard) (Table 3.1). The Frozen treatment was taken as the most similar to “real” community composition, as the snap freezing creates a snapshot of the total RNA, and prevents degradation by inactivating RNAses. This is why soil rather than rhizosphere was used to compare techniques as there was no equivalent “real” rhizosphere. Washing of soil samples was conducted in order to simulate the washing treatment for rhizosphere soil. The other three treatments were compared to the Frozen method to determine the

best method for future large scale rhizosphere metatranscriptomic experiments. In particular collection of rhizosphere soil requires washing of closely adhering soil from roots. These methods provide alternative approaches which can potentially be used to collect rhizosphere soil in the field

Furthermore, differences between the active microbial community composition and functional profiles of bulk soil, rhizosphere soil and root samples, were compared using the Lifeguard method (which was initially presumed to be most effective based on preliminary RNA quality (Suplementary figure 1). Additionally, the extracted RNA from frozen soil samples was divided and half was ribodepleted in order to determine if the mRNA profile of non-ribodepleted samples provided a sufficient representation of the mRNA compared to a non-ribodepleted sample. Each of these techniques was conducted using in triplicates. A summary of the treatments used for these comparative experiments is provided in Table 3.1.

Table 3.1 Sample names for each treatment with abbreviation ID. (*) WRO was not freeze-dried, the roots were snap frozen and homogenised in liquid nitrogen.

Treatment	Bulk soil	Rhizosphere	Root	Ribodepleted Bulk soil
Freeze-dried (Washed in water in field then snap frozen and freeze dried)	WS		WRO*	
Frozen (Direct freezing in liquid N)	NS			RiboNS
Lifeguard Washed in Lifeguard solution)	LGS	LGR	LGRO	
Thawed (Snap frozen, defrosted in Lifeguard)	NLGS			

3.2.2 Sample collection and processing

Samples were collected from a field trial plot on a sandy loam soil at Wellesbourne (Lat 52.211583, Long -1.607942) on Friday 2nd December 2016. The plot was a 6 m x 24 m area of *Brassica napus* planted in late August following 3 prior annual crops of wheat.

In order to minimise time between removal of the samples from the environment and RNA preservation, sampling was conducted using four 'pseudoreplicates', which were combined before RNA extraction. This involved sampling four times in an identical manner and later combining the samples. This procedure was done as quickly as possible and samples processed within 3 minutes of the plant or soil being removed from the ground.

3.2.2.1 RNA extraction

All RNA extractions (bulk soil, rhizosphere and root) were performed using the MOBIO RNA Powersoil extraction kit (MO BIO industries catalogue number 12866-25) according to the manufacturer's instruction. Extractions were completed over the two days after sampling.

3.2.2.2 Soil collection

Bulk soil for each pseudoreplicate was collected by removing and discarding the top 1 cm of soil one metre in from the edge of the plot in an area equidistant from the stems of adjacent plants and collecting a 10 cm depth sample of soil using a trowel. Three further soil samples were collected 2 metres apart, and the four samples were pooled and mixed in a bag. From this pseuodreplicate one of each of the following procedures was conducted;

3.2.2.3 Frozen Soil (NS)

Three level spatulas (approx. 5 g) of homogenised soil was transferred an empty 50 mL falcon tube. The falcon tube containing the soil sample was flash frozen, by submersion in liquid nitrogen and transported on dry ice to a -80°C freezer. Before RNA extraction, the samples were combined and ground using a mortar and pestle in liquid N_2. After homogenisation, 2 g of sample material was used for RNA extraction in triplicate.

3.2.2.4 Lifeguard Soil (LGS)

Three level spatulas (approx. 5 g) of soil were added to a 50 mL falcon tube, containing 30 mL of Lifeguard solution. The tube was shaken for 45 s to allow penetration of the solution into the sample material. The sample was transferred on ice to -20°C. Before RNA extraction the sample was defrosted at 4 ^{o}C overnight and then vortexed for 15 s and centrifuged at 4000 rpm for 10 min at 4^{o}C. The Lifeguard solution supernatant was discarded leaving behind a soil pellet. Wet weight of soil was measured and 1 mL of fresh Lifeguard solution was added per g of soil and was mixed by vortexing, the samples were then combined and mixed. Four millilitres (approx. 2 g soil) of the soil-lifeguard slurry was transferred to the bead tube of the RNA extraction kit in triplicate. Tubes were spun for 1 min at 12000 rpm in a microcentrifuge to pellet the soil, and the Lifeguard solution supernatant was discarded prior to RNA extraction from the soil pellet.

3.2.2.5 Thawed soil (NLGS)

Three level spatulas (approx. 5 g) of homogenised soil was transferred an empty 50 mL falcon tube. The falcon tube containing the soil sample was flash frozen, by submersion in liquid nitrogen and transported on dry ice to a -80°C freezer. Before RNA extraction 30 mL of Lifeguard solution was added to the frozen sample and shaken. The sample was allowed to defrost over 20 minutes, with light shaking every minute. Once thawed the

soil was vortexed for 15 s and then centrifuged at 4000 rpm for 10 min at 4°C. The Lifeguard solution supernatant was discarded leaving behind a soil pellet. Wet weight of soil was measured and 1 mL of fresh Lifeguard solution was added per gram of soil and was mixed by vortexing and then the samples were combined and mixed. Four millilitres (approx. 2 g soil) of the soil-lifeguard slurry was transferred to the bead tube of the RNA extraction kit in triplicate. Tubes were spun for 1 min at 12000 rpm in a microcentrifuge to pellet the soil, and the Lifeguard solution supernatant was discarded prior to RNA extraction from the soil pellet.

3.2.2.6 Freeze-dried soil (WS)

Three level spatulas (approx. 5 g) of homogenised soil were added to a falcon tube, containing 30 mL of sterile distilled water and shaken for 45 s. Washing of soil samples was conducted in order to simulate the washing treatment for rhizosphere soil. The sample was snap frozen in liquid nitrogen and stored on dry ice prior to storage at -80°C. The sample was removed from the freezer and lyophilised using an 'Alpha 1-2 LD plus' freeze drier (Martin Christ, Germany) until all moisture had been removed. Samples were combined and 2 g of freeze-dried soil was used for RNA extraction in triplicate.

3.2.2.7 Ribodepleted frozen soil (RiboNS)

After quantification and quality checking of RNA, 5 μg RNA of each of the NS samples were ribodepleted using a 50:50 combination of the RiboZero Plant/Seed (Illumina, USA) and RiboZero soil kits (Illumina, USA) according to the manufacturer's instructions. RNA concentration was quantified using the Qubit Fluorometer RNA HS (ThermoFisher, USA). RNA quality was assessed using a Bioanalyzer prokaryotic pico chip (Agilent Technologies).

3.2.3 Rhizosphere and root collection.

For each pseudoreplicate, four *B. napus* plants were removed from the soil and lightly shaken to remove loosely adhering soil. The lateral roots were subsequently cut into <5 cm length pieces and pooled and sorted into even sized groups. From this pseuodreplicate one of each of the following procedures was conducted;

3.2.3.1 Lifeguard rhizosphere (LGR)

Roots with adhering soil were transferred to a 50 mL falcon tube containing 25 mL of Lifeguard solution (wash 1) and shaken for 45 s. The sample was transferred on ice to -20°C. Before RNA extraction the sample was defrosted at 4 ^{o}C overnight and then vortexed for 5 s and the roots were transferred to a fresh 50 mL falcon tube containing 15 mL of Lifeguard solution (wash 2). The tube was vortexed for 5 s and roots were transferred to a fresh 50 mL falcon tube containing 30 mL of Lifeguard solution (wash 3) and vortexed for 5 s to completely clean the roots. Roots were transferred to a fresh, empty 50 mL falcon tube, flash frozen and stored at -80°C until use as Lifeguard root (LGRO) samples. Root wash solutions (rhizosphere soil) from wash 1 and 2 were combined and centrifuged at 4000 rpm for 10 min at 4 ^{o}C. The Lifeguard solution supernatant was discarded leaving behind a soil pellet. Wet weight of soil was measured and 1 mL of fresh Lifeguard solution was added per g of soil and was mixed by vortexing and then the samples were combined and mixed. Four millilitres (approx. 2 g soil) of the soil-lifeguard slurry was transferred to the bead tube of the RNA extraction kit in triplicate. Tubes were spun for 1 min at 12000 rpm in a microcentrifuge to pellet the soil, and the Lifeguard solution supernatant was discarded prior to RNA extraction from the soil pellet.

3.2.3.2 Lifeguard root (LGRO)

Root samples from the previously described LGR treatment were combined and ground in liquid nitrogen using a mortar and pestle. Two grams of powdered roots were used for RNA extraction in triplicate.

3.2.3.3 Freeze-dried rhizosphere (WR)

Root material with adhering soil was transferred to a falcon tube containing 25 mL of sterile distilled water and shaken for 25 s. Roots were then transferred to a fresh tube containing 10 mL of SDW and shaken for 20 s. Clean roots were transferred to a fresh tube, flash frozen in liquid nitrogen and stored on dry ice until transfer to a -80^{o}C freezer for use as the Washed frozen root (WRO) material. Root washes (rhizosphere soil) were combined, flash frozen and stored on dry ice until transfer to a -80°C freezer. The rhizosphere sample was then removed from the freezer and lyophilised using an 'Alpha 1-2 LD plus' freeze drier (Martin Christ, Germany) until all moisture had been removed. Samples were combined and 2 g of freeze-dried soil were used for RNA extraction in triplicate.

3.2.3.4 Washed frozen root (WRO)

Root samples (from WR) were removed from -80°C. The samples were combined and ground in liquid nitrogen. Two grams of powdered roots were used for RNA extraction in triplicate.

3.2.4 Quality control

After extraction of total RNA, samples were treated with DNAse enzyme according to the manufacturer's instruction (DNase Max™ Kit ,MO BIO Catalog# UC-15200-50). Samples were purified using Agencourt RNAClean™ according to the manufacturer's instruction (001298v001

Beckman Coulter). The presence of remaining DNA was tested by PCR The reaction mixture comprised forward primer 16S 515f (10 µM)1.25 µl, Reverse primer 16S 806r (10 µM) 1.25 µl, 2 x Q5 readymix 12.5 µl, DEPC treated water 9 µl, 1 µl DNase-treated RNA, total volume of 25 µl. Any samples with visible bands were reprocessed.

Total RNA concentrations were quantified using the Qubit Flurometer RNA HS (ThermoFisher, USA). RNA quality was assessed using a Bioanalyzer prokaryotic nano chip (Agilent Technologies).

3.2.5 Library preparation

An adapted Illumina Truseq library preparation was conducted at the Earlham Insitute (Earlham Institute, Norwich, UK) prior to sequencing using the following procedure: Sequencing libraries were constructed using an adapted TruSeq RNA protocol **(Illumina 15026495 Rev.B).** The library preparation involved QC of the depleted RNA using Bioanalyser with the Pico kit **(Agilent 5067-1513)** on the mRNA setting to detect any potential rRNA contamination and a nano chip for the total RNA to detect the RIN score. The ribo-depleted RNA and total RNA was chemically fragmented and first strand cDNA was synthesised. The ends of the samples were repaired using the 3' to 5' exonuclease activity to remove the 3' overhangs and the polymerase activity to fill in the 5' overhangs creating blunt ends. A single ‘A’ nucleotide was added to the 3’ ends of the blunt fragments to prevent them from ligating to one another during the adapter ligation reaction. A corresponding single ‘T’ nucleotide on the 3’ end of the adapter provided a complementary overhang for ligating the adapter to the fragment. This strategy ensured a low rate of chimera formation. The ligation of a number of indexing adapters to the ends of the DNA fragments prepared them for hybridisation onto a flow cell. The ligated products were subjected to a bead based size selection using Beckman Coulter XP beads **(Beckman Coulter A63880).** This removed the majority of un-ligated adapters, as well as any adapters that may have

ligated to one another. Prior to hybridisation to the flow cell the samples were amplified by PCR to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. The PCR was performed with a proprietary PCR primer cocktail that annealed to the ends of the adapter. The insert size of the libraries was verified by running an aliquot of the DNA library on a PerkinElmer GX using the High Sensitivity DNA chip **(PerkinElmer CLS760672)** and the concentration was determined by using a High Sensitivity Qubit assay and q-PCR.

After library preparation, the samples were pooled and divided evenly over three lanes of the Illumina HiSeq 2500 and sequenced with a 126bp paired-end read metric. Samples were split evenly over three lanes to minimise the impact of lane variation on samples and maximise the depth of sequencing that may have been reduced due to differences in sample type influencing number of sequences per sample.

3.2.6 Bioinformatics pipeline.

Firstly samples were reassembled (from the 3 lane split). The samples were briefly checked using FASTQC and the 5` base was trimmed using trimmomatic (Bolger *et al*., 2014) due to low quality in all samples. Forward and reverse reads were merged prior to SortmeRNA (Kopylova *et al*., 2012) separation of mRNA and rRNA.

3.2.6.1 rRNA analysis

rRNA was trimmed using trimmomatic with the quality control settings of sliding window 4:15, minimum length of 100bp and leading and trailing of 3. Libraries were merged using the multiple_split_libraries_fastq.py script in QIIME. An OTU table was generated from the merged fasta file using pick_closed_reference_otus.py in QIIME. Prokaryotic (16S rRNA) taxonomies were assigned using the Greengenes 97% 13.8 closed

reference database (DeSantis *et al*., 2006) and eukaryotic (18S rRNA) taxonomies were assigned using the Silva 119 database (Quast *et al*., 2013). After assignment the 18S rRNA dataset was filtered to remove plant and animal sequences using the “Metazoa” and “Charophyta” flags and manually searched to remove any plant or animal OTUs missed by these flags.

The microbial community compositions and functional profiles of the soil microbiomes were compared for alpha diversity, beta diversity and community composition. Fisher’s alpha diversity scores were generated in R using the estimate_richness script from the “phyloseq” package (McMurdie and Holmes, 2013). Significant differences overall were calculated using an ANOVA script from the “stats” package (R Core Team, 2013). Post-hoc pairwise comparisons were calculated using the HSD test from the “agricolae” package (de Mendiburu, 2017). Beta diversity was assessed using Bray-Curtis similarity matrices, NMDS, Hierarchical clustering and SIMPROF tests were conducted in order to determine significant differences in beta diversity. NMDS plots and SIMPROF tests were generated in Primer6 (Clarke and Gorley, 2006.) Differences in community composition were tested using Tukeys’s HSD test and visualised using the phyloseq and ggplot (Wickham, 2009) packages. Drivers of differences in community composition between methods were calculated with a White’s non-parametric t-test, with Bonferrroni multiple corrections and visualised in STAMP (Parks *et al*., 2014).

3.2.6.2 mRNA analysis

Trimmed mRNA reads were aligned to the NCBI non-redundant (NR) protein database (retrieved on 22/08/17) (Pruitt *et al*., 2007) using Diamond blastx (Buchfink *et al*., 2015) using the default settings; BLOSUM62 matrix, gap open penalty 11, gap extension penalty 1 and a minimum e value of 0.0001. The 25 top results were kept. The resulting .daa files were converted into blast tabular format (.m8) using DIAMOND

view. In order to separate plant reads, accessions with taxonomy containing embryophyta (green land plants) were extracted from the NCBI protein database using the Entrez API (Sayers, 2008).
Alignments with an accession number matching those on the embryophyta list were separated from other reads using a custom perl script.

SEED function were then assigned to each read using MEGAN6 community edition (Huson *et al*., 2016) using the following parameters: minimum alignment score 50, maximum e value 0.01, minimum percent identity 0, minimum support (number of reads for a taxon to be counted) 1, and the naïve lowest common ancestor algorithm.
Megan files for each library were then loaded into MEGAN6 for comparison, and the assigned reads normalized against the smallest number of reads. The data was then exported to STAMP (Parks *et al*., 2014) where differences in relative abundance of the functional categories were tested for significance using Whites non-parametric t tests, with Bonferroni multiple corrections.

3.3 Results

3.3.1 RNA extraction

Total RNA was successfully extracted from all samples, with sufficient concentration and quality for RNA sequencing (Table 3.2 and Supplementary Figure 1).

Table 3.2 Concentration of RNA (ng/μl) after extraction and purification, measured by Qubit RNA high sensitivity assay, and RIN scores from bioanalyser nano prokarytotic chip.

Sample	RNA concentration (ng/ul)	RIN
WRO1	519	8.7
WRO2	1150	7.9
WRO3	908	7.9
LGRO1	412	7.7
LGRO2	1090	8.9
LGRO3	544	8.8
NS1	244	8.1
NS2	245	7.9
NS3	226	8.1
LGR1	29	8.1
LGR2	11	8.5
LGR3	10	8.3
LGS1	73	6.9
LGS2	104	6.7
LGS3	126	7.8
NLGS1	153	7.5
NLGS2	173	7.2
NLGS3	155	7.2
WS1	175	7
WS2	189	6.2
WS3	163	6.2

Over the three lanes 730,045,865 reads were generated. For the samples that had not been ribodepleted an average of 96% was ribosomal RNA. For the ribodepleted samples 6% of the total reads were attributed to ribosomal RNA.

3.3.2 Ribosomal RNA

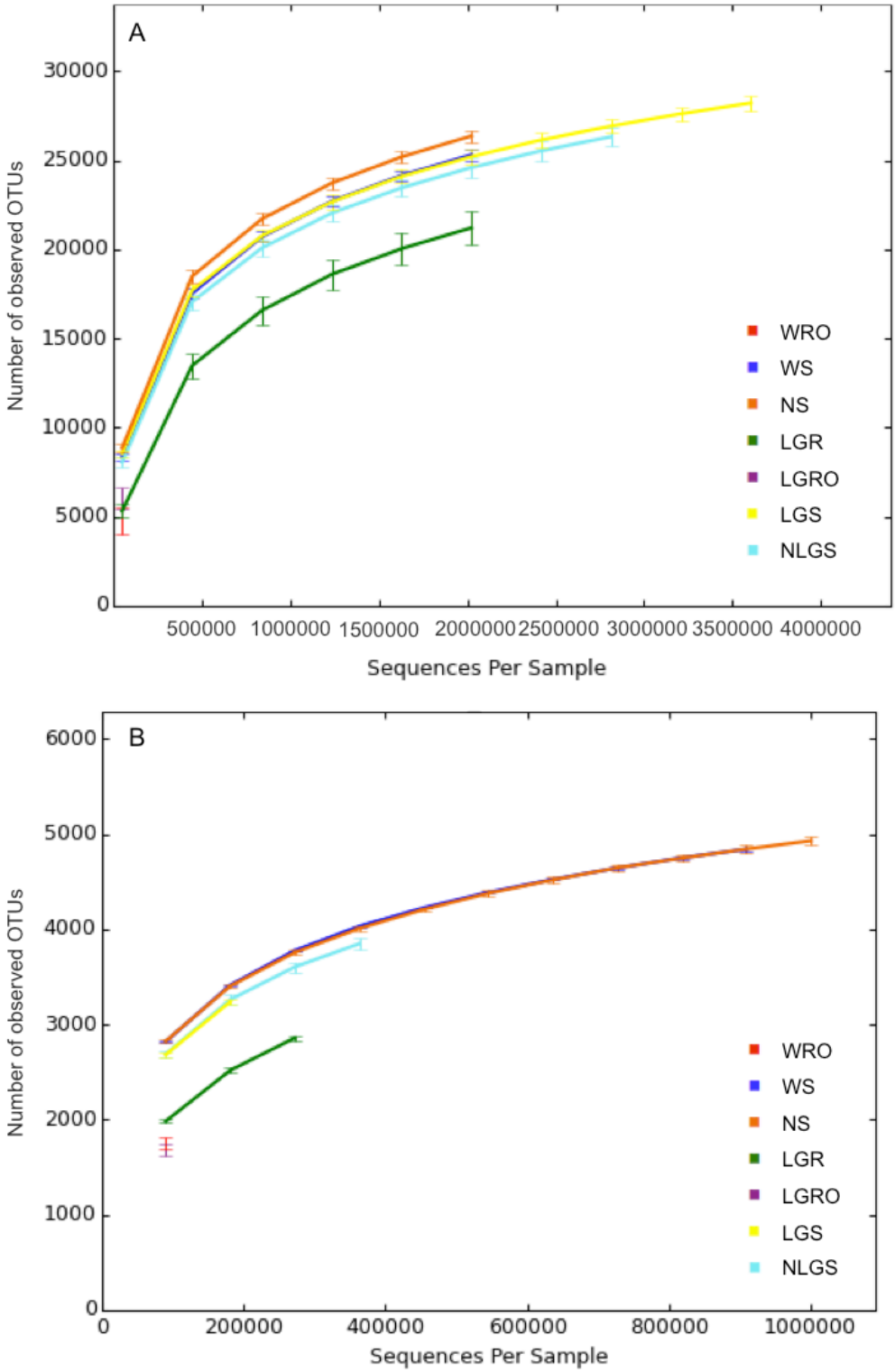


Figure 3.1 Collectors curves displaying number of observed OTUs at increasing sampling depths for 16S rRNA (A) and 18S rRNA (B) assigned taxonomies. Treatment ID's listed in Table 3.2.

In total 47,263,919 sequences were assigned to taxa using the 16S rRNA database across 24 samples. Using the 18S rRNA database, after filtering, 11,137,787 sequences were assigned to taxa. Datasets were subjected to a single random subsampling in order to account for variation in number of sequences per sample. The 16S rRNA dataset was normalised to 2,119,333 sequences per sample for the soils comparison and 53,716 sequences per sample for the compartment comparison. The 18S rRNA dataset was normalised to 245,893 sequences per sample for the soils comparison and 154,703 sequences per sample for the compartment comparison. Collector's curves were constructed to assess the proportion of the diversity that was sampled at a set depth (Figure 3.1). For each amplicon dataset, comparisons were made between 1.) Soil samples using the four different RNA preservation techniques and 2.) Root, rhizosphere and soil samples using the Lifeguard method.

3.3.3 The impact of RNA preservation technique on the diversity and composition of the bacterial soil microbiome.

Fisher's alpha diversity was not significantly different for the bacterial communities from the different RNA preservation treatments, ANOVA P=0.06 (Figure 3.2).

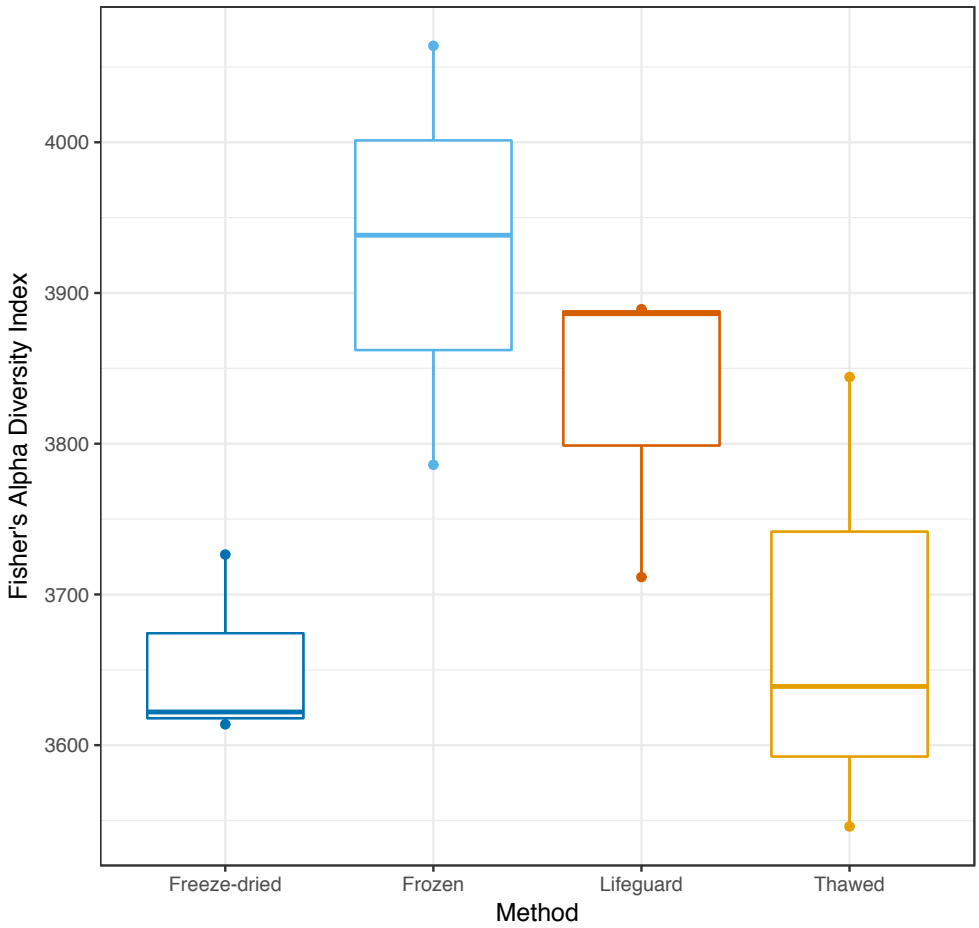


Figure 3.2 Box and whisker plot of Fisher's alpha diversity scores for the bacterial soil microbiome for each the RNA treatment methods; Freeze-dried, frozen, Lifeguard and thawed. No significant differences were identified in alpha diversity between treatments (ANOVA P=0.06). Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5 * of the interquartile range.

Hierarchical cluster analysis was conducted in order to assess beta diversity of samples (Figure 3.3). The bacterial communities of the technical replicates from the four RNA preservation treatments were clustered at 87%, 83%, 88% and 83% for the Lifeguard, Thawed, Freeze-dried and Frozen samples respectively. The Lifeguard and Thawed technical replicates clustered at 79% whereas the Freeze-dried and Frozen technical replicates clustered together at 75% similarity (Figure 3.3). From the cluster analysis and visual representation by NMDS (Figure 3.4) it is possible to conclude that the Freeze-dried technique displays a prokaryotic community most similar to that of the Frozen soil community.

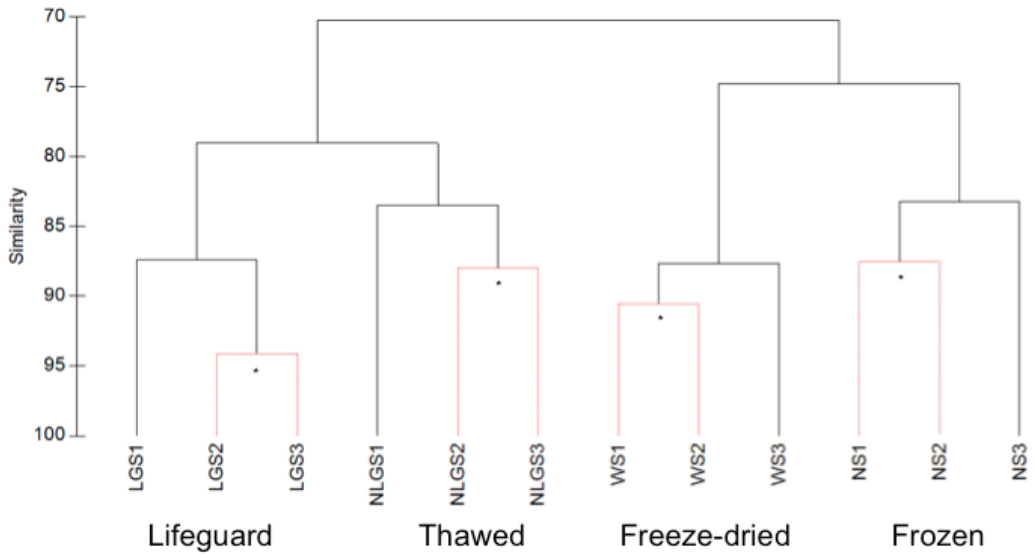


Figure 3.3 Hierarchical Cluster analysis displaying group average Bray Curtis similarity for bacterial soil communities using different RNA preservation methods. Red bars signify and * symbol indicate no significant difference between samples as tested by SIMPROF analysis. Where a branch splits on the y axis displays the percentage of similarity between samples.

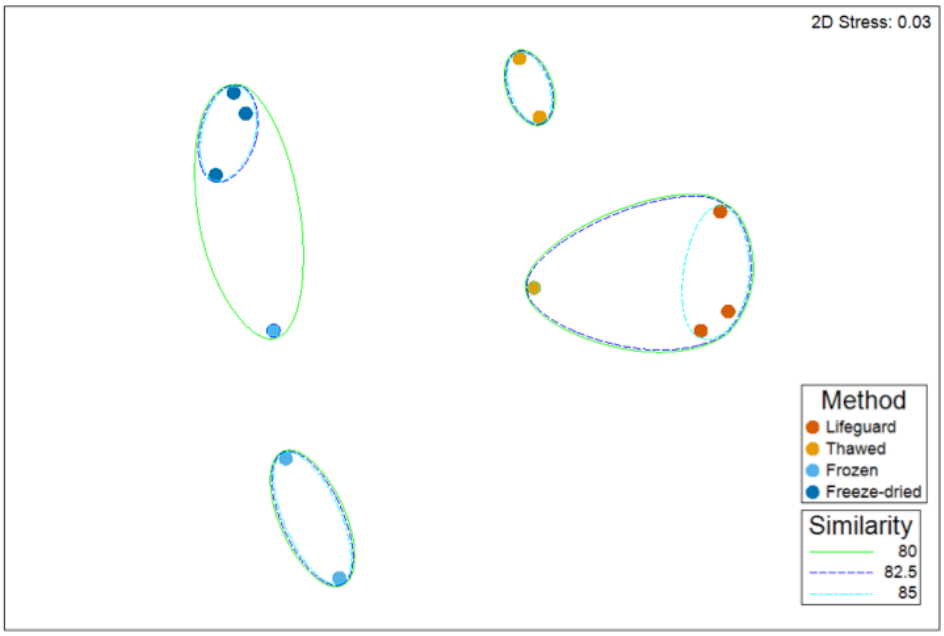
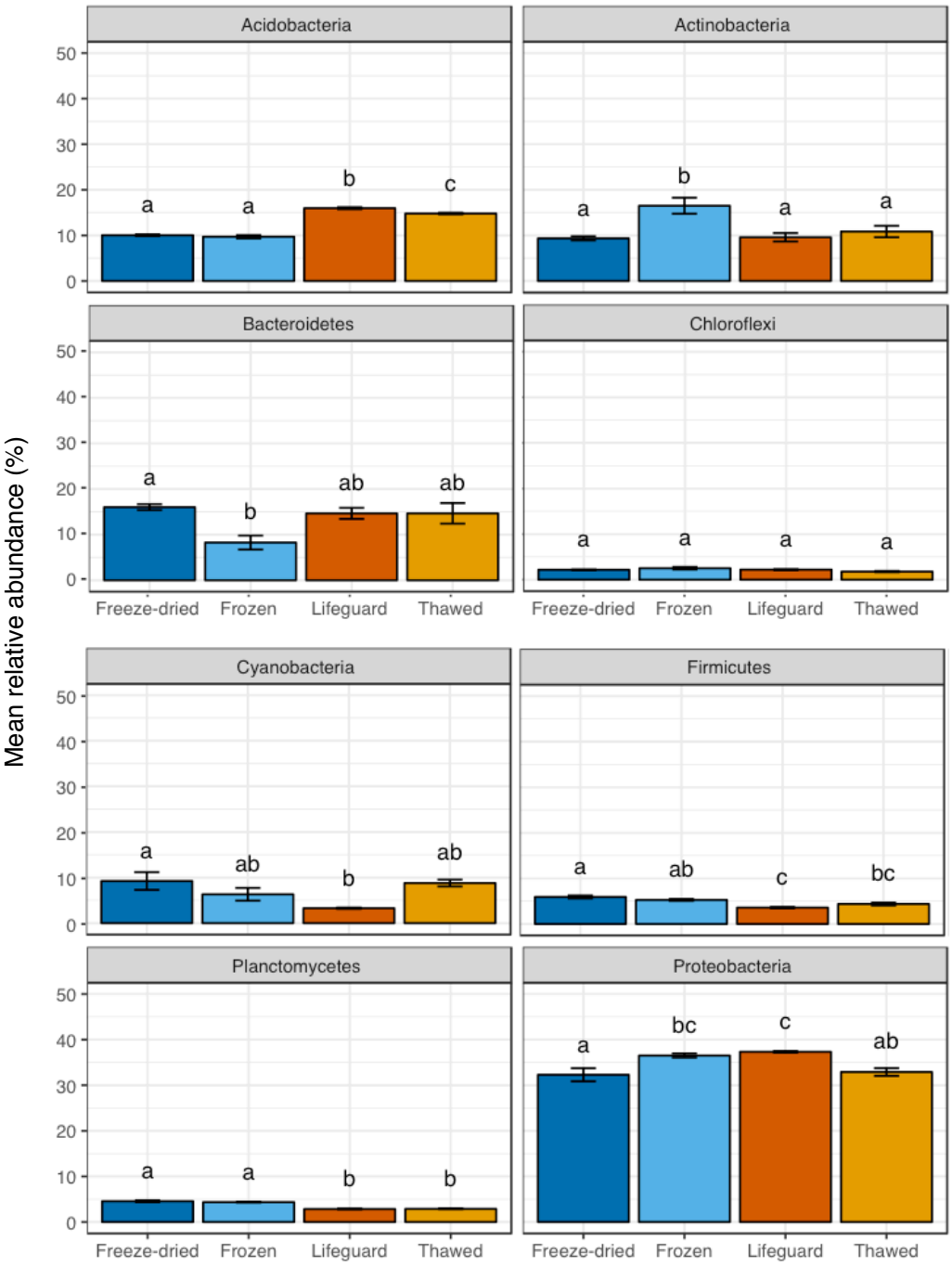


Figure 3.4 NMDS displaying Bray-Curtis similarity between soil sample 16S rRNA community composition under different preservation methods. Bubbles display similarity of samples at given percentage identity

At the highest taxonomic level (phylum) the method of preservation was seen to impact the relative abundance of 8 of the 9 most abundant phyla (Figure 3.5). The Freeze-dried method had the fewest significant ($P<0.05$) differences in relative abundance of bacterial phyla (three) compared to the Frozen samples (as tested by the Tukey's HSD test). The relative abundance of the Actinobacteria was 16.5% when using the frozen method and 9.4% under the Freeze-dried method. The relative abundance of the Proteobacteria was 36.5% in the Frozen method and 32.3% in the Freeze-dried. The relative abundance of the Bacteriodetes was significantly greater (16.0%) when using the Freeze-dried technique than when using the Frozen method (8.3%) (Figure 3.5).

The relative abundance of four Phyla were significantly different in the thawed soil compared to the Frozen soil. The relative abundance of the Acidobacteria was greater by 5.1% in the Thawed compared to the Frozen samples. In contrast the relative abundance of the Actinobacteria, Planctomycetes and Verrumicrobia were greater in the Frozen samples than the Thawed by 5.7%, 1.5% and 1.7% respectively. The relative abundance of five Phlya were significantly different in the Lifeguard soil relative to the Frozen soil. The Lifeguard samples contained 6.3% more Acidobacteria than the Frozen samples. In contrast, relative to Lifeguard samples, the Frozen samples contained higher relative abundances of Actinobacteria, Firmicutes, Planctomycetes and Verrumicrobia, indicated by enrichments of 7.0%, 1.7%, 1.5% and 2.2% respectively (Figure 3.5).


Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Cyanobacteria
Firmicutes
Planctomycetes
Proteobacteria
Mean relative abundance (%)
0
10
20
30
40
50
Freeze-dried
Frozen
Lifeguard
Thawed
a
ab
b
bc
c

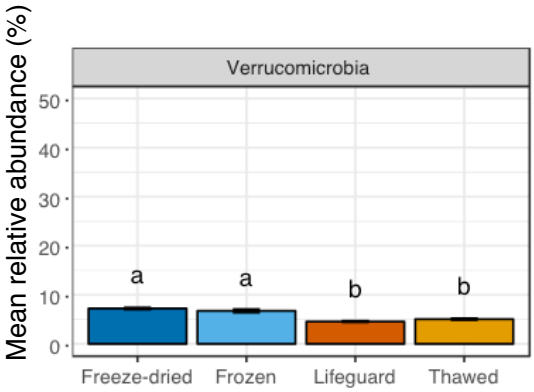


Figure 3.5 Facet bar plots displaying relative abundance of bacterial phyla in the soil for each RNA preservation method. Significance of difference in relative abundance between methods is indicated by Tukey's HSD test P<0.05, treatments with the same letter are not significantly different. Error bars represent standard error of the mean

Analysing the differences in community composition at a finer taxonomic classification (genus) revealed that there were no significant differences (White's non-parametric T-test , P< 0.05, Bonferroni multiple corrections) in the relative abundance of any taxa between the Frozen and Freeze-dried methods. The relative abundance of the *Bradyrhizobium* genus was significantly greater in the Frozen (1.87%) than the Thawed samples (0.96%). The relative abundance of an OTU from the Comamonadaceae family was a significantly greater in the Lifeguard (3.13%) than the Frozen (1.12%) samples.

3.3.4 The impact of RNA preservation technique on the measurement of diversity and composition of the microbial eukaryotic soil microbiome

Fisher's alpha diversity of the microbial eukaryotic community was significantly (P<0.05) greater in the Frozen and Freeze-dried soil samples than both the Thawed and Lifeguard samples when tested using a Tukey's HSD test (Figure 3.6).

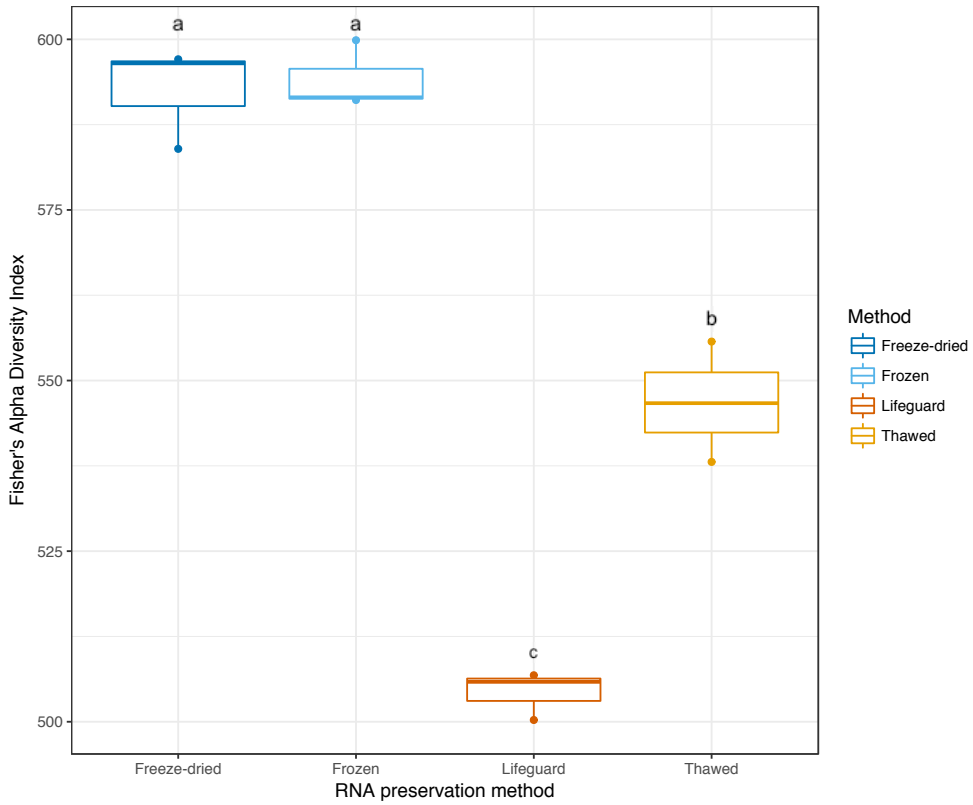


Figure 3.6 Fisher's alpha diversity scores for the microbial eukaryotic soil microbiome for each of the soil RNA treatment methods. Significant differences were identified between treatments (ANOVA P<0.05), treatments with the same letter are not significantly different. Pairwise differences between alpha diversity displayed using letters, groupings attributed by Tukey's HSD test. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range.

Differences in microbial eukaryotic community composition for the different RNA preservations method was assessed using a hierarchical clustering dendogram (Figure 3.7) Technical replicates clustered at 87% , 84%, 90% and 90% similarity for Freeze-dried, Frozen, Thawed and Lifeguard samples respectively. The Freeze-dried method clustered with the Frozen method at 82% similarity. Lifeguard and Thawed samples clustered at 77%. From the cluster analysis and NMDS (Figure 3.8) it is possible to conclude that the Freeze-dried technique displays a eukaryotic community most similar to that of the Frozen soil community.

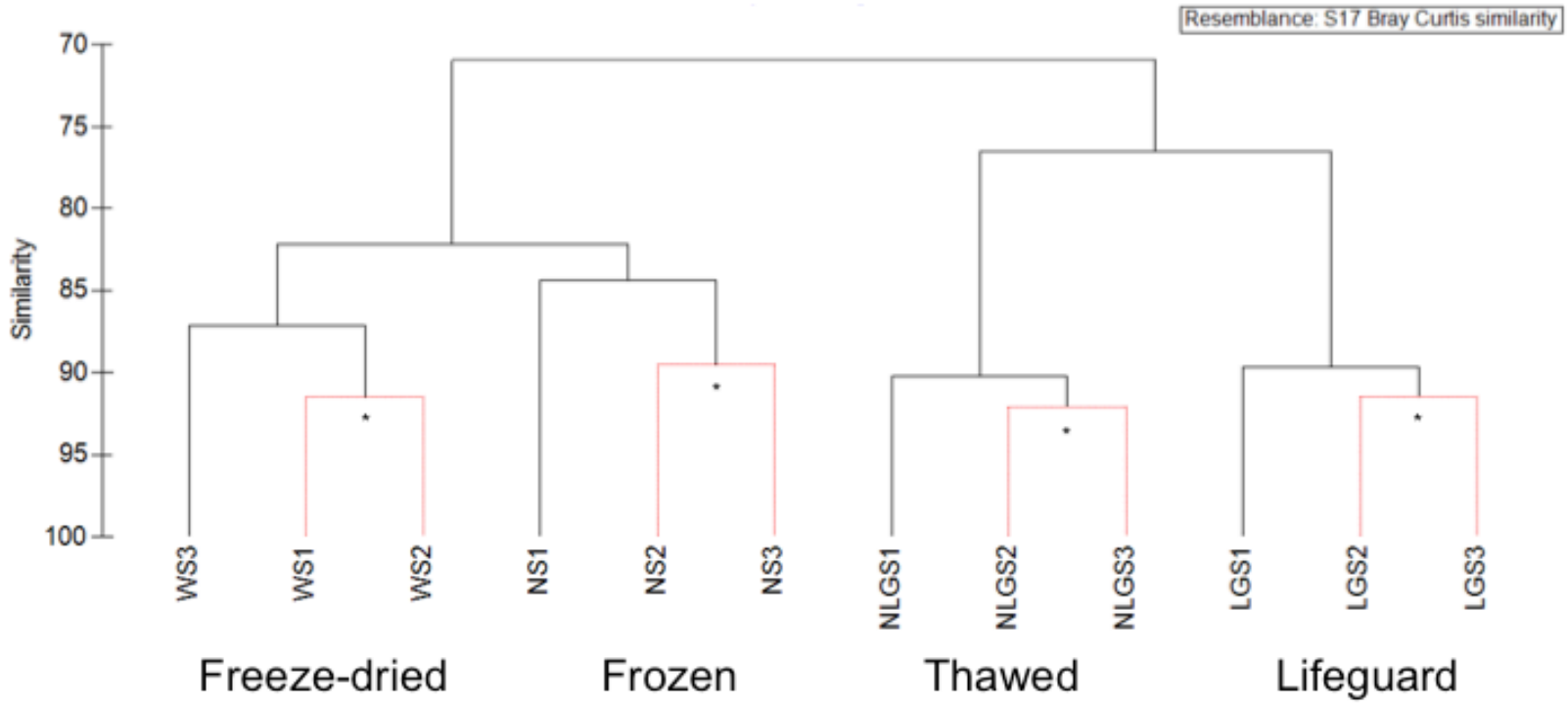


Figure 3.7 Hierarchical Cluster analysis displaying Bray Curtis similarity of microbial eukaryotic soil communities . Red bars signify no significant different between samples as identified by SIMPROF test. SIMPROF analysis is overlaid onto the figure displayed and red bars, which indicate no significant difference between samples. Where a branch splits on the y axis displays the percentage of similarity between samples.

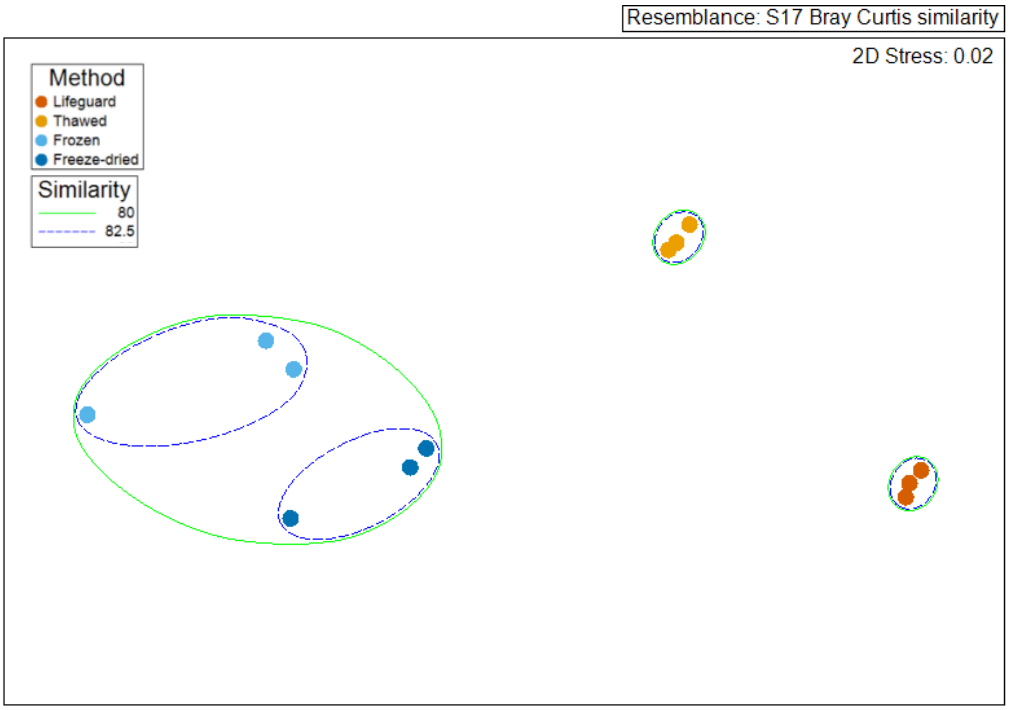
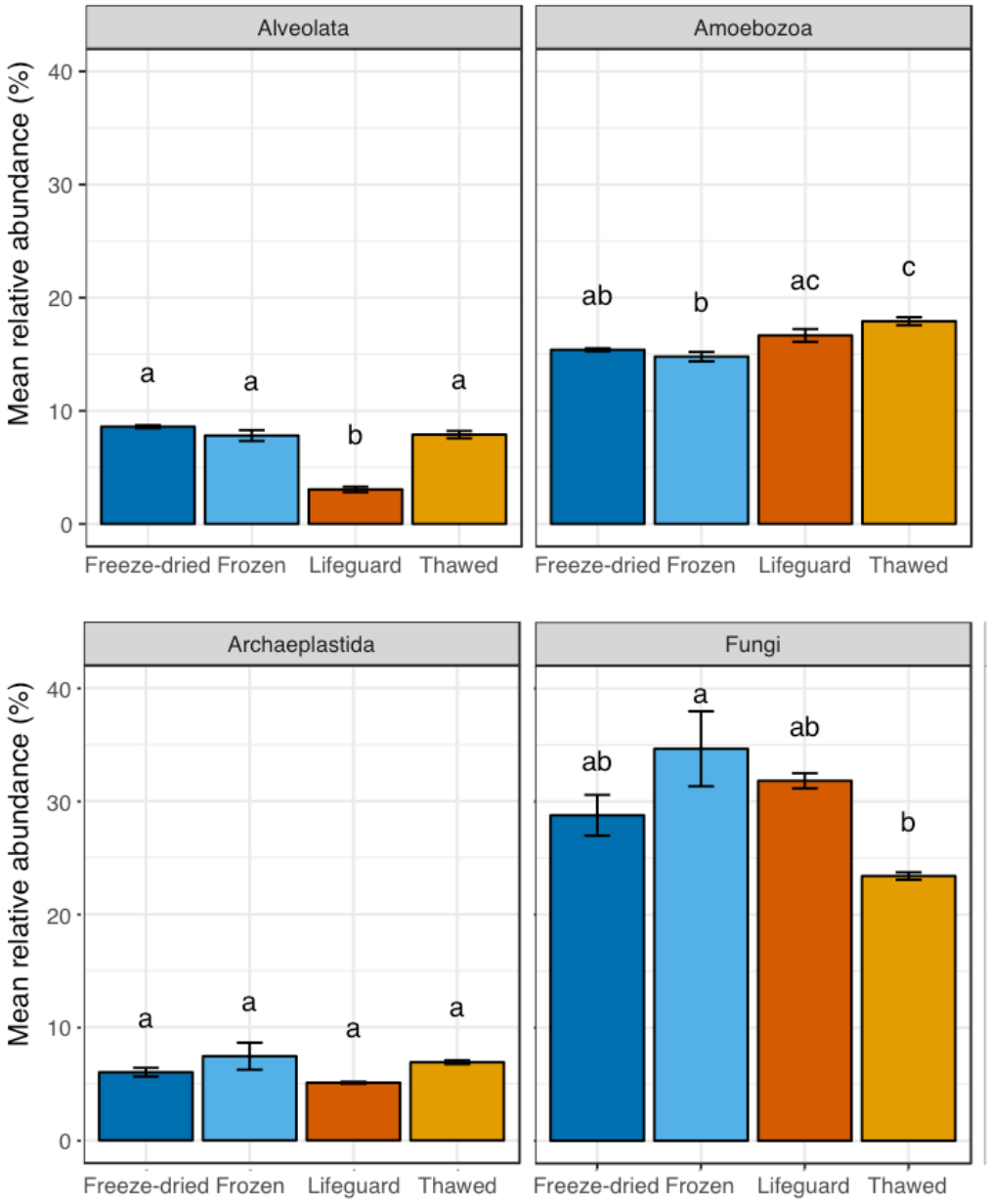


Figure 3.8 NMDS displaying Bray-Curtis similarity between microbial eukaryotic soil community under different preservation methods. Bubbles display similarity of samples at given percentage identity.

RNA preservation method had significant impacts on the microbial 18S rRNA community composition at a broad taxonomic level, for the 6 main taxonomic groups of microbial eukaryotes (Figure 3.9). The community composition of the Frozen treatment was significantly different from the Freeze-dried method for one of the six main groups of taxa. The relative abundance of the Rhizaria was significantly (Tukey's HSD, $P<0.05$) greater in the Freeze-dried samples, relative to the other treatments by an average of 1.68%. The community composition of the Frozen treatment was significantly different from both the Lifeguard and Thawed treatments for three of the six main microbial eukaryotic taxonomic groups. The relative abundance of the Alveolata was significantly greater in the Frozen samples than the Lifeguard samples by 4.76%. The relative abundance of the Amoebozoa and Stramenopiles was significantly greater in the Lifeguard samples relative to the Frozen samples, by 1.86% and 6.12% respectively. The relative abundance of the Amoebozoa was significantly greater in the Frozen samples than the Thawed samples by 3.08%. The relative abundance of the Fungi and Stramenopiles was significantly

greater in the Thawed samples, than the Frozen samples by 11.14% and 6.28% respectively.



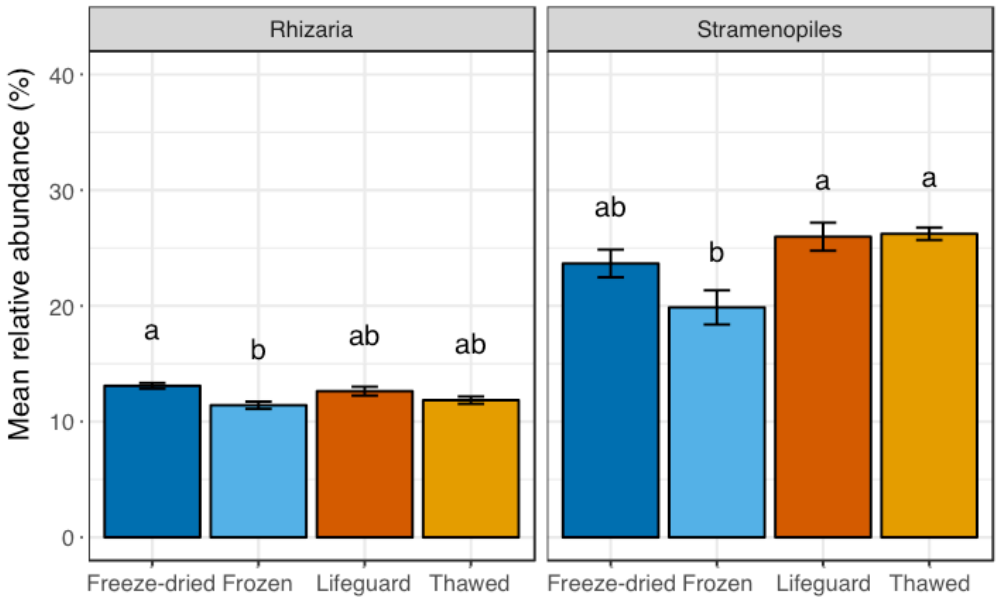


Figure 3.9 Facet bar plots displaying relative abundance of a eukaryotic group in the soil for each RNA preservation method. Significance of difference in relative abundance between methods as indicated by Tukey's HSD test P<0.05, treatments with the same letter are not significantly different. Error bars display standard error of the mean

Analysing the differences in community composition at a finer taxonomic classification (genus) revealed significant differences (White's non-parametric T-test , P< 0.05, Bonferroni multiple corrections) in the relative abundance of taxa between treatment methods. The relative abundance of an OTU from the LEMD255 phylum was significantly greater in the Freeze-dried (1.16%) than the Frozen samples (0.84%). The relative abundance of an OTU from the Zoopagales order was significantly greater in the Thawed (1.13%) than the Frozen samples (0.56%). The relative abundance of an OTU *Acanthamoeba* sp. EFW11 was a significantly greater in the Lifeguard (1.38%) than the Frozen (0.32%) samples.

3.3.5 *The active bacterial communities of the B. napus root, rhizosphere and bulk soil*

The Fisher's alpha diversity of the bacterial community in the soil samples was approximately two fold greater than the root and rhizosphere (Figure 3.10). Tukey's HSD test indicated that the diversity was significantly greater in the soil samples compared to the root and rhizosphere samples, while there was no significant difference in Fisher's alpha diversity between bacterial communities in the root and rhizosphere compartments.

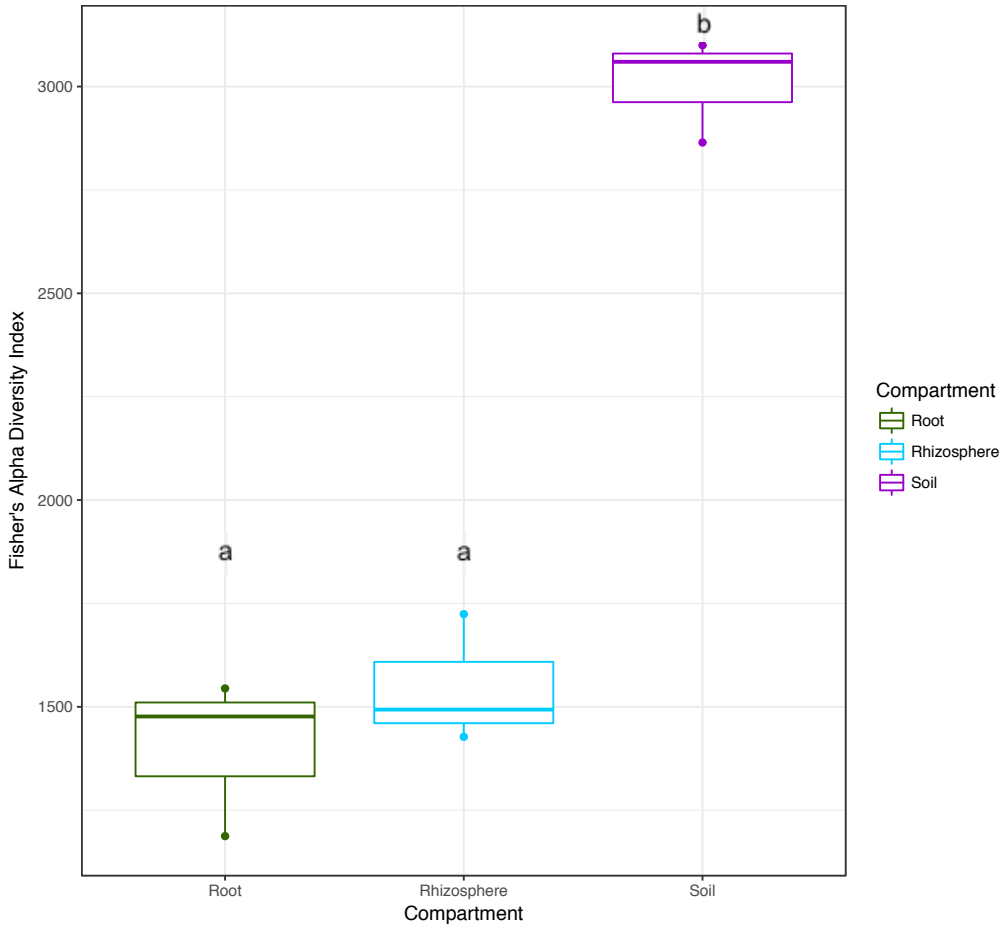


Figure 3.10 Fisher's alpha diversity scores for bacterial communities of the root, rhizosphere and soil compartments. Significant differences identified between treatments (ANOVA P<0.05), groupings identified by Tukey's HSD test, treatments with the same letter are not significantly different. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range.

Hierarchical cluster analysis revealed clustering of technical replicates at 68%, 82% and 76% for the root, rhizosphere and soil respectively. The rhizosphere and root communities displayed 49% similarity. The soil bacterial community was distinct from the rhizosphere and root communities at 29% similarity (Figure 3.11). The cluster analysis and NMDS (Figure 3.12) highlights the dissimilarity of the microbial communities of these compartments.

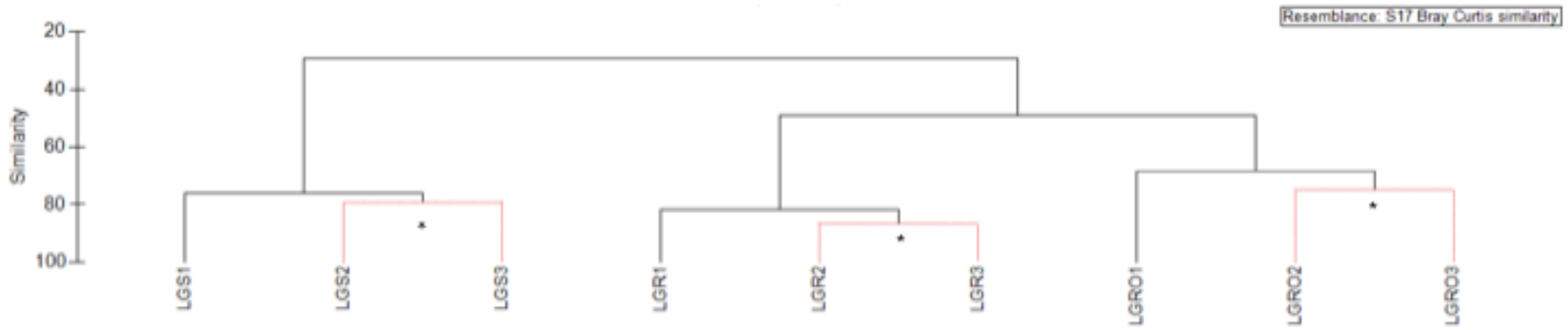


Figure 3.11 Hierarchical Cluster analysis displaying group average Bray Curtis similarity for bacterial soil, rhizosphere and root communities. Red bars signify and * symbol indicate no significant difference between samples as tested by SIMPROF analysis. Where a branch splits on the y axis displays the percentage of similarity between samples. Treatment ID's displayed in Table 1.

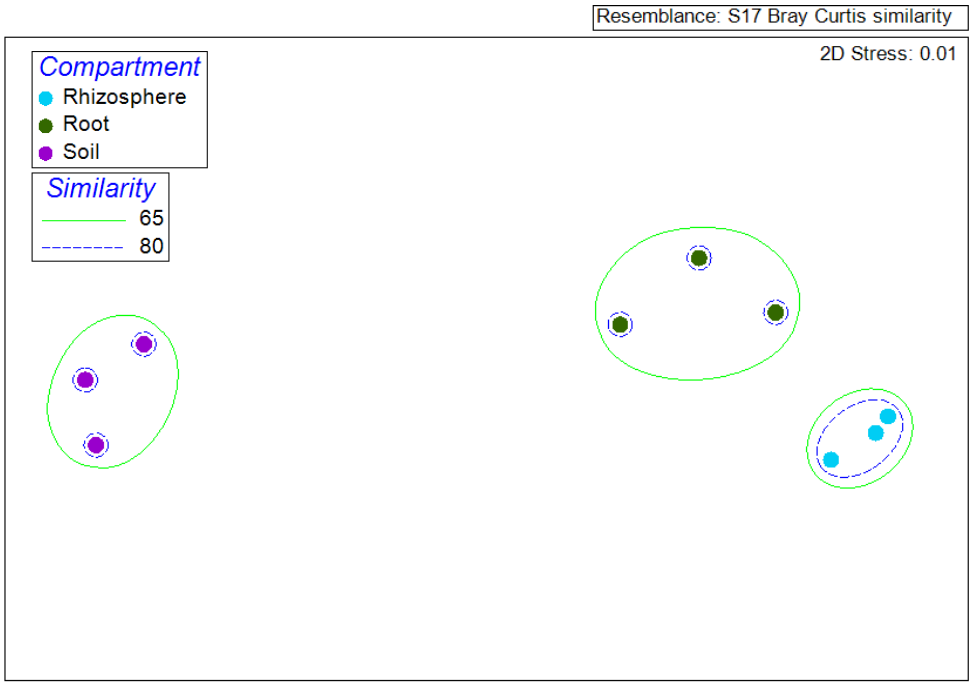
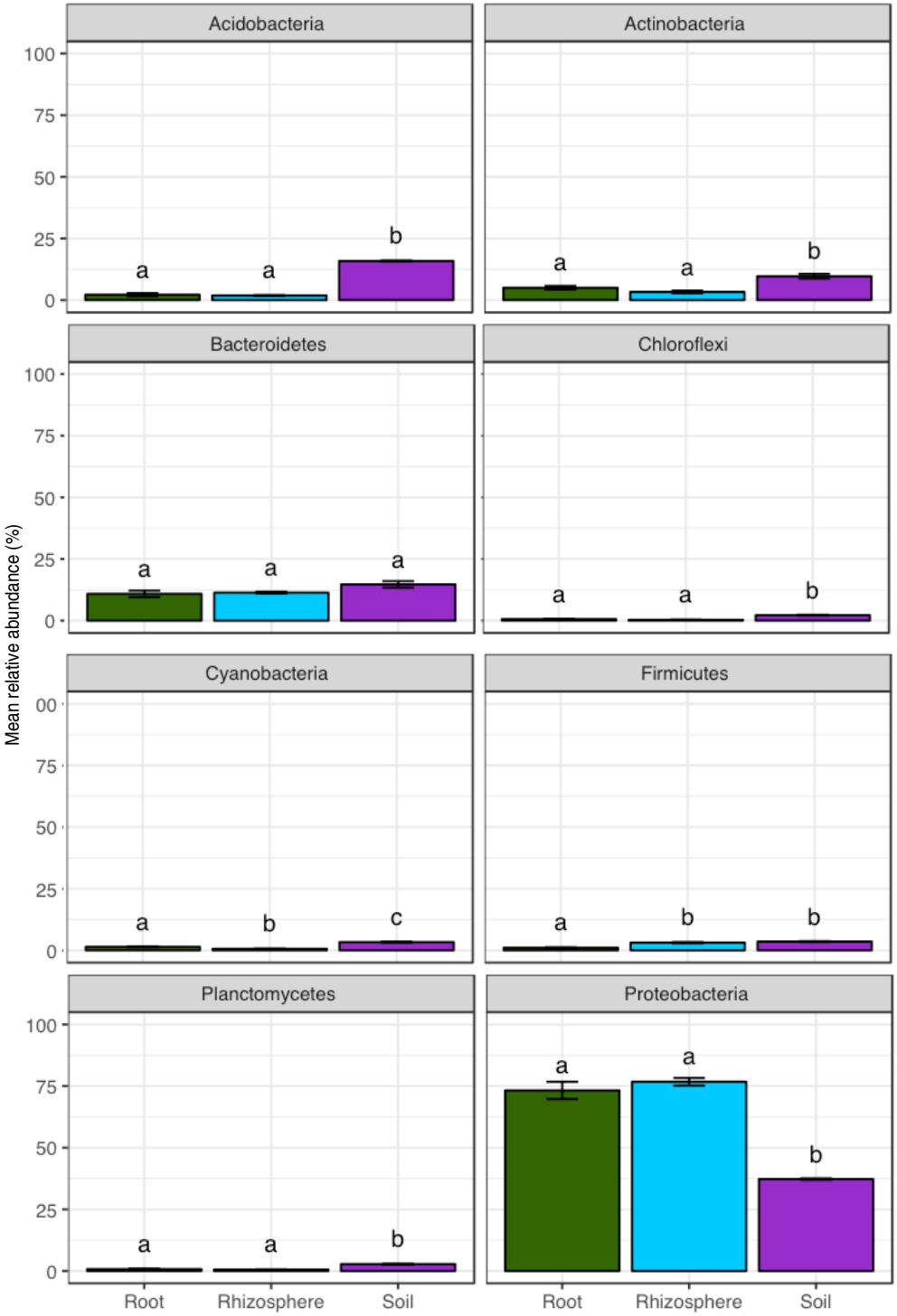


Figure 3.12 NMDS displaying Bray-Curtis similarity between root rhizosphere and soil 16S rRNA community composition. Bubbles display similarity of samples at given percentage identity.

At the Phylum level, there were significant (Tukey's HSD $P<0.05$) differences between root rhizosphere and soil communities (Figure 3.13). The root and rhizosphere compartments contained high relative abundances of Proteobacteria (73.2% and 76.7% respectively) which was significantly greater in these compartments than in the soil (37.3%) (Figure 3.13). The relative abundances in the soil of the Acidobacteria (15.9%), Actinobacteria (9.7%) Chloroflexi (2.2%) and Planctomycetes (2.8%) were all significantly greater than in the root and rhizosphere compartments. The relative abundance of the Cyanobacteria was significantly different across all samples; the relative abundance was greatest in the soil (3.3%), followed by root (1.4%) and then rhizosphere (0.6%). There was no significant difference in the relative abundance of the Bacteroidetes between the root (10.9%), rhizosphere (11.4%) and soil (14.7%) compartments. The relative abundance of Firmicutes in the soil (3.6%) and rhizosphere (3.1%) was significantly greater than in the root (1.1%). The relative abundance of the Verrumicrobia was significantly greater in the root (3.7%) and soil (4.6%) than the rhizosphere by (1.2%)


Acidobacteria
Actinobacteria
Bacteroidetes
Chloroflexi
Cyanobacteria
Firmicutes
Planctomycetes
Proteobacteria
Mean relative abundance (%)
100
75
50
25
0
a
b
c
Root
Rhizosphere
Soil

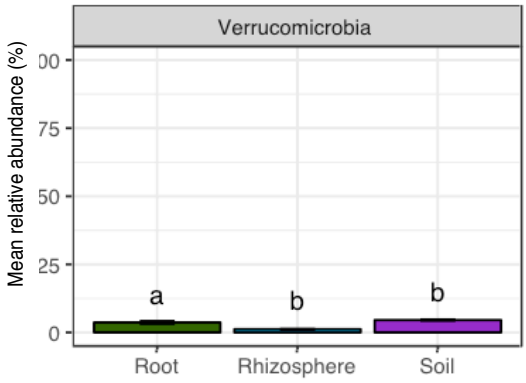


Figure 3.13 Facet bar plots displaying relative abundance of bacterial phyla in the root rhizosphere and soil. Significance of difference in relative abundance between methods as indicated by Tukey's HSD test P<0.05 treatments with the same letter are not significantly different. Error bars display standard error of the mean.

At a finer level, only one genus had significantly different relative abundance between each compartments after Bonferroni multiple corrections. The relative abundance of the *Serratia* genus was significantly greater in the rhizosphere (1.2%) than the root samples (0.18%), the relative abundance of *Cellvibrio* was significantly greater in the root (1.63%) than the soil (0.13%) and the relative abundance of *Janthinobacterium* was significantly greater in the rhizosphere (3.49%) than the soil (0.17%).

3.3.6 Differences in active microbial eukaryotic composition in the root, rhizosphere and soil

Tukey's HSD test indicated that the Fisher's alpha diversity of the microbial eukaryotic community was significantly greater in the soil samples compared to the root and rhizosphere samples. There was no significant difference in the Fisher's alpha diversity of the rhizosphere microbial eukaryotic community and the microbial eukaryotic root community (Figure 3.14).

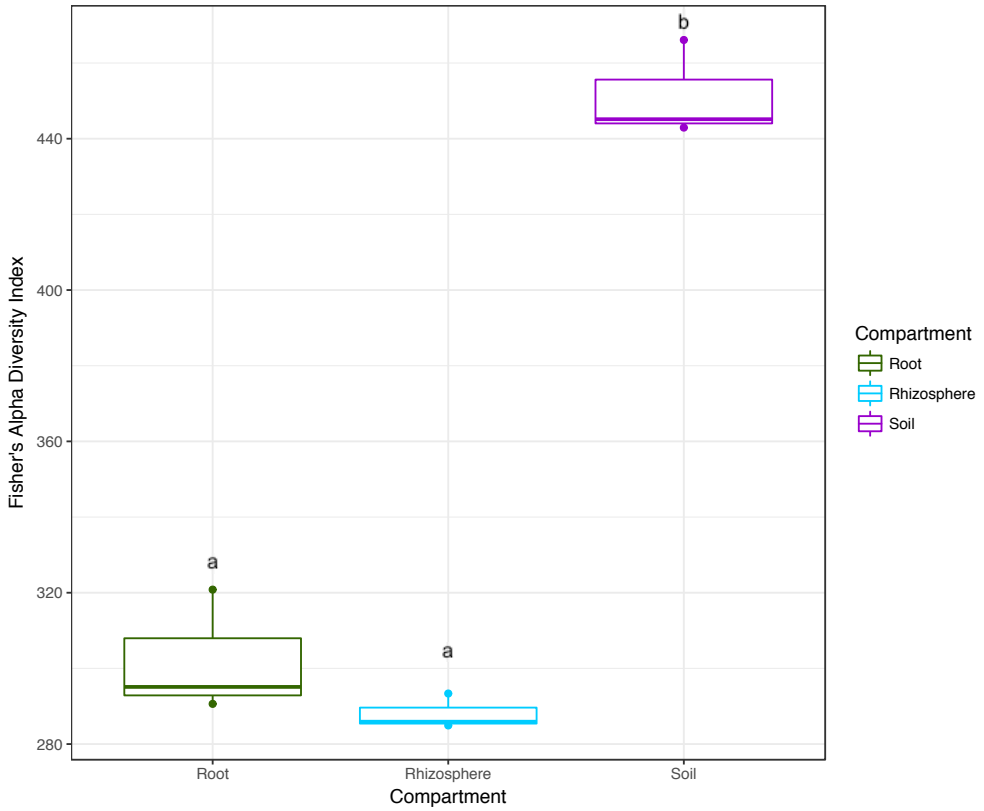


Figure 3.14 Fisher's alpha diversity scores for microbial eukaryotic communities of the root, rhizosphere and soil compartments. Significant differences identified between treatments ANOVA P<0.05. Groupings identified by Tukey's HSD test P<0.05, treatments with the same letter are not significantly different. Hinges are the 25th and 75th percentiles of the data. Whiskers range to the closest value within 1.5* of the interquartile range. Error bars display standard error of the mean

Beta diversity of the distinct root, rhizosphere and soil bacterial communities was revealed using hierarchical clustering dendogram (Figure 3.15). Technical replicates clustered closely together at similarity of 87% for the soil, 90% for the rhizosphere and 83% for the root. SIMPROF test identified no significant difference in the eukaryotic community composition between any of the three rhizosphere replicates. The rhizosphere and soil compartments clustered at 37% similarity. The microbial eukaryotic community of the root was clustered at 21% similarity to the rhizosphere soil and soil compartments, highlighting the value of separation of these samples. The hierarchical cluster analysis and NMDS (Figure 3.16) revealed distinct microbial eukaryotic communities for each of the compartments.

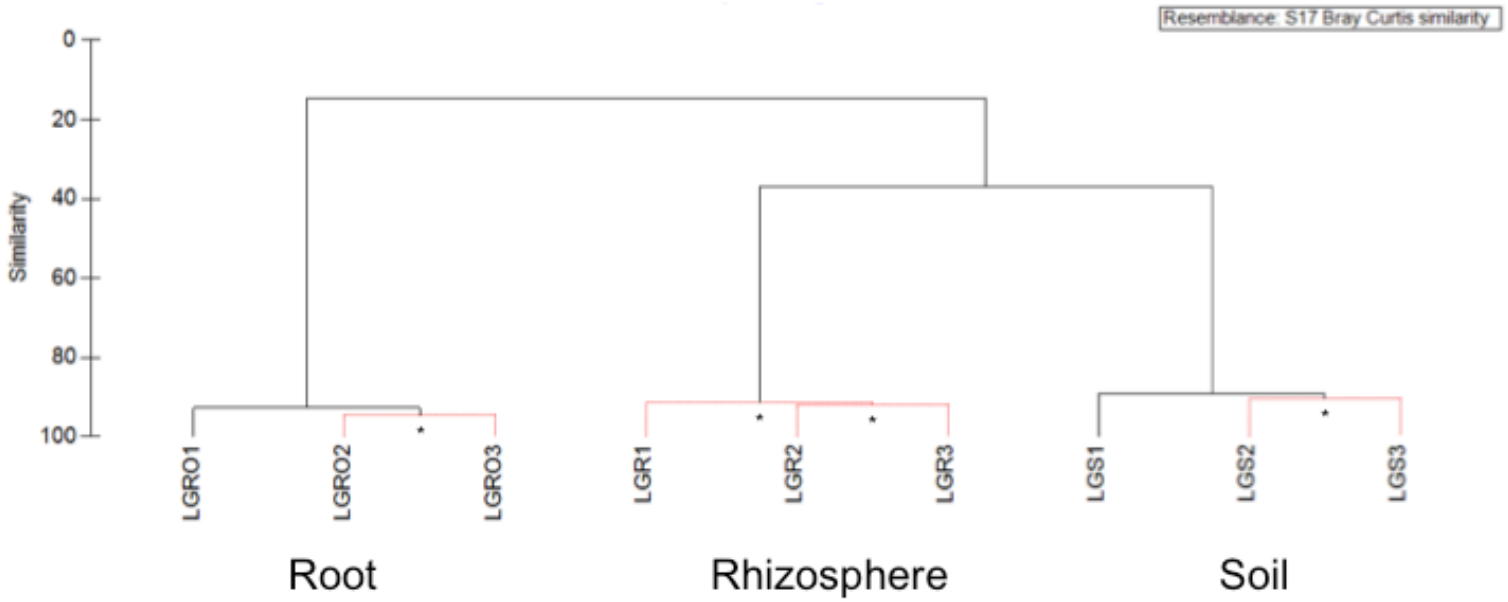


Figure 3.15 Hierarchical Cluster analysis displaying group average Bray Curtis similarity for microbial eukaryotic soil, rhizosphere and root communities. Red bars and * symbol indicate no significant difference between samples as tested by SIMPROF analysis. Where a branch splits on the y axis displays the percentage of similarity between samples.

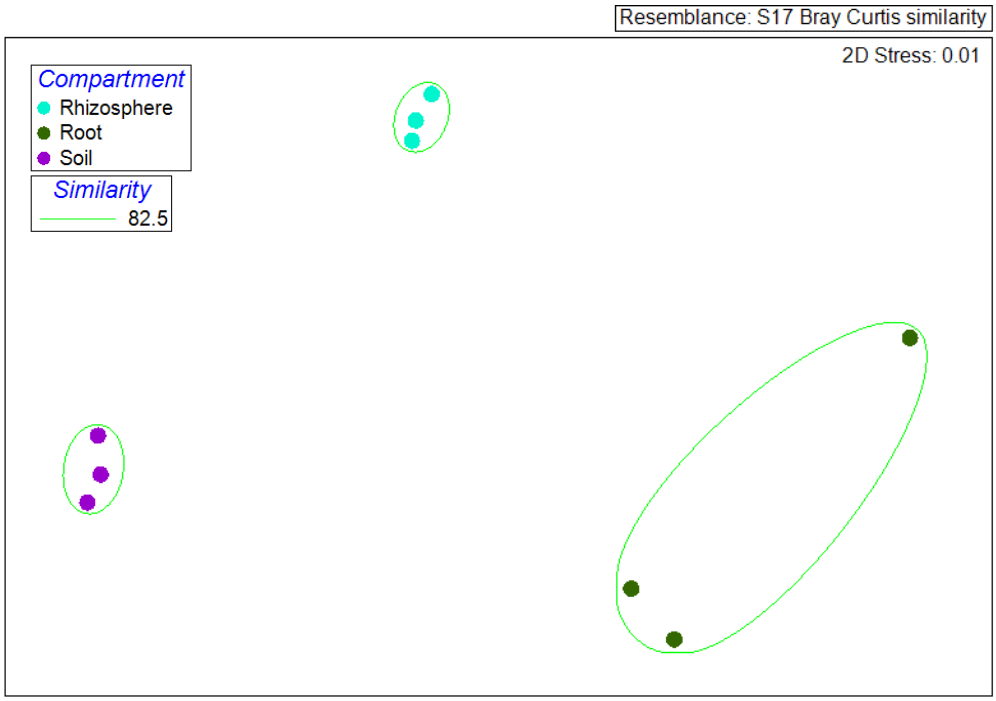


Figure 3.16 NMDS displaying Bray-Curtis similarity between root rhizosphere and soil microbial euakryotic community composition. Bubbles display similarity of samples at given percentage identity.

At the Phylum level, there were significant (Tukey's HSD P<0.05) differences between root rhizosphere and soil microbial eukaryotic communities (Figure 3.17).

The relative abundance of the Alveolata in the root (9.0%) was significantly greater than in the rhizosphere (1.0%) and soil (3.1%) (Figure 3.17). The relative abundance of the Amoebozoa was significantly different across all samples; the relative abundance was greatest in the soil (16.6%), followed by rhizosphere (11.5%) and then root (2.3%). The relative abundances of the Archaeplastida, Stramenopiles and Rhizaria were also significantly different across all samples; the relative abundances of these groups of taxa were greatest in the soil (5.0%, 26.0% and 12.5% respectively), followed by root (3.7%, 14.6% and 12.2% respectively) and rhizosphere (1.2%, 7.6% and 2.8% respectively). The relative abundance of the fungi was greatest in the rhizosphere (74.5%) followed by the root (48.4)% and soil (32%).

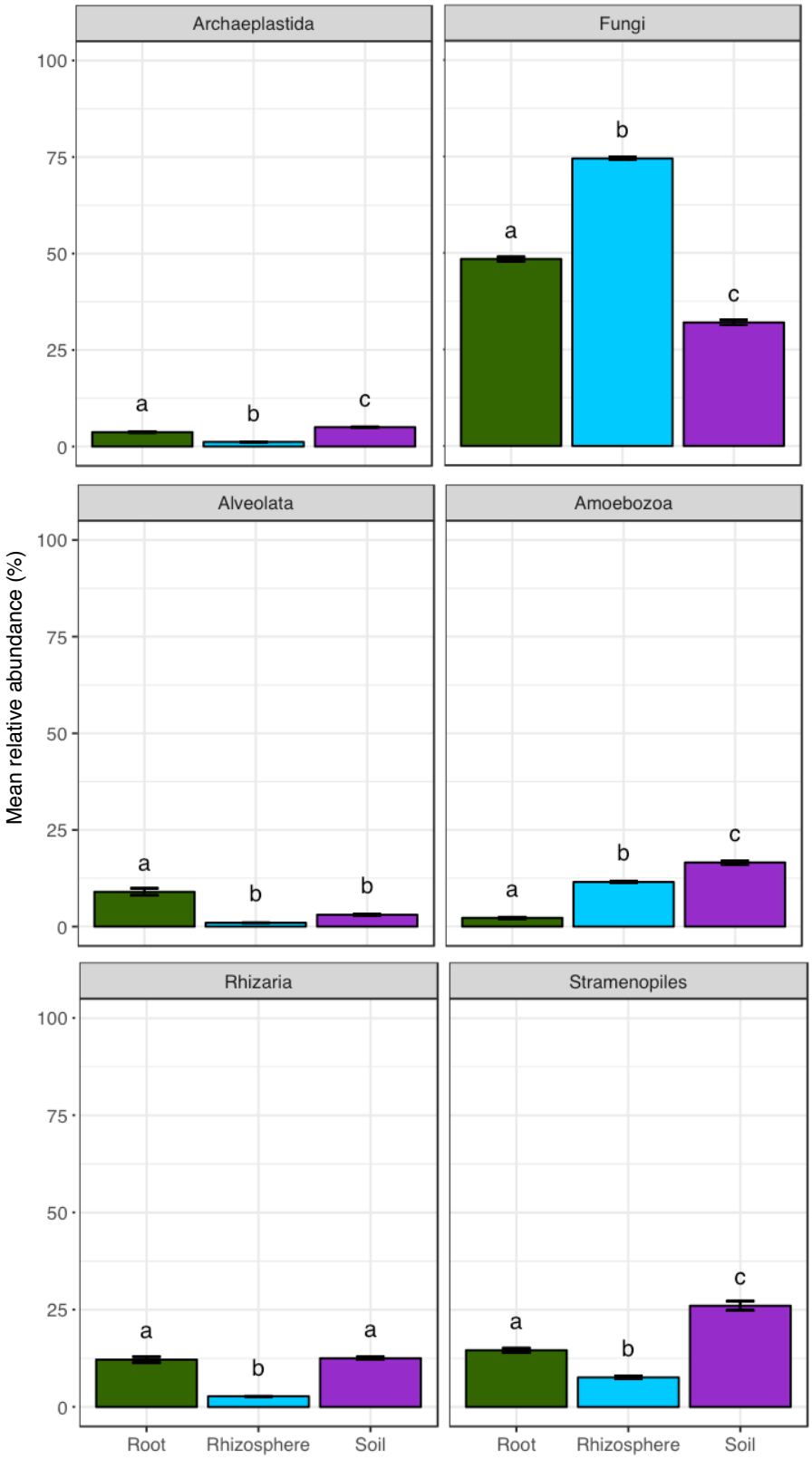


Figure 3.17 Facet bar plots displaying relative abundance of microbial eukaryotic phyla or high level taxonomic group in the root rhizosphere and soil. Significance of difference in relative abundance between methods as indicated by Tukey's HSD test P<0.05 treatments with the same letter are not significantly different. Error bars display standard error of the mean.

At a finer level, only one group of taxa had significantly different relative abundance between each compartments after Bonferroni multiple corrections. The relative abundance of an OTU from the Spizellomycetales order was significantly greater in the root (2.13%) than the rhizosphere (0.13%), the relative abundance of the *Acanthomeoba* genus was significantly greater in the soil (4.84%) than the rhizosphere (0.75%) and the root (0.08%).

3.3.7 The impact of RNA preservation technique on the functional processes of the soil microbial community

Using DIAMOND alignment in MEGAN, total of 6,682,934 microbial functional sequences were identified in the non-ribodepleted samples (Figure 3.18). Filtering of plant functions led to a decrease in the number of sequences in the root.

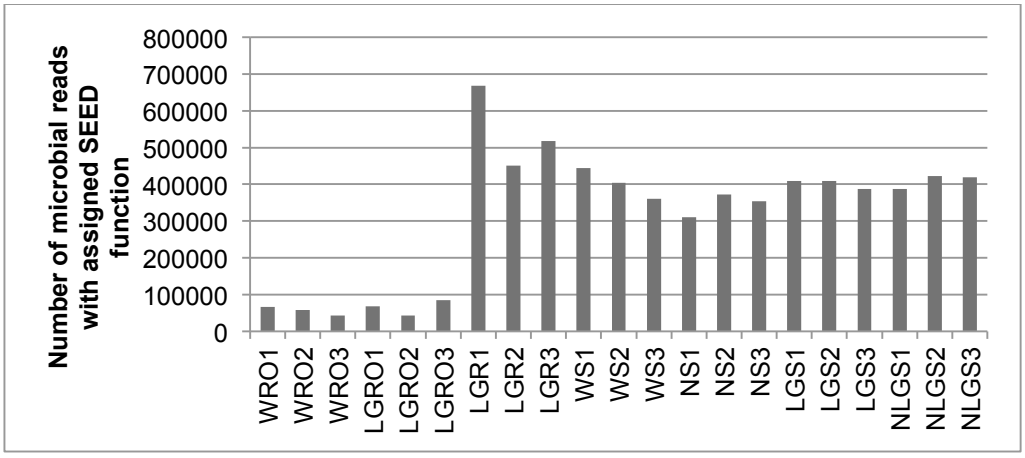
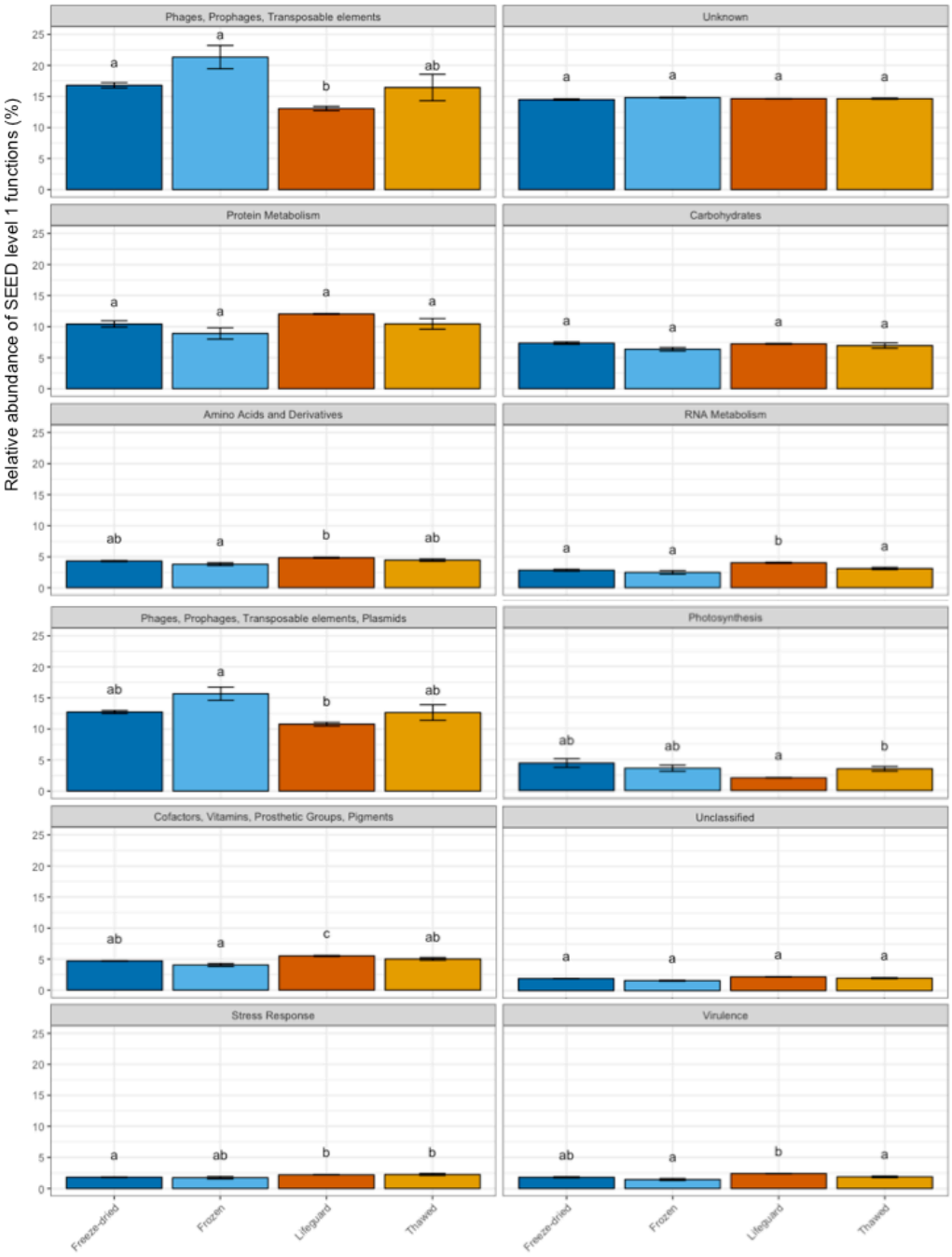


Figure 3.18 Number of microbial sequences with assigned SEED function per sample

Differences were identified between the relative abundance of SEED level 1 functional categories between RNA preservation methods in the soil (Figure 3.19). The relative abundance of two categories; “Respiration” and “Other” were significantly greater in the Freeze-dried soil than the Frozen soil. The relative abundance of three categories; Miscellaneous”, “Other” and “Cofactors, Vitamins, Prosthetic Groups, Pigments” were significantly greater in the Thawed soil treatments, when compared to the Frozen

treatment (Figure 19). The relative abundance of ten SEED level 1 categories; "Other", "Sulfur Metabolism", "Miscellaneous", "Cofactors, Vitamins, Prosthetic Groups, Pigments", "Virulence", "RNA Metabolism", "Cell Wall and Capsule", "Phages, Prophages, Transposable elements, Plasmids", "Phages, Prophages, Transposable elements", and "Amino Acids and Derivatives" were significantly different (P<0.05) between the Frozen and Lifeguard soil preservation methods when tested using White's non-parametric t test (Figure 3.19).



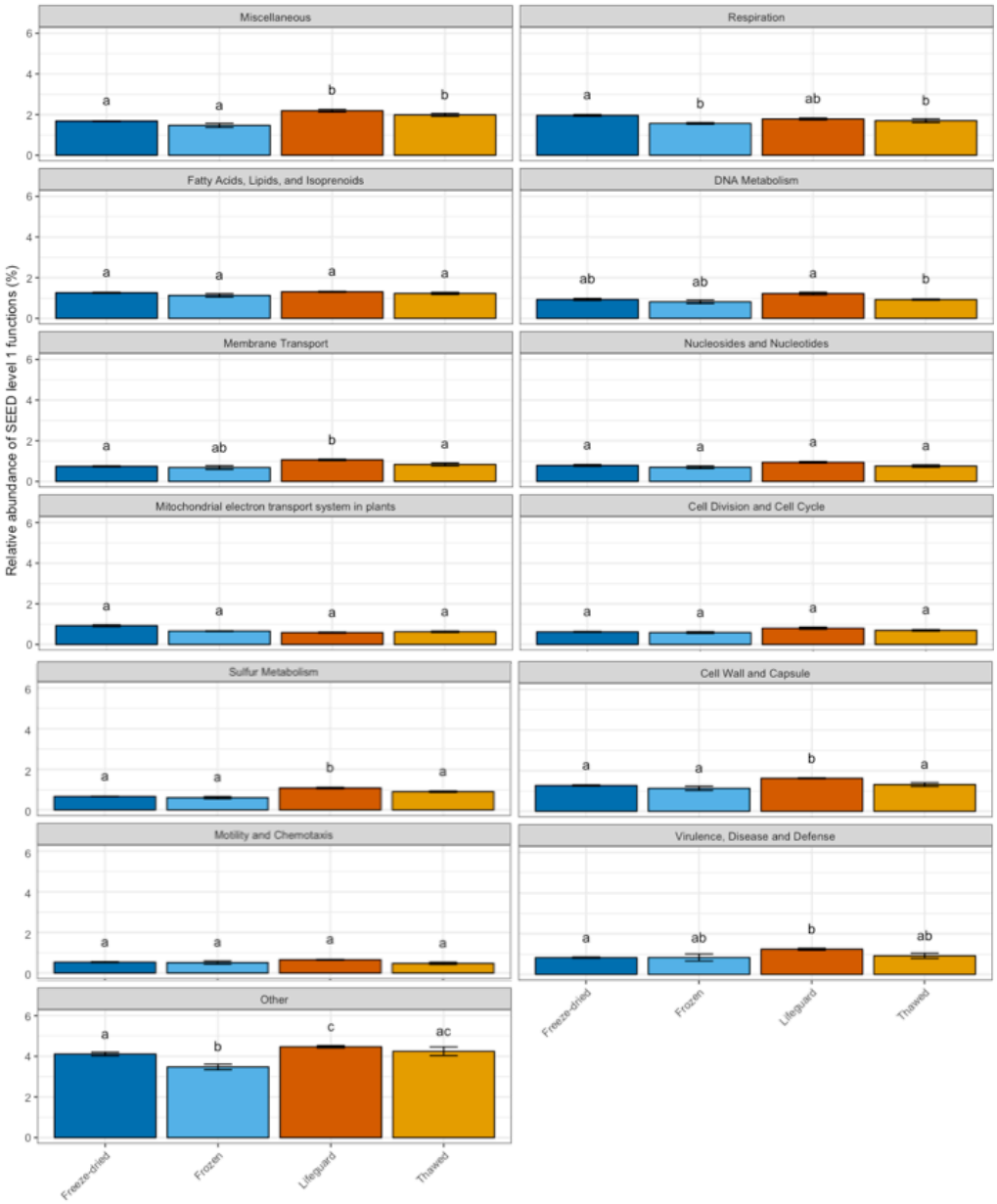
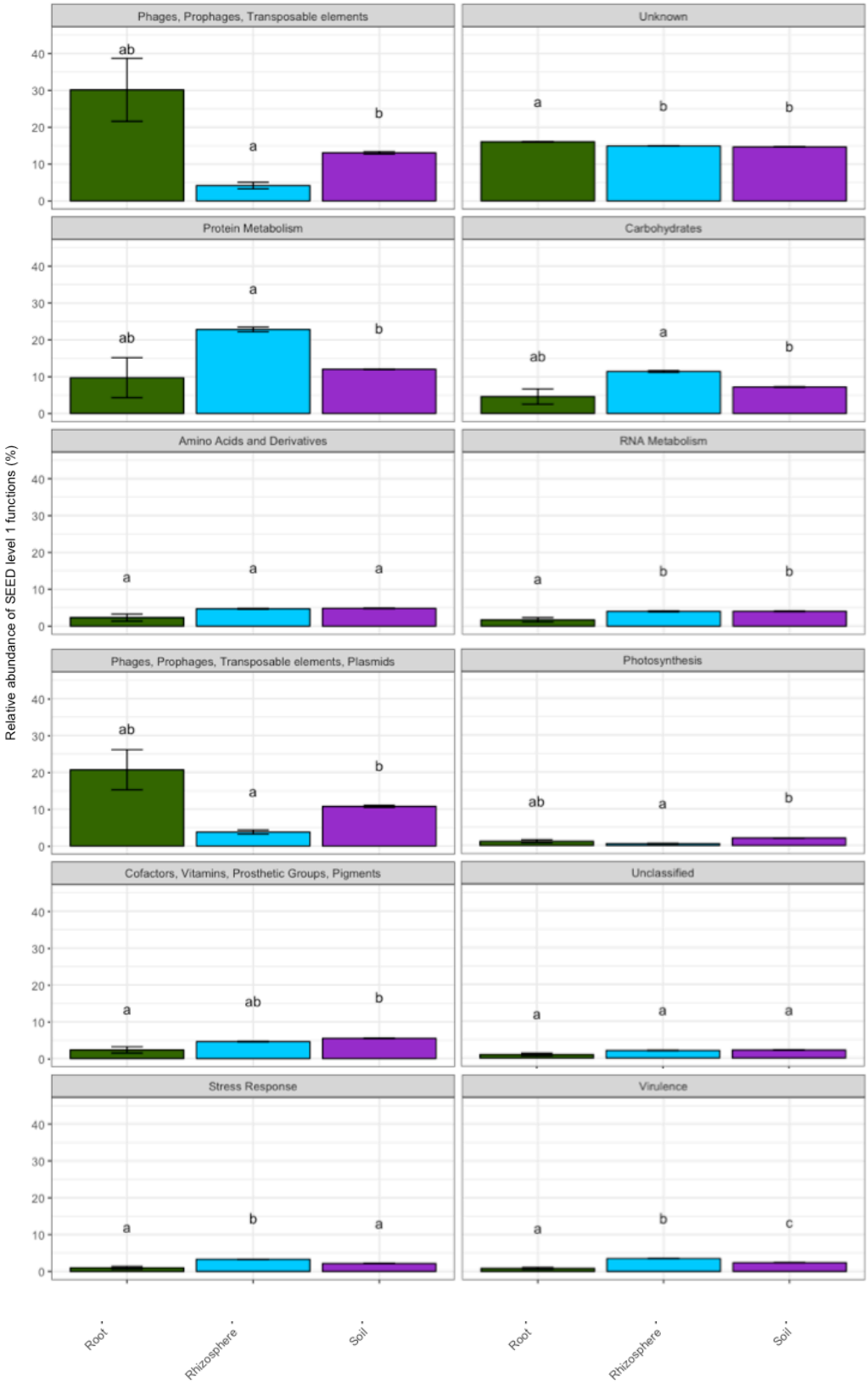


Figure 3.19 Relative abundance of SEED level 1 categories in the microbial transcriptomes of soil samples extracted under the four different RNA preservations methods. Error bars signify standard error of the mean. Treatments that share a letter are not significantly different (P<0.05) as tested by a White's non-parametric T test

3.3.8 Differences in functional fingerprints between compartments

Significant differences were identified between the relative abundance of SEED level 1 functional categories in the root, rhizosphere and soil compartments, as tested by White's non-parametric t test (Figure 3.20).


Phages, Prophages, Transposable elements
Unknown
Protein Metabolism
Carbohydrates
Amino Acids and Derivatives
RNA Metabolism
Phages, Prophages, Transposable elements, Plasmids
Photosynthesis
Cofactors, Vitamins, Prosthetic Groups, Pigments
Unclassified
Stress Response
Virulence
Relative abundance of SEED level 1 functions (%)
Root
Rhizosphere
Soil

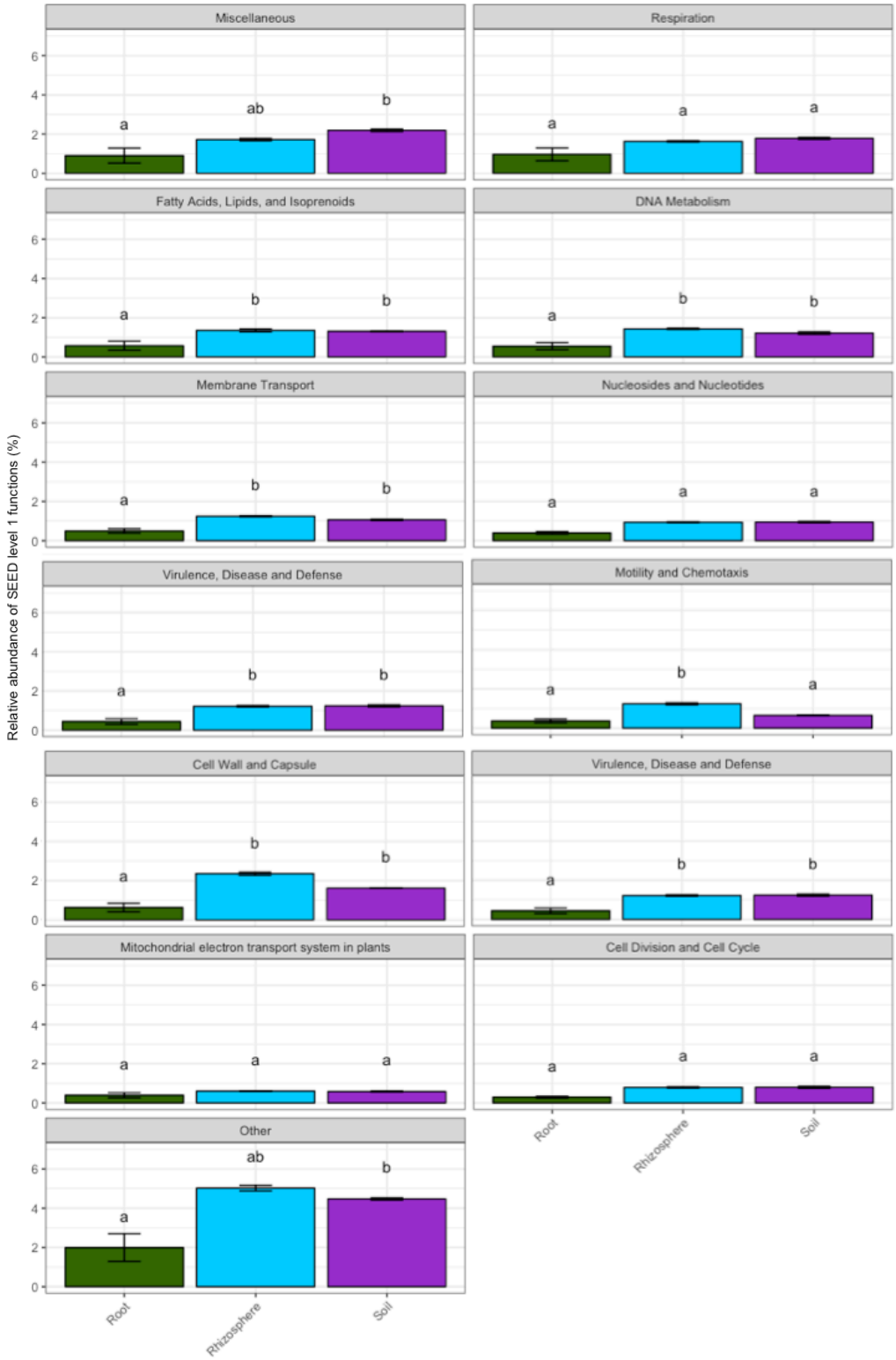


Figure 3.20 Relative abundances of SEED level 1 categories in the microbial transcriptomes of root, rhizosphere and soil samples extracted under the Lifeguard RNA preservations method. Error bars displayed as standard error of the mean. Compartments sharing a letter are not significantly different as tested by a White's non-parametric t test.

3.4 Discussion

3.4.1 The impact of RNA preservation technique on analysis of the composition and function of the soil microbial community

The first aim of this experiment was to determine whether the method of RNA preservation had a significant impact on the revealed microbial community composition of the bulk soil. The method of soil RNA preservation did not significantly impact the alpha diversity of the bacterial microbiome, however it did have a significant impact on the microbial eukaryotic alpha diversity. The alpha diversity of the microbial eukaryotic community of the Frozen soil and Freeze-dried soil were not significantly different, however, these were significantly greater than the alpha diversity of the Thawed and Lifeguard soil samples. It is likely that Lifeguard and Thawed treatment methods were less effective at preventing the degradation of the 18S rRNA which could be due to differences in stability of the 16S and 18S RNAs. The composition of nucleotides in nucleic acid molecules affects its stability, for example higher GC content confers increased thermostability in bacterial RNA compared to eukaryotic RNA (Wang *et al*., 2006). If these techniques were less efficient at protecting RNA degradation, then the less stable 18SrRNA may have been lost to a greater extent, resulting in the loss of rare taxa and therefore diminished alpha diversity. This effect may have been exacerbated by differences in overall diversity and abundance of the samples. The mean Fisher's alpha diversity of the 16S rRNA frozen soil was 3,450, whilst the mean Fisher's alpha diversity of the 18S rRNA frozen soil was 590. The 18S data will therefore be more sensitive to RNA loss than the 16S data.

The assigned composition of the microbial communities was also impacted by RNA preservation method. The relative abundance of eight out of the nine most abundant bacterial phyla were significantly different between RNA preservation techniques. The Freeze-dried method resulted in the fewest significant differences to the Frozen method at the phylum and genus level for both the bacterial and microbial eukaryotic data sets (4 of

15 most abundant phyla). The Thawed and the Lifeguard treatments revealed community compositions with 8 and 9 significantly different phyla respectively.

Overall, for both the bacterial and microbial eukaryotic communities, the Freeze-dried method revealed communities most similar to those of the Frozen soil. NMDS and cluster dendograms (Figures 3 and 12) highlighted the differences in beta diversity of the samples. The addition of the Lifeguard solution, for the Thawed and Lifeguard soils, had a greater distorting effect on the community composition than the Freeze-dry technique. The Lifeguard solution was least effective at maintaining a similar community structure to the Frozen treatment, most likely due to the lack of a freezing step, whilst the Thawed treatment which relies on the solution penetrating into the samples as the thawing process occurs, may have caused the shifts in community composition due to degradation of RNA. The concentration of the Freeze-dried RNA was closest to the Frozen, with Lifeguard having the least RNA. (Figure 4.3), although the quality of the RNA was lowest in the Freeze-dried samples (Table 3.2), the microbial community was most similar to that of the Frozen soil.

The low concentration of Lifeguard RNA may be due to incompatibility with the extraction technique (although it is the recommended procedure for the MOBIO kit), degradation prior to extraction or loss of RNA during the extra centrifugation step. In any case, not all of the environmental RNA was successfully extracted using the Lifeguard method and may account for extreme differences with the Lifeguard samples. Therefore the Freeze-dried technique is suggested as a cost efficient and high throughput method for sampling RNA from separate rhizosphere and root compartments in a field setting.

As with the taxonomic profiles, the Freeze-dried method had the fewest differences in functional classifications of the microbial communities to the Frozen method. Two of the SEED level 1 categories were significantly different between the Freeze-dried and Frozen methods. Three of the

Thawed SEED level 1 categories were significantly different between Thawed and Frozen methods and ten of the SEED level 1 categories were significantly different between lifeguard and thawed treatments. These differences corroborate with the results from the taxonomic study, that the Freeze-drying method is more suitable for high throughput field based rhizosphere RNA studies than the Thawing or Lifeguard methods.

3.4.2 Taxonomic and functional differences of the active root, rhizosphere and soil microbial communities

The second aim of the experiment was to gain an insight into the composition and relative abundances of active microbial communities in the root, rhizosphere and bulk soil. Prior to the start of the experiment, the Lifeguard treatment was presumed to be the best method for extracting rhizosphere RNA due to the highest RIN scores (Table 3.2) so was used for the comparison of soil, root and rhizosphere. However, this technique was subsequently found to lead to underrepresentation of eukaryotic taxa, such as alveolata and fungi (Figure 3.9) and this must be considered when discussing the comparison of rhizosphere root and soil communities.

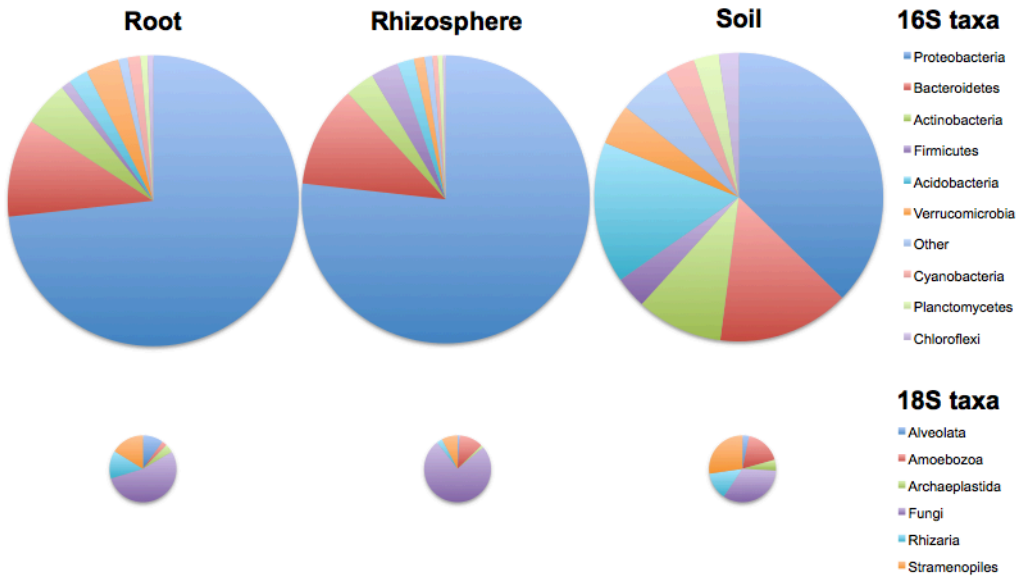


Figure 3.21 Visual representation of microbial communities in root rhizosphere and soil samples scaled to proportion of total sequences within each compartment

Alpha diversity was found to be significantly higher in soil than in root or rhizosphere samples. This was as expected since soil contains the total pool of organisms that can be recruited into the rhizosphere and root. Amplicon based rhizosphere studies have identified significant differences in the alpha diversity of the bulk soil, rhizosphere and root compartments (Edwards *et al*., 2015). The rhizosphere soil may have inflated diversity when measured using DNA due to relic DNA, and this study of RNA should reflect a better representation of rhizosphere diversity. Interestingly for both the bacterial and microbial eukaryotic communities, there were no significant differences between rhizosphere and root beta diversity. This could be due to the presence of fine root hairs in the rhizosphere soil, or perhaps a close similarity of the rhizoplane and adjacent rhizosphere soil communities. As mentioned before, a myriad of factors determine the assembly of rhizosphere microbial communities. Plant species in particular influences community composition. To give a broad indication of previously studied active microbial communities; in a citrus rhizoplane Proteobacteria (74.56%), Actinobacteria (16.80%), Bacteroidetes (2.86 %) were found as the dominant taxa (Zhang *et al*., 2017). In wheat, oat and pea rhizospheres, the dominant phyla were Proteobacteria followed by Actinobacteria Bacteroidetes and Firmicutes (Turner *et al*., 2013).

Here, a similar dominance of Proteobacteria was observed in the root and rhizosphere compartments (76.8% and 73.2% respectively) followed in abundance by Bacteroidetes (10.9% and 11.4.8% respectively) and Actinobacteria (5.0% and 3.3% respectively). In the work by Turner and colleagues, the wheat and pea rhizosphere soil was also enriched in Fungi compared to soil (~55% and ~70% respectively) as seen in this study (75% of eukaryotic sequences post filtering).

Here, possible plant beneficial microbes were identified in the roots and rhizosphere of *B. napus*. *Serratia* and *Janthiobacterium* were enriched in the rhizosphere, whilst *Cellvibrio* was enriched in the roots. *Serratia*

species isolated from the rhizosphere have been shown to have multiple plant growth promoting effects (Muller et al., 2009). *Janthobacterium* species have been identified as having some antimicrobial properties such as chitosanase production (Johnsen et al., 2010). *Cellvibrio* species have been isolated from plant roots and found to have nitrogen fixing capabilities (Suarez et al., 2014).

An interesting missing taxa from the data is *Olpidium brassicae*. This Chytridiomycete fungus has been found in very high abundance in *B. napus* roots. (Hartwright *et al*., 2010). The incidence of this potentially pathogenic taxa increases with continuous cultivation (Hilton *et al*., 2013). The virgin crop, which was sampled from in this study may not yet have been colonised with the *O. brassicae*.

The functional processes of the root microbial community were mainly dominated by phage sequences, an average in LGRO samples (50.1%). Previous metagenomes studies have identified increases in relative abundance of genes related to phage activity in the root and rhizosphere when compared to soil metagenomes (Bulgarelli *et al*., 2015). Endophytic microbes have been said to be "prone to phages" (Hardoim *et al*., 2015), and phages could be in high abundance in the rhizosphere zone as the high turnover of bacteria is beneficial for phage lifestyle. The increased abundance of phage transcripts in the root could be linked to increased competition in the root zone.

No functional SEED level 1 processes were enriched in the root. In the rhizosphere there were three classes of functional process that were enriched compared to the root or soil - 'Virulence', 'Motility & Chemotaxis' and 'Stress Response'. Increased virulence genes in the rhizosphere may be linked to a number of pathogenic microorganisms that are known to be present in the rhizosphere (Raaijmakers *et al*., 2008). In order to colonise the rhizosphere, microorganisms must be motile. Previous rhizosphere motility studies have indicated increased motility of rhizosphere microbes (Czaban *et al*., 2007). The increase of stress response related processes

in the rhizosphere may be linked to the depletion zones present in the rhizosphere as plants uptake the surrounding nutrients and water (York *et al*., 2016), and a subsequent in microbial stress response. The relative abundance of the SEED level 1 processes “Protein metabolism” and “Carbohydrates” were greater in the rhizosphere than the soil. The rhizosphere is considered a more active zone and increases in metabolism genes have previously been linked to the increased diversity and abundance of chemicals in the rhizosphere metagenomes (Ofaim *et al*., 2017).

Annotation of reads from metatranscriptomics depends on gene information from microbes which have been previously sequenced. A limitation of this lies in the number of sequenced microorganisms, from a given environment. When no metagenome has been sequenced in tandem, the reliability of these results depends on the accuracy of the alignments. The assignment methodology may need to be improved by increasing the cut-off percentage similarity for assignment. Coupling of metatranscriptomics with metagenomics may yield to increased information about the communities present and offer mapping of RNA to the metagenome for enhanced identification. One of the limitations of the shotgun metatranscriptomics approach used here is the short length of reads. This short sequence length (125bp) may contribute to a number of poor quality assignment of reads to sequences in databases. An example of this misassignment of taxa is the identification of a *Limnohabitans* OTU which are known marine planktonic bacteria (Kasalicky *et al*., 2013). Alternative approaches, such as the Oxford Nanopore MinION sequencer (Oxford Nanopore, UK) are capable of sequencing longer transcripts which may provide greater accuracy of assignment (Oikonomopoulos *et al*., 2016). However, the current accuracy of MinION is poor compared to Illumina technology (Goodwin *et al*., 2016).

The first aim of this work was to determine the extent to which different RNA preservation techniques influence the active microbial community and transcriptional fingerprint the soil. The Freeze-dried method was

shown to have a smaller distorting effect on the microbial community than the Lifeguard and Thawed treatments. The Freeze-dried methods resulted in both taxonomic and functional profiles most similar to the Frozen method.

The second aim was to determine differences in the active microbial composition and functional fingerprint between root, rhizosphere soil and bulk soil compartments. The function and composition of the rhizosphere root and soil compartments were significantly different. Although the taxonomic profiles of the rhizosphere and root compartments were similar in taxonomic composition, the functional processes were distinct. The root compartment contained a high proportion of phage related functional processes, whilst the rhizosphere contained a significantly greater proportion of protein metabolism processes.

Chapter 4. Identification of putative *PHT1* genes in cultivated *B. oleracea,* wild *B. oleracea* and wild *Brassica* C genome species.

4.1 Introduction

4.1.1 Brassica oleracea L.

Brassica oleracea L. is a species within the Brassicaceae family. *B. oleracea* genomes are diploid and form the CC component of the triangle of U (Nagaharu, 1935), (Figure 1.1). The *B. oleracea* group consists of a diverse range of widely distributed cultivated crop types, such as; cauliflower, kohlrabi, cabbage, broccoli, kale and Brussels sprouts. The crop varieties are globally economically important; over 71 million tonnes were produced in 2014 (FAOSTAT, 2014). The cultivated crops are interfertile with each other and with the wild Brassica C genome lines (Sutton, 1908).

Current theories of Brassica evolution suggest that the wild Brassica C genome varieties share a common ancestor from Western Europe that spread eastward, and underwent divergence through adaptation to unique environments (Hodgkin, 1995). There are 10 recognised species of wild C genome varieties (Mei *et al*., 2010). These wild varieties occupy coastal areas with cliffs, have a limited dispersal range and are threatened to various degrees. *Brassica hilarionis* is listed as endangered, *Brassica macrocarpa* is listed as critically endangered and both *Brassica rupestris* and *Brassica villosa* are listed as near threatened (The IUCN Red List of Threatened Species, 2017).

The diverse range of modern *B. oleracea* cultivated varieties appears to have been produced through an initial domestication from a wild variety in Eastern Europe (Maggioni *et al*., 2017). Subsequent human activity in Western Europe is believed to have driven the diversification of cultivated

B. oleracea through artificial selection. For example the preferential breeding of large stems to eventually form kohlrabi in Germany, and the selection of plants for lateral buds led to Brussels sprouts, cauliflower and broccoli diversification in Italy.

The free-growing sea cabbage *B. oleracea var. oleracea*, may have a "feral" nature, displaying a significantly narrower genetic range compared to the true wild species (Moore *et al*., 2017 personal communication). Because of this, the *B. oleracea var. oleracea* is classified as a separate "wild *B. oleracea* category". This variety is found in coastal regions of the UK, northern France and subsequent studies indicate a spread of this variety eastwards (Drenckhahn, 2017).

4.1.1 The importance of phosphorus for plants

Phosphorus (P) is one of 17 key plant growth nutrients, without which plants would be unable to survive (Mengel and Kirkby, 1987). P has a variety of functions in the plant cell, for example as a component of DNA, RNA, phospholipids and as a key energy transfer unit present in ADP and ATP (Ragothama, 1999). Most plants obtain phosphorus from the surrounding soil through uptake into root hairs (Jungk, 2001). Plants are only able to assimilate P as orthophosphate, either as $H_2PO_4^-$ in acidic soils or HPO_4^{2-} in alkaline soils. Globally phosphorus is second only to nitrogen as a growth limiting plant nutrient (Schactman et al., 1998b), and in some environments and growth stages it can be the limiting factor for plant growth (Agren *et al*., 2012). The amount of readily available inorganic P (Pi) in soil is usually very low, under 10 μM (Bielski, 1973), with a large portion of the P content of soil existing as organic phosphate (Po). Po can account for between 79% and 93% of the P content of soil (Schactman et al., 1998a).

Considering the importance of P to plant growth and the low concentration of accessible Pi, phosphate transport is an important and highly conserved

function in plants. In order to access the available Pi from the soil pool, transporters are needed to transfer the P up a chemical gradient into the plant, and then to translocate the P to where it is required.

Plant uptake and transport of Pi is controlled by phosphate transport proteins (PHT). PHT proteins are transmembrane Pi:H+ symporters (Ullrich and Novacky, 1990). The first identification of a phosphate transporter, was that of PHO84 in yeast (Bun-Ya *et al*., 1991). Homologous phosphate transporters were subsequently identified in the model plant species *Arabidopsis thaliana* (Muchhal *et al*., 1996), which were later classified and named as *PHT1* genes. Plant *PHT1* and yeast *PHO84* have similar sequences and are part of the phosphate transport superfamily (Pao *et al*., 1998). *PHT1* genes have been identified in a diverse range of both cultivated and wild plant species (Baker *et al*., 2015).
A.thaliana PHT proteins have been classified into 4 families, primarily based on the membrane in which they are located. PHT1 family transporters are located in the plasma membrane (Nussaume *et al*., 2011), PHT2 in the plastid inner envelope (Versaw, 2002),(Rausch *et al*., 2004), PHT3 in the mitochondrial inner membrane (Tabata *et al*., 2000), while PHT4 are located in the plastid envelope (Guo *et al*., 2008) or golgi body (Hassler *et al*., 2016).

PHT1 proteins are of primary interest with regards to uptake of P from soil. In *A.thaliana* there are 9 members of the PHT1 family. PHT1;1, PHT1;2, PHT1;3 and PHT1;4 are involved in the uptake of Pi from the soil solution into the plant (Mudge *et al*., 2002). *PHT1;1* and *PHT1;4* have been seen to be up regulated under P stress (Shin *et al*., 2004), whereas the expression of *PHT1;2* was not (Lapis-Gaza *et al*., 2014). PHT1;5 controls translocation of Pi from the root to the shoot tissue (Nagarajan *et al*., 2011). *PHT1;6* and *PHT1;7* expression was identified in pollen grains (Mudge *et al*., 2002). PHT1;8 and PHT1;9 have more recently been shown to have a role in the root to shoot translocation of Pi (Lapis-Gaza *et al*., 2014). The expression profiles of the *PHT1* genes in *A. thaliana* can change at different development growth stages. The expression of *A.thaliana PHT1* genes was

visualized using the Plant eFP viewer from the Bio-analytic Resource for Plant Biology (Austin *et al*., 2016), hosted at (bar.utoronto.ca/eplant) (Supplementary Figure 2). *AtPHT1;1* and *AtPHT1;2* had high expression in the root tissue at the vegetative rosette stage and no expression in any other tissues. *AtPHT1;3* had high levels of expression in root tissue and mature pollen grains, as well as medium level expression across a range of other tissues and growth stages. *AtPHT1;4* had high expression in stamens and medium expression in mature pollen, however other expression studies have also shown expression *AtPHT1;4* in the roots (Peñaloza *et al*., 2016). The expression of *AtPHT1;5* is highest in senescent leaves, with medium expression in mature pollen, petals and dry seed. The expression of *AtPHT1;6* was recorded as high in the flower bud, while *AtPHT1;7* was expressed at high levels in the stamen and medium levels in mature pollen. The expression levels of *AtPHT1;8* were recorded to be highest in the mature pollen grain. The expression level of *AtPHT1;9* was high in roots of the vegetative growth stage and seeds.

The main aims of this work were to 1.) Investigate the overall differences in expression pattern in the roots for wild C genome species and crop types of *Brassica oleracea*. 2.) Identify homologues of *PHT1* genes in the roots of *B. oleracea* based on BLAST searches and sequence alignments. 3.) Determine whether the wild C genome, the wild *B. oleracea* and cultivated *B. oleracea* display any differences in expression of predicted *PHT1* genes, using a wide diversity set of the *B. oleracea* C genome lines.

4.2 Materials and Methods

4.2.1 Brassica category selection

A diverse set of 118 lines from eight *B. oleracea* crop types, nine wild *Brassica* C genome (wild) species and the wild *B. oleracea* species were used in this experiment (Table 4.1). *Brassica* lines were clustered into “Wild”, “Cultivated” and “Wild *B. oleracea*” categories based on the

phylogenetic groupings provided by as yet unpublished work (Moore *et al*., 2017, personal communication).

Table 4.1 Categorisation of wild species and crop types of *Brassica* used in transcriptome study

Wild species/ Crop type	Variety	Category	Number of lines
B. bourgaei	-	Wild	1
B. cretica	-	Wild	8
B. hilarionis	-	Wild	2
B. incana	-	Wild	10
B. insularis	-	Wild	2
B. macrocarpa	-	Wild	17
B. montana	-	Wild	3
B. rupestris	-	Wild	3
B. villosa	-	Wild	8
B. oleracea	*Var. oleracea* (wild cabbage)	Wild *B. oleracea*	24
B. oleracea	Broccoli	Cultivated	11
B. oleracea	Brussels sprouts	Cultivated	4
B. oleracea	Cabbage	Cultivated	6
B. oleracea	Cauliflower	Cultivated	10
B. oleracea	Chinese kale	Cultivated	2
B. oleracea	Collard greens	Cultivated	1
B. oleracea	Kohlrabi	Cultivated	4
B. oleracea	Ornamental kale	Cultivated	1
B. oleracea	White flowered Kale	Cultivated	1

The wild species resources developed in the Defra VeGIN project (https://vegin.warwick.ac.uk/download.php) available for use in this study were collected from wild populations across Europe: *B. bourgaei* from the Canary Islands, *B. cretica* from the eastern Mediterranean, *B. hilarionis* from Cyprus, *B. insularis* from Corsica, *B. montana* from France and *B.*

incana, B. rupestris, and B. macrocarpa from Sicily or the small islands surrounding Sicily (Figure 4.1).

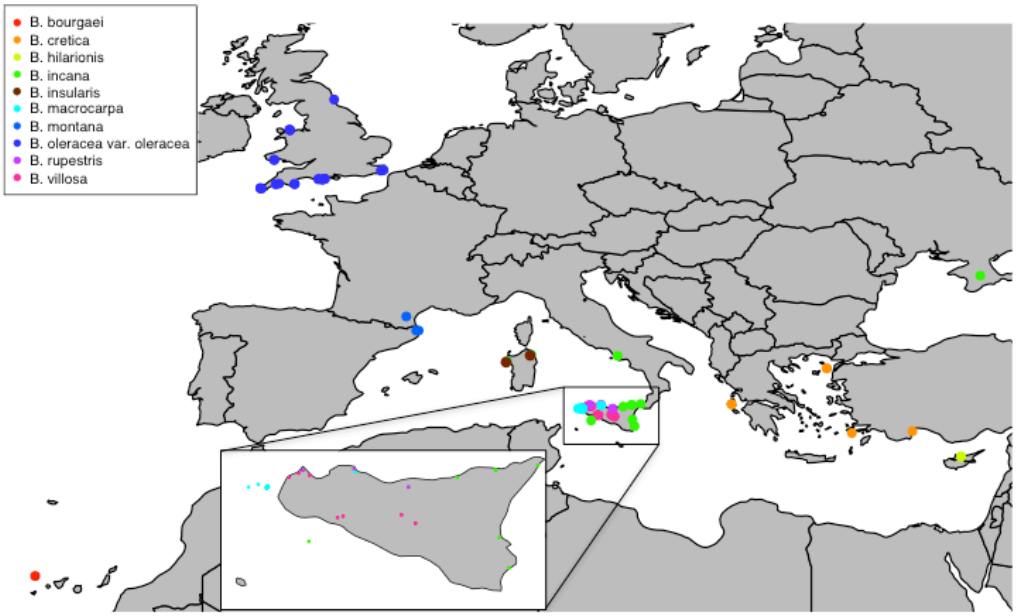


Figure 4.1 Geographical distribution of wild *Brassica* species used in this study

4.2.2 RNA expression data

RNA-seq data was available for use in this project after generation by a *Brassica* C genome diversity transcriptome study (Genebank BioProject accession PRJNA289196: https://www.ncbi.nlm.nih.gov/bioproject/289196) and is described below.

Three replicates plants of each line (generated by one selfing of accessions obtained from the UK Vegetable Genebank at Wellesbourne) were grown on rockwool blocks using a hydroponic nutrient film technique (NFT) with circulation media containing 3.2 mM $Ca(NO_3)2.4H_2O$, 0.8 mM NH_4NO_3, 0.8 mM $Ca(NO_3)2.4H_2O$, 0.2 mM NH_4NO_3,2 mM K_2SO_4, 1 mM KH_2PO_4,0.8 mM $MgSO_4.7H_2O$, 0.1 mM FeNaEDTA, 30 μM H_3BO_3, 10 μM $MnSO_4.H_2O$, 1 μM $ZnSO_4$, 3 μM $CuSO_4$ and 0.5 μM $Na_2MoO_4.2H_2O$. Plants were grown in a glasshouse at 15°C under 16h light: 8h dark lighting regime with top-up light on low-light days.

RNA samples were obtained by harvesting root and leaf tissues at the true-leaf seven stage (21 days) the replicates for each line. The tissues were flash frozen in liquid nitrogen and total RNA was extracted using the Plant RNeasy mini kit (Qiagen Inc., UK) following the manufacturer's instructions. RNA quality was assessed using a BioAnalyzer (Agilent, UK). RNA from the three leaf and root replicate samples was pooled into one sample of root and leaf RNA for each line. For each sample, Illumina library preparations were performed using mRNA-TruSeq sample prep kit v.5 (Illumina Inc., San Diego) according to the manufacturer's protocol. The libraries were sequenced using the Illumina GAIIx sequencing system, which generated 70 base paired-end sequence reads. Bases were quality scored using Illumina CASAVA pipeline version 1.7 (Illumina, 2010). Sequence reads were aligned to the published *B. rapa* genome assembly using the Bowtie (Langmead and Salzberg, 2012) and Tophat (Kim *et al*., 2013) algorithms, and to an assembly of *B. oleracea var*. *alboglabra* line TO1000DH3 (Parkin *et al*., 2014).

4.2.3 BLAST analyses to identify Brassica oleracea gene homologues of known PHT1 genes.

To identify potential *PHT1* genes in *B. oleracea*, genes with known sequences from the most closely related plant species were identified. The *A. thaliana* nucleotide sequence for each of the 9 known *A. thaliana PHT1* genes and a *B. napus PHT1;4* gene, were searched, using BLASTN (Altschu *et al*., 1990), against the *B. oleracea* genome (Parkin *et al*., 2014). Both the target sequences and the *B. oleracea* genome were downloaded from the Ensembl plants database (Kersey *et al*., 2016). Hits were filtered based on a BLAST score of at least 1000, which is a number used to describe similarity of sequences, and a percentage similarity of the alignment of over 80%. Due to erroneous annotation caused by automatic assembly, potential *B. oleracea* homologues of *AtPHT1* genes were visually assessed and manually edited in order to better estimate exon-intron structure (further description in results section).

Potential homologs were searched using BLAST on the *B. oleracea* genome in order to check the homologs did not have a match to a known gene. Predicted *B. oleracea* homologues and known *A. thaliana* PHT coding sequences were aligned using the MUSCLE algorithm (Edgar, 2004) in Megalign Pro (Megalign Pro. Version 11.0. DNASTAR. Madison, WI). A phylogenetic tree was generated on the alignment using Megalign (Megalign. Version 11.0. DNASTAR. Madison, WI) using bootstrapping of 1000 trials.

4.2.4 Analysis of expression data

Gene expression was quantified using cufflinks (Trapnell *et al*., 2010), values for the transcriptome dataset by Jonathan Moore for the *Brassica* C genome diversity transcriptome study. Data was made available for use in this project and displayed as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Heatmaps and bar plots were generated in R (R Core Team, 2013) using the 'ggplot2' (Wickham, 2009) and 'RcolorPalette' packages. NMDS plots and ANOSIM were conducted using a Bray-Curtis resemblance matrix in PRIMER 6 (Clarke and Gorley, 2006.). Statistical tests between root and leaf were conducted using the Kruskall-Wallace test in R using the 'stats' package (R Core Team 2013). Pairwise significance values between Brassica categories (cultivated, wild oleracea and wild) were conducted using the Dunn test from the 'FSA' package (Wickham, 2009, Ogle, 2017). Data was visualized using R and differences between wild, wild *B. oleracea* and cultivated were tested using pairwise Kruskall-Wallace, and corrected for multiple corrections using Dunn method, using GraphPad Prism (GraphPad Software, La Jolla California USA).

4.3 Results

4.3.1 Differences in Brassica root transcriptomes

The differences in expression pattern for the whole root transcriptome from each of the wild species and crop types were tested for significance using an ANOSIM test ($P<0.005$) and visualized using NMDS (Figure 4.2). Significant differences in expression pattern for overall root expressional fingerprint were identified between 58 pairs of wild species and crop types (Table 4.2). Only wild species or crop types with at least 3 lines were included in ANOSIM analysis. Additionally, in order to generate a reliable statistical ANOSIM test result, a threshold of a minimum of 100 permutations between pairs of wild species or crop types was required in order to be included in the results.

Table 4.2 Pairwise ANOSIM results for putative PHT gene expression for wild species or crop types with at least 3 lines and, comparisons containing at least 100 permutations

Group 1	Group 2	R Statistic	Significance level (P)
Brussels sprouts	Cabbage	0.44	0.02
B. incana	*B. oleracea*	0.21	0.02
Cabbage	Broccoli	0.287	0.02
B. oleracea	*B. rupestris*	0.488	0.02
Cabbage	Kohlrabi	0.591	0.01
B. montana	*B. oleracea*	0.586	0.007
B. montana	*B. villosa*	0.962	0.006
B. montana	*B. cretica*	0.715	0.006
B. villosa	*B. rupestris*	0.723	0.006
B. rupestris	*B. cretica*	0.661	0.006
Brussels sprouts	Cauliflower	0.607	0.004

B. montana	*B. macrocarpa*	0.903	0.003
B. rupestris	*B. macrocarpa*	0.865	0.003
B. montana	Broccoli	0.998	0.003
B. rupestris	Broccoli	0.975	0.003
B. montana	*B. incana*	0.951	0.003
B. montana	Cauliflower	0.994	0.003
B. incana	*B. rupestris*	0.832	0.003
B. rupestris	Cauliflower	0.967	0.003
B. macrocarpa	Brussels sprouts	0.916	0.002
Cauliflower	Kohlrabi	0.719	0.002
B. villosa	Brussels sprouts	1	0.002
B. villosa	Kohlrabi	1	0.002
B. cretica	Brussels sprouts	0.614	0.002
B. cretica	Kohlrabi	0.619	0.002
B. incana	*B. villosa*	0.892	0.001
B. incana	*B. macrocarpa*	0.887	0.001
B. incana	*B. cretica*	0.717	0.001
B. incana	Brussels sprouts	0.796	0.001
B. incana	Cabbage	0.853	0.001
B. incana	Cauliflower	0.946	0.001
B. incana	Kohlrabi	0.813	0.001
B. incana	Broccoli	0.829	0.001
B. villosa	*B. oleracea*	0.515	0.001
B. villosa	*B. macrocarpa*	0.832	0.001
B. villosa	*B. cretica*	0.756	0.001
B. villosa	Cabbage	1	0.001
B. villosa	Cauliflower	1	0.001

B. villosa	Broccoli	0.999	0.001
B. oleracea	*B. macrocarpa*	0.602	0.001
B. oleracea	*B. cretica*	0.547	0.001
B. oleracea	Cauliflower	0.405	0.001
B. oleracea	Broccoli	0.403	0.001
B. macrocarpa	*B. cretica*	0.827	0.001
B. macrocarpa	Cabbage	0.914	0.001
B. macrocarpa	Cauliflower	0.943	0.001
B. macrocarpa	Kohlrabi	0.913	0.001
B. macrocarpa	Broccoli	0.952	0.001
B. cretica	Cabbage	0.697	0.001
B. cretica	Cauliflower	0.874	0.001
B. cretica	Broccoli	0.844	0.001
Cabbage	Cauliflower	0.653	0.001
Cauliflower	Broccoli	0.331	0.001

Using ANOSIM, the total expression profiles in the root samples of the wild, cultivated and wild *B. oleracea* groups were all identified as significantly ($P<0.005$) different from each other (Figure 4.2B). The cultivated group was significantly different ($P<0.001$) to both the wild and wild *B. oleracea*, whereas the wild and wild groups were significantly different using the ANOSIM test ($P=0.004$). There was increased variability in expression profile of the wild lines (Figure 4.2). The greater spread of the wild C genome and wild *B. oleracea* lines represents a greater variation in expression patterns than the more tightly clustered cultivated lines, even though the cultivated lines represent a variety of diverse crop types.

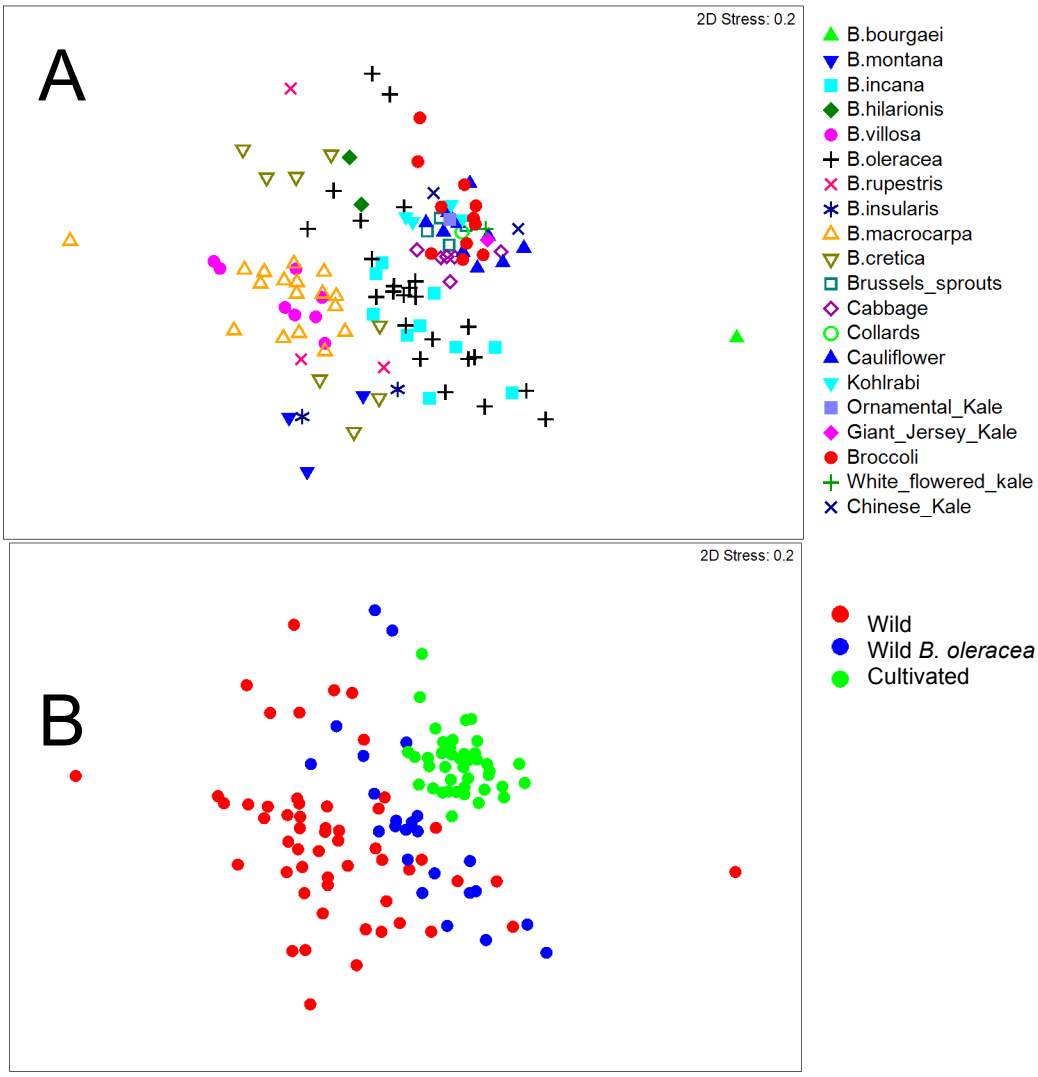


Figure 4.2 Nonmetric Multi-Dimensional Scaling (NMDS) plots displaying the Bray-Curtis similarity of whole root transcriptome profile of each line. A.) Lines grouped by category B.) Lines grouped by category. The distance in 2-dimensional space represents dissimilarity between samples. Therefore the closer together two points on the NMDS are, the more similar the expression profile is.

4.3.2 Identification of Putative PHT1 genes in B. oleracea

The results of the BLAST search generated potential homologues in *B. oleracea* from genomic sequence of the nine *A. thaliana PHT1* genes as well as the single available *Brassica napus* sequence *PHT1;4* sequence (Supplementary Table 2). Twenty-five putative *B. oleracea* genes were identified with a percentage similarity of over 80% to one of the *A. thaliana* PHT1 sequences. Nine of the putative *B. oleracea* sequences showed closest identity to *AtPHT1;1, AtPHT1;2* and *AtPHT1;3* genes. Eight

showed closest identity to *AtPHT1;4*, one had highest similarity to *AtPHT1;5*, two had highest similarity to *AtPHT1;6,* six had highest similarity to *AtPHT1;7*, three had highest similarity to *AtPHT1;8* and two had highest similarity to *AtPHT1;9* (Table 4.3).

Upon investigation of the predicted *B. oleracea PHT1* sequences, errors in annotations of the putative genes were identified. Many genome annotations for *B. oleracea* were generated automatically, and errors in the annotation of genomes are possible due to the process. The coding sequence and genomic sequence homologues were aligned to the coding sequence of *A.thaliana* and manual annotation of the putative homologues was conducted in order to develop coding sequences with better fit to the predicted structure of the *A. thaliana* genes. Manual changes were made to Bo9g059120 in order to remove the repeated N's, and exon 2 was extended to include the last available homologous base (The letter N is used in sequencing when the machine is unable to call a base, this can be due to a range of factors including sequencing error). Exon 1 of Bo9g059110 was set to include the upstream bases that show homology to the start of the *A.thaliana* PHT1;1 genes.

The genomic region covering the predicted Bo4g185840 and Bo4g185850 genes had high similarity to *AtPHT1;4*. The annotation of the start and end points of the coding region was changed to match the known coding region of *AtPHT1;4*. Bo5g030740 and Bo5g030730were annotated as exon 1 and 2 of a potential *PHT1;8* homologue, due to homology to exon 1 and exon 2 of *AtPHT1;8* respectively. An intron was removed from exon 1 of Bo8g070440 to better match the coding region of *AtPHT1;8*. Bo6g120500 and Bo6g120510 were annotated as exon 2 and 1 of a potential *PHT1;9* homologue. An additional annotated exon was removed from Bo6g084900 to better match the coding sequence of *AtPHT1;9*. (Supplementary Figure 3). Additionally the terms "phosphate transporter' and '*B. oleracea'* were queried in the NCBI gene database, leading to the identification of five additional sequences; Bo2g089910 (PHT1;4), Bo2g119350 (PHT1;3) , Bo6g120470 (PHT1;9), Bo6g120540 (PHT1;9)

and Bo7g065380 (PHT1;3). The sequence for Bo6g120470 and Bo6g120480 showed high similarity to exon 1 and 2 of *AtPHT1;9* respectively.

A phylogenetic tree was generated using a MUSCLE alignment (Edgar, 2004) of the manually edited coding sequences of the *A. thaliana* PHT1 sequences, and the top blast hits in the *B. oleracea* genome, after manual annotation (Figure 4.3). The high similarity of the *AtPHT1;1*, *AtPHT1;2* and *AtPHT1;3* coding sequence is revealed by the inclusion of these sequences in a separate cluster displaying high percentage similarity (Figure 4.3). Within this cluster are eleven putative *B. oleracea* genes that share the most similarity to the *AtPHT1;1, AtPHT1;2* and *AtPHT1;3* sequences, and two additional sequences which were most closely related to this group. The *A.thaliana* and *B. napus PHT1;4* coding sequences form a cluster with six putative *B. oleracea* genes. AtPHT1;5 shows high similarity and forms a bifolious cluster with a putative *B. oleracea* gene. AtPHT1;6 shows similarity to two putative *B. oleracea* sequences which form a separate cluster. *AtPHT1;7* forms a cluster with two putative *B. oleracea* genes. *AtPHT1;8* forms a cluster with three putative *B. oleracea* genes. *AtPHT1;9* forms a cluster with five putative *B. oleracea* genes. The clustering of sequences, along with and the percentage similarity scores were used to attribute a predictive *AtPHT;1* homologue, for each of the putative *B. oleracea* genes (Table 4.3).

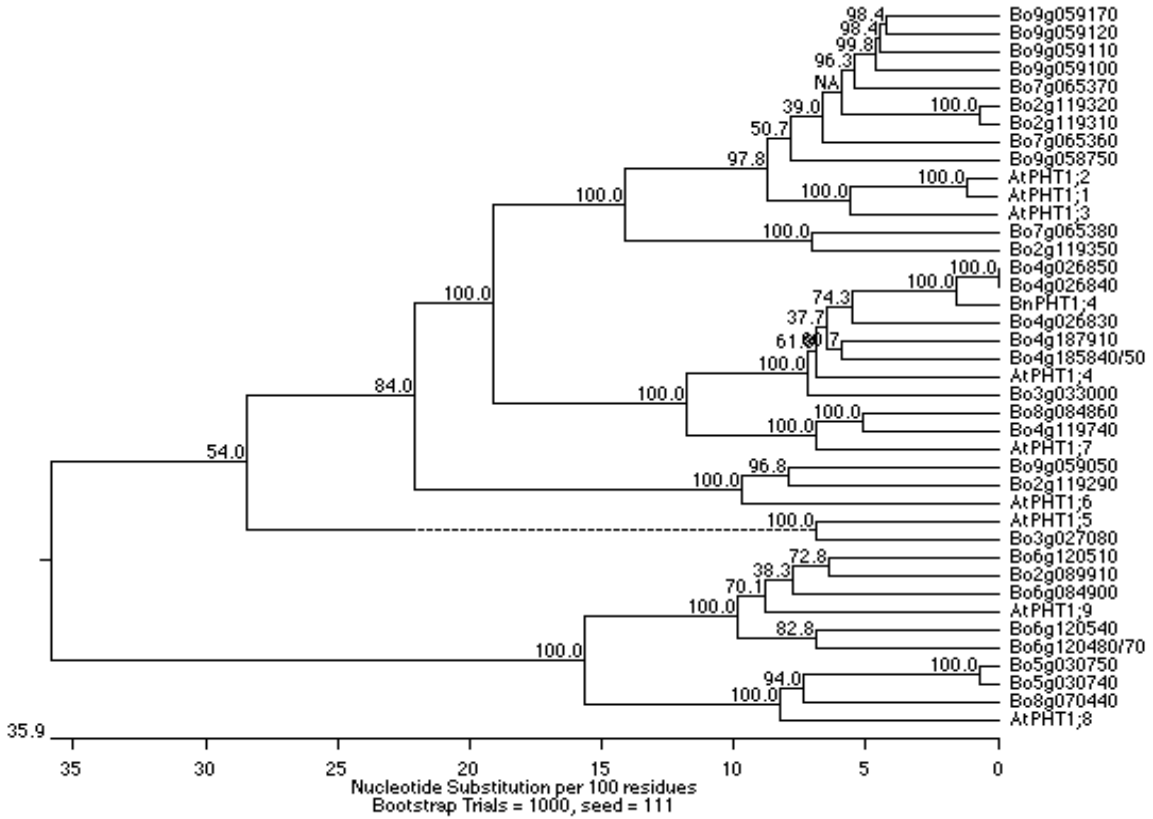


Figure 4.3 Hierarchical clustering dendogram of *A.thaliana* PHT1 coding sequences, *B. napus PHT1;4* and putative *B. oleracea* PHT coding sequences using MUSCLE alignment and 1000 Bootstraps. Phenogram, averages the distance between ancestors in the tree. Dotted line is negative branch length that results from averaging

Table 4.3 Putative *B. oleracea PHT1* genes and closest similarity homologue in *A. thaliana*.

Closest Homologue in *A.thaliana* based on % similarity	Putative B. oleracea gene model number
PHT1-1* PHT1-2* PHT1-3*	Bo2g119310
	Bo2g119320
	Bo2g119350
	Bo7g065360
	Bo7g065370
	Bo7g065380
	Bo9g058750
	Bo9g059100
	Bo9g059110
	Bo9g059120
	Bo9g059170
PHT1-4	Bo3g033000
	Bo4g026830
	Bo4g026840
	Bo4g026850
	Bo4g185840/50
	Bo4g187910
PHT1-5	Bo3g027080
PHT1-6	Bo2g119290
	Bo9g059050
PHT1-7	Bo4g119740
	Bo8g084860
PHT1-8	Bo5g030730/40
	Bo5g030750
	Bo8g070440
PHT1-9	Bo2g089910
	Bo6g084900
	Bo6g120470/80
	Bo6g120500
	Bo6g120540

4.3.3 *Detecting expression of predicted PHT1 genes in root and leaf tissue of Brassica*

Using the transcriptome data, the expression values for each line were extracted for the each of the putative *B. oleracea PHT1* genes (Figure 4.4).

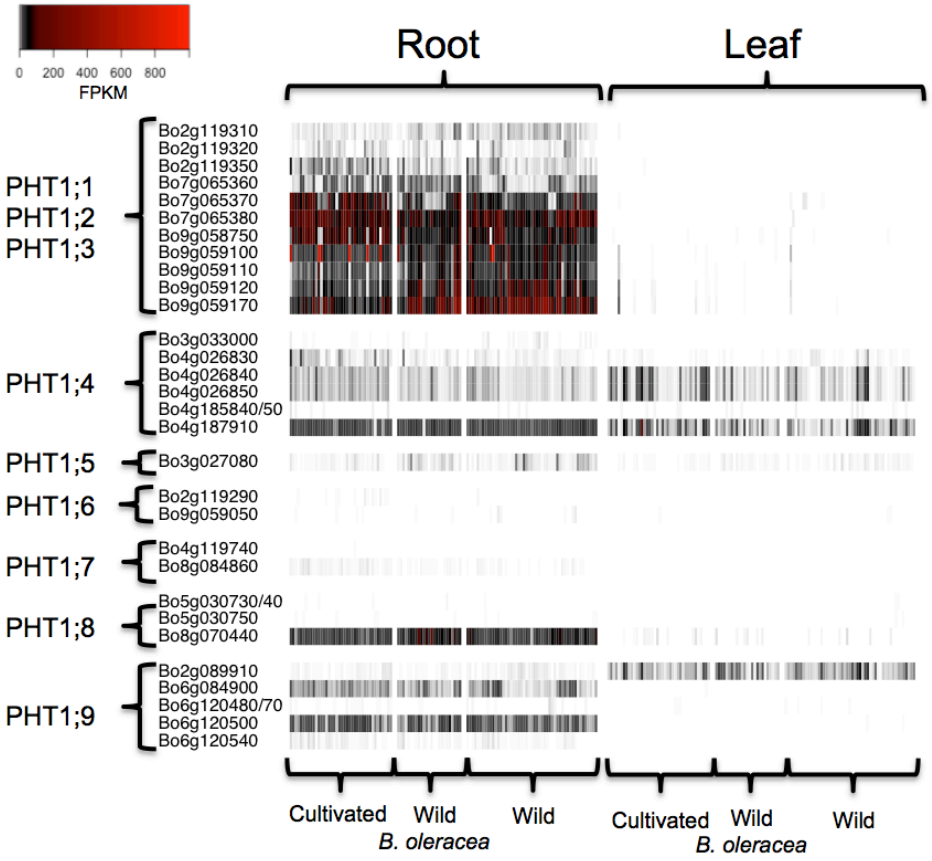


Figure 4.4 Heatmap of RNA expression in the root and leaf tissues of *B. oleracea* lines at true leaf 7 for the putative *PHT1* genes. The FPKM values represented are displayed using a staggered colour key. White represents no expression, black represents expression of 100 FPKM and red is expression of up to 200 FPKM. X axis is grouped based on category and Y axis is grouped based on closest identity of *A. thaliana PHT1* gene.

In order to limit the number of potential targets and account for sequencing noise, only putative genes with expression levels of over 10 FPKM were included in further analysis. Potential *B. oleracea PHT1* homologues were expressed at high levels in the root tissue for eight of the putative *PHT1;1,*

PHT1;2 and *PHT1;3* (Figure 4.5) and (Table 4.4) Expression of one of the *PHT1;4* homologues was identified at over 10 FPKM. No expression of potential *PHT1;5, PHT1;6* or *PHT1;7* genes was identified in the roots or leaf tissue. Expression of one potential *PHT1;8* homologue was identified in the root tissue. Expression of one putative *PHT1;9* gene was identified at above an average of 10 FPKM in the roots across *B. oleracea* lines.

Table 4.4 Mean FPKM expression level in root and leaf tissues across all lines of predicted Brassica PHT homologues.

A. thaliana PHT1 gene	Putative *B. oleracea PHT1* gene	Average expression in root (FPKM) across all lines	Average expression in leaf (FPKM) across all lines
PHT1;1 *PHT1;2* *PHT1;3*	*Bo2g119310*	1.1	0.0
	Bo2g119320	1.4	0.0
	Bo2g119350	6.5	0.0
	Bo7g065360	7.7	0.0
	Bo7g065370	78.2	0.0
	Bo7g065380	94.7	0.0
	Bo9g058750	76.8	0.0
	Bo9g059100	121.3	0.1
	Bo9g059110	16.0	0.0
	Bo9g059120	23.7	0.1
	Bo9g059170	53.0	0.2
PHT1;4	*Bo3g033000*	*0.0*	*0.0*
	Bo4g026830	*3.3*	*0.2*
	Bo4g026840	*3.9*	*5.3*
	Bo4g026850	*3.9*	*5.3*
	Bo4g185840/50	*0.1*	*0.1*
	Bo4g187910	*15.5*	*8.6*
PHT1;5	*Bo3g027080*	*0.4*	*0.3*
PHT1;6	*Bo2g119290*	*0.1*	*0.0*

	Bo9g059050	*0.0*	*0.0*
PHT1;7	*Bo4g119740*	*0.0*	*0.0*
	Bo8g084860	*0.5*	*0.0*
PHT1;8	*Bo5g030730/40*	*0.0*	*0.0*
	Bo5g030750	*0.0*	*0.0*
	Bo8g070440	*18.3*	*0.1*
PHT1;9	*Bo2g089910*	*0.4*	*6.0*
	Bo6g084900	*5.8*	*0.0*
	Bo6g120480/70	*0.1*	*0.0*
	Bo6g120500	*12.7*	*0.0*
	Bo6g120540	*0.6*	*0.0*

Ten of the *PHT1* homologues were found to have expression above an average of 10 FPKM in either the root or leaf tissues and all of these putative genes had significantly ($P<0.005$) more expression in the root tissue than the leaf tissue samples (Figure 4.5).

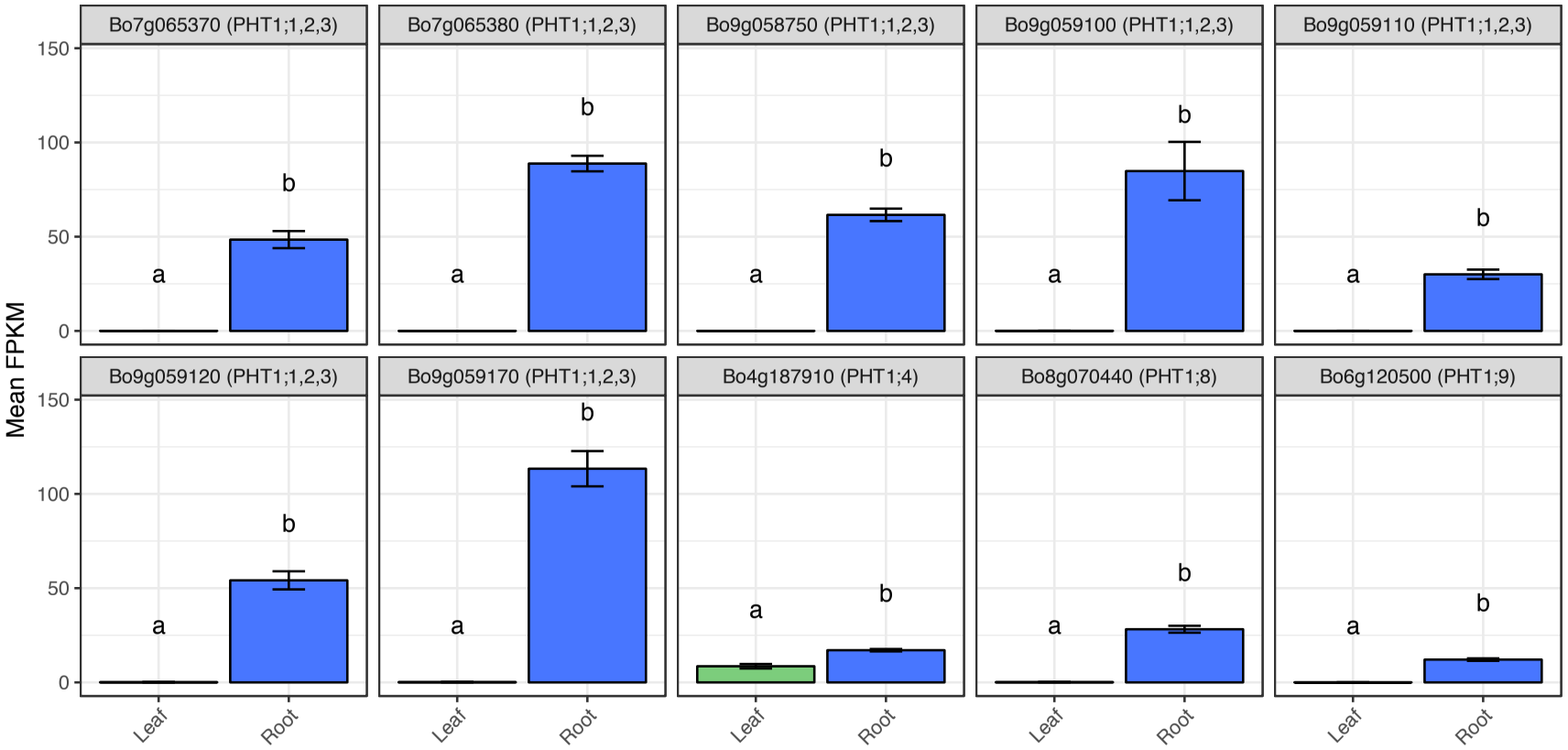


Figure 4.5 Bar plots displaying the mean FPKM values for the putative *PHT1* genes in the root (blue) and leaf (green) tissues of the total *Brassica* diversity set. The gene of highest similarity is displayed in brackets. Error bars are standard error of the mean. Leaf and root with compartments with different letters have significantly different (P<0.05) expression based on a Kruskall Wallace test.

Seven homologues were identified that showed high similarity to *AtPHT1;1*, *AtPHT1;2* and *AtPHT1;3*. The high degree of similarity between these sequences prevents higher resolution identification. The predicted genes had mean FPKM values ranging from 29.9 to 111.9 in the roots, and mean FPKM values in leaves ranging from 0.04 to 0.2, with fold increases in roots relative to leaves ranging from 481x to 2345x (Figure 4.5). Of the identified genes, one was classified as a homologue for *AtPHT1;4*, Bo4g187910. The mean expression of this predicted gene in roots was 16.9 FPKM, which was 2-fold higher than the leaf samples, and was the only putative gene with leaf expression over an average of 5 FPKM. One homologue of *AtPHT1;8* was identified, Bo8g070440, the expression of which in the leaf was 0.14 FPKM, and the expression in the root tissue was 201-fold higher at 28.2 FPKM. One homologue of *AtPHT1;9* was identified, Bo6g120500, which had an expression level in the roots of 11.9 FPKM, with 1161-fold higher expression in roots relative to leaf tissue.

4.3.4 Expression of PHT genes in the roots of wild, cultivated and wild B. oleracea

Differences in expression pattern for the predicted *PHT1* genes between cultivated and wild lines were tested for significance using an ANOSIM test and visualized using NMDS (Figure 4.6). Significant differences in expression patterns for the putative PHT1 homologues were identified between 32 pairs of wild species or crop types (Table 4.5), (Figure 4.6A). Only wild species or crop types with at least 3 lines were included in ANOSIM analysis, additionally, a threshold of a minimum of 100 permutations between paired wild species or crop types was required for inclusion in the ANOSIM results.

Using the cultivated, wild oleracea and wild groupings, the cultivated category was found to have significantly (P=0.001) different expression profiles for the predicted PHT homologues than both the wild *B. oleracea*

and wild categories (Figure 4.6B). The predicted PHT expression profiles for the wild and wild *B. oleracea* groups were not significantly different (P=0.057).

Table 4.5 Pairwise ANOSIM results for putative PHT gene expression for wild species or crop types with at least 3 lines and, comparisons containing at least 100 permutations

Group 1	Group 2	R statistic	Significance level (P)
B. macrocarpa	Broccoli	0.769	0.001
B. incana	*B. macrocarpa*	0.357	0.001
B. macrocarpa	Cauliflower	0.804	0.001
B. macrocarpa	*B. villosa*	0.539	0.001
B. incana	Cauliflower	0.308	0.001
B. macrocarpa	Cabbage	0.718	0.001
B. villosa	Cauliflower	0.439	0.001
B. incana	Broccoli	0.264	0.002
B. villosa	Broccoli	0.381	0.002
B. macrocarpa	Kohlrabi	0.533	0.005
B. cretica	Cauliflower	0.302	0.006
Broccoli	Brussels sprouts	0.523	0.006
B. macrocarpa	Brussels sprouts	0.594	0.008
Brussels sprouts	Cabbage	0.52	0.01
B.villosa	Cabbage	0.354	0.011
B.montana	Broccoli	0.585	0.011
B. oleracea	Broccoli	0.206	0.012
B.cretica	Broccoli	0.306	0.017
Brussels sprouts	Cauliflower	0.47	0.017
B. montana	Cauliflower	0.533	0.018
B. montana	*B. villosa*	0.54	0.018
B. hilarionis	*B. macrocarpa*	0.483	0.02
B. cretica	*B. villosa*	0.217	0.023
B. oleracea	Cauliflower	0.189	0.027
B. incana	*B. villosa*	0.172	0.033

B. cretica	*B. macrocarpa*	0.279	0.034
B. cretica	Cabbage	0.237	0.035
B. rupestris	Broccoli	0.436	0.036
B. macrocarpa	*B. montana*	0.418	0.039
B.villosa	Brussels sprouts	0.408	0.04
B. rupestris	*B. villosa*	0.352	0.048

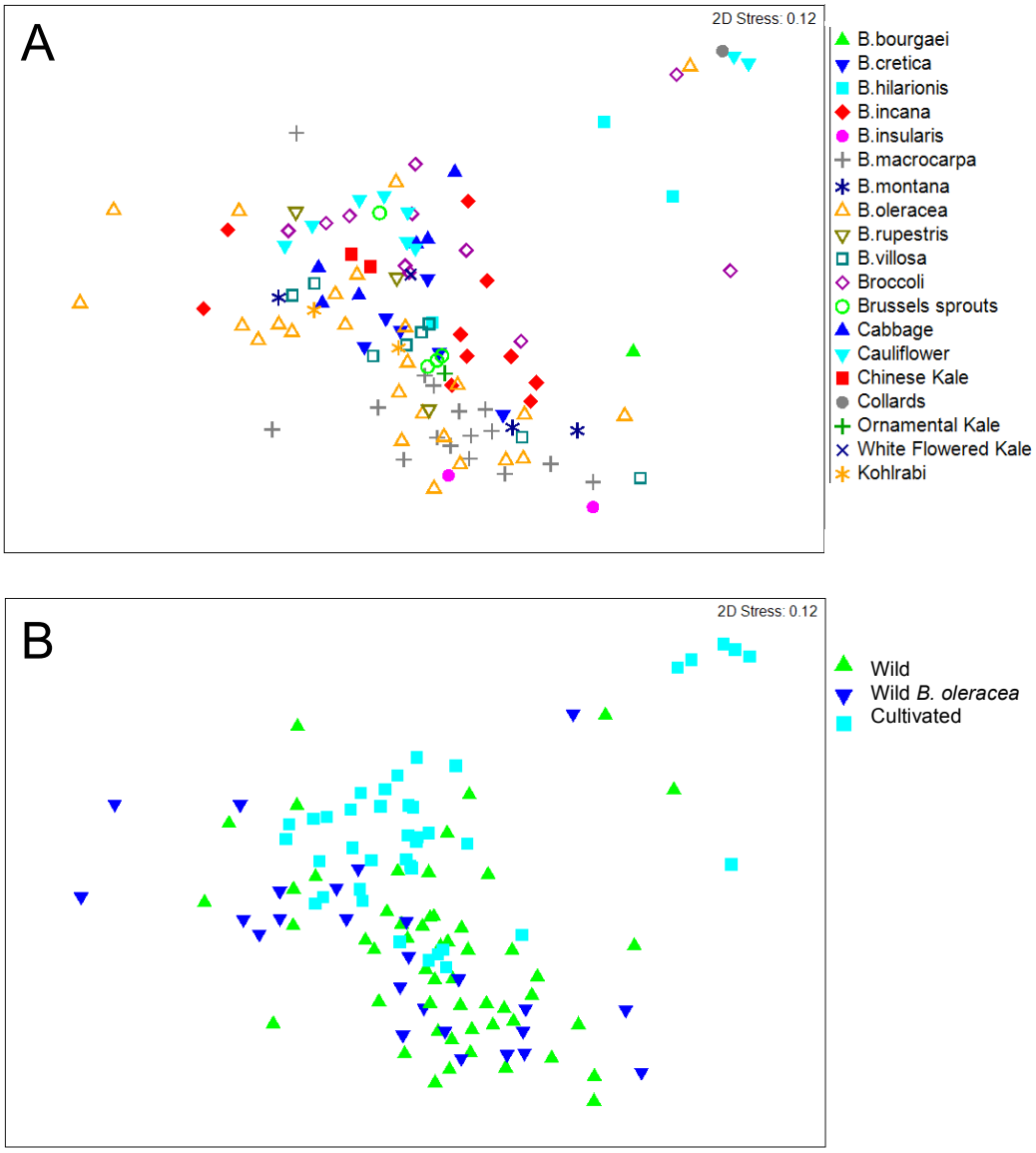


Figure 4.6 Figure 2: Nonmetric Multi-Dimensional Scaling (NMDS) plots displaying the Bray-Curtis similarity of root transcriptome profile for the putative *PHT1* genes for each line. Coloured by A.) Wild species or crop types and B.) Category

Of the 10 *PHT1* homologues showing differential expression between root and leaf tissue, 8 showed differential expression between wild, cultivated and wild *B. oleracea* groups (Figure 4.7). Bo7g065370, Bo9g058750 and Bo9g059100 showed significantly higher expression in the cultivated lines than the wild *B. oleracea* and wild lines. The average fold increase in expression between the cultivated and wild *B. oleracea* lines, was 2.6x, 1.5x and 2.2 respectively. The average fold increase in expression between the cultivated and wild lines, was 2.3x, 1.4x and 1.8 respectively. In contrast the putative genes Bo9g059110, Bo9g059120 and Bo9g059170 had significantly higher expression in the wild and wild *B. oleracea* lines, than the cultivated lines. The differences in expression level were indicated by fold level FPKM increases in the wild *B. oleracea* relative to the cultivated lines of 2.3x, 2.9x and 2.7x respectively. The average fold increase in expression between the wild and cultivated lines was 2.2x, 2.9x and 12.6 respectively.

The expression of Bo7g065380 was significantly higher in the cultivated and wild *B. oleracea* lines than the wild *B. oleracea* lines. The expression was 1.4x higher in cultivated than wild *B. oleracea*, and 1.4x higher in wild than wild *B. oleracea* (Figure 4.7). The mean FPKM expression values for the *AtPHT1;4* homologue, Bo4g187910 did not significantly differ between groups (Figure 4.7). The *AtPHT1;8* homologue Bo8g070440 was significantly different between the three groups. Similarly the mean FPKM expression value of the *AtPHT1;9* homologue Bo6g120500 did not significantly differ between groups.

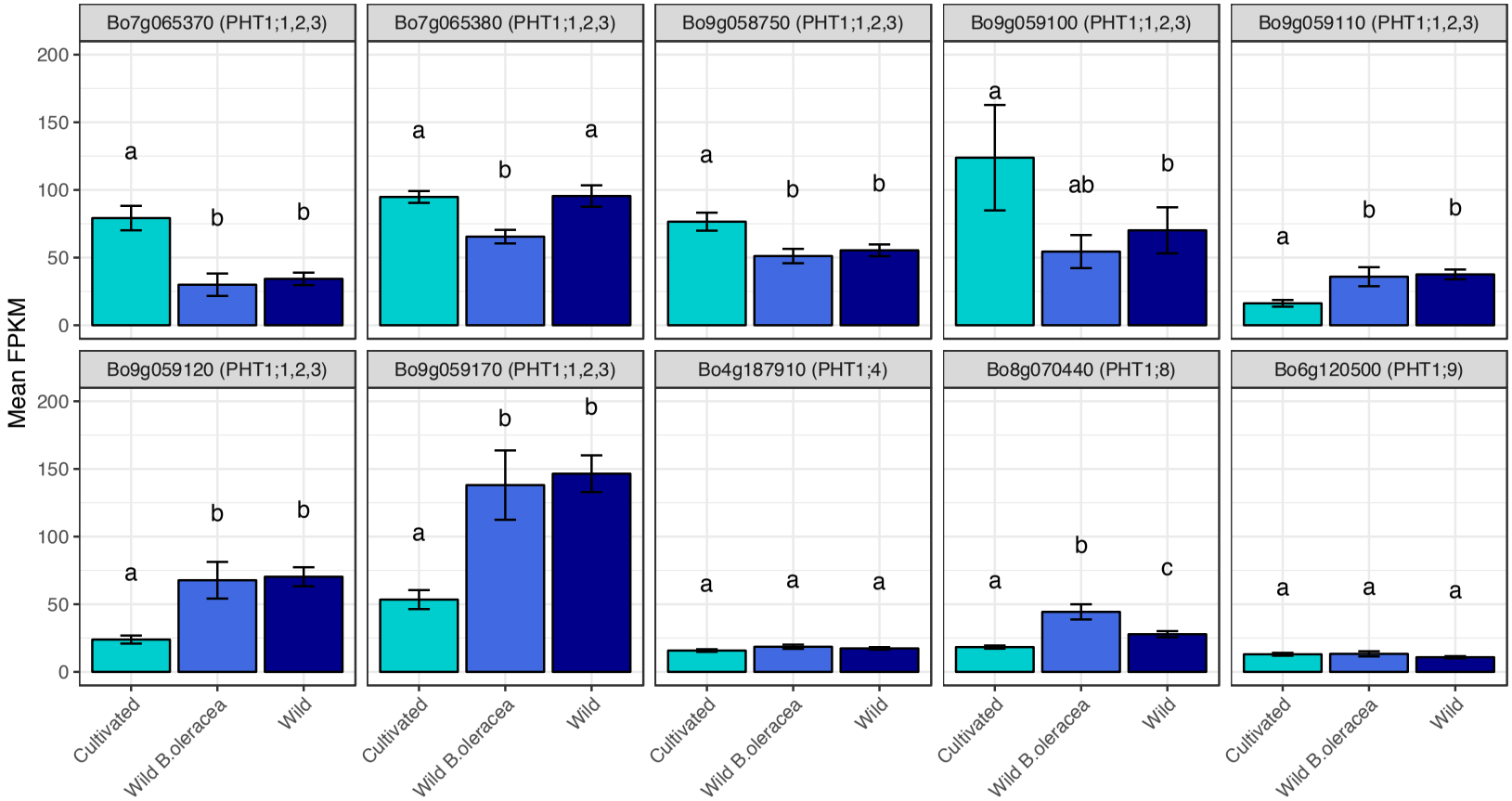
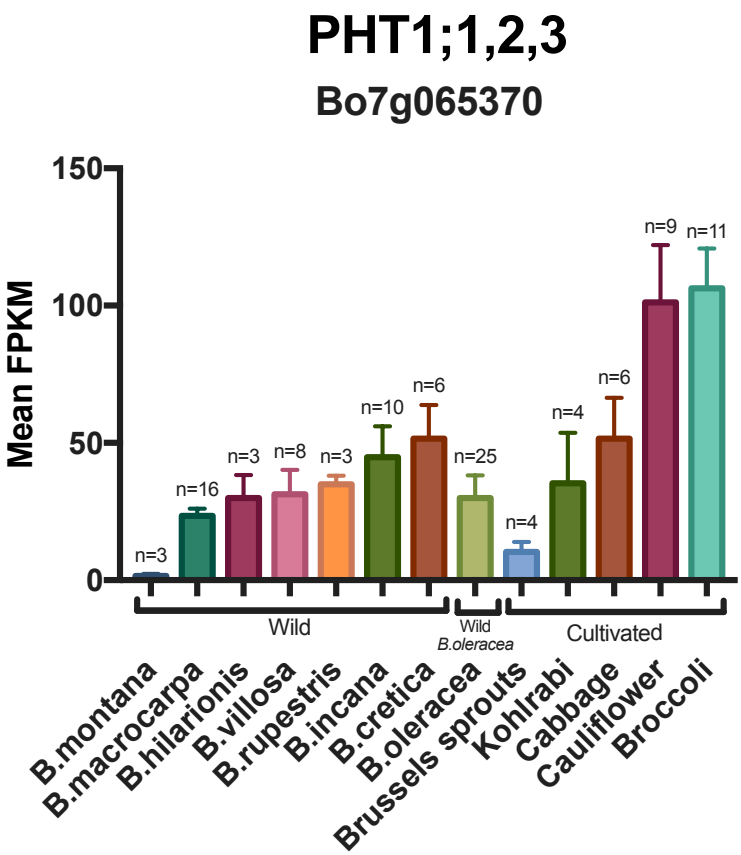


Figure 4.7 Mean FPKM expression values in the roots for cultivated, wild *B. oleracea* and wild lines. The gene of highest similarity is displayed in brackets. Error bars are standard error of the mean. For each putative gene, category of *Brassica* which do not share a letter are significantly different (P<0.05) based on a Dunn test.

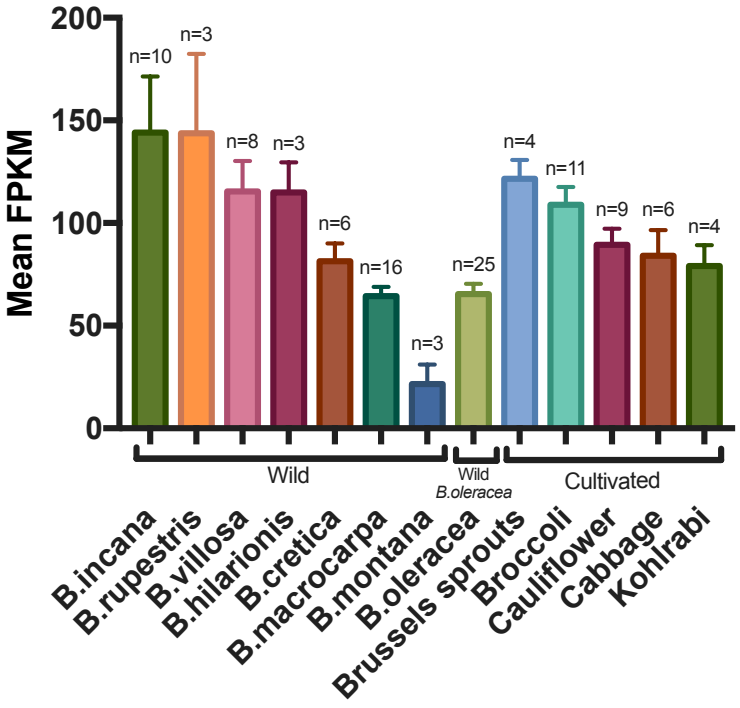
4.3.5 Differences in PHT1 expression for individual species or types between and within Brassica categories

Differences in expression levels were analysed between the individual *Brassica* species/types with at least 3 lines (Figure 4.7). And compared ***between*** species/types from different categories. The binning of the lines into three distinct categories; wild, wild oleracea and cultivated has been justified using phylogenetic trees (Figure 4.1) and similarities of the whole root transcriptome (Figure 4.3). However, the individual expression profiles of species or types can also differ ***within*** categories (Figure 4.8).



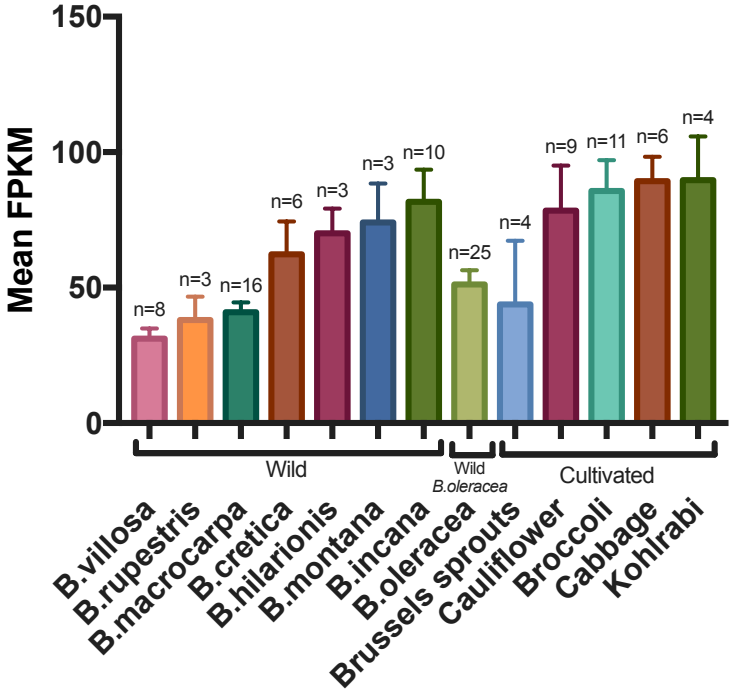
PHT1;1,2,3

Bo7g065380



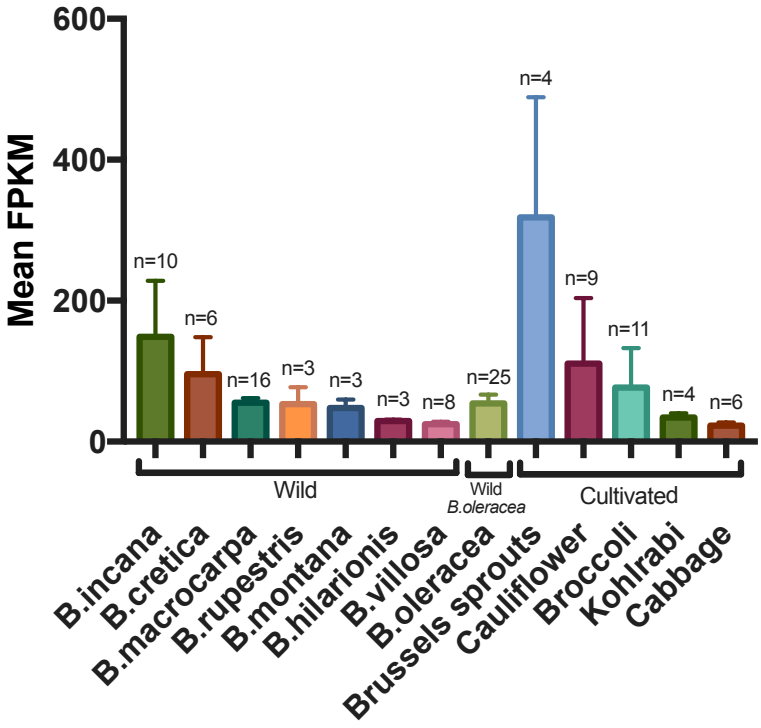
PHT1;1,2,3

Bo9g058750



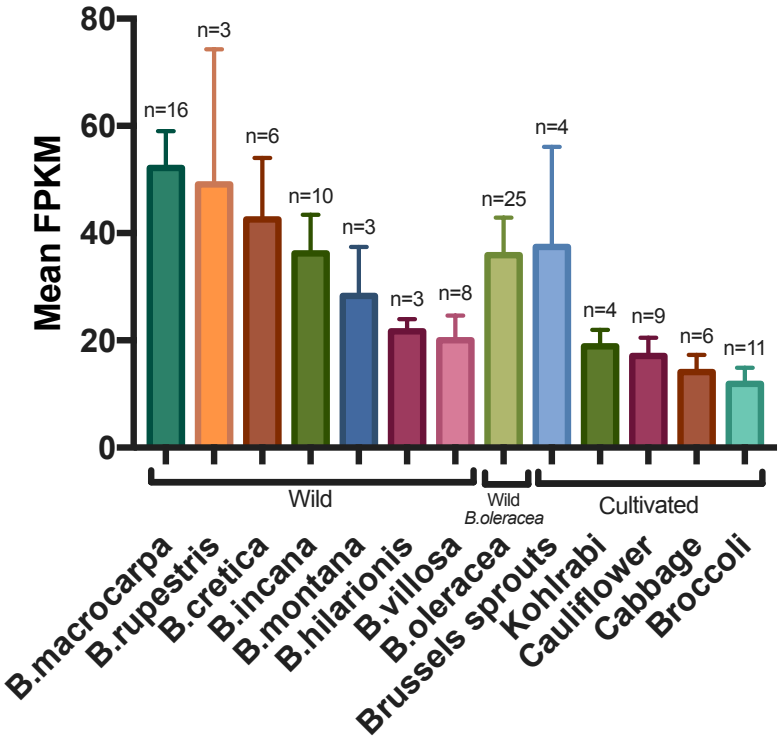
PHT1;1,2,3

Bo9g059100

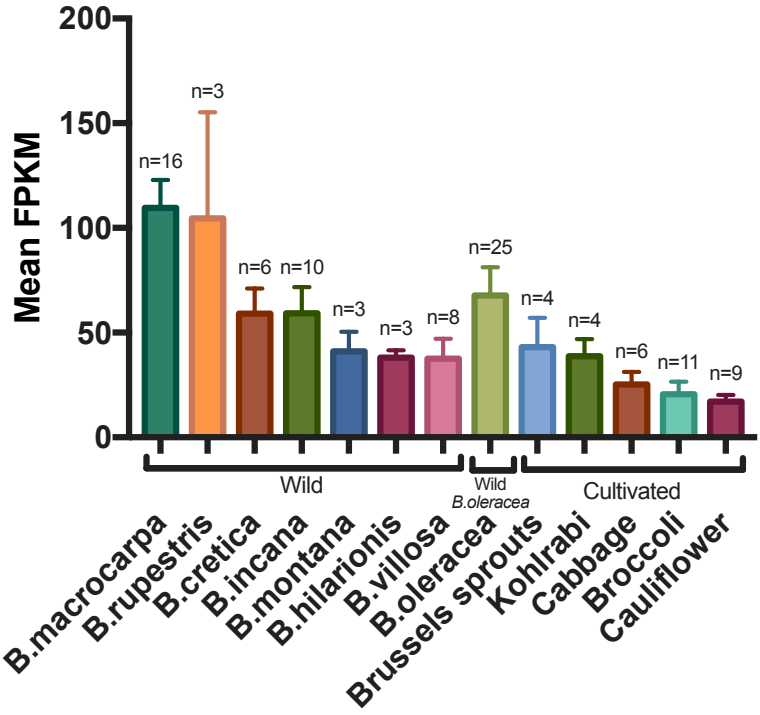


PHT1;1,2,3

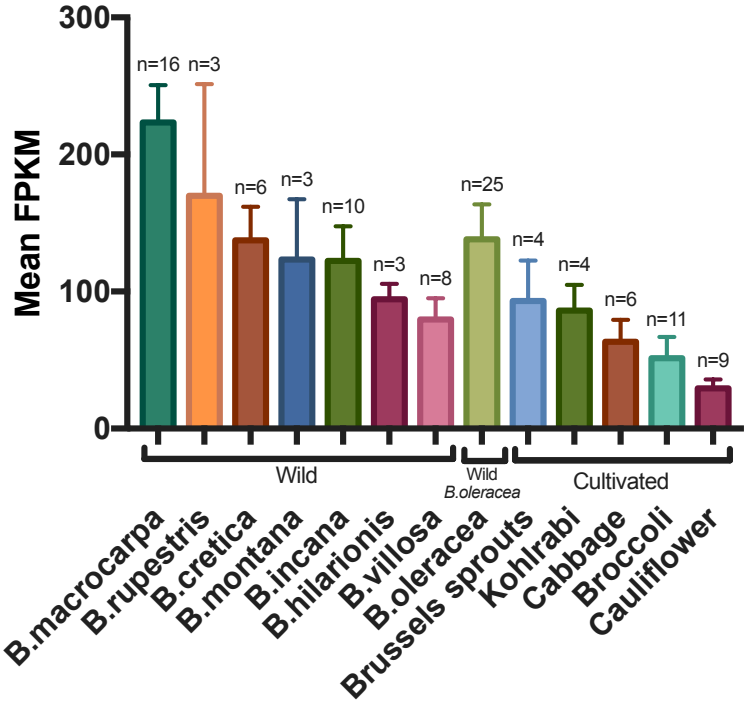
Bo9g059110



PHT1;1,2,3
Bo9g059120



PHT1;1,2,3
Bo9g059170



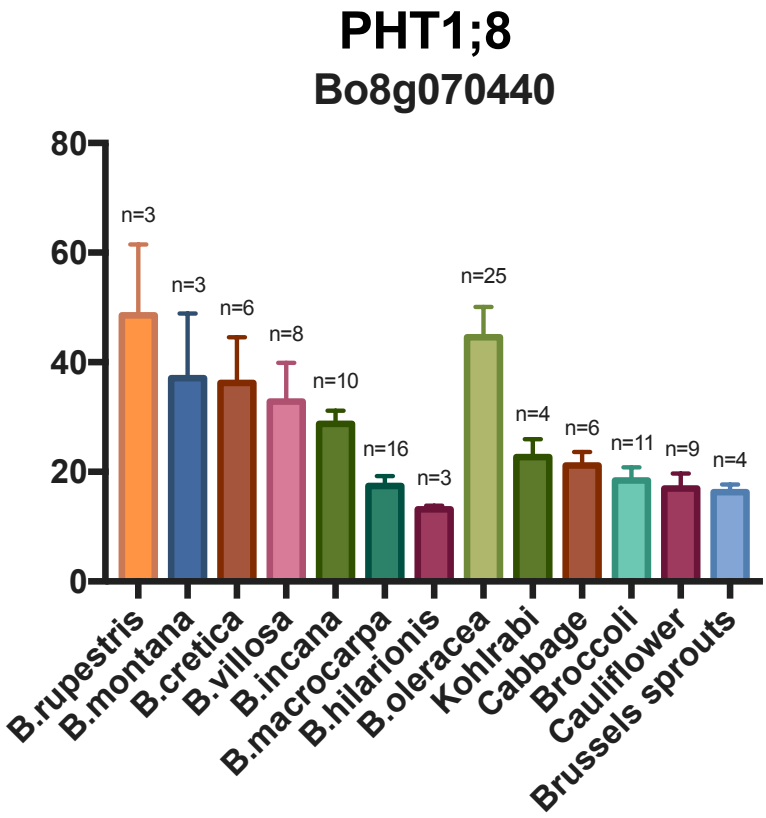


Figure 4.8 Expression of *B. oleracea PHT1* homologues for each wild species or crop type with 3 or more samples, Error bars represent standard error of the mean. N is number of individual lines.

Bo7g065370 displayed differences in expression of species/types between categories; The expression of Bo7g065370 was significantly lower in *B. macrocarpa* (23 FPKM), *B. montana* (2 FPKM) and *B. oleracea* (30 FPKM) than both broccoli (106 FPKM) and cauliflower (101 FPKM). Additionally expression of *B. Montana* was significantly lower than cabbage (52 FPKM) (Figure 9). Bo7g065370 also displayed differences in expression of species/types within categories; The expression of Bo7g065370 was significantly greater in broccoli (106 FPKM) and cauliflower (101 FPKM) than in Brussels sprouts (10 FPKM). The expression of Bo7g065370 was also significantly greater in *B. cretica* (52 FPKM) and *B. incana* (45 FPKM) than in *B. montana* (2 FPKM). Bo7g065380 displayed differences in expression of species/types between categories; The expression of Bo7g065380 was significantly lower in *B. montana* (22 FPKM) than broccoli (109 FPKM) and Brussels sprouts (122 FPKM). Additionally the

expression of Bo7g065380 *B. macrocarpa* (64 FPKM) and *B. oleracea* (65 FPKM) was significantly less than in broccoli. Bo7g065380 also displayed differences in expression of species/types within categories; The expression of Bo7g065380 was significantly lower in *B.montana* (22 FPKM) than *B. incana* (144 FPKM)*, B. rupestris* (144 FPKM) or *B.macrocarpa* (64 FPKM). Bo9g058750 displayed differences in expression of species/types between categories; The expression of Bo9g058750 was significantly lower in *B. villosa* (31 FPKM) and *B. macrocarpa* (41 FPKM) than in broccoli (86 FPKM) and cabbage (89 FPKM). Bo9g058750 also displayed differences in expression of species/types within categories; The expression of Bo9g058750 was significantly greater in *B. incana* (82 FPKM) than in *B.vilosa* (31 FPKM).

The expression of Bo9g059100, Bo9g059120, Bo9g059110, Bo9g059170 and Bo8g070440 were not significantly different between any wild species or crop types from within the same category. The expression of Bo9g059110 was significantly lower in broccoli (12 FPKM) than in *B. macrocarpa* (52 FPKM). The expression of Bo9g059120 was significantly greater in *B. macrocarpa* (110 FPKM) than in broccoli (21 FPKM), cabbage (25 FPKM) or cauliflower (17 FPKM). The expression of Bo9g059170 was significantly greater in *B. macrocarpa* (223 FPKM) than cauliflower (29 FPKM). The expression of the *PHT1;8* homologue Bo8g070440 was significantly greater in *B. oleracea* (45 FPKM) than in *B.macrocarpa* (17 FPKM)*,* broccoli (18 FPKM) and cauliflower (17 FPKM)

4.4 Discussion

The aims of this work were to 1.) Analyse differences in overall root expression profiles of wild and cultivated lines, 2.) Identify potential homologues of the *PHT1* genes in *Brassica oleracea* and 3.) Determine whether there are differences in PHT expression patterns between wild, wild *B. oleracea* and cultivated lines.

4.4.1 Differences in Brassica root transcriptomes

The root and leaf transcriptomes from 119 different lines of *B. oleracea* were analysed. The lines were grouped into cultivated, wild *B. oleracea* and wild based on unpublished phylogenetic research (Moore *et al*., 2017 personal communication). The variability in the in root transcriptional profiles of both the wild and wild *B. oleracea* lines was much greater than the cultivated *B. oleracea* lines (Figure 4.2). The genetic diversity of crop wild relatives has been of increasing interest to plant breeders in recent years (Dempewolf, 2017). Wild relatives of crops can be used as a source of alleles to enhance the resilience of crops to stressful conditions (Schröder, 2013). Crossing with wild lines can improve the agricultural value of cultivated lines. For example increases in glucosinolate production in broccoli have been generated from crosses with wild *B.villosa* relatives (Sarikamis, 2006). Given the known prominent differences between the phenotypes of the crop types, the tight clustering of root transcriptional fingerprints compared to the wild varieties (Figure 4.2) highlights the lack of transcriptional diversity within the cultivated lines, despite the *B. oleracea* crop group representing a highly diverse cultivated species. The wild lines could offer breeding potential to introduce traits, which have been bred out of the cultivated lines as crops have been selected to perform in modern high-input agricultural practices.

4.4.2 Identification of Putative PHT1 genes in B. oleracea

Stringent BLAST searches and sequence alignment were used to identify putative *B. oleracea PHT1* genes with homology to *A. thaliana* and *B. napus* genes (Table 4.3). For each of the *AtPHT1* genes there was variation in the number of hits in the *B. oleracea* genome (Table 4.3). For *AtPHT1;1*, *AtPHT1;2* and *AtPHT1;3* there were eleven genes with high similarity. For *AtPHT1;4* there were six genes with high similarity, for *AtPHT1;5* there was one gene with high similarity, for *AtPHT1;6* there were two genes with high similarity, for *AtPHT1;7* there were four genes

with high similarity, for *AtPHT1;8* there were two genes with high similarity and for *AtPHT1;9* there were four genes with high similarity.
Some other plant species have a greater number of *PHT1* genes than *A. thaliana*. For example thirteen PHT genes have been identified in rice (Teng *et al*., 2017) and at least two of these are involved in Pi uptake mediated through mycorhhizal symbioses (Yang *et al*., 2012). In this study some level of expression was detected for 18 possible *PHT1* genes (Table 4.4). The increased number of potential *PHT1* genes in *B. oleracea* may indicate a number of *PHT1* genes with novel functions, which have been generated over time through gene duplication and subsequent diversification of function.

A key issue which determines the ability to find homologous genes is the quality of the annotations of the genome. Exon-intron structure is typically conserved for orthologous genes (Rogozin *et al*., 2003). The automatic annotation of coding sequences from genomic sequence has lead to a number of irregular intron-exon structures in the *B. oleracea* genome (Supplementary Figure 3). Changes will be made to annotations following validation against expressed gene sequences.

4.4.3 Hypothesising function of putative PHT1 genes.

PHT1 genes are an important family of phosphate transporters in plants, which are responsible for the acquisition of inorganic P from the soil solution. Studies in the model plant *A. thaliana* have attributed distinct functions for members of this family (Supplementary Figure 2). *PHT1;1*, *PHT1;2*, *PHT1;3* and *PHT1;4* have all been seen to be expressed in the roots of *A. thaliana* (Mudge *et al*., 2002). Mudge and colleagues also determined that all four were up-regulated under P stress and *PHT1;1* was constituently expressed. Later research by Shin and coworkers identified the substantial roles of PHT1;1 and PHT1;4 in P uptake under P stress (Shin *et al*., 2004). Further studies showed that *PHT1;2* is constituently expressed (Lapis-Gaza *et al*., 2014), and was not up-regulated under Pi

stress after 2 days. However the expression of *PHT1;1*, *PHT1;3* and *PHT1*;4 was less than the expression of the reference genes under P replete conditions and were then up-regulated during P stress. *PHT1;1* and *PHT1;2* have an almost identical sequence structure (99% sequence similarity), the high sequence similarity of these genes (Ayadi *et al*., 2015) makes a unique identification using transcriptomic approaches difficult. The origin of these genes was probably a gene duplication event (Flagel and Wendel, 2009) that occurred prior to the split of *Arabidopsis* and *Brassica*.

Here, expression was investigated for putative *B. oleracea* genes that displayed high similarity to *A. thaliana PHT1* genes. Seven putative *B. oleracea* genes which had a high similarity to *A.thaliana PHT1;1*, *PHT1;2* and *PHT1;3* genes were seen to be strongly expressed in the roots. Whether the *PHT1* gene function is similar in *B. oleracea* would require further study, but the localized expression in the roots suggest that the Bo7g065380, Bo7g065380, Bo9g058750, B09g0591900, B09g0591910, B09g0591920 and B09g0591970 genes might be involved in P uptake from the soil in a similar functional role to the *A. thaliana PHT1;1*, *PHT1;2* and *PHT1;3* genes. Four putative *B. oleracea* sequences with similarity to *A. thaliana PHT1;4* were identified and one (Bo4g187910) was seen to be expressed in both the root and leaf tissues above background noise levels. The expression of Bo4g187910 was significantly greater in the root tissue than the leaf tissue, suggesting possible similarities to A. thaliana *PHT1;4* (Shin *et al*., 2004), in accessing external P from both low high P environments.

In *A. thaliana* expression of *PHT1;5*, *PHT1;6* and *PHT1;7* have been seen in senescing leaves, pollen grains and Pi deficient leaves respectively. In accordance with this, no high level expression of the putative *B. oleracea PHT1;5*, *PHT1;6* and *PHT1;7* sequences (Bo3g027080, Bo2g119290, Bo9g059050, Bo4g119740 and Bo8g084860) was detected in either root or leaf tissues. *AtPHT1;8* has been shown to have a role in the acquisition of P from the soil (Remy *et al*., 2012) and subsequently

roles in the root to shoot translocation of P (Lapis-Gaza *et al*., 2014). Here, Bo8g070440 was identified as a potential *AtPHT1;8* homologue. The expression of this gene exclusively in the root and not the leaf, supports the hypothesis that the gene may have a similar role to *A. thaliana PHT1;8*, by transporting P from the root endodermis to the xylem (Lapis-Gaza *et al*., 2014). In *A. thaliana PHT1;9* is believed to mainly function in the roots (Lapis-Gaza *et al*., 2014), although low level of expression was located during leaf senescence (Remy *et al*., 2012).Three potential *PHT1;9* genes were identified with some level of expression. Bo2g089910 had low expression in the leaf and root, whereas expression of Bo6g084900 and Bo6g120500 was seen at low levels in the root.
For some putative *B. oleracea* genes, which had high similarity to *A. thalina PHT1* genes there was low level expression (<10FPKM) in root or leaf tissue (Table 4.4). Whilst this may be sequencing noise, these genes could also have roles in P uptake under P limiting conditions. Low level expression was identified in the root tissue for potential PHT1;1, PHT1;2 or PHT1;3 genes Bo2g119310, Bo2g119320, Bo2g119350 and Bo7g065360. Low level expression was seen in both root and leaf tissue was identified for the possible PHT1;4 genes Bo4g026830 Bo4g026840, Bo4g026850 . Low level expression was identified for potential *PHT1;9* the root tissue gene Bo6g084900 and in the leaf tissue for Bo2g089910.

4.4.4 Differences in PHT1 expression between wild, cultivated and wild B. oleracea categories

The expression patterns of the putative *PHT1;1*, *PHT1;2* and *PHT1;3* homologues in the roots of the *B. oleracea* showed significant differences between wild *Brassica* C genome, wild *B. oleracea* and cultivated categories. The putative Bo9g058750 was identified to have a significantly higher level of expression in the cultivated category than both the wild and wild *B. oleracea* categories (Figure 4.7). In contrast the expression of Bo9g059110, Bo9g059120 and Bo9g059170 putative genes was significantly greater in the wild category than the cultivated category (Figure 4.7). The selective breeding of crop lines has caused a loss of

genetic diversity over time (Fu, 2015). This ‘domestication bottleneck’ has reduced the genetic diversity of the crop genotypes as only a subset of alleles were included in the domestication event. The changes in expression level of the PHT homologues could be one of many changes in *B. oleracea* which has been driven by selection for crop lines and provides an example of a possible unintended effect of crop speciation on the expressional profile of a species. The difference in expression of the Bo9g058750 putative *PHT1* gene from the Bo9g059110, Bo9g059120 and Bo9g059170 genes may be an indication of an unseen change in expression profiles driven by cultivation. These changes may reflect differences between wild and cultivated *Brassica* with regards to P uptake, which may help direct future breeding efforts.

4.4.5 Conclusions

Here, fifteen potential PHT1 genes have been identified in *Brassica oleracea*. Possible classifications have been attributed to these genes based on sequence similarity and alignments of protein coding sequence. The location of the expression of these genes have been initially studied in the root and leaf tissues. In order to test whether these differences in PHT1 expression patterns are indicative of an underlying change in P acquisition strategies between wild lines which grow on the phosphate poor environments and the cultivated lines which have been bred under higher nutrient input, it would be beneficial to test the lines under a P gradient and monitor the response in expression patterns in root tissues coupled with P utilisation efficiency. Genes displaying differential expression patterns between high and low P conditions in wild species and crop types could then be investigated further for regulation characteristics such as signal transduction pathways, P sensing and promoter sequences. Through this, mechanisms for improving P uptake in low soils may be discovered.

Chapter 5. General discussion

5.1.1 General overview of findings

The identity and function of soil and rhizosphere microbiological communities are of great importance to agriculture due to the substantial effects that these microorganisms have on plant health and nutrient cycling. The overall aims of this work were to develop understanding of plant and rhizosphere microbial functions and interactions with the ultimate goal of increasing agricultural efficiency and sustainability.

This work has utilised DNA amplicon sequencing to gain information on the identity of microbial rhizosphere and soil communities and used predictive tools to gain and insight into potential functions of these microorganisms. Transcriptomics have been used to analyse expression of a specific rhizosphere function (P uptake) across a range of plant species. Metatranscriptomics have been used to identify the active rhizosphere root and soil microbial communities and gain an insight into the functional expression of these microbes.

Distinct rhizosphere microbial communities were characterised, which were significantly enriched in Actinobacteria, Proteobacteria, Bacteriodetes, Ascomycota, Chytrydiomycota and the SAR group of protists compared to the soil. Some of the most abundant OTUs in the rhizosphere were *Flavobacterium* (16S OTU 8), *Spongospora subterranea f. sp. subterranea* (18S OTU 28) and *Leptosphaeria maculans* (ITS OTU 2). The ability to determine the identity of microbial communities is especially useful when functions are known of the taxa. For example *Spongospora subterranea* (powdery scab) and *Leptosphaeria maculans* (blackleg) are known pathogens, whilst *Flavobacterium* can have plant growth promoting effects (Kolton *et al*., 2016). The identification of these taxa, and subsequent study of what factors influence their relative abundance in the rhizosphere can provide useful tools when developing

methods to improve crop health, such as the knowledge that increased N levels drive increases of these OTUs.

Nitrogen fertilisation was seen to significantly alter soil fungal and protist communities, but had no significant effect on the soil bacterial community. Interestingly, the community composition of bacteria and fungi was affected by N fertilisation in the rhizosphere. There may be fundamentally different responses to N limitation between bacteria and fungi. It is possible that he bacterial rhizosphere community was indirectly altered by soil N status, through interactions with plant root exudates. Composition and concentration of plant root exudates are known to change under nutrient limiting conditions. The bacterial community may be more responsive to these changes in exudation profiles than the fungi. Of particular interest for potential future work are OTUs that were significantly enriched in the rhizosphere under N treatment and not in the soil such as *Promicromonospora* (16S OTU 17) as it may have been responding to changes in plant exudation profiles rather than changes in N level. Attributing function to taxonomy remains one of the greatest challenges in microbial ecology, but identifying candidates that may have beneficial effects for plants helps refine the number of potential taxa for functional assays and inoculation experiments.

Functional prediction tools were used in order to attempt to link the composition and function of microbial communities. For example using PICRUSt to predict metagenomes, a greater abundance of genes associated with antibiotic production were identified in the rhizosphere under low N, whilst in the soil N did not affect the relative abundance of any predicted genes. The additional nutrient depletion of the environment caused by plant N uptake could have driven an increase competiveness of the microbial community. Another example of shifts in community function was the increase in methane metabolism in the rhizosphere under low N conditions. Possible bacterial taxa were identified as *Methylybium (*OTU 3617 *and Methylotenera mobil*is (OTU 40), which may have been enriched in the low N rhizosphere due to capabilities of accessing organic N

compounds (Kalyuzhnaya *et al*., 2006). Significant differences were identified between rhizosphere and soil functional guilds, such as increase in pathotrophic fungi in the rhizosphere and a corresponding decrease of saprophytic fungi. This represents an expected increase of pathogenic taxa closer to potential host biomass. However no significant effect of N was identified on fungal guild composition. As the FUNGuild database grows, a larger percentage of fungi may have a guild annotation, which could influence these results.

No effect of plant genotype on the microbial community composition or function in the rhizosphere was identified. The influence that the various genotypes can have on the assembly of rhizosphere microbial communities may be diluted by the other, stronger drivers of community composition in a field setting. This may stem from the disruption to soil, which is accompanied with preparation techniques for greenhouse experiments. The sieving and storage of environmental soil may decrease the microbial biodiversity. Additionally, the scale of the rhizosphere and soil zone is much smaller in pot conditions than in the field, with field conditions displaying a larger number of factors that may influence the rhizosphere.

In Chapter 3 metatranscriptomics were used to identify the active microbial community composition and function in the bulk soil, rhizosphere soil and root of *B. napus*. Comparisons are difficult between the work from these two experiments as it was not possible to separate the rhizosphere and root samples in the experiments described in Chapter 2 whereas these compartments were separated in Chapter 3. Additionally, the site and season were different. However some trends can still be observed. The greatest difference between the two data sets is the relative abundance of the Proteobacteria. The mean root and rhizosphere relative abundance of the Proteobacteria RNA was 75%, and in the DNA study it was 32%. Many studies have shown high abundances of Proteobacteria in the rhizosphere (Bulgarelli *et al*., 2012, Mendes *et al*., 2013). This may suggest an underrepresentation of the Proteobacteria in the DNA dataset.

Actinobacteria and Proteobacteria were shown to be the most underrepresented Phyla in DNA studies (Carini *et al*., 2016). Although such a comparison between different environments is problematic, the abundance of the Proteobacteria revealed by RNA-seq more likely reflects the true community and highlights the differences between the methods. In Chapter 3 distinct microbial communities were identified within the root rhizosphere and soil compartments. Many of the taxa that had higher relative abundance in the root relative to the rhizosphere and bulk soil such as *Cellvibrio*, *Agrobacterium*, *Candida* and *Flavobacterium* have known functional roles in the rhizosphere. This includes the ability of some *Cellvibrio* spp. to fix nitrogen (Mergaert 2003), the tumour causing abilities of *Agrobacterium* spp. (Pacurar 2011), *Candida* spp. acting as plant growth promoters (Hilber-Bodmer *et al*., 2017), and the plant growth promoting effect of some *Flavobacterium* spp. (Kolton *et al*., 2016).

In Chapter 3, the impact of RNA preservation methods on the community composition and function of the soil microbiome was also analysed. The freeze-dried method was found to distort the composition and function of the microbial community less than the Lifeguard and Frozen methods and is recommended for future *B. napus* field based metatranscriptomic studies.

The identity of the *PHT1* genes is not known for *Brassica oleracea,* which is an important global plant species with high nutritional value and a diversity of crop varieties. The *Brassica* C genome also comprises a number of wild species that offer a largely unexplored genetically diverse resource. In Chapter 4, candidate *PHT1* genes in *B. oleracea* were identified using homology searches from the closely related *A.thaliana.* A root and leaf transcriptome dataset was used to examine the expression of the predicted *PHT1* genes in root and leaf tissue of a variety of wild *Brassica* species, and cultivated *B. oleracea* types. Potential *PHT1* genes were identified in *Brassica oleracea.* Initial results were validated by active expression of these genes in the root tissues from a large range of lines. Significant differences in expression patterns of these putative *PHT1*

genes were found between wild and cultivated lines. This highlights some of the previously unknown differences between wild and cultivated plan species that have developed through domestication.

5.1.2 Implications of work

Here, methods have been developed for understanding the function and composition of rhizosphere microbial communities of *Brassica napus* in a field setting. This work has also identified potential P uptake genes in the closely related *B. oleracea* and utilised a transcriptome study to identify differences across wild and cultivated varieties.

Although genotype has been shown to have the ability to influence rhizosphere microbiome assembly (Wagner *et al*., 2016), here no genotype effect was identified, even though a diverse range of plant lines were used. This supports the use of field trials over laboratory-based cultivation, especially when studying such dynamic communities. The connectivity of microbial communities in this study was seen to be greater in the soil than in the root, whereas processing methods may alter microbial communities in greenhouse conditions.
Microbial communities of *B. napus* were identified using DNA and RNA methods. Taxa were identified which may have potentially important functional roles in the rhizosphere of *B. napus.* Pathogens with known soil based lifestyles such as *Spongospora subterranea* and *Leptosphaeria maculans* were identified, however the leaf pathogen *Pyrenopeziza brassicae* was also identified in the rhizosphere and the lifestyle of this fungus may need revaluating. Taxa with potentially beneficial functional roles such as *Flavobacterium, Pedobacter* and *Promicromonospora* were identified in the rhizosphere and may be good candidates for developing inoculations, or investigation nutrient cycling.

The whole root transcriptomes of the diverse crop types displayed less variation than the wild species. This interesting result highlights the genetic

potential of the wild species for introducing novel traits through breeding. Identification of differences in expression of vital P uptake genes may also help identify areas for future study, for example developing plant lines suitable for growing under lower P levels.

5.1.3 Future directions

The field of microbial ecology will be enhanced by methods that can better determine the function of rhizosphere microbial communities. Here DNA amplicon sequencing was used to gain information on the identity of microbial rhizosphere and soil communities and was used with predictive tools to create potential metagenomes. Plant transcriptomes were used to analyse expression of a specific rhizosphere process, and metatranscriptomics used to identify the active microbial community and compare microbial functions in soil, rhizosphere and roots Whilst RNA studies do offer a representation of the functional processes of the rhizosphere, the study of proteins offers the closest representation of the functional processes in an environment.

Metaproteomics offers the opportunity to identify all of the expressed proteins in a given environment. Studies of metaproteomes have been conducted in both soil (Bastida *et al*., 2016) and rhizosphere (Wang *et al*., 2011) environments and have revealed a depth of phylogenetic and functional information of microbial communities. However these techniques also come with their own difficulties. The greatest challenge of metaproteomics is extracting a sufficient quantity of pure protein (Wang *et al*., 2016). This challenge is especially difficult in complex environments like soil, and in environments like the rhizosphere where volume of sample material can be difficult to access. Additional issues in metaproteomics are due to the identification of proteins, which is limited by quality of reference databases and high costs.

A limitation of both amplicon based sequencing and shotgun meta –omic approaches is the length of the reads. The short read length (100bp – 400bp) makes assembly of this vast number of often similar sequences into contigs difficult without reference genomes. This process is limited by both the number of microorganisms with full genome sequence available and the accuracy of genome annotations. The annotation of genome sequence, for the most part is conducted using *in silico* methods, that use homology to infer function or genetic organisation and are often inaccurate (Devos and Valencia, 2001). Increasing the number of sequenced organisms and improving the annotation of sequences in databases will help improve the accuracy of sequencing methods (Chen *et al*., 2013).

Advances in computational methods have made it possible to assemble individual genomes from metagenomes (Sangwan *et al*., 2016). Although these methods are computationally demanding, this presents the opportunity to determine the impact of environmental conditions on the functional potential of microbial communities. Long read technologies like Oxford Nanopore (Oikonomopoulos *et al*., 2016) circumvent this problem and allow for a more accurate identification of microbial sequences. The development of this handheld technology may revolutionise agriculture by providing a way to monitor microbial communities in real time and detect incidences of pathogens before they can damage crop yield.

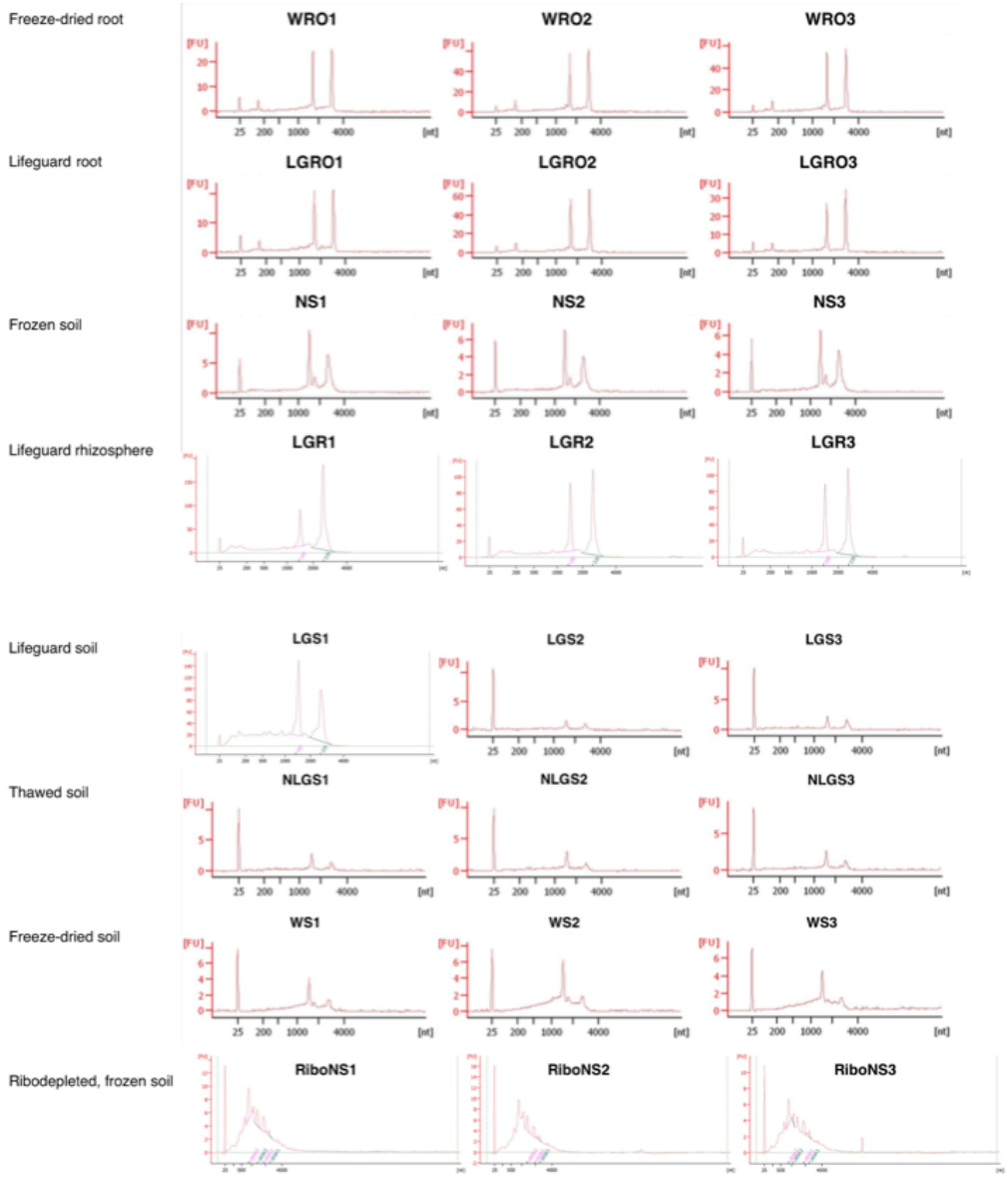
The identification of taxa in Chapter 2 that may be responding to plant exudation patterns, such as the *Promicromionospora* OTU, could inform isolation strategies for rhizosphere bacteria. Utilising isolation and culturing protocols for similar related taxa (Mohammadipanah *et al*., 2014), may facilitate the successful culturing of this potential plant-beneficial organism. Through functional assays and genome sequencing information can be developed on the functional capabilities of this OTU for potential use in inoculations.

The methods developed in Chapter 3 can be applied to field based rhizosphere metatranscriptomic studies. This will allow for investigations

into the factors that influence function of microbial communities on a greater scale. There is still a vast amount of data still available from the RNA-seq data. Comparisons of the ribo-depleted and non-ribodepleted samples could be conducted following activation of MEGAN6 Ultimate edition available (Huson *et al*., 2016). This comparison will determine whether it is possible to gain a realistic view of the function of the microbial community without expensive depletion of ribosomal RNA, especially given the great increases in sampling depth. MEGAN 6 would also offer the use of KEGG classification of function and command line tools. Plant RNA sequences had been filtered out for the analysis in Chapter 3. This data contains a wealth of information concerning the functional activity of the plant. A goal for future work is the simultaneous assessment of plant and microbial functional rhizosphere processes, and determining the impact of agricultural management practices on these functional responses.

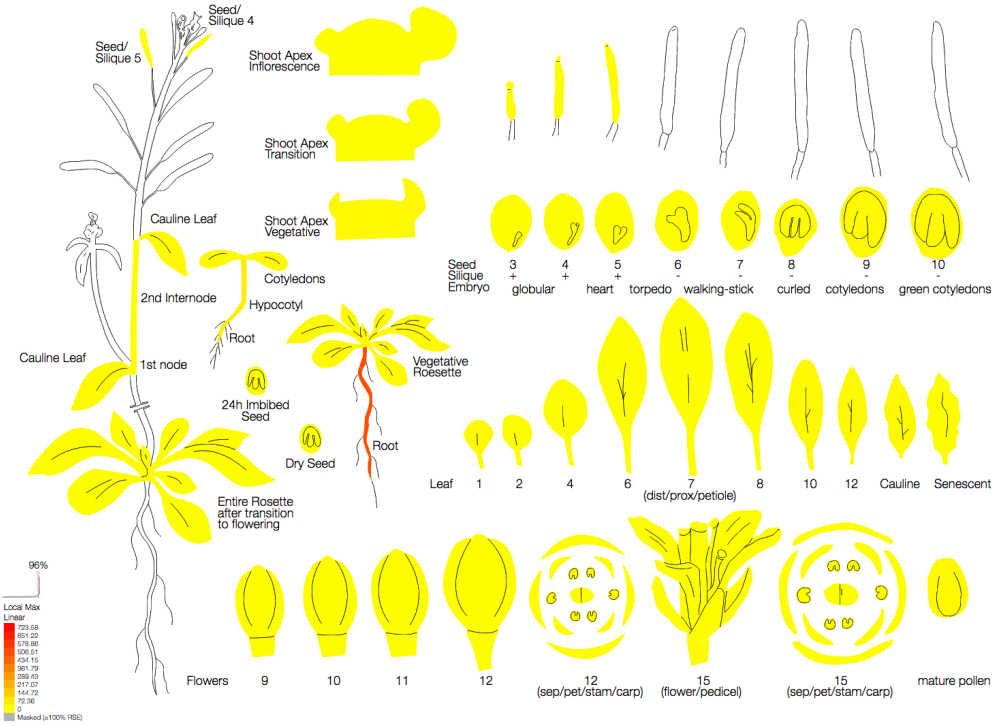
Following on from this work, the putative *B. oleracea PHT1* genes identified from Chapter 4, could be further investigated by conducting qRT-PCR analysis of the putative genes in soil systems under phosphorus replete and deplete conditions. This will allow for a validation of the expression of these putative genes in a soil system, and also detect changes in expression of these genes when the plant is P limited. The differences in expression of these genes between wild species and crop types may offer new breeding targets for development of low input crops.

Supplementary figures

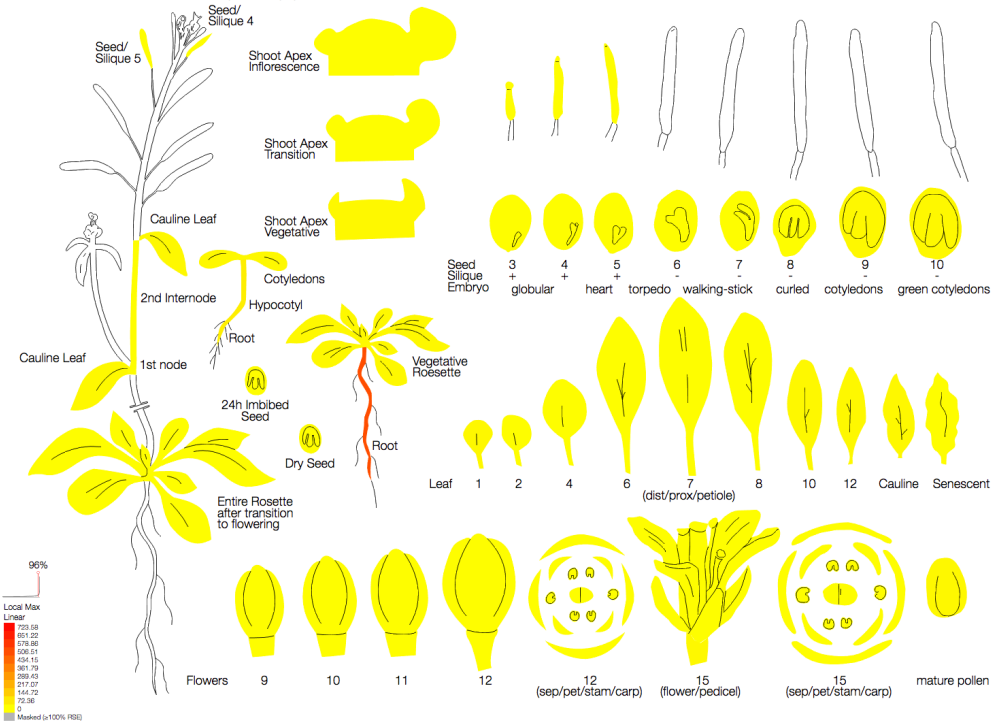


Supplementary Figure 1. Bioanalyser (prokaryotic, nano chip) traces for each sample.

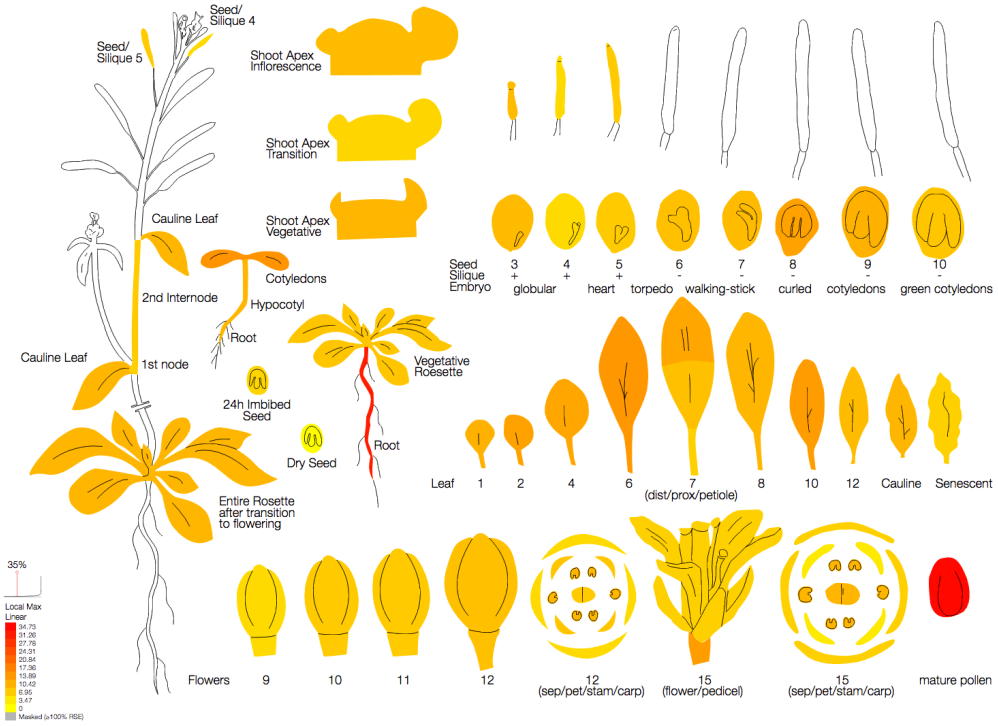
Plant eFP: AT5G43350 / ATPT1, PHT1;1, PT1



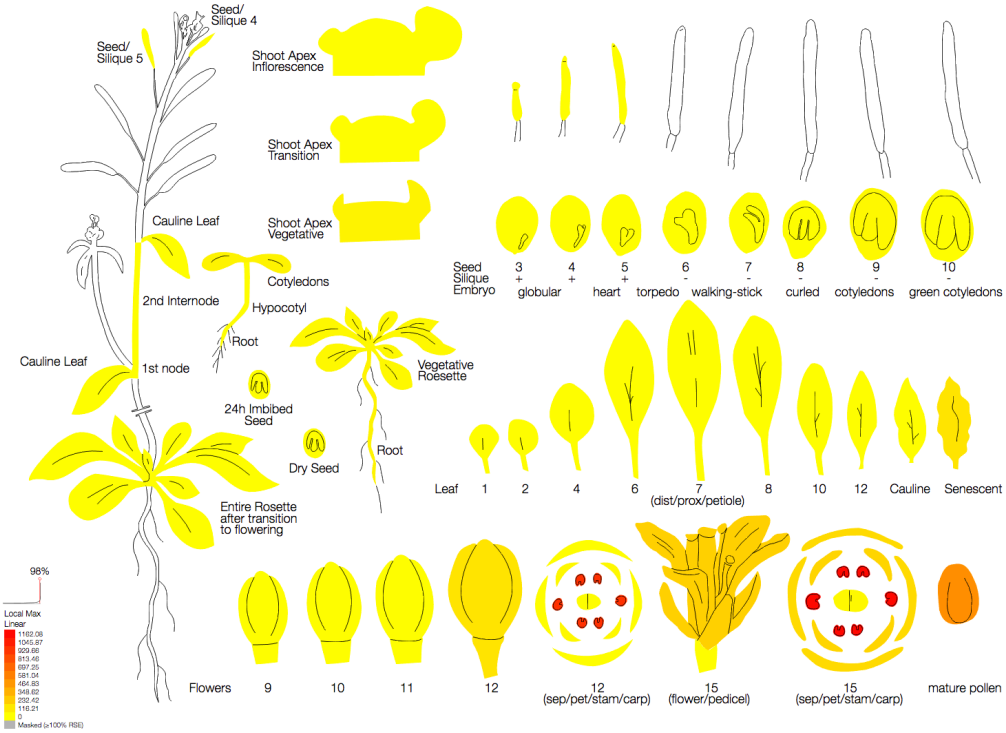
Plant eFP: AT5G43370 / APT1, PHT1;2, PHT2



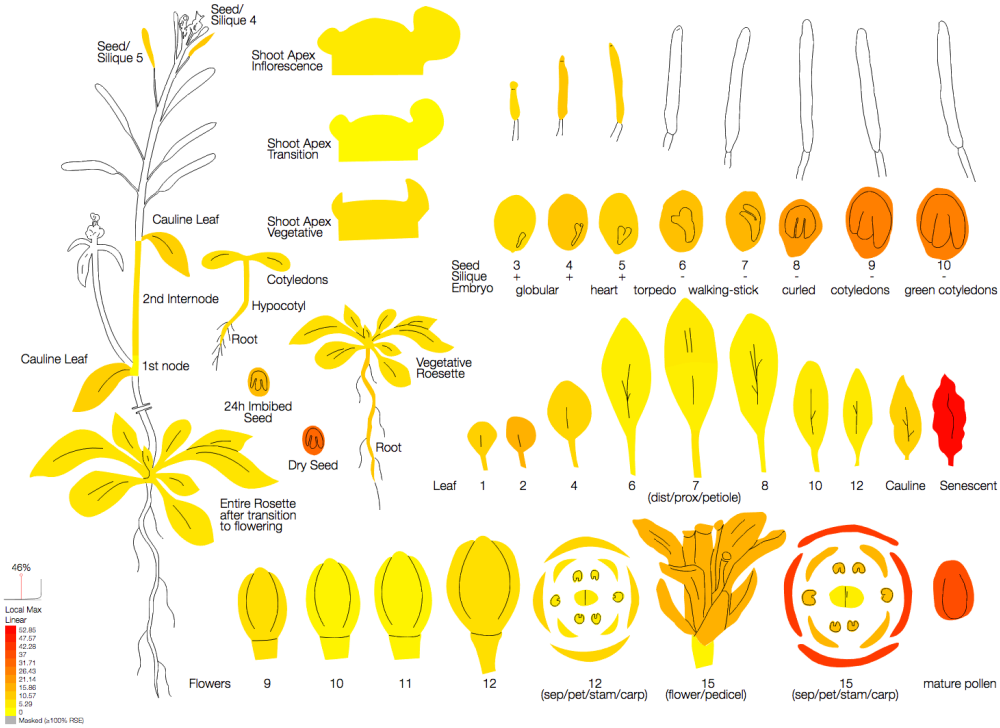
Plant eFP: AT5G43360 / ATPT4, PHT1;3, PHT3



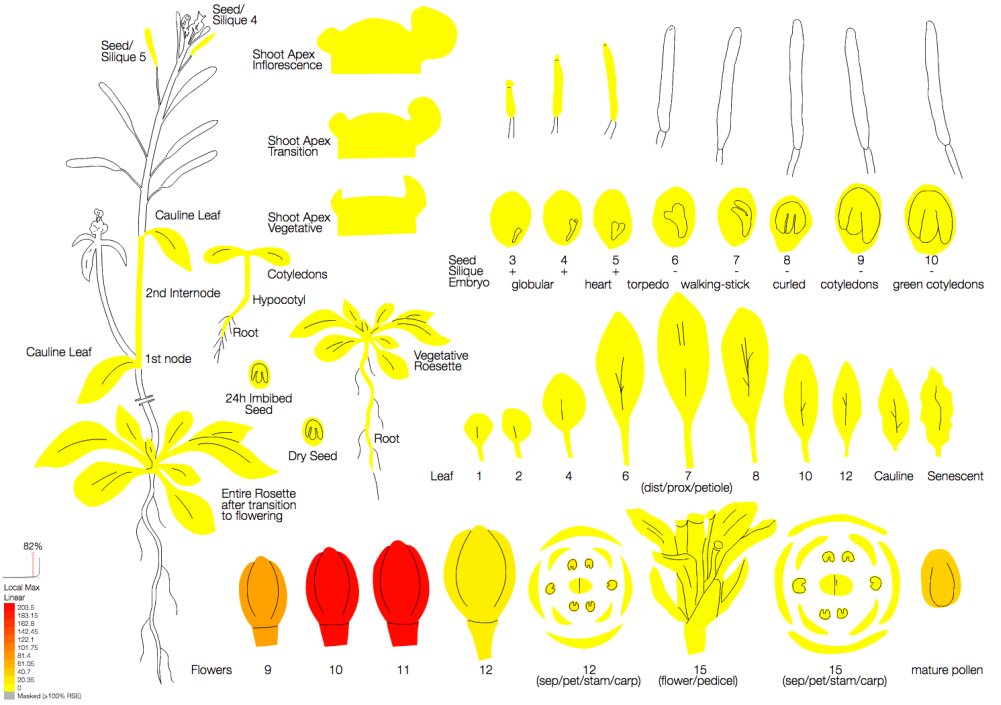
Plant eFP: AT2G38940 / ATPT2, PHT1;4

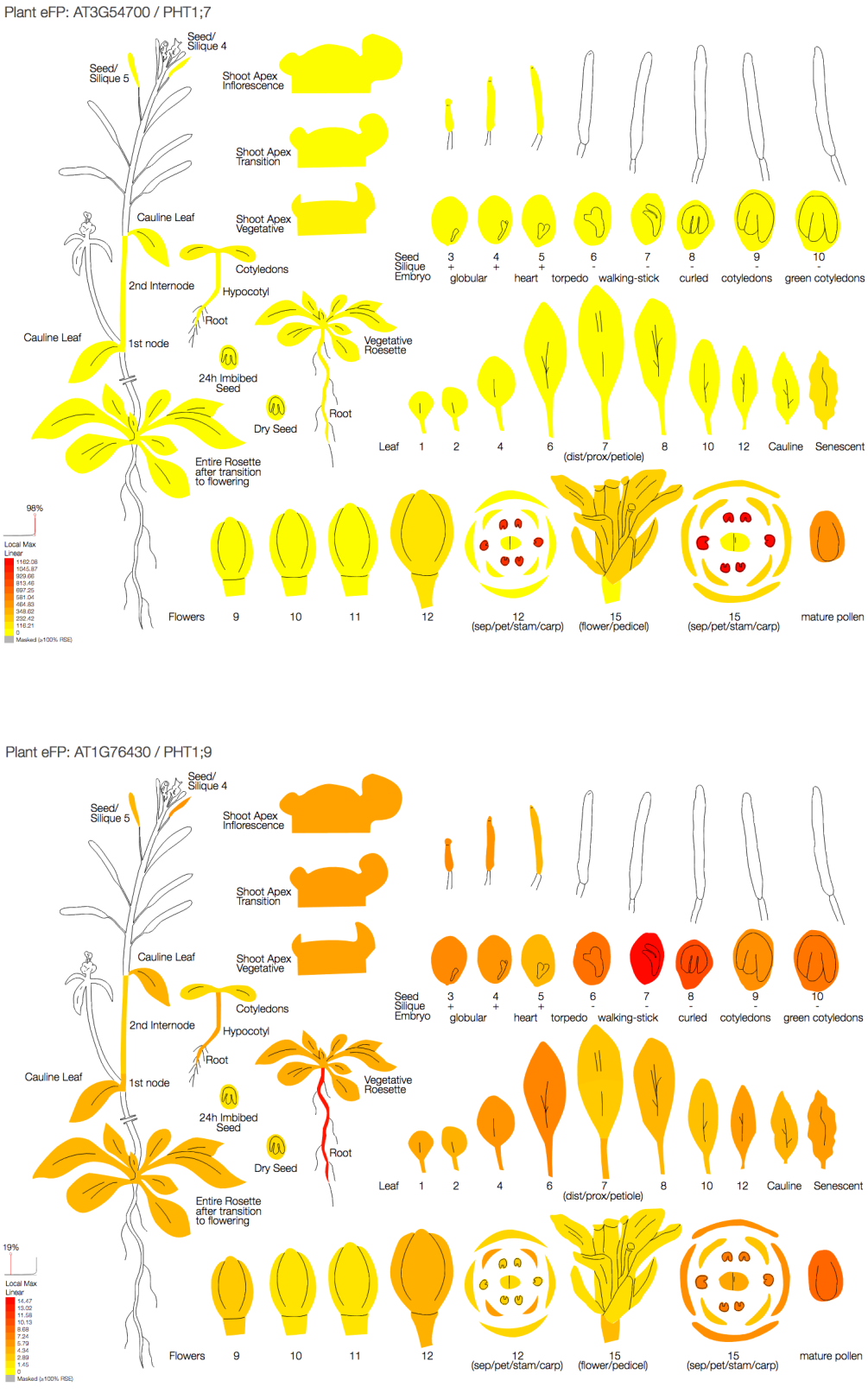


Plant eFP: AT2G32830 / PHT1;5, PHT5

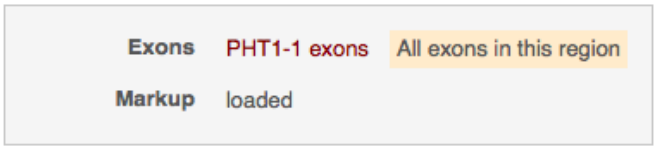


Plant eFP: AT5G43340 / PHT1;6, PHT6

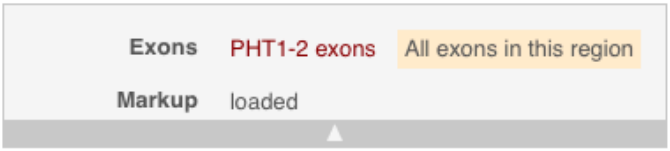




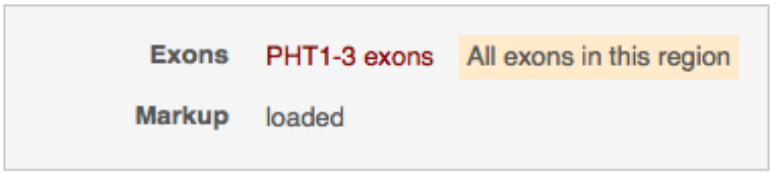
Supplementary Figure 2. Expression of A. *thaliana PHT1* genes in various tissues and growth stages. Data collected from available studies and visualized using using the Plant eFP viewer from the Bio-analytic Resource for Plant Biology (Austin *et al.*, 2016). Expression displayed on a local (relative) linear scale.



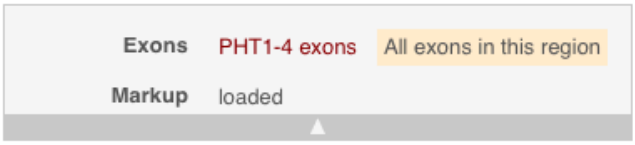
>chromosome:TAIR10:5:17399169:17402758:-1
TTCGACTTATATTGGCTCTAAAATAACTTGTAGTTCAGATTTACACGAAACAATAAATAA
ATGATTAGAGCTTGCAACTCTCCTTATTTAATTTTAGAAATGCATTTTATATTGCGAATA
ATATAATTGCAAAGTTGGAATCAAACATAAATATATATAACTGGTCAACGATTACAAGTA
ATGCCAACAAGATTAACTACAAACTGAATTCGTGTCTACACTACTTCTTTTTGATTTTAT
TTTATTCTTTTTTTGTGTGTGTACAACGTCTAAATTAGCTACTTCTATTTATAGAAATTT
ACATGGAATGCATATCTATAGTAATGCTCATATATCAATCTGTCTCCATATAGTAATTGA
GATGGTTAAACAGAAATTTTTGCTAAAACTTCGTTTCCTGCTTCAAAATAATAATCAGAA
TCCGTAACTATCATACAATAAAAAAATCATCCCATCTCACGTATCCTAGAGAGATTTATT
GTATAAGGTGTTCATATTCTGGAAGGCGTATCGTCCAGCTTAGATGATGTTTTATACTTT
TCAATATAAAATAGACTGTAAATAATATTTTACTTCAGAAACAAATGGTGTTCAAAAATC
TTAAACATCAACTATAAACATACGGAGGGAGTATTGATGACTAAAATTCTAAAATCCAAA
ATGAAATTACAAAATAAATAATAACTATGAAGAAAGAGGCGATCACGTTGCTTACATATC
TTTCCTTTTCAAAAAGCATTTTTTTATTACATTTAACGTGCTTCTCTTTCTTCTTCTCTA
TAAATGGCCAGCATAATTGTTCATTTCTAATGACCACATATAGTCTCTCCTCAACTCTCC
AGAGAAGTTCTTAATTTCTCCTGCCAAGCTGATTAAGGTAATATTACTTTCCTTTGTCGT
CTATATTTAATTTGTGTATATACCATTTCAAGTTGAGTTTTCTTGTTCATAATTCCAAAG
ATATCAAGTTTTGTTTCAAAATTTTGGCCATTGAGAGGTCAAATGCTCTGTTCCATGATA
TTTTTTAAGTCTCCAAATGAAAATATTTCATAAGCGAGAAATAATGACATGAAATCTTGG
ATTCTTTGCGTTGTATGATTATAG**AGCTCTAGGAAATGGCCGAACAACAACTAGGAGTGC**
TAAAGGCACTCGATGTTGCGAAGACGCAACTTTATCATTTCACGGCGATTGTCATCGCCG
GTATGGGTTTCTTTACCGATGCCTACGATCTTTTTTGCGTGTCCTTGGTGACGAAACTCC
TTGGCCGCATCTACTATTTCAATCCGGAGTCAGCGAAGCCTGGCTCACTTCCCCCTCATG
TTGCGGCCGCGGTCAACGGTGTGGCCCTTTGTGGAACCCTTTCTGGTCAACTCTTCTTCG
GTTGGCTCGGTGACAAACTCGGACGGAAAAAAGTGTACGGTCTCACTTTGGTAATGATGA
TCTTGTGCTCTGTCGCTTCTGGTCTCTCTTTTGGCCACGAAGCCAAGGGTGTCATGACCA
CCCTTTGCTTCTTCAGGTACAATTTTCATCCAATTACAATATTATATATACACAATGATT
AACCAATAATCTAATAACGAATGCAAGTTTTAAAAGTTAGTCACGCTTCGAACTGATTTA
GGTATTTTTCTTTGAGAAAGTTTTTTTATATATAAACCATTACGTAG**GTTTTGGTTGGGA**
TTTGGTATTGGAGGTGACTACCCACTTTCTGCCACCATCATGTCTGAATACGCAAACAAG
AAGACCCGTGGGGCTTTCATCGCAGCTGTCTTCGCCATGCAAGGTGTCGGTATCTTGGCT
GGAGGTTTCGTGGCACTCGCAGTATCTTCTATATTCGACAAAAAGTTCCCAGCTCCAACA
TATGCAGTAAACAGGGCCCTCTCAACGCCTCCTCAAGTTGACTACATTTGGCGAATCATC
GTCATGTTTGGTGCTTTACCCGCAGCTTTGACTTACTACTGGCGTATGAAGATGCCTGAA
ACTGCCCGTTACACCGCTTTGGTTGCCAAGAACATCAAACAAGCCACAGCCGACATGTCC
AAGGTCTTACAAACAGATATCGAGCTTGAGGAAAGGGTGGAGGATGACGTCAAAGACCCC
AAACAAAACTATGGCTTGTTCTCCAAGGAATTCCTTAGACGCCATGGGCTTCATCTCCTT
GGAACTACCTCCACATGGTTTTTGCTTGACATTGCCTTCTACAGCCAAAACTTGTTCCAG
AAGGATATTTTCTCGGCCATCGGATGGATCCCAAAGGCAGCCACCATGAACGCCACCCAT
GAGGTTTTCAGGATTGCTAGGGCTCAGACTCTTATCGCCCTTTGCAGTACAGTCCCAGGC
TACTGGTTCACAGTTGCGTTTATTGATACCATTGGAAGGTTTAAGATCCAACTAAATGGA
TTTTTCATGATGACCGTGTTTATGTTTGCCATTGCCTTCCCTTACAACCACTGGATCAAA
CCAGAAAACCGTATCGGATTTGTGGTTATGTACTCTCTTACTTTCTTCTTCGCCAATTTT
GGTCCAAATGCAACCACTTTTATTGTCCCTGCTGAGATATTCCCGGCCAGGCTAAGGTCT
ACATGTCATGGAATATCAGCCGCGGCTGGTAAGGCTGGAGCCATTGTTGGAGCCTTTGGG
TTCCTATATGCGGCTCAATCACAAGACAAGGCCAAGGTAGACGCAGGATACCCACCAGGC
ATCGGAGTTAAGAACTCATTGATCATGCTTGGTGTTCTTAACTTTATCGGTATGCTCTTC
ACCTTCCTTGTCCCAGAGCCCAAAGGCAAGTCCCTTGAAGAACTCTCTGGTGAGGCTGAG
GTTAGCCATGACGAGAAATAATTATGTATGTTTATTTTGTTATTTGGAGTGCGATGTTTG
GTTTTGTTTTCATTTTATTGGCTCGTTGACCTTAAGTTATGATGTTATAAGAATATTTAT
GATATCATTTAAATCTTGTCTAATATTTCGAAATCGACTCACTATTCTGGAAACAAGCTT
ATATATTCACTCCAAAAGCAAGGTTAAAATTAGATCTAGTCGTTTACTTATATTAATATA
TTAGAAAACTAATGAATTTCATTCAAAGTTTAGATCTGATCTATCCAAAGAGAGCAACAT
TTACAGTTACACAGTACTAGAGCTTAGGGCAACAAAATAAAACACAAGAATGTGACTTCT
ACACCATAATAAAAGCTTCCAAAGAATCTTTTCTTATGTCGCTCAAGAAGGAGATTGTGG
ATTCTGTTCGAGCTTCTATCGTGGTGTAGAAGTCCAATGTACCTGATTATTGTTGCGAGA
AATAGGTCTGGACATGCTTTTTTCAGTTAAGTTATATGGTTTAACTTCTTTGAGTTTGGG
GTTTTTAGGTTTAAGCTTTTTAGGAAACATGTCTTAAACTCCAAAATTTTGAATTGTCAA
ATGTAATTTTATTTTGGTATAGGTACCGTACGGGTAGATTTTGAATTGAAAACTGTTACA
AAAATATAAAGTTACATATATTCTGAAAATTTTGGGTATATAAATACAATTTAAATGTAT
TTTATATTTAATTTAAAATATACAGTTTAGTGTATTTTGGTTTGTACGTT



>chromosome:TAIR10:5:17415098:17418572:1
CGTACTATTTTGCAAATCATTTTATATGATTGTATATACCATAGATTATAATCACTACTC
TTAGTCCCATAATTTGTAAATAATAGTAATTCGTAGACATAGATGATCAAAGTTCTTTGT
AATCTTACTACATACACAAAAAGTAATAAAACACAACATTTTTGTTGTATTAAATGTGCT
TGTCCGTTGACAAAAAAGATAAAAGAATAGCTTTGTTTGTCCTTTCTCGACAAGTTTGTA
TTATAATCGTCTCACGTTCATTTCAAGATTACATGCAGAATTTGTCAAATGAAGTCTAAA
TCATCAAAAGTAATTTTTCCTTATTTGTTCTCACTTGAATATAAATTAACGAAATATGAT
TCACCATAAATTACACGAACTATAGGATACGGGTACTTATTTATTCCCTTGCGGTAATAC
GAATAGGTCCAAGATTGGAAATCTCAGAGTTGTCATGAATACTAAGTTAAAATATGTGCA
GACCATACAAAACATAGACGAGCATAGAAAGAGAAGTTGATCAAGTGATGGAAAAAGATT
AAAATAAAATATAAATATGGATATACCCTTAAGATTCCGGAGACTATGCTTCTCCGATCT
CGTCTATAAATTTCAGCCGTCTCAGCCAACACAATCACAACCACCACCTCCCTCCCTCTC
TCTCTCTTAATCTTTCTGCCCCCACCATTAGCGCACAACGGTGAGATTCGTTAGATGTTT
ATTATCTATGCATCCAAATTCACATATAGATAGATAGTACATATATAGTTGTTTATATCA
ATAGAAACTTTTGTTTCTAAATTGTAATCGCTTAAATTATTTGATTTATGTATAGGTTTA
ATTATAAATGTTATTCTGTTGGTTGTATTGGTACTAAACTAAAATAATTTAGTAGCATCT
TTCCAATATTATTATAAAATTATAGTTAAGATATTTATTTTTACTTCATGAAG**CAGGAAA**
GACAAGAGAGGCTTAGATGGCTGAACAACAACTAGGAGTGCTAAAGGCACTCGATGTTGC
GAAGACGCAACTTTATCATTTCACGGCGATTGTCATCGCCGGTATGGGTTTCTTTACCGA
TGCGTACGATCTTTTTTTGCGTGTCCTTGGTAACAAAGCTCCTTGGCCGCATCTACTACTT
CAATCCGGAGTCAGCGAAGCCTGGCTCACTTCCCCCTCATGTTGCGGCCGCTGTCAATGG
TGTGGCCCTTTGTGGAACCCTTTCTGGTCAACTCTTCTTCGGTTGGCTCGGTGACAAACT
CGGACGGAAAAAAGTGTACGGTCTTACTTTGATCATGATGATCTTATGCTCTGTCGCTTC
TGGCCTCTCTTTTGGCAACGAAGCCAAGGGTGTCATGACCACCCTTTGCTTCTTCAGGTA
CAGTTTTCATCCAATTACAATATTATATACACACATGATTAACCAATAATCTAATAACGA
ATGCAAGTTTTAAAAGTTAGTCACGCTTCGAACTGATTTAGGTATTTTTCTTTGAGAAAG
TTTTTTTATATATAAACCATTACGTAG**GTTTTGGTTGGGATTTGGTATTGGAGGTGACTA**
CCCACTTTCTGCCACCATCATGTCTGAATACGCAAACAAGAAGACCCGTGGGGCTTTCAT
CGCAGCTGTCTTCGCCATGCAAGGTGTCGGTATCTTGGCTGGAGGTTTCGTGGCACTCGC
AGTATCTTCTATATTCGACAAAAAGTTCCCAGCTCCAACATATGCAGTAAACAGGGCCCT
CTCAACGCCTCCTCAAGTTGATTACATTTGGCGAATCATCGTCATGTTTGGTGCTTTACC
CGCAGCTTTGACTTACTACTGGCGTATGAAGATGCCTGAAACTGCCCGTTACACCGCTTT
GGTTGCCAAGAACATCAAACAAGCCACAGCCGACATGTCCAAGGTCTTACAAACAGATAT
CGAGCTTGAGGAAAGGGTGGAGGATGACGTCAAAGACCCCAGACAAAACTATGGCTTGTT
CTCCAAGGAATTCCTTAGACGCCATGGACTTCATCTCCTTGGAACTACCTCCACTTGGTT
TTTGCTTGACATTGCCTTCTACAGCCAAAACTTGTTCCAGAAGGATATTTTCTCGGCCAT
CGGATGGATCCCAAAGGCAGCCACCATGAACGCCACCCATGAGGTTTTCAGGATTGCTAG
GGCTCAGACTCTTATCGCCCTTTGCAGTACAGTCCCAGGCTACTGGTTCACAGTTGCGTT
TATTGATACCATTGGAAGGTTTAAGATCCAACTAAATGGATTTTTCATGATGACCGTGTT
TATGTTTGCCATTGCCTTCCCTTACAACCACTGGATCAAACCAGAAAACCGTATCGGATT
TGTGGTTATGTACTCTCTTACTTTCTTCTTCGCCAATTTTGGTCCAAATGCAACCACTTT
TATTGTCCCTGCTGAGATATTCCCGGCCAGGCTAAGGTCAACATGTCATGGAATATCAGC
CGCGGCTGGTAAGGCTGGAGCCATCATTGGAGCCTTCGGGTTCTTATATGCGGCTCAAAA
TCAAGACAAGGCTAAGGTGGATGCAGGATACCCACCAGGTATCGGAGTTAAGAACTCATT
GATCGTGCTTGGTGTTCTTAACTTCATCGGTATGCTCTTCACCTTCCTTGTCCCAGAGCC
CAAGGGCAAGTCCCTCGAAGAACTCTCTGGTGAAGCTGAAGTTAGCCATGACGAGAAATA
ATTTACTTTGTGATCAAATGTGGAGTTGTTTGTGGTTTGTTTGATTTTTTTGTTTCTTTC
CAAATTTTTTTTTTCCTTTCCTTGATGCAGTAAGAAATGTTTGATTGTGTGGAATACTTC
AGTCTTACTGCCGTGAGTGCAAATCTTGTAATCCACTTTTTTGTGTTCTAATTAACATTG
TCAAATAAAGTTTCTTAAACTTTGATTCTAACTTTCCAAACACTAACGAACTCTCTGTTG
TGCGTTTATGTTTATCAATTTGAAGCATGAAAATTTGATTGTTCATATGTTTGCTATATA
CTGAGAAACACCAAAGAATAGTTTTTATGGATTTCGATTATCTGATCAACACATTAGTTG
GTTATGGTGTTCAATGTTCAGTGTCTTGTCATCACTAACGTTATCTCTTTGCCTTATATC
GACACATTCAAGTGGGCGTAGTTGTAGACCATTAGTCTTGTGCTCTGTTTGTAAATAACG
CTATTGGTTAGTTAGCTTTCTTTTTGGTTGGTGTATATATTCTTTTTGGGTTGAAGCTCT
GTGTAGCGTTTAGAAATGTTAAAAGTTCGATCCCAAAATGTAGCAAGATTTGAAATTTCT
TTTTGGATTGATGTAATGTATTTAACTGGCTTCTTATGAATCACGGTGCTTATTGTTTGA
TTAATGTGAACTGAATTGTTCTGTGCTTGAGATTCAAAAAGCTCAAGAGGATCTAATACT
AAATAGTCTCACTAATGCATATGTTTGATTTGTTTCAAATATTGATTGTTCTCTC

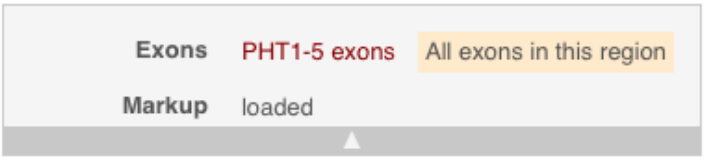


>chromosome:TAIR10:5:17408924:17411814:1
CATAAACAATATAAAAAAATTATTAACGATAGTGAACTCGATTTTTAGTAACTCTATGAA
AATCTTCTTGTGAGATCATAACATTTTAGAAGACTAGGCTTTTCGTGTAATATGATCTGT
AGACCATTCCAATGTTATCATATGCAGAACAGCAACACAATGAAATAAATTAAGAGATGG
AAAATAAATTAAGATCAAATAAAATGTGGATATACCCTTAAGATTCCGGAGACAATACTT
CTCCAGTTCTCCTCTCGTCTATAAATTTCTGCCGTCTCAGCCAACATAATCACAACCACC
ACCTCTCTCAATAATTCTCTCTCTTGTCCTTCACAACCTGTTACTACTTAGGTGAATATC
ATCTACATGTTTATTTTGTCTTAATCCAAATTCACATTTAATTTTGTTTTCACTCTTTTT
ATCATGTATACTTTTTTAATTAGCTAGCCGAGATTATATATGTTAAAACTACCTCTATTT
CTAACACACCCAAGCAATATATCATCTTTCCAATATTTACGAAAGAAGAAATTTCTATAC
AAATATTAACCTTTACTTTGTGCAAAATAGGAGAGAAGAAAAACAAAAAGACTGACTGAG
ATGGCCGATCAACAGCTAGGAGTGCTAAAGGCACTTGATGTTGCGAAGACGCAACTTTAC
CATTTCACGGCTATTGTCATTGCCGGTATGGGCTTCTTTACGGACGCGTACGATCTCTTT
TGTGTGTCCTTGGTGACCAAGCTTCTTGGCCGCCTCTACTACTTCAATCCAACGTCAGCA
AAGCCTGGCTCACTTCCCCCTCATGTTGCGGCTGCAGTCAACGGTGTGGCCCTTTGTGGA
ACCCTTGCCGGTCAACTCTTCTTCGGATGGCTTGGTGACAAACTCGGACGGAAAAAGGTG
TACGGTATCACTTTGATCATGATGATTCTCTGCTCAGTTGCTTCCGGTCTTTCCTTGGGC
AATTCGGCCAAGGGTGTCATGACGACTCTTTGCTTCTTCAGGTACAATTTATTTAGCCAC
AAACCTAATATCACATACGTCACAGATACAAGCTCGAGAGATTAGTCACTATTTCGACCT
AGATTATGGTTACTTAAGATACTGATATCTAGACGATTATATATAG**GTTTTGGCTCGGGT**
TTGGCATTGGAGGTGACTACCCTCTATCTGCCACCATCATGTCTGAATACGCTAACAAGA
AGACTCGTGGGGCTTTCATCGCGGCAGTGTTCGCCATGCAAGGTGTAGGTATCTTGGCGG
GAGGTTTTGTGGCACTTGCAGTTTCTTCCATATTTGACAAAAAGTTCCCATCGCCGACGT
ATGAGCAAGACAGGTTTCTATCAACGCCTCCTCAAGCTGATTACATTTGGCGAATCATCG
TCATGTTTGGTGCTTTACCCGCAGCTTTGACTTACTATTGGCGTATGAAGATGCCTGAAA
CAGCCCGTTACACCGCTTTAGTTGCCAAGAACATCAAACAAGCCACAGCAGACATGTCCA
AGGTCTTACAAACAGATCTCGAGCTTGAGGAAAGGGTGGAGGATGACGTCAAGGACCCCA
AAAAAAACTATGGCTTGTTCTCCAAGGAATTCCTTAGACGCCATGGGCTTCATCTCCTTG
GGACTACCTCCACTTGGTTTTTGCTTGACATCGCCTTCTACAGCCAAAACTTGTTCCAAA
AGGATATTTTCTCGGCCATTGGATGGATCCCAAAGGCAGCCACTATGAACGCCATCCATG
AGGTTTTCAAGATTGCTAGGGCTCAGACTCTCATTGCCCTCTGCAGTACAGTCCCAGGTT
ACTGGTTCACAGTAGCCTTTATTGATATCATTGGAAGGTTTGCGATCCAACTAATGGGAT
TTTTCATGATGACCGTTTTTATGTTTGCTATTGCCTTCCCATACAACCACTGGATTTTAC
CAGATAATCGTATCGGATTCGTGGTTATGTACTCACTCACATTTTTCTTCGCCAACTTTG
GACCCAATGCAACTACTTTCATTGTCCCAGCTGAAATCTTTCCAGCAAGGCTAAGGTCTA
CGTGCCATGGAATATCAGCCGCAACTGGTAAGGCTGGAGCCATCGTTGGAGCCTTCGGGT
TCCTATATGCTGCTCAACCACAGGATAAGACCAAGACAGACGCAGGATACCCACCGGGCA
TCGGAGTCAAGAACTCATTGATCATGCTTGGTGTCATTAACTTTGTTGGTATGCTCTTCA
CCTTCCTCGTCCCTGAGCCCAAGGGCAAGTCCCTTGAAGAACTCTCCGGCGAGGCTGAGG
TTGATAAATGATTATGCCGTCATATATGTTTGTCATTGGTTTTGCGATGTGTGAATTATA
TTTGTAATGGTGTACTACTTTTACGTTTTACGTTCTTTGCGATGAGTGAATTATATTTGT
AACGGTGTACTACTTTCGCTTTTTGTTTAAATGTGTGTGCAAGTGCAACTTGTTAAGATG
TAAACTCATGTTATGGTCATATCCTAGTAATGCTATAAGTTTGGAAGCAATAAAGACATG
AACAATCAAACAAAAAATATGCTTAGTGGAAAGTTTGAAATGAAAGATATAAGGGCTAGT
GGATGTAATTCTAGGGATCAATATTGATAGAACTGATAATACAAAAAATCAGCTTGTACT
GGTATATTAGGTTGAAAAGATAAATGAGATATTTTAGAACTACTGAACTATAAAGATTGG
TATATTGTTGTTCATGTAATCGAAAGATTTTTTTTGTAAGATTGTAGATGAATTTATCAT
CAAATGAACTACAAAGACTAGGGACTCCCAAGAGTCCAAGACCCTACTCTTCACTTGTCT
AAACATTTAAGTGAAGCAATGTCAAATTAGGAGTATGATGCAACAGTCTATTTGATATAT
TTGACCAGTTG

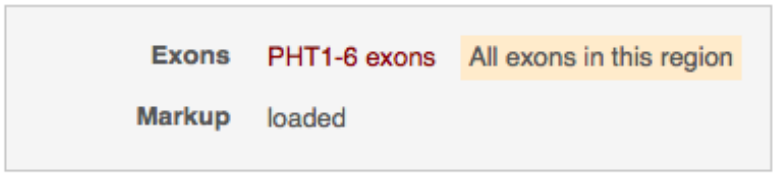


>chromosome:TAIR10:2:16256557:16260853:1

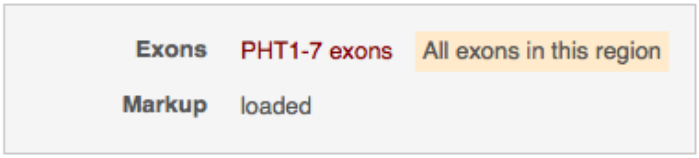
CTATTACACATCCTGCAGTTATTGGCTTATTGCGATCTTTATTAAATCCAAAGATACATA
CTATATCGAAGAAACAAAAAGTCAAGAAATAATAAAACGAAAATAAATGAAGGCATCAAT
AAAAGCTTACCGCTCACATGTTTATTTTCTAATAACTAATTTTTATTTAAAAAGCAGTTT
ATACATCTACCAAATTTATTTCTTAGCATAAATATATATTTGGGTTTTGACTTTTAAGTT
CTTTCTGACTTCTGAGTGATAATCACCAGTTTGCAACTTATATTTGCCTAAACCGCATGC
CAATTGTCATGTATCGTATCTAGTAATGGTATTAATGACGAGGATCCCAAAATTTAAATT
CCACTTTCCAAGCATTGAGCTCTTTAAACAATTCATGGTCAACTTAATTACAAGGAAAAA
AAAGAACTTATTGTTATAGTGGAACAGCTATTTTTTTGGATATTAAAAGAATAATAACAG
CAAAACAGAATTATCGTGTCCTAATAATACCTAAGGTCCTAAACGAAGCAAAAAAGTTGG
TAAATAAGGAAGAGAAAACCTACAAGATATTAAAACGGTGTCGTTGTTCGGAAGAATATA
CCGAAGTAGCAAAAGGAATATCTCATTAGAGAGTCCCTTATAAATGACCGTTTTAATACA
CTTCAACTCTGTCCTTGTTCATAGGCAGCTTCAACGATCATTCCACTTCCTTCTTCCTCT
CTCTCAACATTTTCCCCTGAAAATAAGGAAACTAAAGATTCTTCCTCTCTCTTTCTACAC
TCTTCTGACAATACTAAAACACTTTATCAGATCAGATCCCACATAAACTTTCTGTAAGTG
TTTCTTTTAAACTATGTCTGTTGATGTTAATGTTTATGGATGCATAATGATTTTAGTTAG
GACGTTATCGTTTTTGTTTTGTTAGGTGGTCCTCTCTCCCTGTTCAATCTCTAATTCGTC
ATGTCTTATTCGTTTCTTTCATTGGAAAAAAAACACTGAAAACTAGTTTTTGTTTTTCCA
ACTTTCTTGGATCATTAAATAGGATTGCTCTCTCTTCAAACGTATTTTTGTTGGATCTCT
CCTCTTCAGTTCTAGAAGACTTTTTTTTTTTTCCTTTTGTTGAGTGCATGATTTCAAAAA
ATTTCAATTGAAGAAGTTTGTGATATTTTGTTGTTTTCTTTCTTCAATAAGAGGTGTGTC
CGCAATTAAGACAAAACAACTCGCTGTCATACTCAGTTAAAGTCTTTGTTGCTCGTGTTC
AGATTTTCACTAGACGTCACAAGTCAACTTTTGTGGGTAACTGAAGACAAATAGTAATCG
TTGAAAACTACAATAATAAAATAAGTACTTCCAGTGTCTCCGTTCACCTTTAAAAAGTCG
AGATAAACCCAAAAAAATAGGAATAAACAATCTTTTGTCATATCCGTTAATAAGTTTTAA
CAAATCTTTACAAATTAATAAGAAGAACGTATATAATAAATATTTCAAAATTAACGAGTG
TGGACTAATATTGAGTTGACTTTTATTATCACTTCATTTCAGTTCCTTGAATATGCATAT
ACTGTATATGGTTTCTCTTGCAATAAGTAAGCATCATACATATGTACAGGAACATGAGAA
GTTGTCGTTAACAACTACTTCATCTAAAAAACATACACATGTTATTTTTCTAATATTGCC
TGACTTCAAGTTCGGTGAACATACTTTATTGGGTTGTAATTGTGAGATCAATGAAATGAG
AAAAATTAAATAGAACTTTTATGATATTTTAATGTTAATTGATACTATATAAAGCAGGGT
CTAATTCTATTTGCTTTTAAAACTGCCTAACTTATTTCAAGTTCGGTATAACTAACTACT
TTATTGTTTGTAATGAGTTATTTAAGGGTACTTATATGTCGAACTTTTTTTTGTGTGATG
AAAAATTGCAG**AGGAGAAGAAGAATGGCAAGGGAACAATTACAAGTGTTGAATGCACTTG**
ACGTGGCCAAGACGCAATGGTACCATTTCACGGCGATCATAATCGCCGGAATGGGATTCT
TCACTGATGCTTACGATCTCTTTTGCATCTCTCTCGTAACGAAGCTCCTCGGTCGTATTT
ATTACCACGTGGAAGGCGCACAAAAGCCTGGGACTCTCCCTCCCAACGTCGCAGCCGCCG
TCAATGGCGTTGCCTTCTGTGGGACTCTCGCCGGTCAGCTCTTTTTCGGGTGGCTTGGTG
ATAAGCTCGGGAGGAAGAAAGTTTACGGTATGACGTTGATGGTCATGGTCCTTTGTTCAA
TAGCCTCTGGTCTCTCTTTCGGACATGAGCCAAAAGCTGTGATGGCCACGCTCTGTTTTT
TTCGGTTTTGGCTTGGATTTGGCATCGGTGGTGACTACCCTTTATCCGCAACCATCATGT
CTGAATATGCGAACAAGAAGACTCGCGGAGCCTTTGTCTCTGCGGTTTTTGCTATGCAAG
GGTTCGGAATCATGGCTGGTGGTATTTTCGCTATTATAATTTCCTCTGCTTTTGAAGCTA
AGTTTCCATCCCCGGCCTATGCGGATGATGCCTTGGGATCCACGATTCCTCAAGCTGATT
TGGTATGGCGGATAATCCTGATGGCTGGTGCTATCCCTGCGGCTATGACGTATTACTCAA
GGTCGAAGATGCCTGAGACCGCAAGGTACACGGCTTTGGTTGCTAAGGACGCAAAGCAGG
CAGCTTCGGACATGTCTAAGGTTCTGCAAGTGGAGATAGAGCCAGAACAACAGAAATTGG
AAGAGATCTCAAAGGAGAAGTCCAAGGCCTTTGGATTGTTCTCAAAGGAATTCATGAGTC
GCCATGGGCTTCATTTGCTAGGCACTACATCGACATGGTTCCTTCTCGACATTGCTTTCT
ACAGTCAAAACCTTTTCCAAAAGGATATTTTCAGCGCGATCGGATGGATTCCTCCCGCGC
AAAGCATGAACGCAATTCAAGAGGTTTTCAAGATTGCCCGTGCGCAAACGCTAATCGCCT
TGTGTAGCACGGTACCTGGTTACTGGTTCACAGTTGCGTTCATCGACGTCATTGGAAGAT
TTGCGATTCAGATGATGGGTTTCTTTTTCATGACGGTCTTTATGTTTGCTCTGGCTATTC
CTTACAACCACTGGACTCACAAGGAGAACCGAATCGGATTTGTTATCATGTACTCGTTAA
CATTCTTTTTCGCCAACTTTGGACCCAATGCTACAACCTTCGTTGTGCCGGCCGAAATCT
TCCCAGCCAGGTTCAGATCAACCTGCCACGGTATCTCTGCAGCATCAGGAAAATTAGGAG
CAATGGTTGGTGCGTTCGGGTTCTTGTACTTGGCTCAGAACCCAGACAAGGACAAGACCG
ACGCAGGATACCCTCCAGGGATTGGGGTCAGGAACTCGCTTATTGTGTTGGGTGTAGTTA
ACTTCTTAGGTATCCTCTTCACTTTCTTGGTACCTGAATCTAAAGGTAAGTCACTCGAGG
AAATGTCCGGTGAAAATGAAGACAATGAGAATAGCAACAATGATAGTAGAACGGTCCCAA
TAGTTTAGGTGATATAATACGCCTTTTGTAATAATTTTCGTTTTTTCTTTCTCCTTGTCT
CTAGCAACTCAAGTTGTTCTTTGTGTAATCCATTGATACCTAATTAATGCTAGAGAAATC
AAAATTTTCATGAGTCGATTTTAAATCAGTTCCTAAATTGAAGATTTATTAGACTTGAAG
AAGCCGTGCAAGTTGTGCAGTATATATGTTAAATCACAATTTATGCGTTGCGGGGAAGTT
GAGACCACAGATAAAGGCTTCACTTAAAGATACATTATGGCTTATTAATCTCTAGGGTGA



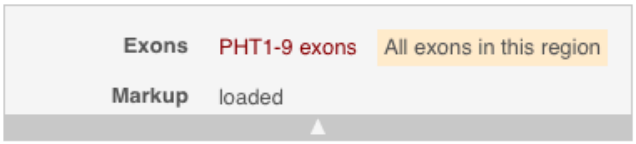
>chromosome:TAIR10:2:13927006:13930363:-1
AAATTGAAATCCAGAGGGTAAGTGAGAGAAGCAATCGACCTCAGCAAGAGGTGTAGTATT
CATAAAAGTTTTACGTTGTATATTTAATGACACTACAAAGTCATTGGAAAATATATTTTT
GTAAGGTATATTCTTATAGATTCTTTTAAATGTACTTTTCGTTTCTAGAGTTCTCTGAAG
ACTTTAGTTCCACTTCAGCGTTTCAACAGTTCAAAGTTTCCACCTCATGCATGTGTTATG
TGTACATATTTCTTCTATACATCCTTTTTTTTCAGTGAACTATATGTTATAAAACAGTAA
CTCTATAGTTAAGTGTAAACTGGTTGGTTTAAGAATAAACTAAATATTTTTTATTAAATT
GTAGTAAATTGATTAGTTTAAGATAGAAACCAAAACCAAGCTTAATAAAGTTAGATTGAT
TCCTAACGTTAGAAAAAAAAAGTTAGATTGATTCCTCACGTTTCAAAAGGGGATCGTTGG
TTCTTGGCCTCACACATACTTTTCCTTGACTCTTTCAAAAAAGACAACCTTTGACAAATC
GCTTATAAACTGCATTTAATTTTACATAGATAACTGCACTCACTGTCACATGCATGTTCA
CAATATCTTGACTATATAATAATTACATTTATAAAATACTAGCCTATGTGATCAAAAATT
GTTGTATCAAGATTTTCTCTCAGGTGACTGAACAACAGAGAAAAACATGGCGAAAAAAGG
AAAAGAAGTGTTGAACGCTTTGGACGCAGCCAAAACGCAGATGTACCATTTCACAGCGAT
TGTAATCGCGGGTATGGGTTTTTTCACTGATGCTTACGATCTATTTAGCATCTCCCTCGT
CACGAAGCTCCTAGGCCGCATATACTACCACGTTGATAGCTCCAAGAAGCCTGGAACGCT
CCCTCCTAATGTTGCTGCCGCTGTTAATGGTGTGGCCTTCTGCGGTACCTTAGCAGGCCA
GCTCTTCTTCGGTTGGCTCGGTGACAAGCTCGGCCGCAAGAAAGTTTACGGTAACCTTTG
CTAATTTCTTGGCCTTAACATATGTTTTAGTAGTTTTACGTTTAAAATGATCTATCTAGA
TTAATTCAAAGTATTTGATCCATTGGTAATACTAGAACTATTAATTATAGTATTTAGTGG
CTAAATTAACTAATGGTATATAATGAAAATTATTGGAATACACATAG**GTATTACCTTGAT**
GCTCATGGTACTATGTTCTCTCGGGTCTGGTCTATCATTTGGTCACTCAGCTAACGGAGT
AATGGCCACACTTTGTTTCTTCCGGTTCTGGCTCGGTTTCGGCATAGGCGGCGACTACCC
TCTCTCAGCCACCATCATGTCGGAATACGCCAACAAGAAAACACGTGGCGCCTTCATAGC
AGCCGTCTTCGCTATGCAAGGCTTCGGAATCTTAGCCGGAGGAATCGTGTCTCTCATCGT
CTCGAGTACATTTGACCATGCCTTCAAAGCGCCCACCTACGAGGTTGATCCTGTTGGCTC
CACTGTTCCTCAAGCGGATTACGTGTGGCGGATTGTGCTCATGTTTGGAGCTATCCCAGC
TTTACTTACTTATTATTGGCGGATGAAAATGCCGGAAACTGCTCGGTATACCGCTCTTGT
AGCTCGGAATACCAAACAAGCCGCCTCCGATATGTCTAAGGTACACGTTCCTTAACTAAT
ACTTAATTTCATCATATACTAAATCATCATACTAAACCATGATGATTAGTGTAAAAGTAA
TTATTATTTACACAACAG**GTGCTTCAAGTGGATCTAATTGCAGAAGAAGAAGCACAATCA**
AATTCAAATTCATCAAACCCTAATTTCACCTTCGGTTTATTCACCAGAGAATTCGCTCGC
CGTCACGGTCTCCACCTTCTCGGCACAACCACCACATGGTTCCTCCTAGACATCGCCTAT
TACAGCAGCAACCTCTTCCAAAAAGACATCTACACAGCAATCGGATGGATTCCCGCCGCC
GAAACAATGAACGCAATCCACGAAGTCTTCACCGTCTCAAAAGCTCAAACCCTAATCGCT
CTCTGCGGCACCGTCCCTGGTTACTGGTTCACCGTCGCATTCATCGATATACTCGGTCGA
TTCTTCATCCAGTTGATGGGATTTATCTTCATGACGATCTTTATGTTCGCTTTAGCGATT
CCTTATGATCATTGGAGACATAGAGAGAATCGAATTGGTTTTTTGATTATGTATTCGTTG
ACTATGTTCTTCGCTAATTTTGGACCTAATGCGACGACGTTTGTTGTTCCTGCTGAGATA
TTTCCGGCGAGGTTGAGATCGACTTGTCATGGGATATCGGCGGCGAGTGGTAAGGCGGGA
GCGATTGTTGGAGCGTTTGGGTTTCTTTACGCGGCGCAGAGTTCGGATTCGGAGAAGACT
GATGCTGGTTATCCTCCTGGGATTGGTGTTAGGAATTCGTTGTTGATGCTTGCTTGTGTT
AATTTCTTGGGGATTGTTTTTACTCTGTTGGTTCCGGAGTCTAAAGGGAAGTCTTTGGAA
GAGATTTCGAGAGAAGATGAGGAACAAAGCGGCGGAGATACGGTGGTGGAGATGACTGTG
GCAAATTCCGGTAGAAAAGTCCCGGTTTGATTATTTTGTGTTACTCATTTTGTAAATTTA
CTCTATATATGAAAGTCCCTTGCATAAAGTTTTTCGATACTCAATATATCAGAATAACAG
TTTTGCTGGAACTTTGCAACCAAATTACTCTGTGAATTGGAAAGTATTATCTTACTTCTT
CCTATAAAGCACCAGTGATCTTTTTGTAAAGAGATTCATAAGTCTATTGGTTTATGGAAC
AATGCCTTCACATGCAAGTCATATTATTTGTCCAGACGACGAAGAGGAAGAATTTGTTGA
ACTAGAAAGCAGTGGCGATAGAATCGTTCATTAAGATAAATGGTTCGTATAAAACCATTA
GGATAAATTCTATGTTTTGTGTTTCATCATTGAATATCTACCAGACCAAATGATTTTATT
TGGTTTTAGTTTCTTTTGTAAATCCCAAAAAAGGTAAAGTTATATCAAAACCAAACTACA
AACCAGATGTTAATCCCTATTGTATTAACCACAAACATGTTACCAGAATATATCAAAAGT
GTATCAGTTTCAGTTTGCTACTAAGTAAAGACTTTCATCTTTGTGTTGCTTAGGCTTCAA
GTTCGTTTTATCTTGTCTTCTCTGTTTTACTCTGCTTTATGATGATGATAAATAACGATG
GAACACAAGAGAGAAGAGCTCCATTATGCACAAGAATTCATTTCCATAATATAGTTGCTT
GTCATGTCGTTCAAATCAAACTCCATTTTCATGAATTTGCATAAACAAAAGAAAAACA



>chromosome:TAIR10:5:17392998:17395748:-1
CATTCCGAATTATTCTTAATTTCCACTCTGATTTTATCAAACCAAAAAGAAATTCCAGTT
TGACGATAATTCTATTCCTTTTCCCAATAGAATATGCTATTGTGATCTCATCTCTATGTG
TGTCTGTCTATATATAAAGCTATAACCCTAAGAATCAAAACTTCAACAATCGTCAGCTCA
AAAAAAAAGGAGTCTTTCCGGCTCTAATCACTTGTTTTGGTAATCACTTCATCCCCTTAA
ATTTACTTTATAAACATATATTTCGTTCATATATTTGTTGGTAACTACTAACTAGCTATA
ATCAGATGTTTCATCACGGAAACTAATTTGTTCAATGGATCATGAATATTGCATGAAAGA
AAACCACGGCATGATCTTTCTGTGTAAAATCTTTTTATCGCTGTTTTTTCAGTATGGTCT
ATAATCCCCAATTTTTTTGGAATAGTTTGTTGATATTGTTTATTTCACGTTTACAAATAA
AATAGTTCCTTATCAAAACTAGAATCGTGCTACGTTTTTGTTACAAAACTTTCTCTATAT
ATTGTAACAATTCTTTCGGCTTGTTTGTTTTTTCAGAGTTGAGAGAACAAAAATATTGCG
ATGGCTAACGAAGAACAAGGAAGCATCTTGAAAGCTCTTGACGTTGCAAAGACTCAATGG
TACCACGTCACGGCGGTTGTAGTCTCCGGTATGGGTTTCTTCACAGACTCATACGATCTC
TTCGTGATATCTCTTATCACGAAGCTTCTTGGCCGGATCTACTATCAGGTTCCTGGCTCC
TCCTCTCCCGGAAGCCTCCCTGACGGCATCTCTGCGGCCGTGAGTGGTGTGGCCTTCGCT
GGAACGTTTATCGGACAGATTTTCTTTGGGTGTCTTGGTGACAAGCTTGGACGTAAGAGA
GTCTATGGTTTGACTCTCTTGATCATGACCATATGCTCCATTTGTTCTGGTCTTTCCCTT
GGAAGAGACCCGAAAACCGTCATGGTAACTCTCTGCTTCTTCCGCTTCTGGCTCGGGTTT
GGCATCGGTGGTGACTATCCTCTCTCGGCTACGATCATGTCCGAGTATTCCAACAAACGT
ACTCGCGGTGCTTTCATAGCAGCTGTTTTCGGTATGCAAGGGATTGGGATCCTTGCCGCA
GGAGCTGTCTCGTTACTTGTCTCAGCTGTTTTCGAGAGTAAGTTTCCGTCCCGAGCCTAT
ATATTGGATGGTGCAGCCTCCACCGTCCCACAGGCGGATTACGTTTGGAGGATCATCCTG
ATGGTCGGCGCCTTACCAGCTCTCTTGACCTACTACTGGCGGATGAAGATGCCCGAAACC
GCTCGTTACACGGCTCTTGTCTCCAAGAACGCGGAGCAAGCGGCTTTAGATATGACTAAG
GTTCTCAACGTAGACATTGAAGCCTCTGCCGCAAAGAATGATCAAGCTAGGGTTTCCTCG
GACGAGTTTGGCTTGTTCTCCATGAAGTTCCTTCGCCGTCACGGACTCCACTTACTCGGA
ACAGCCTCCACATGGTTCCTCCTCGACATCGCTTTCTACAGCCAAAACTTGTTCCAGAAG
GATATCTTCACCACCATTGGGTGGTTACCTTCGGCTAAAACCATGAACGCAATCCAAGAG
CTCTACATGATCGCTAAGGCTCAAACCATAATCGCTTGTTGCAGCACCGTTCCCGGTTAC
TTTTTCACTGTAGGGTTCATCGATTATATGGGACGGAAAAAGATCCAGATCATGGGTTTC
GCTATGATGACCATCTTCATGCTCAGTTTAGCCATACCGTACCATCATTGGACCTTACCG
GCTAACCGGATTGGTTTCGTCGTTCTTTACTCTTTCACATTTTTCTTCTCCAATTTTGGA
CCTAACGCAACTACTTTCATTGTCCCGGCTGAGATTTTCCCGGCGAGGATAAGGTCGACT
TGTCACGGTATATCGGCGGCGTCTGGTAAAGCCGGTGCGATGGTCGGTTCTTTCGGGTTT
TCGGCTTTGGTTAAAGCTTTGGGTATGAGTAACACGTTATACATCATGGCAGGAATCAAT
TTGCTTGGCTTGCTTCTCACGTTTACAATTCCGGAGACTAATGGGAAATCACTTGAGGAG
CTTTCCGGCGAAACTGAGCCGGAGAAAATCAAGGAGAAGATTGTTGTTTAGATTTTTTTA
TTTTCCTTTTGTTTTTTTTGCTTAATTATCGGATTTTGTTAGGATTTCTTTTTGTGTTTT
GTCAGACTTTTAGATTTTTTGTTTAGATTTATTTAGTTGCACTATGAGTTTTCCTACTTT
TTTTGAATTATTATCTTTTTCATCAAACAGATTTTATAAAAAAAAAAAAAATGTTATTGTT
ATTAGAAGCTGATATCAAAAGTAAAACAACAAACAGATTAGTAGATAACATGAGAAAAAT
ATGTATAGCAATATTTATTGCCAATAATAAATGTTAAGGTTGGAATGTATTTAGGTCCTA
TATAGACTTGTCACTAACATTCAATGTCCAGGAAAAAAAAAGAGGACATGAATTAAGCCGT
CGATAATTCATCAGGAATTAGAAGAAAAAAAAAAGAGTTATCAACAATACCCAACAAATCA
TTATCTTCTGCAGCTATATTTGCTCTTTTATAGAAAAAAAATTATATTTAAAAATAAAAT
TACAAGATAAATCAAAGTGACTGCAAGTCTTGGAGAAAAAACTCATAGAATAGGGTTTTA
GTTTTAGATAAAAACCTATGAAATAGGGTTTTACGTTTTCTTTTTTTGATA

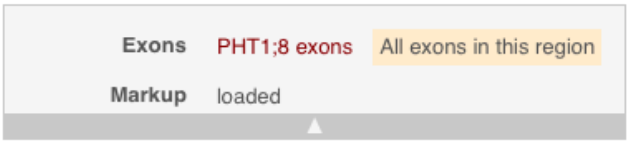


>chromosome:TAIR10:3:20247614:20251548:-1
AATTTATTTATTATTTTGTCATCATTTTCTTCTTTAGTTTATAATATGAATTTTACTATA
TTTACAAATACTTCTGGGAAATTCAACGTTTTTAAAAAACGGTTATACAAAATAATGTAT
GTAAGAAAACACAGAAGACTGGGGTTAAAACGAATCAATTTCCAATTTTCTACTCTTTAT
GGCAAAATAATATCTTGCAAATTCTGCACTTATTACCAAAAAAATAAAAATAACAGCTCA
TCATGTGTCTGAAGTCTGAATCTTATTTTTTCCCCAAAGAATATTCGAAATTATTACTTA
AAAACGACTTCTTGTAACCGTAAGTCTTAAGAATTCGAATAGACAGTAAAACAATTTCCA
AAACAGAAAAAAAATTAACAGCAAAACAGAACCAATAGTGTCTCCTAAAAGAAGTAAAAA
TAAGGTGGAGGAGAAGTAGTCAAATAGAAGAAATAAAAGCGACGTCGTTGTCAGAAGAAT
ATACCACAAGAATATGTCAAAACAAAAACAGAAATATGATTACGTATAAGTAGCCGTTCC
CGACGCTGCCTCTTTTTCATTTTCACAACACCATTTCTTCTTTCTTTCTCTCTTACAGTT
TCTGAAGAAACAACAAACAAACAAACTAATACTCTTCTCTTCTTCTTCTTTTTTCTTTCT
TTAAGACAATCCCATATTAGAGATTGTTTGCGAGAAAGTAAGCAAAAAAACAAAAAAAAC
TTTCTGTAAGTTAATGTCTTCTTCATAACTATGTGCTATTGAATCTTTTATTAACTTTTA
TTATGACTTGTTGTTGAATGAGAAATGAGGTTATGCAAGCGTCTGGTTAGTGTTGTTTAT
AAATTATATAGCTTTTGTGTCTGTTGTGTAGCATGTGTTTGTGTTATGTCTATTAAATAG
AGTAGTTTCCTAAGAAACATTTTCTATGATACCTGATCAATGTTTTGATAGACGACAATA
TGTATGTGGAAGCTAAAACTCAAATCGTCTTTTACAAAATTGTTCAGTAAACTATGATTA
TACAGCTTTTTTGTTGTTGAAGTTCTGAGATTGAAACTTAATTCGTTTACAACTTTCTTT
CTTGGATCAGTTTATTTTGGCACATTCTTGGAACAAAAACAAGATAAGTGGTCTGAAATA
TTTTGTGATTTAAAGGAAAGTTTTGACTGTCATTGAACTGTACATTATGGAATATGATCA
AATCTTTTGTCGGTGTGCGTCACCATCTTACTTCGTGTCCCGTTTTTCATTTCCCACGCG
TCAAGTCAACCTTTACAAATACTTACGCCAAAAAGATATAAAAATTGATGAATTTGACGA
GTTTTTAAACCTTTTTGATGATTATAGATTAAAATTTTACTATAGTTTATTTTCTGCATA
TAAAACGTGTCTGAAACCAGCTTATGTATATGAATTTGATACTGATTACTTAATATGTGT
TACTTAATATTAACGCAG**AGGAAAAGAGAAAGAGAGCAATGGCAGGAGATCAACTAAACG**
TGCTAAACGCACTCGACGTCGCCAAAACGCAATGGTACCATTTCACGGCGATTATAATCG
CCGGAATGGGATTCTTCACCGACGCTTACGACCTTTTTTGTATCTCACTCGTCACAAAGC
TTCTTGGCCGTATTTACTATCACGTGGACGGCTCAGAGAAGCCAGGAACGCTGCCACCAA
ACGTCTCAGCCGCTGTTAATGGAGTCGCCTTCTGCGGTACACTCGCAGGCCAGCTCTTCT
TTGGCTGGCTCGGTGACAAGCTTGGGAGGAAAAAAGTTTACGGTATGACCCTCATGGTCA
TGGTTCTTTGTTCCATCGCTTCGGGTCTATCCTTTGGGAGTAACCCAAAAACAGTGATGA
CTACACTATGTTTCTTCCGGTTCTGGCTTGGTTTCGGCATTGGTGGTGATTATCCGTTAT
CCGCCACGATTATGTCTGAATACGCTAATAAAAAGACCCGTGGTGCGTTTATAGCCGCCG
TTTTTGCTATGCAAGGGTTTGGAATCTTGACCGGGGGGATTTTTGCCATCATCGTGTCTG
CCGCTTTCGAGGCTAAATTTCCAGCTCCAACCTATCAAATAGATGCCTTGGCATCCACGG
TTCCTCAGGCGGATTATGTGTGGCGGATCATACTGATGGTGGGTGCTTTGCCTGCAGCGA
TGACGTATTACTCAAGGTCCAAGATGCCTGAGACTGCCCGGTACACGGCTCTTGTGGCCA
AGGACGCGAAACTAGCGGCTTCAAACATGTCTAAGGTTTTGCAAGTGGAGATCGAAGCAG
AGCAGCAGGGAACTGAAGATAAGTCCAATTCCTTTGGCTTGTTTTCCAAGGAATTCATGA
AACGCCATGGCCTTCATTTGCTAGGAACTACAAGCACATGGTTCCTACTCGACATCGCCT
TCTACAGTCAGAACCTATTCCAGAAAGATATTTTCAGCGCAATCGGATGGATCCCTCCGG
CTCAGACAATGAACGCAATTCAAGAGGTTTTCAAGATTGCACGTGCACAAACCCTAATCG
CCTTGTGTAGCACGGTACCTGGTTACTGGTTCACAGTGGCCTTTATTGACGTCATTGGTA
GATTTGCGATACAGATGATGGGTTTCTTCTTTATGACCGTCTTTATGTTCGCTCTCGCCA
TTCCTTACGACCATTGGACTCACAAGGAGAACAGAATCGGATTCGTTGCTATGTACTCTC
TAACATTCTTCTTCGCCAACTTTGGACCTAATGCAACAACTTTCGTTGTGCCCGCTGAGA
TCTTCCCGGCTAGGTTTAGATCAACCTGCCATGGAATTTCCGCGGCTTCTGGAAAATTAG
GAGCAATGGTTGGTGCGTTTGGGTTCTTGTACTTAGCACAAAGTCCTGACAAGACCAAGA
CAGAGCATGGATATCCTCCAGGTATTGGAGTCAAGAACTCCCTCATTGTTTTGGGTGTGG
TTAATCTTTTGGGGATGGTTTTTACACTGTTGGTTCCTGAATCTAAAGGCAAGTCTCTCG
AGGAAATGTCCGGTGAGAACGAACAAAATGATGAAAGCAGCAGCAGCAGCAACAACAACA
GTAACAACGCAGTGTCAACTGCATAGAGAATTCTAAAAGATTTTGTTTTTTACTCTTGTA
TCTTTGTTTTCTATGTATTTCTTCAGGTTTCTTCTCACTAGCAACTTAAGTTCTCTCTAT
GTTTACTTATTATTGATCGTTAAATGCAATAGGATCTGATTCTTTTTACATATAGAACCG
CTTTTCTTCTGTTCATTGTGACAAACGACAATGGGTTAATGGAAAGGTATTATGATGTAA
CTGTGTTTTTCTTTGTCATCACTGAACTAAAGAAGAATTGTTTGAATCATACATCCATTT



>chromosome:TAIR10:1:28678335:28682662:-1

ACTTTGACCCAAAATGCTATCTTCTGTTCTGTTAATAGATAAGATCAAATGTGAAATTCA
GAGGGAGGGGAGAGTTTGCATCCAAGGAGTTATACCTGAGATTGCGATTCCAAGTGACAC
GGGATGCTTATACAGAATGCAAGTGCAACAGCTCTGTCCGCCTGGTCCATTCTCAATCAC
GTTTAATCTTCCGGGACAAGTGGATCCTCGCTTGTTTTCTCCTAATTTTAGATCTGATGG
GATCTTTGAAGTCGTCGTAGTCAAGCTCGGAGTGCGCATCCCAACTTCATAGTTCCGGTT
TTCATTCAACCACGATATCCTCCATGCTTAATCAGCATTTAGTTGCAGTGGCCGACCTGT
AAGTCTTTGGTCGGATTTAGCAGAGGTTAGAGACTCAAGAATGTTCTAGTTTCTTTTGTG
TAGAACCATGTCCAATGATCCGGTTCCTTAAGATGTTCTTAGACTCTGAAAAAAAGCACC
TTGATGTGAAACGGAACTGTTTTAACCAAGATTTTGGCAATCACATCTTACAACTAAAAC
TCTTTGTACCAAAGAGGGAAATTTAAAACTTATAATCCAAATTTTATAAATAAAGTTGTC
AAAGTCAACTAAAGAAGTGGACTCGATCCAGGGAGGTGTTGTGTTCCTCACAGTCAAACC
AAAGCATATTCCATATTTGTCTCCAATAAGAAAAAATCTCCAAAAACACATTCAAATGAA
AAATACAAAGACTTCTATTGTTCTCATTCTCCGTATCTTGAAAAAATTCTTAAAAAATCG
ACGAAACGATGCCGGAGTTAAGTTTATTATCGGCGTTAGACGCAGCGAGGATACAGTGGT
ACCATTTCAAAGCGATAATAGTCGCCGGAATGGGACTTTTCACCGACGCATACGATCTTT
TCTGTATAGCTCCGATCATGAAAATGATAAGTCAAATATATTACCACAAAGACTCAATCG
GAACCGCTCTTCTCTCAACTTCTTACGCAATCGCTCTTCTTGGCACCGCGTTAGGTCAGC
TCATCTTTGGTTACTTAGGTGACCGAGTCGGACGTCGTAAAGTCTATGGACTTAGTCTCT
TGATCATGGTGTTTAGCTCCTTTGGTTGTGGCTTCTCTGTTTGCACGACTCGACGTTCTT
GTGTTATGGTTAGTCTTGGATTCTTTAGATTCGTTCTTGGACTTGGAATCGGTGGAGATT
ATCCTCTTTCGGCCACGATTATGTCGGAGTTTGCTAACAAAAGGACGCGTGGAGCTTTTA
TCGCTGCCGTGTTCTCGATGCAAGGGTTAGGGATTTTGATGAGCTCTGCTGTGACTATGG
TTGTGTGCTTGGCGTTCAAGAACGCTGGAGAGGGGAGTTCGGAGAAGACGAACGTGGCGG
GGTTAGAGACTTTGGCTCCGCCGGAATCGGATATAGCGTGGAGGTTGATTCTTATGATCG
GAGCTCTTCCCGCCGCGTTAACGTTTTACTGGCGAATGCTCATGCCTGAAACCGCCAGGT
TCGTATTTCTTTTTGTCTTTTTGTTCACCAATTTTTATCCTTTAAATCTTTAATATTACC
ATTGTATAGACTATAATTATATATACATATATATATATATATGCACTATATCTTGTGT
GACTTTGCTTCATATCATCCCACATACAAACCTTGGAAAATTTCTAGACTATTAAATAAC
TCTCTCCTAAAATTTTGTTTTTTAGAAATATTTTCCAAGATTAAAAATATTGAGATTTGAA
ACGGTAAGCATTTTCACCATGGATTGCTTCTATCGCAAACTAGATCGTTAGAGCATATTA
ATCCACAACAATAGTTATTACTTGTCATCACCATTCAAACTTTTCTGTATATAAACCCTT
ATTTTATGTATATCTTGCGTGACTTTGTTAGCTATTCAATGTAGAAATTATAAATTAAAT
CACTAAAAAGTTTAGTCGTGTGCAAAACAAAAGTTCTACTCATACGTCTCCGTACGAAAC
AATGTTGGAGAACCTGAAATTTAACATAAATTCGATCTGATGTTGATCAGATGATAAACA
TAAAATAATATATGTGAGCGATAATGTAAAGAGTGGCAACTTCGTGGTTTGTTACTAGGA
AGTACGTTGACTTGTGAGTTGGGAGGGAGATATACCAAACACAGTGGCATATACCTTTCT
ATGCTACCTCTTTTGATAACATATTTCAATAGTATTTCTTTTTATATGTATAATATATAG
TTCTGGTATGTTTTCCTGTATTAATTATTCCTCCCAGATATAAACTCCAGCTAACTAACA
GAGTTTTATTTCATTATTATAAATAACCACGTACTCTCTTCCTAACAAGTTCTCCCACTT
TAGTTTAATGCTCTAACTTCATACGAAAAAGTCTTAAGAGATTGGAATTTTTACATTCCG
TGTATTGTATGAGTTATGACTTATTATGACTAATTGTTAGCAGTGACATCTCTCACTAAT
AATAAACCTTATCTATGTTGACAAAAAAACCTAATTTGGCAGAAAAAATTGTTTTAAGTA
GCTAATTTGTTAGAAATCTCACATCTAATGTCAACCAACTACAGCTATTGTTACATATTT
TTTACAAAATTATAAATATTTGAACCTTCCTACAAAACTTAAATATGATAACTACTACAT
TTTATGACAATTCATGCAG**ATACACAGCATTGGTTGAGAACAATGTAGTCCAAGCAGCAA**
AAGACATGCAAAGAGTCATGTCCGTATCAATGATATCTCAAATAACCGAAGATTCATCAT
CGGAGCTAGAACAACCACCGTCTTCTTCCTCATACAAACTCTTCTCTCGCCGTTTCCTCA
GTCTCCACGGTCGAGACCTCTTTCGCAGCCTCAGCCAATTGGTTCCTAGTGGACGTCGTCT
TCTATACAAGCAATCTCCTTCTCTCTCAAATCTTTAACTTCTCCAACAAACCCCTCAACT
CCACAAACGTCTACGACTCTGCCTTTGAAGTAGCTAAGCTTGCAGCCATTGTAGCCGCTT
GCTCCACCATCCCTGGTTACTGGTTCACGGTCTATTTCATCGATAAAATCGGTAGAGTCA
AGATTCAGATGATGGGGTTTTTCCTTATGGCGGTCGTTTACTTAGTCGCTGGGATTCCGT
ATAGTTGGTATTGGTCTAAGCATGAGAAGACTAATAAAGGCTTTATGGTTCTCTACGGAT
TGATTTTCTTCTTTAGCAACTTTGGCCCTAACACGACGACGTTTATCATCCCGGCGGAGC
TTTTTCCGGCTAGGTTTAGGTCCACGTGTCACGGGATATCTGGAGCTGCAGGGAAGTTTG
GTGCGATTGTTGGCACTGTCGGTTTCTTGTGGGCCACGAGACACCATGAAGAGGACGGTT
TTCCAGATGTGAAACGTGTGAGAATCGCGTTTTTGATCCTTGGCGGCGTTTGTATCGCGG
GAATGATCGTAACGTATTTGTTCACGCGTGAAACTATGGGAAGATCGTTAGAAGAGAACG
AAGACGAGATTGTTTCCACGTCAGCAGGGTCGTCTCCTGCGAATGAGTTACTTCGGAGAC
AATACTAATTTCAAGATCACGACAAAAGCAATCGTGCTTTGATTTATTACTCTCTAATTA
ACTCTAAATGGTCCTTGGGTTTTTGATTGTTGATACTAAAGTTTCTTTTTTGTTTTTGTT
TTCTCTCAGTCTCATGTTTACATAAATTTTTGATTCCGTGCATATATTTGCAAGAACTCT
TCACTATGATCCTTACCTTTGCAAGAATTCTTCACTTAAGCCCATTTGGGCCCATATAGA
AAGAAAGAAAAATAGATACACGCAGGGCCGGCCCGTGGCATAAGCCACATAAGCGACGGC



>chromosome:TAIR10:1:7253398:7259552:-1

GTTTGGTCCTTTTGTGTGCAGGTGAATTCAGCTGTGAGATTGAATCAGATGGAAAGGTTA
TACTGGAGGGATCAACTACAAGAGGCGAAAAAAATATCAAGAGACATTCTCGGGTGTTCG
AGATGAATATCCGGAAGCTGTGTCCGCCTGGACCATTCAAGCTGTGCTTTAACCTCCCGG
GACCAGTTGATCCGCGGCTATTCTCTCCTAACTTCCGATCAGATGGTATCTTCGAGGGAG
TCATCATCCGACACAAAAACTCTTAATTAAACCGGAGGTTCCTATACAAGTTTAGAACTT
AGGATCTATGTAGATCTTTTATCTTTATGTTACTGGACATAGAAGGAAAGCGAATCAAAC
AGCATGCAAGGATAAATCTTTTACTTTTAGGACGTGACTCTGTAGATTTTGCATGATTTT
GCTAAATATCCGTTTTCACAATGATTTCATCAAACTGAGAATCCGTTAATATTGTTTCTG
TGTTGGCAGTGTTGCTGTTACTTGTTAGTAGTTACTGTTCTCATTTCTCAATAAGTGATC
AAACCGAAACTAAAACAGTACTGAACCGACGGTTTAATTTGGCTAGAAAAATTCAAACAA
TTATGATGTTGTGCCTTAATGTGACAACATAGATAAAAAGAAGCTCTTACAAACAAACTG
ACTAAGTCTTCCTCGGAGTCAAAATCATAAGCATATTCCATGCTCATACCAATAAAGTAA
TGAAAATGACAAAGAGACTTATTAATCGTTATAAAGATCCTTGTTCTGTTTCACTCATCA
TGAATTCGAAATTTCGATATAACAAACCAAAAGGAAAGAAGCAAAAAAAAAAAAAAAAAG
CTCTGGATTCGGAATCTCAGATAAAAAAACCAACGATGCCGATAAAGGTATTGTCGTCGC
TTGATGTAGCTCGAACACAATGGTACCATTTCAAAGCGATCATAGTAGCGGGAATGGGTC
TATTCACCGACGCATACGATCTTTTCTGCATAGCACCGGTCATGAAAATGATTAGTCATG
TTTATTACAACGGTGACTCAATCAACACAGCCGTTCTTTCAACTTCTTACGCCATCGCTT
TACTCGGCACAGCCACGGGTCAACTAGTCTTCGGTTACCTCGGTGACCGAGTCGGTCGTC
GCCGAGTCTACGGACTTTGTTTGATTATTATGATCCTTAGCTCCTTTGGTTGTGGCTTTT
CCGTGTGCACCACTCGTCGGTCTTGTGTTATGGTTAGCCTCGGTTTTTTCCGGTTCTTTC
TTGGTTTAGGTATCGGTGGTGATTACCCGCTCTCGGCTACCATTATGTCGGAGTTTGCTA
ATAAGAGGACACGTGGCGCGTTTATCGCTGCGGTGTTTTCGATGCAAGGGTTGGGGATTT
TGGTTAGCTCTGCCGTGACCATGGCTGTTTGCGTGGCGTTTAAGAGAAGTGGCGGTGGTT
TGGAGGTGGATGCGGCGGCTCCGACGGAAGCTGACTTGGCCTGGAGACTTATACTTATGA
TCGGTGCTCTTCCTGCTGCATTGACGTTCTATTGGCGAATGTTGATGCCTGAAACCGCAA
GGTCCGTCCACTCTTTTCATCCGTTTGTCTTTTTATATATTTTACCTATGACAAATCAGA
AAGGACGTAATGCACAAAGAATGGTCTTGAAAACCTTGAAATCACTATTTGTCCCACATT
AAAAATCTTGAAATCAACTAGATTTGATAATATCGTTTTATAGAAACATAATTGCATTAT
TTGTTCTGTTTATGTGAACGTTTCTCTAGTTTAAGAACATATAGTACCTCATTTCAACCC
ATAAAATGAACATTAAACTATTTTTTTCAAGCAATGATATAACTGTTCCGATCTCTCACA
CACGGACACGCACACACATACTATTGGATTCCTTAAAAGAAGTATTACTAGTTTTGACAT
ATTTTAATAGCTAGGTTTTATAAATATATGTATCAAATATTGAAAGGGTTACAAATTAGT
AGTTTGCTTCAATGGAGTCATTGACTTTATAAGGGATAGATACCATAATTCTGTGGTATA
TACTCTTGTCCTTACGTATTCAACATTAGCAATCCAAAGTCTACACCAATCATGTATTTA
CTCTCTTTTTTATCCATACAAAATTTCATTTATTTATTCTTTCATCTAATATTTATACAT
ACCTCAGGTATCAGATCCTATCAGCTCCAGCTAATCAACCAAGTTTTATTTTTTGATCAC
AACAAATTTAAATTATCCTCACACGTACTCTTCCACTTATTACAACATACACTAACAGAA
AGTCAAAAGCAACAAAAAAATCGATTTCCTAACACTTATGTTTTAGGGTAAATTGTACAC
AGTAGCCATTGACCCACTTTTAATTTCAGATTTTGACAAAGTGGGATGGGCCTTTCAGCT
TTTAGCCATTAAGCCAAAAACTGACCAAAATGCCCCTGACGACCTGACGAAACTGAACCC
GGTTTGTACGTACGGACCGCGTCGTTGGAATCCAAGTGAACCGGGTTCGTACGTACACTG
TTCATCGACGAGAAATGCATTAAATGATTTCTCGGTTTCGGCGAATTTGACACTTGAGAT
TGGAATTAAAAATTTTAAAAATTGGTTGGGCTTAGAAGTTTTTAATGCTAAAAAGTTTTG
ACAAGACTAAAACTCATGGTTACCTATTATAATAGAGATTATATTGTAAGGTATTTACAT
AATTGAGATTAGAATTCTGGGAAAAAAACGAAATTCGAAATTGTGTTACGGTGAAGAATC
TGTAGTTTTCAGTTGCTGTTTTACAAAGTTTGAAAGATTGAAGGGGAGACAATCTGTGTT
CTGGGTAAGAAATTCGTGTTAATCCTATTAATTACACTTGAGAACGTGATATGAAAGTAA
TGATACGTTTTTAAAAATTTCTGGGTTCATAAATTATTTTGTTGGAGATTCTTTTTGTAT
TTGTAAACCCAAAATTATCCAAAGCAAATTGAGAACCGAATGGGTATTCAAAATATAGTT
ATATACTTAACATAAATTATATTTAGACTTTTATATTCCAATTACCCAAATTATAATTAT
AAAATAGAAGTTAACGTACACTCGTACATTTATCAAAGAATATGTATGTACACTACTAAT
TATGTATCTGGTGAATTATGTAAAAATGAATGTTTTACGTACACGAATACATATAGGCAT
GGTAGATGTGAAATAAATTGTAATGGTGTGCAATTTAGTATTGGAAGATCGGTTTTACAT
GAATGAGAGTACGATTCGTTGAAATATTATATATGCATAATCTATATTAAAATATACAGC
ATACATATAATTTACATAATAGCTTAATCTGGATGTCAAAAAAATAAAATTAAAATTTTT
TCGTAGACAATGGTTAGAGAATTTTTAAACTCATTTTAATGTTGTACCATGTACGTTCCA
ATAAATATTTTTGTGTACGTACATAGTATTAGATTTATGTGATTTAGACATTTTTATGAT
GTAACTGTTTAGAGCGGTCGTATACGTGTACGTGTATCTCATATTATCGTGTATGTTTAA
GGGTTTAGTTTAAAATCTTGTTTATTCTCAACAATTTTAACTCACCAACAAAGATGTACG
TACATTATTATACATGTACGTACATAACATTTTTCAAAAGTATGAGGAACAACTATACTT
TGACCATATATTTTTATTTAATTATTATATATGAATAGATTTATGATGAAAACCCCAAAA
GATATTTTATCTAATCAAAATAACCAAAGGTTTCTAATGTACAAACTCAATAAATATTTT

GTATGCCGCATTTACAATAAAACAAAATCTAACATGTTTATTTCTATTTTAAATAATATT
TTATTTATTATCATTTGAAGACAATTCACGATTAGCAATATCATATAAACATATGTGCGT
AACGGTAATAATATATACGTACACGGTAGTAATGTGTACGTACACGGTAGTAATGTGTAC
GTACACGGTAATAATGTGTACATACACGGTAGTAATGTATGTAGTAATGTGTGTGTACGT
ATACCGTAAATTACATATACGTACATTTTTTTTAATGAAGGGTATAATTGTAATTTTGCA
CTCATCTTCAACCTCTGCAACCCTAATTTTGCAGCCGCTTGTTAACTCATTAACATCTGA
AACTCCTCGCCGCCGACTACTCTTCTTCATCTGCGCCATTTGAGTATGAACAATAATTCA
AGAAACCAGAAAATCCTCTTTTTGTTATTCTTTATCACGAGCTCGATCGAAATTTCCTCA
ACTTAAGCAACGACTAGAACTCGCACAGAGCCTCAATCGAAATTTCACTTTCGATCGGAG
TTTGTGTTTCGTCGGAATTGTTCTTCGTCTTCGACAAACAGTTTTTGATTTCCTTTGGTT
GTCAGAAAGTTAATAACGTGTTTTTGATTTAATTTTAAATTTAAATTTTGTAACCAGAAA
AATAAAGAAACGAAACTGAACCGCCTCCTTTCATTTCTCGTTCCAGGGCAAATTTGTCCA
TTTTTGGCTGAATGGCTAAAATCTGGTTGGGCGAACCCACTTTGTCATTTTGTGAAAGAA
AAAGTCAGAAAAGTGTCGAAGGGTGAAAATTTCTTTATGTTTTATTAATTAGTACTAATG
CATGCAG**ATACACAGCACTTGTAGAAAACAATATCGTCCAAGCAGCAAAAGACATGCAAA**
GAGTCATGTCAAGATCTCATATCTCCGACGAAGCCACGACGGATCCACCTCCTCCTCCGC
CGCCACCTTCCTACAAACTCTTCTCCCGCTGCTTCTTCCGCCTCCATGGTCGCGACCTCT
TCGCAGCCTCCTTTAATTGGTTCCTAGTAGACATTGTCTTCTACACAAGCAACCTCCTCC
TGTCCCATATCTTCAGTCATTACTCCAAGAAACCTTCCACTGCAGAAAACGTCTACGACG
CAGCCTTTGAAGTGGCGGAATTAGGAGCCATCATCGCTGCTTGCTCCACCATTCCCGGTT
ATTGGTTCACAGTTTACTTCATCGACAAGATAGGTCGTGTCAAAATCCAGATTATGGGGT
TTTTCTTCATGGCCGTTATTTATTTAGTCGCTGGTATTCCGTACAGTTGGTATTGGTCAA
AGCACGAGCATAATAATAAAGGCTTCATGGTTCTCTACGGTTTGGTTTTCTTCTTTTGCA
ATTTTGGTCCCAACACTACAACTTTTATTATTCCCGCAGAGCATTTCCCGGCTAGGTTTA
GGTCTACTTGTCACGGGATCTCTGGTGCCGCGGGTAAGCTTGGGGCTATTGTGGGCACTG
TTGGGTTCTTGTGGGCTACCAAGAAGATGGAAAGCGATGATAAAAACCAGATTTATCCCG
AAGTAAACCGTATGAGAATTGCTTTCTTGATTCTTGGTGGCGTTTGCATTGCTGGAATCT
TGGTGACGTATTTCTTCACCAAGGAAACTATGGGAAGATCGTTAGAGGAGAACGAGCATG
ACCAAGATAACAATGCTGAGTCTGAAGATGAGCCACAGATTGTTGACGGGCAATCCTCCG
TGAGCACTTTGCTTCAGACACGATGATGATCTCGACATATTGAACACATGATCAGTTCAA

Exons Bo9g059120 exons All exons in this region

Markup loaded

>chromosome:v2.1:C9:17429041:17431382:-1
ATAGATCTTCTATTTTATTATTTGTTGATTTGTGATTAAGTAAATAATTAATGATGTTTT
TGTTTAGAAAATATAAAAAAATTAATGATTTCTTAATCTATGTGCATAATTCTAAAACGA
CTTATATTAAAAAACGGAGGAAGTAACACTTAATGTTATATGACCAATGATAATATATCT
TCTATTTTATTATTGGTTGATTTGTGGTTAGGTAAATAATTAATGATGTTTTTGTTTAGA
AAATATAAAAAAAATTAATGATTTTTTAATCTATATGCATAATTCTAAAACGATTTATAT
TAAAAAATGAGGGAGTAATGTTTAGGTGCTTAACATGCTGTCTCTTAATTAAACACATGC
GTTTCTTTAATGTCTTCCTTTTAGACGTGAAACCCCTAATAAAGAGTCAAACTAGTTTTT
TTCCCCCTCCTATCCGACAGGAGGATCTAAATGTCTCCACATTACAATATTCATGTTTAA
GCCCTTTTTCAAAAAAAAAAATATATATATATATATATATTCATGTTTGCAAGAAAATGT
ATCAACACAAAAGCATTAACGTTTAGTACGTGAAAAGCAGGAGAGACGAGCAAAGCTGAG
ATGGCTGAACAACAACTAGGAGTGCTGAAGGCACTCGATGTTGCGAAAACGCAACTTTAC
CATTTCACAGCGATTGTTATCGCTGGTATGGGTTTCTTTACTGATGCCTACGATCTCTTC
TGCGTCTCCTTGGTAACCAAGCTTCTTGGTCGTCTCTACTACTTCAATCCATTGTCCGAA
AAGCCTGGCTCACTTCCCCCTCATGTCGCGGCTGCAGTCAATGGTGTGGCTCTCTGTGGG
ACTCTTGCTGGTCAGCTATTCTTCGGATGGCTCGGTGACAAACTCGGAAGGAAAAAAGTG
TATGGTATCACTTTGATCATGATGATTGTGTGCTCCGTGGCTTCTGGTCTATCTTTTGGT
AACAAAGCCAAGGGTGTCATGACCACCCTTTGCTTCTTCAGGTACAATTATCATTATAAT
TATATACTTAAACATGTTTTAATAATAATCTCATAGTGAATACGCATGTTTTAAAAATTA
CTGGTTTGGACTAGATATGTCTAGTCAAAATTTATATTAACGTACAATTTTATTAATTTA
G**GTTTTGGCTGGGGTTCGGTATTGGAGGTGATTACCCTCTTTCTGCTACCATTATGTCTG**
AATACGCTAACAAGAAGACTCGTGGTGCTTTCATCGCTGCCGTGTTTGCTATGCAAGGTG
TTGGTATCTTAGCCGGAGGTTTCGTGGCACTTGCTGTCTCTTCCATTTTCGACAAAAAGT
TTCCAGCTCCGACCTATGCAGTCAACAGGGCTCTCTCAACGCCTCCACAAGCTGACTATA
TTTGGAGAATCATCGTCATGTTTGGCGCTCTACCTGCAGCTTTGACTTACTACTGGCGTA
TGAAGATGCCTGAAACTGCTCGTTACACCGCTTTGGTGGCAAAGAACATCAAACAGGCCA
CACAAGACATGTCCAAGGTCTTACAAGTGGAGCTTGAGATGGAGGAAAGAGCAGAGGATA
TCGTTAAAGACCCTAGACTTAACTATGGCTTGTTCTCTAAAGAGTTTGCTAAACGTCATG
GTCTTCCCCTCCTTGGATGTACCTCCACATGGTTCTTGCTTGACATTGCATTTTACAGCC
AAAACTTGTTTCAAAAAATATCTTTTCCGCTATTGGATGGATCCCAAAGGCAGCAACCAT
GAACGCAATCCACGAGGTTTTCAGGATTGGTAGGGCACAAACACTCATTGCACTTTGCAG
TACCGTCCCCGGCTATTGGTTCACGGTTGCATTTATTGATATTATGGGAAGGTTTGCGAT
TCAGCTAATGGGTTTCTTCATGATGACTGTCTTTATGTTTGCAATCGCCTTCCCATACGA
CCACTGGA NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN
NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNTTATAATTTTGACCGAACTATGTTTACAA
GGTTACAAAATTTACGGCAAGGATCTCGTACGGTGGACGAATATGCAGAAGAATTCTATT
CGCTTATGACACGTAATGAAATTCATGACAGTGAAATTCAGCTAGTGTCATGCTTTATTG
GCGGTTTACGGCCTCAACTTCAGAATGCATTGGAACAATTTGATCAAAACACAGTTGCAG
AA

Ensembl Plants release 34 - December 2016 © EMBL-EBI

Exons Bo9g059110 exons All exons in this region

Markup loaded

>chromosome:v2.1:C9:17423760:17426071:-1
GCATGAGTATAATAAATTGAAGTCCCGTAAGATTGGGCCAGTGGAAGTCATTCACCGCAT
CAACGCAAATGCTTATAAGCTTCGTCTTCCTGATCACATTCGTACTTCAGATGTCTTCAA
TGTCAAGCACTTATTTCCATACCATGGTGAAAACGAGGTTCCAGATTCAGGGACGAATCT
TTCTCTACCCGGGGAGACCTGATGCAGCGCATCCTAACCTTAACGAACACGTCTTCCTTT
TCCTTGTTCGATGAGGAACCAATGTTTTCCTTTTCTCTTTACGTTCTACTTTTCTTTAAT
TTTCCTTTTCCTAGATAAGTTAATTTTCCTATTTCTTAGGCTTGGCTTTCTTATAAATAG
TGCTCTAAGGTTTGTAAGAGAGACACGATTGATTGAATTAATAAAGAGTTTTATTGCAAC
CGAGTTTTTACAAGAAGTGTATTCTTGCAAACCTCAAACCTCTGTTTTTCGTGTAGCAAA
ACAGAGCCTCTTTTCTGATTTTCGTGGAGAAAACCAGAAACCTAACTTGTTTCTCTAGCT
TCCCGGAACATGAGACTCAAGCTTTCCATCTATCCCTTCCG ATGCTACCATT
ATGTCTGAATACGCTAACAAGAAGACTCGTGGCGCGTTCATCGCTGCCGTGTTTGCTATG
CAAGGTGTTGGTATCTTAGCCGGAAGTTTCGTGGCACTTGCTGTCTCTTCCATTTTCGAC
AAAAAGTTTCCAGCTCCGACCTATGCAGTCAACAGGGCTCTCTCAACGCCTCCACAAGCT
GACTATATTTGGAGAATCATCGTCATGTTTGGCGCTCTACCTGCAGCTTTGACTTACTAC
TGGCGTATGAAGATGCCTGAAACTGCTCGTTACACCGCTTTGGTGGCAAAGAACATCAAA
CAGGCCACACAAGACATGTCCAAGGTCTTACAAGTGGAGCTTGAGATGGAGGAAAGAGCA
GAGGATATCGTTAAAGACCCTAGACTTAACTATGGCTTGTTCTCTAAAGAGTTTGCTAAA
CGTCATGGTCTTCCCCTCCTTGGATGTACCTCCACATGGTTCTTGCTTGACATTGCATTT
TACAGCCAAAACTTGTTTCAAAAAATATCTTTTCCGCTATTGGATGGATCCCAAAGGCAG
CAACCATGAACGCAATCCACGAGGTTTTCAGGATTGGTAGGGCACAAACACTCATTGCAC
TTTGCAGTACCGTCCCCGGCTATTGGTTCACGGTTGCATTTATTGATATTATGGGAAGGT
TTGCGATTCAGCTAATGGGTTTCTTCATGATGACTGTCTTTATGTTTGCAATCGCCTTCC
CATACGACCACTGGATCAAACCAGACAACCGTATCGGCTTCGTTGTTATGTACTCTCTCA
CTTTCTTCTTTGCTAACTTTGGACCAAATGCAACCACTTTCATTGTCCCAGCTGAGATCT
TCCCGGCCAGGCTAAGGTCCACATGCCATGGAATATCAGCTGCGACCGGTAAGGCTGGAG
CCATCGTGGGAGCTTTCGGGTTCCTGTACGCAGCTCAACCACAGGATAAGACCAAGACGG
ATGCAGGATATCCACCGGGCATTGGAGTCAAGAACTCTCTGATCATGCTTGGTGTCATTA
ACTTTGTCGGCATGCTCTTCACCTTCCTTGTCCCTGAACCCAAGGGTAAGTCACTTGAAG
AGCTCTCTGGTGAGACGGAGGCTGAGAAATGATTAAGCCGTCATATTGTCAATTGTTTTT
GTTGTTTCCCGTTATTTAAATCTTCTTGTGCAATTTATTTTTTGTCAAGTGCAATTTACT
TTTTTCCAATGGGTGATGTGATGATATTTGTAATTGTTCACCTCTACGTTTGTTTTGTGT
GAAAGATCAATCAATACTTGGTGGGCATTGTTATATGGAGTTGATTGCTTACTTTTGTAA
CACGATATGTTTATATTTCGTTAATACATGTCTACCTCTTGCACAACAACTTAACATATA
GCTGAATAAAGTCTCCATATAAAAGTCTCAAAAATAAGTTTACAAAGAGAAACTAATAAA
TAACCTAACTTAAATTAAACGATATCATACACACACAGATATACAATTGGTCTGAGTTGG
CTCGATTGATTTTAAAATATTTACTATGGATATATGATAGCAACTAAGTCCAACTTTATA
TATTAAATTGAATACATCTCCCTAAGTTCTTCCGAATATCCTATGTTTTTCGAATATGCT
TTGTTTTAATAATGGTTTTGACTTTTGCTAAATGATAATTAATCTTCATGCCTATATGTA
TCCAGATTGTCACTCCTCAAATGTCAACTGTT

Ensembl Plants release 34 - December 2016 © EMBL-EBI

Exons Bo4g185840 exons All exons in this region

Markup loaded

>chromosome:v2.1:C4:49031700:49033397:1

TTAGGATTAGTTCTCTTAAAACGTATTTGTTGGAATTTATCTCTTCATGAATACAAGACA
TTTTTTGGGTGCATAAATCAAATGATTCACAAATAAACAAATGACGAAAATAATGTTACT
TCAATAGACAGGAAAAATATAATGTCCTAATTATATAAGAGGAAATGTTTAGTTTGATGA
GTAAATCTGTGAGCAATTACAAACAACTTGCAGTCTCACTCACATTAGTGTTTTGTATCC
TCGTGTTCAGATTTAGATGTCACAAGTCAACCTTTGGGAGTAACAAGAGTAGAGTATACT
CATGTAGTTTATCAATGGCTCTAAAATTGCCAATTACCTCCTTCTTCAGGTTTGGTCAAC
ATGCTCTTTTAAGTATTACGTCAACAAACTTCATGGATAGTGAGTGTGAGATCAACGATA
CGAGAAAATAAAATAAAATTAAAGTAGAGTTTAAATGTTTAATTCTGTTTGCTTCTAAAA
TTATATGATTCATAAGTCAATCTTATAACACTCAAGAATGTTAGCAAGTTACTAATACGG
AATTTTTAGTGAAACTACAGAGCTGAAGACGAATGGCAAGTGATCAGTTACAAGTGTTGA
ATGCACTTGATGTGGCAAAGACGCAATGGTACCATTTTACAGCGATTATAATTGCCGAA
TGGGATTCTTCACCGACGCTTACGATCTCTTTTGCATCTCACTTGGAAG**GCTCTCCGAAG**
CCTGGCACTCTCCCTCCTAACGTTGCTGCTGCCGTTAACGGCGTTGCCTTCTGTGGGACC
ATCGTCGGTCAGCTCTTTTTTGGGTGGCTTGGTGATAAGCTCGGGAGGAAGAAAGTTTAT
GGTATGACGTTGATGGTCATGGTCATTTGCTCCGTAGCCTCCGGTCTCTCTTTCGGGCAC
GAGCCAAAAGCCGTGATGTCCACGCTATGTTTTTTTAGGTTTTGGCTAGGGTTTGGGATC
GGTGGTGACTATCCTTTATCCGCAACGATCATGTCTGAATACGCTAACAAGAAGACTCGT
GGAGCGTTTATTGCTGCGGTTTTCGCGATGCAAGGGTTTGGAATCATGGCTGGTGGTATC
TTTGCTATAATTATTTAATCTGTTTTTGAAGCTAAGTTCCCTGCGCCGGGCCTATGCGGA
TGATGCCTTGAGATCCACGGTATCTCAAGCAGATTTTGTATGACGTATAATCTTGATGGT
TGGTGCTATCCCTGCCGCAATGACGTATTACTCACGGTCCAAGATGCCGGAAACCGCGAG
ATACACAGCTTTGGTTGCTAAAGATGTAAAGCAAGCTGCTTCAGACATGTCTAGGGTCT
GCAAGTGCAGATAGAGACAGAAAAGCAGGAAGTGGAGACCAAAGAGTTTTCCTTGTTTCC
CAAGGAATTTATGAAGCGTCATGGGCTTCATTTGCTCGGCACTACATCAACCTGGTTCCT
CCTCGACATTGCTTTCTTCAGTCAAAACCTCTTTCAAAAGGATATTTACAGCGCAATCGG
GTGGATTCCTCCGGCTCAAACCATGAACGAGATTCAAGAAGTGTTCAAGATCGCATGTGC
ACAAACCATCATAGCCTTGTGTAGCGCGGTGCCTGGTTACTGGTTCACAGTTGCGTTCAT
CGACGTTAAGATTCACGATTCAGATAATGGGATTTTTCTTCATGATGGTTTTTATGTTTG
CTCTGGCTATACCCTACA

Exons Bo4g185850 exons All exons in this region

Markup loaded

>chromosome:v2.1:C4:49032295:49034398:1

GTTGAATGCACTTGATGTGGCAAAGACGCAATGGTACCATTTTACAGCGATTATAATTGC
CGGAATGGGATTCTTCACCGACGCTTACGATCTCTTTTGCATCTCACTTGGAAGGCTCTC
CGAAGCCTGGCACTCTCCCTCCTAACGTTGCTGCTGCCGTTAACGGCGTTGCCTTCTGTG
GGACCATCGTCGGTCAGCTCTTTTTTGGGTGGCTTGGTGATAAGCTCGGGAGGAAGAAAG
TTTATGGTATGACGTTGATGGTCATGGTCATTTGCTCCGTAGCCTCCGGTCTCTCTTTCG
GGCACGAGCCAAAAGCCGTGATGTCCACGCTATGTTTTTTTTAGGTTTTGGCTAGGGTTTG
GGATCGGTGGTGACTATCCTTTATCCGCAACGATCATGTCTGAATACGCTAACAAGAAGA
CTCGTGGAGCGTTTATTGCTGCGGTTTTCGCGATGCAAGGGTTTGGAATCATGGCTGGTG
GTATCTTTGCTATAATTATTTAATCTGTTTTTGAAGCTAAGTTCCCTGCGCCGGGCCTAT
GCGGATGATGCCTTGAGATCCACGGTATCTCAAGCAGATTTTGTATGACGTATAATCTTG
ATGGTTGGTGCTATCCCTGCCGCAATGACGTATTACTCACGGTCCAAGATGCCGGAAACC
GCGAGATACACAGCTTTGGTTGCTAAAGATGTAAAGCAAGCTGCTTCAGACATGTCTAGG
GTTCTGCAAGTGCAGATAGAGACAGAAAAGCAGGAAGTGGAGACCAAAGAGTTTTCCTTG
TTTCCCAAGGAATTTATGAAGCGTCATGGGCTTCATTTGCTCGGCACTACATCAACCTGG
TTCCTCCTCGACATTGCTTTCTTCAGTCAAAACCTCTTTCAAAAGGATATTTACAGCGCA
ATCGGGTGGATTCCTCCGGCTCAAACCATGAACGAGATTCAAGAAGTGTTCAAGATCGCA
TGTGCACAAACCATCATAGCCTTGTGTAGCGCGGTGCCTGGTTACTGGTTCACAGTTGCG
TTCATCGACGTTAAGATTCACGATTCAGATAATGGGATTTTTCTTCATGATGGTTTTTAT
GTTTGCTCTGGCTATACCCTACAACCACTGGACTCACAAG**GAGAACCGGATCGGTTTTGT**
TATCATGTACTCGTTAACATTCTTTTTCGCCAACTTTGGACCTAATGCTACAACCTTTAT
CGTACCGGCCGAAATCTTCCCGGCCCGGTTTAGATCCACTTGCCATGGTATCTCTGCAGC
ATCAGGAAAAGTAGGTGCAATAGTTGGTGCTTTTGGGTTCTTGTACTTGTCCCAAAGTCC
TGACAAGGACAAGACTGATGCAGGATACCCTCCAGGGACTGGGGTCAGGAACTCACTTAT
TGTGTTAGGGGTGGTTAACTTCTTGGTTTTCTCTTCACTTTCTTGGTGCCTGAAGCTAAG
GGAAAGTCGCTTGAGGAAATGTCGGGTGAAAATGAAGACAACAAAATTAGAACGGTCCCA
ATAGTATATTAGGTTATATAGTACACCTTTTGTTTCTTCTTCTTCTCTTGTTGTGTTCAG
CAACTTAAATTGTTACTTCTTCATGTTGATACGTGCAGATGGGCGCAACATGAGGTGGAA
GGTGGGGACTATTACCATCGGGAAATATCAATTTTACGATTCTCAGATTTTAAAACATCT
AGTATTATAAAGTGGGTAAGTGTAATATCTCCATTAATTGTATAAGTGGAGCTTTTTAAA
TGGTCAGTATGAATTTAAGACGAGCCCATTTAAGCCTATTTTTACCAAAATTTGATCCAA
ACTTGTATAAATACCATTATTATGAATTGAACTCTAAGTTTGATTTCTTAAAGAGATTTC
CAAGATCTTGTGTTCGTCATAATGGATGAATCAAACTTCAATCTTTTTGCTAATCCTACT
ATATATTAAGTGAGAAGTCACTTAAATGACTTGTCTTAAGTGTCATTCATAAGTGAAGT
TTAGAATTAATTTTTTTAATTTGATTGGTTGTTGGTATTTGAATTTTTATTTATTTAATT
AAAGTTTTAAAAAGAAAACTAATACTCTTCCGAATCTAATCTATACTACCATACAAACAA
TTAA

Exons Bo5g030740 exons All exons in this region

Markup loaded

>chromosome:v2.1:C5:10579022:10581313:-1

ATAATATAATCTAGCTAAACTCTAAAATCTTGTCTTTTTTTTGTAACACATAACTCTAAA
ATCTTGTCTAATTGCGTCTTTTCCCATGAAATGAACATGTTCCATATAAAACCACTAAAG
AAAACAAAAAAAAAAGTCCAAAAGTCGAAACAAATCAATCCGATTAAAAAGCAAAGCAAA
AGCTCAAAGAAACACAACACATCACTTCTGACCTTCGGTCTACGTAGTGTTGTTTCCGCG
AGGTTTTCATGTTAAGACTGTCTGCATGTTGTTAACACATCCCTTCACCAGTGAGTTAAA
CCAAAATAAATGGATTAATACAAAAGCATATGGCTTAGTCTGCCGGAGATTCTCAAACAA
TTAATAATGTGCACGTCGTGACAACACAAAATAGAATATAACAACTCACCAAGTCTTTCT
CGGATTCAAGCATAAGCAGCATATTTCATGAGCATAGAAATAATGTAATGAAAAACAGAC
AGAGACTTCTTGTACCTAACTTATTATAAGTACAATCAATAATGTCCTGTTTCAATTTAT
TGTGCCTACGAAGTCCAAACCTTCGATAAACAAAACCAAATAGAGGTGAAGAAGGAAGCT
ATGAATTCAGAATCGCAGATCAAGAAAAGGACGTTGACACCAAAGGTAATGTTGTCTCTT
GATGCAGCTAGAATACAATGGTACCATCTCAAAGCGATCATAATAGCAGGAATGGGACTA
TTCACCGACGCTTACGACCTCTTCTGCATCGCACCGATCATGAAAATGATTAGCCATATT
TATTACCACGGAGACTCCATCACCACAGCCGTTCTCTCCCTTTCTTACGCCATCGCTCTT
CTCGGGACAGCAACGGGTCAACTCGTGTTCGGTTACCTCGGCGACCGAGTTGGTCGCCGC
CGAGTTCTACGGGATTTGTCTGATCATCATGATCCTAAGCTCCTTTGGTTGTGGCTTTTCC
GTGTGCACGAGTAGTAGGTCTTGCGTTATGGTCAGTCTCGGTTTCTTTCGGTTTATTCTC
GGTTTGGGTATCGGTGGTGATTACCCGCTCTCGGCTACCATCATGTCGGAGTTTTCGAAT
AAGAGAAGACGTGGCGCCTTTATAGCTGCCGTGTTCTCCATGCAAGGGTCGGGGATTCTT
GTGAGCTCAGCCGTGACGATGGCTGTTTGTGTGGCGTTTAAAACTTTGGCGGCGGTTCG
AAGGCTGATGGAGAGGCTCCGGTTGAAGCTGACGTGGCGTGGAGGCTGATCCTTATGATC
GGTGCTCTTCCCGCTGCACTGACGTTGTACTGGCGAACGTTGATGCCTGAAACCGCGAGG
TTCGTCCAAAACTTTCTCCATATTTTTGTACTGATATTAATTAATTAGTATTATTTAAAA
AAAGTCTCTGTGCTTTATTTATATATTTTATATTTCCAATCACAAATCAGAAGGGACGTA
CGGCATAACGTTTTTGCCTCACTATTTATTGGCTACTTGTCCCACATATAAAATATTGGA
AGCTAGAAGATTTAAGCTTATATCATTCTATAGAACTAATATCATTTTTACCCTATTTTG
TAGTTTTCTTTCCTTTGGGAAGAAAACAGCAGTTTCATATATTTTACAATTTTTTGCAGT
TTTAACATATGTTGGAAGATTTAAATATAAAATTATTCAAAAAGACATTCTGCTTCTTCAA
CATATATTTTAAAAACAAAACGATAGATGCCATGATTTCAATTAAACATTATTTAAACTC
TAGTTTTTAATAAATATATTGTACTTTAACCAATGAAGACCATAAAAACGTTTTTCTAGT
CTATTGAAAATGTTAAGAGTTGATGGTGGGTTTAGAACTATGGTTTGATGACAATCTATA
TATTGTATATAATATTAGATTCCTTAACAATTTTTGTTAGTTGTATATTTTAATAGCTAT
CTTTTATTGATATATATATCAGTTATTCAGAGGGTTAACAAATTAGTTGTTTGTTTCAAA
TGAGTCATTGACATTGTAAGGTTAACAAATTAGCATCTATGAATACTAAAATATTTTTTT
GGAATGTCCGCGTATTGAATGATCCGGACAAGCACACCCTTTTTGTCAATGGCTAGCGAG
CCAACAACCTTCTTTTGACATCATTTTGGAATTACATCAAGAATCATAACCTGAGCCC
TTTAATGACCAAACTTTGTAGAGGCTGGAACTATGCCTCTAATCACTCTATGGACGTTGA
AGGTCGCATAATTGTTATCTGGAAGGAAACTGTCTCCGTACGCATTCTGCATCAATCGAG
GCAAGCTGTAAC

Exons Bo5g030730 exons All exons in this region

Markup loaded

>chromosome:v2.1:C5:10569947:10572010:-1

AGCAGGGTCGCAGCTAATAATTTGAGAAATAGTGTAATAAATGGAACACCAGATTTAACG
TGGAAAACCCCAAAGGGTAAAAACCACGGACAAGGTAAAAGAATTTATTACGAGAAAAAT
AATAGGAGTTACAACCTCTCAATTATAACGGAGAAATCGATTAATAATTTTCTCGTTATA
ATAGGAGAAAATACTCAAACTCTCTAAATAATTTTCTCTCAACCCGATCCCAAATACCCG
TAAGAAAAAAGAATTTTTTTTTTTCTCTCTCTCACGTTCTTCTTCTTCTTCACTTCTCTT
TGTGATTTTGATTTTGGGATATCTTTCCTCTTCAGCTTAGCTCTCCTTTTATAGAAGAAG
TTTGGACATGTGAAAGCCATAATTATTGACAAAATCACCTCCTTCACCGTCCAACTTCAC
CGTCCATCACCGTCCATTATGAAAAACGTGTTATTTGACTTACAGTTTGTACTCTAGTCC
GTACG**ATAGAAAAGGAAAAAGT**CAAAATACCAAAAACTCGATTTCGTAAATGT
AACCATGCATGCAGGTACACAGCACTTGTGGAACACAATATCGTCCAAGCAGCAAAAGAC
ATGCAAAAAGTCATGTCAATATCGAATATCTCTGACGAAACAACGACGGAGCCACCTTCT
CCTTCACCACCGCCTACCTACAAGCTCTTCTCCATTCGCTTCCTCCGCCTCCACGGCCGT
GACCTCTTTGCCGCATCCGCTAACTGGTTCCTAGTAGACGTCGTCTTCTACACAAGCAAC
TTCCTCCTCTCCCATATTTTCAGTAGCTACTCCCACAACCCTTCTTCTTCCGCCACCACA
AACGTCTACGACGCAACCTTCGAAGTGGCGAAACTAGGAGCCATCATCGCCGCTTGCTCT
ACCATTCCCGGTTACTGGTTCACAGTTTTCTTCATCGACAAGGTAGGTCGTGTCAGGATC
CAGATTATGGGGTTTTTCTTCATGGCCGTTATTTTTTTGGCCGCGGGGATTCCGTACTGT
TGGTATTGGTCTAAGCATGAGCAGACAAAGAAAGGCTTTATGGTTCTTTACGGTTTCGTT
TTCTTCTTTTGTAACTTTGGTCCCAACACTACAACTTTTATCATTCCCGCGGAGCTTTTC
CCGGCTAGGTTTAGGTCGACTTGCCCACGGGATATCTGGAGCCGCGGGTAAGCTTGGGGCT
ATTGTGGGCACTGTTGGATTCTTGTGGGCTACGAGAGAGGTAGAAGACGGCGATGATATA
AACCAGGTTTATCCTGATGCGAACCGAATGAGAACTGCTTTCTTGATTCTTGGTGGAGTT
TGTATTGCTGGAATGTTTGTGACGTATTTCTTCACGAAGGAAACTATGGGAAGATCGTTA
GAAGAGAATGAGCACGAGCACGAACGAGATAAGAATGTTGAGGACGAAGAGATTGGTGAC
GGGCAATCCTTGGCAAGCACTTGATGACCAAGATACCAAACTCTTGAATCAGTTCAAGTT
TACTCAGACAAAAGGAAGCCAAATGTATATTTCTTACCAAATAAAACCGTTTAACATGTA
AGATCACTTCCAATGAGGTTCTACCCATTGGAGTTCTAAACATGAATATGTGTATATGGA
TATATTGTATATGTGTATGTAGTTATCGGATCCACTAAATAGTGTAGAACCCCTACCTAA
AGGTTAGAGCACTTTCAATGGGGTTCTACCCTTTGGGTAAGGGTTCTACACTATTTAGTG
AACCCCACAACTACATATACATATACAATATATACATACATACACACATATTCATATTTA
GAACTCCAATAGGTAAAACCTCCATGACAGGTGCTCTAAGCTTCCTCTGTTTGTACTTTC
TACACATGCAACTGTTCAATATACAACAGTCTTTCTTTGCTTTTCATCAGTTTAGTTAAT
TCTCTCTGTTTCTCTTCAGTCATCATTTTATATTCAGTTCTATAAAGGTCATCATCACTG
GGTAGAAAATTGAGTTTCTCAACCAAGTTTTAATAATAAATAATACTGGACAATGAAAAA
ATGACAAATATTAAAATAGTTTTT

Exons Bo5g030740 exons All exons in this region

Markup loaded

>chromosome:v2.1:C5:10579022:10581313:-1

ATAATATAATCTAGCTAAACTCTAAAATCTTGTCTTTTTTTTTGTAACACATAACTCTAAA
ATCTTGTCTAATTGCGTCTTTTCCCATGAAATGAACATGTTCCATATAAAACCACTAAAG
AAAACAAAAAAAAAAGTCCAAAAGTCGAAACAAATCAATCCGATTAAAAAGCAAAGCAAA
AGCTCAAAGAAACACAACACATCACTTCTGACCTTCGGTCTACGTAGTGTTGTTTCCGCG
AGGTTTTCATGTTAAGACTGTCTGCATGTTGTTAACACATCCCTTCACCAGTGAGTTAAA
CCAAAATAAATGGATTAATACAAAAGCATATGGCTTAGTCTGCCGGAGATTCTCAAACAA
TTAATAATGTGCACGTCGTGACAACACAAAATAGAATATAACAACTCACCAAGTCTTTCT
CGGATTCAAGCATAAGCAGCATATTTCATGAGCATAGAAATAATGTAATGAAAAACAGAC
AGAGACTTCTTGTACCTAACTTATTATAAGTACAATCAATAATGTCCTGTTTCAATTTAT
TGTGCCTACGAAGTCCAAACCTTCGATAAACAAAACCAAATAGAGGTGAAGAAGGAAGCT
ATGAATTCAGAATCGCAGATCAAGAAAAGGACGTTGACACCAAAGGTAATGTTGTCTCTT
GATGCAGCTAGAATACAATGGTACCATCTCAAAGCGATCATAATAGCAGGAATGGGACTA
TTCACCGACGCTTACGACCTCTTCTGCATCGCACCGATCATGAAAATGATTAGCCATATT
TATTACCACGGAGACTCCATCACCACAGCCGTTCTCTCCCTTTCTTACGCCATCGCTCTT
CTCGGGACAGCAACGGGTCAACTCGTGTTCGGTTACCTCGGCGACCGAGTTGGTCGCCGC
CGAGTCTACGGGATTTGTCTGATCATCATGATCCTAAGCTCCTTTGGTTGTGGCTTTTCC
GTGTGCACGAGTAGTAGGTCTTGCGTTATGGTCAGTCTCGGTTTCTTTCGGTTTATTCTC
GGTTTGGGTATCGGTGGTGATTACCCGCTCTCGGCTACCATCATGTCGGAGTTTTCGAAT
AAGAGAAGACGTGGCGCCTTTATAGCTGCCGTGTTCTCCATGCAAGGGTCGGGGATTCTT
GTGAGCTCAGCCGTGACGATGGCTGTTTGTGTGGCGTTTAAAAACTTTGGCGGCGGTTCG
AAGGCTGATGGAGAGGCTCCGGTTGAAGCTGACGTGGCGTGGAGGCTGATCCTTATGATC
GGTGCTCTTCCCGCTGCACTGACGTTGTACTGGCGAACGTTGATGCCTGAAACCGCGAGG
TTCGTCCAAAACTTTCTCCATATTTTTGTACTGATATTAATTAATTAGTATTATTTAAAA
AAAGTCCTGTGCTTTATTTATATATTTTATTCCAATCACAAATCAGAAGGGACGTA
CGGCATAACGTTTTTGCCTCACTATTTATTGGCTACTTGTCCCACATATAAAATATTGGA
AGCTAGAAGATTTAAGCTTATATCATTCTATAGAACCTAATATCATTTTACCCTATTTTG
TAGTTTTCTTTCCTTTGGGAAGAAAACAGCAGTTTCATATATTTTACAATTTTTTGCAG**T**
TTAACATATGTTGGAAGATTTAAATATAAAATTATTCAAAAAGACATTCTGCTTCTTCAA
CATATATTTTAAAAACAAAACGATAGATGCCATGATTTCAATTAAACATTATTTAAACTC
TAGTTTTTAATAAATATATTGTACTTTAACCAATGAAGACCATAAAAACGTTTTTCTAGT
CTATTGAAAATGTTAAGAGTTGATGGTGGGTTTAGAACTATGGTTTGATGACAATCTATA
TATTGTATATAATATTAGATTCCTTAACAATTTTTGTTAGTTGTATATTTTAATAGCTAT
CTTTTATTGATATATATCAGTTATTCAGAGGGTTAACAAATTAGTTGTTTGTTTCAAA
TGAGTCATTGACATTGTAAGGTTAACAAATTAGCATCTATGAATACTAAAATATTTTTTT
GGAATGTCCGCGTATTGAATGATCCGGACAAGCACACCCTTTTTGTCAATGGCTAGCGAG
CCAACAACCTTCTTTTGACATCATTTTGGAATCACATATCAAGAATCATAACCTGAGCCC
TTTAATGACCAAACTTTGTAGAGGCTGGAACTATGCCTCTAATCACTCTATGGACGTTGA
AGGTCGCATAATTGTTATCTGGAAGGAAACTGTCTCCGTACGCATTCTGCATCAATCGAG
GCAAGCTGTAAC

Exons Bo5g030730 exons All exons in this region

Markup loaded

>chromosome:v2.1:C5:10569947:10572010:-1

AGCAGGGTCGCAGCTAATAATTTGAGAAATAGTGTAATAAATGGAACACCAGATTTAACG
TGGAAAACCCCAAAGGGTAAAAACCACGGACAAGGTAAAAGAATTTATTACGAGAAAAAT
AATAGGAGTTACAACCTCTCAATTATAACGGAGAAATCGATTAATAATTTTCTCGTTATA
ATAGGAGAAAATACTCAAACTCTCTAAATAATTTTCTCTCAACCCGATCCCAAATACCCG
TAAGAAAAAAGAATTTTTTTTTTTCTCTCTCTCACGTTCTTCTTCTTCTTCACTTCTCTT
TGTGATTTTGATTTTGGGATATCTTTCCTCTTCAGCTTAGCTCTCCTTTTATAGAAGAAG
TTTGGACATGTGAAAGCCATAATTATTGACAAAATCACCTCCTTCACCGTCCAACTTCAC
CGTCCATCACCGTCCATTATGAAAAACGTGTTATTTGACTTACAGTTTGTACTCTAGTCC
GTACGG**ATG**AATAGAAAAGGAAAAAGTCAAAATACCAAAAACTCGATTTCGTAAATGT
AACCATGCATGCAGGTACACAGCACTTGTGGAACACAATATCGTCCAAGCAGCAAAAGAC
ATGCAAAAAGTCATGTCAATATCGAATATCTCTGACGAAACAACGACGGAGCCACCTTCT
CCTTCACCACCGCCTACCTACAAGCTCTTCTCCATTCGCTTCCTCCGCCTCCACGGCCGT
GACCTCTTTGCCGCATCCGCTAACTGGTTCCTAGTAGACGTCGTCTTCTACACAAGCAAC
TTCCTCCTCTCCCATATTTTCAGTAGCTACTCCCACAACCCTTCTTCTTCCGCCACCACA
AACGTCTACGACGCAACCTTCGAAGTGGCGAAACTAGGAGCCATCATCGCCGCTTGCTCT
ACCATTCCCGGTTACTGGTTCACAGTTTTCTTCATCGACAAGGTAGGTCGTGTCAGGATC
CAGATTATGGGGTTTTTCTTCATGGCCGTTATTTTTTTGGCCGCGGGGATTCCGTACTGT
TGGTATTGGTCTAAGCATGAGCAGACAAAGAAAGGCTTTATGGTTCTTTACGGTTTCGTT
TTCTTCTTTTGTAACTTTGGTCCCAACACTACAACTTTTATCATTCCCGCGGAGCTTTTC
CCGGCTAGGTTTAGGTCGACTTGCCACGGGATATCTGGAGCCGCGGGTAAGCTTGGGGCT
ATTGTGGGCACTGTTGGATTCTTGTGGGCTACGAGAGAGGTAGAAGACGGCGATGATATA
AACCAGGTTTATCCTGATGCGAACCGAATGAGAACTGCTTTCTTGATTCTTGGTGGAGTT
TGTATTGCTGGAATGTTTGTGACGTATTTCTTCACGAAGGAAACTATGGGAAGATCGTTA
GAAGAGAATGAGCACGAGCACGAACGAGATAAGAATGTTGAGGACGAAGAGATTGGTGAC
GGGCAATCCTTGGCAAGCACTTGATGACCAAGATACCAAACTCTTGAATCAGTTCAAGTT
TACTCAGACAAAAGGAAGCCAAATGTATATTTCTTACCAAATAAAACCGTTTAACAATGTA
AGATCACTTCCAATGAGGTTCTACCCATTGGAGTTCTAAACATGAATATGTGTATATGGA
TATATTGTATATGTGTATGTAGTTATCGGATCCACTAAATAGTGTAGAACCCCTACCTAA
AGGTTAGAGCACTTTCAATGGGGTTCTACCCTTTGGGTAAGGGTTCTACACTATTTAGTG
AACCCCACAACTACATATACATATACAATATATACATACACACACATTCATATTTA
GAACTCCAATAGGTAAAACCTCCATGACAGGTGCTCTAAGCTTCCTCTGTTTGTACTTTC
TACACATGCAACTGTTCAATATACAACAGTCTTTCTTTGCTTTTCATCAGTTTAGTTAAT
TCTCTCTGTTTCTCTTCAGTCATCATTTTATATTCAGTTCTATAAAGGTCATCATCACTG
GGTAGAAAATTGAGTTTCTCAACCAAGTTTTAATAATAAATAATACTGGACAATGAAAAA
ATGACAAATATTAAAATAGTTTTT

Exons Bo6g120500 exons All exons in this region

Markup loaded

>chromosome:v2.1:C6:37997479:37999524:-1

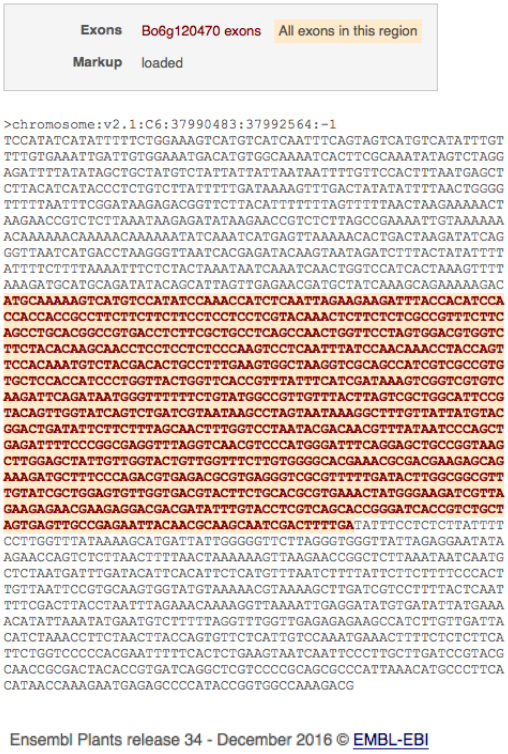
TTTTTTTTTAGTTTTAACTAAAACTAGGATAAAAATATCTTTTCACGTGTAGGAAAGTGT
AGTTTTCAAAAAAAATTAAAAAATAAATGTATTTTTCAAATTCTAAATCATAAATGGTATA
TTTTTCAAATTATCCCTTGATATTATGTTGAAATATATATTATGAAAGCTAATATACTAC
TCCATTCGTTTTATTTTAAGTGTCGTTTTACATTTATGTTCACAAATTACGAAAATATTT
AATTTTTTCTATTTTCAAAATAAAAATAGTATTACATGTACATCTGACTATATTTTAACT
AATAATAAAAATAGGAGTATAAGTTTTTCTCGAAAAATGTTGAATGTTATATAACTTAGTT
TTAAAATGCTGCCACATAACTTTACAAAGTTTTAATACATGTTCTACTAACTAAAATTAC
ACTTAGAGGATTTTACAGGGAATTACGAGTAACAAATCTTTACTAAACTTTCCTTTTAAA
AATACTTTTCTCCTAAATAATAAAATCAGTAGTTCAGTTACTAAAATTTAAAGATGCTGC
AGATACACAGCATTAGTTGAGAACAATGCAACTCAAGCAGCAAGAGACATGCAAAAAGTC
ATGTCCGAATCCATATTATCTCAAATAGCCGAAGATTCATCGTCGGAGCCACCACAACCA
CCCTCTTCTTCCTCATACAAACTCTTCTCTCGCCGTTTCTTCAGCCTCCATGGCCGTGAC
CTCTTCGCTGCCTCAGCCAATTGGTTCCTAGTCGACGTCGTCTTCTACACAAGCAACCTC
CTCCTCTCTCAAATCTTCACCTTCTCCAACAAACCTCCCAATTCCACAAACGTCTACGAC
TCAGCCTTCGAAGTGGCTAAGGTCGCTGCCATCGTAGCCGCGTGCTCAACCATCCCCGGT
TACTGGTTCACGGTTTACTTCATTGATAGAGTTGGTCGAGTCAAGATTCAGATAATGGGG
TTTTTTCTCATGGCCGTTGTTTACTTAGCCGCTGGCATCCCCTACAGTTGGTACTGGTCT
AAGCATGAGAAGACTAATAAAGGCTTCATGGTTCTCTACGGACTGATCTTCTTCTTTAGC
AACTTTGGTCCAAATACGACAACGTTTATAATCCCAGCTGAGCTTTTTCCGACAAGGTTT
AGGTCTACGTGTCATGGTATTTCCGGAGCTGCCGGGAAGCTTGGGGCGATTGTTGGGACT
GTTGGTTTCTTGTGGGCCACGAAACACGAGGACGAGGACAAAGAAGATATTTCCCAGAC
GTGAAACGGGTGAGGATCGCGTTTTTGATCCTTGGCGGCGTTTGTATCGCTGGAGTGTTG
GTGACGTACTTCTTCACGCGTGAAACTATGGGAAGATCGCTAGAAGAGAACGAAGAGGAC
GAGAATGGTACCACGGTAGCACCAGGAGGGTCATCATCTGCTAATGAGTTGCTTCCAAGA
CAATAGTGATTTTGATATCACAACATATATAATCAATCTTCTTTTGATATTTCCTCTCTA
ATATTCCTTGGTTTTTGATGTTCGATACTCACATATGCTTGTTTCCTCGAATGTTTTCTT
CTTTTTCACATTTTAATTTGGTGCATATTGAAGGGATATGTAAATGTCAAACCTTGATCA
TAATTTTCTTTCGTTTTTTTACCAACCCAGTTTAGAAACAAAAGAAAGTTAAAGTTGAGG
ATATGTGATATTAAAACATGATTATATGAATATCACAGCCAGTAGCTTTTCGTTTTTATA
TTAGTTTCATGGATGTGTTAGCCTCTGAAATTTTCCGCGAACCATGATTGTCGAATGCTT
TAGTCGGGATATAAGACCAGCTCCATCAACGAGTTCAGGAGAGGTTCTAAGGGAAAAAAG
ACAAAAAAAATAAATGAAAAATGTAAAAATACGAAAGAACCGGTGTTTAATTAAGAGAAA
GTAGAAGTCGTAAAAACCCCTAACGTGGCGGATCCGTATTGGCTTAAAAAACGGTGCTGA
AAGAACCGGTGTTTAATTAAGAGAAAGTAGAAGTGGTAAAAACCCCTAACGTGGCGGATC
CGTATT

Exons Bo6g120510 exons All exons in this region

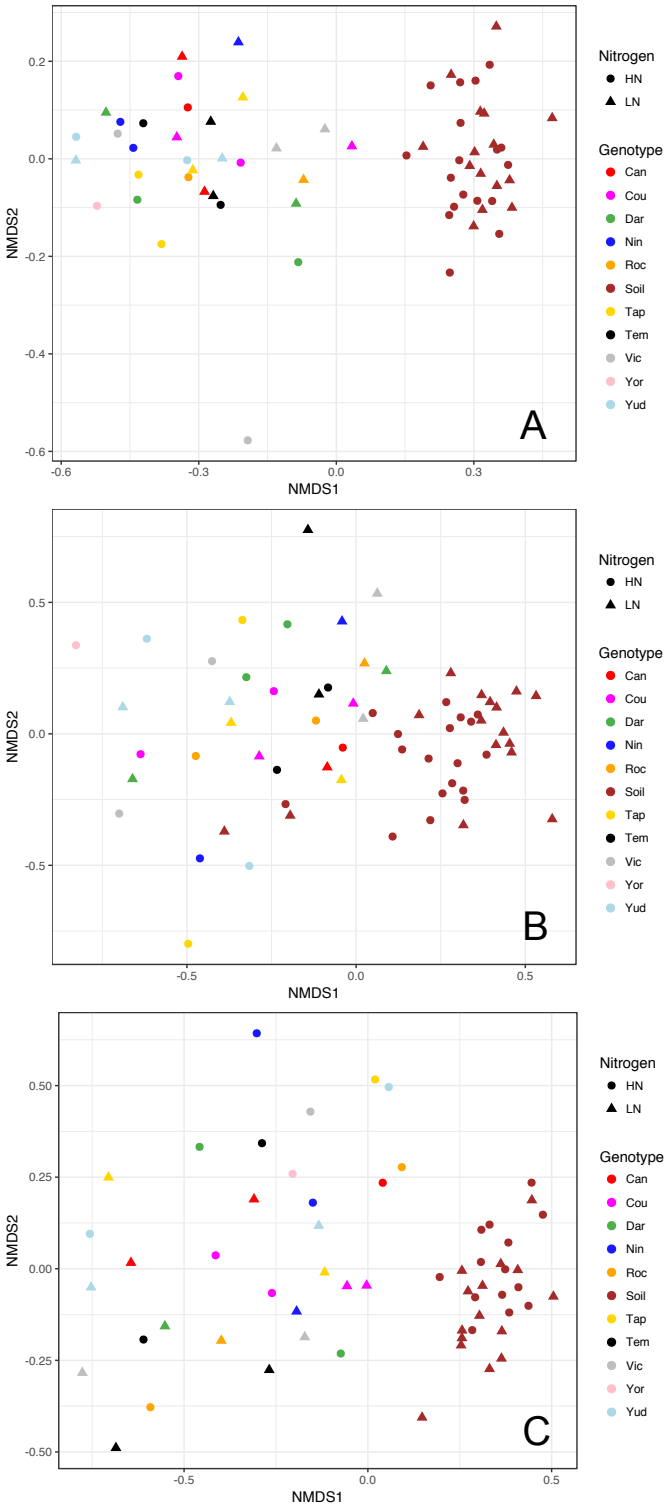
Markup loaded

>chromosome:v2.1:C6:38004377:38006553:-1

ACATTATGACGAAAAGAAACAGATAAACAAACTTTATTTTAAGCTATAATCCAAACTATA
TACGTCATGCTCTTCATTTTTCAAACCAATTTTTCATTTAGAAAATGTATAAACATATAG
GTTCTAGCTAATAATAAAGGTTATTGTAAAACATATTTTAAACTTGCTATCTATTAGGA
TTTGATATTTT**TG**TATAATTTATAAAAGCTTATAACTATAGTTTCTGATTTATT
AGAAAATATAATTTGTTGATTTATAATATTTTCATTACTACATATTAGTTCTTAGGAATG
AGTAGCATATTATACCCGCGTAAATATAATTTGTGTTGTGCCCATCGTGAGAACAGAAGT
TACAAAGGTAGTGGAAGTCAACTAAAGAAAGTGGACTCGATCCAGGGGGAGGTGTTGTGT
TCCTCACAGTAAAAACAAAAGCATATTCCATGGTGACTTCTTTGAACAAGAAACTCTTCA
GACATTAATAGCTAAGGGACAGAGACTTCCATTGTTCTCAGTTTGTGTATCCACGAACTC
CACAAAACCTAATAAGAAAAAACAAAAACAACTTTTATTAAAACAATAAAGTCATTAACG
ATGCCTGCGCTAAAAGTGTTATCGGCGTTAGATGTGGCGAGGATACAGTGGTACCATTTC
AAGGCGATAATCGTCGCCGGAATGGGTCTATTCACCGACGCTTACGATCTATTTGTATA
GCTCCGATCATGAAAATGATCGGCCATATCTATTACCACAAAGACTCCATCGGAACCGCT
GTTCTCTCCACTTCTTACGCCATCGCTCTCCTCGGCACCGCCTTAGGTCAGCTCATCTTC
GGCTACTTAGGCGACCGTGTCGGCCGCCGACGAGTCTACGGTCTTTGTCTCATGATCATG
GTCTTAAGCTCCTTTGGCTGTGGATTCTCCGTCTGCACGACTCGCCGTTCTTGCGTCATA
GCAAGTCTTGGCTTCTTTAGATTCGTCCTCGGGCTAGGGATTGGTGGAGACTACCCTCTC
TCCGCCACGATCATGTCGGAGTTCGCTAATAAACGGTCGCGTGGGACATTTATCGCCGCC
GTGTTCTCCATGCAGGGCTAGGAATCTTGATGAGCTCTGCCGTTACGATGGCTGTGTGT
GAGGCGTTCAAGAACGCCGGAAAAGGGAGTTTGGAGAAAATGAGAGCGGCGGGGGAGGAG
ACTTTGCTCCATGGGAATCGGATATTGCTTGGAGGTTGATTCTGATGATTGGTGCTCTT
CCCGCCGCGTTAACGTTCTACTGGCGAATGCTTATGCCTGAAACTGCCAGGTTCCGTTCA
ATAGTCTCTTCTATAGTTTTTTGTTTTTTTTCCTCAAGATTTTAATTATTTCCGGCGT
AAAAGCAAAAATACAAATATTCGCGTATTACAAAATAAATTAAATATTCTATTGCTTATA
CATCATCATCTATATACTTACTTATGTGTGTCTCTCTTTTTCTTTTCTGTTCCGTGATCC
TTTTAAAATCTAATTAATTTCCTGTCTTTTTCGCTAATTTTTTATCTATTATTATATCAC
CGTAATATAG**AATATGA**TAATAGCATATGCACTATCTTGTGTGACTTTACTTCCTACCTA
GCCCACATACATAAAGGTTTCTATACGTTATTAAAGTCTAACCAACCATTAGTATAACAT
AACCGCGAGAGAGATGAACCATATTTAAATAACTCTTTTTAACGGTAACAATTGTCACTG
GCTGTGAATTACTTCGGTTACAAACTAGATCGTTCGATTCTTAGTCGAGCAATATTCTTT
ATCATATAGTATGCAAATGGATATACTGCCTTGTCTTTGTTTGTGATTATCCCTTGGAT
AATTTCAGATTTTTAATGACACCAAAATGTATTTTTTTTTTTGAACAAACAAACAATGC
ATTAACTAGAAGGGTTCTATCCACCGTGAAAGGGGTTCTTACAAACTGAGCGAAAACTG
ATTTTGCCACAGAATCAGCCTCCAAATTAAAGTTGCTAGAAATGAAAGCAAAAGAAATAA
CATCAAAATAAGACATAAAGTGATAGATATCAAACATGATACCGAACATGACACCAAAAT
GTATAGAATTATAAATTAAATCGTAAAAAAGTTTAGTCGTGTGTTATGCAAAAGTTCTAC
CCACGTTTCCGTACCAT



Supplementary Figure 3, exon-intron structure of *A. thaliana PHT1* genes and putative *PHT1* genes in *B. oleracea from EMBL-EBI plant Ensembl (www.ensemblgenomes.org)*



Supplementary Figure 4 , NMDS displaying beta diversity of the microvial communities in Chapter 2 for Bacterial (A), Fungal (B) and Eukaryotic (C) communities displayed with associated genotype

Supplementary table 1. OTUs of greatest relative abundance in the Rhizosphere low N samples (Chapter 2).

Species	OTU ID	Rhiz HN Av.Abund	Rhiz LN Av.Abund	Contrib %
Methylibium	3617	0.99	1.07	0.52
Methylotenera mobilis	40	0.86	1.05	0.47
Xanthomonadaceae	11	0.86	0.91	0.61
DA101	4	0.76	0.86	0.57
Bradyrhizobiaceae	13	0.71	0.86	0.27
Bacillus	7	0.63	0.72	0.37

Supplementary table 2. BLAST results (Chapter 4)

Query	% identity	alignment length	mismatches	gap opens	evalue	bit score	Hit in B. oleracea genome	Bol genes overlapping hit	genomic length (bp)	Uniprot (A0A0D3+)	Protein length (aa)	Blast back to genome
AtPHT 1;1	84.261	1760	245	23	0	1687	C9:17421369-17419634	Bo9g059100	1566	E686	521	PHT1-1,2,3
AtPHT 1;1	84.082	1715	243	22	0	1628	C9:17492953-17494646	Bo9g059170	1470	E693	489	PHT1-1,2,3
AtPHT 1;1	83.481	1356	194	22	0	1236	C9:17430782-17429448	Bo9g059120	1002	E688	333	PHT1-1,2,3
AtPHT 1;1	85.265	1113	161	3	0	1144	C9:17425481-17424371	Bo9g059110	1017	E687	338	PHT1-1,2,3
AtPHT 1;1	82.605	1282	207	9	0	1118	C9:17198241-17199519	Bo9g058750	1173	E651	390	PHT1-1,2,3
AtPHT 1;1	83.875	1538	219	21	0	1439	C2:36985956-36984439	Bo2g119310	1467	ATS1	488	PHT1-1,2,3
AtPHT 1;1	86.121	1160	161	0	0	1251	C2:37017690-37018849	Bo2g119320	1566	ATS2	521	PHT1-1,2,3
AtPHT 1;1	85.517	1160	166	2	0	1210	C7:25008700-25009858	Bo7g065370	1566	D8MO	521	PHT1-1,2,3
AtPHT 1;1	84.508	1149	178	0	0	1136	C7:24991059-24989911	Bo7g065360	1608	D8L9	535	PHT1-1,2,3
AtPHT 1;2	83.815	1730	248	25	0	1615	C9:17492932-17494638	Bo9g059170	1470	E693	489	PHT1-1,2,3
AtPHT 1;2	83.986	1711	243	22	0	1613	C9:17421323-17419636	Bo9g059100	1566	E686	521	PHT1-1,2,3
AtPHT 1;2	83.563	1381	198	22	0	1266	C9:17430808-17429448	Bo9g059120	1002	E688	333	PHT1-1,2,3
AtPHT 1;2	84.887	1105	164	3	0	1112	C9:17425481-17424379	Bo9g059110	1017	E687	338	PHT1-1,2,3
AtPHT 1;2	83.725	1149	187	0	0	1086	C9:17198241-17199389	Bo9g058750	1173	E651	390	PHT1-1,2,3
AtPHT 1;2	83.854	1536	222	19	0	1439	C2:36985956-36984439	Bo2g119310	1467	ATS1	488	PHT1-1,2,3
AtPHT 1;2	86.034	1160	162	0	0	1245	C2:37017690-37018849	Bo2g119320	1566	ATS2	521	PHT1-1,2,3
AtPHT 1;2	84.828	1160	174	2	0	1166	C7:25008700-25009858	Bo7g065370	1566	D8MO	521	PHT1-1,2,3
AtPHT 1;2	84.238	1161	183	0	0	1131	C7:24991059-24989899	Bo7g065360	1608	D8L9	535	PHT1-1,2,3
AtPHT 1;3	83.635	1711	252	25	0	1583	C9:17421323-17419622	Bo9g059100	1566	E686	521	PHT1-1,2,3
AtPHT 1;3	85.617	1175	165	4	0	1230	C9:17493484-17494655	Bo9g059170	1470	E693	489	PHT1-1,2,3
AtPHT 1;3	85.34	1146	163	5	0	1181	C9:17425485-17424343	Bo9g059110	1017	E687	338	PHT1-1,2,3
AtPHT 1;3	84.143	1173	186	0	0	1136	C9:17198241-17199413	Bo9g058750	1173	E651	390	PHT1-1,2,3
AtPHT 1;3	86.672	1178	153	4	0	1303	C2:37017691-37018866	Bo2g119320	1566	ATS2	521	PHT1-1,2,3
AtPHT 1;3	86.409	1192	157	5	0	1299	C2:36985604-36984416	Bo2g119310	1467	ATS1	488	PHT1-1,2,3
AtPHT 1;3	85.995	1171	160	4	0	1251	C7:25008700-25009868	Bo7g065370	1566	D8MO	521	PHT1-1,2,3
AtPHT 1;3	84.402	1154	178	2	0	1133	C7:24991063-24989911	Bo7g065360	1608	D8L9	535	PHT1-1,2,3
AtPHT 1;4	87.38	1767	211	8	0	2017	C4:5346365-5344607	Bo4g026850	1605	BQQ6	534	PHT1-4
AtPHT 1;4	87.731	1728	199	12	0	2004	C4:50391565-50393284	Bo4g187910	1605	C525	534	PHT1-4
AtPHT 1;4	88.525	1647	186	3	0	1991	C4:5340373-5338728	Bo4g026840	1605	BQQ6	534	PHT1-4
AtPHT	85.2	1779	237	15	0	180	C4:5335393-	Bo4g026	1602	BQQ5	533	PHT1-4 and

1;4	16					5	5333639	830				PHT1-7
AtPHT 1;4	86.01	1401	171	17	0	1478	C4:49032399-49033775	Bo4g185840	417	C4G8	417	PHT1-4
AtPHT 1;4	80.401	1597	287	24	0	1192	C4:30760796-30762380	Bo4g119740	1623	BY14	540	PHT1-7
AtPHT 1;4	86.086	1689	223	8	0	1807	C3:12708028-12709709	Bo3g033000	1608	B5S4	353	PHT1-4
AtPHT 1;4	79.874	1585	302	16	0	1144	C8:29248309-29246732	Bo8g084860	1596	DT29	531	PHT1-7
AtPHT 1;5	87.266	801	97	4	0	909	C3:10083638-10084434	Bo3g027080	1641	B4N0	546	PHT1-5
AtPHT 1;6	82.908	1726	264	19	0	1524	C9:17400386-17398676	Bo9g059050	1536	E681	511	PHT1-6
AtPHT 1;6	80.947	1879	298	45	0	1432	C2:36952927-36951089	Bo2g119290	1554	ATR9	517	PHT1-6
AtPHT 1;7	86.361	1833	214	25	0	1967	C8:29248351-29246529	Bo8g084860	1596	DT29	531	PHT1-7
AtPHT 1;7	86.163	1814	216	28	0	1927	C4:30760777-30762586	Bo4g119740	1623	BY14	540	PHT1-7
AtPHT 1;7	81.218	1576	271	20	0	1247	C4:5340353-5338785	Bo4g026840	1605	BQQ6	534	PHT1-4
AtPHT 1;7	81.218	1576	271	20	0	1247	C4:5346345-5344777	Bo4g026850	1605	BQQ6	534	PHT1-4
AtPHT 1;7	80.919	1588	273	25	0	1227	C4:50391597-50393175	Bo4g187910	1605	C525	534	PHT1-4
AtPHT 1;7	80.421	1568	286	17	0	1175	C4:5335363-5333803	Bo4g026830	1602	BQQ5	533	PHT1-4 and PHT1-7
AtPHT 1;8	85.823	917	117	11	0	961	C5:10584370-10583458	Bo5g030750	1629	CBW6	542	PHT1-8
AtPHT 1;8	85.466	922	120	12	0	948	C5:10571466-10570549	Bo5g030730	864	CBW4	287	PHT1-8
AtPHT 1;8	84.237	996	113	16	0	929	C8:23076848-23075873	Bo8g070440	1521	DQN1	506	PHT1-8
AtPHT 1;9	85.197	966	120	15	0	970	C6:37998987-37998023	Bo6g120500	846	D155	281	PHT1-9
AtPHT 1;9	83.493	1042	129	25	0	931	C6:26437574-26438596	Bo6g084900	1656	CWc6	551	PHT1-9

References

ADL, S. M., HABURA, A. & EGLIT, Y. 2014. Amplification primers of SSU rDNA for soil protists. *Soil Biology and Biochemistry,* 69**,** 328-342.

AGREN, G. I., WETTERSTEDT, J. A. & BILLBERGER, M. F. 2012. Nutrient limitation on terrestrial plant growth--modeling the interaction between nitrogen and phosphorus. *New Phytol,* 194**,** 953-60.

AKIYAMA, K., MATSUOKA, H. & HAYASHI, H. 2001. Isolation and Identification of a Phosphate Deficiency-Induced C-Glycosylflavonoid That Stimulates Arbuscular Mycorrhiza Formation in Melon Roots. *Molecular Plant-Microbe Interactions,* 15**,** 334-340.

ALDEN, L., DEMOLING, F. & BAATH, E. 2001. Rapid method of determining factors limiting bacterial growth in soil. *Appl Environ Microbiol,* 67**,** 1830-8.

ALEXANDER, R. B., SMITH, R. A., SCHWARZ, G. E., BOYER, E. W., NOLAN, J. V. & BRAKEBILL, J. W. 2008. Differences in Phosphorus and Nitrogen Delivery to The Gulf of Mexico from the Mississippi River Basin. *Envionmental Science and Technology,* 42**,** 822-830.

ALTSCHU, S. F., GISH, W., MILLER, W., MYERS, E. W. & LIPMAN, D. J. 1990. Basic Local Alignment Search Tool. *Journal of Molecular Biology,* 215**,** 403-410.

AMARAL-ZETTLER, L. A., MCCLIMENT, E. A., DUCKLOW, H. W. & HUSE, S. M. 2009. A method for studying protistan diversity using massively parallel sequencing of V9 hypervariable regions of small-subunit ribosomal RNA genes. *PLoS One,* 4**,** e6372.

ANDREWS, M. & ANDREWS, M. E. 2017. Specificity in Legume-Rhizobia Symbioses. *Int J Mol Sci,* 18.

ANGUS, J. F., HERWAARDEN, A. F. V. & HOWE, G. N. 1991. Productivity and break crop effects of winter growing oilseeds. *Austrailian Journal of Experimental Agriculture,* 31**,** 669-677.

AUSTIN, R. S., HIU, S., WAESE, J., IERULLO, M., PASHA, A., WANG, T. T., FAN, J., FOONG, C., BREIT, R., DESVEAUX, D., MOSES, A. & PROVART, N. J. 2016. New BAR tools for mining expression data and exploring Cis-elements in Arabidopsis thaliana. *Plant J,* 88**,** 490-504.

AYADI, A., DAVID, P., ARRIGHI, J. F., CHIARENZA, S., THIBAUD, M. C., NUSSAUME, L. & MARIN, E. 2015. Reducing the genetic redundancy of Arabidopsis PHOSPHATE TRANSPORTER1 transporters to study phosphate uptake and signaling. *Plant Physiol,* 167**,** 1511-26.

BABIKOVA, Z., GILBERT, L., BRUCE, T. J., BIRKETT, M., CAULFIELD, J. C., WOODCOCK, C., PICKETT, J. A. & JOHNSON, D. 2013. Underground signals carried through common mycelial networks warn neighbouring plants of aphid attack. *Ecol Lett,* 16**,** 835-43.

BADRI, D. V. & VIVANCO, J. M. 2009. Regulation and function of root exudates. *Plant, Cell & Environment,* 32**,** 666-681.

BAKER, A., CEASAR, S. A., PALMER, A. J., PATERSON, J. B., QI, W., MUENCH, S. P. & BALDWIN, S. A. 2015. Replace, reuse, recycle: improving the sustainable use of phosphorus by plants. *J Exp Bot,* 66**,** 3523-40.

BASHIARDES, S., ZILBERMAN-SCHAPIRA, G. & ELINAV, E. 2016. Use of Metatranscriptomics in Microbiome Research. *Bioinform Biol Insights,* 10**,** 19-25.

BASTIDA, F., JEHMLICH, N., LIMA, K., MORRIS, B. E. L., RICHNOW, H. H., HERNANDEZ, T., VON BERGEN, M. & GARCIA, C. 2016. The ecological and physiological responses of the microbial community from a semiarid soil to hydrocarbon contamination and its bioremediation using compost amendment. *J Proteomics,* 135**,** 162-169.

BEKKU, Y., KIMURA, M., IKEDA, H. & KOIZUMI, H. 1997. Carbon input from plant to soil through root exudation in Digitaria adscendens and Ambrosia artemisiifolia. *Ecological Research,* 12**,** 305-312.

BELL, C. W., ASAO, S., CALDERON, F., WOLK, B. & WALLENSTEIN, M. D. 2015. Plant nitrogen uptake drives rhizosphere bacterial community assembly during plant growth. *Soil Biology and Biochemistry,* 85**,** 170-182.

BERENDSEN, R. L., PIETERSE, C. M. & BAKKER, P. A. 2012. The rhizosphere microbiome and plant health. *Trends Plant Sci,* 17**,** 478-86.

BERG, G. & SMALLA, K. 2009. Plant species and soil type cooperatively shape the structure and function of microbial communities in the rhizosphere. *FEMS Microbiol Ecol,* 68**,** 1-13.

BOLGER, A., LOHSE, M. & USADEL, B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics,* 30**,** 2114-2120.

BONFANTE, P. & GENRE, A. 2010. Mechanisms underlying beneficial plant-fungus interactions in mycorrhizal symbiosis. *Nat Commun,* 1**,** 48.

BONKOWSKI, M. & CLARHOLM, M. 2012. Stimulation of Plant Growth through Interactions of Bacteria and Protozoa: Testing the Auxiliary Microbial Loop Hypothesis. *Acta Protozoologica,* 51**,** 237-247.

BOYS, E. F., ROQUES, S. E., ASHBY, A. M., EVANS, N., LATUNDE-DADA, A. O., THOMAS, J. E., WEST, J. S. & FITT, B. D. L. 2007. Resistance to infection by stealth: Brassica napus (winter oilseed rape) and Pyrenopeziza brassicae (light leaf spot). *European Journal of Plant Pathology,* 118**,** 307-321.

BROWN, L. R. 1970. Nobel Peace Prize: Developer of High-Yield Wheat Receives Award (Norman Ernest Borlaug). *Science* 170**,** 518-519.

BUCHFINK, B., XIE, C. & HUSON, D. H. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods,* 12**,** 59-60.

BULGARELLI, D., GARRIDO-OTER, R., MUNCH, P. C., WEIMAN, A., DROGE, J., PAN, Y., MCHARDY, A. C. & SCHULZE-LEFERT, P. 2015. Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe,* 17**,** 392-403.

BULGARELLI, D., ROTT, M., SCHLAEPPI, K., VER LOREN VAN THEMAAT, E., AHMADINEJAD, N., ASSENZA, F., RAUF, P., HUETTEL, B., REINHARDT, R., SCHMELZER, E., PEPLIES, J.,

GLOECKNER, F. O., AMANN, R., EICKHORST, T. & SCHULZE-LEFERT, P. 2012. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. *Nature,* 488**,** 91-5.
BULGARELLI, D., SCHLAEPPI, K., SPAEPEN, S., VER LOREN VAN THEMAAT, E. & SCHULZE-LEFERT, P. 2013. Structure and functions of the bacterial microbiota of plants. *Annu Rev Plant Biol,* 64**,** 807-38.
BUN-YA, M., NISHIMURA, M., HARASHIMA, S. & OSHIMA, Y. 1991. The PH084 Gene of Saccharomyces cerevisiae Encodes an Inorganic Phosphate Transporter. *Molecular and Cellular Biology,* 11**,** 3229-3283.
BUSBY, P. E., SOMAN, C., WAGNER, M. R., FRIESEN, M. L., KREMER, J., BENNETT, A., MORSY, M., EISEN, J. A., LEACH, J. E. & DANGL, J. L. 2017. Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLoS Biol,* 15**,** e2001793.
CAMERON, D., OSBORNE, C., HORTON, P. & SINCLAIR, M. 2015. A sustainable model for intensive agriculture. *In:* SHEFFIELD, T. U. O. (ed.) *Grantham Centre for Sustainable Futures.*
CAPORASO, J. G., LAUBER, C. L., WALTERS, W. A., BERG-LYONS, D., HUNTLEY, J., FIERER, N., OWENS, S. M., BETLEY, J., FRASER, L., BAUER, M., GORMLEY, N., GILBERT, J. A., SMITH, G. & KNIGHT, R. 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *ISME J,* 6**,** 1621-4.
CARBONERO, F., OAKLEY, B. B. & PURDY, K. J. 2014. Metabolic flexibility as a major predictor of spatial distribution in microbial communities. *PLoS One,* 9**,** e85105.
CARINI, P., MARSDEN, P. J., LEFF, J. W., MORGAN, E. E., STRICKLAND, M. S. & FIERER, N. 2016. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. *Nat Microbiol,* 2**,** 16242.
CARVALHAIS, L. C., DENNIS, P. G., FEDOSEYENKO, D., HAJIREZAEI, M.-R., BORRISS, R. & VON WIRÉN, N. 2011. Root exudation of sugars, amino acids, and organic acids by maize as affected by nitrogen, phosphorus, potassium, and iron deficiency. *Journal of Plant Nutrition and Soil Science,* 174**,** 3-11.
CHAPARRO, J. M., BADRI, D. V., BAKKER, M. G., SUGIYAMA, A., MANTER, D. K. & VIVANCO, J. M. 2013. Root exudation of phytochemicals in Arabidopsis follows specific patterns that are developmentally programmed and correlate with soil microbial functions. *PLoS One,* 8**,** e55731.
CHAPARRO, J. M., BADRI, D. V. & VIVANCO, J. M. 2014. Rhizosphere microbiome assemblage is affected by plant development. *ISME J,* 8**,** 790-803.
CHEN, I. M., MARKOWITZ, V. M., CHU, K., ANDERSON, I., MAVROMATIS, K., KYRPIDES, N. C. & IVANOVA, N. N. 2013. Improving microbial genome annotations in an integrated database context. *PLoS One,* 8**,** e54859.
CLARKE, K. & GORLEY, R. 2006. PRIMER v6: User Manual/Tutorial. PRIMER-E. Plymouth, UK.

CORDELL, D., DRANGERT, J.-O. & WHITE, S. 2009. The story of phosphorus: Global food security and food for thought. *Global Environmental Change,* 19**,** 292-305.
CRICK, F. 1970. Central Dogma of Molecular Biology. *Nature,* 227**,** 561-563.
CROWTHER, T. W., MAYNARD, D. S., LEFF, J. W., OLDFIELD, E. E., MCCULLEY, R. L., FIERER, N. & BRADFORD, M. A. 2014. Predicting the responsiveness of soil biodiversity to deforestation: a cross-biome study. *Glob Chang Biol,* 20**,** 2983-94.
CURCI, P. L., AIESE CIGLIANO, R., ZULUAGA, D. L., JANNI, M., SANSEVERINO, W. & SONNANTE, G. 2017. Transcriptomic response of durum wheat to nitrogen starvation. *Sci Rep,* 7**,** 1176.
CZABAN, J., GAJDA, A. & WRÓBLEWSKA, B. 2007. The Motility of Bacteria from Rhizosphere and Different Zones of Winter Wheat Roots. *Polish journal of Environmental studies,* 16**,** 301-308.
DE MENDIBURU, F. 2017. Statistical Procedures for Agricultural Research.
DENG, Y., JIANG, Y.-H., YANG, Y., HE, Z., LUO, F. & ZHOU, J. 2012. Molecular ecological network analysis. *BMC Bioinformatics,* 13.
DESANTIS, T. Z., HUGENHOLTZ, P., LARSEN, N., ROJAS, M., BRODIE, E. L., KELLER, K., HUBER, T., DALEVI, D., HU, P. & ANDERSEN, G. L. 2006. Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol,* 72**,** 5069-72.
DEVOS, D. & VALENCIA, A. 2001. Intrinsic errors in genome annotation. *Trends in Genetics,* 17**,** 429-431.
DOEBLEY, J. F., GAUT, B. S. & SMITH, B. D. 2006. The Molecular Genetics of Crop Domestication. *Cell,* 127**,** 1309-1321.
DRENCKHAHN, D. 2017. Occurrence of Atlantic wild cabbage (Brassica oleracea L. subsp. oleracea) at the limestone cliffs of the baltic island of Rügen, Germany. *Forum geobotanicum,* 7**,** 18-26.
EDGAR, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res,* 32**,** 1792-7.
EDWARDS, J., JOHNSON, C., SANTOS-MEDELLIN, C., LURIE, E., PODISHETTY, N. K., BHATNAGAR, S., EISEN, J. A. & SUNDARESAN, V. 2015. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A,* 112**,** E911-20.
EUROPEAN_COMMISION 2017. OILSEEDS and PROTEIN CROPS market situation. *In:* MARKETS, C. F. T. C. O. O. A. (ed.).
FAN, K., CARDONA, C., LI, Y., SHI, Y., XIANG, X., SHEN, C., WANG, H., GILBERT, J. A. & CHU, H. 2017. Rhizosphere-associated bacterial network structure and spatial distribution differ significantly from bulk soil in wheat crop fields. *Soil Biology and Biochemistry,* 113**,** 275-284.
FAO. 2011. *Calories per person* [Online]. Available: http://www.nationalgeographic.com/what-the-world-eats/ [Accessed].
FAO 2015. World fertilizer trends and outlook to 2018. Rome: FOOD AND AGRICULTURE ORGANIZATION OF THE UNITED NATIONS.

FARRAR, K., BRYANT, D. & COPE-SELBY, N. 2014. Understanding and engineering beneficial plant-microbe interactions: plant growth promotion in energy crops. *Plant Biotechnol J,* 12**,** 1193-206.
FAUST, K., SATHIRAPONGSASUTI, J. F., IZARD, J., SEGATA, N., GEVERS, D., RAES, J. & HUTTENHOWER, C. 2012. Microbial co-occurrence relationships in the human microbiome. *PLoS Comput Biol,* 8**,** e1002606.
FIERER, N. 2017. Embracing the unknown: disentangling the complexities of the soil microbiome. *Nat Rev Microbiol*.
FIERER, N., LAUBER, C. L., RAMIREZ, K. S., ZANEVELD, J., BRADFORD, M. A. & KNIGHT, R. 2012. Comparative metagenomic, phylogenetic and physiological analyses of soil microbial communities across nitrogen gradients. *ISME J,* 6**,** 1007-17.
FLAGEL, L. E. & WENDEL, J. F. 2009. Gene duplication and evolutionary novelty in plants. *New Phytol,* 183**,** 557-64.
FOX, G. E., WISOTZKEY, J. D. & JURTSHUK, P. 1992. How Close Is Close: 16s rRNA Sequence Identity May Not Be Sufficient To Guarantee Species Identity. *International Journal of Systematic Bacteriology,* 42**,** 166-170.
FRIEDMAN, J. & ALM, E. J. 2012. Inferring Correlation Networks from Genomic Survey Data. *PLoS Comput Biol,* 8.
FU, Y.-B. 2015. Understanding crop genetic diversity under modern plant breeding. *Theoretical and Applied Genetics,* 128**,** 1231-2142.
GEISEN, S., TVEIT, A. T., CLARK, I. M., RICHTER, A., SVENNING, M. M., BONKOWSKI, M. & URICH, T. 2015. Metatranscriptomic census of active protists in soils. *ISME J,* 9**,** 2178-90.
GEORGE, T. S., RICHARDSON, A. E., HADOBAS, P. A. & SIMPSON, R. J. 2004. Characterization of transgenic *Trifolium subterraneum* L. which expresses phyA and releases extracelluar phytase: growth and P nutrition in labratory media andsoil. *Plant, Cell and Environment,* 27**,** 1351-1361.
GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-51.
GRIFFITHS, R. I., THOMSON, B. C., JAMES, P., BELL, T., BAILEY, M. & WHITELEY, A. S. 2011. The bacterial biogeography of British soils. *Environ Microbiol,* 13**,** 1642-54.
GUO, B., JIN, Y., WUSSLER, C., BLANCAFLOR, E. B., MOTES, C. M. & VERSAW, W. K. 2008. Functional analysis of the Arabidopsis PHT4 family of intracellular phosphate transporters. *New Phytol,* 177**,** 889-98.
HAAS, D. & KEEL, C. 2003. Regulation of antibiotic production in root-colonizing Peudomonas spp. and relevance for biological control of plant disease. *Annu Rev Phytopathol,* 41**,** 117-53.
HABER, F. The Synthesis of Ammonia From It's Elements. Bobel Lecture in Chemistry, 1920.
HARDOIM, P. R., VAN OVERBEEK, L. S., BERG, G., PIRTTILA, A. M., COMPANT, S., CAMPISANO, A., DORING, M. & SESSITSCH, A. 2015. The Hidden World within Plants: Ecological and Evolutionary

Considerations for Defining Functioning of Microbial Endophytes. *Microbiol Mol Biol Rev,* 79**,** 293-320.
HARTWRIGHT, L. M., HUNTER, P. J. & WALSH, J. A. 2010. A comparison of Olpidium isolates from a range of host plants using internal transcribed spacer sequence analysis and host range studies. *Fungal Biol,* 114**,** 26-33.
HASSLER, S., JUNG, B., LEMKE, L., NOVAK, O., STRNAD, M., MARTINOIA, E. & NEUHAUS, H. E. 2016. Function of the Golgi-located phosphate transporter PHT4;6 is critical for senescence-associated processes in Arabidopsis. *J Exp Bot,* 67**,** 4671-84.
HERBERT, L. T., VAZQUEZ, D. E., ARENAS, A. & FARINA, W. M. 2014. Effects of field-realistic doses of glyphosate on honeybee appetitive behaviour. *J Exp Biol,* 217**,** 3457-64.
HIBBING, M. E., FUQUA, C., PARSEK, M. R. & PETERSON, S. B. 2010. Bacterial competition: surviving and thriving in the microbial jungle. *Nat Rev Microbiol,* 8**,** 15-25.
HILBER-BODMER, M., SCHMID, M., AHRENS, C. H. & FREIMOSER, F. M. 2017. Competition assays and physiological experiments of soil and phyllosphere yeasts identify Candida subhashii as a novel antagonist of filamentous fungi. *BMC Microbiol,* 17**,** 4.
HILTNER, L. 1904. Uber neue erfahrungen und probleme auf dem gebiete der bodenbakteriologie. *Arbeiten der Deutschen Landwirtschaft Gesellschaft,* 98**,** 59-78.
HILTON, S., BENNETT, A. J., KEANE, G., BENDING, G. D., CHANDLER, D., STOBART, R. & MILLS, P. 2013. Impact of shortened crop rotation of oilseed rape on soil and rhizosphere microbial diversity in relation to yield decline. *PLoS One,* 8**,** e59859.
HIRSCH, P. R. & MAUCHLINE, T. H. 2015. The Importance of the Microbial N Cycle in Soil for Crop Plant Nutrition. *Adv Appl Microbiol,* 93**,** 45-71.
HODGKIN, T. 1995. Cabbages, kales, etc. *In:* SMARTT, J. & SIMMONDS, N. (eds.) *Evolution of Crop Plants, 2nd Edition*
. Wiley-Blackwell.
HUANG, X.-F., CHAPARRO, J. M., REARDON, K. F., ZHANG, R., SHEN, Q. & VIVANCO, J. M. 2014. Rhizosphere interactions: root exudates, microbes, and microbial communities. *Botany,* 92**,** 267-275.
HUBER, M., BONT, Z., FRICKE, J., BRILLATZ, T., AZIZ, Z., GERSHENZON, J. & ERB, M. 2016. A below-ground herbivore shapes root defensive chemistry in natural plant populations. *Proc Biol Sci,* 283**,** 20160285.
HUGENHOLTZ, P. 2002. Exploring prokaryotic diversity in the genomic era. *Genome Biology,* 3**,** 1-8.
HUNTER, M. C., SMITH, R. G., SCHIPANSKI, M. E., ATWOOD, L. W. & MORTENSEN, D. A. 2017. Agriculture in 2050: Recalibrating Targets for Sustainable Intensification. *BioScience,* 67**,** 386-391.
HUSON, D. H., BEIER, S., FLADE, I., GORSKA, A., EL-HADIDI, M., MITRA, S., RUSCHEWEYH, H. J. & TAPPU, R. 2016. MEGAN Community Edition - Interactive Exploration and Analysis of Large-

Scale Microbiome Sequencing Data. *PLoS Comput Biol,* 12**,** e1004957.
HWANG, S. F., STRELKOV, S. E., FENG, J., GOSSEN, B. D. & HOWARD, R. J. 2012. Plasmodiophora brassicae: a review of an emerging pathogen of the Canadian canola (Brassica napus) crop. *Mol Plant Pathol,* 13**,** 105-13.
ILLUMINA 2010. CASAVA Version 1.7. *In:* ILLUMINA (ed.).
IRVINE, I. C., VIVANCO, L., BENTLEY, P. N. & MARTINY, J. B. 2012. The effect of nitrogen enrichment on c(1)-cycling microorganisms and methane flux in salt marsh sediments. *Front Microbiol,* 3**,** 90.
JANIAK, A., KWASNIEWSKI, M. & SZAREJKO, I. 2016. Gene expression regulation in roots under drought. *J Exp Bot,* 67**,** 1003-14.
JOHNSEN, M. G., HANSEN, O. C. & STOUGAARD, P. 2010. Isolation, characterization and heterologous expression of a novel chitosanase from Janthinobacterium sp. strain 4239. *Microbial Cell Factories,* 9.
JONES, D. L., HODGE, A. & KUZYAKOV, Y. 2004. Plant and mycorrhizal regulation of rhizodeposition. *New Phytologist,* 163**,** 459-480.
JONES, D. L., NGUYEN, C. & FINLAY, R. D. 2009. Carbon flow in the rhizosphere: carbon trading at the soil–root interface. *Plant and Soil,* 321**,** 5-33.
JORQUERA, M. A., HERNÁNDEZ, M. T., RENGEL, Z., MARSCHNER, P. & DE LA LUZ MORA, M. 2008. Isolation of culturable phosphobacteria with both phytate-mineralization and phosphate-solubilization activity from the rhizosphere of plants grown in a volcanic soil. *Biology and Fertility of Soils,* 44**,** 1025-1034.
JOUSSET, A., ROCHAT, L., PECHY-TARR, M., KEEL, C., SCHEU, S. & BONKOWSKI, M. 2009. Predators promote defence of rhizosphere bacterial populations by selective feeding on non-toxic cheaters. *ISME J,* 3**,** 666-74.
JUNGK, A. 2001. Root hairs and the acquisition of plant nutrients from soil. *Journal of Plant Nutrition and Soil Science,* 164**,** 121-129.
KALYUZHNAYA, M. G., BOWERMAN, S., LARA, J. C., LIDSTROM, M. E. & CHISTOSERDOVA, L. 2006. Methylotenera mobilis gen. nov., sp. nov., an obligately methylamine-utilizing bacterium within the family Methylophilaceae. *Int J Syst Evol Microbiol,* 56**,** 2819-23.
KAMBOJ, D., KUMAR, R., KUMARI, A., KUNDU, B. S., PATHAK, D. & SHARMA, P. K. 2008. Rhizobia, NOD factors and nodulation-A review. *Agricultural Reviews,* 3**,** 200-206.
KANG, S. M., KHAN, A. L., HAMAYUN, M., HUSSAIN, J., JOO, G. J., YOU, Y. H., KIM, J. G. & LEE, I. J. 2012. Gibberellin-producing Promicromonospora sp. SE188 improves Solanum lycopersicum plant growth and influences endogenous plant hormones. *J Microbiol,* 50**,** 902-9.
KASALICKY, V., JEZBERA, J., HAHN, M. W. & SIMEK, K. 2013. The diversity of the Limnohabitans genus, an important group of freshwater bacterioplankton, by characterization of 35 isolated strains. *PLoS One,* 8**,** e58209.

KENT, A. D. & TRIPLETT, E. W. 2002. Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annu Rev Microbiol,* 56**,** 211-36.
KERSEY, P. J., ALLEN, J. E., ARMEAN, I., BODDU, S., BOLT, B. J., CARVALHO-SILVA, D., CHRISTENSEN, M., DAVIS, P., FALIN, L. J., GRABMUELLER, C., HUMPHREY, J., KERHORNOU, A., KHOBOVA, J., ARANGANATHAN, N. K., LANGRIDGE, N., LOWY, E., MCDOWALL, M. D., MAHESWARI, U., NUHN, M., ONG, C. K., OVERDUIN, B., PAULINI, M., PEDRO, H., PERRY, E., SPUDICH, G., TAPANARI, E., WALTS, B., WILLIAMS, G., TELLO-RUIZ, M., STEIN, J., WEI, S., WARE, D., BOLSER, D. M., HOWE, K. L., KULESHA, E., LAWSON, D., MASLEN, G. & STAINES, D. M. 2016. Ensembl Genomes 2016: more genomes, more complexity. *Nucleic Acids Res,* 44**,** D574-80.
KIM, D., PERTEA, G., TRAPNELL, C., PIMENTAL, H., KELLEY, R. & SALZBERG, S. L. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology,* 14.
KINSELLA, K., SCHULTHESS, C. P., MORRIS, T. F. & STUART, J. D. 2009. Rapid quantification of Bacillus subtilis antibiotics in the rhizosphere. *Soil Biology and Biochemistry,* 41**,** 374-379.
KOLTON, M., ERLACHER, A., BERG, G. & CYTRYN, E. 2016. The Flavobacterium Genus in the Plant 9 Holobiont: Ecological, Physiological,
and Applicative Insights. *In:* CASTRO-SOWINSKI, S. (ed.) *Microbial Models: From Environments to Industrial Sustainability.* Singapore: Springer Science.
KOPYLOVA, E., NOE, L. & TOUZET, H. 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics,* 28**,** 3211-7.
KROME, K., ROSENBERG, K., BONKOWSKI, M. & SCHEU, S. 2009. Grazing of protozoa on rhizosphere bacteria alters growth and reproduction of Arabidopsis thaliana. *Soil Biology and Biochemistry,* 41**,** 1866-1873.
LANGILLE, M. G., ZANEVELD, J., CAPORASO, J. G., MCDONALD, D., KNIGHTS, D., REYES, J. A., CLEMENTE, J. C., BURKEPILE, D. E., VEGA THURBER, R. L., KNIGHT, R., BEIKO, R. G. & HUTTENHOWER, C. 2013. Predictive functional profiling of microbial communities using 16S rRNA marker gene sequences. *Nat Biotechnol,* 31**,** 814-21.
LANGMEAD, B. & SALZBERG, S. L. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods,* 9**,** 357-9.
LAPIS-GAZA, H., JOST, R. & FINNEGAN, P. 2014. Arabidopsis Phosphate Transporter1 genes PHT1;8 and PHT1;9 are involved in root-to-shoot translocation of orthophosphate. *BMC Plant Biology,* 14.
LAUBER, C. L., HAMADY, M., KNIGHT, R. & FIERER, N. 2009. Pyrosequencing-based assessment of soil pH as a predictor of soil bacterial community structure at the continental scale. *Appl Environ Microbiol,* 75**,** 5111-20.

LEE, Z. M., BUSSEMA, C., 3RD & SCHMIDT, T. M. 2009. rrnDB: documenting the number of rRNA and tRNA genes in bacteria and archaea. *Nucleic Acids Res,* 37**,** D489-93.
LOCEY, K. J. & LENNON, J. T. 2016. Scaling laws predict global microbial diversity. *PNAS,* 113**,** 5970-5975.
LUNDBERG, D. S., LEBEIS, S. L., PAREDES, S. H., YOURSTONE, S., GEHRING, J., MALFATTI, S., TREMBLAY, J., ENGELBREKTSON, A., KUNIN, V., DEL RIO, T. G., EDGAR, R. C., EICKHORST, T., LEY, R. E., HUGENHOLTZ, P., TRINGE, S. G. & DANGL, J. L. 2012. Defining the core Arabidopsis thaliana root microbiome. *Nature,* 488**,** 86-90.
LYSAK, M. A., KOCH, M. A., PECINKA, A. & I, S. 2005. Chromosome triplication found across the tribe Brassiceae. *Genome Research,* 15.
MA, B., WANG, H., DSOUZA, M., LOU, J., HE, Y., DAI, Z., BROOKES, P. C., XU, J. & GILBERT, J. A. 2016. Geographic patterns of co-occurrence network topological features for soil microbiota at continental scale in eastern China. *ISME J,* 10**,** 1891-901.
MAGGIONI, L., VON BOTHMER, R., POULSEN, G. & LIPMAN, E. 2017. Domestication, diversity and use of Brassica oleracea L., based on ancient Greek and Latin texts. *Genetic Resources and Crop Evolution*.
MARSCHNER, P., CROWLEY, D. E. & YANG, C. H. 2004. Development of specific rhizosphere bacterial communities in relation to plant species, nutrition and soil type. *Plant and Soil,* 261**,** 199-208.
MCMURDIE, P. J. & HOLMES, S. 2013. phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. *PLoS One,* 8**,** e61217.
MEI, J., LI, Q., YANG, X., QIAN, L., LIU, L., YIN, J., FRAUEN, M., LI, J. & QIAN, W. 2010. Genomic relationships between wild and cultivated Brassica oleracea L. with emphasis on the origination of cultivated crops. *Genetic Resources and Crop Evolution,* 57**,** 687-692.
MENDES, L. W., KURAMAE, E. E., NAVARRETE, A. A., VAN VEEN, J. A. & TSAI, S. M. 2014. Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J,* 8**,** 1577-87.
MENDES, R., GARBEVA, P. & RAAIJMAKERS, J. M. 2013. The rhizosphere microbiome: significance of plant beneficial, plant pathogenic, and human pathogenic microorganisms. *FEMS Microbiol Review,* 37**,** 634-663.
MENDES, R., KRUIJT, M., DE BRUIJN, I., DEKKERS, E., VAN DER VOORT, M., SCNEIDER, J. H. M., PICENO, Y. M., DESANTIS, T., Z, ANDERSEN, G. L., BAKKER, P. A. & RAAIJMAKERS, J. M. 2011. Deciphering the Rhizosphere Microbiome for Disease-Suppressive Bacteria. *Science,* 332**,** 1097-1100.
MENGEL, K. & KIRKBY, E. A. 1987. *Principles of plant nutrition*.
METTEL, C., KIM, Y., SHRESTHA, P. M. & LIESACK, W. 2010. Extraction of mRNA from soil. *Appl Environ Microbiol,* 76**,** 5995-6000.
MICALLEF, S. A., CHANNER, S., SHIARIS, M. P. & COLÓN-CARMANA, A. 2009. Plant age and genotype impact the progression of

bacterial community succession in the Arabidopsis rhizosphere. *Plant signalling & behaviour,* 4**,** 777-780.
MOHAMMADIPANAH, F., HAMEDI, J., SPROER, C., MONTERO-CALASANZ MDEL, C., SCHUMANN, P. & KLENK, H. P. 2014. Promicromonospora iranensis sp. nov., an actinobacterium isolated from rhizospheric soil. *Int J Syst Evol Microbiol,* 64**,** 3314-9.
MUCHHAL, U. S., PARDO, J. M. & RAGHOTHAMA, K. G. 1996. Phosphate transporters from the higher plant Arabidopsis thaliana. *Proc. Natl. Acad. Sci. U S A,* 93**,** 10519-10523.
MUDGE, S. R., RAE, A. L., DIATLOFF, E. & SMITH, F. W. 2002. Expression analysis suggests novel roles for members of the Pht1 family of phosphate transporters in Arabidopsis. *The Plant Journal,* 31**,** 341-353.
MULLER, H., BERG, C., LANDA, B. B., AUERBACH, A., MOISSL-EICHINGER, C. & BERG, G. 2015. Plant genotype-specific archaeal and bacterial endophytes but similar Bacillus antagonists colonize Mediterranean olive trees. *Front Microbiol,* 6**,** 138.
MULLER, H., WESTENDORF, C., LEITNER, E., CHERNIN, L., RIEDEL, K., SCHMIDT, S., EBERL, L. & BERG, G. 2009. Quorum-sensing effects in the antagonistic rhizosphere bacterium Serratia plymuthica HRO-C48. *FEMS Microbiol Ecol,* 67**,** 468-78.
NAGAHARU, U. 1935. Genome analysis in Brassica with special reference to the experimental formation of B. napus and peculiar mode of fertilization. *Japanese Journal of Botany,* 7**,** 389–452.
NAGARAJAN, V. K., JAIN, A., POLING, M. D., LEWIS, A. J., RAGHOTHAMA, K. G. & SMITH, A. P. 2011. Arabidopsis Pht1;5 mobilizes phosphate between source and sink organs and influences the interaction between phosphate homeostasis and ethylene signaling. *Plant Physiol,* 156**,** 1149-63.
NEAL, A. L., AHMAD, S., GORDON-WEEKS, R. & TON, J. 2012. Benzoxazinoids in root exudates of maize attract Pseudomonas putida to the rhizosphere. *PLoS One,* 7**,** e35498.
NEWMAN, M. M., LORENZ, N., HOILETT, N., LEE, N. R., DICK, R. P., LILES, M. R., RAMSIER, C. & KLOEPPER, J. W. 2016. Changes in rhizosphere bacterial gene expression following glyphosate treatment. *Sci Total Environ,* 553**,** 32-41.
NGUYEN, N. H., SONG, Z., BATES, S. T., BRANCO, S., TEDERSOO, L., MENKE, J., SCHILLING, J. S. & KENNEDY, P. G. 2016. FUNGuild: An open annotation tool for parsing fungal community datasets by ecological guild. *Fungal Ecology,* 20**,** 241-248.
NUSSAUME, L., KANNO, S., JAVOT, H., MARIN, E., POCHON, N., AYADI, A., NAKANISHI, T. M. & THIBAUD, M. C. 2011. Phosphate Import in Plants: Focus on the PHT1 Transporters. *Front Plant Sci,* 2**,** 83.
OERKE, E. C. 2005. Crop losses to pests. *The Journal of Agricultural Science,* 144**,** 31.
OFAIM, S., OFEK-LALZAR, M., SELA, N., JINAG, J., KASHI, Y., MINZ, D. & FREILICH, S. 2017. Analysis of Microbial Functions in the Rhizosphere Using a Metabolic-Network Based Framework for Metagenomics Interpretation. *Front Microbiol,* 8**,** 1606.

OFEK-LALZAR, M., SELA, N., GOLDMAN-VORONOV, M., GREEN, S. J., HADAR, Y. & MINZ, D. 2014. Niche and host-associated functional signatures of the root surface microbiome. *Nat Commun,* 5**,** 4950.
OGLE, D. H. 2017. FSA: Fisheries Stock Analysis. *R package version 0.8.13.*
OGURA, A. 2014. *Genome Analysis: Current Procedures and Applications*, Horizon Scientific Press.
OIKONOMOPOULOS, S., WANG, Y. C., DJAMBAZIAN, H., BADESCU, D. & RAGOUSSIS, J. 2016. Benchmarking of the Oxford Nanopore MinION sequencing for quantitative and qualitative assessment of cDNA populations. *Sci Rep,* 6**,** 31602.
OLIVERIO, A. M., BRADFORD, M. A. & FIERER, N. 2017. Identifying the microbial taxa that consistently respond to soil warming across time and space. *Glob Chang Biol,* 23**,** 2117-2129.
PANDIT, S. N., KOLASA, K. & COTTENIE, K. 2009. Contrasts between habitat generalists and specialists:
an empirical extension to the basic metacommunity framework. *Ecology,* 90**,** 2253-2262.
PAO, S. S., PAULSEN, I. T. & SAIER , M. H. J. 1998. Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews,* 62**,** 1-34.
PARKIN, I. A., KOH, C., TANG, H., ROBINSON, S. J., KAGALE, S., CLARKE, W. E., TOWN, C. D., NIXON, J., KRISHNAKUMAR, V., BIDWELL, S. L., DENOEUD, F., BELCRAM, H., LINKS, M. G., JUST, J., CLARKE, C., BENDER, T., HUEBERT, T., MASON, A. S., PIRES, J. C., BARKER, G., MOORE, J., WALLEY, P. G., MANOLI, S., BATLEY, J., EDWARDS, D., NELSON, M. N., WANG, X., PATERSON, A. H., KING, G. J., BANCROFT, I., CHALHOUB, B. & SHARPE, A. G. 2014. Transcriptome and methylome profiling reveals relics of genome dominance in the mesopolyploid Brassica oleracea. *Genome Biology,* 15.
PARKS, D. H., TYSON, G. W., HUGENHOLTZ, P. & BEIKO, R. G. 2014. STAMP: statistical analysis of taxonomic and functional profiles. *Bioinformatics,* 30**,** 3123-4.
PARTIDA-MARTINEZ, L. P. & HERTWECK, C. 2005. Pathogenic fungus harbours endosymbiotic bacteria for toxin production. *Nature,* 437**,** 884-8.
PAUNGFOO-LONHIENNE, C., YEOH, Y. K., KASINADHUNI, N. R., LONHIENNE, T. G., ROBINSON, N., HUGENHOLTZ, P., RAGAN, M. A. & SCHMIDT, S. 2015. Nitrogen fertilizer dose alters fungal communities in sugarcane soil and rhizosphere. *Sci Rep,* 5**,** 8678.
PEÑALOZA, E., SANTIAGO, M., CABRERA, S., MUÑOZ, G., CORCUERA, L. J. & SILVA, H. 2016. Characterization of the high-affinity phosphate transporter PHT1;4 gene promoter of Arabidopsis thaliana in transgenic wheat. *Biologia Plantarum,* 61**,** 453-462.
PEOPLES, M. & CRASWELL, E. 1992. Biological nitrogen fixation: Investments, expectations and actual contributions to agriculture. *Plant and Soil,* 141**,** 13-39.

PEREIRA E SILVA, M. C., DIAS, A. C., VAN ELSAS, J. D. & SALLES, J. F. 2012. Spatial and temporal variation of archaeal, bacterial and fungal communities in agricultural soils. *PLoS One,* 7**,** e51554.
PHILIPPOT, L., RAAIJMAKERS, J. M., LEMANCEAU, P. & VAN DER PUTTEN, W. H. 2013. Going back to the roots: the microbial ecology of the rhizosphere. *Nat Rev Microbiol,* 11**,** 789-99.
PONNIAH, S. K., THIMMAPURAM, J., BHIDE, K., KALAVACHARLA, V. K. & MANOHARAN, M. 2017. Comparative analysis of the root transcriptomes of cultivated sweetpotato (Ipomoea batatas [L.] Lam) and its wild ancestor (Ipomoea trifida [Kunth] G. Don). *BMC Plant Biol,* 17**,** 9.
PRUITT, K. D., TATUSOVA, T. & MAGLOTT, D. R. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res,* 35**,** D61-5.
QU, X. & CHRIST, B. J. 2006. Host Range of Spongospora subterranea f. sp. subterranea in the United States. v. 83.
QUAST, C., PRUESSE, E., YILMAZ, P., GERKEN, J., SCHWEER, T., YARZA, P., PEPLIES, J. & GLOCKNER, F. O. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. *Nucleic Acids Res,* 41**,** D590-6.
R CORE TEAM. 2013. *R: A language and environment for statistical computing.* [Online]. Vienna, Austria.: R Foundation for Statistical Computing. Available: http://www.R-project.org/ [Accessed].
RAAIJMAKERS, J. M., PAULITZ, T. C., STEINBERG, C., ALABOUVETTE, C. & MOËNNE-LOCCOZ, Y. 2008. The rhizosphere: a playground and battlefield for soilborne pathogens and beneficial microorganisms. *Plant and Soil,* 321**,** 341-361.
RAGOTHAMA, K. G. 1999. Phosphate acquisition. *Ann. Rev. Plant Physiol.,* 50**,** 665-693.
RAMIREZ, K. S., LAUBER, C. L., KNIGHT, R., BRADFORD, M. A. & FIERER, N. 2010. Consistent effects of nitrogen fertilization on soil bacterial communities in contrasting systems. *Ecology,* 91**,** 3463-3470.
RANJAN, R., RANI, A., METWALLY, A., MCGEE, H. S. & PERKINS, D. L. 2016. Analysis of the microbiome: Advantages of whole genome shotgun versus 16S amplicon sequencing. *Biochem Biophys Res Commun,* 469**,** 967-77.
RAPPÉ, M. S. & GIOVANNONI, S. J. 2003. The Uncultured Microbial Majority. *Annual Review of Microbiology,* 57**,** 369-394.
RASMANN, S., KÖLLNER, T. G., DEGENHARDT, J., HILTPOLD, I., TOEPFER, S., KUHLMANN, U., GERSHENZON, J. & TURLINGS, T. 2005. Recruitment of entomopathogenic nematodes by insect-damaged maize roots. *Nature,* 434**,** 732-737.
RAUSCH, C., ZIMMERMANN, P., AMRHEIN, N. & BUCHER, M. 2004. Expression analysis suggests novel roles for the plastidic phosphate transporter Pht2;1 in auto- and heterotrophic tissues in potato and Arabidopsis. *Plant J,* 39**,** 13-28.
REMY, E., CABRITO, T. R., BATISTA, R. A., TEIXEIRA, M. C., SA-CORREIA, I. & DUQUE, P. 2012. The Pht1;9 and Pht1;8 transporters mediate inorganic phosphate acquisition by the

Arabidopsis thaliana root during phosphorus starvation. *New Phytol,* 195**,** 356-71.
RICHARDSON, A. E., BAREA, J.-M., MCNEILL, A. M. & PRIGENT-COMBARET, C. 2009. Acquisition of phosphorus and nitrogen in the rhizosphere and plant growth promotion by microorganisms. *Plant and Soil,* 321**,** 305-339.
ROCKSTROM, J., WILLIAMS, J., DAILY, G., NOBLE, A., MATTHEWS, N., GORDON, L., WETTERSTRAND, H., DECLERCK, F., SHAH, M., STEDUTO, P., DE FRAITURE, C., HATIBU, N., UNVER, O., BIRD, J., SIBANDA, L. & SMITH, J. 2017. Sustainable intensification of agriculture for human prosperity and global sustainability. *Ambio,* 46**,** 4-17.
ROESCH, L. F., FULTHORPE, R. R., RIVA, A., CASELLA, G., HADWIN, A. K., KENT, A. D., DAROUB, S. H., CAMARGO, F. A., FARMERIE, W. G. & TRIPLETT, E. W. 2007. Pyrosequencing enumerates and contrasts soil microbial diversity. *ISME J,* 1**,** 283-90.
ROGOZIN, I. B., WOLF, Y. I., SOROKIN, A. V., MIRKIN, B. G. & KOONIN, E. V. 2003. Remarkable Interkingdom Conservation of Intron Positions and Massive, Lineage-Specific Intron Loss and Gain in Eukaryotic Evolution. *Current Biology,* 13**,** 1512-1517.
ROSENBERG, K., BERTAUX, J., KROME, K., HARTMANN, A., SCHEU, S. & BONKOWSKI, M. 2009. Soil amoebae rapidly change bacterial community composition in the rhizosphere of Arabidopsis thaliana. *ISME J,* 3**,** 675-84.
ROUXEL, T. & BALESDENT, M. H. 2005. The stem canker (blackleg) fungus, Leptosphaeria maculans, enters the genomic era. *Molecular plant pathology,* 6**,** 225-241.
RUDRAPPA, T., CZYMMEK, K. J., PARE, P. W. & BAIS, H. P. 2008. Root-secreted malic acid recruits beneficial soil bacteria. *Plant Physiol,* 148**,** 1547-56.
SALEEM, M., LAW, A. D. & MOE, L. A. 2015. Nicotiana Roots Recruit Rare Rhizosphere Taxa as Major Root-Inhabiting Microbes. *Microb Ecol*.
SANGWAN, N., XIA, F. & GILBERT, J. A. 2016. Recovering complete and draft population genomes from metagenome datasets. *Microbiome,* 4**,** 8.
SASAKI, K., IKEDA, S., OHKUBO, T., KISARA, C., SATO, T. & MINAMISAWA, K. 2013. Effects of Plant Genotype and Nitrogen Level on Bacterial Communities in Rice Shoots and Roots. *Microbes and Environments,* 28**,** 391-395.
SAYERS, E. 2008. E-Utilities Quick Start. *In: Entrez Programming Utilities Help* [Online].
SCHACTMAN, D. P., REID, R. J. & AYLING, S. M. 1998a. Phosphorus Uptake by Plants: From Soil to Cell. *Plant Physiol,* 116**,** 447-453.
SCHACTMAN, D. P., REID, R. J. & AYLING, S. M. 1998b. Phosphorus uptake by plants: From soil to cell. *Plant Physiology,* 116**,** 447 -453.
SCHIRMER, M., IJAZ, U. Z., D'AMORE, R., HALL, N., SLOAN, W. T. & QUINCE, C. 2015. Insight into biases and sequencing errors for

amplicon sequencing with the Illumina MiSeq platform. *Nucleic Acids Res,* 43**,** e37.
SCHLOSS, P. D. & HANDELSMAN, J. 2006. Toward a census of bacteria in soil. *PLoS Comput Biol,* 2**,** e92.
SCHMIDT, T. S., MATIAS RODRIGUES, J. F. & VON MERING, C. 2014. Ecological consistency of SSU rRNA-based operational taxonomic units at a global scale. *PLoS Comput Biol,* 10**,** e1003594.
SELINGER, D. W., SAXENA, R. M., CHEUNG, K. J., CHURCH, G. M. & ROSENOW, C. 2003. Global RNA Half-Life Analysis in Escherichia coli Reveals Positional Patterns of Transcript Degradation. *Genome Research,* 13**,** 216-223.
SEMCHENKO, M., SAAR, S. & LEPIK, A. 2014. Plant root exudates mediate neighbour recognition and trigger complex behavioural changes. *New Phytol,* 204**,** 631-7.
SHI, L., SHI, T., BROADLEY, M. R., WHITE, P. J., LONG, Y., MENG, J., XU, F. & HAMMOND, J. P. 2013. High-throughput root phenotyping screens identify genetic loci associated with root architectural traits in Brassica napus under contrasting phosphate availabilities. *Ann Bot,* 112**,** 381-9.
SHI, S., NUCCIO, E. E., SHI, Z. J., HE, Z., ZHOU, J. & FIRESTONE, M. K. 2016. The interconnected rhizosphere: High network complexity dominates rhizosphere assemblages. *Ecol Lett*.
SHIN, H., SHIN, H. S., DEWBRE, G. R. & HARRISON, M. J. 2004. Phosphate transport in Arabidopsis: Pht1;1 and Pht1;4 play a major role in phosphate acquisition from both low- and high-phosphate environments. *Plant J,* 39**,** 629-42.
SINGH, A., KUMAR, P., GAUTAM, V., RENGASAMY, B., ADHIKARI, B., UDAYAKUMAR, M. & SARKAR, A. K. 2016. Root transcriptome of two contrasting indica rice cultivars uncovers regulators of root development and physiological responses. *Sci Rep,* 6**,** 39266.
SINGH, H. P., BATISH, D., KAUR, S., RAMEZANI, H. & KOHLI, R. K. 2002. Comparative phytotoxicity of four monoterpenes against Cassia occidentalis. *Association of Applied Biologists,* 141**,** 111-116.
SOLTANI, A., KHAVAZI, H., ASADI-RAHMANI, H., OMIDVARI, M., DAHAJI, P. & MIRHOSEYNI, H. 2010. Plant Growth Promoting Characteristics in Some Flavobacterium spp. Isolated from Soils of Iran. *Journal of Agricultural Science,* 2**,** 106-115.
STEWART, E. J. 2012. Growing unculturable bacteria. *J Bacteriol,* 194**,** 4151-60.
SUAREZ, C., RATERING, S., KRAMER, I. & SCHNELL, S. 2014. Cellvibrio diazotrophicus sp. nov., a nitrogen-fixing bacteria isolated from the rhizosphere of salt meadow plants and emended description of the genus Cellvibrio. *Int J Syst Evol Microbiol,* 64**,** 481-6.
SUGIYAMA, A., UEDA, Y., TAKASE, H. & YAZAKI, K. 2014a. Pyrosequencing assessment of rhizosphere fungal communities from a soybean field. *Can J Microbiol,* 60**,** 687-90.

SUGIYAMA, A., UEDA, Y., ZUSHI, T., TAKASE, H. & YAZAKI, K. 2014b. Changes in the bacterial community of soybean rhizospheres during growth in the field. *PLoS One,* 9**,** e100709.
SUTRA, L., RISÈDE, J. M. & GARDAN, L. 2000. Isolation of fluorescent pseudomonads from the rhizosphere of banana plants antagonistic towards root necrosing fungi. *Letters in Applied Microbiology,* 31**,** 289-293.
SUTTON, A. W. 1908. Brassica Crosses. *Botanical Journal of the Linnean Society,* 38**,** 337-349.
TABATA, S., KANEKO, T. & NAKAMURA Y, K. H., KATO T, ASAMIZU E, MIYAJIMA N, SASAMOTO S, KIMURA T, HOSOUCHI T, KAWASHIMA K, KOHARA M, MATSUMOTO M, MATSUNO A, MURAKI A, NAKAYAMA S, NAKAZAKI N, NARUO K, OKUMURA S, SHINPO S, TAKEUCHI C, WADA T, WATANABE A, YAMADA M, YASUDA M, SATO S, DE LA BASTIDE M, HUANG E, SPIEGEL L, GNOJ L, O'SHAUGHNESSY A, PRESTON R, HABERMANN K, MURRAY J, JOHNSON D, ROHLFING T, NELSON J, STONEKING T, PEPIN K, SPIETH J, SEKHON M, ARMSTRONG J, BECKER M, BELTER E, CORDUM H, CORDES M, COURTNEY L, COURTNEY W, DANTE M, DU H, EDWARDS J, FRYMAN J, HAAKENSEN B, LAMAR E, LATREILLE P, LEONARD S, MEYER R, MULVANEY E, OZERSKY P, RILEY A, STROWMATT C, WAGNER-MCPHERSON C, WOLLAM A, YOAKUM M, BELL M, DEDHIA N, PARNELL L, SHAH R, RODRIGUEZ M, SEE LH, VIL D, BAKER J, KIRCHOFF K, TOTH K, KING L, BAHRET A, MILLER B, MARRA M, MARTIENSSEN R, MCCOMBIE WR, WILSON RK, MURPHY G, BANCROFT I, VOLCKAERT G, WAMBUTT R, DÜSTERHÖFT A, STIEKEMA W, POHL T, ENTIAN KD, TERRYN N, HARTLEY N, BENT E, JOHNSON S, LANGHAM SA, MCCULLAGH B, ROBBEN J, GRYMONPREZ B, ZIMMERMANN W, RAMSPERGER U, WEDLER H, BALKE K, WEDLER E, PETERS S, VAN STAVEREN M, DIRKSE W, MOOIJMAN P, LANKHORST RK, WEITZENEGGER T, BOTHE G, ROSE M, HAUF J, BERNEISER S, HEMPEL S, FELDPAUSCH M, LAMBERTH S, VILLARROEL R, GIELEN J, ARDILES W, BENTS O, LEMCKE K, KOLESOV G, MAYER K, RUDD S, SCHOOF H, SCHUELLER C, ZACCARIA P, MEWES HW, BEVAN M, FRANSZ P 2000. Sequence and analysis of chromosome 5 of the plant Arabidopsis thaliana. *Nature,* 408**,** 823-826.
TADANO, T., OZAWA, K., SAKAI, H., OSAKI, M. & MATSUI, H. 1993. Secretion of acid phosphatase by the roots of crop plants under phosphorus-deficient conditions and some properties of the enzyme secreted by lupin roots. *Plant and soil,* 155/156**,** 95-98.
TAWARAYA, K., HORIE, R., SAITO, S., WAGATSUMA, T., SAITO, K. & OIKAWA, A. 2014. Metabolite Profiling of Root Exudates of Common Bean under Phosphorus Deficiency. *Metabolites,* 4**,** 599-611.
TEDERSOO, L., BAHRAM, M., PÕLME, S., KÕLJALG, U., YOROU, N. S., WIJESUNDERA, R., RUIZ, L. V., VASCO-PALACIOS, A. M., THU, P. Q., SUIJA, A., SMITH, M. E., SHARP, C., SALUVEER, E.,

SAITTA, A., ROSAS, M., RIIT, T., RATKOWSKY, D., PRITSCH, K., PÕLDMAA, K., PIEPENBRING, M., PHOSRI, C., PETERSON, M., PARTS, K., PÄRTEL, K., OTSING, E., NOUHRA, E., NJOUONKOU, A. L., NILSSON, R. H., MORGADO, L. N., MAYOR, J., MAY, T. W., MAJUAKIM, L., LODGE, D. J., LEE, S. S., LARSSON, K.-H., KOHOUT, P., HOSAKA, K., HIIESALU, I., HENKEL, T. W., HAREND, H., GUO, L.-D., GRESLEBIN, A., GRELET, G., GEML, J., GATES, G., DUNSTAN, W., DUNK, C., DRENKHAN, R., DEARNALEY, J., DE KESEL, A., DANG, T., CHEN, X., BUEGGER, F., BREARLEY, F. Q., BONITO, G., ANSLAN, S., ABELL, S. & ABARENKOV, K. 2014. Global diversity and geography of soil fungi. *Science,* 346.

TENG, W., ZHAO, Y. Y., ZHAO, X. Q., HE, X., MA, W. Y., DENG, Y., CHEN, X. P. & TONG, Y. P. 2017. Genome-wide Identification, Characterization, and Expression Analysis of PHT1 Phosphate Transporters in Wheat. *Front Plant Sci,* 8**,** 543.

THIELE-BRUHN, S. 2003. Pharmaceutical antibiotic compounds in soils - a review. *J. Plant Nutr. Soil Sci.,* 166**,** 145-167.

THOMAS, F. & CEBRON, A. 2016. Short-Term Rhizosphere Effect on Available Carbon Sources, Phenanthrene Degradation, and Active Microbiome in an Aged-Contaminated Industrial Soil. *Front Microbiol,* 7**,** 92.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol,* 28**,** 511-5.

TURNER, B. L. & HAYGARTH, P. M. 2005. Phosphatase activity in temperate pasture soils: Potential regulation of labile organic phosphorus turnover by phosphodiesterase activity. *Sci Total Environ,* 344**,** 27-36.

TURNER, T. R., RAMAKRISHNAN, K., WALSHAW, J., HEAVENS, D., ALSTON, M., SWARBRECK, D., OSBOURN, A., GRANT, A. & POOLE, P. S. 2013. Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants. *ISME J,* 7**,** 2248-58.

ULLRICH, C. I. & NOVACKY, A. J. 1990. Extra- and Intracellular pH and Membrane Potential Changes Induced by K+, Cl-, H2PO4-, and NO3 Uptake and Fusicoccin in Root Hairs of Limnobium stoloniferum. *Plant Physiology,* 94**,** 1561-1567.

UMAMAHESWARI, T., ANBUKKARASI, K., HEMALATHA, T. & CHENDRAYAN, K. 2013. Studies on phytohormone producing ability of indigenous endophytic bacteria isolated from tropical legume crops. *International journal of Current Microbiology and Applied Sciences,* 2**,** 127-136.

UNITED NATIONS, D. O. E. A. S. A. 2015. World population prospects: THe 2015 revision. New York: United Nations.

VAN DE PEER, Y., CHAPELLA, S. & DE WACHTER, R. 1996. A quantitative map of nucleotide substitution rates in bacterial rRNA. *Nucleic Acids Res,* 24**,** 3381-3391.

VAN DER HEIJDEN, M. G., DE BRUIN, S., LUCKERHOFF, L., VAN LOGTESTIJN, R. S. & SCHLAEPPI, K. 2016. A widespread plant-fungal-bacterial symbiosis promotes plant biodiversity, plant nutrition and seedling recruitment. *ISME J,* 10**,** 389-99.
VERSAW, W. K. 2002. A Chloroplast Phosphate Transporter, PHT2;1, Influences Allocation of Phosphate within the Plant and Phosphate-Starvation Responses. *The Plant Cell Online,* 14**,** 1751-1766.
VETROVSKY, T. & BALDRIAN, P. 2013. The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses. *PLoS One,* 8**,** e57923.
WAGNER, M. R., LUNDBERG, D. S., DEL RIO, T. G., TRINGE, S. G., DANGL, J. L. & MITCHELL-OLDS, T. 2016. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat Commun,* 7**,** 12151.
WANG, D. Z., KONG, L. F., LI, Y. Y. & XIE, Z. X. 2016. Environmental Microbial Community Proteomics: Status, Challenges and Perspectives. *Int J Mol Sci,* 17.
WANG, H. B., ZHANG, Z. X., LI, H., HE, H. B., FANG, C. X., ZHANG, A. J., LI, Q. S., CHEN, R. S., GUO, X. K., LIN, H. F., WU, L. K., LIN, S., CHEN, T., LIN, R. Y., PENG, X. X. & LIN, W. X. 2011. Characterization of metaproteomics in crop rhizospheric soil. *J Proteome Res,* 10**,** 932-40.
WANG, H. C., XIA, X. & HICKEY, D. 2006. Thermal adaptation of the small subunit ribosomal RNA gene: a comparative study. *J Mol Evol,* 63**,** 120-6.
WANG, Z., GERSTEIN, M. & SNYDER, M. 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics,* 10**,** 57-63.
WHITE, T. J., BRUNS, T., LEE, S. & TAYLOR, J. 1990. *PCR Protocols: a guide to methods and applications.,* New York, USA, Academic Press.
WICKHAM, H. 2009. *ggplot2: Elegant Graphics for Data Analysis,* New York, Springer-Verlag.
WILKINSON, D. M., CREEVY, A. L. & VALENTINE, J. 2012. The Past, Present and Future of Soil Protist Ecology. *Acta Protozoologica,* 51**,** 189-199.
WOESE, C., KANDLER, O. & WHEELIS, M. L. 1990. Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci U S A,* 87**,** 4576-4579.
XU, C., JIAO, C., ZHENG, Y., SUN, H., LIU, W., CAI, X., WANG, X., LIU, S., XU, Y., MOU, B., DAI, S., FEI, Z. & WANG, Q. 2015. De novo and comparative transcriptome analysis of cultivated and wild spinach. *Sci Rep,* 5**,** 17706.
YAN, Y., KURAMAE, E. E., DE HOLLANDER, M., KLINKHAMER, P. G. & VAN VEEN, J. A. 2016. Functional traits dominate the diversity-related selection of bacterial communities in the rhizosphere. *ISME J*.
YANG, S. Y., GRONLUND, M., JAKOBSEN, I., GROTEMEYER, M. S., RENTSCH, D., MIYAO, A., HIROCHIKA, H., KUMAR, C. S., SUNDARESAN, V., SALAMIN, N., CATAUSAN, S., MATTES, N.,

HEUER, S. & PASZKOWSKI, U. 2012. Nonredundant regulation of rice arbuscular mycorrhizal symbiosis by two members of the phosphate transporter1 gene family. *Plant Cell,* 24**,** 4236-51.
YEOH, Y. K., PAUNGFOO-LONHIENNE, C., DENNIS, P. G., ROBINSON, N., RAGAN, M. A., SCHMIDT, S. & HUGENHOLTZ, P. 2016. The core root microbiome of sugarcanes cultivated under varying nitrogen fertilizer application. *Environ Microbiol,* 18**,** 1338-51.
YERGEAU, E., SANSCHAGRIN, S., MAYNARD, C., ST-ARNAUD, M. & GREER, C. W. 2014. Microbial expression profiles in the rhizosphere of willows depend on soil contamination. *ISME J,* 8**,** 344-58.
YORK, L. M., CARMINATI, A., MOONEY, S. J., RITZ, K. & BENNETT, M. J. 2016. The holistic rhizosphere: integrating zones, processes, and semantics in the soil influenced by roots. *J Exp Bot,* 67**,** 3629-43.
ZANCARINI, A., MOUGEL, C., VOISIN, A. S., PRUDENT, M., SALON, C. & MUNIER-JOLAIN, N. 2012. Soil nitrogen availability and plant genotype modify the nutrition strategies of M. truncatula and the associated rhizosphere microbial communities. *PLoS One,* 7**,** e47096.
ZARRAONAINDIA, I., OWENS, S. M., WEISENHORN, P., WEST, K., HAMPTON-MARCELL, J., LAX, S., BOKULICH, N. A., MILLS, D. A., MARTIN, G., TAGHAVI, S., VAN DER LELIE, D. & GILBERT, J. A. 2015. The soil microbiome influences grapevine-associated microbiota. *MBio,* 6.
ZHANG, J., SUBRAMANIAN, S., STACEY, G. & YU, O. 2009. Flavones and flavonols play distinct critical roles during nodulation of Medicago truncatula by Sinorhizobium meliloti. *Plant J,* 57**,** 171-83.
ZHANG, Y., XU, J., RIERA, N., JIN, T., LI, J. & WANG, N. 2017. Huanglongbing impairs the rhizosphere-to-rhizoplane enrichment process of the citrus root-associated microbiome. *Microbiome,* 5**,** 97.
ZHU, S., VIVANCO, J. M. & MANTER, D. K. 2016. Nitrogen fertilizer rate affects root exudation, the rhizosphere microbiome and nitrogen-use-efficiency of maize. *Applied Soil Ecology,* 107**,** 324-333.