School of Environmental Sciences

Metatranscriptomic analysis of community structure and metabolism of the rhizosphere microbiome

by

Thomas Richard Turner

Submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy, September 2013

Declaration

I declare that this is an account of my own research and has not been submitted for a degree at any other university. The use of material from other sources has been properly and fully acknowledged, where appropriate.

Thomas Richard Turner

Acknowledgements

I would like to thank my supervisors, Phil Poole and Alastair Grant, for their continued support and guidance over the past four years. I'm grateful to all members of my lab, both past and present, for advice and friendship. Graham Hood, I don't know how we put up with each other, but I don't think I could have done this without you. Cheers Salt! KK, thank you for all your help in the lab, and for Uma's biryanis! Andrzej Tkatcz, thanks for the useful discussions about our projects. Alison East, thank you for all your support, particularly ensuring Graham and I did not kill each other.

I'm grateful to Allan Downie and Colin Murrell for advice. For sequencing support, I'd like to thank TGAC, particularly Darren Heavens, Sophie Janacek, Kirsten McKlay and Melanie Febrer, as well as John Walshaw, Mark Alston and David Swarbreck for bioinformatic support. To all of the friends I've made during my time here, particularly Matt, Jack, Henry and Ellis for our regular pub visits. To Ade, Kate and Beth for many successful pub quiz nights, and to Joy, Dave, Hannah, Tori and Millie for giving me a home away from home. Finally, I'm very grateful to my family for constantly supporting me, particularly my Mum and Dad, my sister Ellie and soon-to-be-brother-in-law Gary, as well as my late grandparents Fred and Hilda Winifred Turner, who inspired me to do this PhD project.

Abstract

Plant-microbe interactions in the rhizosphere, the region of soil influenced by plant roots, are integral to biogeochemical cycling, and maintenance of plant health and productivity. Interactions between model plants and microbes are well understood, but relatively little is about the plant microbiome. Here, comparative metatranscriptomics was used to determine taxonomic compositions and metabolic responses of microbes in soil and the rhizospheres of wheat, oat and pea. Additionally a wild-type oat was compared to a mutant (*sad1*) deficient in production of antifungal avenacins.

Analyses of taxonomic compositions and functions based on rRNA and protein coding genes agreed that rhizosphere microbiomes differed from soil and between plant species. Pea had a stronger effect than wheat and oat, suggesting distinct cereal and legume microbiomes. Proportions of eukaryotic rRNA in the oat and pea rhizospheres were more than fivefold higher than in the wheat rhizosphere or soil. Nematodes and bacterivorous protozoa were enriched in all rhizospheres, while the pea rhizosphere was highly enriched for fungi. Only the eukaryotic community was distinct from wild-type oat in the *sad1* mutant, suggesting avenacins have a broader role than protecting from fungal pathogens.

The addition of an internal RNA standard allowed quantitative determination of global transcriptional activity in each environment. This was generally higher in the rhizospheres, particularly pea, than in soil. Taxa known to possess metabolic traits potentially important for rhizosphere colonisation, plant growth promotion and pathogenesis were selected by plants. Such traits included cellulose and other plant polymer degradation, nitrogen fixation, hydrogen oxidation, methylotrophy and antibiotic production. These functions were also more highly expressed in rhizospheres than soil. Microbes also induced functions involved in chemotaxis, motility, attachment, pathogenesis, responses to oxidative stress, cycling of nitrogen and sulphur, acquisition of phosphorous, iron and other metals, as well as metabolism of a variety of sugars, aromatics, organic and amino acids, many plant species specific.

Profiling microbial communities with metatranscriptomics allowed comparison of relative and quantitative abundance of microbes and their metabolism, from multiple samples, across all domains of life, without PCR bias. This revealed profound differences in the taxonomic composition and metabolic functions of rhizosphere microbiomes between crop plants and soil.

Contents

Chapter 1: The plant microbiome

1.1 Introduction

Microbes are the most phylogenetically and functionally diverse organisms on the planet. They are fundamental to the maintenance of life on Earth, yet we understand little about the uncultured majority of microbes in environments such as soils, oceans, the atmosphere and even those living on and in our own bodies. Culture-dependent techniques have long allowed the study of microbial isolates in great detail, albeit in artificial laboratory environments, while culture-independent molecular techniques are allowing whole microbial communities to be studied in their natural environments. Microbial community profiling of environments has become common place with high-throughput techniques such as 16S rRNA gene amplicon sequencing. The microbial communities, or microbiomes, of diverse environments have been studied in this way, with the goal of understanding their ecological function (Gilbert *et al.*, 2010). Interest in the human microbiome has increased in recent years (Turnbaugh *et al.*, 2007). Numerous associations have been made between different microbial groups and host traits such as disease (Greenblum *et al.*, 2012), diet (Martinez *et al.*, 2012; Turnbaugh *et al.*, 2009) and genetics (Spor *et al.*, 2011), and manipulation of the human microbiome has recently shown efficacy in treating diseases (Brandt, 2013).

Similarly, the plant microbiome is considered a key determinant of plant health and productivity (Berendsen *et al.*, 2012), and efforts to increase understanding of it are being made (Bulgarelli *et al.*, 2013; Lebeis *et al.*, 2012). As the most important terrestrial primary producers, plants perform a vital step of the carbon cycle (i.e. photosynthesis). The translocation of fixed carbon (photosynthate) to roots and their associated microbes in the soil is another important part of the cycle. The diverse, and sometimes unique, metabolic capabilities of microbes, particularly the prokaryotic Bacteria and Archaea, means they are involved in cycling of nitrogen, phosphorous, sulphur, and other elements. Plant associated microbes are therefore key players in biogeochemical cycles globally.

The tissues and surfaces of a plant that can host a microbial community can be grouped into three main niches: the rhizosphere, phyllosphere, and endosphere. The rhizosphere is the interface between soil and roots. A region of rich, largely soil derived, microbial diversity, influenced by deposition of plant mucilage and root exudates (Kent and Triplett, 2002). By contrast, the phyllosphere, or aerial surfaces, are relatively nutrient poor and are

subject to extremes of temperature, radiation and moisture (Vorholt, 2012). Microbial inhabitants of the rhizosphere and phyllosphere are considered epiphytes, while microbes residing within plant tissues are considered endophytes. Microbes in these niches can establish beneficial, neutral or detrimental associations of varying intimacy with their host plants. The importance of particular plant-microbe interactions has been known for centuries. Most notably that of *Rhizobium*-legume symbioses, which contributed to the development of crop rotation systems that led to increased agricultural production. Such model systems are well understood (Oldroyd *et al.*, 2011), but overall the plant microbiome, which can be considered an extended phenotype of the host plant, is not as yet well defined

Manipulation of the plant microbiome has the potential to reduce incidence of plant disease (Andrews, 1992; Bloemberg and Lugtenberg, 2001) and increase agricultural production (Bakker *et al.*, 2012), while reducing chemical inputs (Adesemoye *et al.*, 2009) and emissions of greenhouse gasses (Singh *et al.*, 2010), resulting in more sustainable agricultural practices. This goal is seen as vital for sustaining the world's growing population and reducing some contributors to anthropogenic climate change.

1.2 Approaches used to study the plant microbiome

1.2.1 Culture dependent approaches

Classic microbiology involves isolating and culturing microbes from an environment using different nutrient media and growth conditions depending on the target organisms. While obtaining a pure culture of an organism is required for detailed studies of its genetics and physiology, culture-dependent techniques miss the vast majority of microbial diversity in an environment. Single cell sequencing (Hutchison and Venter, 2006) and mini-metagenome approaches (McLean *et al.*, 2013) are bridging the gap between culture dependent and independent methods, but these techniques are still in their infancy. It is estimated from diversity of DNA in soils that as little as 0.1% to 1% of microbial species from soil are culturable in any given set of conditions (Torsvik *et al.*, 1990; Torsvik and Ovreas, 2002). Similarly the "great plate count anomaly" reflects the differences between what can be seen under a microscope and what can be observed growing on nutrient agar in a petri dish (Staley and Konopka, 1985).

A major limitation to culturing elusive microbial taxa from soil is the presence of fast growing microbes. Given a rich media, they will out-compete the majority of other species. Because of their high growth rate they are easier to isolate and subsequently study. However it is suspected the vast majority of microbes are slow growing and are rarely growing at optimum rate in their natural environment. Using nutrient poor media and long incubation periods has allowed culturing of novel strains of soil microbes, reflecting those that are detected using molecular techniques in soil and plant rhizospheres, such as the Acidobacteria (George *et al.*, 2011), Verrucomicrobia (da Rocha *et al.*, 2010) and others (da Rocha *et al.*, 2009; Davis *et al.*, 2011; Stewart, 2012). Determining the key carbon sources being metabolised by the dominant microbial species in a nematode gut environment allowed it to be cultured where previous attempts had failed (Bomar *et al.*, 2011). While the nematode gut environment is much less complex in terms of host influence and microbiome structure compared to an environment such the rhizosphere, culturing soil microbes on known plant derived carbon sources could be used to isolate novel rhizosphere microbes. Such compounds could include carbon sources such as cellulose, pectin, phenolics or terpenes, nitrogen sources such as aminocyclopropane carboxylic acid (ACC) or alkaloids, or sulphur sources such as glucosinolates or arylsulphates. Tolerance of plant defence compounds could be considered an additional screen for isolating novel microbes from plant niches. Successful rhizosphere colonisers should be able to tolerate general and widespread plant chemical defences such as salicylic acid, and reactive oxygen species (ROS), and plant specific toxins such as the avenacins of oat(Maizel *et al.*, 1964), and pisatin of pea (Perrin and Bottomley, 1961), if they are associated with such plants. These approaches require physiologically relevant concentrations of such compounds.

Growth temperature is another key factor influencing microbial survivability and growth. Microbes tolerate a wide range of temperatures in their natural environment but typical isolations are incubated between 27°C and 37°C. Incubating at lower temperatures, or temperatures comparable to the isolation site will likely improve discovery rates of new species. Of some concern is the overlap in ability of some bacteria to promote plant growth and also cause opportunistic infections in humans, such as in species of *Pseudomonas* (Wu et al., 2012), *Stenotrophomnonas* (Ryan *et al.*, 2009) and *Acinetobacter* (Peleg et al., 2008; Rokhbakhsh-Zamin et al., 2011). Growth at 37°C is a prerequisite for human pathogenesis, and it has been suggested that potential PGPRs are screened for this ability and only those unable to do so are considered for further study and application in the field. A further challenge to isolating novel microbes is their sheer diversity in soil and the rhizosphere. The

life styles of archaea and eukaryotes can be quite different to those of bacteria, and they have typically been neglected from attempts to culture novel organisms. Microbes living in complex communities can in some respects be considered part of a superorganism, each group carrying out specialised biochemical transformations. Some bacteria therefore are unable survive without other microbes, making their isolation particularly difficult because they require co-culture.

1.2.2 Ribosomal RNA and other genes as phylogenetic markers

Ribosomal RNA genes (or rDNA) are ubiquitous in cellular organisms, including bacteria. They encode structural RNA components of the ribosome, the protein synthesis machinery of the cell and are therefore essential. In prokaryotes, three genes encode the subunits of rRNA characterised by the sedimentation properties of the RNA (i.e. 5S, 16S and 23S). In eukaryotes, rRNA genes are arranged differently with 5S, 5.8S, 18S and 26S or 28S products. Modern molecular taxonomy, particularly for prokaryotes, is based on the relatedness of these sequences between organisms (Woese, 1987). Differences in 16S rDNA sequence were first used to propose what we now know to be Archaea as a separate domain of life, distinct from Eubacteria and Eukaryota (Woese *et al.*, 1990). Ribosomal RNA genes have become the benchmark in culture-independent analyses of microbial communities, although increasingly other marker genes and even whole genomes are being used.

1.2.3 Genetic fingerprinting

Variation in target DNA sequence allows identification of organisms at different taxonomic levels depending on the evolutionary rate of the target sequence and the sensitivity of the technique. Before sequencing of nucleic acids was widely available and affordable on the scales needed for microbial ecology, other techniques were developed to examine differences in the sequences of members of the community being studied. Commonly, a marker gene of interest is amplified by polymerase chain reaction (PCR)(Mullis *et al.*, 1986) from an environmental DNA sample. The amplified product is then exposed to a denaturing treatment or restriction enzymes resulting in a fragmentation pattern, when separated with electrophoresis, which is reflective of the community structure. Such techniques include denaturing gradient gel electrophoresis (DGGE) (Muyzer *et al.*, 1993) and terminal

restriction fragment length polymorphism (TRFLP) (Liu *et al.*, 1997). A variation on this is to amplify a size-variable region of DNA such as the internally transcribed spacer (ITS) between the 16S and 23S rRNA genes, as in automated ribosomal intergenic spacer analysis (ARISA) (Fisher and Triplett, 1999; Garcia-Martinez *et al.*, 1999). ARISA requires no additional treatment after the initial PCR. It is possible to measure the size and abundance of the fragments, and these data can be used to generate graphs based on principal component analysis (PCA) or multidimensional scaling (MDS), allowing the community structure, or more often differences between several community structures to be visualised. Fragment bands that are different between communities can be gel extracted and sequenced to identify the organism. These fingerprinting techniques have been extensively used with the 16S rRNA gene, or cDNA derived from reverse transcription of 16S rRNA, to study rhizosphere microbial communities (Costa *et al.*, 2006; Garbeva *et al.*, 2008; Haichar *et al.*, 2008; Kuske *et al.*, 2002; Tkacz *et al.*, 2013b), although other phylogenetic and functional markers genes have also been used (Bremer *et al.*, 2009; Haichar *et al.*, 2012).

1.2.4 High throughput analysis of 16S rRNA gene sequences

The recent availability and affordability of high-throughput sequencing technologies such as Roche's 454 (Margulies *et al.*, 2005) and Illumina's HiSeq and MiSeq platforms (Bentley *et al.*, 2008) has revolutionised microbial ecology. Their wide adoption by the scientific community is due to their generation of huge amounts of sequence data at a greatly reduced cost per base-pair (bp) compared to traditional Sanger sequencing (Sanger *et al.*, 1977). In addition, they do not require cloning of PCR products prior to sequencing as was traditionally done. Multiple samples can be pooled and sequenced then sorted downstream based on a unique barcode, a technique termed multiplexing. While the majority of microbial ecology studies so far have been carried out using 454 pyrosequencing, Illumina's HiSeq and MiSeq platforms are likely to dominate in future (Caporaso *et al.*, 2012; Degnan and Ochman, 2012).

Amplification and sequencing of a variable region of the 16S rRNA gene from environmental samples is now routine and has contributed to our understanding of microbial diversity in several rhizosphere environments. These include the rhizospheres of model organisms such as *Arabidopsis thaliana* (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012;

Tkacz *et al.*, 2013b), *Medicago truncatula* and *Brachypodium diastychon* (Tkacz *et al.*, 2013b), crop plant such as potato (*Solanum tuberosum*)(Inceoglu *et al.*, 2011) and maize (*Zea mays*)(Peiffer *et al.*, 2013), and trees such as Oak (*Quercus sp.*)(Uroz *et al.*, 2010) and Poplar (*Populus deltoides*)(Gottel *et al.*, 2011). Aside from pyrosequencing, microarray technologies have also been used to study rhizosphere microbiomes of maize (*Zea mays*) (Bouffaud *et al.*, 2012) and sugar beet (*Beta vulgaris*)(Mendes *et al.*, 2011), while 16S rRNA gene clone libraries have been used in conjunction with shotgun metagenomics to study bacteria in the rice (*Oryza sativa*) rhizosphere(Knief *et al.*, 2012). These studies have all highlighted remarkable consistency in the dominance of Proteobacteria among rhizosphere bacterial phyla, which other large contributions from Actinobacteria, Firmicutes, Planctomycetes, Bacteroidetes and also Acidobacteria.

An important limitation of these approaches is that PCR amplification of genomic DNA is inherently biased by primer design. It is generally only capable of detecting a particular target group of organisms, and even then gives a biased sample within that target group (Hong *et al.*, 2009; Pinto and Raskin, 2012). However, complex environments are occupied by organisms from all domains of life. Eukaryotes, including fungi, protozoa, oomycetes, and nematodes, are ubiquitous in soils and can be important plant pathogens or symbionts, while others are bacterial grazers. The Archaea, carry out important biogeochemical reactions, particularly in agricultural soils, such as ammonia oxidation (Leininger *et al.*, 2006) and methanogenesis (Conrad *et al.*, 2006). Viruses too are found wherever there are cellular organisms, and these can affect the population dynamics of their hosts (Williams, 2013). In the rhizosphere, members of a community interact with each other as well as the plant host (Barea *et al.*, 2005), so it is important to try and capture the entire diversity of a microbiome. To do so requires the use of global analyses such as metagenomics, metatranscriptomics and metaproteomics which allow simultaneous assessment and comparison of microbial populations across all domains of life.

1.2.5 Metagenomics

A metagenome, in the strictest sense, is the combined genomes of all organisms within a particular environment. In practice only fractions of genomes from many organisms are sampled, but this approach is far more encompassing that a targeted approach using PCR for example. The original metagenomic studies of oceans and soils and others, cloned

genomic DNA (gDNA) from the environment into a heterologous host, typically *Escherichia coli*. The bacterial artificial chromosomes (BACs) could be sequenced to identify the nature of the insert, and which organism it came from. Such studies detected a wide range of microbial taxa, the expression of a wide range of functional genes, and also gene products such as antibiotics and enzymes (Donato *et al.*, 2010; Gillespie *et al.*, 2002; Jiang *et al.*, 2009; Rondon *et al.*, 2000; Venter *et al.*, 2004). Alternatively or additionally, the heterologous host containing the BAC or other vector could be functionally screened for a particular product (Tett *et al.*, 2012). Examples from the rhizosphere environment include novel lipases (Lee *et al.*, 2010), antibiotics (Chung *et al.*, 2008) and nickel resistance genes (Mirete *et al.*, 2007). Cloning into a heterologous host has some major limitations (Temperton *et al.*, 2009). Firstly, the size of insert is limited by the type of construct it is cloned into, resulting in bias against larger inserts. Additionally, introduction of foreign gDNA may result in production of a product toxic to the cell. Cells containing these inserts are then not recovered and therefore not represented in the subsequent analysis. There are also limitations to the number of colonies that can be picked and sequenced, and the majority of sequenced DNA is derived from the insert. Quantitative information is also lost due to the different replication rates of the plasmids within the host.

The advent of high-throughput, direct sequencing technologies has vastly improved the depth of information and accuracy of true shot-gun metagenomic approaches. They have been demonstrated to more accurately represent known simulated microbial communities than PCR based amplicon studies (Shakya *et al.*, 2013). Near complete genomes from the dominant bacteria in low diversity environments such as acid mine drains have been sequenced in this way (Tyson *et al.*, 2004). In more complex samples this is still not realistic, but it is now possible to obtain vast amounts of information on the presence and abundance of genes encoding for particular metabolic pathways (Handelsman, 2004) and even non-coding RNA species (Weinberg *et al.*, 2009).

Taxonomic information can be provided by potentially all sequences, but more commonly ubiquitous, essential, and slowly evolving genes are markers providing a general overview of taxonomic composition, such as those used by MetaPhyler (Liu *et al.*, 2010). These include rRNA genes, *rpoB,* EF-Tu, *dnaG*, HSP70, *recA* and others (Wu and Eisen, 2008). Alternatively, functionally important genes might be considered as taxonomic markers, revealing the organisms behind respective processes. Genes encoding enzymes for key steps in nutrient cycling are often used for this purpose, for example, the *nifH* gene

encoding the catalytic subunit of nitrogenase (Ueda *et al.*, 1995).

Metagenomics allows detection of organisms from all domains of life (Bacteria, Archaea and Eukaryotes) and also viruses, avoiding the bias associated with primer annealing and PCR amplification (Hong *et al.*, 2009; Pinto and Raskin, 2012). However, it is limited to detecting the presence of an organism. The activity of many organisms in an environment such as soil may be very low, and thus they contribute little to the functioning of that ecosystem at that particular time. The rhizosphere microbiome is selected from that of surrounding soil. The plant causes an increase in the abundance of some taxa but a reduction in the abundance of others. Plants will also influence the activity of microbes by providing sources of carbon and energy. The use of stable isotope techniques (SIP) (Radajewski *et al.*, 2000) with 16S rRNA based DGGE has shown a that a subset of the rhizosphere bacterial community is primarily utilising plant derived carbon (Haichar *et al.*, 2008; Lu *et al.*, 2006). Coupling SIP with metagenomics would give a more global picture of this subset, though this is still limited in its ability to provide detailed information on those microbes most active in the rhizosphere.

1.2.6 Metatranscriptomicsand the challenge of mRNA enrichment

A metatranscriptome, or the total pool of RNA, from a microbial community, provides a snap-shot of community wide gene expression. The dominance of rRNA in a metatranscriptome allows robust community profiling of organisms from all domains of life. This has been applied to study the oceans (Ottesen *et al.*, 2011; Shi *et al.*, 2010), soil (Urich *et al.*, 2008) and recently the rhizospheres of crop plants (Turner *et al.*, 2013). Metatranscriptomes also provide information on the expression of non-coding and small RNA species (ncRNA and sRNA) (Shi *et al.*, 2009b) which have important regulatory roles in bacteria (Narberhaus and Vogel, 2009).The main focus of metatranscriptomics however, has been to provide information on the active metabolic pathways in the studied environment.

The transcriptomes of actively growing organisms, whether in pure culture or complex communities, are dominated by ribosomal RNA (rRNA) (Hewson *et al.*, 2009; Neidhardt and Umbarger, 1996), which can represent over 90% of RNA species. Even with the depth of sequencing now possible with high-throughput technologies (Bentley *et al.*, 2008; Margulies *et al.*, 2005), the enrichment of messenger RNA (mRNA) is needed for studies of

the transcriptome. The dominance of rRNA, particularly the 16S and 23S subunits of prokaryotes, can be visualised on native agarose gels. Excision and purification of all but these two dominant bands been used to enrich for mRNA (McGrath *et al.*, 2008). While enriching to some extent, this method would have removed any mRNA with similar molecular weight to the rRNA subunits. It would also fail to remove processed or degraded rRNA, or the smaller 5S subunit. Additionally, large quantities of input RNA are required and the extraction can result in degradation of the RNA sample. Such a risky procedure is not recommended for precious environmental samples, the yields of which are often low and sampling effort high.

The 3' ends of most eukaryotic mRNA transcripts are poly-adenylated, resulting in a poly-A tail (Zhao *et al.*, 1999), allowing specific, efficient and straight-forward recovery using complementary poly-thymidylated (poly-T) columns or magnetic beads. This technique has been applied to study the metatranscriptomes of several soil environments (Bailly *et al.*, 2007; Damon *et al.*, 2012; Takasaki *et al.*, 2013). However, prokaryotic mRNA lacks poly-A tails so cannot be recovered in this way. Polyadenylation of prokaryotic mRNA using a poly-A polymerase enzyme from *Escherichia coli* has been used in studies of marine metranscriptomes (Frias-Lopez *et al.*, 2008; Shi *et al.*, 2009b), but was only partially successful with a soil metatranscriptome (Botero *et al.*, 2005). Archaeal transcripts were not present in the mRNA enriched sample, but were shown to be expressed using quantitative reverse transcriptome polymerase chain reaction (qRT-PCR) (Botero *et al.*, 2005). For studies of prokaryote or mixed population transcriptomes (metatranscriptomes), the depletion of rRNA is favoured. Though this often results in incomplete removal of rRNA from the sample, it is preferable to introducing bias by taking only a sub-set of the mRNA. In single species pure cultures, the rRNA sequences are identical, allowing highly efficient removal. This has been exploited by a number of commercial kits, which are often tested on *E. coli* and *Bacillus subtilis*. Compatibility with other species is variable and lists on the respective providers' websites are updated when user information is available for other microbes. Mixed populations contain enough variation in rRNA sequence to become a challenge to most sequence dependent depletion methods. More importantly, the population is largely unknown, so the compatibility lists are not particularly useful.

Sequence dependent rRNA depletion methods are based on subtractive hybridisation, whereby complimentary rRNA oligos or longer probes bind to the rRNA in the sample. Both are subsequently removed with the use of magnetic beads or microspheres. Such methods

have been shown to be both more effective and introduce less bias than enzymatic treatments such as terminator exonuclease (mRNA-ONLY, Epicentre) (He *et al.*, 2010). Subtractive hybridisation is employed by various commercial kits, including MICROBExpress (Ambion) and Ribo-Zero (Epicentre). MICROBExpress is available as a single kit that claims to remove >95% of 16S and 23S rRNA. Ribo-Zero kits are available for Gram-negative and Gram-positive bacteria, a "meta-bacteria" kit (Bacteria) and also several eukaryotes, including human, yeast, mouse and plant. The Bacteria kits claims to remove 99% of 16S, 23S and 5S rRNA from cultures of *E. coli* and *B. subtilis*. MICROBExpress was for a long time the only kit available for such purposes, and it has been used in several marine studies (Gilbert *et al.*, 2008; Shrestha *et al.*, 2009).

Commercial kits are limited by the sequence diversity of their capture probes. The generation of sample specific capture probes has proved effective in depleting rRNA in ocean samples (Stewart *et al.*, 2010a; Stewart *et al.*, 2012). This involved PCR amplification of the rDNA from environment to be studied. The reverse primer contained a 5' T7 promoter, which allowed subsequent *in vitro* transcription, resulting in a high yield of rRNA probes. Incorporation of biotinylated cytosine and uracil allowed recovery of the probes using streptavidin coated magnetic beads (Stewart *et al.*, 2010a). The advantage of this method is the specificity of the probes to the sample. However, to truly capture all the rRNA in a sample, probes need to be generated for bacteria, archaea and eukaryotes, and even then the primers available for the initial amplification are not universal and will miss some of the diversity. Probe generation is also labour intensive, particularly if multiple sets are required. Full removal of 5S rRNA is often not successful with such methods either and if 5S rRNA probes are generated, the workload is further increased.

Alternatives to subtractive hybridisation include not-so-random priming reverse transcription (Ovation RNA Seq System, NuGen) to bias against rRNA during cDNA synthesis, and enzymatic degradation of rRNA. The mRNA-ONLY kit (Epicentre) employs terminator exonuclease to degrade transcripts without a 5' monophosphate, leaving mRNA intact. However, RNA from environmental samples is often in different states of degradation and while this kit has been used to deplete rRNA in marine metatranscriptomes, it has not been successful when applied to soil (Karunakaran Ramakrishnan, personal communication). It has also been shown to be less effective at removing rRNA and also introduced greater bias than MICROBExpress (He *et al.*, 2010). Several studies have used a combination of both mRNA-ONLY and MICROBExpress (Gifford

et al., 2011; Poretsky *et al.*, 2009), although there is evidence of a synergistic increase in bias introduced at least when MICROBExpress is used after treatment with mRNA-ONLY (He *et al.*, 2010).

Another enzyme capable of depleting rRNA is duplex specific nuclease (DSN), which is used extensively in normalisation of eukaryotic gDNA (Shagina *et al.*, 2010) and cDNA (Zhulidov *et al.*, 2004) libraries. It is capable of degrading any double stranded nucleic acid molecule, i.e. DNA:DNA, DNA:RNA and RNA:RNA. Its efficacy in removal of rRNA has recently been demonstrated (Ciulla *et al.*, 2010; Yi *et al.*, 2011). Depletion of rRNA using DSN involves denaturing a cDNA sample to remove secondary structure, resulting in single stranded molecules. The denatured sample is then maintained at a lower temperature for a specific length of time, after which DSN is added. The highly abundant and self-homologous rRNA derived cDNA molecules re-form their duplexes and become a target for DSN, while medium and low abundant mRNA transcripts are unaffected. DSN has been shown to be more effective at removing rRNA than MICROBExpress and it also introduced less bias (Yi *et al.*, 2011). A disadvantage of many rRNA depletion methods is the requirement for large amounts of input RNA, which are often difficult to obtain from environmental samples. Additionally, one treatment can remove nearly all the RNA in the sample, meaning there may be insufficient left to generate a sequencing library. The DSN protocol described (Yi *et al.*, 2011) overcomes this by generating cDNA from the RNA, using conserved tails which can then be used to amplify the depleted cDNA resulting in large quantities of mRNA enriched cDNA which can be used to generate sequencing libraries directly.

Overcoming low yields of RNA from environmental samples would allow multiple rounds of enrichments and ensure there is sufficient remaining after treatment for validation, quantification and downstream processing. Non-biased amplification of RNA with kits such as SENSATION (Genisphere, Hatfield, PA, USA), have been used successfully in microarray analysis (Poole lab, microarray database). Amplification may also be useful in metatranscriptomics to generate large amounts of RNA from low amounts of precious starting material, or after an mRNA enrichment step prior to sequencing.

It is important to validate the success of any mRNA enrichment before proceeding to sequencing. The most accurate way to obtain the proportion of mRNA in a sample is to sequence it, but this is not always practical due to time and financial constraints. Capillary

electrophoresis is employed by bioanalysers which are typically used to determine rRNA depletion based on the reduction or loss of the dominant peaks representing 16S and 23S rRNA. However, even high sensitivity assays performed on such instruments to do not accurately determine enrichment levels. Quantitative PCR and qRT-PCR can also be used to assess relative abundances of rRNA in a sample both before and after a treatment. The amount of template RNA has to be the same for treated and untreated samples, so accurate quantification, with an RNA specific fluorescent dye for example, is required.

The vast majority of metatranscriptomic studies to date have focused on the marine environment (Gifford *et al.*, 2011; Gilbert *et al.*, 2008; Mason *et al.*, 2012; McCarren *et al.*, 2010; Ottesen *et al.*, 2011; Poretsky *et al.*, 2009; Shi *et al.*, 2010; Shi *et al.*, 2009b; Stewart *et al.*, 2012), where microbial diversity, density and activity is low compared with that in soil, this typically results in lower proportions (<90%) of rRNA in marine metatranscriptomes (Frias-Lopez *et al.*, 2008; Poretsky *et al.*, 2009; Stewart *et al.*, 2010a). Though more recently other environments have been studied with this approach, including deep-sea hydrothermal vents (Lesniewski *et al.*, 2010; Lesniewski *et al.*, 2012), freshwater lakes (Vila-Costa *et al.*, 2013), and the guts of humans (Gosalbes *et al.*, 2011; Ponten, 2011; Ursell and Knight, 2013), mice (Xiong *et al.*, 2012), termites (Raychoudhury *et al.*, 2011; Tartar *et al.*, 2009) and nematodes (Bomar *et al.*, 2011). The metatranscriptomes of complex terrestrial environments such as soils and plant rhizospheres studied to date have been limited to eukaryotes (Bailly *et al.*, 2007; Damon *et al.*, 2012; de Menezes *et al.*, 2012; Takasaki *et al.*, 2013). This is largely due to straightforward enrichment of mRNA, taking advantage of the poly-A of eukaryotic mRNA transcripts. An additional challenge presented by the soil environment is the presence of humic acids, breakdown products of lignin, which co-purify with nucleic acids and are inhibitory to many enzymes used in molecular biology (Wang *et al.*, 2012b).

A current limitation in metatranscriptomic studies has been that few studies have included biological replication or comparisons between different environments. Some studies have compared day and night metatranscriptomes in marine (Poretsky *et al.*, 2009) and lake communities (Vila-Costa *et al.*, 2013), while others have compared changes in transcriptomes due to perturbations (de Menezes *et al.*, 2012; Ursell and Knight, 2013). The temporal dynamics of marine metatranscriptomes have been assessed using automated collection and preservation equipment (Ottesen *et al.*, 2011). Counting the number of sequencing reads that match a particular taxonomic group, or hit a gene in a metabolic

pathway of two different environments can provide only relative comparisons, because sequencing depth is unknown. The addition of an internal RNA standard allows the determination of sequencing depth and absolute transcript abundance (Gifford *et al.*, 2011; Moran *et al.*, 2013). The adoption of this protocol, further recent improvements in mRNA enrichment (Ciulla *et al.*, 2010), and the vast amount of sequence provided by Illumina's HiSeq platform now makes it possible to statistically and quantitatively compare metatranscriptomes from multiple complex environments.

1.3 The phyllosphere environment

The phyllosphere, or aerial surface of a plant, is considered relatively nutrient poor compared to the rhizosphere. Microbial colonisation of leaves is not homogenous, but is affected by leaf structures such as veins, hairs, and stomata. Leaf surfaces are colonised by up to 10^7 microbes per cm^2 (Lindow and Brandl, 2003). The phyllosphere is a much more dynamic environment than the rhizosphere, with resident microbes subjected to large fluxes in temperature, moisture and radiation throughout the day and night. These abiotic factors also indirectly affect the phyllosphere microbiome through changes in plant metabolism. Precipitation and wind in particularly are thought to contribute to the temporal variability in resident phyllosphere microbes (Lindow, 1996). Interestingly, leaf metabolite profiles of *A. thaliana* have been altered by application of soil microbes to roots. Increased concentration of several amino acids in the leaf metabolome were correlated with increased herbivory by insects (Badri *et al.*, 2013b), suggesting cross-talk between above and below ground parts of the plant.

Bacterial and fungal communities in the phyllospheres of various plants have been profiled using PCR amplification of rRNA genes. Microbial richness appears to be greater in warmer, more humid, climates than in temperate ones. Proteobacteria are consistently the dominant bacterial phylum (namely Alpha and Gamma classes), with Bacteroidetes and Actinobacteria also commonly found (Bodenhausen *et al.*, 2013; Vorholt, 2012). The phyllospheres of several plants in the Mediterranean were found to be dominated by lactic acid bacteria (Firmicutes) during summer. Their mode of metabolism was proposed to allow them to tolerate the hot and dry weather conditions (Vokou *et al.*, 2012), although this was not compared to other seasons. At high microbial taxonomic levels, phyllosphere

microbiomes of different plants can appear similar, but at the level of microbial species and strains stark difference are apparent, reflective of the finely tuned metabolic adaptations required to live in such an environment (Vorholt, 2012). While rhizosphere microbiomes are comparable to soil, little similarity has been found between phyllosphere microbiomes and those of air (Vokou *et al.*, 2012).

Proteogenomic analyses of various phyllosphere microbiomes, including those of wild *A. thaliana*, rice, clover and soybean, have revealed species that assimilate plant derived ammonium, amino acids and simple carbohydrates, implicating these compounds as primary nitrogen and carbon sources in the phyllosphere. Expression of microbial stress response proteins, porins, components of ABC transporters, and TonB-dependent receptors, particularly those from *Sphingomonas spp*., was high (Delmotte *et al.*, 2009; Knief *et al.*, 2012), indicating a nutrient poor environment. These studies also determined that *Methylobacterium spp*, and other methylotrophs were widely abundant phyllosphere microbes, and that they were actively assimilating and metabolising methanol, derived from plant pectin (Galbally and Kirstine, 2002). Metagenomic analysis of taxonomically diverse plant species has identified an abundance of various known and novel microbial rhodopsins present in the phyllosphere. These light-sensing proteins, and proton pumps showed non-overlapping absorption spectra with their host plant (Atamna-Ismaeel *et al.*, 2012), indicating that energy metabolism in the phyllosphere is not entirely dependent on the plant.

1.4 The rhizosphere environment

The rhizosphere is the region of soil influenced by plant roots through rhizodeposition of exudates and mucilage. Root exudates have been implicated as key determinants of rhizosphere microbiome structure (Badri *et al.*, 2013a; Bais *et al.*, 2006; Broeckling *et al.*, 2008; Shi *et al.*, 2011). Root exudate compositions in *Arabidopsis thaliana* have shown variation across different accessions resulting in correspondingly different rhizosphere bacterial communities (Micallef *et al.*, 2009). The mutation of an ABC transporter in one accession induced changes both root exudate composition and the rhizosphere bacterial communities (Micallef *et al.*, 2009).

Root exudates contain a variety of compounds, predominately organic acids and sugars, but also amino acids, fatty acids, vitamins, growth factors, hormones and antimicrobial

compounds (Bertin *et al.*, 2003). The composition of root exudates varies spatially and temporally with a number of biotic and abiotic factors. These include plant species and cultivar (Mark *et al.*, 2005; Micallef *et al.*, 2009), as well as plant age and developmental stage (Cavaglieri *et al.*, 2009; Chaparro *et al.*, 2013; Houlden *et al.*, 2008). In the potato rhizosphere, bacterial microbiomes were shown to be different at three developmental stages (young leaf development, florescence and senescence), however, the numerous potato cultivars tested only showed differences in microbiota structure at the early stage of development (Inceoglu *et al.*, 2011). This may be due to carryover of microbes from the tubers, or initial bursts of different root exudates which then stabilize as the plant ages. At locations along the roots of wild oat (*Avena fatua*), 8% of bacterial taxa were found to be enriched in root zones compared to soil, and higher numbers of live cells were isolated from growing root tips and hairs compared to mature root zone (DeAngelis *et al.*, 2009). Usually, attempts are made to sample the entire rhizosphere, but microbes enriched specifically at different root zones may be diluted by this approach, giving the overall impression that they are weakly or not at all enriched. This is an important consideration when sampling rhizosphere soil.

Plants grown anexically have markedly different exudate compositions from those influenced by microbes. Metabolomic analysis of the root exudates from anexically grown pea showed levels of sugars and sugar alcohols (Poole lab, unpublished data), while the microarray analysis *of Rhizobium leguminosarum* during colonisation of the pea rhizosphere revealed up-regulation of genes required for the transport and metabolism of organic acids, particularly aromatic amino acid, as well as C1 and C2 compounds (Ramachandran *et al.*, 2011). Furthermore, the rhizospheres of pea, alfalfa and sugar beet all induced gluconeogenesis, which is repressed by the presence of sugars *in R. leguminosarum* (Ramachandran *et al.*, 2011). These observations, and the fact that plant nutritional status determines how much carbon is allocated to roots (Dakora and Phillips, 2002) make the extraction of physiologically relevant root exudates a particular challenge.

In attempts to recreate rhizosphere effects, the addition of carbon sources such as glucose, glycine, and citrate to different soils has resulted in enrichment of Beta- and Gammaproteobacteria, as well as Actinobacteria (Eilers *et al.*, 2010). These are often also enriched in rhizospheres compared to bulk soils. Although root exudates contain a variety of carbon sources, enrichment of these taxa using a single carbon source suggests some rhizosphere colonisers may be opportunistic fast growers. However the low taxonomic

resolution of the study prevented the genera or species responding to the carbon sources to be identified. These fine taxonomic levels are where real differences in metabolic capabilities thought to be required for rhizosphere colonisation would occur.

Although important, exudates are not the only component of rhizodepostion, and there is evidence to suggest they may be only important at growing root tips (Dennis *et al.*, 2010). The sloughing of root cells and the release of mucilage deposits a large amount of material into the rhizosphere, including plant cell wall polymers such as cellulose and pectin. Cellulose degradation is a widespread trait among microbial residents of high organic matter soils (Haichar *et al.*, 2007; Stursova *et al.*, 2012). The decomposition of pectin releases methanol (Galbally and Kirstine, 2002) which can be used as a carbon source by microbes. Active metabolism of C1 compounds in the rhizosphere has been observed (Knief *et al.*, 2012; Matilla *et al.*, 2007; Ramachandran *et al.*, 2011).

1.4.1 Molecular determinants of rhizosphere colonisation

The ability to utilise plant derived carbon is not much use if an organism is unable to locate a plant in the soil. Thus it is thought that both chemotaxis and motility are key ability for rhizosphere competent microbes. However this is complicated by the fact that attachment to the plant root surface involves a switch from motile to sedentary lifestyle. Genes involved in chemotaxis, flagellar assembly and function were up-regulated in *P. putida* in the maize rhizosphere, but were down-regulated in *R. leguminosarum* in the rhizospheres of pea, alfalfa and sugarbeet (Ramachandran *et al.*, 2011). Exposure of *Pseudomonas aeruginosa* to sugar beet root exudates also down-regulated motility-related genes (Mark *et al.*, 2005). The plant pathogenic *Ralstonia solanacearum* responds chemotactatically to tomato root exudates of host plants, particularly the organic and amino acid components. The loss of either one of two key regulators of chemotaxis, *cheW* or *cheA*, resulted in strains with wild-type motility but reduced virulence. However, they were able to cause disease when directly inoculated into the plant stem (Yao and Allen, 2006). The plant growth promoting rhizobacteria (PGPR) *Pseudomonas fluorecscens* WCS365 also required *cheA* for chemotaxis in the tomato rhizosphere, where it responds to malate and citrate. Mutants of *cheA* showed reduced competition in rhizosphere colonization (de Weert *et al.*, 2002). Another species, *Pseudomonas putida*, is attracted to the maize rhizosphere by benzoxazinoids (Neal *et al.*, 2012). It might be particularly useful to compare the expression

of chemotaxis related genes in large rhizosphere metatranscriptomic data sets with what is currently known from model systems. This might allow determination of what plant derived chemical signals are attracting different groups of microbes to the rhizosphere, and how they are distributed across different plants. For example, up-regulation of a methyl-accepting chemotaxis protein for serine in a rhizosphere compared soil might indicate serine as compound in the root exudates. This might then allow programming of plants to produce chemicals known to be chemoattractants for PGPRs, or stop producing those that attract pathogens. Alternatively, PGPR strains could be genetically engineered to respond to a molecule produced by their host plant.

Direct contact with the plant roots at the rhizoplane could be considered optimum for the acquisition of plant derived carbon, and it is a prerequisite for the colonisation of internal tissues by endophytes. It could also be considered the competitive goal of all rhizosphere colonising microbes, although the effects of the plant defence response would likely be felt more strongly at the rhizoplane, adding further selection pressure. A significant overlap (around 40% of operational taxonomic units (OTUs)) was seen in those bacteria attaching to a root and to an inert wooden structure (Bulgarelli *et al.*, 2012), suggesting that the rhizosphere effect is in part due to transitions to a sedentary lifestyle. Plant cell walls contain proteoglycans such as arabinogalactan proteins (AGPs), which are thought to be important for bacterial attachment and biofilm formation. For example, an AGP from pea root exudates can induce biofilm formation in *R. leguminosarum* (Xie *et al.*, 2012), and a mutant of *Arabidopsis thaliana* (*rat1*) deficient in production of a lysine-rich AGP is resistant to transformation by *Agrobacterium tumefaciens* (Gaspar *et al.*, 2004). These two bacteria are closely related (family Rhizobiaceae) and form intimate associations with host plants. It is not yet known if and how important AGPs and other proteoglycans are for attachment of other root associated microbes.

1.4.2 The arbuscular mycorrhizal fungi

The majority of rhizosphere microbiome studies have focused on bacteria (Bulgarelli *et al.*, 2013). This neglects key members of the rhizosphere microbiome, the arbuscular mycorrhizal fungi (AMF, Phylum Glomeromycota). They are a testament to the importance of plant-microbe interactions, as it is thought that association of green algae with ancient fungal lineages was fundamental to the evolution of land plants c700 million years ago

(Heckman *et al.*, 2001). Today, most plants, though notably not *Arabidopsis thaliana* and other Brassicas (Smith and Smith, 2011), maintain this association as mycorrhizal symbiosis. They form an extensive network of hyphae that dramatically increases the surface area of the below ground parts of a plant, allowing increased nutrient acquisition, particularly for phosphorous. Mycorrhizal hyphae also maintain their own microbiomes at the soil interface and stimulate the decomposition of organic matter (Bonfante, 2010). This raises some important questions. Do plants that form mycorrhizal symbioses have more similar rhizosphere microbiomes than those that don't? Or are there particular groups either enriched or depleted by the presence of mycorriza?

There are a number of ways in which mycorrhiza can influence the rhizosphere microbiome directly and also indirectly via changes in the host plant. Hyphal exudates containing organic acids and sugars can act as carbon sources for microbial growth (Toljander *et al.*, 2007). Additionally, the protein Glomalin is known to provide a significant amount of nitrogen for soil microbes (Wright and Upadhyaya, 1996). Infection by mycorrhiza triggers broad transcriptional changes in roots, including activation of the plant defence system (Wang *et al.*, 2012a), which will affect other microbes. The improved nutritional status of the plant conferred by the AMF will have an impact on plant root exudate compositions, root mass and structure, increasing the niche available to rhizosphere colonising microbes. Few studies have shown changes in relative abundance of some bacterial groups when a plant is colonised by AMF compared to when it is not (Nuccio *et al.*, 2013). In addition, inoculation of leek (*Allium porrum*) with the AMF (*Glomus intraradices*) resulted in increased translocation to the shoot and survivability of two food-borne human pathogenic bacteria in the plant (Gurtler *et al.*, 2013), though the experiment was carried out in the absence of other bacteria. Thus there is little information available on whether there are absolute changes in total abundance of bacteria or indeed other fungi and eukaryotes when a plant forms mycorrhizal symbioses.

1.4.3 Nutrient cycling

Plant associated microbes are key players in global biogeochemical cycles (Philippot *et al.*, 2009). A significant amount, 5% to 20% depending on plant species, age, and nutrient status, of photosynthate is released, mainly through roots (Marschner, 1995), while 100 Tg of methanol and 500 Tg of isoprene are released into the atmosphere by plants annually

(Galbally and Kirstine, 2002; Wang and Shallcross, 2000). For methanol this corresponds to between 0.016% and 0.14% of photosynthate depending on plant type (Galbally and Kirstine, 2002). Both methanol and isoprene are potential sources of both carbon and energy for microbes. In agricultural soils in particular, plants stimulate microbial methanogenesis which contribute to emissions of methane (Conrad *et al.*, 2006). This represents a loss of carbon from the system and contributes to the greenhouse effect. It also stimulates the growth of other microbes, methanotrophs, which mitigate some of the methane emissions from soil (Holmes *et al.*, 1999).

After carbon dioxide and water, nitrogen and phosphorous are considered two of the most important nutrients limiting plant growth, and represent the main constituents of artificial fertilisers. Microbes that make these nutrients more bioavailable to plants have gained significant interest. Diazotrophic bacteria are considered a major group of PGPRs due to their ability to fix nitrogen from the atmosphere (N_2) into a form usable by plants (NH_4^+). They include both free-living and symbiotic organism such as *Azospirillum* and the Rhizobiales respectively. Isolation of diazotrophs is performed on N-free media to encourage the growth of bacteria with such a capability. Strains of *Azoarcus* and *Klebsiella* lacking a functional nitrogenase do not show the same plant growth promotion as that of their respective wild-types (Hurek *et al.*, 2002; Iniguez *et al.*, 2004). However, it is speculated that the main contribution of diazotrophs to plant productivity is from endophytes, and that non-endophytic rhizosphere colonisers are simply responding to nitrogen limitation in the soil. Ammonium produced by nitrogen fixation is highly soluble and thus easily taken up by roots, but bacteria can rapidly incorporate it into amino acids via glutamine synthetase or oxidise it to hydroxylamine via ammonium monooxygenases. This has lead to the idea that rhizosphere diazotrophs contribute fixed nitrogen to the plant indirectly, after cell death, through mineralisation.

Another widely available form of nitrogen, particularly in agricultural soils, is nitrate (NO_3^-), which can be used as an alternative electron acceptor to oxygen to oxidize compound such as methanol (Kalyuhznaya *et al.*, 2009). With plants providing a carbon source, significant microbial respiration could take place in the rhizosphere with nitrate as an electron acceptor. This results in denitrification, leading initially to the production of nitrite (NO_2^-), then nitrous oxide (N_2O), nitric oxide (NO), or even back to ammonia or N_2. Nitrous oxide is fairly inert and escapes into the atmosphere where it contributes to the greenhouse effect (Wrage *et al.*, 2001). Nitric oxide however is a highly reactive radical and also a signalling

molecule in both plants and animals (Wendehenne *et al.*, 2001). It has been demonstrated to induce auxin responses leading to increase root proliferation. Due to its toxicity, its production by bacteria must be tightly regulated, and it is detoxified by nitric oxide reductases (Tucker *et al.*, 2010). It may thus act as an antimicrobial in the rhizosphere, targeting those organisms unable to detoxify it.

Nitrogen cycling within the rhizosphere is more complex than exchange between root and bacteria. Mycorrhizal fungi have been shown to transfer nitrogen to both plant roots (Hodge and Fitter, 2010) and rhizosphere bacteria (Nuccio *et al.*, 2013). Another fungus, *Metarhizium*, transfers nitrogen to plants that it obtains from parasitising insects, while receiving carbon in return from the plant (Behie *et al.*, 2012; Fang and St Leger, 2010). Plants have shown uptake preference for amino acid homodi- and trimers compared to the respective monomers (Farrell *et al.*, 2013), and they have even been observed taking up whole microbial cells (Paungfoo-Lonhienne *et al.*, 2010), though the idea of widespread mixotrophy in plants is not widely accepted.

Phosphorous is another key element often limiting plant growth, due mainly to the insolubility and thus poor biovailablilty of over 95% of soil phosphorous, the majority which is in the form of phytic acid (inositol hexakisphosphate).Plants, AMF and rhizobacteria can secrete organic acids such as acetate, succinate, citrate and gluconate, which reduce the pH of the rhizosphere increasing solubility of phosphate minerals (Rodriguez and Fraga, 1999). Secretion of phosphatases and phytases liberates orthophosphate (PO_4^{-3}) which is readily taken up by phosphate transporters in plant roots (Rodriguez *et al.*, 2006). Some transporters have been shown to be specific to cells harbouring mycorrhiza (Rausch *et al.*, 2001), further reinforcing the importance of AMF in plant phosphorous acquisition.

Other important nutrients that can limit plant growth include, sulphur, boron and silicon, as well as many metals, particularly sodium, potassium, calcium, iron, magnesium, manganese, copper and zinc (Marschner, 1995). In high concentrations, nutrients and other elements, particularly heavy metals, can become toxic for the plant. The microbiome is thought to play a role in both acquiring important nutrients (George *et al.*, 1994; Lemanceau *et al.*, 2009) while mitigating the effects when levels become toxic (Burd *et al.*, 2000; Farinati *et al.*, 2011; Yang *et al.*, 2013).

1.4.4 Disease suppression

Many PGPRs are antagonistic towards plant pathogens through production of antibiotics, though the use of type III secretion systems to secret effectors that interfere with virulence has also been documented (Rezzonico *et al.*, 2005). Actinomycetes in particularly are known to produce a wide array of compounds with antibacterial, antifungal, antiviral, nematicidal and insecticidal properties. They are often found as one of the most abundant bacterial classes in soil and rhizospheres, and are notably enriched in endophytic communities (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Sessitsch *et al.*, 2002). Other disease antagonists include *Pseudomonas fluorescens* which produces the antifungal compound diacetylphloroglucinol (DAPG). *Pseudomonas spp.* producing DAPG have also been shown to modulate transcription of another PGPR, *Azospirillum brasilense*, increasing expression of genes involved in wheat root colonisation and growth promotion (Combes-Meynet *et al.*, 2011). DAPG also affects other microbiota, including nematodes where it was found to be toxic to some species, while stimulatory to others (Meyer *et al.*, 2009). The presence of DAPG producing pseudomonads in soils has been implicated in the phenomenon of take-all decline, whereby disease severity of take-all reduces with time and the soil becomes suppressive (Raaijmakers and Weller, 1998). Other pseudomonads, producing lipopeptides, hydrogen cyanide, phenazines and other bioactive compounds, contribute to soils suppressive to other disease (Haas and Keel, 2003). Shifts in the microbiome have also been associated with soils suppressive towards *Fusarium* (Klein *et al.*, 2013), *Rhizoctonia* (Mendes *et al.*, 2011), and *Streptomyces scabies* (Rosenzweig *et al.*, 2012). This suggests a consortium of microbes may contribute to suppressiveness, though cause and effect are often not distinguishable. A rich and diverse microbiota may alone be sufficient to prevent infection, by limiting availability of space and nutrients. This is observed in mammalian systems, where antibiotic treatment can subsequently increase susceptibility to infections (Croswell *et al.*, 2009).

1.5 Plant host factors determining microbiome structure and function

1.5.1 Antimicrobials

Plants produce a wide variety of antimicrobial compounds which play a role in protecting them from disease causing organisms including viruses, bacteria, fungi and oomycetes, as well as from herbivory by insects and other animals. Plant antimicrobials that are

preformed are known as phytoanticipins (Broekaert *et al.*, 1995), while those that are synthesised in response to pathogens are termed phytoalexins (Darvill and Albersheim, 1984). There is huge chemical diversity among plant antimicrobials, from the simplest reactive oxygen species (ROS) to the complex glycosylated saponins. Other classes include phenolics, terpenoids and alkaloids, which are widespread in the plant kingdom, while others are restricted to particular groups such as glucosinolates from Brassicas (Bednarek, 2012; Bednarek and Osbourn, 2009).

Antimicrobials released from plant roots are, depending on their mechanism of action, thought be a key determinant of rhizosphere microbiome structure. One way in which this can be demonstrated experimentally is by artificially applying antimicrobials or comparing wild-type plants with those genetically modified in a way that affects the production or release of a particular antimicrobial compound. Any changes in the microbiomes are assumed to be due to the presence or absence of the antimicrobial. Although the reality is much more complex, such studies can provide insight into the broader role of plant defences in shaping the microbiome. Cucumber (*Cucumis sativus*) produces the autotoxin p-coumaric acid, which when added to the rhizosphere, altered the bacterial community and increased the population of a pathogenic fungus (Zhou and Wu, 2012). Maize genetically modified to constitutively express the insecticidal *Bt* toxin showed a different bacterial community and a reduced mycorrhizal population to that of the wild type (Castaldini *et al.*, 2005). It is not known whether this is an indirect effect of changes in insect populations or whether *Bt* toxin directly affects bacteria. As *Bt* toxin is protein, it is not unlikely that some bacteria can degrade it for use as a carbon source. More recently however, two other studies showed no differences in rhizosphere microbial communities between wild-type and *Bt* toxin producing maize (Cotta *et al.*, 2013; Dohrmann *et al.*, 2013). This highlights that different experimental methods can produce different results when used to study a similar system. *Arabidopsis thaliana* produces glucosinolates naturally, but genetic modification resulting in production of an exogenous glucosinolate usually produced by white mustard (*Sinapsis alba*) altered the rhizosphere microbiome (Bressan *et al.*, 2009). This highlights the different activity spectrum of antimicrobials even within the same class, presumably with the same mechanism of action. Additionally, the bacterial and fungal rhizosphere microbiomes of an *Arabidopsis* mutant deficient in aliphatic glucosinolate production were different to that of the wild-type (Tkacz *et al.*, 2013a). Other compounds produced by a broad range of plants that have *in vitro* antimicrobial activity are the methyl halides, methyl chloride, methyl bromide and methyl

iodide (Rhew *et al.*, 2003). One proposal for their production by plants has been to protect from diseases. However, no difference was shown between the bacterial rhizosphere communities of a wild-type *Arabidopsis*, a methyl halide deficient mutant or a methyl halide over-expressing line (Andrzej Tkatcz, unpublished data).

A major class of plant antimicrobial compounds are the saponins. Structurally they consist of an aglycone sapogenin (either a steroid or a triterpenoid) which is glycosidically linked to sugar moieties. The sugar moieties can vary considerably, but often include glucose, galactose, xylose and glucoronic acid. Diversity of aglycone structure and sugar moieties has resulted in the huge array of biological activities of the saponins. In plants, they are thought to play a role in defence, and have shown toxicity against a number of potential plant pathogens (Vincken *et al.*, 2007). Pea (*Pisum sativum*), for example, produces a triterpenoid saponin that specifically inhibits diguanylate cyclase (Ohana *et al.*, 1998).

Oats (*Avena* spp) produce avenacins, which are triterpenoid saponins with broad antifungal (Carter *et al.*, 1999; Maizel *et al.*, 1964) and anti-oomycete (Deacon and Mitchell, 1985) activity. They are thought to protect oat from root pathogens (Papadopoulou *et al.*, 1999) including *Gaeumannomyces graminis*, the causative agent of take-all (Osbourn et al., 1994; Turner, 1953). Take all is a particularly devastating disease that affects wheat, a non-saponin producing cereal, and cropping systems including oat have been used to reduce disease incidence of subsequent wheat crops (Seymour *et al.*, 2012).

The avenacins are biosynthesised from 2,3 oxidosqualene in a reaction catalysed by β-Amyrin synthase, encoded on the Sad1 genetic locus (Haralampidis *et al.*, 2001). Enzymes involved in subsequent steps include cytochrome P450-dependent monooxygenases (Sad2) (Papadopoulou *et al.*, 1999), acyltransferases and glycosylases (Sad3, 4) (Trojanowska *et al.*, 2000). Potentially a number of sugars can be linked to the aglycone, resulting in various forms of avenacin (e.g. A1, A2, B1 and B2). The most abundant avenacin, with two glucose molecules linked to an arabinose molecule is avenacin A-1. The sugar moieties are required for avenacin to associate with membrane sterols, which is the mechanism of its antifungal activity (Armah *et al.*, 1999).

The ability to degrade avenacins has been documented in fungi, including root colonising endophytes (Carter *et al.*, 1999) and *Gaeumannomyces graminis* var *avenae* which can infect oat (Osbourn *et al.*, 1991). The action of avenacinase enzymes removes the sugar moieties required for activity (Armah *et al.*, 1999) and avenacinase mutants of

Gaeumannomyces graminis are unable to infect oat, but retain pathogenicity to wheat (Bowyer *et al.*, 1995). There is high similarity in amino acid sequence and physicochemical properties between saponin-detoxifying enzymes, but they retain their specificity (Osbourn *et al.*, 1995). The degradation products of some saponins are able to suppress the plant immune system, resulting in a twofold benefit for a pathogen possessing a saponin-detoxifying enzyme (Bouarab *et al.*, 2002).

Mutation of the Sad1 locus resulted in an oat deficient in production of avenacins, which was sensitive to fungal root pathogens (Papadopoulou *et al.*, 1999) and showed differences in isolated root endophytic fungi compared to wild type oat (Carter *et al.*, 1999). The mutant also showed elevated levels of avenacin precursors, squalene and 2,3, oxidosqualene, as well as the sterols Δ-7-campesterol and Δ7-avenasterol in its root tissue (Qin *et al.*, 2010).

Avenacin is primarily accumulated in the root epidermis, but while no export system is known, it has been measured in the oat rhizosphere at concentrations known to inhibit fungi. Here it would be in contact with a vast diversity of soil microbes and has therefore been proposed to play a role in shaping the rhizosphere microbiome of oats.

1.5.2 The plant immune system

The plant immune system has co-evolved with the plant microbiome and thus is thought to play a key role in determining its structure. Plant innate immunity is triggered by exposure to microbes via microbe associated molecular patterns (MAMPS) (Bittel and Robatzek, 2007). These are wide-spread, slowly evolving features of bacteria and other microbes such as bacterial flagellin, peptidoglycan, elongation factor Tu (EF-Tu) and fungal chitin. A component of flagellin, *flg22*, can be used alone to stimulate the plant immune system, specifically via the LRR-receptor kinase FLS2. Similarly, Ef-Tu is recognized by other LRR-kinase called EFR. Interestingly, responses to both of these molecules trigger nearly identical transcriptional responses in the plant (Jones and Dangl, 2006).

Originally studied in plant pathogenic microbes, MAMPs were termed pathogen associated molecular patterns (PAMPs). The plant response to MAMPs, or PAMP triggered immunity (PTI), includes production of reactive oxygen species (ROS), callose deposition leading to strengthening of cell walls, and activation of signalling and defence genes. Pathogens can

affect these responses through secretion of effectors (Dou and Zhou, 2012), which trigger a further response from the plant, known as effector triggered immunity (ETI) (Spoel and Dong, 2012). Systemic acquired resistance (SAR), or priming, is activated by both MAMP recognition and ETI. It is a plant-wide response involving the accumulation of broad-spectrum antimicrobials in healthy tissue, thus limiting the spread of the infection (Ryals *et al.*, 1996). A similar priming response is induced systemic resistance (ISR), which results in similar responses to SAR but is triggered by different stimuli. Plant defence signalling is coordinated by hormones depending on the type of pathogen (Bari and Jones, 2009). Salicylic acid is produced in response to attack by biotrophic pathogens while jasmonate controls responses to insect herbivores and necrotrophic pathogens. Ethylene is another plant hormone produced in response to herbivores and necrotrophic pathogens, environmental and developmental signals. It can also modulate jasmonate and salicylic acid signalling pathways. Microbes trigger responses from the plant immune system and are then subject to its effects. Many member of the microbiome may also have the ability to modulate or suppress the plant immune system via producing or degradation of hormones or manipulation of signalling cascades. The latter usually occur via effector molecules, which are recognised by plant receptors known as NB-LRR proteins because they contain nucleotide binding (NB) and leucine-rich repeat (LRR) domains. Due to the presence of diverse microbes, plants grown in soil are thought to be already primed to elicity a response against pathogens. But such a broad response is also detrimental to other, potentially beneficial microbes. A number of microbes that are plant associated, such as *Pseudomonas, Xanthomonas, Agrobacterium and Sinorhizobium spp.*, have evolved ways to adapt to this. For example, *P. syringae* effectors mimic or inhibit components of the immune response (Jones and Dangl, 2006). The interactions between a plant's immune system and its microbiome are thus highly complex and dynamic.

The effects of some components of the plant immune system on the plant microbiota have been studied. Mutants of Arabidopsis deficient in systemic acquired resistance (SAR) have shown different rhizosphere bacterial communities compared to wild-type (Hein *et al.*, 2008), while chemical activation of SAR and ISR did not result in significant shifts in the rhizosphere bacterial community (Doornbos *et al.*, 2011). In the phyllosphere of Arabidopsis, induction of salicylic acid mediated defence reduced diversity of endophytes, while plants deficient in jasmonate mediated defence showed higher epiphytic diversity (Kniskern *et al.*, 2007). These reports suggest that the effects of plant physiological

processes on the microbiome are location dependent and that SAR and ISR are responsible for controlling the populations of some bacteria. Additionally, Arabidopsis mutants of a receptor-like-kinase required for innate immunity (Roux *et al.*, 2011) have shown altered bacterial communities in the rhizosphere compared to wild-type (Tkacz et al., in preparation).

1.5.3 Chemical signalling between plants and their microbiota

Microbial production of plant hormones and hormone analogues is widespread. Production of indole-3-acetic acid (IAA) and other auxins is common among rhizosphere bacteria, particularly the rhizobia (Ghosh *et al.*, 2011). Some *Bacillus* spp. are able to produce gibberellins (Gutierrez-Manero *et al.*, 2001). *Pseudomonas syringae*, produces hormone analogues that interfere with jasmonate and ethylene signalling, resulting in stomatal opening and pathogen entry (Melotto *et al.*, 2006). Degradation of hormones or hormone precursors by bacteria is also documented. For example, microbial deamination of 1-aminocyclopropane 1-carboxylic acid (ACC) prevents plant ethylene signalling, resulting in plants more tolerant to environmental stress (Glick, 2005).

Though some chemical signals released by plants facilitate specific interactions, many are recognised by other organisms. For example, flavonoids trigger diverse responses in rhizobia, mycorrhiza, root pathogens and other plants (Hassan and Mathesius, 2012). Strigolactones induce hyphal branching in mycorrhizal fungi and promote seed germination of parastitic plants (Akiyama and Hayashi, 2006). Some plant genes and pathways play roles in establishment of multiple interactions with different microbes, such as the shared developmental pathways for both mycorrhizal and rhizobial symbioses (Stracke *et al.*, 2002), mycorrhizal symbiosis and infection by oomycetes (Wang *et al.*, 2012a) and rhizobial symbiosis and infection by nematodes (Damiani *et al.*, 2012). Components of these pathways could potentially interact with and be manipulated by other members of the microbiome.

1.6 Perspective

The microbiome can be considered the extended phenotype of its host. Microbiomes associated with above ground (phyllosphere), below ground (rhizosphere) and internal

tissues (endophytes) of the same plant are distinct. In addition, the same niche of different plants can host widely dissimilar microbiomes, particularly when viewed at fine taxonomic levels such as microbial genera, species, and strains. This is where specific metabolic capabilities are required to utilise host derived carbon sources and tolerate host defences. Abiotic conditions, such as temperature, moisture, and pH, have broad affects on the microbiome directly and indirectly though the host. Phyllosphere microbial communities are subject to large fluxes in abiotic conditions, and so rates of microbial turnover are different between areas of the plant. Additionally there is cross-talk between above and below ground plant tissues, which can impact on other external factors such as herbivory by insects. Even small changes in the host can influence the plant microbiome, which feed-back to modulate the behaviour of the host. Despite its complexity and dynamism, particularly in natural environments, it is important not to overlook the importance and potential uses of the plant microbiome. Genetic modification of plants, to resist disease for example, may have unforeseen consequences for the rest of the microbiome, which may or may not be physiologically relevant to the plant. The role of the microbiome and its relationship to plant health, productivity, and biogeochemical cycles should be considered as much as the plant itself. An extension of this notion is that molecular breeding or genetic modification of plants could be used to modulate the microbiome intentionally, recruiting disease antagonists and plant growth promoters to improve agricultural production. Before this can happen, knowledge of the plant microbiome and how it varies with different hosts and host factors needs to be expanded to allow a number of important questions to be answered.

1.6.1 Aims of this project

The aims of this project are to determine:

i. What are the global differences in rhizosphere microbiome structure between different crop plants?
ii. Are there differences in abundance or activity of bacteria, archaea, fungi or other microbes in the different rhizospheres compared to soil?
iii. What are the microbes doing in the rhizosphere, and how is this different from those living in plant-free soil?
iv. What compounds are the microbes sensing and metabolising?

v. How are they dealing with competition and the plant defence responses they have induced?

vi. Are they promoting plant growth, and if so how?

vii. Which taxonomic groups are contributing to cycling of nutrients and key elements?

viii. Are there changes that could be considered general to any rhizosphere, and those which are plant species specific?

To do this, comparative metatranscriptomic will be used. Total RNA will be isolated from soil and the rhizospheres of wheat, oat and pea, grown in the same soil. Sequencing total RNA, the majority of which will be rRNA, will allow robust assessment of the active microbes in these environments. This approach avoids the need to target taxonomic groups and avoids PCR bias. The use of RNA over DNA provides information on activity and also a much higher abundance of rRNA from which to characterise. To study the metabolism of the soil and rhizosphere microbes, the depletion of rRNA from complex environmental samples will be optimised and applied, before samples are sequenced with next-generation sequencing technologies. Resulting data will be analysed bioinformatically to determine homology to known organisms, genes, proteins and metabolic pathways. The abundance, activity and expression of these will be compared between environments. This approach will help to answer the above questions resulting in an advancement of understanding and harnessing the plant microbiome for sustainable agricultural production.

Chapter 2: Materials and methods

2.1 General considerations

Prior to laboratory work, surfaces were cleaned with 70% ethanol, and protective nitrile gloves were worn at all times. Safety goggles were worn when using phenol and work was carried out in a fume hood for as much time as was feasible. All water used was nuclease free, molecular biology grade water unless stated otherwise, i.e. distilled water for plants. When working with RNA, surfaces, utensils and gloves were cleaned with an RNase deactivating product such as RNase Zap (Ambion, Austin, TX, USA) or RNase Erase (MP Biomedicals, Solon, OH, USA), and any water used was RNAase free. Any reagents stated were used directly from respective kits unless otherwise stated and chemically defined. Tubes were centrifuged in either an Eppendorf Minispin Plus (1.5 and 2 ml) or an Eppendorf 5810 (15 and 50 ml) centrifuges (Eppendorf, Hamburg, Germany). Standard PCR reactions were carried out in an MJ Research FTC200 thermocycler (Bio-Rad, Hercules, CA).

2.2 Plant growth and harvesting

2.2.1 Soil sampling

Soil was collected from two agricultural fields. One was an experimental field plot at the John Innes Centre (JIC) Norfolk, UK (52°62’29”N, 1°21’81”E) in March 2009. This had been left fallow over the winter, and no major vegetation was present. The second site was

located next to the “Antirrhinum wall” at Bawburgh Farm, Norfolk, UK (52°63’33”N, 1°18’33”E). Soil was harvested from here in June 2011 and July 2012. The top 10 cm of soil, including the mixed wild grass vegetation, was removed. Soil was harvested down to a depth of 20 cm from several sub-sites within the vicinity and each sub-sample (≈20 kg) was placed in an opaque plastic bag. Collected soil was laid out in a glasshouse for 2-3 days to air dry then passed through a 5 mm^2 sieve into opaque plastic bags. Stones, plant material, insects and other debris were removed. All sieved soil from each collection date was loaded into a cement mixer and mixed thoroughly for 10 minutes. A sample of the sieved, mixed soil was chemically analysed by Macaulay Soils at the James Hutton Institute (Aberdeen, UK), with the remaining soil stored in plastic bags in a glasshouse, away from direct sunlight, for a few days until required for planting. See Appendix table A33 for soil analysis data.

2.2.1 Plant seed sterilisation, germination, and planting

Seeds of spring wheat (*Triticum aestivum* var. Paragon), diploid wild oat (*Avena strigosa* accession S75) and the avenacin deficient oat mutant (*sad1*) were surface sterilised by soaking in 5% sodium hypochlorite solution for 1 minute then rinsed several times in distilled water. Pea (*Pisum sativum* var. Avolar) seeds were surface sterilised by soaking in 95% ethanol for 1 minute, washed once with distilled water then soaked for 5 minutes in 2% sodium hypochlorite solution. Pea seeds were then rinsed several times with distilled water

Surface sterilised cereal seeds were checked by eye for contamination (i.e. visible fungal growth), and seeds free of contamination were placed on single moist filter paper discs in petri dishes using sterile forceps, with up to 10 seeds per disc. Pea seeds required more space and water to germinate, so up to 5 pea seeds were placed on 3 water saturated filter paper discs in petri dishes. Seeds were then covered with aluminium foil and left at room temperature in the dark to germinate for 2-3 days, until roots were visible.

Seedlings grown in JIC soil were planted in 500 ml closed pots and grown for 4 weeks in a glasshouse, with distilled water added as necessary. Plants were harvested and loosely attached soil was discarded by shaking. Stems were removed and the dry roots were vortexed briefly in a 50 ml tube to remove soil closely adhering to the roots, which was considered to be rhizosphere soil. Roots were removed with sterile forceps before

approximately 1 g soil was weighed out and added to a Bead Tube from the RNA PowerSoil Total RNA Isolation kit (Mobio, Carlsbad, CA, USA), and was RNA extracted according to the manufacturer's instructions.

Seedlings grown in Bawburgh soils were planted in 450 ml open pots and grown for 4-6 weeks in a glasshouse (see experimental chapters for specific times). Plastic watering tubes were made by cutting 2 cm off the tip of a 5 ml pipette tip. These were then sterilised by autoclaving and one tube was placed inside each pot. Each pot was then covered with a layer of autoclaved perlite to prevent growth of autotrophs on the soil surface and to reduce moisture loss from the soil. Distilled water was added to the watering tubes as necessary. At harvest, loosely attached soil was removed by shaking and stems were removed. Roots were placed in a 50 ml tube, one plant per tube, before 10 ml distilled water was added to each. Tubes were placed on a Heidolph Multi Reax vortex adaptor set at maximum speed (10) for 10 minutes. Roots were then removed using sterile forceps and tubes were centrifuged at 4000 rpm for 10 minutes. Supernatants were removed and the remaining soil was mixed before ≈5 g (wet weight) soil was added to a Bead Tube from the RNA PowerSoil Total RNA Isolation kit, and RNA extracted according to the manufacturer's instructions. DNA was extracted from the same samples using the RNA PowerSoil DNA Elution Accessory kit (Mobio, Carlsbad, CA, USA), according to manufacturer's instructions.

2.3 Nucleic acid isolation and manipulations

2.3.1 Nucleic acid extraction with PowerSoil

Nucleic acid extractions were carried out using the PowerSoil Total RNA isolation and DNA elution accessory kits (MoBio, Carslbasd, CA, USA) according to manufacturer's instructions as follows. For samples grown in JIC soil, 2.5 ml of Bead Solution was added to a Bead Tube followed by 0.25 ml of Solution SR1, vortexing at each stage to mix. For samples grown in Bawburgh soil, 40 ng (in 1 µl) RNA Internal Standard (RIS) was added to 2.5 ml Bead Solution per sample before 2.5 ml of this mix was added to a Bead Tube followed by 0.25 ml of Solution SR1, vortexing at each stage to mix.

The downstream protocol was the same for all samples. Solution SR2 (0.8 ml) was added and tubes were vortexed at maximum speed (10) on a Heidolph Multi Reax vortex adaptor for 5 minutes. Phenol:chloroform:isoamyl alcohol (3.5 ml, pH 6.5) was added and tubes

were vortexed until the biphasic layer disappeared. Tubes were placed back on the vortex adaptor at maximum speed for 10 minutes then centrifuged at 4000 rpm for 10 minutes at room temperature. The upper aqueous phase of each sample was transferred to a new 15 ml Collection Tube, avoiding the interphase and lower phenol layer. Solution SR3 (1.5 ml) was added to the aqueous phase, vortexed to mix then incubated at 4 °C for 10 minutes. Tubes were centrifuged at 4000 rpm for 10 minutes at room temperature and supernatants transferred to new 15 ml collection tubes. Solution SR4 (5 ml) was added to the collection tube containing the supernatant, vortexed to mix and incubated at -20 °C for 30 minutes. Tubes were centrifuged at 4000 rpm for 30 minutes at room temperature, supernatants were discarded and tubes were inverted on a paper towel for 5 minutes to dry. Solution SR5 was shaken to mix, and 1 ml was added to each of the 15 ml tubes. Pellets were resuspended completely by vortexing, followed by incubating in a water bath at 45 °C for 10 minutes, vortexing again, repeating until pellets were completely resuspended.

One RNA Capture Column was prepared for each sample by placing a Capture Column inside a 15 ml Collection Tube. Solution SR5 (2 ml) was added to the column and allowed to gravity flow through, collecting in the 15 ml tubes. Solution SR5 was allowed to completely flow through before adding the RNA samples onto the columns. Samples were allowed to gravity flow through the columns, collecting the flow through in the tubes. Columns were each washed with 1 ml of Solution SR5 allowing it to gravity flow and collect in the collection tubes. The columns were transferred to new 15 ml tubes and 1 ml Solution SR6 was added to each to elute the RNA. The RNA eluates (≈ 1 ml) were transferred to separate 2.2 ml tubes on ice and 1 ml of Solution SR4 was added. Tubes were inverted several times to mix and incubated at -20 °C for a minimum of 30 minutes, before centrifuging at 13,000 rpm for 15 minutes at room temperature to pellet the RNA. Supernatants were carefully discarded and tubes inverted onto a paper towel for 5 minutes, then placed in a heat block at 37 °C for 5 minutes to dry the pellets. Pellets were resuspended in 100 µl of Solution SR7 and stored at -80 °C.

To elute DNA from the same samples, columns were placed in new 15 ml tubes, and 1 ml Solution SR8 was added to the columns and allowed to gravity flow through to elute the DNA. The DNA samples were transferred to separate 2.2 ml tubes on ice and 1 ml of Solution SR4 was added. Tubes were inverted several times to mix and incubated at -20 °C for a minimum of 30 minutes, then centrifuged at 13,000 rpm for 15 minutes at room temperature to pellet the DNA. Supernatants were discarded and tubes inverted onto a

paper towel for 5 minutes, then placed in a heat block at 37 °C for 5 minutes to dry the pellets, which were then resuspended in 100 µl of Solution SR7 and stored at -20 °C.

2.3.2 Nucleic acid quantification

The yield and quality of RNA samples were determined using an Experion bioanalyser (Bio-Rad, Hercules, CA, USA) according to the manufacturer's instructions as follows. Around 20 minutes prior to use, the RNA StdSens reagent kit was removed from storage at 4 °C and equilibrated to room temperature. A gel mix was prepared by adding 600 µl of RNA gel matrix to a spin filter and centrifuging at 4000 rpm for 10 minutes. Filtered gel (65 µl) was transferred to a 0.5 ml tube. The RNA dye concentrate was vortexed thoroughly and briefly centrifuged, then 1 µl of dye was added to the 65 µl aliquot of filtered gel. This gel-dye mix was vortexed thoroughly and centrifuged at 13,000 rpm for 1 minute.

The RNA samples to be quantified and the RNA ladder from the kit were thawed on ice before 1.5 µl of each were aliquoted into separate 0.5 ml tubes on ice. Samples and ladder were denatured in a heat-block at 70 °C for 2 minutes, returned to ice for 5 minutes then centrifuged briefly. Electrodes were decontaminated by adding 800 µl of Experion electrode cleaner to the electrode cleaning chip, which was then placed in the Experion machine, closing the lid for 2 minutes. This was repeated using 800 µl of water for 5 minutes, then allowing the electrodes to air dry with the lid open for 15 seconds.

A new RNA chip was placed in the chip priming station, the time function was set to "1" and the pressure function to "B". Gel-dye mix (9 µl) was added to the well marked GS (orange band). The lid was closed and plunger pressed. Gel-dye mix (9 µl) was added to the second well marked "GS" (no orange band) and 9 µl of gel was pipetted into the well marked in "G". RNA Loading Buffer (5 µl) was added to the ladder well and in all 12 sample wells. Prepared RNA ladder (1 µl) was pipetted into the ladder well and 1 µl of samples were added to each of the 12 sample wells. If less than 12 samples were analysed, 1 µl of water was added to each unused well. The chip was placed in the Experion vortex adapter and vortexed for 1 minute (default setting), then placed in the Experion machine. A new run was selected (Total RNA prokaryote Std Sens) from Assay menu, and the run initiated.

Some nucleic acid samples were quantified on a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Software was initialised and "Nucleic Acid" selected. The sample pedestal was cleaned gently with a damp Kimwipe (Kimberly-Clark,

Irving, TX, USA), before a 1.5 µl blank sample was loaded (either distilled water or appropriate buffer). The sampling arm was lowered to the sample pedestal and the "Blank" button clicked. Residual blank sample was wiped off with a dry Kimwipe and 1.5 µl of sample were loaded in the same manner, pressing the "Measure" button to read the absorbance. For DNA quantification the concentration provided by the Nanodrop was used directly. For RNA quantification samples were diluted 1 in 50 with TE and the A_{260nm} was multiplied by an extinction coefficient of 40 and the dilution factor (50) to obtain concentration in ng/µl.

Some nucleic acid samples were quantified on a Qubit 2.0 fluorometer (Invitrogen, Paisely, UK) using the appropriate kits (dsDNA and RNA). A dye Working Solution was prepared by mixing 199 µl of buffer with 1 µl of dye for each sample. The Working Solution (190 µl) was aliquoted into two assay tubes for standards, which were made every time a new set of measurements was taken. Standard 1 (10 µl) was added to one tube and 10 µl Standard 2 was added to the other, before both were mixed by vortexing. For each sample, 180 – 190 µl of Working Solution and 1 – 20 µl of sample were aliquoted into assay tubes depending on the estimated concentration, bringing the total volume to 200 µl. Initially 1 µl sample was added, but if the resulting concentration was too low this was increased to a maximum of 20 µl depending on the volume of sample available and the amount required for downstream processed. After vortexing to mix, samples were incubated for 2 minutes at room temperature before being placed in the Qubit 2.0 fluorometer for measurement.

2.3.3 DNase treatment of total RNA

To remove contaminating DNA, total RNA samples were treated with TurboDNase (Ambion, Austin, TX, USA) according to manufacturer's instructions as follows. A 50 µl reaction was made up for each sample with 5 µl 10X TurboDNase buffer, 1 µl TurboDNase enzyme, and the desired amounts of RNA sample and water to 50 µl. These were mixed and incubated at 37 °C for 20 – 30 minutes. DNase inactivation reagent (5 µl) was added and reactions mixed and incubated at room temperature for 2 minutes. Reactions were then centrifuged at 11,000 rpm at 4 °C, and supernatants (≈ 45 µl) were transferred to clean 1.5 ml tubes. The RNEasy MinElute kit (Qiagen, Venlo, NL) was used to clean and concentrate the RNA. Water (100µl) was added to each sample, followed by 350 µl RLT buffer (without β-mercaptoethanol) and 250 µl ethanol (99%), mixing at each stage. The total volume

(≈750µl) was added to RNEasy spin columns and centrifuged at 12,000 rpm for 30 seconds, discarding the supernatents. Buffer RPE (500 µl) was added and tubes centrifuged at 12,000 rpm for 15 seconds. Supernatants were discarded and another RPE wash step performed. After discarding the supernatants, tubes were centrifuged for 2 minutes at 12,000 rpm and columns were then transferred to clean 1.5 ml tubes. The desired volume of water was added and tubes left to stand for 2 minutes before centrifuging at 12,000 rpm for 1 minute. TurboDNase treated RNA was stored at -80°C.

2.3.4 Whole transcriptome amplification with Rubicon

Generation of cDNA was performed using strand displacement whole transcriptome amplification (WTA2, Rubicon Genomics, Ann Arbor, MI, USA) on DNase treated RNA, according to manufacturer's instructions as follows. For the Library Synthesis reactions, 2 µl Library Synthesis Solution was added to at least 25 ng of total RNA and brought to 16.6 µl with water in 0.2 ml PCR tubes. Samples were mixed and incubated in a thermocycler programmed at 70 °C for 5 minutes then 18 °C for at least 2 minutes. Library Synthesis Buffer (2.5 µl), 3.9 µl water and 2 µl Library Synthesis Enzyme were added to the cooled-primed RNA and immediately incubated in a thermocycler using the following cycle parameters: 18 °C for 10 minutes, 25 °C for 10 minutes, 37 °C for 30 minutes, 42 °C for 10 minutes, 70 °C for 20 minutes then 4 °C.

For the amplification reaction, a master mix was prepared by mixing 301 µl water, 37.5 µl Amplification Mix, 7.5 µl WTA dNTP Mix and 3.75 µl Amplification Enzyme. The entire Library Synthesis reaction was added to the 25 µl master mix then divided into five 75 µl reactions in 0.2ml PCR tubes. Reactions were incubated in a thermocycler using the following cycle parameters: 94 °C for 2 minutes followed by 17 cycles of 94 °C for 30 seconds, 70 °C for 5 minutes, then 4 °C. The cDNA product was purified using the Qiaquick PCR clean-up kit (Qiagen, Venlo, NL) and quantified using a Nanodrop1000 spectrophotometer or a Qubit 2.0 fluorometer.

2.3.5 Purification of DNA from PCR and enzymatic reactions

Purification of DNA was performed using the Qiaquick PCR clean-up kit (Qiagen, Venlo, NL). Five volumes of Buffer PB were added to 1 volume of the PCR or enzyme treated samples

and mixed. Qiaquick spin columns were placed in 2 ml collection tubes and samples were applied to the column before centrifuging for 30 seconds at 13,000 rpm. Flow-throughs were discarded and the columns placed back into the same tubes. Buffer PE (0.75 ml) was added to the columns and centrifuged for 30 seconds. Flow-throughs were discarded and columns placed back in the same tubes before centrifuging for an additional 1 minute at 13,000 rpm. Columns were placed in clean 1.5 ml tubes and the desired volumes of Buffer EB (10 mM TrisHCl, pH 8.5) were added to the centre of membrane before incubating at room temperature for 1 minute then centrifuging for 1 minute at 13,000 rpm. DNA eluate was stored at -20 °C.

2.3.6 Amplification of RNA using SENSATION

Amplification of RNA was performed using the SENSATION RNA amplification kit (Genisphere, Hatfield, PA, USA) according to the manufacturer's instructions as follows. The first strand cDNA was synthesised with MMLV reverse transcriptase provided in the kit. The volume of total RNA was adjusted by performing RNA cleanup using the RNEasy MinElute kit, eluting in 9 µl water. On ice, 4 µl RT primer mix was added to the 9 µl total RNA to make a 13 µl RNA-primer mix. This was incubated in a thermocycler at 80 °C for 10 minutes then at 4 °C for 2 minutes. A separate RT master mix reaction was made on ice by mixing 5 µl RT buffer mix and 2 µl RT enzyme mix. This was then added to the 13 µl RNA-primer mix and incubated in a thermocycler at 42 °C for 1 hour then at 25 °C for 2 minutes.

Resulting cDNA was purified with the RNeasy MinElute kit using a modified protocol as follows. The RNeasy MinElute columns were prewashed by applying 500 µl of water to the spin columns, incubating at room temperature for 1 minute then centrifuging at ≥12,000 rpm for 1 minute, discarding the flow-through. Water (20 µl) was applied to the spin columns before incubating at room temperature for 1 minute and centrifuging at ≥12,000 rpm for 1 minute. This step was repeated once more before 350 µl RLT buffer (no β-mercaptoethanol) was added to the 100 µl cDNA sample and mixed well by pipetting up and down. Ethanol (250 µl, 100%) was added and mixed well by pipetting up and down. Each sample was transferred to an RNeasy spin column and centrifuged at ≥12,000 rpm for 15 seconds, discarding the flow-through. Each column was transferred to a new 2 ml collection tube and 500 µl of RPE buffer was added to each spin column then centrifuged at ≥12,000 rpm for 15 seconds, discarding the flow-through. Ethanol (250 µl, 80%) was added

to each spin column and centrifuged at ≥12,000 rpm for 2 minutes discarding the flow-through. Each spin column was placed in a new collection tube and centrifuged with lids open for 5 minutes at ≥12,000 rpm to dry and remove any residual ethanol. Each spin column was placed in a new 1.5 ml tube and 12 µl water was added to the centre of the filter disk in each spin column. Tubes were incubated at room temperature for 2 minutes then centrifuged at ≥12,000 rpm for 1 minute to elute the sample.

For the cDNA promoter synthesis reaction the 12 µl purified cDNA was incubated in a thermocycler at 80 °C for 10 minutes then at 4 °C for 2 minutes before transferring the cDNA to ice. For each reaction, a tailing master mix was prepared in a separate tube on ice by mixing 6 µl tailing buffer mix with 2 µl tailing enzyme mix. On ice, the 8 µl tailing master mix was added to the 12 µl purified cDNA then incubated in a thermocycler at 37 °C for 2 minutes, 80 °C for 10 minutes then 4 °C for 2 minutes. For each reaction, a promoter synthesis master mix was prepared in a separate tube on ice by mixing 4 µl promoter synthesis buffer mix with 1 µl promoter synthesis enzyme mix. This was added to the 20 µl tailed cDNA and incubated in a thermocycler for 25 °C for 30 minutes. The resulting primed cDNA was then used as a template for an *in vitro* transcription reaction. For this, an IVT Master Mix was prepared in a separate tube at room temperature by mixing 16 µl T7 nucleotide mix, 5 µl 10X T7 reaction buffer, and 9 µl T7 enzyme mix. This was added to the 25 µl promoter-modified cDNA then incubated in a thermocycler at 37 °C for 16 - 18 hours. Reactions were stopped by placing samples at –20 °C until ready to proceed with purification.

Purification of the sense RNA from the *in vitro* transcription reactions was performed with the RNeasy MinElute kit as follows. Each sample was brought to 100 µl with water. RLT buffer (350 µl, no β-mercaptoethanol) was added to the 100 µl cDNA sample and mixed well by pipetting up and down. Ethanol (250 µl, 100%) was added and mixed well by pipetting up and down. Each sample was transferred to an RNeasy spin columns and centrifuged at ≥12,000 rpm for 15 seconds, discarding the flow-through. Each column was transferred to a new collection tube and 500 µl of RPE buffer was added to each spin column then centrifuged at ≥12,000 rpm for 15 seconds, discarding the flow-through. Ethanol (500 µl, 80%) was added to each spin column and centrifuged at ≥12,000 rpm for 2 minutes discarding the flow-through. Each spin column was placed in a new collection tube and centrifuged with lids open for 5 minutes at ≥12,000 rpm to dry and remove any residual ethanol. Each spin column was placed in a new 1.5 ml tube and 25 µl water pre-

heated to 50 °C was added to the centre of the filter disk in each spin column. Tubes were incubated at room temperature for 1 minute then centrifuged at ≥12,000 rpm for 1 minute. Elution was repeated by pipetting the eluate back onto each spin column, incubating at room temperature for 2 minutes then centrifuging at ≥12,000 rpm for 1 minute. Final RNA eluate was quantified using a Qubit 2.0 fluorometer, assessed using a high-sensitivity Agilent bioanalyser (Agilent, Santa Clara, CA, USA) at The Genome Analysis Centre (TGAC, Norwich, UK) then stored at -80°C.

2.3.7 Quantitative and reverse transcription quantitative PCR of nucleic acids

Quantitative and reverse transcription quantitative PCR were used to determine the abundance of rRNA genes or transcripts in native and rRNA depleted samples. For bacterial 16S rRNA, primers EB_27 and EB_338R were used, for plant 18S rRNA primers Plant_18S_F and Plant_18S_R primers were used, and for eukaryotic 18S rRNA, primers EUK_1427F and EUK_1616R primers were used. The annealing temperature (T_a) of reactions was 57 °C for bacterial 16S primers, 57°C for plant 18S primers, and 64°C for eukaryotic 18S primers. See 2.6 for primer sequences.

For qPCR of cDNA, 2 ng cDNA was used as a template for rRNA treated samples and 0.2 ng untreated cDNA was used as a template for untreated rRNA samples. This difference was taken into account when estimating abundance of rRNA. Templates were mixed with 12.5 µl 2X iQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 10 pmol each forward and reverse primers and brought to a total volume of 25 µl with water. Samples were placed in clear walled wells of a 96 well plate, distributing samples so as to minimise cross-well fluorescence, then covered with strips of clear plastic PCR tube lids. The plate was placed in a MiniOpticon thermocycler (Bio-Rad, Hercules, CA, USA) and the following program run: 95 °C for 3 minutes, then 35 cycles of 95 °C for 15 seconds, T_a for 15 seconds, 72 °C for 15 seconds, reading at the end of each 72 °C step.

For qPCR of gDNA, 5 ng gDNA was used as a template. Templates were mixed with 12.5 µl 2X iQ SYBR Green Supermix (Bio-Rad, Hercules, CA, USA), 10 pmol each forward and reverse primers and brought to a total volume of 25 µl with water. Samples were placed in white opaque walled wells of a 96 well plate and covered with a clear plastic film. The plate was placed in a CFX96 thermocycler (Bio-Rad, Hercules, CA, USA) and the following

program run: 95 °C for 5 minutes, then 40 cycles of 95 °C for 15 seconds, T_a °C for 30 seconds, 72 °C for 30 seconds, reading at the end of each 72 °C step.

For qRT-PCR, 0.2 ng Ribo-Zero treated RNA and respective untreated samples were used as templates. These were mixed with 12.5 µl 2X QuantiTect SYBR Green mix (Qiagen, Venlo, NL), 0.25 µl QuantiTect RT Mix, 10 pmol each forward and reverse primers, and brought to a total volume of 25 µl with water. Samples were placed in white opaque walled wells of a 96 well plate and covered with a clear plastic film. The plate was placed in a CFX96 thermocycler and the following program run: 50 °C for 15 minutes, 95 °C for 3 minutes, then 40 cycles of 95 °C for 15 seconds, T_a for 30 seconds, 72 °C for 30 seconds, reading at the end of each 72 °C step.

2.3.8 In vitro transcription using MEGAScript

The MEGAScript T7 kit (Ambion, Austin, TX, USA) was used to transcribe RNA *in vitro* according to manufacturer's instructions as follows. Reactions were set up at room temperature with 2 µl ATP, 2 µl GTP, 2 µl CTP, 2 µl UTP, 2 µl 10X buffer, 0.5 µl SUPERase RNAse inhibitor, 2 µl T7 RNA polymerase, 0.1 µg to 1 µg template DNA, and brought to 20 µl with water. Reactions were incubated at 37 °C overnight, before 1 µl TurboDNase was added and mixed then incubated at 37 °C for a further 30 minutes.

In vitro transcribed RNA samples were purified using the MEGAClear kit (Ambion, Austin, TX, USA) according to the manufacturer's instructions as follows. Samples were brought to 100 µL with elution solution and mixed gently. Binding solution concentrate (350 µl) was added and mixed gently by pipetting. Ethanol (250 µl, 100%) was added to the samples and mixed gently by pipetting. Filter cartridges were placed in collection tubes and the RNA mixtures were pipetted onto the filter cartridges then centrifuged for 15 seconds at 13,000 rpm. Flow-throughs were discarded and the collection tubes reused for the washing steps. Wash solution (500 µl) was applied to the columns and centrifuged for 15 seconds at 13,000 rpm. This was repeated with a second 500 µl aliquot of wash solution. After discarding the wash solution, tubes were centrifuged for 1 minute at 13,000 rpm. Filter cartridges were placed into new collection tubes and the desired volume of elution solution was added to the centre of the filter cartridges. Tubes were incubated in a heat block at 65 °C for 10 minutes, then centrifuged for 1 minute at 13,000 rpm. This elution step was repeated and resulting RNA eluate was stored at -80 °C.

2.3.9 Generation of an RNA internal standard (RIS)

Generation of an RNA internal standard (RIS) was modified from (Gifford *et al.*, 2011). A strain of *Escherichia coli* containing the plasmid pGEM-3Z was streaked from a glycerol stock onto a Luria-Bertani (LB) agar plate with 50 µg/ml carbenicillin and grown overnight at 37 °C. Two single colonies were used to inoculate two 10 ml LB liquid cultures, which were incubated shaking (250 rpm) at 37 °C for 6 hours. Cells from 1 ml of culture were pelleted by centrifuging at 4000 rpm for 10 minutes and resuspended in 250 µl resuspension solution from the GeneJET plasmid purification kit (Fermentas, Waltham, MA, USA). Plasmid DNA was isolated according to the manufacturer's instructions as follows. Lysis solution (250 µl, containing RNAse A) was added to the samples which were then mixed thoroughly by inverting the tubes 4 - 6 times until the solutions became viscous and slightly clear. Neutralization solution (350 µl) was added and mixed immediately and thoroughly by inverting the tubes 4 - 6 times or until the neutralized bacterial lysate became cloudy. Tubes were centrifuged at 13,000 rpm for 5 minutes to pellet cell debris and chromosomal DNA. Supernatants (500 µl at a time) were transferred to GeneJET spin columns avoiding the white precipitate, before centrifuging for 1 minute at 13,000 rpm. Flow-throughs were discarded and columns placed back into the collection tubes. Wash Solution (500 µl at a time) was added to the spin column and centrifuged for 1 minute. Flow-throughs were discarded, columns placed back into the collection tubes and the wash procedure was repeated. Flow-throughs were discarded and tubes centrifuged for an additional 1 minute to remove residual Wash Solution. Columns were transferred to new 1.5 ml tubes and 50 µl of elution buffer was added to the centre of columns. Tubes were incubated for 2 minutes at room temperature then centrifuged for 2 minutes at 13,000 rpm to elute the plasmid DNA. The two purified plasmid DNA samples were pooled and quantified using a Nanodrop 1000 spectrophotometer (See 2.3.2). Around 1 µg plasmid DNA (in 12 µl) was mixed with 5 µl 10X Green buffer, 1 µl 50X S-adenosyl methionine (SAM), 1 µl *Acu*I restriction enzyme (Fermentas, Waltham, MA, USA) then brought to 50 µl with water. The reaction was incubated at 37 °C for 4 hours then stopped by increasing the temperature to 65 °C for 20 minutes. Successful digestion reaction was determined by agarose gel electrophoresis. Two *Acu*I sites were located within pGEM-3Z, resulting in two bands of 752 bp and 1991 bp. The 1991 bp fragment was extracted from the gel and purified using the Qiaquick gel extraction kit (Qiagen, Venlo, NL) according to

manufacturer's instructions as follows. Excised gel slices were weighed in a 1.5 ml tube then 3 volumes of buffer QG were added to 1 volume of gel (100 mg ≈ 100 µl). Tubes were incubated at 50 °C for 10 minutes, or until the gel slices had completely dissolved, vortexing every 2 – 3 minutes. One gel volume of isopropanol was added to the samples and mixed. Qiaquick spin columns were placed in 2 ml collection tubes and the samples were applied to the columns before centrifuging for 1 minute at 13,000 rpm. Flow-throughs were discarded and columns placed back in the collection tubes. An additional 0.5 ml of buffer QG was added to the columns and centrifuged for 1 minute to remove traces of agarose. To wash, 0.75 ml buffer PE was added to the columns and incubated at room temperature for 2 – 5 minutes then centrifuged for 1 minute at 13,000 rpm. Flow-throughs were discarded and columns centrifuged for an additional 1 minute at 13,000 rpm. Columns were placed into clean 1.5 ml tubes, 15 µl water was added to the centre of the columns, incubated at room temperature for 1 minute and centrifuged for 1 minute at 13, 000 rpm. This was repeated with another 15 µl water to elute the DNA in a total volume of 30 µl. The resulting DNA was quantified on a Nanodrop 1000 spectrophotometer (See 2.3.2).

Two *in vitro* transcription reactions were set up using the MEGAScript kit (See 2.3.8) with 7 µl template DNA (≈140 ng). Reactions were incubated at 37 °C overnight before adding 1 µl of TurboDNase and incubated for a further 15 minutes at 37 °C. *In vitro* transcribed RNA was purified using the MEGAClear kit eluting in two washes of 50 µl EB. Reactions were combined and quantified on an Experion bioanalyser, revealing a single peak of 967 bp. Concentration was measured using a Qubit 2.0 fluorometer (See 2.3.2). A 1 in 10 dilution of this was made with water as a working solution of RIS with a concentration of 40 ng/µl.

2.4 Ribosomal RNA depletion methods

2.4.1 Depletion of ribosomal RNA using MICROBExpress

MICROBExpress (Ambion, Austin, TX, USA) was used according to manufacturer's instructions as follows. Approximately 500 ng (in 15µl) DNase treated RNA samples were added to 200 µL Binding Buffer in 1.5 ml tubes and vortexed. Capture oligo mix (4 µl) was added before tubes were vortexed and centrifuged briefly. Samples were heated to 70 °C for 10 minutes to denature secondary structures in the rRNA then incubated at 37 °C for 15 minutes to allow capture oligos to hybridise to homologous regions of the 16S and 23S rRNAs.

Oligo MagBeads were vortexed thoroughly and 50 µl per sample were added to the 1.5 ml tubes. Tubes were placed on a magnetic rack for 3 minutes to pellet the MagBeads, supernatants were then discarded. An equal volume of water was added to the tubes and vortexed to resuspend the MagBeads. Tubes were placed on a magnetic stand for 3 minutes to pellet the beads and supernatents were discarded. An equal volume of binding buffer was added to the MagBeads and then vortexed to resuspend. Tubes were placed on a magnetic rack for 3 minutes to pellet the beads and supernantents were discarded. An equal volume of binding buffer was added to the MagBeads, vortexed to resuspend then incubated at 37°C.

Prepared Oligo MagBeads (50 µl) were added to the RNA-capture oligo mix and incubated at 37 °C for 15 minutes. Tubes were placed on a magnetic rack for 3 minutes to pellet the beads. The rRNA depleted supernatants were removed and transferred to 0.5 ml tubes on ice. Wash Solution (100 µl, pre-heated to 37 °C) was added to the beads, then gently vortexed and placed on a magnetic rack for 3 minutes. The remaining supernatants were added to the rRNA depleted samples on ice to a total volume of ≈350 µl.

To precipitate and resuspend the rRNA depleted samples 35 µl sodium acetate (3 M), 6 µl glycogen (5 mg / ml) and 1175 µl ethanol (100%, ice cold) were added before briefly vortexing to mix. Tubes were incubated at –20 °C for at least 1 hour, then centrifuged for 30 minutes at 13,000 rpm. Supernatants were carefully discarded before 750 µL ethanol (70%, ice cold) was added and vortexed briefly, then centrifuged for 5 minutes at 13,000 rpm, discarding the supernatants. This ethanol wash was then repeated. Tubes were centrifuged briefly after the second ethanol wash and remaining supernatants were carefully removed with a pipette, avoiding the pellets. The pellets were air dried for no longer than 5 minutes before being resuspended in 25 µL TE (10 mM Tris-HCl pH 8, 1 mM EDTA) and incubated at room temperature for 15 minutes, vortexed to mix, then centrifuged briefly to collect the samples. The entire volume from this initial treatment was used as template for a second treatment with MICROBExpress, repeating the protocol above. Twice rRNA depleted RNA was quantified on an Experion bioanalyser and stored at -80°C.

2.4.2 Subtractive hybridisation using sample specific, anti-sense RNA capture probes

Generation of sample specific anti-sense RNA capture probes was modified from (Stewart *et al.*, 2010a). For generation of 16S rRNA capture probes, four 50 µl PCR reactions were set up using 25 µl GoTaq Green master mix (Promega, Madison, WI, USA), 50 pmol primer EB_27F, 50 pmol primer EB_1492R_T7 (see 2.6), 50 ng template DNA and brought to 50 µl with water. Cycle parameters were 95 °C for 2 minutes, 35 cycles of 95 °C for 1 minute, 57 °C for 1 minute, 72 °C for 3 minutes, followed by a final extension step of 72 °C for 10 minutes. For the 23S probes, the same reactions were set up using primers EB_189F and EB_2490R_T7 (see 2.6), and an annealing temperature of 40 °C. Amplification of the correct fragment size was determined using agarose gel electrophoresis. Replicates for each set of probes were pooled and cleaned up using the Qiaquick PCR purification kit (Qiagen, Venlo, NL), eluting in 50 µl EB. *In vitro* transcription was carried out for both 16S and 23S probes separately using the MEGAScript kit (see 2.3.8) with 1 µl PCR amplicons (250 to 500 ng), 2 µl ATP, 2 µl GTP, 1.5 µl CTP, 1.5 µl UTP, 3.75 µl biotin-11-CTP (10mM, Roche, Basel, Switzerland) 3.75 µl biotin-16-UTP (10mM, Roche), 2 µl 2X buffer, 0.5 µl SUPERase RNAse inhibitor (Ambion) , and 2 µl T7 RNA polymerase. Reactions were set up at room temperature then incubated at 37 °C overnight. TurboDNase enzyme (1 µl) was added and reactions mixed and incubated at 37 °C for a further 30 minutes. Reactions were purified using the MEGAClear kit, eluting in 50 µl elution solution. Concentrations were determined by diluting 1 in 50 with TE and measuring A_{260nm} using a Nanodrop 1000 spectrophotometer (see 2.3.2).

For the subtractive hybridisation reactions, streptavidin coated magnetic beads (NEB, Ipswich, MA, USA) were pre-treated to remove RNases. Magnetic beads were mixed thoroughly by vortexing then 100 µl beads per sample were transferred to a 1.5 ml tube. The tube was placed on a magnetic rack for 1 minute until the beads had pelleted. The supernatant was removed and beads were resuspended in an equal volume of 0.1 N (0.1 g / l) sodium hydroxide to deactivate RNases. The tube was vortexed briefly then placed on a magnetic rack for 1 minute to pellet the beads. The supernatant was removed and beads were washed twice with 1X saline sodium citrate (SSC) buffer then aliquoted (100 µl per sample) into separate 1.5 ml tubes on ice.

Hybridisation reactions were set up in 0.2 ml PCR tubes with 250 – 500 ng total RNA, 500 – 1000 ng each 16S and 23S RNA probes, to a maximum volume of 36.5 µl. SUPERase RNase inhibitor (1 µl), 2.5 µl 20X SSC and 10 µl formamide were added and volumes brought to 50 µl with water. Reactions were incubated in a thermocycler at 70 °C for 5 minutes, then at

65 °C to 25 °C at 5 °C increments for 1 minute each. Reactions were then incubated at room temperature for 5 minutes. Pre-aliquoted beads were placed on a magnetic rack for 1 minute and supernatants were removed before 50 µl 1X SSC 20% formamide were added to the hybridisation reactions. These were then added to the dry magnetic beads. The bead mixes were incubated at room temperature for 10 minutes with occasionally flicking to mix. Tubes were centrifuged briefly then placed on a magnetic rack for 3 minutes. Resulting rRNA depleted supernatants were transferred to 1.5 ml tubes on ice. The beads were resuspended with 100 µl 1X SSC, and placed on a magnetic rack for 3 minutes. The residual rRNA depleted supernatants were added to the tubes containing the rRNA depleted samples on ice. Treated RNA samples were cleaned using the RNEasy MinElute kit (see 2.3.6), eluting in 15 µl water then quantified using an Experion bioanalyser (see 2.3.2).

2.4.3 Duplex specific nuclease treatment

Duplex specific nuclease (DSN) enzyme (Evrogen, Moscow, Russia) was used according to a protocol modified from (Yi *et al.*, 2011). Template cDNA was generated using the whole transcriptome amplification kit (Rubicon) according to the manufacturer's instructions (see 2.3.4). A 4X hybridization buffer was prepared by mixing 200 µl 1 M HEPES buffer solution, 400 µl 5 M sodium choride solution and 400 µl water per sample (total 1000 µl). A DSN buffer was prepared by diluting the 10X DSN master buffer supplied with the DSN enzyme to 2X concentration with water. The DSN enzyme was prepared by adding DSN storage buffer to the lyophilized DSN enzyme (5 µl buffer per 10 units DSN). Contents were mixed by gently flicking the tube, before centrifuging briefly. The tube was incubated at room temperature for 5 minutes before an equal volume of 100% glycerol (to 50% final glycerol concentration) was added and contents mixed by gently flicking the tube, centrifuging briefly and storing at -20 °C.

For the DSN treatment, reactions were prepared in 200 µl PCR tubes on ice for each sample with 13.5 µl template cDNA (250 – 300 ng) and 4.5 µl 4X hybridization buffer (total 18 µl). Reactions were pipetted up and down 10 times to mix, and centrifuged briefly, before being transferred directly to the bottom of new 200 µl PCR tubes using a pipette, then incubated in a thermocycler for 2 minutes at 98 °C followed by 5 hours at 68 °C. Following incubation, the thermocycler lid was kept closed and the temperature held at 68 °C. The

reactions were not removed from the thermocycler prior to or during DSN treatment. Pre-heated 2X DSN buffer (20 µl, 68 °C) was quickly added to the first reaction tube. With the reaction mix tube remaining in the thermocycler, the entire volume was pipetted up and down 10 times to mix thoroughly using a pipette set to 40 µl. The thermocycler lid was immediately closed afterwards. This was repeated for each sample, keeping the lid closed between each addition. Reactions were incubated in the thermocycler at 68 °C for 10 minutes before quickly adding 2 µl DSN enzyme to the first reaction tube. With the reaction tube remaining in the thermocycler, the entire volume was pipetted up and down 10 times to mix thoroughly using a pipette set to 40 µl. The thermocycler lid was immediately closed afterwards. This was repeated for each sample before reactions were incubated in the thermocycler at 68 °C for 25 minutes. The DSN Stop Solution (40 µl, 2X) was added to each reaction, gently pipetting up and down to mix thoroughly before tubes were placed on ice. The DSN reactions were cleaned using Qiaquick PCR purification kit (see 2.3.5), eluting in 11.5 µl EB. A 1.5 µl aliquot was quantified on a Nanodrop 1000 spectrophotometer (see 2.3.2) and the remaining 10 µl re-amplified in a PCR reaction with 20 pmol Rubicon primer (see 2.6), 11 µl GoTaq Green and broought to 25 µl with water. Cycle parameters were as follows 95 °C for 2 minutes, 35 cycles of 95 °C for 1 minute, 57 °C for 1 minute, 72 °C for 3 minutes, followed by 72 °C for 10 minutes. Reactions were cleaned using the Qiaquick PCR purification kit (see 2.3.5), eluting in 100 µl EB. The resulting rRNA depleted cDNA was stored at -20 °C.

2.4.4 Ribosomal RNA depletion with Ribo-Zero Metabacteria (non-magnetic)

The Ribo-Zero Metabacteria (non-magnetic) kit (Epicentre, Madison, WI, USA) was used according to manufacturer's instructions as follows. Microspheres were vigorously mixed at room temperature for 20 seconds by vortexing to produce a homogeneous suspension. For each reaction, 65 µl of microspheres were transferred to separate 2 ml Microsphere Wash Tubes. Tubes were centrifuged at 13,000 rpm for 3 minutes and supernatants were discarded. Microspheres were washed by adding 130 µl of Microsphere Wash Solution to each tube and vortexing o resuspend the microspheres. Tubes were centrifuged at 13,000 rpm for 3 minutes and supernatants were carefully discarded using a pipette. Microsphere Resuspension Solution (65 µl) was added to the tubes before vortexing at maximum speed until homogeneous suspensions were produced. RiboGuard RNase inhibitor (1 µl) was

added to each tube of resuspended microspheres, which were then mixed by vortexing briefly and kept at room temperature until required.

The amount of rRNA Removal solution required was dependent on the amount of total RNA used in the reaction. For 1 - 2.5 µg template RNA in a maximum volume of 28 µl, 8 µl of rRNA removal solution were used. For 2.5 - 5 µg template RNA in a maximum volume of 26 µl, 10 µl of rRNA removal solution were used. For each reaction, the following were combined in a 0.5 ml tube in the order given: 4 µl Ribo-Zero Reaction Buffer, 1 - 5 µg total RNA sample, 8.5 or 10 µl Ribo-Zero rRNA Removal Solution, then brought to 40 µl with water. Reactions were gently mixed by pipetting then incubated at 68 °C for 10 minutes, then at room temperature for 15 minutes.

The washed microspheres were briefly vortexed and centrifuged to collect, then homogenised by gently pipetting up and down. The hybridized RNA samples were added to the microspheres and, without changing the pipette tip, immediately mixed by rapidly pipetting 10 - 15 times. Contents were immediately vortexed for 5 seconds and placed at room temperature before proceeding to the next sample. Tubes were incubated at room temperature for 10 minutes with vortexing for 5 seconds every 3 to 4 minutes. At the end of the 10 minute incubation at room temperature, tubes were mixed by vortexing for 5 seconds and then placed at 50 °C for 10 minutes in a heat block. The RNA-microsphere suspensions were then immediately transferred to Microsphere Removal Units and centrifuged at 13,000 rpm for 1 minute at room temperature. The eluates containing the rRNA depleted RNA were placed on ice.

To purify the treated RNA, the volume was adjusted to 180 µl with water before adding 18 µl of 3 M sodium acetate and 2 µl glycogen (10 mg/ml) to each tube and mixing by gentle vortexing. Three volumes (600 µl) of ethanol (100%, ice cold) were added to each tube and mixed thoroughly by vortexing. Tubes were incubated at –20 °C for at least 1 hour then centrifuged at 13,000 rpm for 30 minutes, carefully discarding the supernatants. The pellets were washed twice with ethanol (70%, ice cold) and centrifuged at 13,000 rpm for 5 minutes, carefully discarding the supernatants. Tubes were centrifuged briefly to collect any residual supernatants which were carefully discarded. The pellets were allowed to air dry at room temperature for 5 minutes before being resuspended in 18 µl water, quantified using a Qubit 2.0 fluorometer (see 2.3.2) then stored at -80 °C.

2.4.5 Ribosomal RNA depletion with Ribo-Zero Bacteria and Plant Seed / Root (magnetic)

The Ribo-Zero (magnetic) (Epicentre, Madison, WI, USA) reactions were performed according to manufacturer's instructions as follows. Magnetic beads were mixed thoroughly by vortexing, then for each Ribo-Zero reaction 225 µl magnetic beads were added to a 1.5 ml tube (maximum 6 reactions per tube). The tube was placed on a magnetic rack for at least 1 minute until the solution appeared clear before the supernatant was discarded. The tube was removed from the stand and an equal volume of water was added and mixed well by vortexing briefly. The tube was placed on a magnetic rack for at least 1 minute until the solution appeared clear before the supernatant was discarded. The tube was removed from the magnetic rack and a volume of Magnetic Bead Resuspension Solution was added equal to the number of reactions x 60 µl. This was mixed well by vortexing before 65 µl washed magnetic beads per sample were transferred to new 1.5 ml tubes along with 1 µl of RiboGuard RNase Inhibitor. These were mixed by briefly vortexing then left at room temperature until required.

The amount of rRNA removal solution required was dependent on the amount of total RNA. For 1 - 2.5 µg template RNA in a maximum volume of 28 µl, 8 µl of rRNA removal solution were used. For 2.5 - 5 µg template RNA in a maximum volume of 26 µl, 10 µl of rRNA removal solution were used. The rRNA removal solution was made up of a 4:1 ratio of Ribo-Zero Bacteria and Ribo-Zero Plant Seed / Root kits. For each sample, the following were combined in a 0.5 ml tube in the order given: 4 µl Ribo-Zero Reaction Buffer, 1-5 µg total RNA sample, 8.5 or 10 µl Ribo-Zero rRNA Removal Solution before being brought to 40 µl with water. Reactions were gently mixed by pipetting then incubated at 68 °C for 10 minutes then at room temperature for 5 minutes.

The treated RNA samples were added to the 1.5 ml tubes containing the washed magnetic beads, and without changing the pipette tip immediately and thoroughly mixed by pipetting up and down at least 10 times. Tubes were immediately vortexed for 10 seconds then incubated at room temperature for 5 minutes. Following incubation, reactions were mixed by vortexing for 5 seconds then placed in a 50 °C heat block for 5 minutes. Tubes were then immediately placed on a magnetic rack for at least 1 minute, until the solutions appeared clear. The rRNA depleted supernatants (85-90 µl) were carefully removed and transferred to new 1.5 ml tubes on ice.

Treated RNA samples were then cleaned using the RNEasy MinElute kit using a modified procedure as follows. Samples were brought to 100 µl with water before adding 350 µl of Buffer RLT then mixing. Ethanol (550 µl, 100%) were added to the diluted RNA and mixed well by pipetting before half of each sample (≈500 µl) was transferred to RNeasy MinElute spin columns placed in 2 ml tubes. Tubes were centrifuged at 10,000 rpm for 15 seconds and flow-throughs discarded. The remainder of each sample was transferred to the same spin column and centrifugation repeated, discarding the flow-throughs. The spin columns were placed in new 2 ml tubes before 500 µl Buffer RPE were added and the tubes centrifuged at 10,000 rpm for 15 seconds, discarding the flow-throughs. Ethanol (500 µl, 80%) was added to the spin columns and centrifuged at 10,000 rpm for 2 minutes, discarding the flow-throughs. The spin columns were placed in new 2 ml tubes and centrifuged at 13,000 rpm for 5 minutes, discarding the flow-throughs. The spin columns were placed in new 1.5 ml tubes and 12.5 µl water were applied directly to the centre of the spin column membrane. Tubes were centrifuged for 1 minute at 13,000 rpm to elute the RNA. This elution step was repeated with 12.5 µl water. The RNA samples were then quantified using a Qubit 2.0 fluorometer (see 2.3.2) and stored at -80 °C.

2.5 Sequencing and bioinformatic analyses

2.5.1 Preparation of samples for sequencing

For 454 pyrosequencing, Rubicon generated cDNA was submitted TGAC as part of a Capacity and Capability Challenge (CCC) project. Here it was assessed with a high-sensitivity Agilent bioanalyser to ensure size profiles fragments were consistent across samples. Multiplexing and sequencing were carried out on a 454 GS Flx sequencer using Titanium chemistry (Roche). For Illumina HiSeq (Illumina, San Diego, CA, USA) sequencing DNA, and rRNA depleted RNA samples were submitted to TGAC for library construction and sequencing with 100 bp paired end reads, again as part of a CCC project.

2.5.2 Bioinformatic analysis of 454 pyrosequencing data

Sequences were quality filtered using standard 454 Newbler parameters during conversion of .fna to .fasta files formats. They were then de-multiplexed to provide individual .fasta files for each sample. The conserved tail generated by the Rubicon procedure for cDNA synthesis was removed using a Perl script that removed the first 22bp of each read. The

emulsion PCR step during library preparation of 454 sequencing has been shown to introduce a bias resulting in artificial replicate sequences which can be filtered out (Gomez-Alvarez *et al.*, 2009). However due to the dominance of rRNA sequences and their similarity, particularly at their transcription start sites, the filtering step would likely have removed genuine biological replicates which would have down-weighted abundant taxa. Therefore reads were used in downstream analyses without filtering artificial replicates, as in another metatranscriptomic study (Ottesen *et al.*, 2011).

Read files were used as queries against a cleaned and de-replicated (95% identity) set of sequences in a single database derived from the small sub-unit (SSU) SILVA (Pruesse *et al.*, 2007) and RDP (Cole *et al.*, 2009) rRNA databases using USEARCH in UBLAST mode (Edgar, 2010). An E-value cut-off of 10^{-7} was applied, and the top 100 hits were recorded in an output file, short reads (<10 bp) were discarded in the process. Output files were uploaded into MEGAN (Huson *et al.*, 2007) using default parameters, except that Min. Support was set to 1, and Top Percent to 5.

To compare groups of samples, comparison files were generated in MEGAN for all relevant samples using absolute counts, and numbers of assigned reads per taxa were extracted for different taxonomic levels. Reads were normalised by expressing as a percentage of the total number of reads assigned in MEGAN minus any reads that were assigned to Viridiplantae. Means were calculated for each group of samples from the same environment and differences between environments were statistically validated using an unpaired t-test. Pair-wise comparisons were made between each of the plant rhizospheres with soil, and for the wild-type oat versus the *sad1* oat mutant. Statistically significant differences were further filtered using an abundance cut off of 0.01% of assigned reads for the environment in which they were more abundant. For example a taxon statistically more abundant in the wheat rhizosphere compared to bulk soil would be ignored unless it contributed at least 0.01% of the reads assigned to the wheat rhizosphere community.

Rarefaction analyses were performed separately on prokaryotes and eukaryotes at the phylum and genus levels for each sample using MEGAN. Data were extracted and absolute read numbers were calculated. Means for both number of reads sampled and number of taxa detected were generated for each group of samples, and then used to plot rarefaction curves.

Additional analyses were performed by Mark Alston (TGAC) as part of the bioinfomatic support accompanying the CCC project agreement. Between-classes principal component analysis (PCA) was carried out using the R package ade4 (Dray and Dufour, 2007). Before analysis, the taxon abundance counts for each sample were normalised to 100,000 reads within MEGAN (Huson *et al.*, 2007) and low abundance taxa removed if the average abundance across all the samples was < 0.01% or < 0.1% depending on the taxonomic level being tested. PCAs were performed at both phylum and genus level for both prokaryotes and eukaryotes, and also at genus level for four major eukaryotic groups (Fungi, Nematoda, Amoebozoa, Alveolata).

2.5.3 Analysis of Illumina HiSeq sequencing data

All samples were de-multiplexed and quality filtered as standard, and data analysed by two different approaches, at the The European Bioinformatics Institute (EBI, Hinxton, UK) and at TGAC. At TGAC, analyses were largely performed by Mark Alston as part of the bioinfomatic support accompanying the CCC project agreement. Sequence data from DNA samples was analysed using Metaphlan (Segata *et al.*, 2012) and Metaphyler (Liu *et al.*, 2010) to determine taxonomic composition based on protein coding genes. Data were also uploaded to MG-RAST (Meyer *et al.*, 2008) to assign functional information based on the SEED database (Overbeek *et al.*, 2005) and analysed using default paramaters, i.e. an E-value cut-off of $1E^{-5}$, minimum identity cut-off of 60%, and a minimum alignment length cut-off of 15 bp.

For the RNA data, residual rRNA sequences were removed from samples *in silico* using SortMeRNA (Kopylova *et al.*, 2012) and the number of copies of RIS recovered was determined using USEARCH with an identity cut-off of 1. Sequencing depth and transcriptional activity per gram of soil were then calculated (see 2.5.4). Non-rRNA reads were filtered using Sickle (https://github.com/najoshi/sickle) then analysed using Metaphyler to determine taxonomic composition. A subset of the data (25 million reads based on the lowest read count sample) were analysed using rapsearch2 (Zhao *et al.*, 2012), a reduced alphabet BLAST-like algorithm, against the non-redundant nucleotide collection at the National Centre for Biotechnology Information (NCBI). Output files were uploaded into MEGAN (Huson *et al.*, 2007) using default parameters (min support = 5, min score = 50, top% 10) to visualise and compared samples based on taxonomic composition,

SEED and KEGG (Kanehisa and Goto, 2000) assignments. Pair-wise comparisons were made between each plant rhizosphere and soil using un-paired t-tests with a 95% confidence interval. Some multiple comparisons were made using analysis of variance (ANOVA). In addition, all samples in full were uploaded to MG-RAST and analysed using default parameters. Multidimensional scaling analysis was performed in PRIMER6. Data were normalised to a percentage then square root transformed before a Bray-Curtis similarity matrix was generated and used to plot data on x and y axis to generate the plot in Excel.

At EBI, at subset of reads (mean 92 million) were analysed using the EBI Metagenomics Portal courtesy of Peter Sterk. SeqPrep (https://github.com/jstjohn/SeqPrep) was used to merge mate pairs and perform additional quality filtering. The parameters used were as follows: -f -r -1 -2 -3 -4. If reads did not overlap, both reads were used in the analysis. Further filtering, including a 100 bp cut-off was applied using Trimmomatic (http://www.usadellab.org/cms/?page=trimmomatic) with default parameters. Residual rRNA sequences were removed from the RNA sample *in silico* using rRNASelector (Lee *et al.*, 2011). Non-rRNA reads were analysed by InterProScan 5 (Quevillon *et al.*, 2005; Zdobnov and Apweiler, 2001) to generate InterPro and Gene Ontology (GO) assignments. Pair-wise comparisons were made between each plant rhizosphere and soil using unpaired t-tests with 95% confidence interval.

2.5.4 Calculation of sequencing depth and transcript abundances

The length of the RIS generated (967 bp), as determined by the Experion bioanalyser (2.3.2), allowed the sequence to be estimated, based on the number of base pairs downstream of the T7 promoter, which in turn allowed calculation of the molecular weight. This was used to determine the number of copies of RIS per µl of the stock solution, and thus how many copies were added to each RNA sample during extraction. Post-sequencing, USEARCH was used to determine the % of a subset of reads from each sample that matched the RIS sequence with 100% identity. The % of the subset was used to calculate the number of RIS sequences recovered in the whole sample. Sequencing depth was calculated using the following equation (Gifford *et al.*, 2011):

(Standards recovered / Standards added) x 100%

The % non-rRNA in the samples was determined by Mark Alston at TGAC using SortMeRNA (Kopylova *et al.*, 2012), and transcript abundance per sample was calculated using the following equation (Moran *et al.*, 2013):

(Standards added / Standards recovered) x non-rRNA transcripts sequenced

This value was then divided by the total mass of input soil for each RNA extraction to obtain a value for transcripts per g soil.

Subsequent analyses provided numbers of reads matching particular protein coding genes or taxonomic groups in a database. To convert this to a quantitative value of number of transcripts per g a modification of the above equation was applied as follows:

(Standards added / Standards recovered) x "specific protein coding transcript" sequenced

Again, this value was then divided by the total mass of input soil for each RNA extraction to obtain a value for transcripts per g soil.

2.6 Primer oligonucleotide sequences

Table 1; Primer oligonucleotide sequences used in the study. All target rDNA with the exception of the Rubicon primer. Bac. refers to Bacteria, and euk. refers to eukaryotes. For primers EB_1492R_T7 and EB_2490R_7, the T7 promotor sequences are underlined, and preceded by a sequence to encouraging polymerase binding. References are as follows: [1](Hamady *et al.*, 2008), [2](van Hannen *et al.*, 1998), [3](Stewart *et al.*, 2010a), and [4](Kim *et al.*, 2003).

Primer Name	Target	Sequence (5' to 3')	Reference
EB_27F	Bac.16S	GCCTTGCCAGCCCGCTCAGTCAGAGTTTGATCCTGGCTCAG	1
EB_338R	Bac.16S	CATGCTGCCTCCCGTAGGAGT	
EUK_1616R	Euk.18S	GCGGTGTGTACAAAGGGCAGGG	2
EUK_1427F	Euk.18S	TCTGTGATGCCCTTAGATGTTCTGGG	
EB_1492R_T7	Bac. 16S	GCCAGTGAATTGTAATACGACTCACTATAGGGACGGCTACCTTGTTACGACTT	3
EB_189F	Bac. 23S	GAASTGAAACATCTHAGTA	
EB_2490R_T7	Bac. 23S	GCCAGTGAATTGTAATACGACTCACTATAGGGCGACATCGAGGTGCCAAAC	
Plant_18S_F	Plant 18S	ATGATAACTCGACGGATCGC	4

Plant_18S_R	Plant 18S	CTTGGATGTGGTAGCCGTTT
Rubicon	Tail	GTGGTGTGTTGGGTGTGTTTGG

Chapter 3: Ribosomal RNA based community analysis of crop plant rhizosphere microbiomes

3.1 Introduction

Interactions between plants and microbes in the rhizosphere are of global importance to biogeochemical cycling (Philippot *et al.*, 2009), plant health and productivity (Bloemberg and Lugtenberg, 2001). Colonisation of the rhizosphere, the region of soil influenced by plant roots, is necessary for both plant pathogens and plant growth promoting rhizobacteria (PGPR). The latter aid plants by providing nutrients, modulating growth, and defending against diseases (Lugtenberg and Kamilova, 2009), thus contributing to disease suppressive soils (Mendes *et al.*, 2011). Many plant associated microbes are known and well studied, including the symbiotic nitrogen-fixing *Rhizobium leguminosarum* (Young *et*

al., 2006), and both beneficial and pathogenic *Pseudomonas spp.* (Feil *et al.*, 2005; Paulsen *et al.*, 2005). Also, association of arbuscular mycorrhizal fungi (AMF) with most land plants is fundamental to their acquisition of mineral nutrients such as phosphate (Bonfante, 2010) . However, little is known about how these organisms interact at the community level.

Every gram of soil is estimated to contain in excess of 50,000 species of bacteria (Roesch *et al.*, 2007), the vast majority of which are uncultured (Handelsman, 2004). Sequencing of PCR-amplified 16S rDNA has been extensively used to examine rhizosphere bacterial communities of various plants, and recently high-throughput pyrosequencing (Margulies *et al.*, 2005) has revealed these communities in previously unobtainable detail. Plants studied include the important crops potato , *Solanum tuberosum*, (Inceoglu *et al.*, 2011) and maize, *Zea mays*, (Peiffer *et al.*, 2013), plants of the Antarctic (Teixeira *et al.*, 2010), and recently the model dicot *Arabidopsis thaliana* (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012). However, a significant limitation of such approaches is that PCR amplification of genomic DNA (gDNA) is inherently biased by primer design (Hong *et al.*, 2009; Pinto and Raskin, 2012) and is limited to the targeted division of life (Bacteria, Archaea, Eukarya or a smaller taxonomic group). In addition, studies to date have largely focused on either bacteria or fungi, often neglecting other eukaryotes and archaea. These organisms may represent significant proportions of a community in terms of their abundance or they may play a functional role in their environments. For example, methanogenesis, a key step in the carbon cycle, is an ability specifically possessed by archaea (Friedrich, 2005).

High-throughput sequencing has also enabled the use of metagenomic strategies where total genomic DNA from the rhizosphere is sequenced (Tett *et al.*, 2012). While encompassing all domains of life and indicating the metabolic potential of a microbiome, a metagenome contains relatively few rRNA genes, reducing the strength of taxonomic assignments. Metatranscriptomics, where total RNA from the environment is sequenced,

reveals active community members and metabolic pathways (Urich *et al.*, 2008). Many applications of metatranscriptomics are focused on the latter but a significant challenge is posed in the requirement for enrichment of mRNA (Stewart *et al.*, 2010a; Yi *et al.*, 2011). However, the dominance of rRNA in a metatranscriptomic sample (≥95% for soil) allows robust assessment of the phylogenetic structure of the entire microbiome, without prior selection of taxonomic groups for study. This is technically much less challenging than enrichment of mRNA, avoids PCR bias and can be carried out straightforwardly on multiple samples.

Here, comparative metatranscriptomics was used to study the rhizosphere microbiomes of three crop plants grown in the same soil; wheat (*Triticum aestivum*) a major world food staple, oat (*Avena strigosa*) a cereal that produces antifungal avenacins (Maizel *et al.*, 1964), and pea (*Pisum sativum*) a widely grown crop legume nodulated by N_2-fixing *Rhizobium leguminosarum*. In addition the rhizosphere microbiomes of the wild type oat (*Avena strigosa*) and an avenacin deficient mutant (*sad1*) (Papadopoulou *et al.*, 1999) were compared.

3.2 Materials and methods

Seeds were planted in soil (JIC1, see Appendix table A33) and grown for 4 weeks (See 2.2.2). Total RNA and DNA were extracted from rhizospheres of wheat, *wt* and *sad1* mutant oats and pea, as well as unplanted bulk soil using the RNA PowerSoil RNA isolation kit and DNA elution accessory kit (See 2.3.1). RNA samples were treated with TurboDNase (See 2.3.3), converted to cDNA using Rubicon stand displacement amplification (See 2.3.4) and multiplexed and sequenced on a 454 GS Flx sequencer (See 2.5.1). Read files were de-multiplexed and searched against a custom SSU database using USEARCH (Edgar, 2010), then uploaded and analysed in MEGAN (Huson *et al.*, 2007) to produce read counts for different taxonomic groups (See 2.5.2). For some taxonomic groups, data were used to

generate between-class principle component analyses using the R package ade4 (Dray and Dufour, 2007). Quantitative PCR was performed on DNA samples using the iQ SYBR Green Supermix, with bacterial 16S rRNA primers and plant 18S rRNA primers (See 2.3.7).

3.3 Results and Discussion

3.3.1 Sequencing and analysis summary

A total of 1,674,231 reads were generated for the 19 biologically independent samples (4 each from soil and oat, 3 each from wheat and pea, 5 from the sad1 mutant oat line). The PHRED quality scores (±1SD) were 32.3 ±8.3 for the first plate containing soil and wheat rhizosphere samples, and 34.3 ±7.7 for the second plate containing rhizosphere samples from wild-type oat, *sad1* oat mutant and pea. Average read lengths were 260 bp and 235 bp respectively.

There was large variation in read output across samples (Table 3.1), particularly for the soil and wheat samples which were sequenced on the same sequencing plate. Each barcode contains a single 6-carboxyfluoroscein (FAM) label, so measuring the fluorescence of each sample library should have accurately determined its concentration, allowing samples to be equimolar pooled. This was done for the first sequencing plate containing the soil and wheat samples. The variation in read counts from the sequence data showed that fluorescence is a poor determinant of library concentration (Figure 3.1). Quantitative PCR (qPCR) was performed at The Genome Analysis Centre (TGAC) on the mixed library to determine the concentrations of the different samples. Although the qPCR tended to overestimate the concentration of libraries with low numbers of reads, the results were more similar to the sequence data (Figure 3.1), and this method has since been used to determine library concentrations for future multiplexing experiments. The variation in read number for the second plate containing oat and pea samples was much less than that of the first plate containing soil and wheat samples (56709±37974 compared to 11272±84145) (Table 3.1). Amplification bias introduced by pyrosequencing barcodes has

been documented elsewhere (Alon *et al.*, 2011). Small variations in input DNA levels may also have contributed to the variation. This is important to account for when designing multiplexed pyrosequencing experiments, but by normalising data and making only relative comparisons, statistical challenges arising from this variation were avoided here.

Table 3.1: Summary of sequencing output, USEARCH and MEGAN analyses. Samples B110, B111, B53, B54, W75, W76 and W77 correspond to MIDs 1, 2, 11, 12, 3, 4, and 5 respectively (see Figure 3.1). SSU Hits are those reads that matched a sequence in the small subunit rRNA database. These were uploaded into MEGAN and Viridiplantae sequences removed.

Environment	Sample ID	Total Reads	SSU Hits		Assigned in MEGAN		Viridiplantae Hits	Assigned in MEGAN (minus Viridiplantae)
			Number	% of Total Reads	Number	% of SSU Hits		
Soil	B110	6821	3342	49	3321	99.4	20	3301
	B111	141957	80184	56.5	79750	99.5	439	79311
	B53	249373	144346	57.9	140350	97.2	483	139867
	B54	121413	67486	55.6	65655	97.3	221	65434
Wheat	W75	161443	88886	55.1	88371	99.4	558	87813
	W76	19272	11567	60	11518	99.6	57	11461
	W77	89125	51115	57.4	50863	99.5	335	50528
wt Oat	O41	8428	3855	45.7	3838	99.6	60	3778
	O42	66227	30807	46.5	30664	99.5	1122	29542
	O43	10793	5359	49.7	5334	99.5	58	5276
	O45	112731	53466	47.4	53253	99.6	4118	49135
sad1 Oat	S47	112096	46608	41.6	45172	96.9	786	44386
	S48	72452	33655	46.5	33548	99.7	1387	32161
	S49	94136	38034	40.4	37821	99.4	658	37163
	S51	85204	44772	52.5	44626	99.7	4405	40221
	S52	118283	54281	45.9	53982	99.4	3180	50802
Pea	P63	88005	41795	47.5	41657	99.7	2556	39101
	P64	56072	27888	49.7	27805	99.7	2071	25734
	P66	54705	21447	39.2	21325	99.4	417	20908

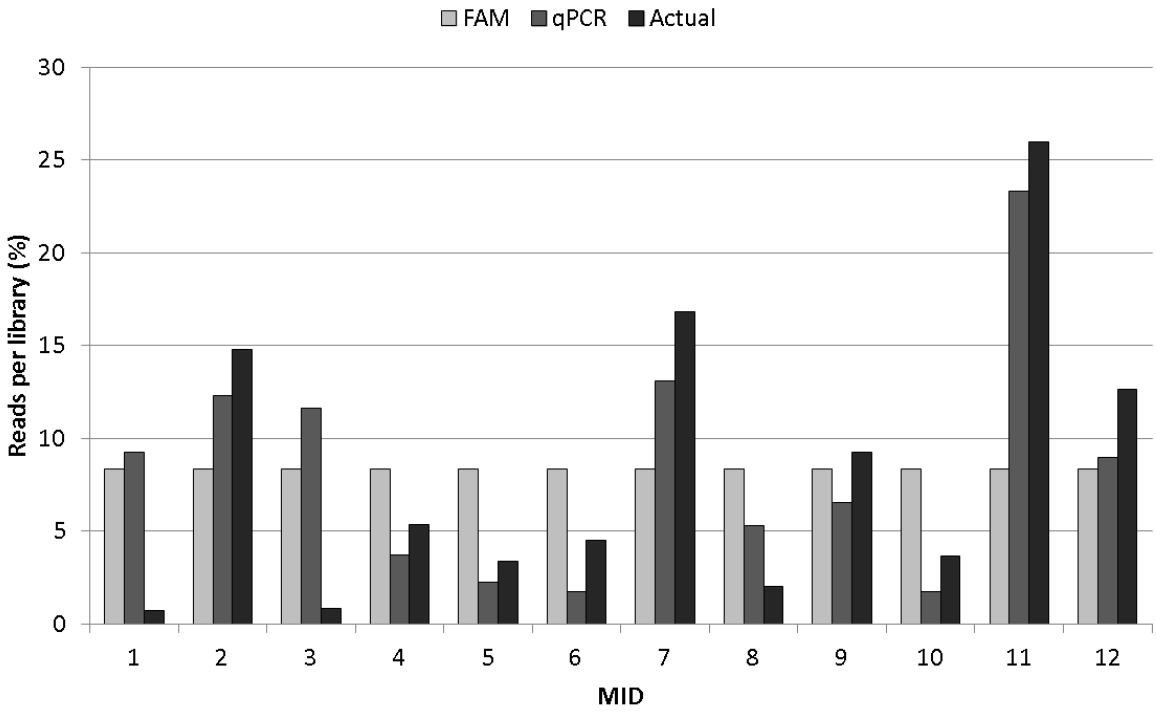


Figure 3.1: Variation in predicted and actual proportions of reads for the soil and wheat rhizospsphere sequencing libraries. Libraries were equimolar pooled based on fluorescence from a 6-carboxyfluoroscein (FAM) label. Quantitative PCR (qPCR) of the barcodes generally produced a more accurate prediction, except for MIDs 1 and 3 which had very low actual read abundances. MID refers to multiplex identifier. MIDs 1, 2, 11, 12, 3, 4, and 5 correspond to samples B110, B111, B53, B54, W75, W76 and W77 respectively (see Table 3.1).

3.3.3 Identification of a human contaminated sample

Initially the read files from the first sequencing run (bulk soil and wheat rhizosphere samples) were searched against the entire NCBI nr nucleotide database. This was computationally intensive, taking several weeks and producing output files that were tens

of gigabytes (GB) in size. The output files each took several hours to load into MEGAN. The advantage of this approach is that it provided a taxonomic assignment for the majority of reads in the samples, from small and large-subunit rRNA and also mRNA, including sequences from RNA viruses. When uploaded into MEGAN, one of the wheat rhizosphere samples (W78, not included in subsequent analyses) showed a high proportion of human sequences(Figure 3.2b), and also sequences of human associated organisms such as *Escherichia*, *Yersinia* (Figure 3.2a) and Variola virus (Figure 3.2c). When this sample was analysed using USEARCH against the custom small-subunit (SSU) database, only 6.1% of reads had a hit. If human or human-associated microbial RNA was the contamination, this would be reflected in the SSU analyses. However, the MEGAN output did not reveal any major differences between sample W78 and the other samples, suggesting that human DNA was the contaminant. Reinforcing this, the human sequences matched to human bacterial artificial chromosome (BAC) clones, and Variola virus is a DNA virus with no RNA stage. This contamination must have occurred during handling of the cDNA or RNA after DNase treatment.

Interestingly, a number of sequences matched closely related primates (*Pan*, *Gorilla*), other mammals (*Mus*, *Sus*) and animals (*Drosophila*), which were clearly not present in the soil. Upon closer inspection, reads that hit these organisms had a much higher score against human BAC clone sequences, but were interpreted by MEGAN as being “unknown” and therefore did not contribute to the assignment of the read. While this sample was not included in subsequent analyses of the microbiome, it has highlighted some important limitations of metagenomic approaches and the use of MEGAN. The presence of human DNA in a sample from the wheat rhizosphere was unexpected, and easily noticeable. However, it would be difficult to identify, let alone account for, such contamination in a microbiome sample from a human, or another mammal. Considering these limitations, the

SSU rRNA based analysis using USEARCH and MEGAN was developed (Walshaw and Grant, in preparation), and performed on all samples.

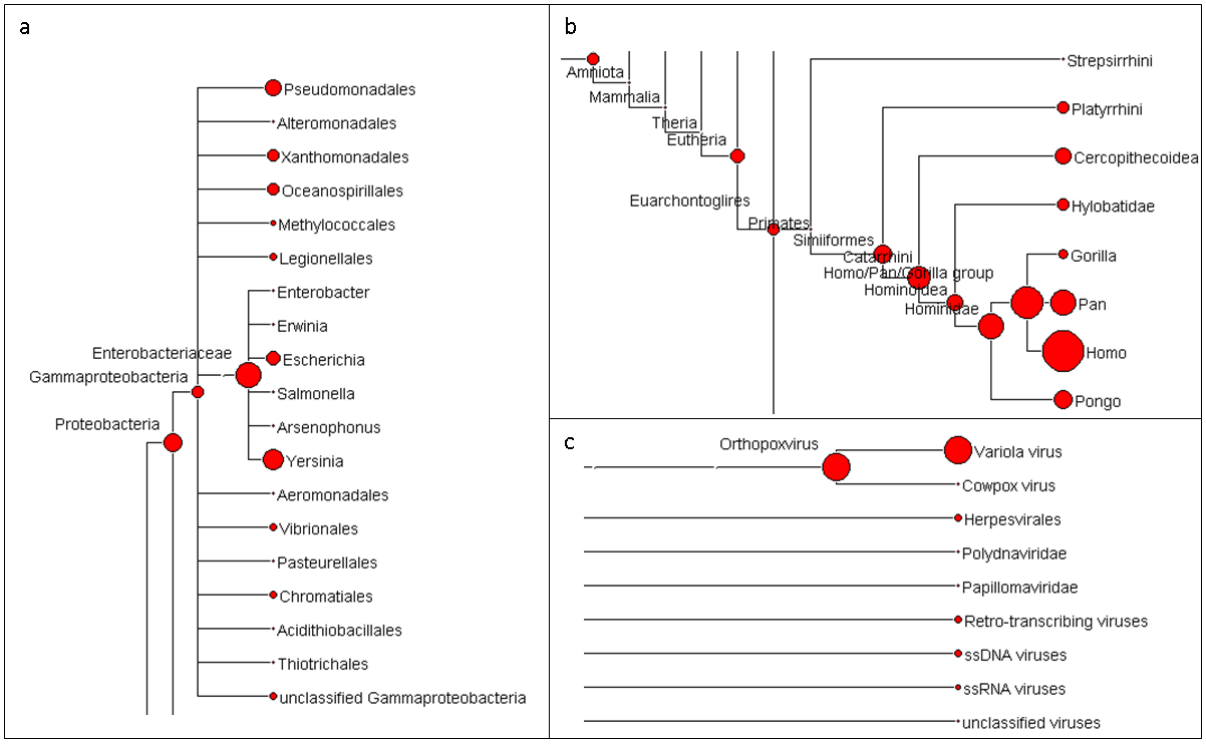


Figure 3.2: MEGAN analysis of DNA contaminated sample W78, showing high abundances of Enterobacteriaceae (a), human and primate (b) and DNA virus (c) sequences.

3.3.4 Lowest common ancestor analysis with MEGAN

The USEARCH SSU analysis allowed data to be processed in a few hours on a desktop computer, and output files were of a much more manageable size. The top 100 hits were recorded in a USEARCH output file for each sample. This provided a highly similar result to when the top 500 hits were recorded, but in a much shorter time. The default parameters for MEGAN are optimised for metagenomic samples based on DNA, the nature of a metatranscriptome is quite different, and parameters were adjusted accordingly. Minimum support was reduced from 5 to 1. This means that a taxon was represented in the final output even if only one read was assigned to it, allowing even comparison between

samples with different read counts. The default minimum bit score (Min. Score) of 35 was retained, although the stringency of the 10^{-7} E value cut-off used in the USEARCH analyses meant that the majority of reads input into MEGAN would meet this requirement. Indeed >99% of SSU hits had were assigned to a taxon in MEGAN (Table 3.1). Because all the reads input into MEGAN were derived from SSU rRNA, they would have been much more similar to each other than to reads from a shot-gun metagenome. At the default Top percent parameter of 10%, a large number of reads would be assigned to high taxonomic levels. By reducing this to 5%, fewer reads are assigned overall, but they are able to be assigned to much more specific taxa, resulting in a reduction in reads assigned to cellular organisms. No sequences where observed to be assigned to mitochondria or chloroplasts, despite the separate nodes designated for these organelles within MEGAN and their likely presence in samples, assuming they are efficiently extracted by the methods used. It is possible however that their rRNA sequences share similarity with other Ricketsialles and Cyanobacteria respectively so that they are placed witin a higher level taxonomic node where they would contribute to the relative abundance.

On average, roughly half the reads had a hit in the custom SSU rRNA database and the majority of those were assigned to a taxon in MEGAN (Table 3.1). In a similar study of soil, only 38% of reads matched SSU rRNA (Urich *et al.*, 2008), lower than in the current study. Reasons for this difference are likely to be a combination of the differences in database and analysis parameters. The remainder were mostly large sub-unit rRNA or rarely mRNA sequences. These can be taxonomically assigned, but the available databases are less comprehensive and less well curated, particularly for mRNA derived sequences. Therefore only sequences matching small sub-unit rRNA were analysed further. This allowed easier comparison with published studies where 16S and 18S PCR have been used.

3.3.5 Analysis of plant rRNA in soil and the rhizospheres

The presence of plant rRNA in a sample may indicate a seed bank present in the soil, however, rhizosphere samples would inevitably be contaminated with host plant material, which is indistinguishable from other plants based on 18S rRNA sequence. Because the analysis method is limited to making relative comparisons, any amount of plant rRNA in a sample would down-weight other taxa, giving the impression they were depleted in a particular rhizosphere. Reads were therefore normalised by expressing as a percentage of the total number of reads assigned in MEGAN minus any reads that were assigned to Viridiplantae (Table 3.1).

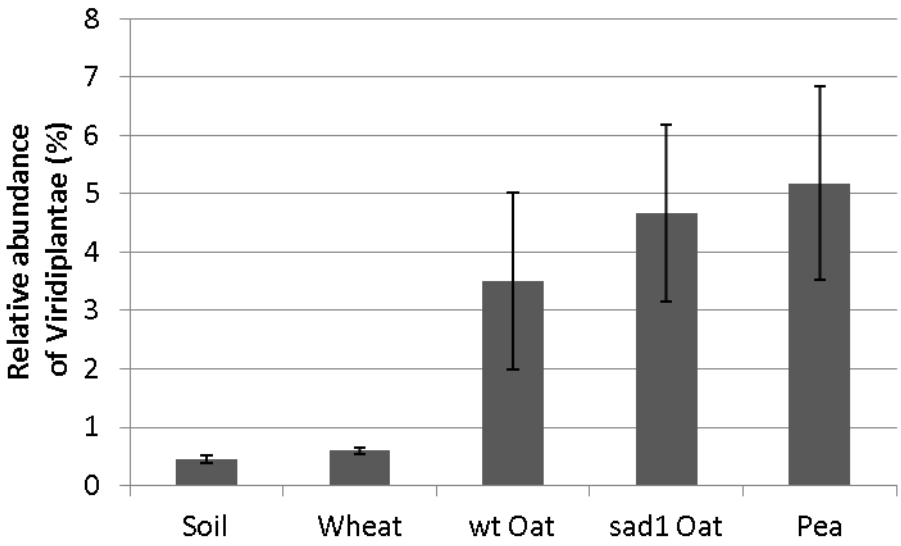


Figure 3.3: Relative abundance of Viridiplantae (plant) sequences in samples. Differences between a rhizosphere and soil where only significant for the pea rhizosphere (P values were 0.171, 0.090, 0.027 and 0.035 for wheat, *wt* oat, *sad1* oat and pea respectively).

It might be expected that the rhizospheres would contain higher amounts of plant rRNA compared to soil. However this was only the case for the *sad1* oat and pea rhizosphere (Figure 3.3). A number of factors would contribute to the measurement of plant rRNA including the abundance of plant cells in the rhizosphere samples after extraction, the

activity of those cells and the stability of their rRNA. Specifically, root cap border cells and also root hairs would contribute to the abundance of plant rRNA in the rhizospheres. The large variation in abundance of plant rRNA further reinforces the need for its removal before analysing the microbiome. There may also have been plant chloroplast and mitochondrial rRNA which may have been misclassified as Cyanobacteria or Ricketsialles respectively. To test this, a single sample (P66) was analysed using the RDP Classifier. Of 74 Cyanobacterial sequences, 10 were actually misclassified as chloroplast. However no Ricketsia sequences were identified, suggesting at least mitochondria sequences were not mis-assigned.

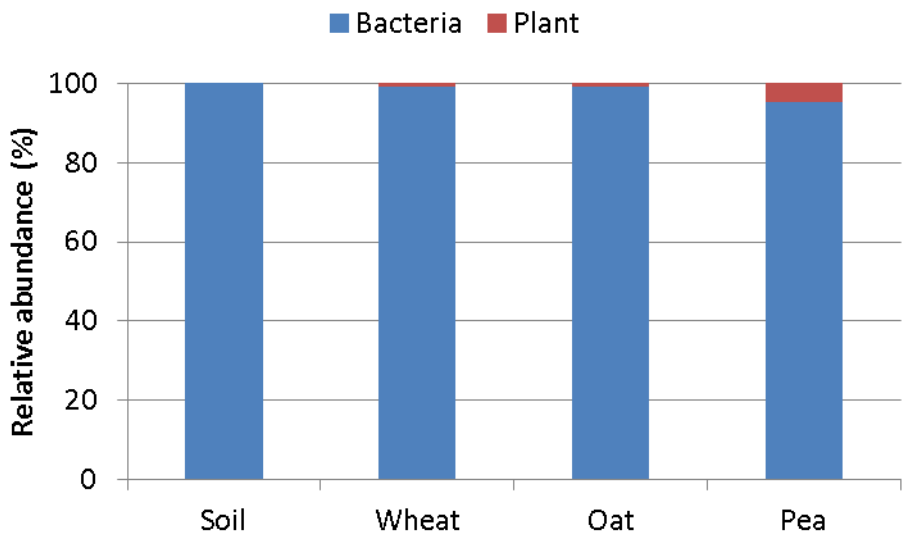


Figure 3.4: Relative abundance of bacteria and plants determined by quantitative PCR analysis of bacterial 16S and plant 18S in soil, wheat, oat and pea rhizospheres.

Quantitative PCR analysis of bacterial 16S and plant 18S rRNA genes showed that wheat, oat and pea rhizospheres had a lower relative abundance of bacteria than soil (P= 0.0105, 0.0308 and 0.017 respectively) and a higher relative abundance of plant tissue than soil (P= 0.014, 0.0002, 0.0003 respectively) (Figure 3.4). Plant 18S was 1575 and 1866 fold more abundant in wheat and oat rhizospheres respectively, compared to soil, while this value

rose to 9201fold in the pea rhizospheres, suggesting legume rhizospheres contain more plant cells than those of cereals. There was also little plant material overall (0.005% in soil) compared to bacteria, as is to be expected, though in the pea rhizosphere this value was higher at 4.7%. It is likely that the presence of plant material down weights the relative abundance of bacteria. This targeted approach produced a clearer result than the analysis of Viridiplantae from MEGAN, suggesting that MEGAN analysis of eukaryotes based on SSU rRNA is not optimum. Different set of parameters might be required to take into account the increased length and reduced diversity of the 18S rRNA compared to 16S rRNA, or alternatively the longer 28S rRNA could be used.

The 16S rRNA genes of prokaryotes are routinely used for phylogenetics and community profiling, resulting in a huge amount of available sequence data. A consequence of this is that any 16S rRNA gene or transcript will likely be similar to many other 16S genes or transcripts. MEGAN implements a lowest common ancestor (LCA) algorithm that means that reads are not simply assigned to the taxon with the best hit. This LCA algorithm takes into account other highly similar sequences (determined by Top % score set by the user), resulting in the read being assigned to a taxonomic node that would be the lowest common ancestor of the best hit and the most highly similar group of taxa to it. Here the analysed rRNA sequences were random fragments derived from full length rRNA transcripts, so the use of the LCA algorithm is particularly important for avoiding false assignment of conserved sequences to low taxonomic ranks.

A large proportion of reads were therefore assigned to high taxonomic ranks. Averaged across all samples, 5.5% of reads could not be classified more specifically than as cellular organisms, 81.4% were assigned to prokaryotes and 13.1% were assigned to eukaryotes. Within prokaryotes, 54.46% of reads were assigned to phylum level, while only 13.44% were assigned to genus level. For eukaryotes, 60.35% of reads were assigned to phylum

level and 36.86% were assigned to genus level (Figure 3.5). The 18S rRNA genes of eukaryotes are less well represented in sequence database. This is due to them being less well studied than bacteria and the more widespread use of the hypervariable internally transcribed spacer (ITS) region for both phylogenetics and community profiling. The consequence of this is that any read derived from a eukaryote is more likely to be similar to one taxonomic group than to many, resulting in a large proportion of reads being assigned at genus level. This is consistent with a high proportion of eukaryotic taxa from one phyla being derived from a particular genus rather than several. These data suggest metatranscriptomics is less discriminating at low level taxonomic ranks than PCR targeting of a variable rRNA gene region or an ITS. With sufficient sequence coverage, significant numbers of reads can be confidently assigned to species and even strain level, for both prokaryotes and eukaryotes, but the amount of sequence effort required for the same level of detail is greater than typically required for amplicon studies.

The use of MEGAN allowed community profiles to be generated and compared at any chosen taxonomic rank. The main focus here was on kingdom, phylum, and genus levels. Analysing at kingdom level is a unique opportunity provided by metatranscriptomics to reveal global differences between samples. Analysing phylum level is commonly used in 16S PCR based studies of microbial communities and allows any broad changes in community structure to be identified. While some reads are able to be assigned to species and strain level, taking genus level as a fine taxonomic rank has two advantages. Firstly, because of the LCA algorithm, more reads are assigned to genus level than species, particularly for prokaryotes. This increases the robustness of statistics applied and reduces the skewing of data due to differences in sequence depth. Secondly, the available sequences of 16S rRNA genes are biased towards those organisms that have been easy to culture and are of clinical or agronomical importance. The better studied genera, such as *Bacillus* and *Pseudomonas*, tend to be represented by a large number of species and

strains, while others, often recently cultured genera are monotypic, i.e. they have only one representative species such as *Desulforudis audaxviator*. The assigment of a read by MEGAN is biased against well characterised taxa, which will be assigned to higher taxonomic ranks. By focussing on genus level and using a dereplicated rRNA database, with more even taxonomic coverage, some of this bias is reduced.

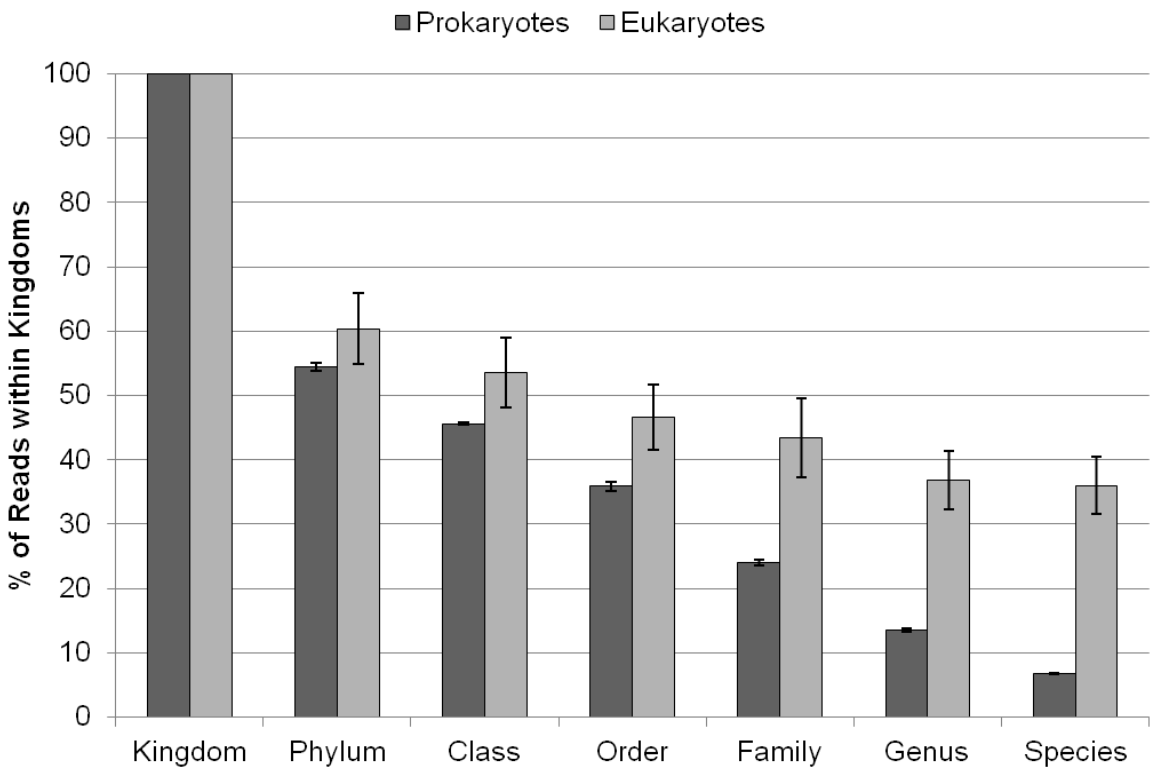


Figure 3.5: Effect of the Lowest Common Ancestor MEGAN analysis on assigning reads to taxonomic ranks. Values are means across all 19 samples, error bars represent ±1SEM.

3.3.6 Total Community Structure and Diversity

The proportion of sequences derived from Bacteria varied between environments, ranging from 91.0% for bulk soil, 88.5% for wheat rhizosphere, 77.3% for wild-type oat rhizosphere and 73.7% for pea rhizosphere (Figure 3.6a). These were significantly different from bulk

soil for all three rhizospheres (p= 0.0465, 0.0201 and 0.0107 respectively). Eukaryotes made up 2.8% of the bulk soil community and 3.3%, 16.6% and 20.7% of the wheat, oat and pea rhizosphere communities respectively (Figure 3.6a). These were only significantly different from bulk soil for oat and pea rhizospheres (p= 0.0380 and 0.0167 respectively). This striking difference in relative abundance of eukaryotes demonstrates the strength of a metatranscriptomic approach, rather than a PCR based strategy, to detect kingdom level differences between microbiomes. Such information can be achieved with qPCR targeting several different kingdoms, but is much less straightforward and is still affected by primer bias.

Archaea were consistently represented at around 0.5% (Figure 3.6a) for all environments, comparable to a study of soil (Urich *et al.*, 2008). The Crenarchaeota were the dominant phylum in soil and the three rhizospheres, followed by the Euryarcheaota. There were a large proportion of reads assigned to those organisms only able to be classified as Archaea.

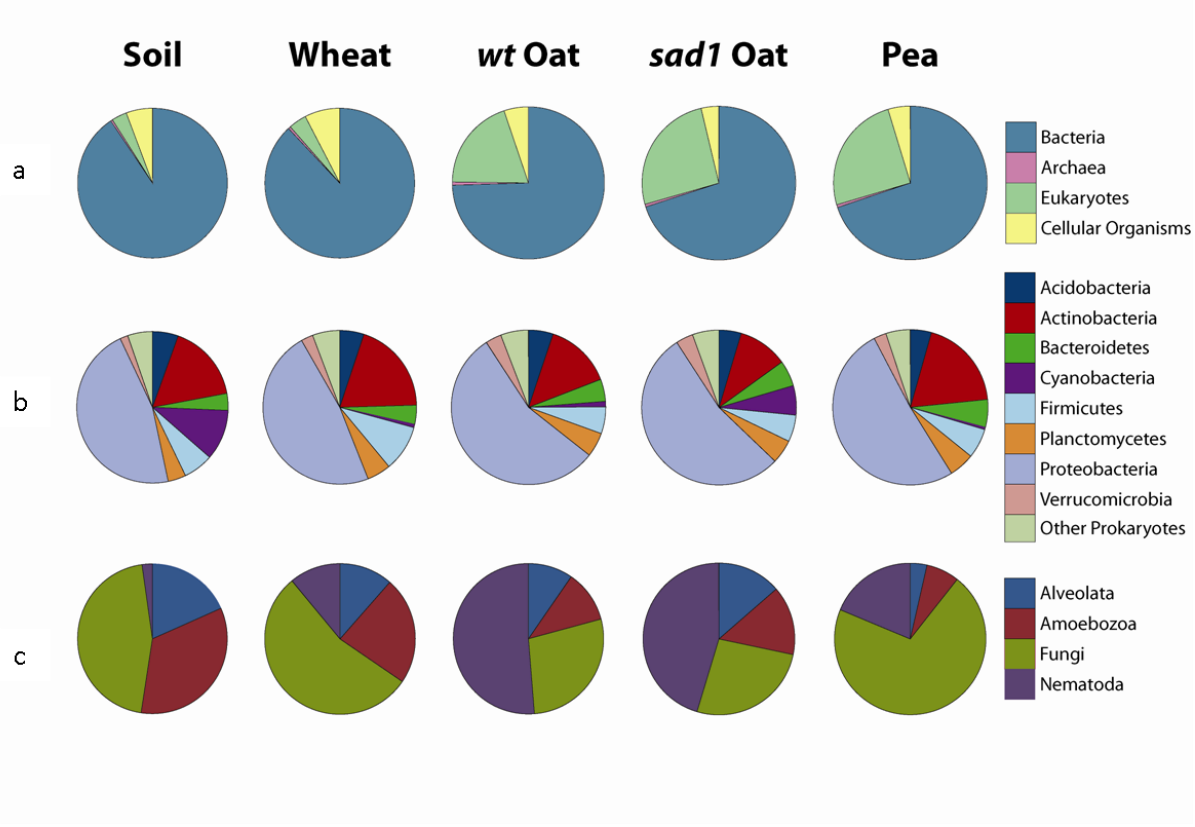


Figure 3.6: Relative abundance of taxonomic groups in soil, wheat, *wt* oat, *sad1* oat and pea rhizospheres. (a) all taxa at the domain level, the pie excludes sequences from Viridiplantae. (b) prokaryotic phyla, (c) four major eukaryotic groups that are studied in soil. Both (b) and (c) only include confidently assigned reads, i.e. not reads that cannot be assigned more specifically than to Bacteria, Archaea (b) or eukaryotes (c). Values are means of biological replicates where n=3 for wheat and pea, n=4 for soil and *wt* oat and n=5 for *sad1* oat.

Differences between microbiomes were visualised using between-classes principal component analysis (PCA). Prokaryotic and eukaryotic communities were analysed separately, at both phylum and genus level. At phylum level, prokaryotic communities of oat and pea were distinct from bulk soil, while that of wheat was not (Figure 3.7). At genus level, prokaryotic communities in all wild-type plant rhizospheres were different from soil, and more different from each other. This suggests selection of rhizosphere microbiomes was largely plant specific. Interestingly, the large changes seen in the eukaryotic communites did not appear to have a knock on effect of the prokaryotic communities. The pea rhizosphere microbiome was strikingly distinct from that of the other plants, suggesting a strong difference in the effect of a legume versus a cereal, though comparison of additional cereals and legumes would be needed to confirm this. Eukaryotic communities of bulk soil, wheat and pea were similar at phylum level, while that of oat was distinct (Figure 3.7). At genus level the eukaryotic communities of wheat and bulk soil were similar, while oat and especially pea were different, most noticeably for Fungi and Amoebozoa (Figure 3.8). Communities of Alveolata and Nematoda in the oat and pea rhizospheres were somewhat different to those of soil and wheat, although they were also highly variable (Figure 3.8). A Bray-Curtis similarity tree generated with the same data was consistnent with the PCA analysis (Supplementary

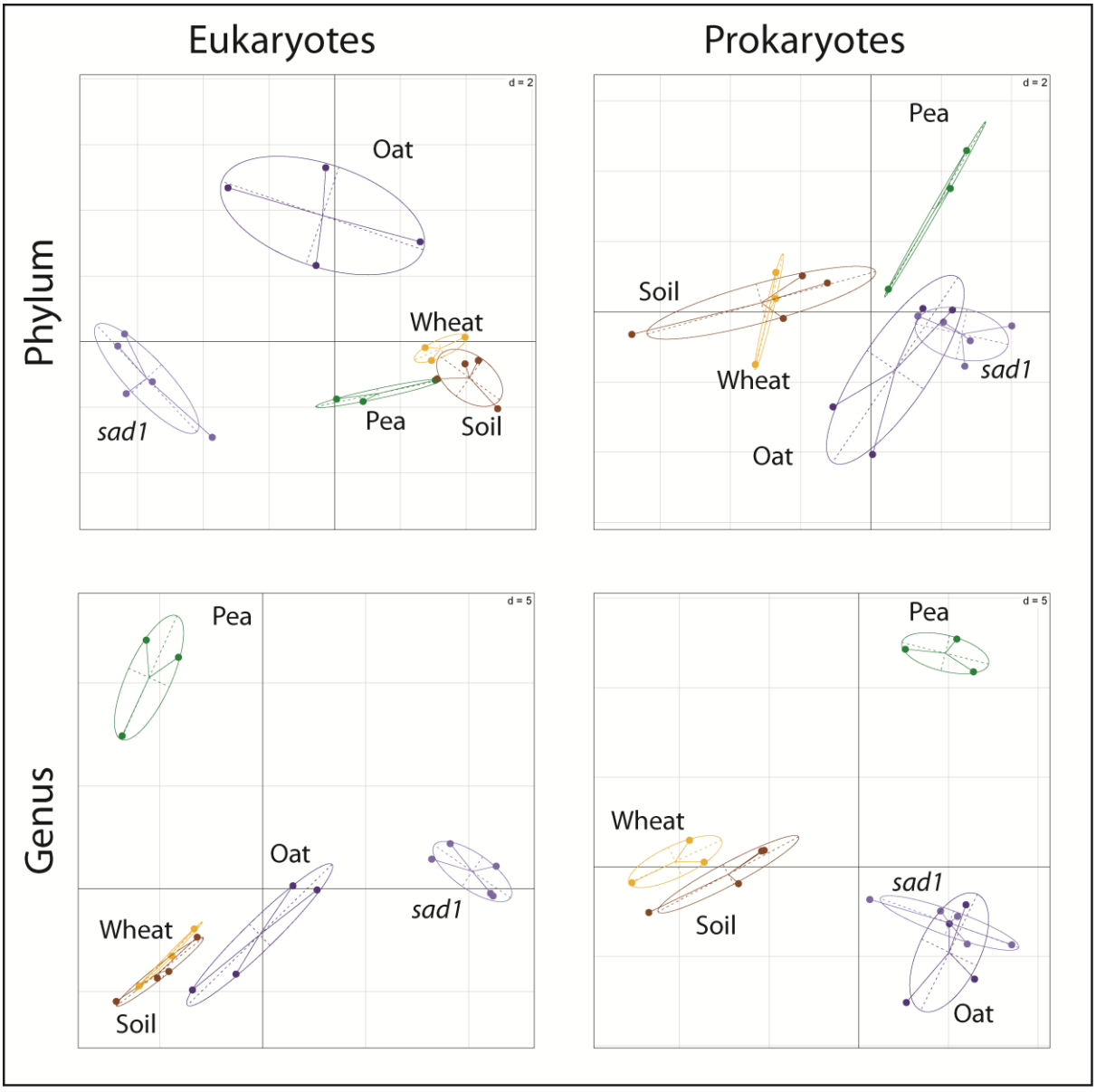


Figure 3.7: Taxonomic differences between rhizospheres as revealed by between-classes PCA. The ade4 R package (Dray and Dufour, 2007) was used to plot sample locations on two principal components based on data from prokaryotic and eukaryotic communities at the phylum and genus levels. The centre of gravity for each class of rhizosphere is given by the mid-point of all samples and the ellipse covers ±1 standard deviation of the samples belonging to that class. The closer two ellipses appear, the greater the similarity in their community profile. The statistical significance of the between-classes PCA, as determined

by a Monte-Carlo test [n = 999], gave P<0.01 in all four cases indicating the existence of highly significant differences between the groups

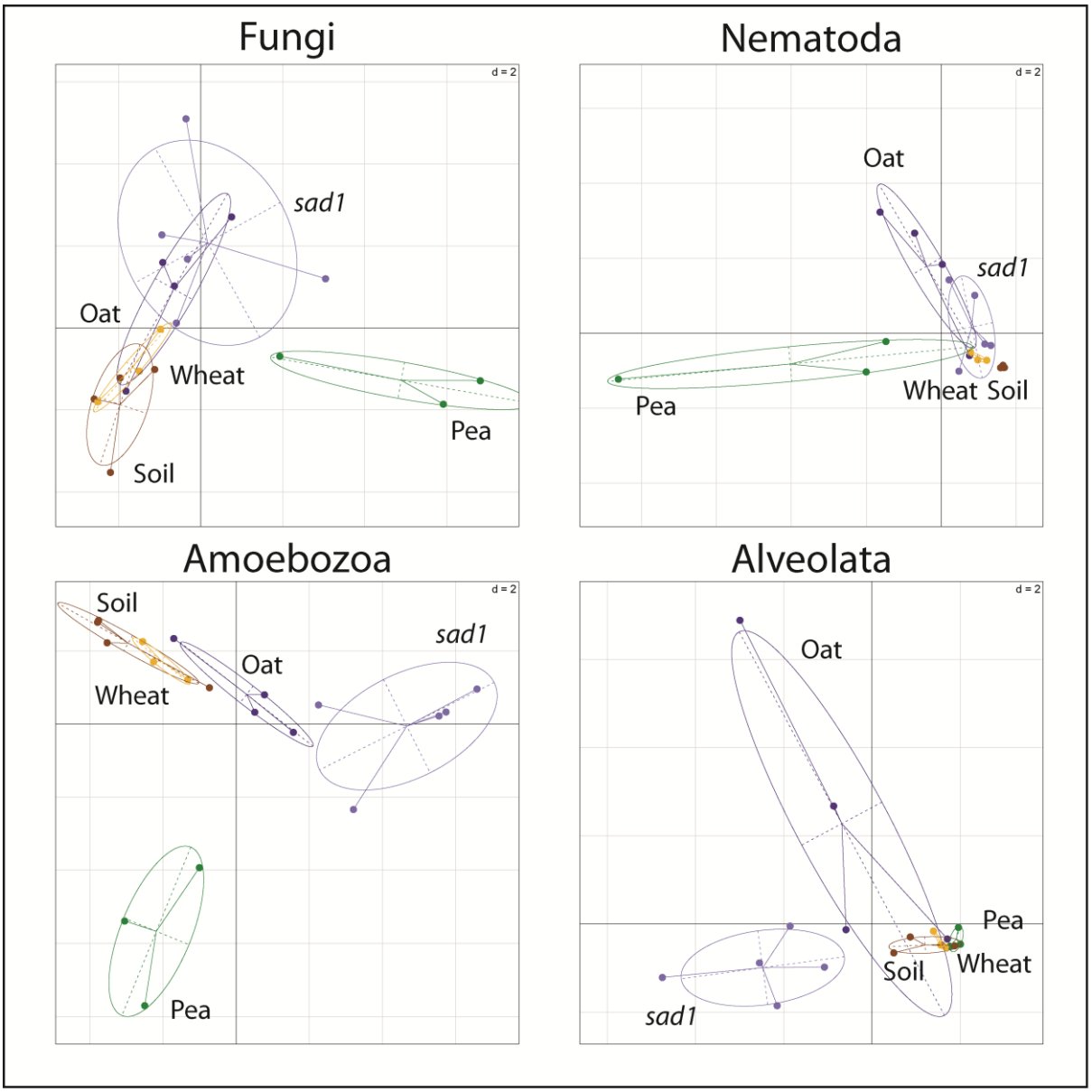


Figure 3.8: Genus level differences between major eukaryotic groups (Figure 3.1C) in the rhizospheres as revealed by between-classes PCA. The ade4 R package (Dray and Dufour, 2007) was used to plot sample locations on two principal components based on data from prokaryotic and eukaryotic communities at the phylum and genus levels. Other details are as Figure 3.7. The statistical significance of the between-classes PCA, as determined by a Monte-Carlo test [n = 999], gave P<0.004 for Fungi, P<0.002 for Nematoda, P<0.001 for

Amoebozoa and P<0.035 for Alveolata, indicating the existence of highly significant differences between the groups.

Rarefaction analyses of prokaryotic communities (Figure 3.9) showed that phylum level diversity of bulk soil and rhizospheres was similar, while at genus level the oat rhizosphere had slightly reduced diversity compared to the other environments. Diversity of eukaryotic phyla and genera in rhizosphere samples was greater than in bulk soil. Eukaryotic diversity was surprisingly high, equalling or exceeding that of prokaryotes at phylum level. These observations are interesting because the rhizosphere is a soil derived environment selected by a plant and is considered less diverse. The increased abundance of food sources (e.g. root exudates and bacteria) for some eukaryotes might have allowed them to attain population numbers higher than they would in bulk soil, making them more likely to be detected. If this is the case, further sequencing should result in rhizosphere rarefaction curves reaching asymptote before that of bulk soil. An important consideration here is the measurement of activity rather than abundance. It is entirely possible that diversity based on abundance is different from the based on activity. Not all soil bacteria would colonise the rhizosphere (reduced diversity), but those that succeeded would have higher activity and growth rates due to the increased availability of nutrients. The data here suggest there might be low abundance prokaryotes whose activity is greatly enhanced in the rhizosphere. Previous estimates of bacterial diversity suggest almost different 700 genera per gram of soil (Roesch *et al.*, 2007), greater than that seen in the current samples. Although differences in sample and analysis methods prevent a direct comparison with this study, the rarefaction curves do not reach asymptote at genus level, indicating more sequencing would detect additional genera. Interestingly, the rarefaction analyses (Figure 3.9) do not correspond well with the PCA analyses (Figure 3.7), with the exception of the eukaryotic communities of the two oat genotypes. This indicates that diversity is not the only driver of community structure. For example, two environments could be equally diverse in terms of

numbers of different genera or phyla but they could be made up of entirely different taxa. It is likely that the diversity in the rhizospheres is maintained by the variety and levels of carbon available to the microbes from the plants, but the specific compounds are being utilised by very different microbial taxa. With respect to the sad1 oat, the lack of avenacin in the rhizosphere appears to have allowed an increase in diversity, indicating that avenaci suppresses many eukaryotes in the rhizosphere of the wild-type.

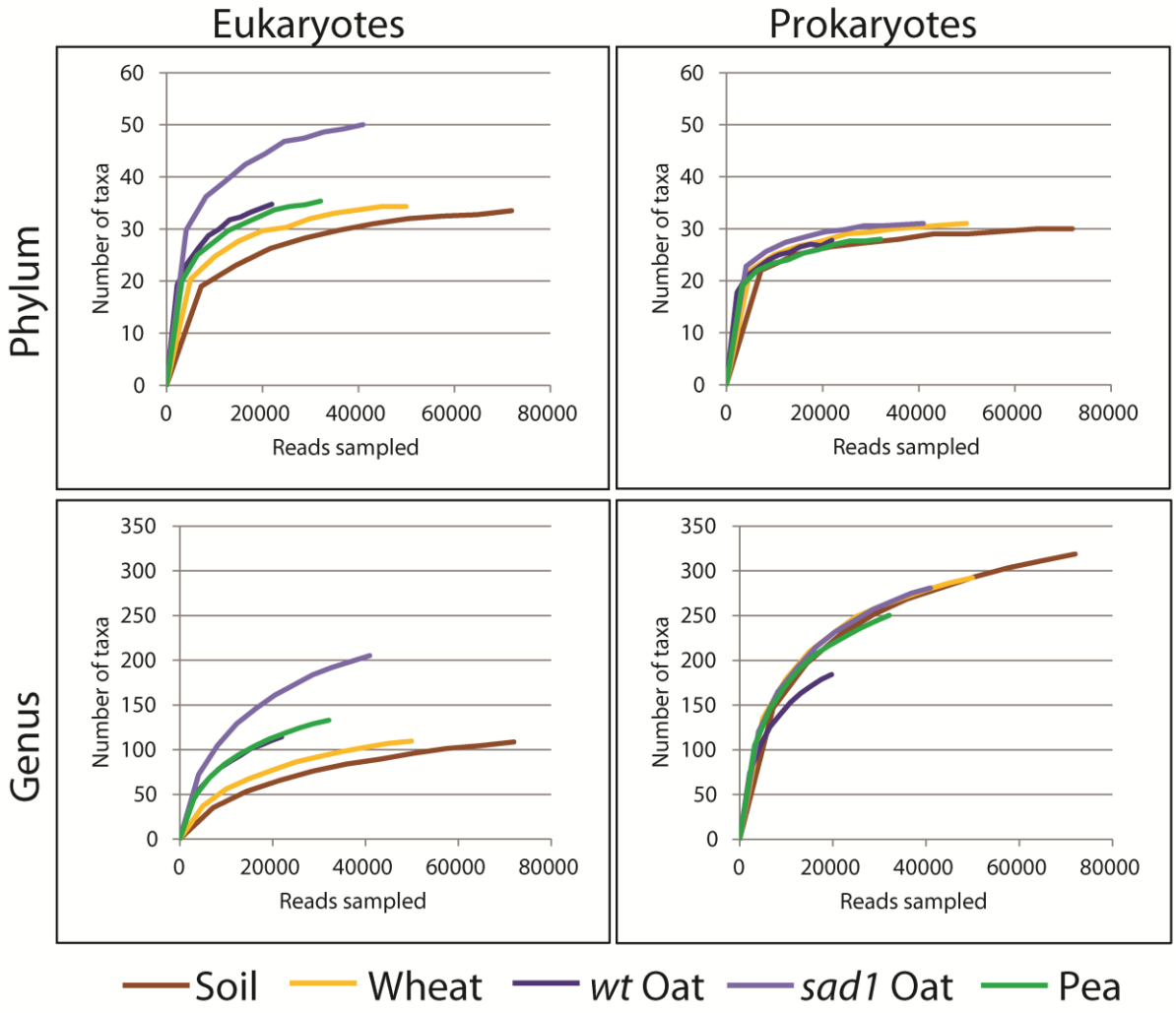


Figure 3.9: Rarefaction analysis of prokaryote and eukaryote communities at the phylum and genus levels. Values are means of biological replicates where n=3 for wheat and pea, n=4 for soil and wt oat and n=5 for sad1 oat.

3.3.6 Highly abundant microbes in soil and rhizospheres

In bulk soil and all plant rhizospheres, the most abundant prokaryotes were the Proteobacteria, followed by Actinobacteria, Firmicutes, Acidobacteria, Planctomycetes and Bacteroidetes (Figure 3.6b). Cyanobacteria were also highly abundant in bulk soil. These major groups were identified in a metatranscriptomic study of soil (Urich *et al.*, 2008), and have been well represented in PCR based analyses of soil (Roesch *et al.*, 2007) and the rhizosphere (Bulgarelli *et al.*, 2012; Inceoglu *et al.*, 2011; Lundberg *et al.*, 2012). In our study the Acidimicrobiales, Actinomycetales and Bacillales contributed most to their respective phyla (Figure 3.10b), while representation of Proteobacteria was more diverse. Major contributions came from the Myxococcales and unclassified Deltaproteobacteria, as well as Alpha-, Beta-, and Gamma- subdivisions, namely the Rhizobiales, Burkholderiales and Pseudomonadales respectively (Figure 3.10a). These taxa are well known for their interactions with plants and have previously been detected in soil (Janssen, 2006), rhizosphere (Lu *et al.*, 2006) and phyllosphere (Yashiro *et al.*, 2011) environments.

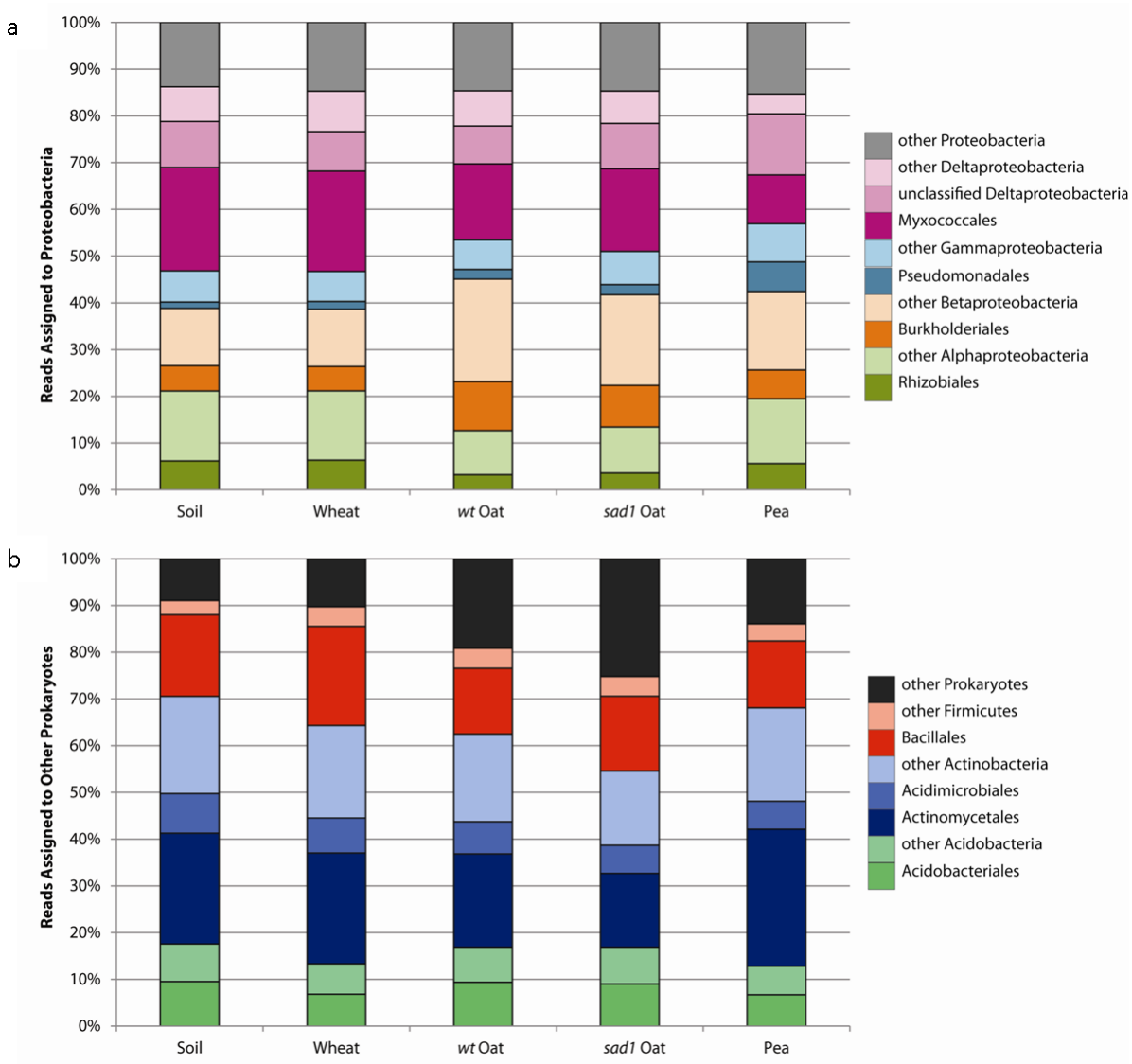


Figure 3.10: Contribution of prokaryotic orders and classes to their respective phyla. Proteobactetia (a) are derived from the Proteobacteria wedge of Figure 3.6b, while other prokaryotes (b) are derived from the remaning wedges of Figure 3.6b. Values are means of biological replicates where n=3 for wheat and pea, n=4 for soil and wt oat and n=5 for sad1 oat.

The majority of eukaryotic sequences were derived from Fungi and Nematoda, with some contribution from Amoebozoa and Alveolata (Figure 3.6c). In addition to eukaryotes being over 5fold more abundant in oat and pea rhizospheres compared to both wheat rhizosphere and soil (Figure 3.6a), the proportion of major eukaryotic groups was different.

Nematoda were more abundant in all rhizospheres compared to soil, while the pea rhizosphere was highly enriched for fungi (Figure 3.6c).

3.3.7 Indepdendent comparison of metatranscriptomic data with qPCR

To test reproduciblity of the total rRNA sequence data, qPCR was performed on DNA from samples of Bawburgh soil (Baw4), and the wheat, oat and pea rhizospheres, from a separate harvest using universal primers for 16S and 18S rRNA genes. The Ct values (Figure 3.11a) were used to calculate relative abundance of bacteria and eukaryotes (Figure 3.11b). Bulk soil was comprised of 3.73% eukaryotes and wheat only 7.22%, while oat and pea had much higher proportions at 32.07 and 38.38 respectively. This was highly consisent with the rRNA data (Figure 3.6a). The 18S primer would also amplify plant DNA, so the qPCR data was compared against the RNA data without removing Viridplantae sequences. Cellular organisms and Archaea were also ignored and % abundance recalculated accordingly. The RNA and DNA values for relative abundance correllated well (Figure 3.11c) (R^2=0.994 for both bacteria and eukaryotes). This demonstrates a high level of reproducibility for the rhizosphere effect, at least at the kingdom level, as although the same plant cultivars were used, the soil and analyses were different. By calculating the ratio of rRNA to rDNA, the activity of organisms can be at least approximately normalised to their abudance (Figure 3.11d). In bulk soil, both bacteria and eukaryotes were represented similarly in the two data sets, with ratios close to 1. Interestingly however, eukaryotes in the rhizosphere were on average 50% more abundant in the DNA data than in the RNA data. This suggests that although are they more abundant, they are less active, even for the wheat rhizosphere. Bacteria in contrast had ratios above 1, suggesting they

are more active at least relative to eukayotes.

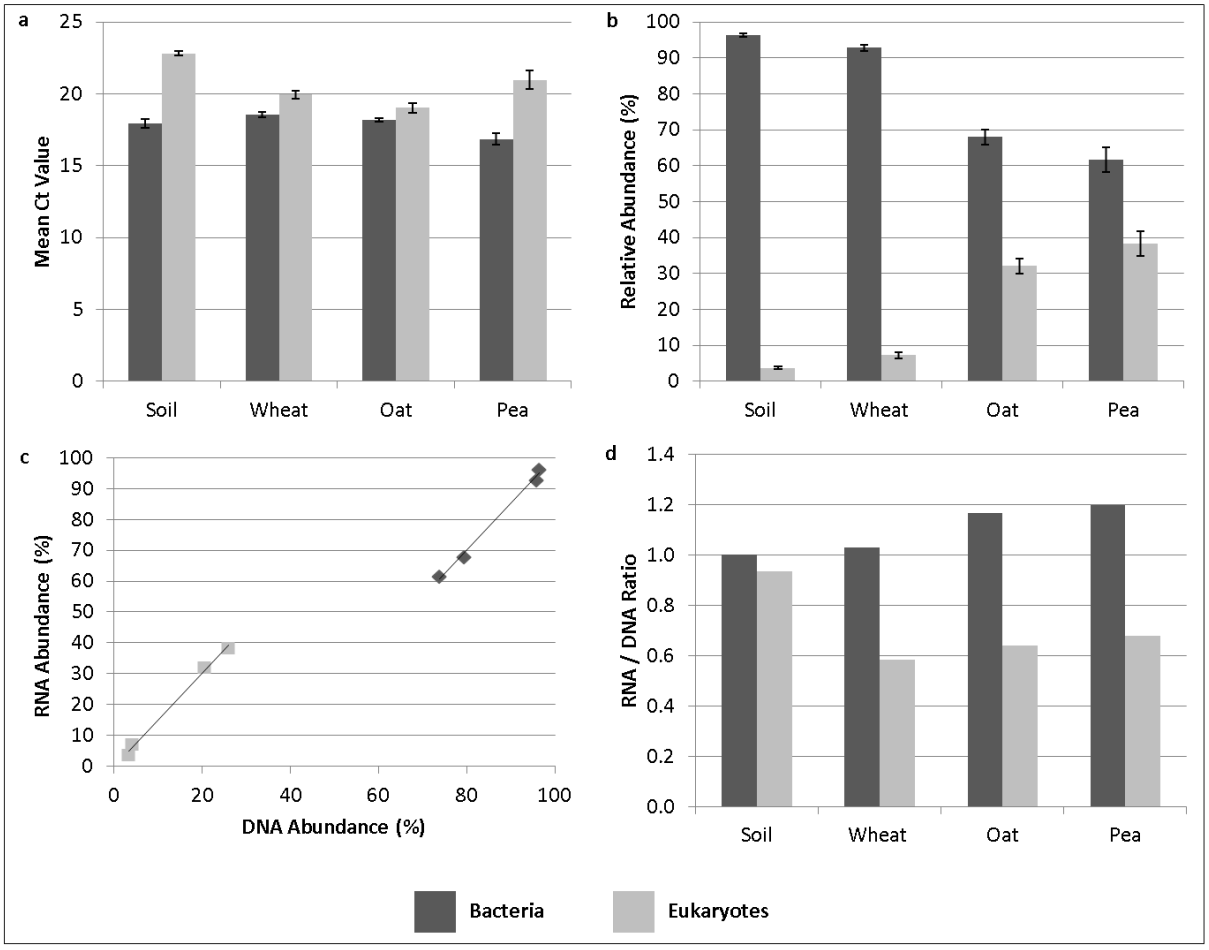


Figure 3.11: (a) Quantitative PCR of bacterial 16S and eukaryotic 18S rRNA genes. (b) Relative abundnance of bacteria and eukaryotes calculated from qPCR data. (c) Comparison of data from metatranscriptomic RNA with qPCR DNA. (d) Expression ratio of rRNA transcripts to genes. All values are means of3 technical replicates of 4 biological replicates, and error bars are ±1SEM where shown.

3.3.8 Plant selection of microbes

Between-classes PCA of total community structure showed that plants had specific effects on their rhizospheres. The disimiliarity in community structure between a rhizosphere and bulk soil can largely be attributed to those taxa most strongly selected or depleted by the

plant. Here, taxa were considered strongly selected if they were ≥5fold more abundant in a rhizosphere compared to bulk soil (Tables 2 and 3). Full lists of differentially abundant taxa can be found in Appendix Tables A1 through to A6. Some information on close relatives of the taxa enriched in the different rhizospheres compared to soil was available. However, caution is required when interpreting what lifestyles and abilities the different taxa possess. Not all species in a genus will be able to carry out a particular process described in one species. In addition, there is no direct indication of particular processes being carried out, only the activity of the organism. The following observations of enriched taxa and their functional traits are therefore speculative.

Prokaryotic taxa most strongly selected by wheat included *Dyadobacter*, Fibrobacteraceae and *Verrucomicrobium*. Firmicutes, including *Bacillus* and *Lysinibacillus* were also enriched, as were *Catellatospora*, which are known to associate with plants (Saracchi *et al.*, 2004), *Chitinophaga*, which have the ability to degrade chitin (Sangkhobol and Skerman, 1981), *Tetrasphaera*, which can accumulate polyphosphate (Nguyen *et al.*, 2011), and the methanotrophic thermophile *Methylocaldum* (Eshinimaev *et al.*, 2004). The enrichment of a thermophile is surprising. Perhaps *Methylocaldum* are able to grow at a range of temperatures, or this particular species is not a thermophile. This highlights that while naming microbes with regard to their lifestyle can be useful, it can also be deceptive. Strongly enriched eukaryotes included the bacterivorous nematode *Acrobeloides*, and Eurotiomycete fungi.

In the oat rhizosphere, Actinobacteria were depleted, but an unclassified group of Actinobacteridae was strongly enriched. Unclassified members of taxonomic groups are particularly important targets for future isolation and study. It is only by isolation that they can be understood in detail. Members of the Comamonadaceae and Pseudomonadaceae, both known to associate with plants were enriched, as were the Euryarcheaota. The

Euryarcheota are a broad phylum of archaea, containing all known methanogens, as well as some thermophiles and halophiles. Though it is likely a more specific group of Eurarcheaota were enriched in the oat rhizosphere rather than the whole phylum, this is not possible to tell from the current data. Strongly selected eukaryotes included Euglenozoa and the Amoebozoa *Glaeseria* and *Leptomyxa*.

The pea rhizosphere was strongly enriched for *Masillia*, *Dyadobacter*, *Flavobacterium* and *Streptomyces*. *Masillia* are consistently detected in studies of rhizosphere microbiomes and some members have shown production of siderophores and auxins and displayed antagonism against the oomycete plant pathogen *Phytophthora infestans in vitro* (Ofek *et al.*, 2012). *Streptomyces spp* are well known for their production secondary metabolites, including antibacterials, antifungals and nematicidals, and thus may play a role in protecting plants from pathogens. They have also been shown to increase iron availability and nodule size in pea plants (Tokala *et al.*, 2002). Other taxa selected included *Azospira,* which can fix nitrogen, *Kineosporia*, which include plant associated species (Pagani and Parenti, 1978), known plant growth promoters *Stenotrophomonas* (Ryan *et al.*, 2009) and *Variovorax* (Han *et al.*, 2011). Strains of *Variovorax*, and also *Flavobacterium* have been isolated from soil adjacent to N_2-fixing nodules of soybean (*Glycine max*), were shown to oxidise H_2, a bi-product of N_2 fixation (Hunt and Layzell, 1993), and promote plant growth (Maimaiti *et al.*, 2007). *Flavobacterium johnsoniae* strains also produce indoles and the anti-oomycete compound 2,4-di-tert-butylphenol, which contributes to their ability to protect pepper (*Capsicum spp*) plants from *Phytophthora capsici* (Sang and Kim, 2012).

Eukaryotes strongly selected by pea included the ciliate *Bresslaua*, flagellate *Dimastigella,* and root-knot nematode *Meloidogyne*, in addition to numerous fungi. These included *Tetracladium*, *Fusarium* and *Exophiala,* which have been associated with diseased peas (Yu *et al.*, 2012), as well as the mycorrhizal Glomeromycota. The latter is of particular

significance given the shared developmental pathways of *Rhizobium*-legume and mycorrhizal symbioses (Stracke *et al.*, 2002), hinting that legumes may have enhanced associations with mycorrhiza compared to non-legumes. There were increases in abundance of Glomeromycota in wheat and oat rhizospheres, but these differences were not statistically significant. It would be interesting to compare the abundances of Glomeromycota in the rhizospheres of a number of both legumes and non-legumes, using targeted qPCR for example, to determine if this is the case.

Selection of other taxa with metabolic capabilities potentially important in rhizosphere colonisation was observed. Cellulolytic bacteria such as Fibrobacteres (Ransom-Jones *et al.*, 2012) and *Cellvibrio* were selected by both wheat and oat. This suggested the presence of plant cell wall material in the cereal rhizospheres, which is often overlooked in the context of shaping the rhizosphere microbiome despite providing a large source of carbon (Dennis *et al.*, 2010) as well as a surface for attachment (Bulgarelli *et al.*, 2012). The Clostridiales enriched in the wheat rhizosphere may also be capable of anaerobic degradation of cellulose, within anaerobic niches of the generally aerobic soil. The cereals also selected for unclassified Methylophilaceae, which can reduce nitrate, which was high in soil used here (see Appendix table A33), in the presence of methanol (Kalyuhznaya *et al.*, 2009). Methanol is produced by demethylation of pectin in plant cell walls (Galbally and Kirstine, 2002), and metabolism of C1 compounds has been observed during bacterial colonisation of the phyllosphere (Delmotte *et al.*, 2009) and rhizosphere (Knief *et al.*, 2012). The *Methylocaldum* enriched in the wheat rhizosphere would also be capable of methanol utilisation. Enrichment of other methylotrophs in the pea rhizosphere, including Beijerinckiaceae, *Variovorax,* and the archaeon *Methanosarcina* was also observed, emphasising the role of C1 metabolism in rhizosphere colonisation. Aerobic methanol oxidising bacteria possessing methanol dehydrogenase (*mxaF*) are diverse, with isolates representing Alpha-, Beta-, and Gammaproteobacteria, Verrucomicrobia, Firmicutes, and

Actinobacteria (Kolb, 2009). Additionally, alternate methanol oxidising enzymes may be employed by rhizosphere microbes, such as catalases or peroxidises (Harrington and Kallio, 1960). Interestingly, formate dehydrogenase genes were up-regulated in *Rhizobium leguminosarum* and *Pseudomonas aeruginosa* during rhizosphere colonisation (Ramachandran *et al.*, 2011) and on exposure to root exudates (Mark *et al.*, 2005) respectively. Neither bacteria are known to possess methanol dehydrogenases, although recently members of the Rhizobiales have been shown to incorporate C^{14} labelled methanol and possess divergent MxaF-like genes (Stacheter *et al.*, 2012). In many plant associated bacteria, catalases and peroxidises are involved in resistance to reactive oxygen species (ROS) produced by the plant immune response during colonisation. Both these enzymes can couple detoxification of ROS with oxidation of methanol, which might contribute to the success of these microbes in the rhizosphere. It is also feasible that ROS directly oxidise methanol in the rhizosphere independently of the bacteria.

Other taxa were selected by more than one plant suggesting they may be more general rhizosphere colonisers. Among prokaryotes, only the Verrucomicrobiaceae were more abundant in all rhizospheres compared to bulk soil. There are few cultured representatives of this group, and thus they remain poorly understood, however there are some methylotrophic representatives (Kolb, 2009). Fungi, particularly Ascomycota, were selected by oat and strongly selected by pea, while fungi closely related to those that form lichens (Chaetothyriomycetidae), bacterivorous protozoa (Cercomonadida and Kinetoplastida) and nematodes (Criconematoidea) were selected by all plants. No eukaryotic taxa were statistically less abundant in a rhizosphere sample compared to bulk soil.

The production of antimicrobial compounds by plants and competition or inability to use plant derived carbon sources inevitably leads to some taxa being depleted in the rhizosphere relative to soil. Sphingomondales, Roseomonas, Gemmatinomas and

potentially plant pathogenic Xanthomonadales were less abundant in the wheat rhizosphere. Actinomycetes, Methylocystaceae and Rhizobiales were less abundant in the oat rhizosphere, while sulphate-reducing Desulphobacterales, bacterivorous Bdellovibrionaceae, and nitrite oxidising *Nitrospira* were less abundant in the pea rhizosphere.

Photosynthetic Noscocales, including *Anabaena*, were depleted in all rhizospheres. Because plants provide the major carbon flux into the rhizosphere, microbial autotrophs may lose their selective advantage. It is possible that a plant canopy prevents light reaching the soil, reducing the ability of such organisms to photosynthesise. This is unlikely however, as plants were relatively small at harvest, and the thin leaves of the cereals in particular did not cast a significant shadow on the soil. Additionally, the Chloroflexaceae, which are anyoxygenic photoautotrophs, were enriched in the oat rhizosphere. Over the course of the four weeks of plant growth, a green mat was observed on the bulk soil samples. While only subsurface bulk soil was harvested, some of these autotrophs may have been collected. Future plant growth protocols have included a layer of perlite to reduce colonisation of surface soil by these organisms.

A range of taxa, including Alphaproteobacteria, Rhodobacterales, Bacillaceae and Acidimicrobiales, were depleted in both oat and pea rhizospheres compared to soil. Oat and pea both produce antimicrobial saponins (Crombie and Crombie, 1986; Morita *et al.*, 2000), while wheat does not (Haralampidis *et al.*, 2001). However, this is unlikely to explain these differences in community structure because these taxa were also less abundant in the *sad1* oat mutant compared to soil. A more likely explanation could be that these bacteria were grazed upon by nematodes and protozoa. Introduction of the protozoan predator, *Acrobeloides maximus*, into the *Arabidopsis thaliana* rhizosphere altered the relative abundance of specific bacterial groups (Rosenberg *et al.*, 2009). Nematodes and

protozoa were more abundant in oat and pea rhizospheres than in wheat rhizosphere or bulk soil, and their feeding preference may be the reason behind the depletion of some bacterial taxa.

Table 3.2: Prokaryotic taxa strongly (P ≤0.05, ≥5fold) selected or depleted in the rhizosphere of wheat, *wt* oat and pea relative to bulk soil. A full list of differentially abundant taxa can be found in Appendix Table A1 to A3.

Taxonomic group	**Soil mean (% of reads)**	**Wheat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Dyadobacter	0.00	0.01	0.042	6.14
Fibrobacteraceae	0.00	0.02	0.007	5.90
Verrucomicrobium	0.00	0.01	0.048	5.61
Roseomonas	0.02	0.00	0.027	0.20
Anabaena	0.03	0.01	0.003	0.18
Anabaena lemmermannii	0.03	0.01	0.002	0.17
Taxonomic group	**Soil mean (% of reads)**	**Oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
unclassified Actinobacteridae	0.01	0.29	0.000	29.77
Peptostreptococcaceae	0.00	0.03	0.001	8.94
alpha proteobacterium FH6	0.06	0.01	0.005	0.20
Bacillus subtilis	0.05	0.01	0.032	0.14
Nostocaceae	0.06	0.01	0.006	0.13
Nostocales	0.08	0.01	0.000	0.11
unclassified Intrasporangiaceae	0.02	0.00	0.008	0.11
Anabaena	0.03	0.00	0.000	0.00
Anabaena lemmermannii	0.03	0.00	0.000	0.00
Taxonomic group	**Soil mean (% of reads)**	**Pea mean (% of reads)**	**P value**	**Fold change (vs soil)**
Flavobacterium johnsoniae	0.00	0.01	0.025	Absent in soil
Flavobacterium sp. DB3.3-15	0.00	0.01	0.029	Absent in soil
Dyadobacter koreensis	0.00	0.03	0.006	84.58
Chryseobacterium sp. HX31	0.00	0.01	0.045	81.02
Dyadobacter	0.00	0.10	0.038	58.00
Massilia	0.00	0.01	0.006	34.56
Methanosarcina barkeri str. Fusaro	0.00	0.01	0.034	14.46
Methanosarcina	0.00	0.01	0.045	10.44
Streptomyces	0.03	0.22	0.045	7.61
Peptostreptococcaceae	0.00	0.02	0.014	5.97
Herpetosiphon	0.00	0.02	0.002	5.84
Pseudoxanthomonas	0.01	0.03	0.018	5.64
Kribbella	0.00	0.02	0.044	5.60
Flavobacterium	0.06	0.32	0.047	5.02
Nitrosospira	0.02	0.00	0.011	0.10

Table 3.3: Eukaryotic taxa strongly (P=≤0.05, ≥5fold) selected or depleted in the rhizosphere of wheat, *wt* oat and pea relative to bulk soil. A full list of differentially abundant taxa can be found in Appendix Tables A4 to A6.

Taxonomic group	**Soil mean (% of reads)**	**Wheat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Paratylenchidae	0.00	0.02	0.000	24.45
Paratylenchus	0.00	0.02	0.000	24.45
Paratylenchus dianthus	0.00	0.02	0.000	24.45
Tylenchulidae	0.00	0.02	0.000	24.45
Criconematoidea	0.00	0.03	0.006	16.29
Cephalobidae	0.00	0.06	0.027	12.68
Cephaloboidea	0.00	0.06	0.027	12.68
Acrobeloides	0.00	0.05	0.025	12.02
Acrobeloides maximus	0.00	0.05	0.025	12.02
Rhabditida	0.01	0.07	0.044	10.47
Chaetothyriomycetidae	0.00	0.01	0.000	6.46
Eurotiomycetes	0.00	0.01	0.004	6.14
Chromadorea	0.03	0.19	0.001	5.81
Nematoda	0.03	0.19	0.001	5.75
Pseudocoelomata	0.03	0.20	0.001	5.58
Taxonomic group	**Soil mean (% of reads)**	**Oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Belonolaimidae	0.00	0.01	0.050	Absent in soil
Tylenchorhynchus	0.00	0.01	0.050	Absent in soil
Tylenchulidae	0.00	0.04	0.002	47.46
Paratylenchidae	0.00	0.03	0.000	36.58
Paratylenchus	0.00	0.03	0.000	36.58
Paratylenchus dianthus	0.00	0.03	0.000	36.58
Endopterygota	0.00	0.06	0.039	30.63
Criconematoidea	0.00	0.05	0.001	26.36
Diplonemida	0.00	0.03	0.044	20.97
Sphenomonadidae	0.00	0.04	0.018	16.39
Sphenomonadales	0.01	0.07	0.007	12.24
Chaetothyriomycetidae	0.00	0.02	0.010	10.61
Cnidaria	0.00	0.02	0.049	10.43
Heteromitidae	0.00	0.02	0.000	9.22
Euglenida	0.01	0.08	0.001	8.61
Leptomyxa	0.00	0.03	0.007	7.48
Leptomyxa reticulata	0.00	0.03	0.007	7.48
Leptomyxidae	0.00	0.03	0.007	7.48
Euglenozoa	0.04	0.33	0.003	7.32
Kinetoplastida	0.03	0.19	0.013	6.37
Bodonidae	0.03	0.16	0.020	6.22
Glaeseria	0.02	0.09	0.025	5.78
Glaeseria mira	0.02	0.09	0.025	5.78
Cercomonadidae	0.01	0.08	0.009	5.76
Cercomonadida	0.02	0.11	0.001	5.73
Taxonomic group	**Soil mean (% of reads)**	**Pea mean (% of reads)**	**P value**	**Fold change (vs soil)**
Tylenchulidae	0.00	0.06	0.008	79.70

Paratylenchidae	0.00	0.06	0.007	78.59
Paratylenchus	0.00	0.06	0.007	78.59
Paratylenchus dianthus	0.00	0.06	0.007	78.59
Hypocreales	0.02	1.78	0.017	73.49
Hypocreomycetidae	0.03	1.82	0.016	67.21
Meloidogyne	0.02	1.13	0.040	63.19
Meloidogyne incognita	0.02	1.13	0.040	63.19
Meloidogynidae	0.02	1.13	0.040	63.19
Meloidogyninae	0.02	1.13	0.040	63.19
Chromadorea	0.03	1.88	0.042	58.04
Nematoda	0.03	1.89	0.041	57.65
Criconematoidea	0.00	0.10	0.002	55.84
Tylenchoidea	0.02	1.15	0.041	55.61
Tylenchina	0.02	1.25	0.034	55.06
Pseudocoelomata	0.03	1.91	0.041	54.55
Tylenchida	0.02	1.33	0.035	54.46
mitosporic Hypocreales	0.01	0.33	0.013	31.22
Fusarium	0.01	0.32	0.012	30.97
Fusarium oxysporum	0.01	0.32	0.012	30.97
Fusarium oxysporum f. sp. lycopersici	0.01	0.32	0.012	30.97
Fusarium oxysporum f. sp. lycopersici 4286	0.01	0.32	0.012	30.97
Fusarium oxysporum species complex	0.01	0.32	0.012	30.97
Sordariomycetes	0.11	2.94	0.029	26.33
Eurotiomycetes	0.00	0.05	0.041	23.16
Chaetothyriomycetidae	0.00	0.03	0.029	21.67
Pezizomycotina	0.19	3.91	0.027	20.85
Ascomycota	0.27	5.18	0.028	19.52
Exophiala	0.00	0.02	0.012	18.86
Exophiala oligosperma	0.00	0.02	0.012	18.86
mitosporic Herpotrichiellaceae	0.00	0.02	0.012	18.86
Chaetothyriales	0.00	0.03	0.010	18.85
Herpotrichiellaceae	0.00	0.02	0.007	17.90
Dikarya	0.32	5.45	0.026	16.81
Bilateria	0.15	2.15	0.038	14.16
Eumetazoa	0.16	2.16	0.037	13.40
Pezizales	0.00	0.01	0.005	12.64
Pezizomycetes	0.00	0.01	0.005	12.64
Metazoa	0.17	2.18	0.036	12.57
Fungi/Metazoa group	0.89	9.64	0.022	10.79
Fungi	0.67	7.15	0.020	10.72
Glomus	0.00	0.04	0.001	10.18
Glomeraceae	0.00	0.04	0.000	10.09
Glomerales	0.00	0.04	0.000	10.09
Glomeromycetes	0.01	0.05	0.001	9.70
Glomeromycota	0.01	0.05	0.001	9.70
mitosporic Ascomycota	0.01	0.07	0.007	9.53
Tetracladium	0.01	0.07	0.012	9.13
Dimastigella	0.01	0.03	0.039	6.15
Dimastigella trypaniformis	0.01	0.03	0.039	6.15
Colpodida	0.01	0.03	0.005	5.72
Colpodea	0.01	0.04	0.005	5.66
Cyrtolophosidida	0.00	0.01	0.012	5.51
Bresslaua	0.00	0.02	0.023	5.42
Bresslaua vorax	0.00	0.02	0.023	5.42

Colpodidae	0.00	0.02	0.023	5.42

3.3.9 Comparison of the wild-type oat with the sad1 oat mutant

The proportion of sequences derived from Bacteria was similar for both oat genotypes at 77.3% for wild-type oat rhizosphere and 73.7% for the *sad1* oat rhizosphere (Figure 3.6a). Both rhizospheres were significantly different from bulk soil (p= 0.0201 and 0.0118 respectively) but not from each other (p=0.590). Eukaryotes made up 2.8% of the bulk soil community, 16.6% of the oat community and 22.0% of the sad1 community (Figure 3.6a). These differences were significantly different from bulk soil (p= 0.0380 and 0.0134 respectively) but not from each other (p= 0.51). Archaea were represented at 0.6% for the wild-type oat rhizosphere and 0.7% for the *sad1* oat rhizosphere (Figure 3.6a). At the kingdom level, there were no significant differences between the oat genotypes.

Rarefaction analysis showed eukaryotic diversity in the *sad1* rhizosphere to be greater than that of the wild-type and bulk soil at both phylum and genus level, while prokaryotic diversity was unaltered (Figure 3.9). The wild-type oat rhizosphere likely contains a higher proportion of avenacin resistant organisms than the *sad1* rhizosphere. Although this is difficult to test directly, root endophytic fungi of a wild-type oat where found to be almost all resistant to avenacin (Carter *et al.*, 1999). The constant arms race between plant defence and microbial resistance may result in a stabilisation of diversity in any given rhizosphere. Perturbation of one component, i.e. loss of avenacin production, may shift the balance, possibly explaining the rise in eukaryotic diversity within the *sad1* rhizosphere. Eukaryotic diversity was greater than that of prokaryotes at phylum level, and about half of that of prokaryotes at genus level. This observation is likely due to the divisive nature of eukaryotic taxonomy. While the 16S rRNA gene is considered the bench-mark for prokaryotic phylogenetics, the 18S rRNA gene of eukaryotes is much more highly

conserved. Because of this, it is difficult to discriminate between finer taxonomic differences using this marker. Instead, the ITS region between the small and large rRNA gene subunits has been used to generate eukaryotic phylogenetic trees. The ITS is not under selection pressure, although the tRNA genes sometime located within it are, so sufficient diversity has been introduced in the relatively short evolutionary history of eukaryotes.

Differentially abundant taxa in the *sad1* rhizosphere compared to both wild-type and bulk soil are listed in Tables 4 and 5 and Appendix Table A7 through to A10. There were differences in abundance of some bacteria between the *sad1* oat rhizosphere and soil that were not seen between soil and the wild-type oat. These included an enrichment of *Opitutus,* found abundantly in rice paddy fields and capable of producing acetate, propionate and H_2 from plant derived polymers such as pectin (Chin and Janssen, 2002). These are substrates for methanogenic archaea, such as Methanosarcinales which were also enriched in the *sad1* rhizosphere. Interestingly *Opitutus* have shown higher growth rates when co-cultured with methanogens (Chin and Janssen, 2002), implying they may be methanotrophic, or the removal of acetate, propionate or H_2 prevents a build up that may be toxic to the *Opitutus*. Furthermore, the Methanosarcinales are also methylotrophic, and are therefore able to metabolise methanol, a breakdown product of pectin. This further reinforces the synergism between these two organisms, and is another example of the complex, multi-level interactions occurring in the rhizosphere. Bacteria specifically depleted in the *sad1* rhizosphere included Desulfobacterales, *Nitrospira*, and Candidate Divisions OP10 and WS3.

The PCA analyses (Figure 3.7) showed that at phylum and genus level, the eukaryotic communities of the two oat genotypes were distinct from each other and bulk soil, particular at genus level for the *sad1* oat. A number of eukaryotes were selected by both

oat genotypes (Tables 3 and 5) including Fungi, Euglenozoa and Cercozoa. The *sad1* mutant specifically selected a variety of eukaryotes, including Alveolata and numerous Amoebozoa, while the wild type specifically selected the plant pathogenic nematode Paratylenchus. No eukaryotes were depleted by either of the oat genotypes relative to soil, but there were differences in abundance of some eukaryotes between the oats. Neoptera, Criconematoidea, and Heteromitidae were less abundant in the *sad1* rhizosphere than wild-type. The *sad1* rhizosphere was enriched for two fungal taxa, Mucoromycotina and Pezizomycetes (Table 3.5), but there was no significant difference in overall fungal community compared to the wild-type (Figure 3.8). The fungal community in the wild-type oat rhizosphere was intermediate between the bulk soil and *sad1* mutant, which were distinct from each other. Differences in communities of Amoebozoa and Alveolata were more strongly affected in the mutant, while that of Nematoda was not altered (Figure 3.8).

Avenacins have broad-spectrum anti-fungal activity (Carter *et al.*, 1999; Maizel *et al.*, 1964) and plants defective in their production have compromised resistance to fungal pathogens (Osbourn *et al.*, 1994; Papadopoulou *et al.*, 1999). While avenacins are primarily thought to occur in the root tissue, they have been detected in soil at levels known to inhibit fungal growth. It is surprising therefore that there was little difference between the fungal community of the wild-type and the avenacin-deficient mutant. Avenacins' anti-fungal activity is attributed to their ability to bind to sterols, forming a pore that disrupts the cell membrane (Armah *et al.*, 1999). Sterols are almost exclusively eukaryotic (Desmond and Gribaldo, 2009), which may account for the shift in eukaryotic community between the two oat genotypes tested here. Perhaps avenacins have higher affinity for sterols in Amoebozoa and Alveolata than for those in fungi, or the inhibitory concentrations of avenacin are lower for these organisms than for fungi, which are protected by a chitin cell wall. Avenacins accumulate in the root endodermis at higher concentrations than they are found outside roots, suggesting they protect from fungal pathogens after infection.

There were some differences in abundance of prokaryotic taxa between the two oat genotypes (Table 3.5 and Appendix Tables A7 and A8), suggesting that avenacin may possess some antibacterial activity. Sterol production in prokaryotes is rare, and those few taxa known to produce sterols (*Pleistocystis*, *Gemmata*, *Methylococcus*, *Stigmatella*) were not differentially abundant between the two oat genotypes. The cell membranes of some prokaryotes contain functional analogues of sterols, known as hopanoids (Kannenberg and Poralla, 1999). Their biosynthesis is widespread in both Gram-positive and Gram-negative bacteria, but is not conserved within groups. For example, within the Rhizobiales, *Bradyrhizobium japonicum* produces hopanoids while *Rhizobium leguminosarum* does not. Hopanoids have been implicated in oxygen homeostasis in *Frankia* (Berry *et al.*, 1993), stress responses in *Burkholderia* (Schmerk *et al.*, 2011), and development in *Streptomyces coelicolor* (Poralla *et al.*, 2000) but not *Streptomyces scabies* (Seipke and Loria, 2009). If avenacins can bind to hopanoids they may be able to disrupt bacterial membranes. Two groups of bacteria that were more abundant in the *sad1* rhizosphere have representatives with a squalene-hopene cyclase (*shc*), indicating they may produce hopanoids. However there is currently no evidence that avenacins and hopanoids interact. While the presence of hopanoids in a membrane may make bacteria susceptible to avenacins, increased expression of a squalene-hopene cyclase gene has been related to antibiotic resistance (Sass *et al.*, 2011). This may indicate that if enough hopanoids are produced, they strengthen the membrane, or sequester avenacins sufficiently to reduce their concentration below the level at which they are disruptive to the membrane. There may also be an analogous scenario for the sterols in eukaryotic membranes.

Metabolic changes in the *sad1* mutant may affect the rhizosphere microbiome in addition to the lack of avenacin production itself. Avenacins are synthesised from an oxidosqualene cyclase, which the *sad1* line is mutated in. Two precursor metabolites, squalene and 2,3-oxidosqualene, have been measured at elevated levels in the *sad1* oat roots compared to

wild-type (Qin *et al.*, 2010). Both are precursors of membrane sterols, while squalene is also a precursor for hopanoids. Elevated levels of these metabolites may have an effect on the rhizosphere microbiome, possibly providing a carbon source for microbes, or feeding into sterol and hopanoid biosynthetic pathways. There is no evidence that they are released from the root, but their hydrophobicity may allow them to cross them membrane, or they may be released from root cells that are deposited in the rhizosphere (Dennis *et al.*, 2010).

While not able to produce sterols themselves, some Spirochaetes, intracellular animal pathogens, require host sterols for growth (Lemcke and Burrows, 1980). These bacteria were more abundant in the *sad1* rhizosphere than wild-type. Avenacins may target the sterols in the membranes of these bacteria directly, or the depletion of animal host numbers in the wild-type may have led to a reduction in abundance of these bacteria.

The ability to degrade avenacins has been documented in fungi, including root colonising endophytes (Carter *et al.*, 1999) and *Gaeumannomyces graminis* var *avenae* which can infect oat (Osbourn *et al.*, 1991). The action of avenacinase enzymes removes the sugar moieties which are required for activity (Armah *et al.*, 1999) and avenacinase mutants of *Gaeumannomyces graminis* are unable to infect oat, but retain pathogenicity to wheat (Bowyer *et al.*, 1995). There is remarkable similarity in amino acid sequence and physicochemical properties between saponin-detoxifying enzymes, but they retain their specificity (Osbourn *et al.*, 1995). In addition, the degradation products of some saponins are able to suppress the plant immune system, resulting in a twofold benefit for a pathogen possessing a saponin-detoxifying enzyme (Bouarab *et al.*, 2002) There are likely a wide variety of avenacin degrading enzymes in soils, but no taxa known to produce them were more abundant in the wild-type compared to the *sad1* mutant. To determine this would

require a targeted approach, i.e. using PCR to detect known avenacinase producers or the avenacinase genes themselves.

Table 3.4: Prokaryotic taxa strongly (P≤0.05, ≥5fold) selected or depleted in the rhizosphere of the *sad1* oat mutant relative to *wt* oat and bulk soil. A full list of differentially abundant taxa can be found in Appendix Table A7 and A8.

Taxonomic group	***wt* oat mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs *wt*oat)**
Haliea	0.00	0.01	0.008	20.79
Leptospira	0.00	0.03	0.011	17.96
Leptospiraceae	0.00	0.04	0.019	8.98
Thiotrichales	0.00	0.01	0.002	8.09
uc.Intrasporangiaceae	0.00	0.02	0.034	6.19
Cytophaga	0.00	0.01	0.044	5.81
Cyanobacteria	0.47	2.52	0.050	5.37
Gemmata-like str. JW3-8s0	0.01	0.00	0.034	0.13
Taxonomic group	**Soil mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Leadbetterella	0.00	0.01	0.02	Absent in soil
Chitinimonas	0.00	0.02	0.02	50.19
Sorangium cellulosum	0.00	0.01	0.01	15.14
Methanosarcinaceae	0.00	0.01	0.04	6.81
Methanosarcinales	0.00	0.01	0.04	6.20
Cytophaga	0.00	0.01	0.03	5.99
Bacteriovorax	0.00	0.01	0.00	5.44
Blastococcus	0.06	0.01	0.00	0.19

Table 3.5: Eukaryotic taxa strongly (P≤0.05, ≥5fold) selected or depleted in the rhizosphere of the *sad1* oat mutant relative to *wt* oat and bulk soil. A full list of differentially abundant taxa can be found in Appendix Table A9 and A10.

Taxonomic group	***wt* oat mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Aplanochytrium	0.00	0.01	0.044	Absent in *wt* oat
Pessonella	0.00	0.01	0.009	15.12
Pessonella sp. PRA-29	0.00	0.01	0.009	15.12
Thaumatomonas	0.00	0.02	0.005	14.80
Thaumatomonas seravini	0.00	0.02	0.005	14.80
Rotaliina	0.00	0.01	0.038	12.13
Cercomonadida environmental sample	0.00	0.01	0.006	11.79
environmental samples	0.00	0.01	0.006	11.79

Nuclearia	0.00	0.03	0.049	11.13
Paramoebidae	0.01	0.13	0.008	10.95
Mayorella	0.01	0.11	0.009	10.90
Mayorella sp. JJP-2003	0.01	0.11	0.009	10.90
Korotnevella	0.00	0.02	0.021	10.68
Nucleariidae	0.00	0.03	0.042	7.84
Korotnevella stella	0.00	0.01	0.019	7.26
Dactylopodida	0.02	0.14	0.010	6.65
Mucoromycotina	0.00	0.01	0.012	5.94
Pezizales	0.00	0.02	0.020	5.66
Pezizomycetes	0.00	0.02	0.020	5.66
Foraminifera	0.01	0.07	0.012	5.23

Taxonomic group	**Soil mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Anaplectus	0.00	0.02	0.047	Absent in soil
Anaplectus sp. PDL-2005	0.00	0.02	0.047	Absent in soil
Aplanochytrium	0.00	0.01	0.044	Absent in soil
Araeolaimida	0.00	0.03	0.006	Absent in soil
Cyrtolophosididae	0.00	0.02	0.043	Absent in soil
Plectidae	0.00	0.03	0.006	Absent in soil
Plectoidea	0.00	0.03	0.006	Absent in soil
Mayorella	0.00	0.11	0.005	213.18
Mayorella sp. JJP-2003	0.00	0.11	0.005	213.18
Vannella	0.00	0.04	0.010	112.45
Paraphysomonadaceae	0.00	0.03	0.013	84.24
Paraphysomonas	0.00	0.03	0.013	84.24
Prorodontida	0.00	0.05	0.007	79.90
Prostomatea	0.00	0.05	0.007	79.90
environmental samples	0.00	0.05	0.008	79.16
Prorodontidae environmental sample	0.00	0.05	0.008	79.16
unclassified Vannella	0.00	0.02	0.032	50.15
Bodo	0.00	0.02	0.013	48.30
Paraphysomonas foraminifera	0.00	0.01	0.045	47.24
Chromulinales	0.00	0.03	0.004	45.68
Glomus	0.00	0.15	0.024	43.48
Codonosigidae	0.00	0.01	0.043	43.17
Pessonella	0.00	0.01	0.006	40.59
Pessonella sp. PRA-29	0.00	0.01	0.006	40.59
Sphenomonadidae	0.00	0.10	0.003	38.79
Glomeraceae	0.00	0.17	0.022	38.63
Glomerales	0.00	0.17	0.022	38.63
Glomeromycetes	0.01	0.19	0.020	34.98
Glomeromycota	0.01	0.19	0.020	34.98
Nucleariidae	0.00	0.03	0.025	33.63
Paramoebidae	0.00	0.13	0.006	30.94
Nuclearia	0.00	0.03	0.038	30.54
Leptomyxa	0.00	0.10	0.006	28.91
Leptomyxa reticulata	0.00	0.10	0.006	28.91
Leptomyxidae	0.00	0.10	0.006	28.91
Sphenomonadales	0.01	0.16	0.001	28.69
Chrysophyceae	0.00	0.04	0.006	27.00
Glomus eburneum	0.00	0.02	0.005	25.34
Petalomonas	0.00	0.05	0.008	23.76
Petalomonas cantuscygni	0.00	0.05	0.008	23.76

Dactylopodida	0.01	0.14	0.005	23.06
Pezizales	0.00	0.02	0.007	20.37
Pezizomycetes	0.00	0.02	0.007	20.37
Choanoflagellida	0.00	0.02	0.039	18.97
Rotaliina	0.00	0.01	0.031	18.35
Euglenida	0.01	0.18	0.002	18.15
Diplonemida	0.00	0.03	0.005	17.40
Tracheleuglypha	0.00	0.04	0.038	14.50
Tracheleuglypha dentata	0.00	0.04	0.038	14.50
Glomus etunicatum	0.00	0.01	0.004	13.47
Boletales	0.00	0.01	0.008	13.05
Euglyphidae	0.01	0.07	0.022	12.42
Haptoria	0.00	0.02	0.027	11.76
Thraustochytriidae	0.00	0.02	0.012	11.01
Cryptophyta	0.00	0.02	0.046	10.66
Euglypha	0.00	0.03	0.013	10.65
Agaricomycetidae	0.00	0.01	0.021	10.32
Leptomyxida	0.02	0.23	0.005	9.74
Cyrtolophosidida	0.00	0.02	0.048	9.56
Mucoromycotina	0.00	0.01	0.002	9.54
Euglenozoa	0.04	0.39	0.001	8.87
Euglypha tuberculata	0.00	0.02	0.014	8.80
Rhynchomonas	0.00	0.03	0.016	8.78
Rhynchomonas nasuta	0.00	0.03	0.016	8.78
Cnidaria	0.00	0.01	0.008	8.77
Vannellidae	0.01	0.11	0.027	8.72
Chaetothyriomycetidae	0.00	0.01	0.036	8.53
Labyrinthulida	0.00	0.03	0.009	8.36
Glaeseria	0.02	0.12	0.009	8.11
Glaeseria mira	0.02	0.12	0.009	8.11
Thaumatomonas	0.00	0.02	0.006	8.08
Thaumatomonas seravini	0.00	0.02	0.006	8.08
Fungi/Metazoa group	0.89	7.22	0.021	8.07
Ciliophora	0.08	0.63	0.007	7.84
Dinophyceae	0.00	0.01	0.000	7.80
Intramacronucleata	0.08	0.61	0.008	7.74
Eurotiomycetes	0.00	0.02	0.024	7.32
Endopterygota	0.00	0.02	0.048	7.30
Haptorida	0.00	0.01	0.032	6.94
Neobodo	0.00	0.03	0.017	6.88
Neobodo designis	0.00	0.03	0.017	6.88
Lobosea sp. Mb_5C	0.01	0.04	0.034	6.83
Spirotrichea	0.01	0.05	0.000	6.82
Cercomonadida environmental sample	0.00	0.01	0.007	6.69
environmental samples	0.00	0.01	0.007	6.69
Paraflabellula	0.01	0.07	0.008	6.59
Paraflabellula hoguae	0.01	0.07	0.008	6.59
Oligohymenophorea	0.04	0.27	0.001	6.59
Flabellulidae	0.02	0.13	0.006	6.53
unclassified Lobosea	0.01	0.05	0.028	6.48
Cercomonadidae	0.01	0.09	0.000	6.44
Stichotrichida	0.00	0.02	0.011	6.41
Flabellinea	0.06	0.35	0.005	6.27
Stichotrichia	0.01	0.04	0.000	6.26

Korotnevella	0.00	0.02	0.040	5.96
Cercomonadida	0.02	0.12	0.000	5.96
Agaricomycetes	0.02	0.14	0.036	5.89
Kinetoplastida	0.03	0.18	0.000	5.79
Adineta	0.00	0.01	0.034	5.72
Adineta vaga	0.00	0.01	0.034	5.72
Adinetida	0.00	0.01	0.034	5.72
Adinetidae	0.00	0.01	0.034	5.72
Oxytrichidae	0.00	0.02	0.017	5.71
Plasmodiophorida	0.00	0.01	0.039	5.42
Plasmodiophoridae	0.00	0.01	0.039	5.42
Agaricomycetes incertae sedis	0.02	0.09	0.036	5.11
Bodonidae	0.03	0.13	0.000	5.05

3.4 Conclusion

Using metatranscriptomics, the rhizosphere microbiomes of several plants were characterised. Arguably, the most important organisms in an environment are those that are metabolically active and metatranscriptomic analysis of rRNA is an indicator of this community. Unlike analysis of rDNA, analysis of rRNA is not biased by gene copy number, which is hugely variable in bacteria, ranging from 1 (e.g. in *Bradyrhizobium*) to 14 (e.g. in *Bacillus*) (Klappenbach *et al.*, 2001). In fact, global changes in community rDNA copy number have been observed in environments that have been perturbed (Klappenbach *et al.*, 2000). It is thought that organisms with higher rDNA copy numbers are able to respond more rapidly to changes by translating the relevant proteins faster. The rhizosphere is considered a dynamic environment, and rapid adaptation to utilise plant carbon and resist plant defences is advantageous. There may then be global differences in the presence of rDNA gene copy number in a rhizosphere compared to bulk soil, something suggested by qPCR data here (Figure 3.11b). This has implications for the use of 16S PCR based studies of soil and rhizosphere, few of which take into account rDNA copy number variation. Analysis of RNA has been shown to be more sensitive than DNA for detecting differences in microbial communities treated with pollutants (Lillis *et al.*, 2009; Mengoni *et al.*, 2005) and for pathogen detection (Kim and Wang, 2009). Metatranscriptomic studies have focused on

enriching mRNA from oceans (Poretsky *et al.*, 2009; Shi *et al.*, 2009b), human gut (Gosalbes *et al.*, 2011), and soil (Bailly *et al.*, 2007), identifying active metabolic pathways. Soil studies have been restricted to eukaryotes (Bailly *et al.*, 2007), due to the ease of mRNA enrichment. Thus the global composition and metabolism of the rhizosphere microbiome is poorly understood.

After only 4 weeks of growth the microbiomes of three crop plants were different from each other and from bulk soil, with a profound change in the balance of prokaryotes and eukaryotes between plants. Differences in the field are likely to be greater, as crops are typically grown for several months. Oat and pea exerted strong selection on eukaryotes, while selection by wheat was much weaker. Oat and pea are used extensively as break-crops in crop rotation systems. A recent meta-analysis showed increased wheat yield after seasons of either oat or pea (Seymour *et al.*, 2012). The effect of pea was largely attributed to improved soil nitrogen status, while oat reduced disease incidence. The large shifts in rhizosphere microbiota seen here for oat and pea may also contribute to their positive effects when used in crop rotation systems (Figure 3.1).

No eukaryotic taxa were significantly less abundant in the rhizosphere compared to bulk soil (Figure 3.6a), with all rhizosphere samples showing higher eukaryotic diversity than bulk soil (Figure 3.9). This enrichment of eukaryotes, particularly nematodes and protozoa, implies more resources are available from either the plant or its microbiome. Both these groups of eukaryotes contain known grazers of bacteria, which are known to have feeding preferences. *Acanthamoeba castellanii*, for example, appears to not predate Verrucomicrobia (Rosenberg *et al.*, 2009). Both Verrucomicrobia and protozoa were enriched in the rhizospheres in the current study. Relative abundance of some taxa may have increased because they avoided predation rather than they were selected directly by the plant. This emphasises the complexity of interactions occurring in the rhizosphere, and

reinforces the need for global approaches to analysis community structure. However, the large changes seen in the eukaryotic communites did not appear to have a knock on effect of the prokaryotic communities. This was particularly so for the oat rhizospheres, and suggests that prokaryotes were buffered to some extent from changes in the eukaryotic community, arguing against complex interactions across these groups.

Insight into functional roles in environments can be provided by the presence of well characterised taxonomic groups. Cereal (wheat and oat) rhizospheres were enriched for cellulose degraders, while a legume (pea) rhizosphere was enriched for H_2-oxidisers. Importantly, H_2 is a by-product of N_2 reduction by nitrogenase (Hunt and Layzell, 1993) and is thought to drive selection of plant beneficial microbes in legume rhizospheres (Dong *et al.*, 2003). Different methylotrophs were enriched in rhizospheres, depending on whether the plant was a cereal or a legume. An obvious source of methanol in the rhizosphere is pectin from plant cell walls (Galbally and Kirstine, 2002). The presence of an organism with potential to carry out a biochemical process is not evidence for the activity of the metabolic process itself. Functional metatranscriptomics (i.e. analysis of mRNA) would be required to determine this and will provide a useful comparison.

The lab work for a metatranscriptomic study can be more time consuming than for a PCR based one, although much of the difference is due to the targeting of RNA. The costs of sequencing are the same, and the use of strand displacement amplification (WTA, Rubicon, see 2.3.4)) makes it feasible on relatively small samples. Strand displacement amplification can incorporate barcoded primers to rapidly multiplex samples ready for sequencing. Bioinformatic analysis required to interpret data obtained from a metatranscriptomic study is more challenging than that required by a simple 16S PCR based study. This is because randomly primed RNA sequences cannot be clustered into OTUs based on similarity, nor can they be assembled into contiguous sequences because of the conserved regions.

However, the approach taken here, combining a reduced representation database, USEARCH and MEGAN makes this feasible on large datasets using desktop computers. Only relative comparisons based on read counts can be made here due to unknown sequencing depth. Spiking samples with a known RNA transcript can allow calculation of depth and absolute transcript number per gram of soil (Gifford *et al.*, 2011), and will be a useful addition to this approach.

Metatranscriptomic analysis was sensitive enough to detect differences due to a single mutation in host plant. The loss of avenacin production in the *sad1* mutant oat had broad effects on the eukaryotic community, while prokaryotes were weakly affected (Figure 3.7). The diversity of eukaryotes in the *sad1* rhizosphere was consistently greater than the wild-type and bulk soil (Figure 3.9). These differences would likely have been missed by PCR based analysis of bacteria or fungi because it was the non-fungal eukaryotic community that were most strongly influenced (Figure 3.8). This finding has important implications for the use of genetic modification in crop plants. A small change host genotype can produce complex and unexpected changes in the microbiome. Programming plants to produce antimicrobials, even those considered narrow-spectrum, in an attempt to prevent disease, may have undesired effects on plant health mediated through the microbiome.

Metatranscriptomics is a powerful tool in microbial ecology and can provide an initial, comprehensive picture of community structure from an environment. If desired, targeted PCR can then be used to focus on important groups. Here, this global approach highlighted the complexity of the rhizosphere microbiome and revealed profound differences in community structure particularly at the kingdom level and between plants.

Chapter 4: Protein coding gene based community analysis of crop plant rhizosphere microbiomes

4.1 Introduction

The majority of soil and rhizosphere microbiomes have been profiled using 16S rRNA genes (Bulgarelli *et al.*, 2013), while others have utilised 16S rRNA transcripts (Kim and Wang, 2009; Lillis *et al.*, 2009; Mengoni *et al.*, 2005). The ubiquity of 16S rRNA, and its conserved and variable regions have made it particularly useful as a genetic marker. In addition, it is

not translated into protein, meaning that only nucleotide sequence similarity need be compared. This makes analyses of such sequences more straightforward than for protein coding genes. However 16S rRNA genes (rDNA) are not ideal markers. There is considerable variability in rDNA copy number within prokaryotes, ranging from 1 in *Bradyrhizobium* spp for example, to 14 in some *Bacillus* spp (Klappenbach *et al.*, 2001). The differences in rDNA copy number are not proportional to genome size, so any 16S rDNA based community profile will be biased towards those organisms with higher rDNA copy number. Knowing the number of gene copies can allow more accurate estimates of community structure (Farrelly *et al.*, 1995; Kembel *et al.*, 2012), but the problem is that this information is unknown for the majority of taxa detected in complex environments. There is evidence to suggest higher rDNA copy numbers provide a selective advantage to an organism, allowing them to adapt quicker to perturbations (Klappenbach *et al.*, 2000). This is an important consideration when comparing environments such as plant rhizosphere and soil, where rDNA copy number might have ecological consequences.

In eukaryotes, where rDNA can reach copy numbers of hundreds to thousands, the use of the equivalent small subunit 18S rDNA is less common. Also, the close evolutionary relationship between all eukaryotes means that their 18S rDNA is not discriminating enough, so the large subunit (LSU) 26S or 28S rDNA or the hypervariable internally transcribed spacer (ITS) are used to provide sufficient taxonomic detail. Use of the LSU or ITS does not overcome the problem of gene copy number variability, but the strong positive correlation between rDNA copy number and genome size in eukaryotes (Prokopowich *et al.*, 2003) somewhat reduces the bias when profiling eukaryotic communities.

Metatranscriptomes are dominated by rRNA allowing robust assessment of taxonomic composition (Ottesen *et al.*, 2011; Shi *et al.*, 2010; Turner *et al.*, 2013; Urich *et al.*, 2008). However, while rDNA sequences are some of the most abundant in metagenomes, their relative abundance is greatly reduced compared to that in metatranscriptomes. Consequently the use of rDNA alone is less favourable for taxonomic profiling of metagenomes. A number of other alternative marker genes are used for both profiling communities and for determining phylogeny. These are often wide-spread, slowly evolving, single copy, protein coding sequences such as genes encoding the β subunit of RNA polymerase (*rpo*B), heat shock and stress response proteins (HSP70, GroEL), elongation factors (EF-Tu) (Wu and Eisen, 2008). They can be used individually, but more often are

used together and in conjunction with rRNA gene data. In this study the protein coding genes of metagenomic and metatranscriptomic samples from soil and the rhizospheres of wheat, oat and pea were analysed and the taxonomic compositions inferred and compared.

4.2 Materials and methods

Seeds were planted in Bawburgh farm soil (Baw3) and grown for 5 weeks before rhizosphere soil was harvested (see 2.2.2). Nucleic acids (DNA and RNA) were extracted from wheat, oat and pea rhizospheres, plus unplanted controls, using the PowerSoil RNA isolation DNA elution accessory kits (see 2.3.1). Each soil sample was spiked with an RNA internal standard (RIS) prior to extraction (see 2.3.9). Five DNA samples from each environment were pooled, and one pooled sample from each environment was submitted to The Genome Analysis Centre (TGAC, Norwich, UK) for sequencing on the Illumina HiSeq (one lane each) (see 2.5.1). Five RNA samples from each environment were treated with RiboZero Bacteria and Plant seed/root kits in a 4:1 ratio (see 2.4.5). Success of rRNA depletion was determined using qRT-PCR (see 2.3.7) before samples were submitted to TGAC for sequencing on the Illumina HiSeq (two samples per lane) (see 2.5.1).

All samples were de-multiplexed and quality filtered as standard, and data was in part analysed by Mark Alston at TGAC as part of a collaborative CCC project (see 2.5.3). Sequence data from DNA samples was analysed using Metaphlan (Segata *et al.*, 2012) and Metaphyler (Liu *et al.*, 2010) to determine taxonomic composition based on protein coding and rRNA genes. Residual rRNA sequences were removed from the RNA sample *in silico* using SortMeRNA (Kopylova *et al.*, 2012) and the number of copies of RIS recovered was determined using USEARCH (Edgar, 2010). Sequencing depth and transcriptional activity per gram of soil were then calculated (see 2.5.4). Non-rRNA sequences were analysed using Metaphyler to determine taxonomic composition. Sub-sampled files from each sample (based on the lowest read count sample) were analysed using rapsearch2 (Zhao *et al.*, 2012), a reduced alphabet BLAST-like algorithm, against the non-redundant nucleotide collection at the National Centre for Biotechnology Information (NCBI). Output files were uploaded into MEGAN (Huson *et al.*, 2007) to visualise and compared samples based on taxonomic composition. Sequences from the Viridiplantae were excluded from analysis of eukaryotes. Pair-wise comparisons were made between each plant rhizosphere and soil

using the t-test with a 95% confidence interval. Multidimensional scaling was performed in PRIMER6 (Clarke, 1993). Count data for different taxa were normalised to a percentage within their respective samples then square root transformed before a Bray-Curtis similarity matrix was generated and used to plot data on x and y axis to generate the plot.

4.3 Results and discussion

4.3.1 Analysis of metagenomic data with Metaphyler and Metaphlan

For the metagenomic sequencing, read counts were 64,156,233 for soil, 73,210,555 for the wheat rhizosphere, 87,541,019 for the oat rhizosphere and 92,177,163 for the pea rhizosphere. Due to differences here, read numbers assigned to different taxonomic groups were expressed as percentages. Based on the Metaphyler analysis (Figure 4.1) which only allows identification of prokaryotes, the Proteobacteria were the dominant phylum in all environments, followed by Actinobacteria. Bacteroidetes, Firmicutes, Verrucomicrobia and Acidobacteria also contributed at least 1 %. The remaining ≈ 2 % of reads were derived from 12 different phyla. The Alphaproteobacteria were the dominant class in all environments, followed by Actinobacteria, Betaproteobacteria, Deltaproteobacteria, Gammaproteobacteria and Flavobacteria. The remaining ≈ 20 % of reads were derived from 28 different classes. Actinomycetes, represented by Mycobacteriaceae and Streptomyceteaceae, were the most abundant order. Rhizobiales (particularly Bradyrhizobiaceae), Spingomonadales (particularly Sphingomonadaceae) and Rhodobacterales contributed most to the Alphaproteobacteria. Burkholderialases (particularly Burkholderiaceae and Commamonadaceae) contributed most to the Betaproteobacteria. Myxococcales and Caulobacterales contributed most to Deltaproteobacteria, while Xanthomonadales and Pseudomonadales contributed most to the Gammaproteobacteria. The other ≈ 43 % and ≈ 83 % of orders and families were represented by 54 and 99 taxa respectively. At genus level (Figure 4.2), the majority of the population, represented by 116 taxa, did not represent at least 1%.The most abundant genera included *Mycobacterium*, *Streptomyces* and *Burkholderia* in all environments. *Xanthomonas* and *Flavobacterium* were also highly abundant in wheat rhizosphere and *Pseudomonas*, *Mesorhizobium* and *Variovorax* were highly abundant in pea rhizosphere.

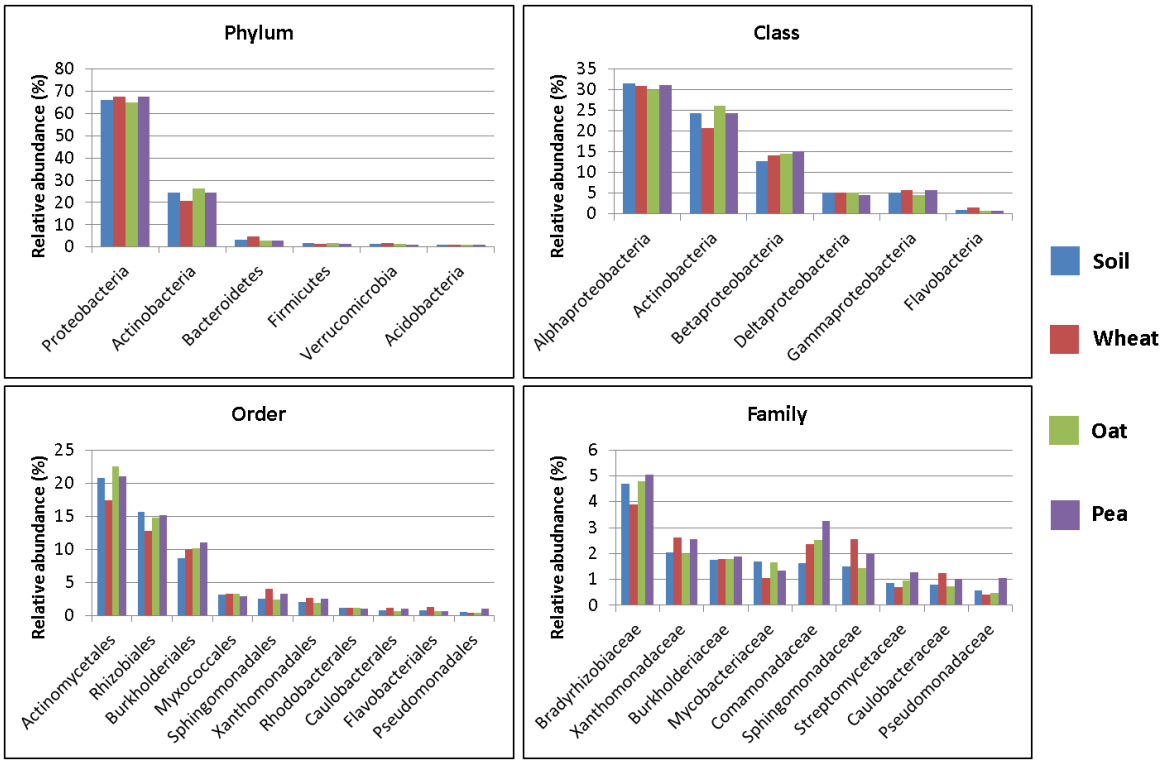


Figure 4.1: Relative abundance of high level taxonomic ranks derived from Metaphyler analysis of metagenomic data. The most abundant phyla, classes, orders and families representing ≥1% in at least one environment are shown. Total read counts are 40631 for soil, 52921 for wheat, 60268 for oat and 36296 for pea.

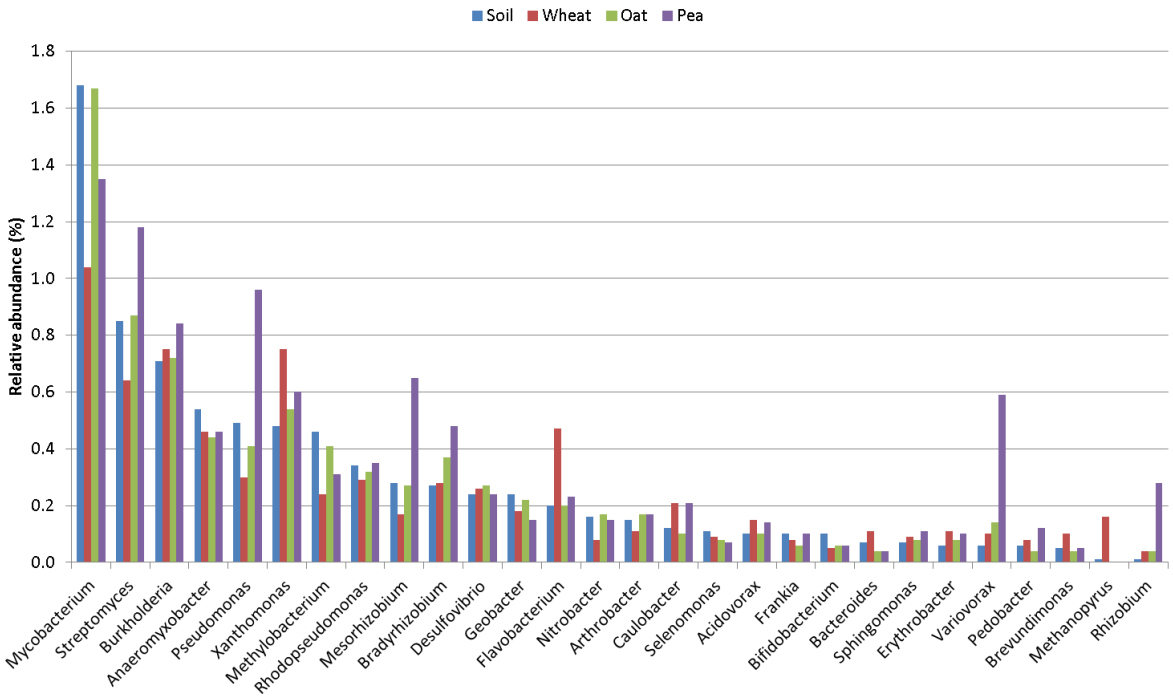


Figure 4.2: Relative abundance genera from Metaphyler analysis of metagenomic data. The most abundant (at least 0.1% in at least one environment) genera are shown. Total read counts are 40631 for soil, 52921 for wheat, 60268 for oat and 36296 for pea.

There were some differences between soil and the three rhizospheres, particularly at lower taxonomic levels (Table 4.1), however they could not be validated statistically due to the absence of biological replicates in the DNA samples. There was considerable overlap in the selection (≥1.5 fold) or depletion (≤0.5 fold) of different taxa. For example, *Rhizobium*, *Variovorax*, *Ralstonia* and *Cupriavidus* were more abundant in all three rhizospheres compared to soil. These taxa are well known for their interactions with plants. The former two are plant growth promoting rhizobacteria (PGPR) while the latter two are plant pathogens. *Rhizobium* was enriched 4 fold in the cereal rhizospheres but 28 fold in the pea rhizosphere, its host legume. *Variovorax* was also more enriched in the pea rhizosphere (9 fold) compared to the cereals. This is consistent with the observation in a recent rRNA based community analysis of the pea rhizosphere (Turner *et al.*, 2013). Also selected by all three plants were *Alicyclobacillus*, *Chloroflexus*, Lactobacillales, *Neisseria*, Oceanospirillales, *Polaromonas* and Verrucomicrobiaceae. Comamonadaceae were selected only by oat and pea, while several taxa were selected by either wheat and oat, or wheat and pea. These included *Halobacteria*, *Prevotella* and *Stenotrophomonas* by wheat and oat, and *Caulobacter*, *Erythrobacter*, Methylophilaceae and *Xanthomonas* by wheat and pea.

A number of taxa were depleted relative to soil by all plants (Table 4.2), including *Clostridium, Gluconacetobacter*, *Legionella*, *Paenibacillus*, and *Slakia*. Desulfobacteraceae were depleted by only wheat and pea, Cytophagaceae, Euryarchaeota, Methanobacteria and *Segniliparus* were depleted by only oat and pea, while only wheat and oat depleted *Planctomyces*.

Table 4.1: Comparison of taxa selected (≥1.5 fold) by wheat, oat and pea relative to soil, based on Metaphyler analysis of metagenomic DNA. Only taxa contributing at least 0.01% in at least one environment are shown.

Comparison	Taxa	Relative abundance (%)		Fold difference
		Soil	Rhizosphere	vs soil
Soil vs wheat	Methanopyri	0.01	0.16	16.0
	Methanopyrales	0.01	0.16	16.0
	Methanopyraceae	0.01	0.16	16.0

	Methanopyrus	0.01	0.16	16.0
	Polaromonas	0.01	0.07	7.0
	Neisseria	0.01	0.05	5.0
	Methylophilales	0.01	0.04	4.0
	Oceanospirillales	0.01	0.04	4.0
	Rhizobium	0.01	0.04	4.0
	Methanobacteria	0.15	0.51	3.4
	Methanobacteriales	0.15	0.51	3.4
	Methanobacteriaceae	0.15	0.51	3.4
	Oxalobacteraceae	0.06	0.19	3.2
	Euryarchaeota	0.26	0.82	3.2
	Lactobacillales	0.01	0.03	3.0
	Methylophilaceae	0.01	0.03	3.0
	Moraxellaceae	0.01	0.03	3.0
	Halobacteria	0.02	0.05	2.5
	Halobacteriales	0.02	0.05	2.5
	Halobacteriaceae	0.02	0.05	2.5
	Verrucomicrobiaceae	0.02	0.05	2.5
	Cupriavidus	0.02	0.05	2.5
	Sphingobium	0.02	0.05	2.5
	Flavobacterium	0.2	0.47	2.4
	Alicyclobacillaceae	0.01	0.02	2.0
	Porphyromonadaceae	0.01	0.02	2.0
	Staphylococcaceae	0.01	0.02	2.0
	Alicyclobacillus	0.01	0.02	2.0
	Brevundimonas	0.05	0.1	2.0
	Chloroflexus	0.01	0.02	2.0
	Ralstonia	0.01	0.02	2.0
	Staphylococcus	0.01	0.02	2.0
	Thioalkalivibrio	0.01	0.02	2.0
	Polaromonas	0.01	0.05	5.0
	Verrucomicrobiaceae	0.02	0.08	4.0
	Rhizobium	0.01	0.04	4.0
	Lactobacillales	0.01	0.03	3.0
	Variovorax	0.06	0.14	2.3
	Verrucomicrobiae	0.04	0.09	2.3
	Verrucomicrobiales	0.04	0.09	2.3
Soil vs oat	Halobacteria	0.02	0.04	2.0
	Halobacteriales	0.02	0.04	2.0
	Oceanospirillales	0.01	0.02	2.0
	Alicyclobacillaceae	0.01	0.02	2.0
	Alteromonadaceae	0.01	0.02	2.0
	Halobacteriaceae	0.02	0.04	2.0
	Pelobacteraceae	0.01	0.02	2.0
	Porphyromonadaceae	0.01	0.02	2.0

	Alicyclobacillus	0.01	0.02	2.0
	Chloroflexus	0.01	0.02	2.0
	Meiothermus	0.01	0.02	2.0
	Neisseria	0.01	0.02	2.0
	Pelobacter	0.01	0.02	2.0
	Ralstonia	0.01	0.02	2.0
Soil vs pea	Rhizobium	0.01	0.28	28.0
	Variovorax	0.06	0.59	9.8
	Methylophilales	0.01	0.09	9.0
	Methylophilaceae	0.01	0.09	9.0
	Polaromonas	0.01	0.05	5.0
	Neisseria	0.01	0.04	4.0
	Rhizobiaceae	0.15	0.53	3.5
	Ralstonia	0.01	0.03	3.0
	Sphingobacteriaceae	0.17	0.47	2.8
	Verrucomicrobiaceae	0.02	0.05	2.5
	Oxalobacteraceae	0.06	0.14	2.3
	Mesorhizobium	0.28	0.65	2.3
	Comamonadaceae	1.62	3.27	2.0
	Lactobacillales	0.01	0.02	2.0
	Oceanospirillales	0.01	0.02	2.0
	Syntrophobacterales	0.01	0.02	2.0
	Alicyclobacillaceae	0.01	0.02	2.0
	Alicyclobacillus	0.01	0.02	2.0
	Chlorobaculum	0.01	0.02	2.0
	Chloroflexus	0.01	0.02	2.0
	Pedobacter	0.06	0.12	2.0
	Pseudomonas	0.49	0.96	2.0

Table 4.2: Comparison of taxa depleted (≤0.5 fold) by wheat, oat and pea relative to soil, based on Metaphyler analysis of metagenomic DNA. Only taxa contributing at least 0.01% in at least one environment are shown.

Comparison	Taxa	Relative abundance (%)		Fold difference
		Soil	Rhizosphere	vs soil
Soil vs wheat	Methylobacteriaceae	0.46	0.24	0.5
	Methylobacterium	0.46	0.24	0.5
	Chlamydiae	0.02	0.01	0.5
	Chlamydiae	0.02	0.01	0.5
	Chlamydiales	0.02	0.01	0.5
	Desulfobacteraceae	0.02	0.01	0.5
	Legionellaceae	0.06	0.03	0.5

	Nocardiaceae	0.06	0.03	0.5
	Pseudonocardiaceae	0.04	0.02	0.5
	Actinomyces	0.04	0.02	0.5
	Bifidobacterium	0.1	0.05	0.5
	Gluconacetobacter	0.02	0.01	0.5
	Legionella	0.06	0.03	0.5
	Magnetospirillum	0.04	0.02	0.5
	Nitrobacter	0.16	0.08	0.5
	Planctomyces	0.02	0.01	0.5
	Slackia	0.02	0.01	0.5
	Thermus	0.04	0.02	0.5
	Legionellales	0.07	0.03	0.4
	Paenibacillaceae	0.05	0.02	0.4
	Nitrosomonadales	0.03	0.01	0.3
	Nitrosomonadaceae	0.03	0.01	0.3
	Clostridiaceae	0.04	0.01	0.3
	Clostridium	0.04	0.01	0.3
	Rhodococcus	0.04	0.01	0.3
	Paenibacillus	0.05	0.01	0.2
Soil vs oat	Euryarchaeota	0.26	0.14	0.5
	Clostridiaceae	0.04	0.02	0.5
	Lachnospiraceae	0.06	0.03	0.5
	Segniliparaceae	0.02	0.01	0.5
	Clostridium	0.04	0.02	0.5
	Gluconacetobacter	0.02	0.01	0.5
	Planctomyces	0.02	0.01	0.5
	Segniliparus	0.02	0.01	0.5
	Slackia	0.02	0.01	0.5
	Cytophagaceae	0.14	0.06	0.4
	Paenibacillaceae	0.05	0.02	0.4
	Paenibacillus	0.05	0.02	0.4
	Legionellales	0.07	0.02	0.3
	Methanobacteria	0.15	0.04	0.3
	Methanobacteriales	0.15	0.04	0.3
	Methanobacteriaceae	0.15	0.04	0.3
	Legionellaceae	0.06	0.01	0.2
	Legionella	0.06	0.01	0.2
Soil vs pea	Desulfobacteraceae	0.02	0.01	0.5
	Prevotellaceae	0.02	0.01	0.5
	Rikenellaceae	0.02	0.01	0.5
	Segniliparaceae	0.02	0.01	0.5
	Alistipes	0.02	0.01	0.5
	Gluconacetobacter	0.02	0.01	0.5
	Oribacterium	0.02	0.01	0.5
	Prevotella	0.02	0.01	0.5

Segniliparus	0.02	0.01	0.5
Slackia	0.02	0.01	0.5
Legionellales	0.07	0.03	0.4
Paenibacillaceae	0.05	0.02	0.4
Paenibacillus	0.05	0.02	0.4
Cytophagia	0.19	0.07	0.4
Cytophagales	0.19	0.07	0.4
Cytophagaceae	0.14	0.05	0.4
Euryarchaeota	0.26	0.09	0.3
Legionellaceae	0.06	0.02	0.3
Legionella	0.06	0.02	0.3
Clostridiaceae	0.04	0.01	0.3
Clostridium	0.04	0.01	0.3
Methanobacteria	0.15	0.03	0.2
Methanobacteriales	0.15	0.03	0.2
Methanobacteriaceae	0.15	0.03	0.2
Lachnospiraceae	0.06	0.01	0.2

4.3.2 Comparison of Metaphyler and Metaphlan outputs

Metaphlan (Segata *et al.*, 2012) is a protein coding gene marker based tool to profile microbial communities, similar to Metaphyler (Liu *et al.*, 2010). Comparing the output of Metaphlan with that of Metaphyler showed the two produced similar results at phylum level (R^2 = 0.9485) (Figure 4.3a). Twelve phyla were detected by both, while three phyla (Gemmatimonadetes, Nitrospirae and Thermi) were only detected by Metaphlan, and eight phyla (Aquificae, Crenarchaeota, Deinococcus-Thermus, Spirochaetes, Synergistetes, Tenericutes, Thaumarchaeota, Thermotogae) were only detected by Metaphyler. Actinobacteria and Proteobacteria were slightly over-represented by Metaphyler while Chloroflexi were largely over-represented by Metaphlan.

At genus level (Figure 4.3b) the two produced widely different results (R^2 = 0.2874). A total of 85 genera were detected by Metaphlan and 119 by Metaphyler, but only 16 of these were shared between both. Metaphlan was able to assign more reads at genus level than Metaphyler, possibly due to its use of clade-specific marker genes. This led to the over-representation of many genera including *Burkholderia*, *Mycobacterium* and *Pseudomonas* by Metaphlan. Some genera were slightly over-represented by Metaphyler, such as *Streptomyces* and *Xanthomonas*.

Both Metaphlan and Metaphyler use multiple marker genes and can analyse data orders of magnitude faster than conventional algorithms such as BLAST (Altschul *et al.*, 1990), and other programs such as WEBCARMA (Gerlach *et al.*, 2009) and PhymmBL (Brady and Salzberg, 2009). The different results generated for the same data set here highlight the need for caution when interpreting results from any study analysed by such methods. The marker genes used in Metaphlyler do not include rRNA genes. This is important when analysing communities based on RNA that has been depleted of rRNA. Although residual rRNA sequences were removed *in silico*, the comparison to the DNA based analysis is more straightforward if rRNA genes were not used as markers. For this reason, the RNA data was analysed using Metaphyler alone.

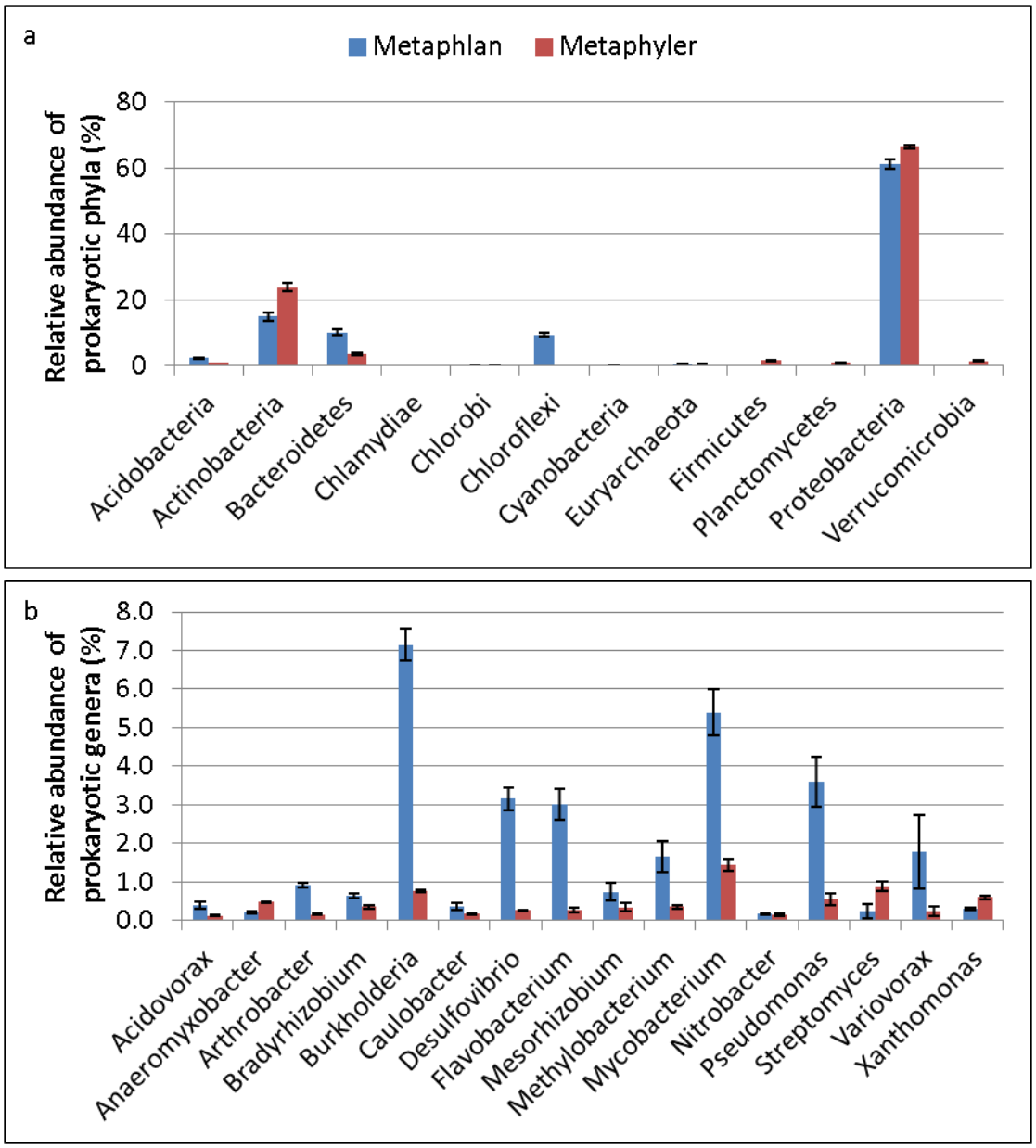


Figure 4.3: Comparison of Metaphlan and Metaphyler analyses of prokaryotic phyla (a) and genera (b). Values are means of 4 biological replicates one from each soil, wheat, oat and pea rhizospheres. Only taxa detected by both are shown.

4.3.3 Comparison of DNA and RNA based taxonomic profiles from Metaphyler

Multidimensional scaling allowed representation of the community profiles based on the Metaphyler analysis for both DNA and RNA (Figure 4.4). Although the data cannot be statistically validated due to lack of biological replication for the DNA samples, the separation of the points suggests that communities are distinct for each environment

based on both DNA and RNA. The wheat rhizosphere was similar to soil, particularly in the RNA dataset, while pea and oat where distinct from both wheat and soil, and each other, particularly in the RNA dataset.

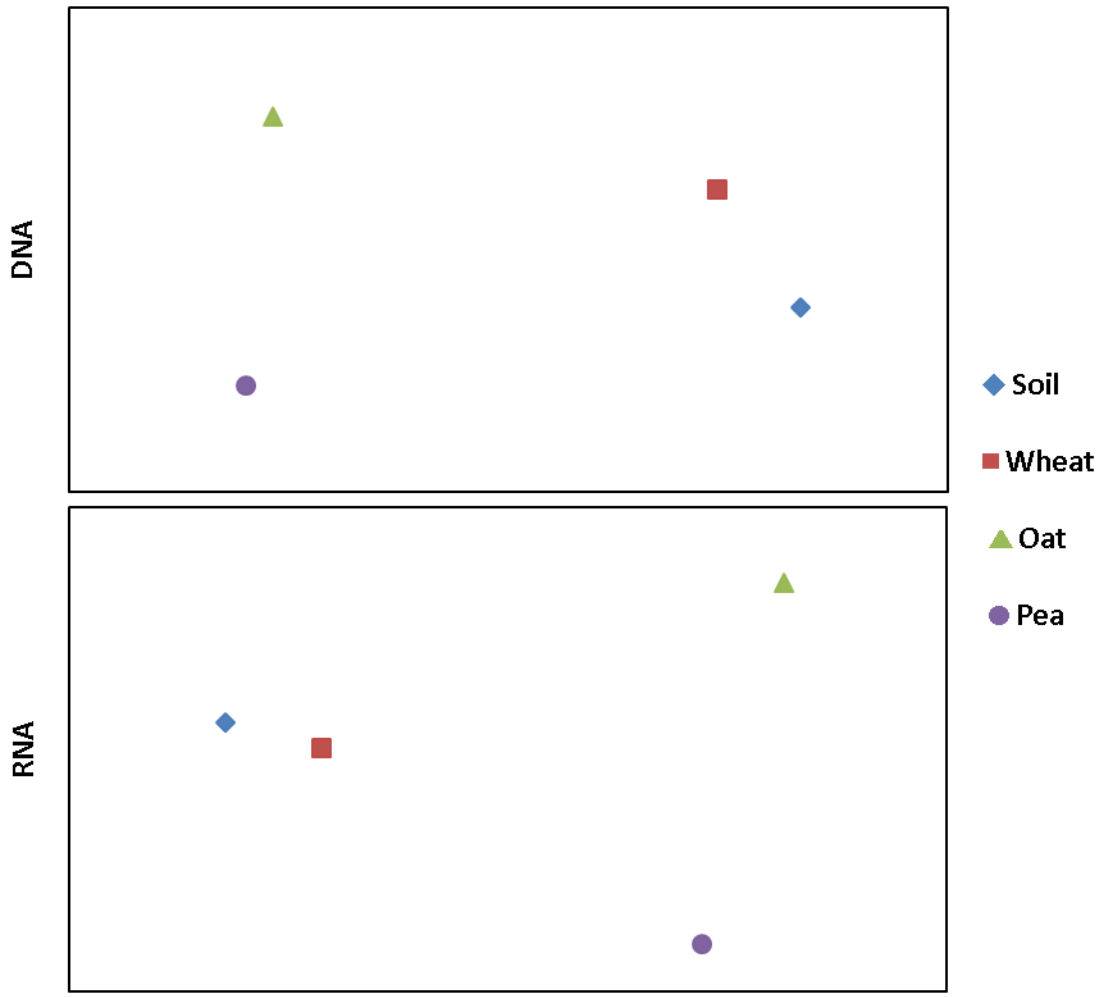


Figure 4.4. Multidimensional scaling representation of community structure in soil and the rhizospheres of wheat, oat and pea. Data were generated in PRIMER6 and plotted in Excel. Data are means of biological replicates for RNA, where n=5 for soil, oat and pea, and n=4 for wheat, but single samples for DNA. Because of the way the analysis was performed, no error bars could be shown. The plots are based on relative abundance of all taxa analysed in Metaphyler. The x,y coordinates were generated and exported by PRIMER6 and scales, although arbitrary, were standardised across plots. The closer two points are together, the more similar their communities are.

A comparison of the RNA and DNA outputs from Metaphyler was also performed at the class level. From the metatranscriptomes, 10 classes represented the majority of taxa (Figure 4.5a), with between 5% and 10% comprised of other less abundant taxa. The 6 dominant classes from the metagenomes were represented in the transcriptomes, in addition, Sphingobacteria, Cytophagia, Opitutae and Bacilli represent at least 1% in at least one environment. Differences between environments were present and there was also intra-sample variability, particularly for the wheat rhizosphere. From the metagenomes 6 classes represented the majority of taxa (Figure 4.5b), while around 20% was comprised of other less abundant taxa. The dominant classes were Alphaproteobacteria, followed by Actinobacteria, Betaproteobacteria, Gammaproteobacteria, Deltaproteobacteria then Flavobacteria. There was general consistency across environments, but intra-sample variability cannot be assessed as no biological replicates were available. A larger proportion of sequences were grouped as “other” in the DNA compared to RNA, possibly indicating a large and diverse population of present, but poorly active bacteria. Alternatively, it could be due to the conserved nature and high expression of housekeeping genes.

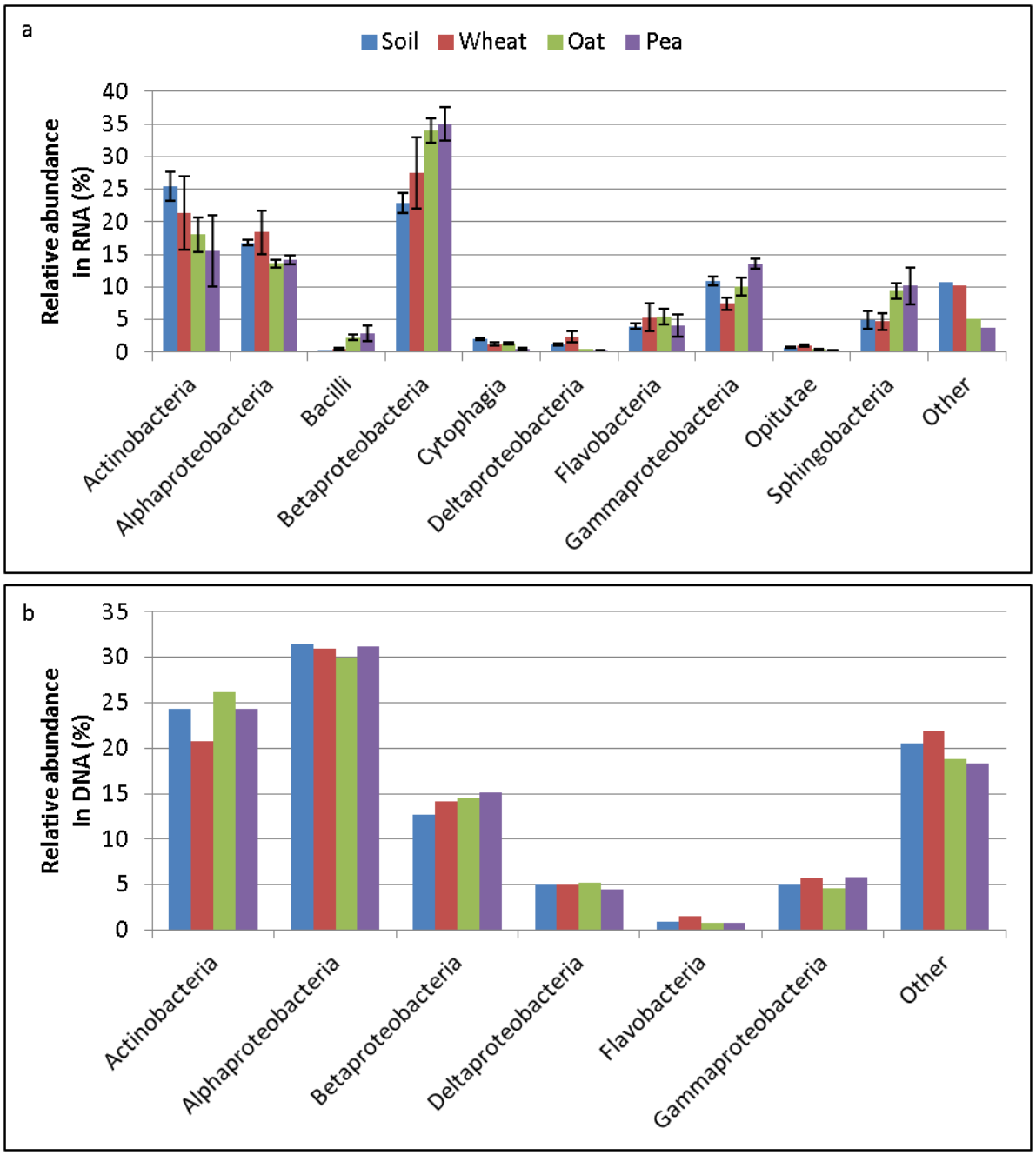


Figure 4.5: Relative abundances of the most highly abundant (at least 1% in at least one environment) bacterial classes analysed by Metaphyler on RNA (a) and DNA (b). Data are means (±1SEM) of biological replicates for RNA, where n=5 for soil, oat and pea, and n=4 for wheat, but single samples for DNA.

By comparing the abundance ratio of taxa based on RNA to those based on DNA, it was possible to determine which taxa were present but not particularly active, and also those which are less abundant but more active (Figure 4.6). Generally there was a consensus across environments, with taxa such as Alphaproteobacteria and Methanobacteria being over-represented in the DNA pool while Betaproteobacteria and Flavobacteria were over-represented in the RNA pool. However, there was variability in the degree of over-representation for some taxa, for example Sphingobacteria were over-represented 9.39 fold in soil, 5.47 fold in wheat rhizosphere, 21.25 fold in oat rhizosphere and 14.74 fold in pea rhizosphere. Bacilli stood out as being over-represented in the RNA pool for the rhizospheres (1.21, 5.52, 9.08 fold for wheat, oat and pea respectively) but under-represented in soil (0.75 fold). Their representation in the DNA pool was similar, suggesting that they become more active in the rhizosphere, possibility due to availability of the influence of plant derived carbon and energy sources. Thermoprotei were represented in the metagenomes at low abundance, and were largely absent from the metatranscriptomes, though they were present at low abundance in wheat rhizosphere, which explains their over-representation in the DNA for wheat rhizosphere only. Interestingly there were more shared classes over-represented in the DNA than the RNA.

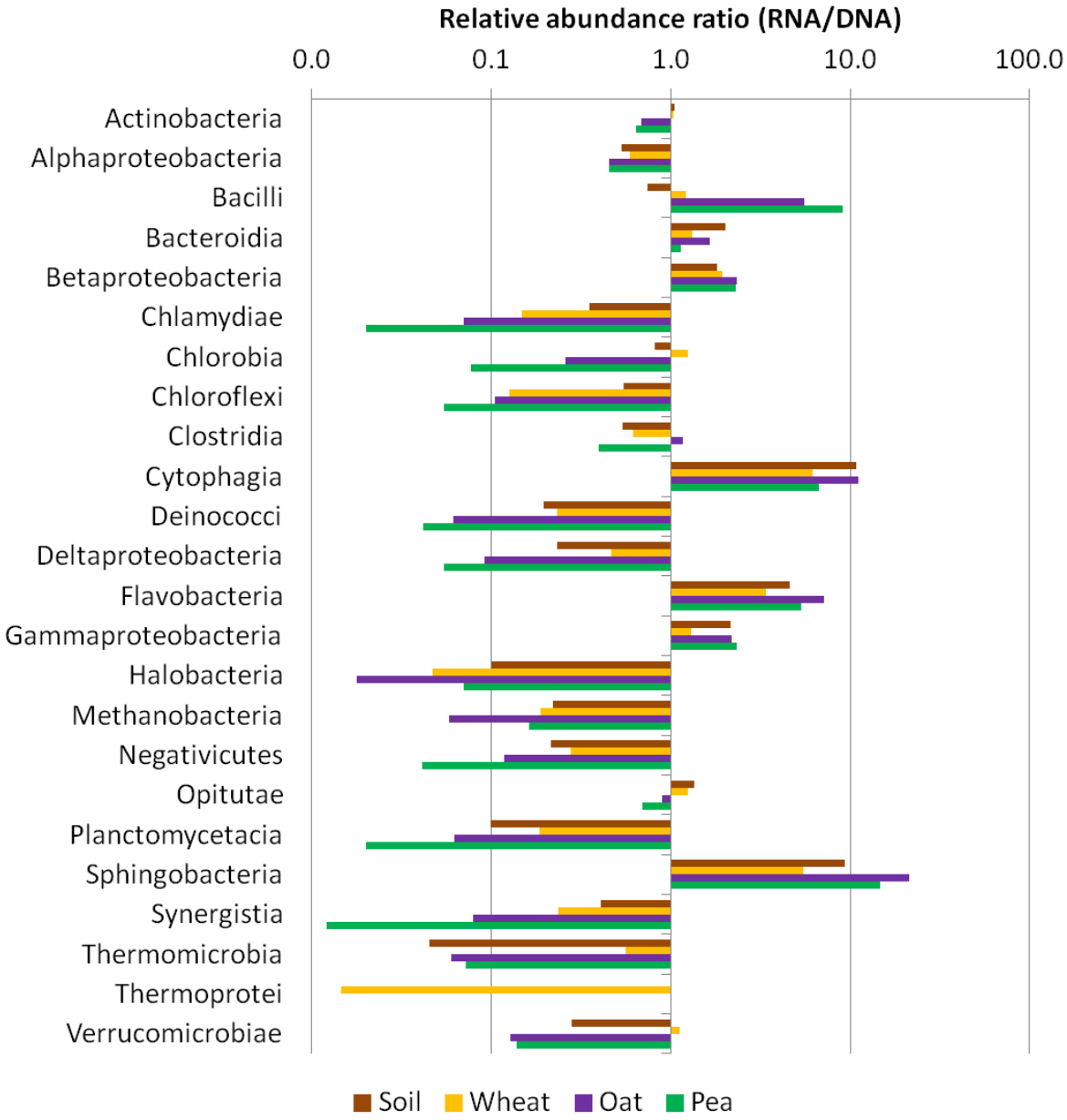


Figure 4.6: Ratio of taxonomic abundance in RNA to DNA, based on class level and only those taxa detected by Metaphyler in both the metagenome and metatranscriptome.

Based on the Metaphyler analysis of RNA, which could be statistically validated due to biological replication, a few bacterial classes were selected in a rhizosphere compared to soil (Table 4.3). Relatively few taxa were differentially abundant between wheat and soil, while more differences were observed between both oat and pea compared to soil, consistent with the results seen in Turner *et al.* (2013). Verrucomicrobiae and Fibrobacteres were enriched in wheat rhizosphere RNA and also in the DNA. Bacilli,

Clostridia and Sphingobacteria were enriched in the oat rhizosphere, although they were slightly reduced in the DNA. Gammaproteobacteria were enriched in the pea rhizosphere in both the RNA and DNA. Betaproteobacteria were enriched in both oat and pea rhizospheres in both the RNA and DNA. Consistent with the observations in Turner *et al.* (2013), a number of taxa were depleted in oat and pea rhizospheres relative to soil. These included Alphaproteobacteria, Bacteroidia, Cytophagia and Deltaproteobacteria and the majority of these were also depleted in the DNA.

Table 4.3: Differentially abundant (P≤0.05) taxa in the rhizospheres of wheat, oat and pea compared to soil, based on relative abundance from Metaphyler analysis of RNA and DNA. No difference in from the DNA is denoted by -.

Comparison	Taxa	Relative abundance		P value	Fold difference in RNA	Fold difference in DNA
		Soil (%)	Rhizosphere (%)			
Soil vs wheat	Verrucomicrobiae	0.01	0.07	0.012	5.99	1.50
	Fibrobacteres	0.00	0.01	0.023	2.91	-
	Thaumarchaeota	0.01	0.00	0.028	0.11	-
	Bacteroidetes	3.64	1.89	0.022	0.52	-
	Gammaproteobacteria	10.95	7.48	0.018	0.68	1.13
Soil vs oat	Bacilli	0.32	2.26	0.001	7.06	0.95
	Betaproteobacteria	22.91	34.00	0.002	1.48	1.14
	Clostridia	0.33	0.64	0.004	1.95	0.90
	Erysipelotrichi	0.00	0.02	0.026	14.34	1.00
	Sphingobacteria	4.97	9.35	0.041	1.88	0.83
	Actinobacteria	0.72	0.20	0.043	0.28	-
	Thaumarchaeota	0.01	0.00	0.034	0.15	-
	Verrucomicrobia	0.41	0.15	0.032	0.35	-
	Negativicutes	0.04	0.02	0.024	0.47	0.86
	Cytophagia	2.04	1.32	0.019	0.65	0.63
	Acidobacteria	0.33	0.17	0.017	0.51	-
	Bacteroidetes	3.64	1.65	0.014	0.45	-
	Proteobacteria	1.89	0.62	0.011	0.33	-
	Bacteroidia	0.89	0.51	0.008	0.57	0.70
	Deinococcus-Thermus	0.04	0.01	0.005	0.20	-
	Alphaproteobacteria	16.77	13.58	0.003	0.81	0.95
	Gemmatimonadetes	0.83	0.35	0.002	0.43	-
	Deltaproteobacteria	1.18	0.47	0.002	0.40	1.02
	Planctomycetes	0.30	0.08	0.000	0.28	-
Soil vs pea	Betaproteobacteria	22.91	34.99	0.004	1.53	1.19
	Gammaproteobacteria	10.95	13.52	0.034	1.23	1.15

Synergistia	0.03	0.00	0.049	0.04	1.43
Opitutae	0.68	0.24	0.032	0.35	0.68
Deinococci	0.04	0.01	0.030	0.21	1.00
Bacteroidetes	3.64	1.61	0.028	0.44	-
Chlorobia	0.05	0.01	0.027	0.11	1.17
Actinobacteria	0.72	0.14	0.026	0.19	-
Cyanobacteria	0.14	0.02	0.023	0.15	-
Verrucomicrobia	0.41	0.12	0.023	0.29	-
Alphaproteobacteria	16.77	14.15	0.015	0.84	0.99
Deinococcus-Thermus	0.04	0.01	0.010	0.31	-
Proteobacteria	1.89	0.56	0.008	0.30	-
Chlorobi	0.06	0.01	0.005	0.12	-
Acidobacteria	0.33	0.15	0.001	0.45	-
Bacteroidia	0.89	0.27	0.001	0.30	0.55
Negativicutes	0.04	0.01	0.001	0.18	0.95
Planctomycetes	0.30	0.06	0.000	0.19	-
Deltaproteobacteria	1.18	0.24	0.000	0.21	0.88
Cytophagia	2.04	0.47	0.000	0.23	0.37
Gemmatimonadetes	0.83	0.05	0.000	0.06	-

4.3.4 Analysis of metatranscriptomic data by rapsearch2

In addition to Metaphyler, the metatranscriptomic data was assigned to taxonomic groups with MEGAN. The number of reads per sample ranged from 31 million to 133 million, with a mean of 95 million (Figure 4.7a). Variation in read number was independent of the origin of the sample (P=0.972) All samples were successfully depleted of rRNA by treatment with Ribozero, to around 10%, with the exception of one soil sample which retained 43% of its rRNA. Overall there were no differences between sample types and the level of rRNA depletion (Figure 4.7b), suggesting using bacterial and plant Ribo-Zero kits was optimal for both soil and rhizosphere environments. USEARCH was used to filter out non-rRNA sequences which were searched against the non-redundant nucleotide database at NCBI using rapsearch2 then used as input for MEGAN. On average 4.49 % (94% of rapseach2 hits) could be assigned to a taxon, including cellular organisms, in MEGAN (Figure 4.7c). This was slightly higher in the rhizospheres compared to soil, which may reflect an enrichment of better characterised organisms, or at least proteins, in the rhizosphere. These could include well characterised plant growth promoting bacteria or plant and human pathogens. The overall assignment rate is low, suggesting the majority of protein coding genes in the soil

and rhizosphere are either entirely novel or have not yet been well characterised. Further analysis of the data will utilise tools to determine the origin of such sequences. Tools such INFERNAL might allow discovery of ncRNAs (Nawrocki *et al.*, 2009), while PhAnToMe, an extension of the SEED database might allow detection of viral sequences. The short read length of 100 bp might also have contributed to low assignment rate, so the assembly of sequences into longer contigs might improve the assignment rate. It would be interesting also to see the distributions of such sequences to see if any are particularly dominant.

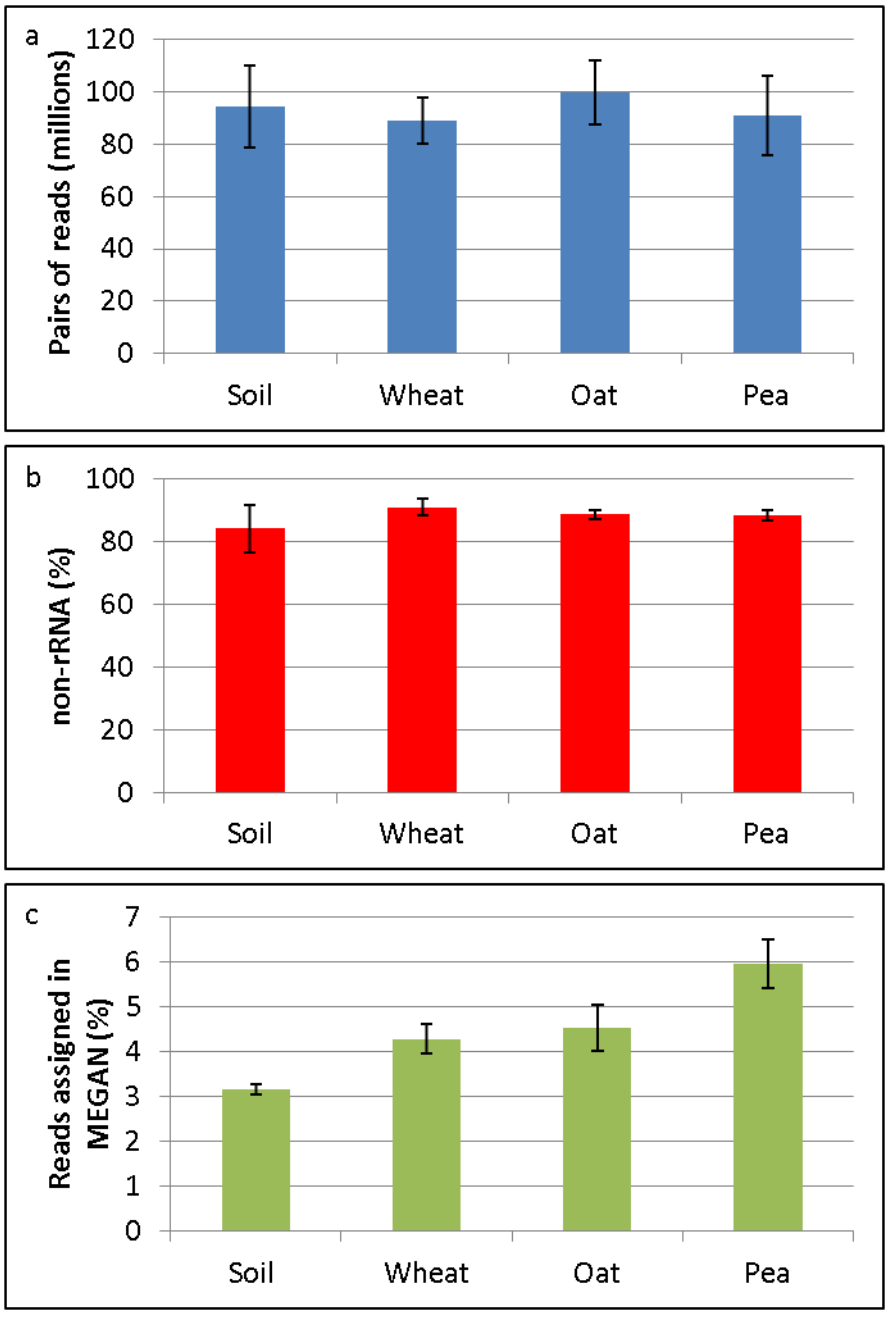


Figure 4.7: Sequencing summary showing read numbers (a), proportion of non-rRNA (b) and the proportion of reads assigned to a taxon in MEGAN (c) as a % of the non-rRNA reads uploaded to rapsearch2.

4.3.5 Community structure and highly abundant microbes

Multidimensional scaling allowed the similarities in community structure to be visualised for different taxonomic groups at different taxonomic levels. Based on this, all environments had distinct prokaryotic and eukaryotic communities at both phylum and genus level. For prokaryotic communities, the wheat rhizosphere was more similar to soil than either oat or pea rhizospheres, with the pea rhizosphere least similar to soil at both phylum and genus level. Wheat and oat rhizospheres were more similar to each other than either was to either soil or pea rhizosphere particularly at genus level. The pattern was similar for eukaryotes with pea most dissimilar to soil, and wheat and oat fairly similar, though more so at phylum level. Generally the variability within samples was greater in the rhizospheres, particularly that of pea, than in soil.

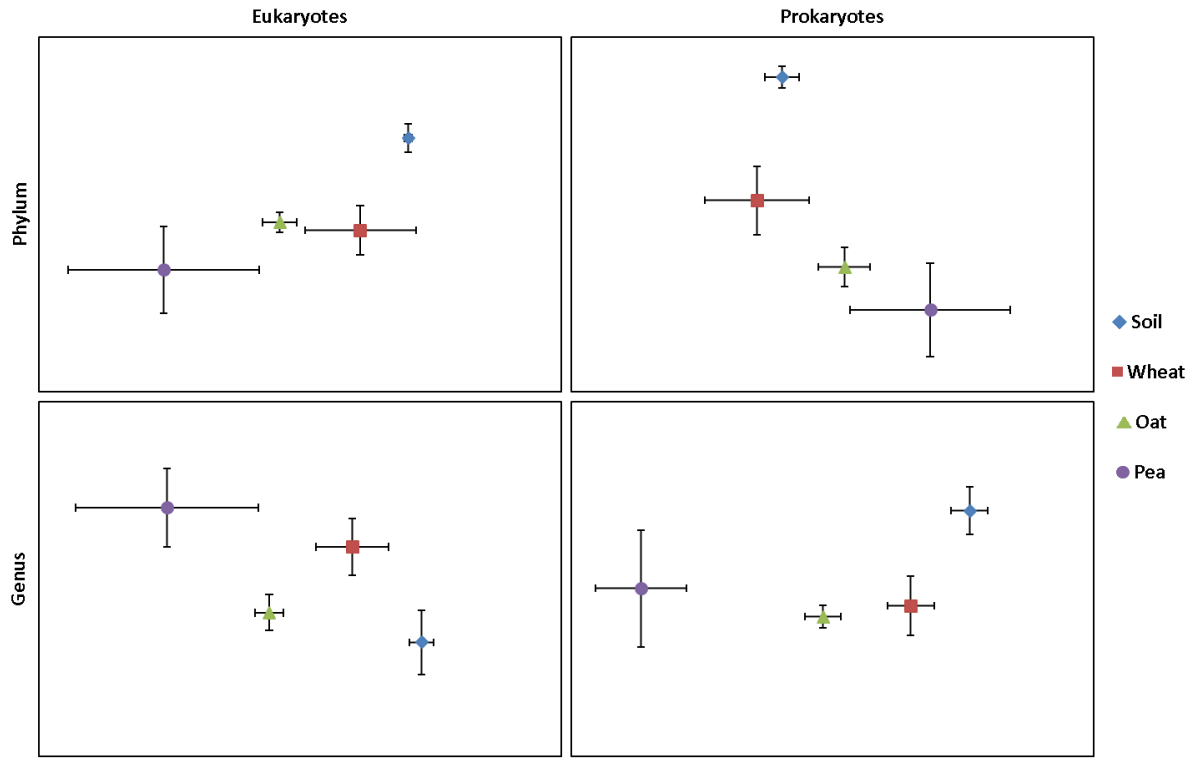


Figure 4.8. Multidimensional scaling representation of community structure in soil and the rhizospheres of wheat, oat and pea. Data were generated in PRIMER6 and plotted in Excel. Data are means (±1SEM) of biological replicates, where n=5 for soil, oat and pea, and n=4 for wheat. Although the coordinates were generated with the quantitative data of transcripts per g soil, the standard transformations for generating resemblance matrices include a normalisation step, so the plots can be considered to be based on relative abundance. The x,y coordinates were generated and exported by PRIMER6 and scales, although arbitrary, were standardised across plots. The closer two points are together, the more similar their communities are.

The use of rapsearch2 allowed the abundance of eukaryotic sequences to be determined, something that was not possible with Metaphyler. At kingdom level (Figure 4.9a) there were no significant differences ($P>0.05$) in the number of transcripts per gram of soil between any plant rhizosphere and bulk soil for bacteria, eukaryotes, archaea or cellular organisms. However the relative abundance of some groups was significantly different. Archaea were around a third less abundant in all rhizospheres compared to soil ($P<0.001$). The relative abundances of eukaryotic proteins were reduced by 27% ($P=0.022$) and 49% ($P<0.001$) in the wheat and oat rhizospheres respectively, but were relatively increased 51% ($P=0.039$) in the pea rhizosphere. These observations of relative abundance are not consistent with the observations in Turner et al (2013), but this is likely due to the different biomarkers used, i.e. mRNA vs rRNA. There was some positive correlation ($R^2=0.7799$) between the total transcripts per gram of soil and the relative abundance of eukaryotic rRNA, based on the data from Turner et al (2013). If the microbial activity in the rhizosphere is related to the amount of carbon released into the rhizosphere by the plant, this might indicate that the largest factor contributing to the proportion of eukaryotes is the amount of carbon released by the plant. However, there was no such correlation between the relative abundance of transcripts per gram of soil and eukaryotic mRNA ($R^2=0.008$). To confirm this relationship, a quantitative rRNA based analysis would need to be carried out.

The prokaryotic phylum level taxonomic profiles of all environments based on protein coding genes (Figure 4.9b) were dominated by those from Proteobacteria, representing 42% in soil and up to 53% in the pea rhizosphere. Bacteroidetes and Actinobacteria were also highly represented, on average at 23% and 19% respectively. Other well represented phyla included Verrucomicrobia, Crenarchaeota, Planctomycetes, Acidobacteria,

Firmicutes, at between 1% and 4%. The remaining 1% to 3% were represented by 32 other phyla, including many candidate divisions.

At the genus level (Figure 4.9c) proteins from a number of well characterised soil and plant-associated bacteria were dominant. These included *Mycobacterium, Bradyrhizobium, Streptomyces, Pseudomonas and Bacillus*. There was more variation between environments at genus level compared to phylum level. *Mycobacterium* were dominant in soil and wheat rhizosphere, representing 10.4% and 7.4% respectively, while in the oat and pea rhizospheres, *Streptomyces* were dominant at 8.1% and 16.3% respectively. The proportion of other taxa (from 609 genera) was variable between environments, at 63% in soil, which reduced to 52%, 42% and 30% in the wheat, oat and pea rhizospheres respectively. These data suggest that there is reduced diversity, in the rhizosphere compared to soil, and that this reduction is plant species dependent. This was reinforced by lower Shannon-Weaver diversity indexes for all three rhizospheres compared to soil (Figure 4.10). Additionally, diversity was generally higher and less variable for prokaryotic genera than it was for eukaryotic genera.

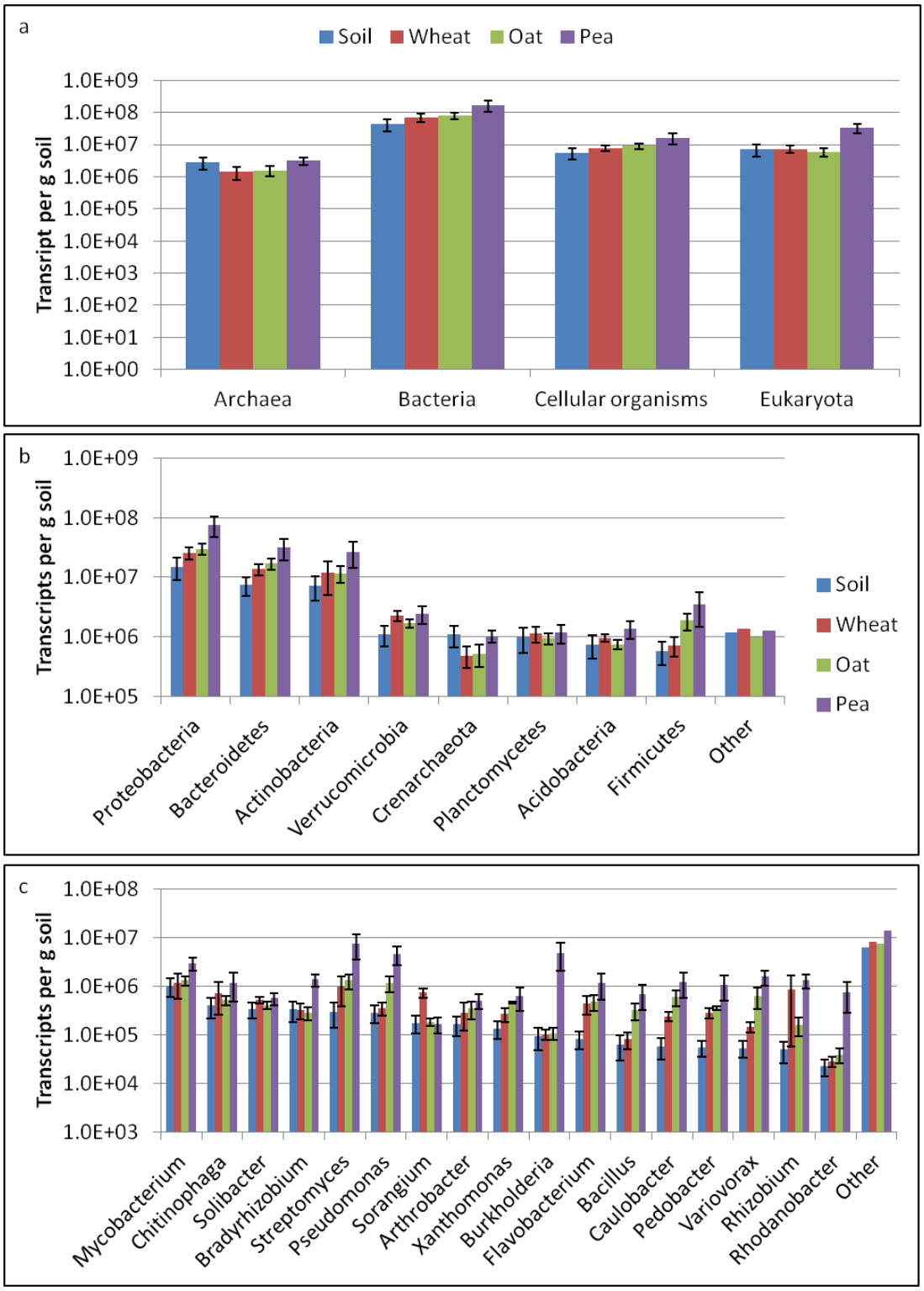


Figure 4.9: Abundance of kingdoms (a), highly abundant prokaryotic phyla (b) and genera (c) in soil and the rhizospheres of wheat, oat and pea. Values are means of biological replicates (±1SEM), where n = 5 for soil, oat and pea rhizospheres and n = 4 for wheat rhizosphere. Eukaryota exclude reads assigned to Viridiplanteae.

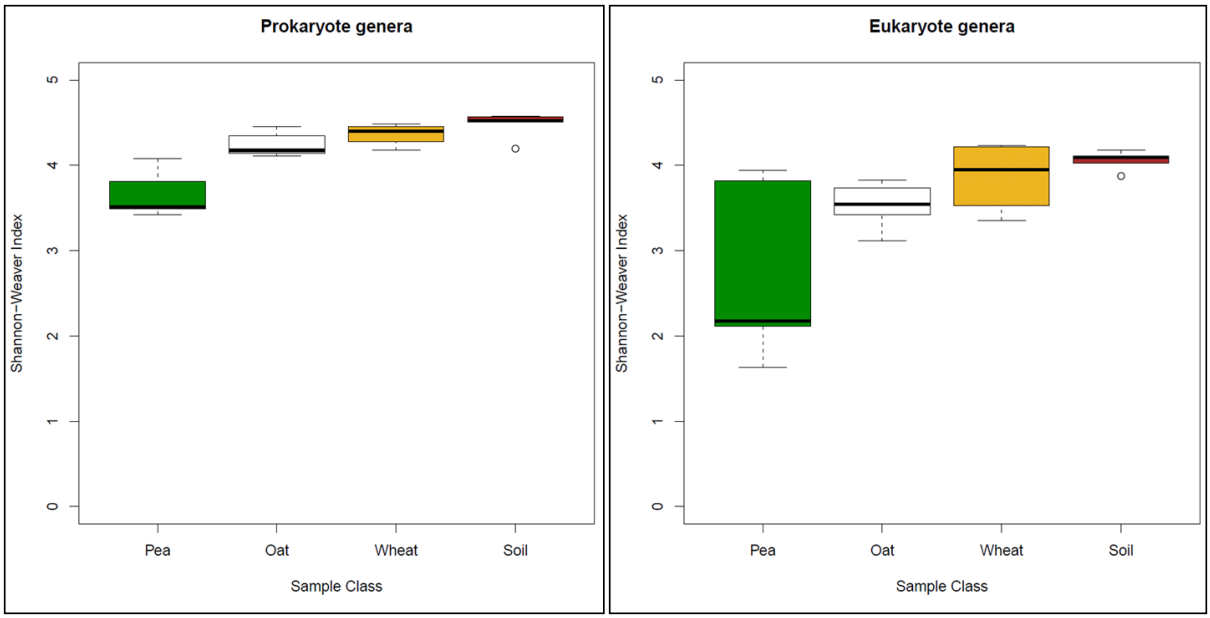


Figure 4.10: Shannon-Weaver diversity indexes of prokaryotic and eukaryotic genera. Generated by Mark Alston (TGAC).

The eukaryotic phylum level taxonomic profiles of all environments based on protein coding genes (Figure 4.11a) were dominated by those from Ascomycote fungi, which contributed 39% to the soil, 56% to both wheat and oat rhizospheres and 78% to the pea rhizosphere. Chordata and Amoebozoa were also well represented in soil at 13% and 14% respectively, but less so in the rhizospheres, at less than 10% on average. Other well represented phyla included potential plant pathogens and herbivores and bacterivores in the Oomycetes, Arthropoda and Nematoda. There were 66 other eukaryotic phyla that cumulatively contributed 14% to soil. This was lower in the rhizospheres, at 10%, 11% and 3% for wheat, oat and pea respectively.

At genus level (Figure 4.11b), proteins from a number of well characterised soil and plant-associated were dominant. These included plant pathogens such as *Phytophtora* and *Phaeosphaeria*, symbionts such as *Metarhizium* and *Glomus*, and bacterivorous nematodes and protozoa including *Acanthameoba* and *Caenorhabditis*, and other model eukaryotes such as *Homo*, *Saccharomyces* and *Drosophila*. There were a number of differences across environments, including, the surprising absence of *Saccharomyces* in the oat rhizosphere. *Penicillium* represented only 1% of soil but was the dominant genus in rhizosphere samples, 14% and 12% to wheat and oat rhizospheres respectively, and 50% of pea rhizosphere. There were 521 other eukaryotic genera that contributed 45% to soil, 40% to wheat rhizosphere, 33% to oat rhizosphere and only 17% to pea rhizosphere. This is

consistent with the data from prokaryotic genera, that the rhizospheres, particularly that of pea, are less diverse than soil (Figure 4.11).

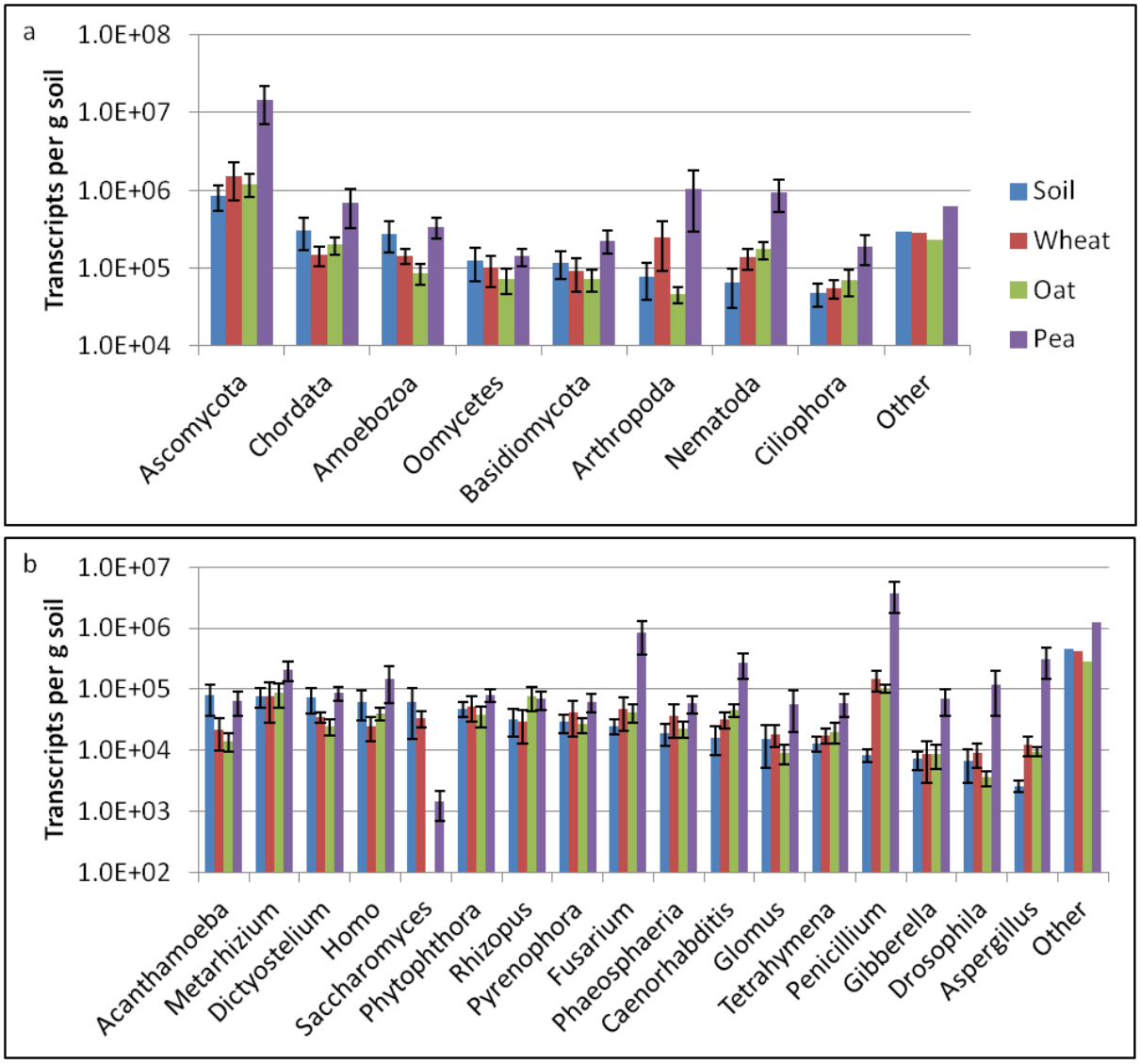


Figure 4.11: Highly abundant prokaryotic phyla (a) and genera (b) in soil and the rhizospheres of wheat, oat and pea. Values are means of biological replicates (±1SEM), where n = 5 for soil, oat and pea rhizospheres and n = 4 for wheat rhizosphere.

An advantage of using non-rRNA sequences as taxonomic marker is that they can provide information on the abundances of viruses in the environment from both DNA and RNA. Viral genomes contain no rRNA genes and so are not able to be detected by methods that rely on this marker gene. Here, only the RNA samples were searched against a protein database, so the abundance of viruses can be determined from these samples only. It

should be noted that the DNase treatment prior to Ribo Zero treatment and sequencing would have removed viral DNA. This means that any sequences assigned by MEGAN to Viruses would have had an RNA virus origin. Separate RNase treatment of DNA is required to detect DNA viruses, which is typically how they are analysed in metagenomic studies (Edwards and Rohwer, 2005). The mean number of viral transcripts in soil was 2.1, 7.7 and 2.6 fold higher than in the rhizospheres of wheat, oat and pea respectively (Figure 4.12). However, due to the large variability in soil there were no significant differences between any of the rhizospheres and soil. The variability and abundance of viral transcripts was lower in the rhizosphere samples, particularly in that of oat, suggesting that plants have some mechanisms to limit their populations. Such mechanisms could include the ROS burst of the plant immune system which could deactivate viruses. Many of the viruses in the rhizosphere would be infecting other organisms such as bacteria, fungi and nematodes, which may also have their own mechanism to cope with viral infection. Viral populations are thought to play an important role in the population dynamics of their hosts, but they themselves are dynamic and largely understudied (Williams, 2013). In fact a significant proportion of sequences unable to be assigned to a protein or organism are likely to be derived from completely novel viruses.

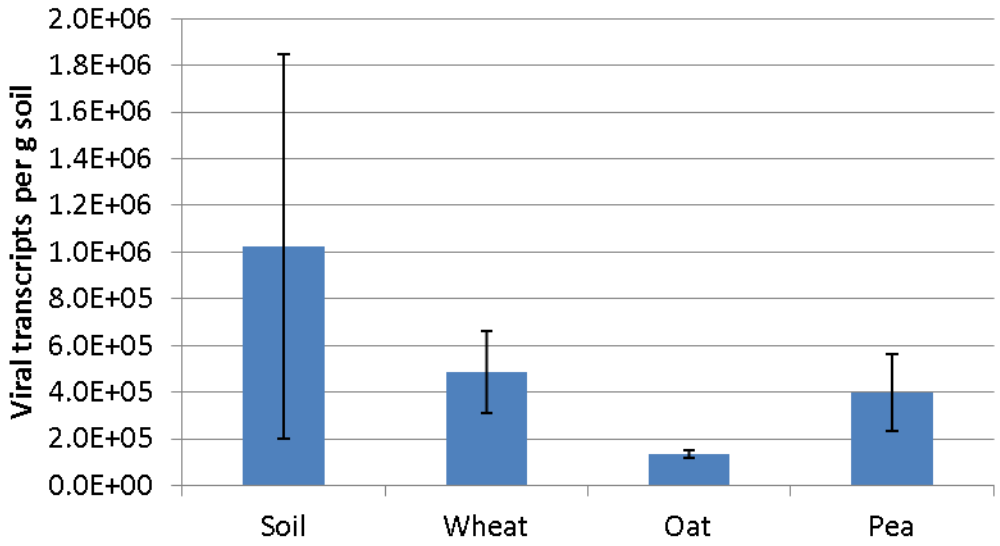


Figure 4.12: Abundance of viral transcripts in soil and plant rhizospheres. Values are means biological replicates (±1SEM) where n = 5 for soil, oat and pea, and n=4 for wheat.

4.3.6 Selection of microbes by plants

The plant effects on the microbiome community structure based on protein coding genes expressed in the rhizosphere was highly dependent on plant species (Figure 4.13). For prokaryotes, only 1.7% of 462 taxa were selected in all three rhizospheres. There was little overlap between wheat and pea (2.2%), and pea and oat (4.8%), but considerable shared selection between the two cereals, wheat and oat (16%). Oat and pea showed a similar number of specifically selected taxa (22%), though this was slightly higher for wheat (31%). The pattern was similar for eukaryotic taxa, but the plant specificity was less strong. Of the 120 selected eukaryotic taxa, 3.3% were selected by all three plants, 3.3% were shared between wheat and pea, 9.2% between the two cereals and 6.6% between oat and pea. Wheat and oat selected similar proportions of unique taxa, at 18.3% and 12.5% respectively, while the pea rhizosphere specifically selected a much higher number of eukaryotes (46.7%). Full lists of differentially abundant taxa can be found in Appendix Tables A11 throught to A24.

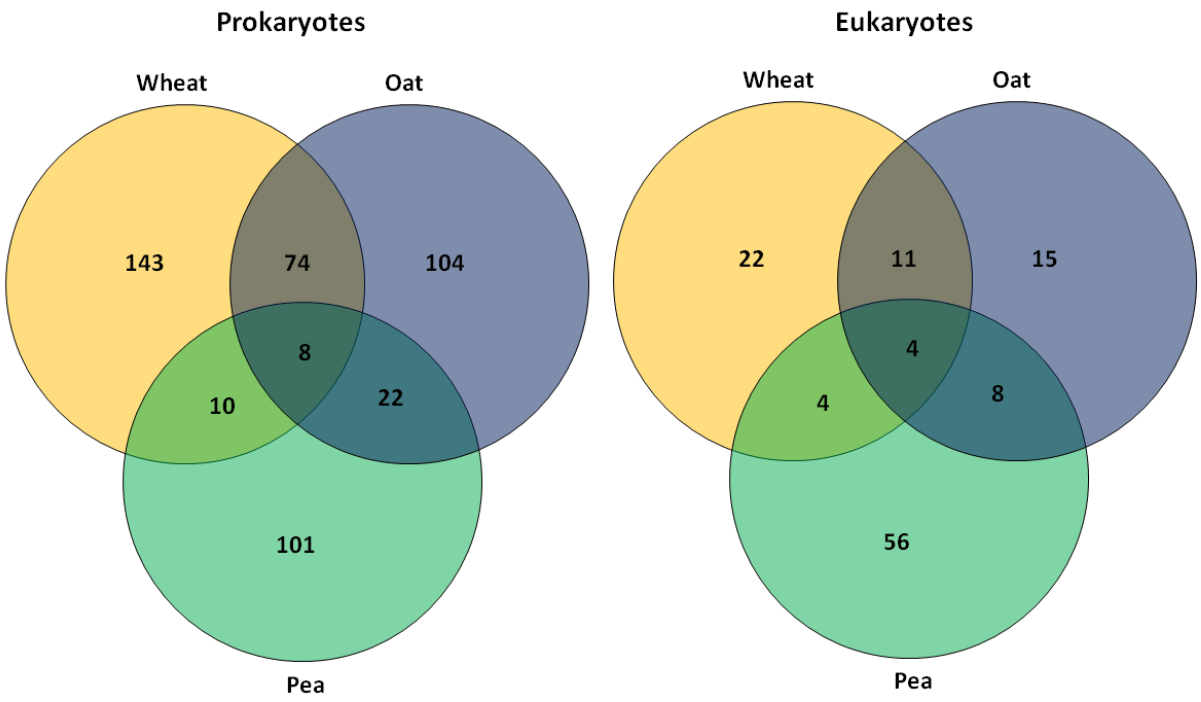


Figure 4.13: Shared, selected (P≤0.05) taxonomic groups between the rhizosphere of wheat, oat and pea compared to soil as determined by rapsearch2 and MEGAN, converted to transcripts per gram of soil for each taxon.

4.3.7 Taxa selected by all plants

The eight prokaryotic taxa selected by all plants were *Burkholderia phytofirmans*, *Burkholderia thailandensis*, *Fluoribacter*, *Janthinobacterium*, *Janthinobacterium* sp. Marseille, *Novosphingobium aromaticivorans*, *Ruegeria*, and Sphingobacteriaceae (Figure 4.14a and Appendix Table A17). Among eukaryotes, Euglenozoa, particularly *Trypanozoma*, were selected by all plants (Figure 4.14b and Appendix Table A24). *Burkholderia* spp. are well known for their interactions with plants, with some strains capable of nodulating legumes and fixing nitrogen (Chen *et al.*, 2005). *B. phytofirmans* is a known plant growth promoting rhizobacteria (PGPR) (Sessitsch *et al.*, 2005), while *B. thailandensis* can metabolise the plant sugar arabinose (Smith *et al.*, 1997). *Fluoribacter* closely resemble the human pathogen, *Legionella pneumophilia* (Garrity *et al.*, 1980), and have been detected in anoxic rice paddy fields (Weber *et al.*, 2001). *Janthinoobacterium* spp. produce violacein (Pantanella *et al.*, 2007), a secondary metabolite generally produced from biofilms, with broad antimicrobial activity, and shown to protect bacteria from predation by protozoa (Duran *et al.*, 2007). Given the increased abundance of such predators in the rhizosphere, those bacteria protected from them would be at a selective advantage. Interestingly, predation of non-violacein producing bacteria was reduced when co-cultured with violacein producers (Matz *et al.*, 2004), suggesting other rhizosphere bacteria may benefit from association with violacein producing organisms. *Novosphingobium aromaticivorans* and relatives have glycosphingolipids in their cell membranes which are more common in eukaryotes and are thought to aid in microbial colonisation of eukaryotes including plants (Heung *et al.*, 2006). Closely related *Sphingmonas* spp are among the most abundant organisms on leaf surfaces (Vorholt, 2012) *N. aromaticivorans* is capable of degrading a number of aromatic compounds (Bell and Wong, 2007), which plants are prolific producers of. *Ruegeria* spp. are typically marine bacteria that play a role in the sulphur cycle as metabolisers of dimethylsulphionoproprionate (DMSP) (Todd *et al.*, 2012). Other strains, producing antibacterial cyclic peptides, have been isolated from marine sponges (Mitova *et al.*, 2004). Terrestrial strains may also play a role in the sulphur cycle, as DMSP is known to be produced by plants (Hanson and Gage, 1996). Alternatively or additionally, they may act as biocontrol agents. Sphingobacteriaceae are a broad family including many isolates from soil. The eukaryotic *Trypanosoma* spp. are obligate animal parasites which could be infecting nematodes and other small animals that were enriched in plant rhizospheres. Interestingly, their abundance was increased despite the presence of violacein producing bacteria. Violacein is known to be toxic to trypanosomes (Duran *et al.*, 2007).

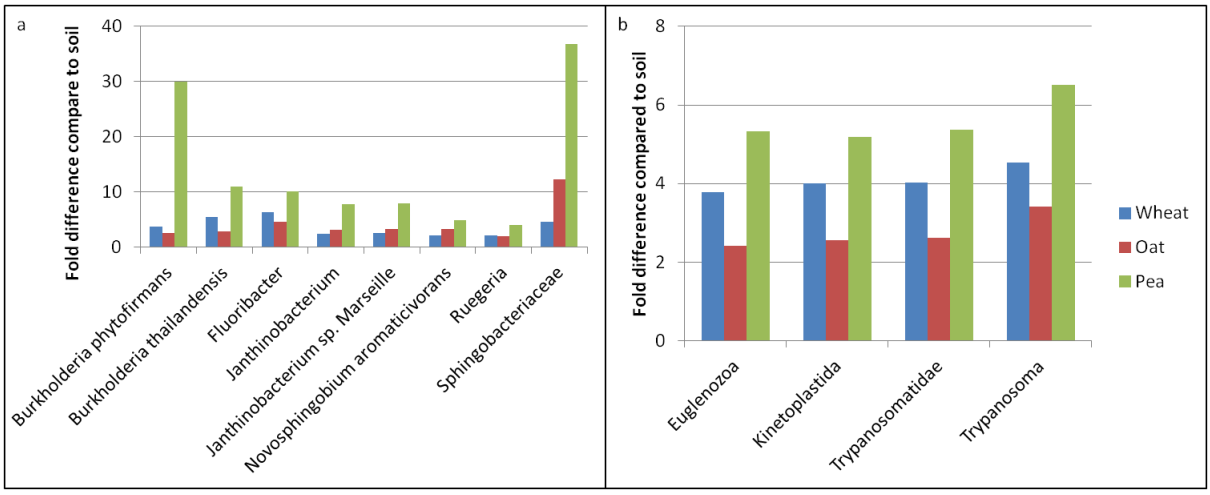


Figure 4.14: Prokaryotic (a) and eukaryotic (b) taxa selected by all plants compared to soil. Values are fold differences of means of biological replicates where n = 5 for soil, oat and pea, and n=4 for wheat.

4.3.8 Taxa specifically enriched in the wheat rhizosphere

Some of the prokaryotes selected specifically by wheat (Appendix Table A11) included the nitrogen fixing *Herbaspirillum seropedicae*, and the hydrogen oxidising *Hydrogenivirga* sp. 128-5-R1-1. Hydrogen is a by product of nitrogen fixation (Hunt and Layzell, 1993) and its utilisation has been associated with plant growth promotion by particular bacteria (Dong *et al.*, 2003), including *Flavobacterium* spp. (Maimaiti *et al.*, 2007) three strains of which, MS024-3C, BBFL7 andHTCC2170 were also enriched in the wheat rhizosphere. A number of plant polymer degrading bacteria were also selected. These included pectin degrading and potentially pathogenic *Pectobacterium*, cellulolytic Fibrobacteres (Ransom-Jones *et al.*, 2012) and *Sorangium cellulosum*, and proteolytic *Coprothermobacter proteolyticus*, suggesting the presence of plant cell wall material in the wheat rhizosphere. Sulphur oxidisers (*Hydrogenovirga* 128-5-R1-1 and *Thiomonas intermedia*), and reducers (*Desulfitobacterium hafniense* and *Desulfatibacillum alkenivorans*) were selected, suggesting enhanced sulphur cycling in the wheat rhizosphere. Antibiotic producing (Pradella *et al.*, 2002), *Streptomyces* sp. SPB74 were selected, possibly acting as biocontrol agents. Other selected prokaryotes included *Burkholderia ambifaria, Stenotrophomonas* sp. SKA14 and *Verrucomicrobiaceae*. Selected eukaryotes (Appendix Table A18) included plant pathogenic fungi, *Aspergillus oryzae* RIB40 and *Verticillium albo-atrum*, a mycorrhizal fungus *Glomus mosseae*, bacterivorous protozoa Cercomonadida, the phenolic degrading

alga *Ochromonas danica* (Semple and Cain, 1996), and mammalian pathogens *Cryptosporidium muris* (a Chromoalveoate protozoan) and Trichostrongyloidea (a nematode).

4.3.9 Taxa specifically enriched in the oat rhizosphere

Some of the prokaryotes specifically selected by oat (Appendix Table A12)included four potentially plant pathogenic *Xanthomonas* spp (X. *axonopodis*, X. *albilineans*, X.oryzae and X *camprestris*) and *Pseudomonas syringae* pathovars (pv. phaseolicola and pv. tomato). As well as known plant associated Sphingobacteria, Burkholderiales and Flavobacteria. A specific member of the Flavobacteria, *Leeuwenhoekiella blandensis*, known to degrade organic matter (Pinhassi *et al.*, 2006), was also selected. A number or strains capable of degrading aromatics were selected, including *Phenylobacterium zucineum*, *Polaromonas naphthalenivorans* and *Novosphingobium* sp. Rr 2-17. The latter has been isolated from crown galls of grapevine (*Vitis vinifera*) (Gan *et al.*, 2012). Plant cell wall polymer degraders such as *Clostridium* spp (*C. beijerinckii*, *C. carboxidivorans* and *C. acetobutylicum*) were selected, as well as potential biocontrol agents *Streptomyces griseoflavus*, *Pseudomonas fluorescens Pf0-1*, *Pseudomonas putida F1*, *Paenibacillus polymyxa* and *Lysobacter enzymogenes*. These two *Psuedomonas* spp. are well known PGPRs, while *P. polymyxa* is known to produce antifungal peptides, plant hormones and growth modulators (e.g. 2,3 butanediol) and fix nitrogen (Bitas *et al.*, 2013; Dijksterhuis *et al.*, 1999; Lebuhn *et al.*, 1997; Timmusk and Wagner, 1999), while *L. enzymogenes* produces the antifungal 2,4-diacylphloroglucinol (DAPG) as well as lactam and peptide antibiotics (Folman *et al.*, 2003), and has been associated with soils suppressive to *Rhizoctonia* (Postma *et al.*, 2008). Another selected bacteria, *Ramlibacter tataouinensis*, is highly resistant to reactive oxygen species (De Luca *et al.*, 2011), which are known to be produced by plants in response to microbes. The majority of eukaryotes specifically selected by oat (Appendix Table A19) were fungi, with the exception of the nematode *Caenorhabditis briggsae.* Eurotiomycete fungi, including *Aspergillus spp. (A. terreus*, *A. clavatus* and *A. nidulans)* were selected, as well as other fungi such as *Fusarium oxysporum* and plant pathogenic *Alternaria*.

4.3.10 Taxa specifically enriched in the pea rhizosphere

Some of the prokaryotes specifically selected by pea (Appendix Table A13) included a number of *Streptomyces* spp. (*S. lividans, S. roseosporus, S. ambofaciens, S. filamentosus* and *S. ghanaensis*) which could act as biocontrol agents by producing antimicrobials. Several *Pseudomona*s spp, including beneficial *P. fluorescens SBW25* and *P. putida*, and pathogenic *P. aeruginosa* and *P. syringae* were selected. As were nitrogen fixing *Klebsiella pneumoniae*, *Variovorax paradoxus, Rhodopseudomonas palustris* BisB18, and the Rhizobiales. Specifically within the Rhizobiales, the symbiont of pea, *Rhizobium leguminosarum* was selected but also others such as *R. etli* and *Bradyrhizobium*, and the related pathogen *Agrobacterim vitis*. Other nitrogen cycling bacteria such as, *Nitobacter*, which oxidise nitrite to nitrate, and denitryfing Methylophilales (Kalyuhznaya *et al.*, 2009) were selected, suggesting enhanced nitrogen cycling in the pea rhizosphere. Nitrogen fixation produces hydrogen as a by-product (Hunt and Layzell, 1993), and the selection of hydrogen oxidisers such as *V. paradoxus*, *R. palustris* BisB18, was observed. A number of methylotrophs, including *Methylotenera*, *Methylophilales*, some Rhizobiales, and *Variovorax paradoxus* were selected. Methylotrophy, particularly methanol oxidation is a widespread trait among soil and rhizosphere bacteria (Kolb, 2009) due to the production of methanol from pectin during growth and decomposition of plant material (Galbally and Kirstine, 2002). A number of *Mycobacterium* spp. (*M. avium, M. gastri, M. tuberculosis, M. intracellulare* and *M. smegmatis*) were selected. While these are known to cause disease in animals, some are known to aid phytoremediation by catabolising aromatics and other complex hydrocarbons (Cheung and Kinkle, 2001; Toyama *et al.*, 2011). Other selected bacteria included the halogenated aromatic compound degrader *Arthobacter chlorophenolicus*, insecticide producing *Photorhabdus luminescens*, plant cell wall polymer degrading *Clostridium phytofermentans*, and sugar fermenting *Gluconobacter*. Eukaryotes selected by pea (Appendix Table A20) included numerous fungi, such as the phylum Chytridiomycota as well as Pezizomycte, Lasiosphaeriaceae, mitosporic Phyllachoraceae and Tricholomataceae, *Neurospora*, *Podospora anserina*, *Ajellomyces*, truffle forming *Tuber*, and plant pathogens *Colletotrichum higginsianum, Moniliophthora perniciosa* and Sclerotiniaceae. Other eukaryotes included the amoeba *Naegleria*, alga Raphidophyceae, euglenozoa *Trypanosoma cruzi* strain CL Brener, and a nematode pathogen of soybean, *Heterodera glycines*.

4.3.11 Taxa enriched in the wheat and oat rhizospheres

Some of the prokaryotes selected by the cereals (Appendix Table A14) included plant pathogens *Acidovorax avenae* and *Ralstonia solanacearum*, methylotrophic *Methylibium petroleiphilum*, as well as anoxygenic phototrophs *Chlorobium ferrooxidans* and *Chloroherpeton thalassium*. *C. ferrooxidans* oxidises iron while *Rhodoferax ferrireducens*, also selected by cereals, reduces it, suggesting enhanced cycling of iron oxidation states in cereal rhizospheres. Potential PGPRs selected included *Asticcacaulis*, a producer of insectical toxins (Liu *et al.*, 1996), and *Flavobacteria* bacterium MS024-2A. Other selected prokaryotes included the neutral endophytic *Dyadobacter fermentans* DSM 18053 (Lang *et al.*, 2009), saprophytic spirochaete *Leptospira biflexa* (Picardeau *et al.*, 2008), and *Pseudoalteromonas tunicate*, a marine bacterium that is adapted to surface attached growth (Thomas *et al.*, 2008) and produces antifouling compounds (Holmstrom *et al.*, 1998), which could potentially act against eukaryotes in the rhizosphere. Eukaryotes selected by the cereals (Appendix Table A21) included nematodes Panagrolaimoidea, β-lactam antibiotic producing *Penicillium chrysogenum*, and the cryptomonad *Pyrenomonas helgolandii*.

4.3.12 Taxa enriched in the wheat and pea rhizospheres

Prokaryotes selected by wheat and pea (Appendix Table A15) included the cyanobacterium *Anabaena*, diazotroph *Azotobacter vinelandii*, silicate degrading *Bacillus mucilaginosus* (Mo and Lian, 2011) and *Neptuniibacter caesariensis*, a marine bacterium that assimilates taurine with the release of sulfoacetic acid (Krejcik *et al.*, 2008). Selected eukaryotes (Appendix Table A22) included the trypanosome *Schizotrypanum*, and the bacterivorous flagellate *Heteromita*, which has been shown to enhance the metabolism of aromatics by *Pseudomonas* spp. (Mattison and Harayama, 2001).

4.3.13 Taxa enriched in the oat and pea rhizospheres

Prokaryotes selected by oat and pea (Appendix Table A16) included nitrogen fixing *Azospirillum* sp. B510, and two strains of *Rhizobium etli* (8C-3 and IE4771). These might have been expected to be selected by pea but not necessarily by oat. Other selected bacteria included the Comamonadaceae which contains a number of plant associated

bacteria, as well as the marine *Hyphomonas neptunium*, xyulolytic *Paenibacillus* sp. JDR-2 (Chow *et al.*, 2012) and *Sphingomonas wittichii*, capable of degrading complex organic molecules such as dioxins (Wittich *et al.*, 1992). *Magnetospirillum gryphiswaldense*, which utilise ferrihydrite to produce magnetite, was also selected (Fdez-Gubieda *et al.*, 2013). Ferrihydrite is a widespread iron mineral, which may be made available to plants by the actions of *M. gryphiwaldense*, suggesting an alternative mechanism for iron acquisition. Selected eukaryotes (Appendix Table A23) were mostly nematodes, such as plant pathogenic Aphelenchoidoidea, including *Bursaphelenchus xylophilus*, as well as others such as *Caenorhabditis brenneri* and Heteroderidae.

4.4 Conclusion

The analysis of expressed protein coding sequences (i.e. mRNA) allowed the active microbes in each sample to be profiled, independent of rRNA. The taxonomic composition of samples were determined using both a high throughput marker gene based program, Metaphyler (Liu *et al.*, 2010), and a full analysis of all sequences using rapsearch2 against the whole non-redundant nucleotide collection at NCBI. This mRNA based approach produced some results consistent with the rRNA based approached used in Turner et al. (2013), but also some different ones. These differences would be largely due to the databases for each of the respective biomarkers, the way in which they were analysed and genuine biological differences between samples, such as soil type. Protein databases are much more biased towards cultured organisms than rRNA databases, and the majority of mRNA sequences remained unassigned. This suggests that for efficient use of sequences and ease of analysis, rRNA is a better taxonomic marker than mRNA, namely because the databases are smaller yet taxonomically more extensive and better curated, allowing much higher proportions of reads to be assigned.

The relative abundance of eukaryotes was only greater than soil in the pea rhizosphere, with the oat and wheat rhizospheres showing a lower relative abundance of eukaryotic proteins. This was not consistent with the higher proportions of eukaryotes in the oat and pea rhizospheres based on rRNA in Turner et al (2013). However, after converting relative abundances to transcripts per gram of soil, there were no significant differences between any rhizosphere and soil at the kingdom level. It was not possible to compare this directly to the data in Turner et al. (2013) because that data did not include an RNA internal

standard, and the rRNA was excluded from this dataset because it was biased by the Ribo-Zero treatment. Therefore, a quantitative rRNA based community profile would be a useful next step.

By comparing the data analysed by Metaphyler for both RNA and DNA it was possible to somewhat normalise their activity to their abundance in the different environments. This normalisation was based on a single sequenced DNA sample, pooled from several biological samples, for each environment. Additionally the DNA was not extracted from the same plant harvest as the RNA. These factors mean that a fully integrated metatranscriptomic and metagenomic analysis was not possible. However, the normalisation revealed that certain bacterial classes, such as Betaproteobacteria, were more active than they were abundant, while the opposite was true for other classes such as Methanobacteria. The majority, but not all, of the RNA to DNA ratios were consistent across environments, suggesting that some bacterial classes responded to plant rhizospheres differently. In the future it would be useful to perform integrated analyses of metagenomes and metatranscriptomes, as has been done in the marine environment (Shi *et al.*, 2010), using the same biological replicates for each, to determine if the observations here are reproducible. The use of rRNA to rDNA ratios should be used with caution however. The regulation of rRNA quantities has been demonstrated for a few species in pure culture, laboratory conditions. But very little is known how the majority of microbes regulate their rRNA *in situ*. There have been conflicting reports of relationships between levels of rRNA, growth rate and activity (Blacewicz *et al.* 2013).

Based on the taxonomic assignments of mRNA, rhizosphere microbiomes were distinct from soil, although the wheat rhizosphere remained the most similar. The effects of plants on the rhizosphere microbiome were largely plant species specific, although there was some shared selection, particularly between the cereals and also between both oat and pea. There were very few taxa quantitatively less abundant in the rhizspheres compared to soil. This observation suggests that generally plants did not select against particular microbes, rather they selected some more strongly than others. This does of course mean that those more strongly selected are relatively more abundant. All plants selected different taxa with potential plant growth promoting or pathogenic properties. Potential plant growth promoters included nitrogen fixers and antibiotic producers. The selection of the same functions within different taxonomic groups suggest that microbial responses to the rhizosphere are very specific, and are likely controlled by specific compounds exuded

by particular plant species. Supporting this, a number of different strains of the same species, such as *Streptomyces* and *Pseudomonas*, were specifically selected by either wheat, oat or pea compared to soil.

The high frequency of sequences from plant pathogens was interesting given the apparent health of the plants at harvest. This suggests that many plant pathogens are in fact opportunists, or that the strains detected were non-virulent and had adapted to a commensal lifestyle. There is considerable similarity between plant pathogenic and plant beneficial microbes, exemplified by numerous *Pseudomonas* spp. (Feil *et al.*, 2005; Paulsen *et al.*, 2005) and the *Rhizobium*/*Agrobacterium* group, and also by how plants respond to such associations (Damiani *et al.*, 2012; Stracke *et al.*, 2002; Wang *et al.*, 2012a).

The selection of a number of microbes by plants with known functional traits such as plant cell wall polymer degradation, methylotrophy, hydrogen oxidation, and various redox transformations of key nutrients such as nitrogen and sulphur was observed. However, just as with rRNA, the presence of mRNA derived from an organism known to carry out a particular biochemical transformation is not an indication that it was carrying out that process. The quantitative analysis of mRNA provided a useful method for determining microbiome structure in soil and the rhizospheres of crop plants and allowed some insight into the function that might be important in the rhizosphere. The functional information within that mRNA sequence will be used to determine what processes are being carried out. Some functions are known to be carried out only by specific organisms, for example, methanogenesis is an ability unique to some members of the Euryarcheaota, while others such as nitrogen fixation and antibiotic production have no taxonomic constraints. The continued use of the taxonomic information within the mRNA sequence will be to determine which taxa are carrying out which functions. This is probably the best use for such information and when fully integrated with metagenomics, these approaches will be particularly useful for comparing the microbiomes of plants in future.

Chapter 5: Optimising mRNA enrichment for soil and rhizosphere metatranscriptomics

5.1 Introduction

Ribonucleic acid (RNA) samples from both pure culture and complex environments are dominated by ribosomal RNA (rRNA). Enrichment of messenger RNA (mRNA) is a major barrier to applying metatranscriptomics to complex environments such as soil and plant rhizospheres. Several methods have been developed to enrich in for mRNA, including those based on subtractive hybridisation of rRNA (Giannoukos *et al.*, 2012; Stewart *et al.*, 2010a), enzymatic treatments (He *et al.*, 2010; Yi *et al.*, 2011) and other methods (Botero *et al.*, 2005; McGrath *et al.*, 2008). An additional challenge is how to determine the degree of success of an enrichment method. Here, several methods for enriching mRNA were tested and validated using molecular weight profiling, quantitative PCR and next generation sequencing.

5.2 Materials and methods

Total RNA samples from *Rhizobium leguminosarum* (provided by Karunakaran Ramkrishnan), soil or the rhizospheres of wheat and oat (see 2.2.2 and 2.3.1) were used to test the efficacy of mRNA enrichment by several different methods. MICROBExpress was used according to manufacturer's instructions (see 2.4.1). Samples were treated with two rounds of MICROBExpress before being converted to cDNA using Rubicon (2.3.4). Biotinylated anti-sense capture probes were generated from soil (Stewart *et al.*, 2010a). In short, primers containing a 5' T7 promoter sequence were used to amplify full length bacterial 16S and 23S rDNA. An *in vitro* transcription reaction using MEGAScript was performed to generate large quantities of anti-sense rRNA probes incorporating biotin labelled UTP and CTP. These were then hybridised to total RNA, which was removed with streptavidin coated magnetic beads (see 2.4.2). Enriched RNA was converted to cDNA using Rubicon (see 2.3.4). For duplex specific nuclease treatment, cDNA was generated from total RNA using Rubicon (see 2.3.4), and treated with DSN (Yi *et al.*, 2011). Re-amplification from the conserved tail generated by WTA2 was then performed (see 2.4.3). Ribo-Zero (Epicentre) kits were used according to manufacturer's instructions. The bacteria kit was used alone (see 2.4.4) and also in combination with the plant seed / root kit in a 4:1 ratio

(see 2.4.5). Initially the non-magnetic Meta-bacteria kit was use, but this was replaced by a magnetic version referred to as Ribo-Zero Bacteria. The amount of RNA present before and after a treatment was determined by either an Experion bioanalyser or Qubit 2.0 fluorometer (see 2.3.2). Ribosomal RNA depletion was assessed with an Experion bioanalyser, q(RT)-PCR and next generation sequencing (at TGAC). Quantitative PCR and qRT-PCR were performed with the iQ SYBR Green Supermix or Quantitech SYBR Green kit respectively (see 2.3.7). Amplification using bacterial 16S primers and plant 18S primers was used to compare treated samples with their respective un-treated samples. The proportion of rRNA in sequence data, and its taxonomic composition was determined using USEARCH (Edgar, 2010). Taxonomic data for remaining rRNA was analysed in MEGAN (Huson *et al.*, 2007) (see 2.5.2).

5.3 Results and discussion

5.3.1 Determination of rRNA depletion efficacy with sequencing tests

Based on sequencing tests, all treatments depleted rRNA to some extent (Figure 5.1). MICROBExpress provided a 4.3 fold enrichment of mRNA compared to un-treated. Biotinylated anti-sense capture probes provided a 10.7 fold enrichment of mRNA, comparable to their original implementation in marine samples (Stewart *et al.*, 2010a). Duplex specific nuclease was almost twice as effective, providing a 20 fold enrichment. Ribo-Zero Meta-bacteria enriched 24.7 fold, while addition of the Ribo-Zero Plant kit provided an enrichment of 29.3 fold compared to un-treated samples.

The increased efficacy of duplex specific nuclease over MICROBExpress and the biotinylated anti-sense capture probes might be because it works independently of sequence, so would remove archaeal and eukaryotic rRNA in addition to that of bacteria. The high efficacy of Ribo-Zero Meta-bacteria treatment was likely a result of high capture probe concentration and diversity. Eukaryotic rRNA has been shown to make up a significant proportion of rRNA in soil and plant rhizospheres, particularly those of oat and pea (Turner *et al.*, 2013). There is also the risk of host plant contamination in rhizosphere samples, at levels that would be variable and potentially high. For example, assuming 97% % rRNA in a sample, even if 90% of prokaryotic rRNA was removed from a sample containing 20% eukaryotic rRNA, the total amount of rRNA would not be able to be reduced below ≈27%. This explains the efficacy of

the combination of bacterial and plant Ribo-Zero kits, which will be used in subsequent work.

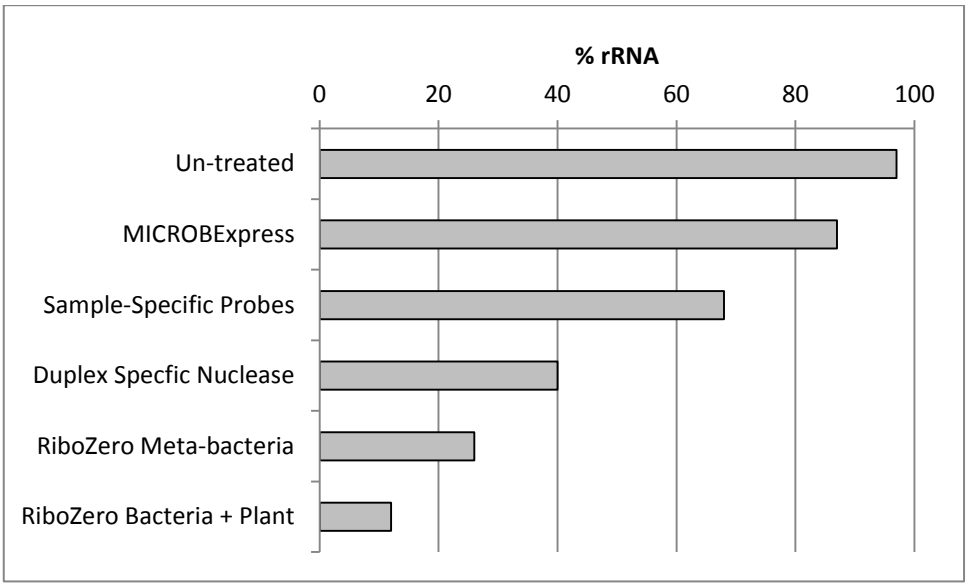


Figure 5.1: Summary of sequencing tests on rRNA depleted samples. Proportion of rRNA was determined using USEARCH and an rRNA database. Values represent single values from various test samples, with the exception of the Ribo-Zero Bacteria + Plant bar which is a mean of 20 samples.

5.3.2 Differences between subtractive hybridisation based methods

It is interesting and potentially useful to consider the differences in the subtractive hybridisation kits and how this relates to their effectiveness. When the capture probes for MICROBExpress were analysed on an Experion bioanalyser they were shown to be short oligos. In contrast, the biotinylated anti-sense capture probes and those of the Ribo-Zero kits were much larger (Figure 5.2), resembling nearly full length rRNA transcripts. Short oligos may be less effective at forming sTable 5.duplexes with longer transcripts, which may contribute to the improved performance of the longer probe methods. The ratio of probes to template is also an important factor in efficacy of depletion. The capture oligos of MICROBExpress had a concentration of ≈ 75 ng/µl. Depending on the amount of template used (recommended between 1 µg and 10 µg) this correspond to a 3 to 33 fold excess of template. By contrast there is a 4 fold excess of capture probes in the biotinylated anti-sense method. The Ribo-Zero Bacteria capture probe solution had a concentration of ≈ 1.5 µg/µl, corresponding to an excess of probes between 4.8 and 6 depending on the amount

of template (1 µg to 5 µg) and capture probes used (8 µl or 10 µl). The probes are 100 % rRNA, while the majority, but not all, of the sample template would be rRNA. This will affect the ratios of probe to true template slightly. An excess of probes over template is more desirable for efficient removal and probably contributed to the improved efficacy of the biotinylated anti-sense method and Ribo-Zero. The longer probes of Ribo-Zero and also the biotinylated anti-sense 23S probe showed additional peaks that usually correspond to 5S rRNA, and therefore may additionally remove some of this from samples. While this is claimed for Ribo-Zero kits, it not expected for the biotinylated anti-sense capture probes, where only 16S and 23S primers were used in their generation. There is known to be some homology between 23S and 5S rRNA (Ko et al., 2001), suggesting the smaller probes may be true 5S, however this homology must occur within the 23S primer sequences to allow a T7 promoter to be incorporated, allowing subsequent transcription to occur. Alternatively there may have been non-specific binding of the primers resulting in a smaller product capable of being *in vitro* transcribed by T7 polymerase, although no such sized product was observed on agarose gels. A more likely explanation therefore is premature termination of transcription resulting in a shorter transcript.

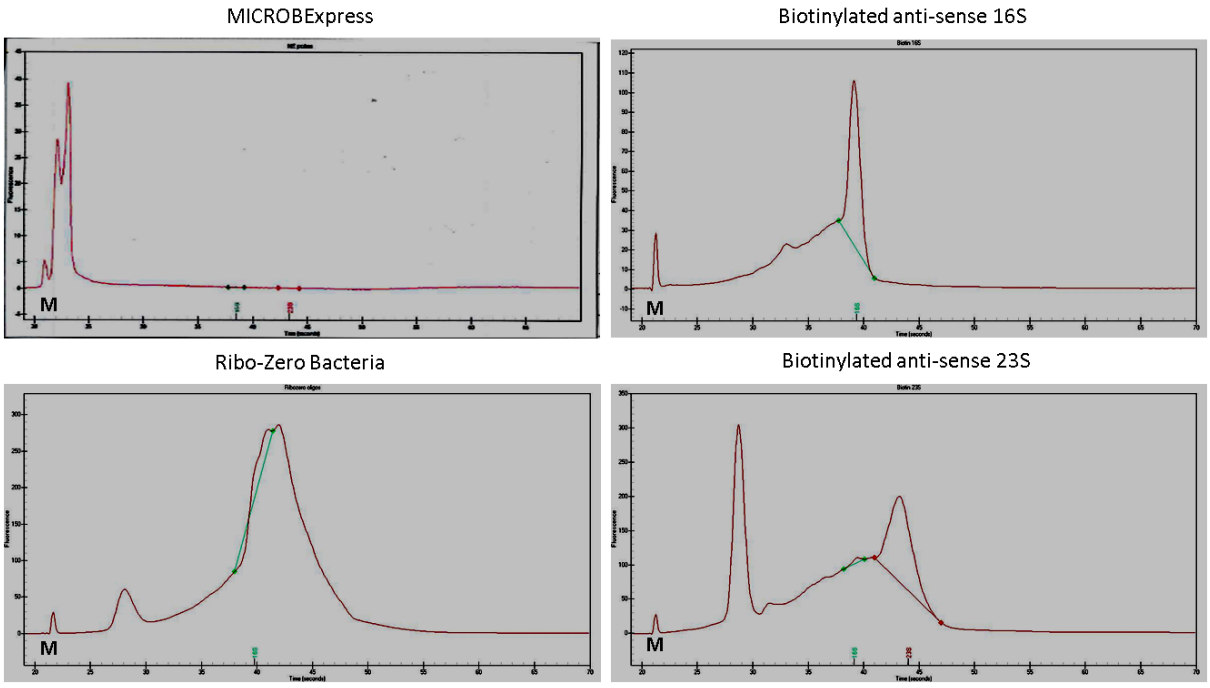


Figure 5.2: Molecular weight distributions of subtractive hybridisation capture probes determined by capillary electrophoresis on an Experion bioanalyser. The sharp, low molecular weight peak (M) represents a marker. The y-axis is relative fluorescence units,

proportional to concentration. The x-axis is time taken to run though the capillary gel, proportional to molecular weight.

5.3.3 Determining rRNA depletion using molecular weight profiles

The most accurate way to determine the level of rRNA depletion is to sequence a treated sample. However, this is both expensive and time consuming. By estimating the amount of rRNA remaining after a treatment and prior to sequencing can save considerable time, money and effort. A simple way to do this is to compare the molecular weight profiles of un-treated and treated samples. In high quality, un-treated samples, the dominance of two distinct peaks of the correct molecular weights (based on the time taken to run through the capillary gel) is indicative of intact 16S and 23S rRNA as seen in the un-treated samples. RNA from environmental samples is usually of lower quality that of pure cultures. The large, broad based peak near the marker on the molecular weight profiles shows the presence of a diversity of low molecular weight RNA species, which will include 5S rRNA, tRNAs, ncRNAs and degraded transcripts (Figure 5.2).

Treatment of total RNA with MICROBExpress, biotinylated anti-sense capture probes (both 16S and 23S) and Ribo-Zero Bacteria resulted in low concentrations of remaining RNA, which generated flat profiles that were indistinguishable from each other (Figure 5.2) even though they gave widely different levels of enrichment based on sequencing tests (Figure 5.1). This demonstrated that molecular weight profiles re unsuiTable 5.for accurately measuring the effectiveness of rRNA depletion methods.

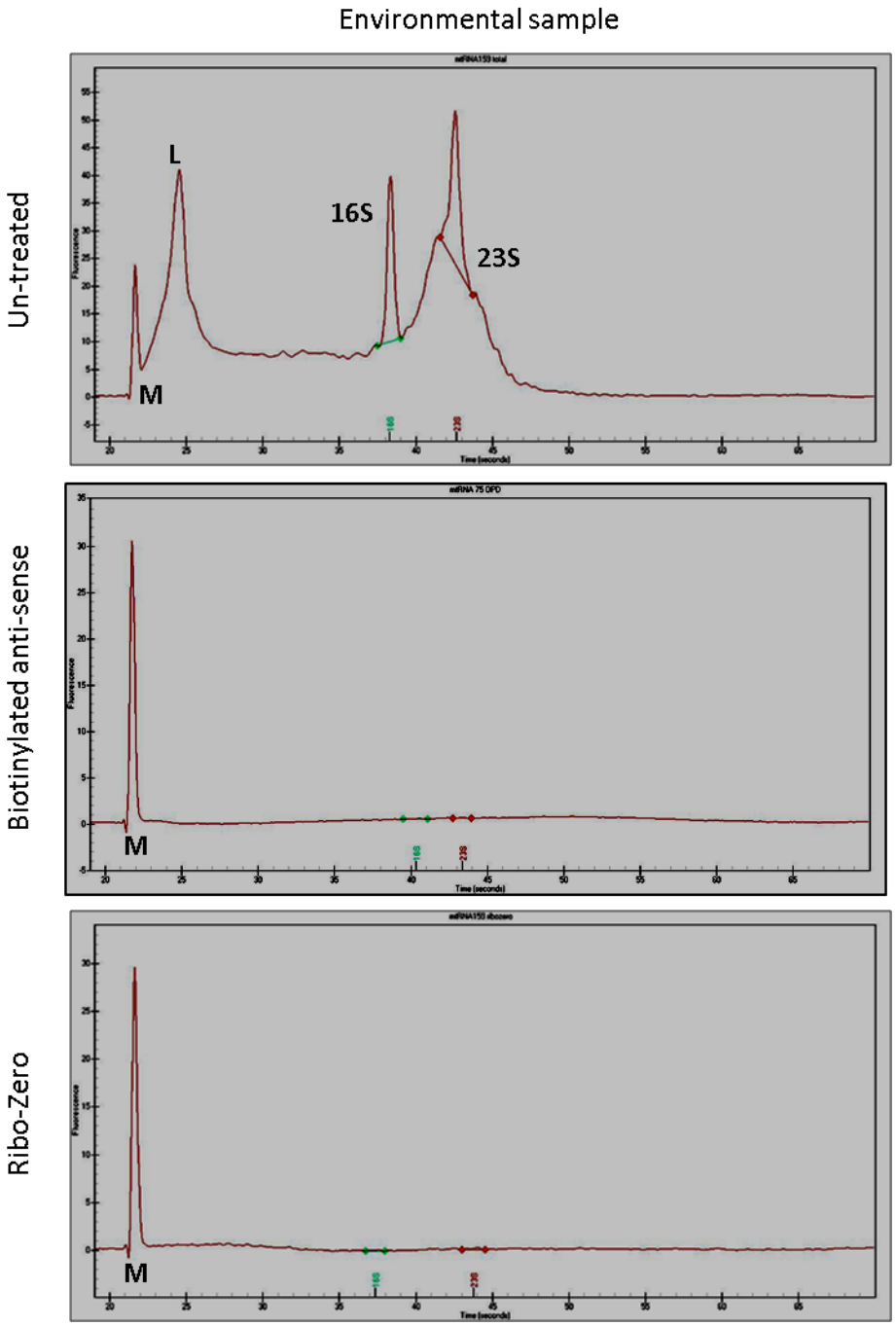


Figure 5.3: Molecular weight profiles generated by an Experion bioanalyser. The sharp, low molecular weight peak represents a marker (M), while the green and red markings highlight 16S and 23S rRNA respectively. The y-axis is relative fluorescence units, proportional to concentration. The x-axis is time taken to run though the capillary gel, proportional to molecular weight.

5.3.4 Spiking RNA to calculate efficacy of rRNA depletion

Using the bioanalyser software it was possible to calculate the proportion of rRNA in un-treated samples from the area under the peaks and the total area of the trace. However the concentration of RNA after depletion treatments was much lower than that of the starting RNA template. Even when adjusting for this, the software is not sensitive enough to allow accurate calculation of the proportion of rRNA. In attempt to overcome this, a known amount of an RNA molecular weight standard (RNA250, Ambion) was added to samples prior to treatment with MICROBExpress (Figure 5.4). The standard was of an intermediate molecular weight to that of the 16S and 23S rRNA. Treatment with MICROBExpress somewhat reduced the height of the rRNA peaks relative to the marker. However the peak corresponding to the RNA spike also reduced in height relative to the marker (Figure 5.4), suggesting that the capture oligos shared sufficient homology to remove it. The sequence of the standard was not provided by Ambion, but it was revealed that both the spike and the capture oligos contained poly-A tails, and the magnetic beads were coated in poly-T molecules. The excess of beads present in the sample would have resulted in the removal of the spike at the similar rate to the capture probes bound to rRNA. This further highlighted the inability of molecular weight profiles to accurately determine RNA depletion. In addition, the presence of poly-T molecules on the magnetic beads would remove eukaryotic mRNA, meaning that MICROBExpress is not appropriate for complex environmental samples.

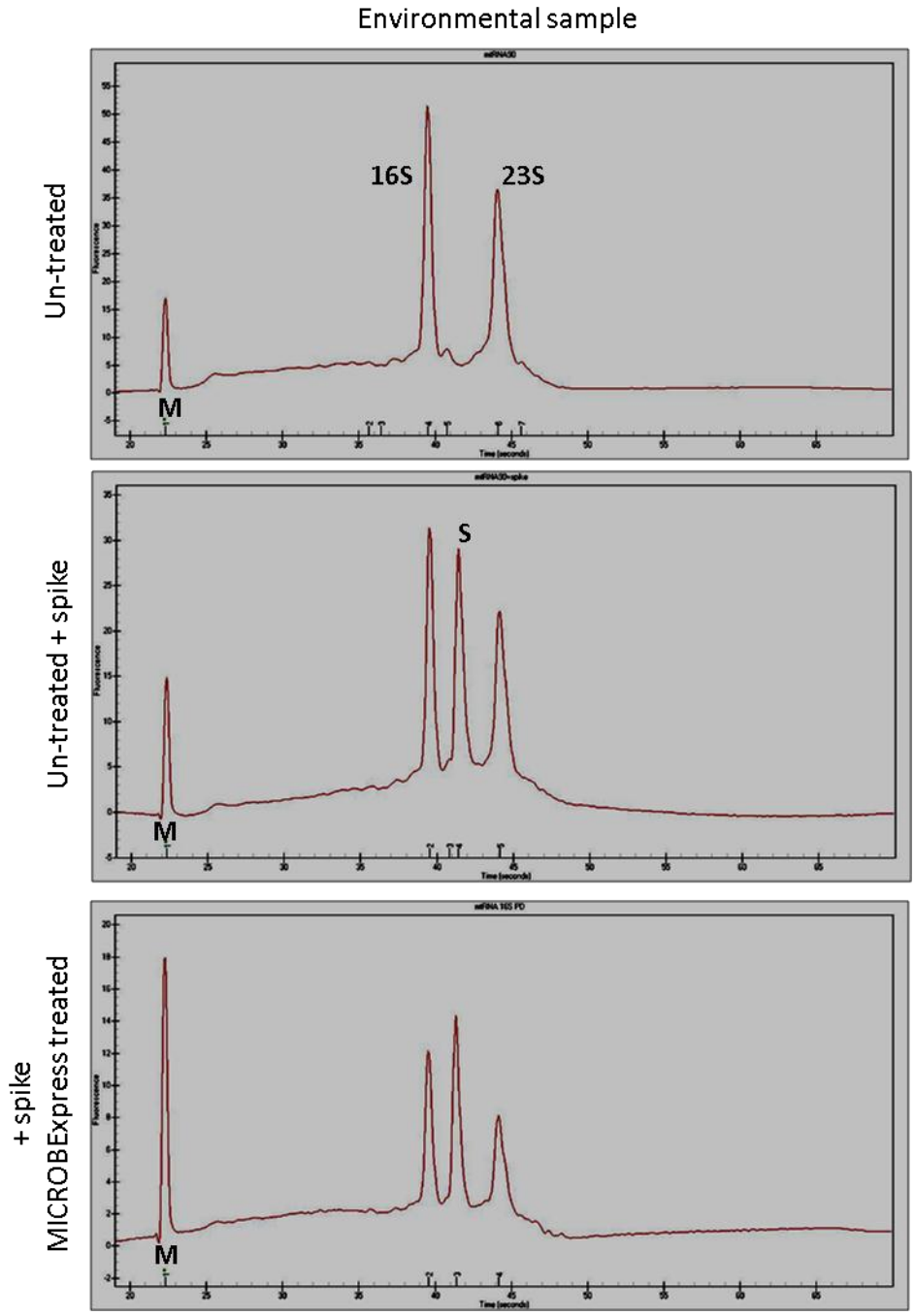


Figure 5.4: Molecular weight profiles generated by an Experion bioanalyser. The sharp, low molecular weight peak represents a marker (M). Peaks represent 16S rRNA, RNA250 standard (S), and 23S as shown. The y-axis is relative fluorescence units, proportional to concentration. The x-axis is time taken to run though the capillary gel, proportional to molecular weight.

5.3.5 Determining rRNA depletion using qPCR

As an alternative to molecular weight profiles, qPCR was used to compare the changes in bacterial 16S rRNA abundance when samples were treated with DSN and Ribo-Zero Bacteria. Because DSN treatment is performed on cDNA, an un-treated RNA sample and a sample of Ribo-Zero treated RNA were converted to cDNA to allow direct comparison between the samples. Duplex specific nuclease was effective at removing rRNA, resulting in increases in Ct values of ≈ 7 and ≈ 10 compared to un-treated for pure culture and environmental sample respectively (Figure 5.5). Treatment with Ribo-Zero Bacteria was even more effective, with Ct value increases of ≈ 18 and ≈ 16 for pure culture and environmental sample respectively (Figure 5.5). Increases in Ct value reflect a higher number of cycles required to reach a threshold level of fluorescence, indicating a lower abundance of the targeted sequence. DSN appeared to be more effective with the environmental sample than with the pure culture sample while the opposite was true for Ribo-Zero (Figure 5.5). The qPCR data for the environmental sample was consistent with the data provided by the sequencing tests (Figure 5.1), suggesting that qPCR is an effective and useful pre-sequencing step.

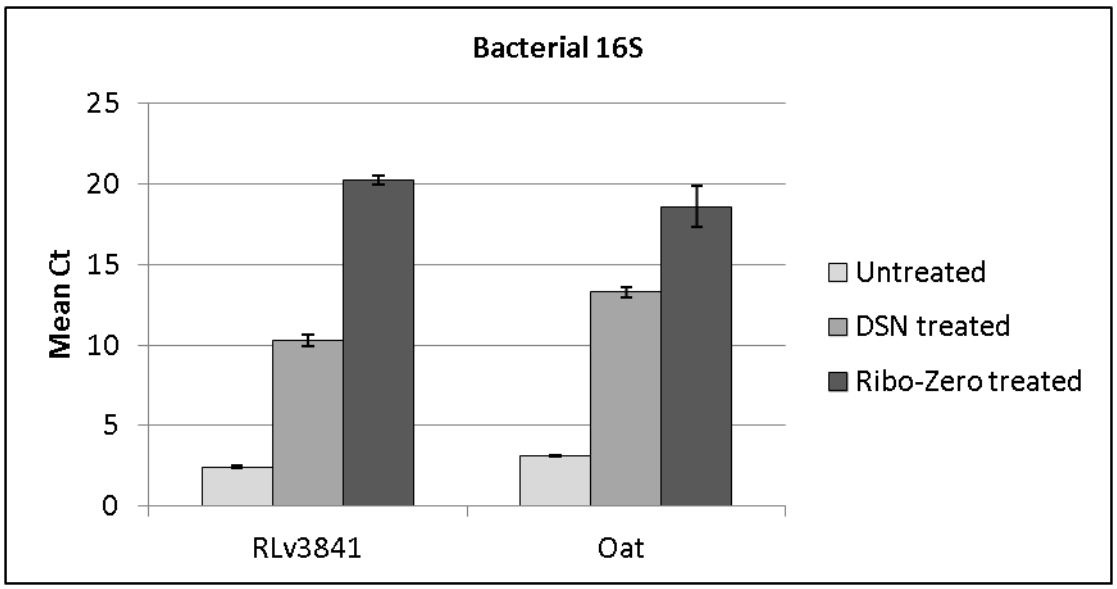


Figure 5.5: Relative abundance of bacterial rRNA in un-treated, DSN treated and Ribo-Zero treated samples from *Rhizobium leguminosarum* bv. *viceae* 3841 (RLv3841) and oat rhizosphere determined by qPCR of Rubicon generated cDNA. The Ribo-Zero Meta-Bacteria magnetic kit was used here. Values are means of 6 technical replicates, error bars represent ±1SEM. 2ng template cDNA was used for each treated sample and 0.02ng was used for the un-treated samples and Ct values were adjusted accordingly.

The RNA samples treated with the Ribo-Zero bacteria and plant combination were also validated with qPCR prior to sequencing. They were not converted to cDNA, so qRT-PCR was performed directly on the RNA using primers for both bacterial 16S and also plant 18S rRNA. Ribo-Zero treatment was effective in all tested samples, reflected in increased Ct values of between 10 and 15, resulting in Ct values in treated samples reaching ≈30, nearing the upper limit of detection (Figure 5.6). Plant 18S rRNA removal was effective for both wheat and oat rhizosphere samples, resulting in increased in Ct values of ≈8 and ≈7 respectively. Again, the Ct values reached ≈30, suggesting very low abundance of plant 18S rRNA transcripts (Figure 5.6). Interestingly the Ct value for soil did not change after treatment, but was ≈30 in the un-treated sample, indicating very little plant rRNA in the soil, and supporting the idea of host plant contamination of the rhizosphere (Figure 5.6).

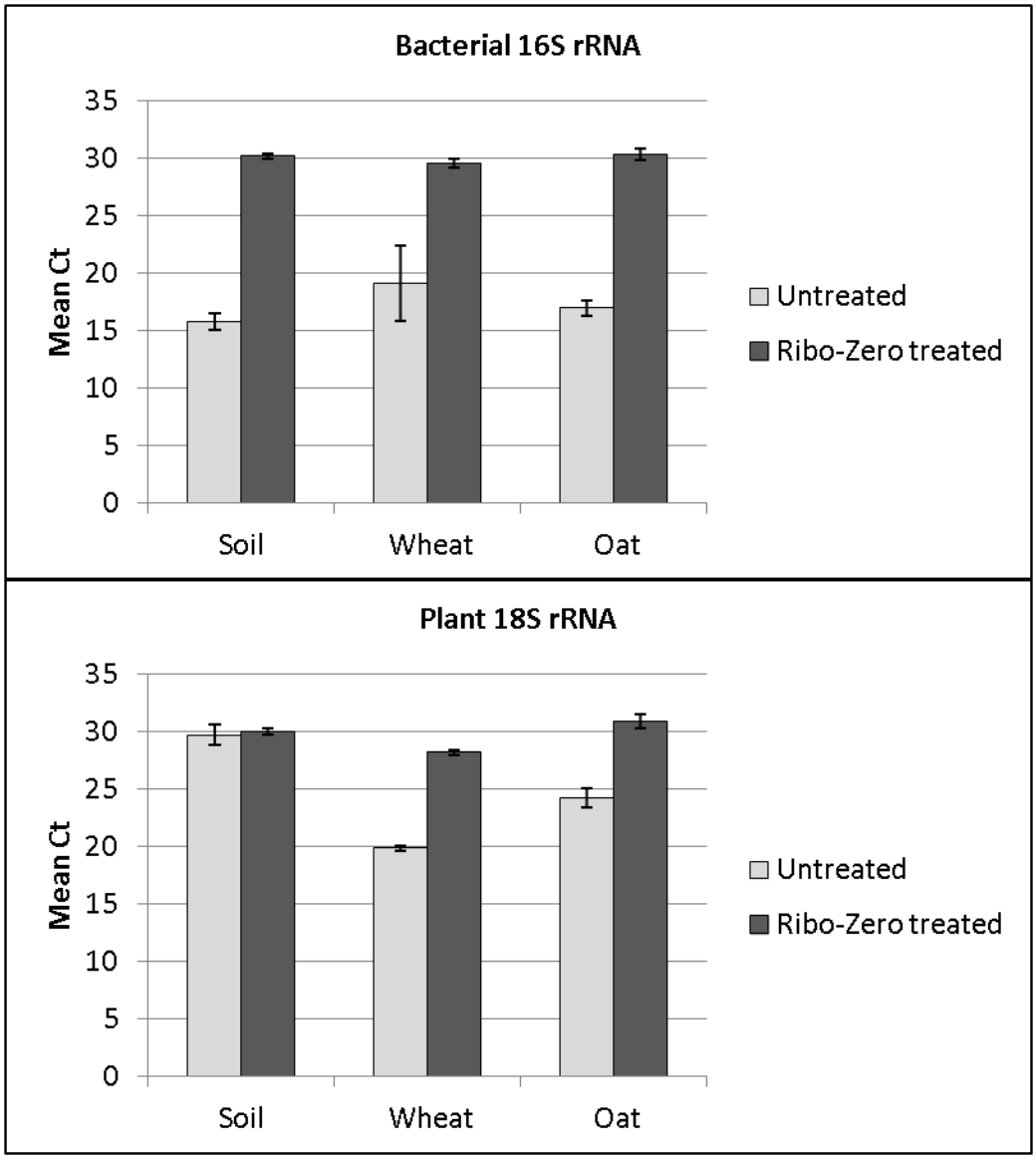


Figure 5.6: Relative abundance of bacterial and plant rRNA in un-treated and Ribo-Zero treated samples determined by qRT-PCR of RNA. The Ribo-Zero treatment consisted of a 4:1 ratio of Bacteria to Plant root and seed kits. Values are means of 3 technical replicates, error bars represent ±1 SEM.

3.3.6 Amplification of RNA using SENSATION

The high input of RNA required by Ribo-Zero, and the low yield recovered after treatment, is another potential barrier to metatranscriptomic analyses of soils, where RNA yield is typically low. Unbiased amplification of RNA provides a way to increase yield so multiple enrichments can be performed, and or precious samples are preserved. Amplification of soil and rhizosphere RNA using SENSATION (Genisphere) resulted in large quantities of RNA, however the random priming during the initial reverse transcription phase resulted in a shift in the molecular weight profile of the RNA (Figure 5.7). The distinct 16S and 23S peaks were no longer visible, and the sample consisted entirely of low molecular weight transcripts. This is the same profile that would be generated from degraded RNA, and made it impossible to tell if the Ribo-Zero treatment had removed rRNA. The amplified, Ribo-Zero treated sample was sequenced, revealing that Ribo-Zero treatment on amplified RNA failed to deplete rRNA. This may have been due to the altered stoichiometry and or thermodynamics of the reaction. High numbers of short transcripts may bind differently to the capture probes than full-length transcripts. Additionally the optimum temperatures for binding would be different for short transcripts. The compatibility lists provided by Epicentre often recommend using the highest quality RNA possible, possibly because degraded RNA is unable to be depleted.

While amplification was shown to be incompatible prior to Ribo-Zero, it might be possible to be implemented post-Ribo-Zero to amplify remaining RNA. This would avoid having to repeat Ribo-Zero reactions if there was insufficient RNA for sequencing, particularly if there was limited template available. Illumina RNA-Seq libraries can be generated from between 10 ng and 400 ng of RNA, while SENSATION amplification requires as little as 25 ng RNA. This could perhaps be advantageous over alternative methods of transcriptome amplification such as Rubicon WTA2.

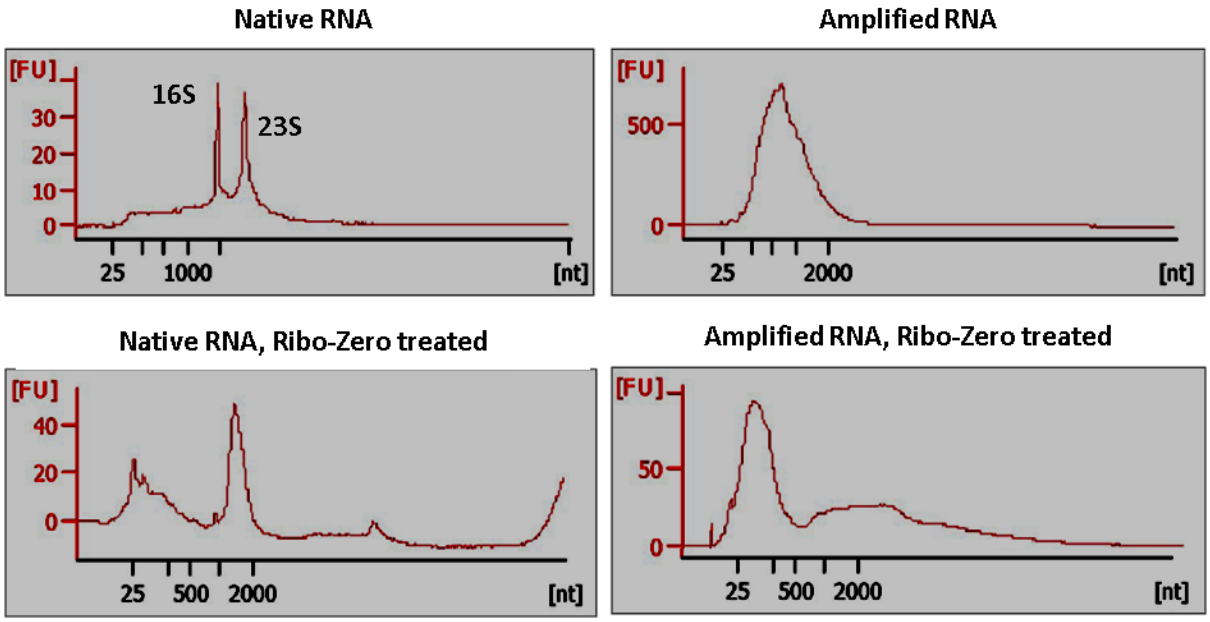


Figure 5.7: High-sensitivity molecular weight profiles generated by an Agilent 2100 bioanalyser. In the native RNA panel, distinct 16S and 23S are present. The y-axis is relative fluorescence units, proportional to concentration and the x-axis is molecular weight.

5.3.7 Removal of residual rRNA in silico

Assessing rRNA depletion post-sequencing and subsequent removal of rRNA-derived sequence reads are important steps in bioinformatic analysis of metatranscriptomic data. This reduces the amount of data input to computationally intensive analyses such as BLASTx, increasing efficiency. It also avoids the problem of translated rRNA sequences, which, depending on the alignments cut-offs used, can have significant matches to protein coding genes, such as ribokinase and group II introns associated genes. Several methods are available to do this, both standalone and also built into analysis pipelines such as MG-RAST (Meyer *et al.*, 2008). USEARCH, combined with an rRNA database can be used to assess the proportion of rRNA, and using several other commands it can produce an output file of either rRNA or non-rRNA reads. Other programs include rRNASelector (Lee *et al.*, 2011), riboPICKER (Schmieder *et al.*, 2012) and SortMeRNA (Kopylova *et al.*, 2012). USEARCH and SortMeRNA produced highly comparable results (R^2 = 0.9833, n = 20), with the SortMeRNA method slightly overestimating the level of depletion, while rRNASelector was less comparable (R^2 = 0.919, n = 19) to USEARCH and overestimated the level of depletion more than SortMeRNA (Figure 5.8).

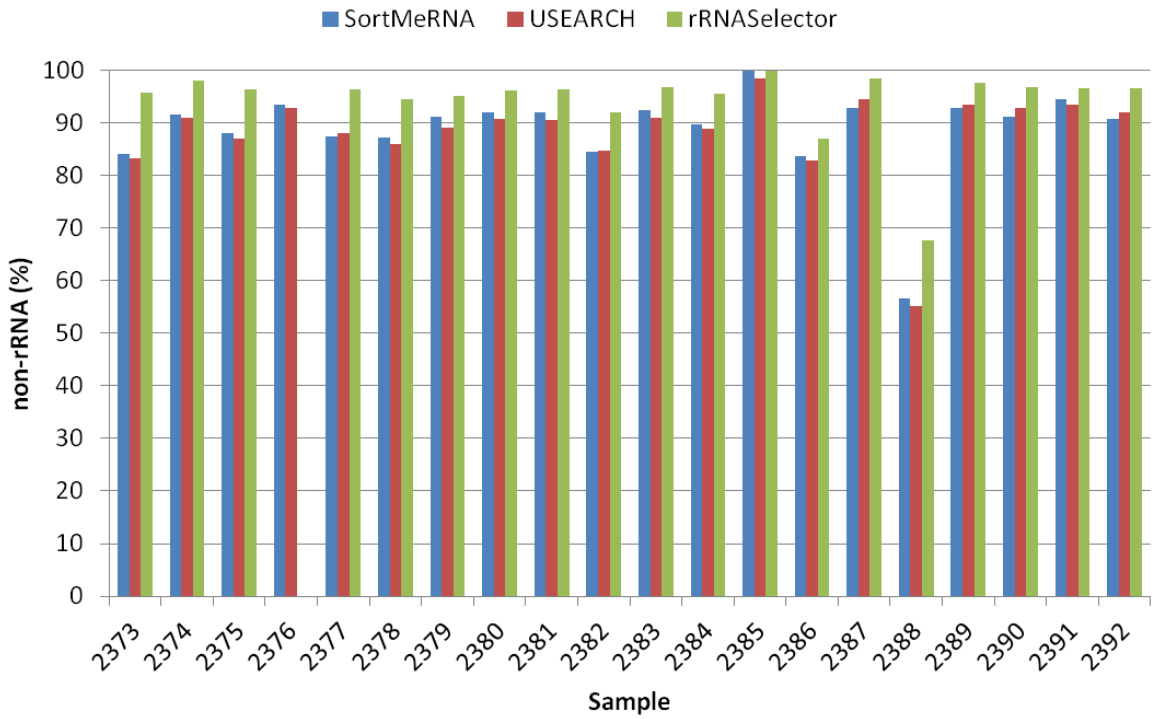


Figure 5.8: Comparison of *in silico* rRNA detection and removal approaches, SortMeRNA, USEARCH and rRNASelector. rRNASelector data was not available for sample 2376.

5.3.8 Differences in rRNA depletion due sample origin

The improved enrichment from using the additional Ribo-Zero plant kit was clear (Figure 5.1). The plant kit was chosen over other eukaryotic depletion kits because contamination of rhizosphere samples with host plant tissue could potentially introduce a high amount of plant rRNA as demonstrated by the qPCR analysis of plant 18S (Figure 5.6). However, the similarity of eukaryotic rRNA sequences may be such that the plant kit would also have removed other eukaryotic rRNAs. It has, as yet, only been documented to be compatible with higher plants, algae and mosses (Epicentre). The kits do however claim to also remove chloroplast rRNA, which is closely related to prokaryotic Cyanobacteria, and mitochondrial rRNA, closely related prokaryotic Rickettsiales. The proportion of non-plant eukaryotes is known to change in the rhizospheres of oat and pea, compared to soil and wheat rhizosphere (Turner *et al.*, 2013), additionally soil contains much lower amounts of plant 18S rRNA (Figure 5.6). At low proportions of eukaryotes, the inclusion of a eukaryotic kit may be disadvantageous as removal of bacterial rRNA is sacrificed. However, no differences in rRNA depletion were seen between samples groups (Figure 5.9). Because of the excess of

probes over template, it would almost always be worthwhile to include capture probes from a eukaryotic kit when dealing with complex environments such as soil.

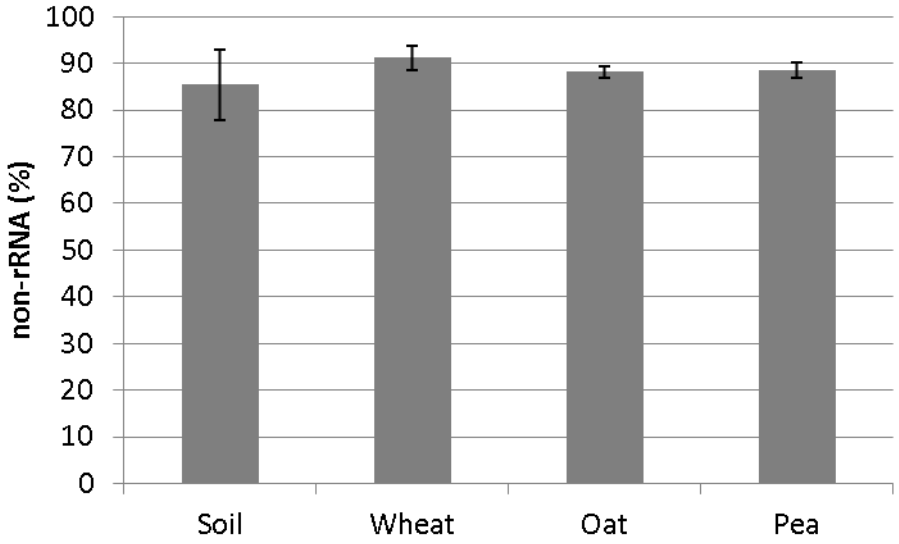


Figure 5.9: Consistency of rRNA depletion across environments using Ribo-Zero Bacteria + Plant kits. Data are means (±1SEM) of biological replicates (n=5 for each) determined using SortMeRNA.

5.3.9 Analysing the poorly depleted soil sample 2388

One soil sample (2388) showed lower level of depletion than the others, but excluding this, the remaining soil samples showed greater depletion than the rhizosphere samples (Figure 5.9). This suggests that the 4:1 ratio of kits is optimal for soil, while it could be adjusted slightly to improve depletion in rhizosphere samples. It also suggests that when not provided with plant rRNA, the probes are capable of binding to RNA of other eukaryotes. The reason for one soil sample showing lower levels of depletion might have been human error during the Ribo-Zero treatment, or due to a genuine difference in the community structure of the sample. Comparing the soil samples revealed general consistency across the well-depleted samples (2389, 2390, 2391 and 2392), while the poorly depleted sample (2388) contained higher proportions of eukaryotes (Figure 5.10a), particularly Amoebozoa and Fungi (Figure 5.10b). The comparison also highlighted that even in rRNA depleted bulk soil, Viridiplantae sequences are among the most abundant within the eukaryotes (Figure 5.10b), strongly indicating the presence of either decaying plant matter and or a seed bank in the soil. Comparing the soil eukaryotes at genus level revealed that two plant genera

(*Zea* and *Sorghum*) as well as the mycorrhizal fungal genus *Glomus*, were enriched in sample 2388 (Figure 5.10c). This suggests that a plant was growing in the soil sample at the time of harvesting for RNA extraction. While any weed plants emerging from the surface of soil were removed throughout the growth period, and any visible root or shoot structures discarded from the sampled soil , it appears that some plant material was carried over. This highlights the importance of vigilance when dealing with un-planted control samples, to ensure that as little plant material as possible is carried over. But it also raises an important point about the nature of soil used for such experiments. It is very difficult to find soil where plants are not, or have not been growing. The shaping of a soil microbiome by a plant may be maintained after a plant has been removed, although the timescales of reversion to a less plant influenced community are not known. When interpreting results, it is therefore important to recognise and account for the fact that all soils have potentially been plant influenced at some point.

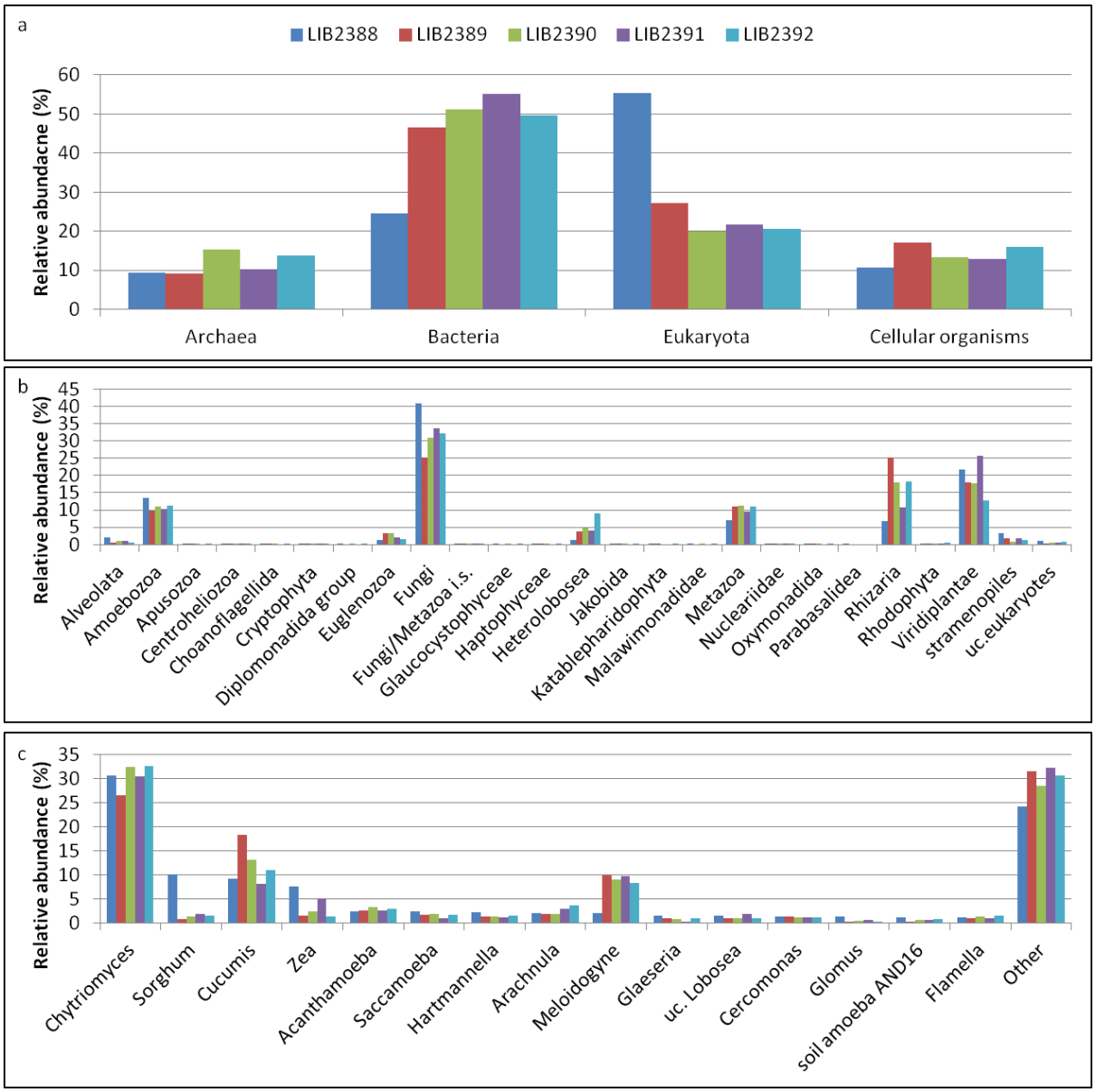


Figure 5.10: Analysis of SSU rRNA remaining in soil samples treated with Ribo-Zero Bacteria + Plant kits using USEARCH and MEGAN. Data for kingdom and eukaryotic phylum and genus levels are shown by panels a, b and c respectively. Percentage relative abundances are within each taxonomic rank. i.s. refers to incertae sedis, uc. refers to unclassified.

5.3.10 Identifying rRNA not removed by MICROBExpress and Ribo-Zero Bacteria

Comparing the distribution of remaining rRNA sequences from various treatments is also interesting. MICROBExpress treatment resulted in a decrease in relative abundance of bacterial sequences, and a corresponding relative increase in archaeal and eukaryotic sequences (Figure 5.11a and 11b). This shift is more pronounced in a sample treated with

Ribo-Zero Bacteria alone (Figure 5.11c). The vast majority (84 %) of sequences in the Ribo-Zero treated sample were eukaryotic, while bacteria represented < 3 % of sequences. Archaeal sequences were also reduced almost 5fold compared to the un-treated samples, suggesting that Ribo-Zero Bacteria is highly effective at removing prokaryotic rRNA from both bacteria and archaea. This is consistent with the species compatibility list provided by Epicentre which states that the Bacteria kit is effective with two species of *Sulfobus* (Crenarchaeota) and will remove 16S and 23S rRNA from three methanogenic species (Euryarchaeota). It is also stated to work with the eukaryotic protozoan *Acanthamoeba*.

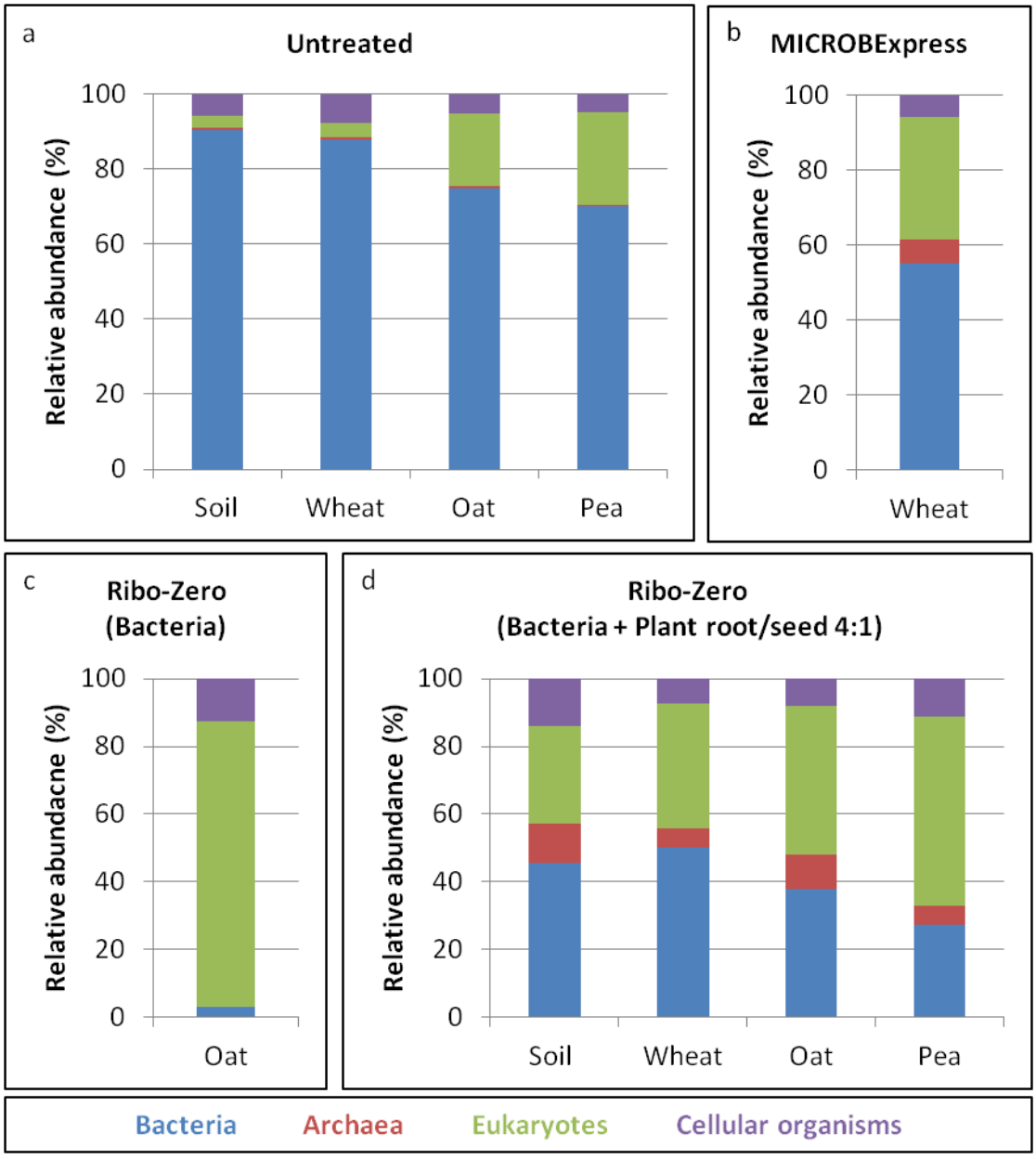


Figure 5.11: Relative abundance of domains in un-treated and rRNA depleted samples, based on UESEARCH and MEGAN analysis of SSU rRNA sequences. For MICROBExpress and Ribo-Zero (Bacteria) only single test samples were sent for sequencing.

There were higher proportions of Cellular organism in all the rRNA depleted samples (Figure 5.11). These are assigned from the most highly conserved rRNA sequences, which would be expected to be abundant in the capture probes of the various kits. So their enrichment after rRNA depletion is surprising. MICROBExpress and Ribo-Zero Bacteria kits are both designed to remove bacterial rRNA and are highly effective with organisms such as *E. coli* and *B. subtilis*. Ribo-Zero outperforms MICROBExpress when dealing with both pure culture (Giannoukos *et al.*, 2012) and environmental samples (Figure 5.1). The capture probes of Ribo-Zero are longer and more concentrated than those of MICROBExpress. They also presumably possess a higher diversity of rRNA sequences. Both are capable of removing rRNA from Acidobacteria, Actinobacteria, Gemmatimonadetes, Nitrospirae, Proteobacteria, and also unclassified Bacteria (Figure 5.12b). Proteobacteria and Actinobacteria phyla include some of the most well studied organisms, so it is to be expected that even MICROBExpress is effective with these organisms. The depletion of Acidobacteria and unclassified Bacteria is somewhat surprising given they are poorly characterised organisms. Some organisms were resistant to depletion, particularly, as expected, the Crenarchaeota, Euryarchaeota and environmental samples, but also Firmicutes (Figure 5.12b). There were differences in depletion of some taxa by the two kits, Bacteroidetes and Cyanobacteria were depleted by MICROBExpress and enriched by Ribo-Zero, while the opposite was true for Chloroflexi and Planctomycetes (Figure 5.12b).

Unexpectedly there was some depletion of eukaryotic rRNA sequences with both kits and there were variations in the depletion of eukaryotes. Alveolata, Metazoa, and to a lesser extent Fungi and Stramenopiles were depleted by both kits, while some including Ameobozoa, Centroheliozoa, Choanoflagellida, and Malawimonadidae were enriched by both kits (Figure 5.12a). Cryptophyta and Oxymonadidia were depleted by MICROBExpress and enriched Ribo-Zero, while the opposite was true for Euglenozoa and Heterolobosea. The depletion of eukaryotes might be due to genuine similarities between the capture probes of a kit and the rRNA sequences of an organism, or potentially mitochondrial and chloroplast rRNA, which share common ancestry with those of bacteria. Ribosomal RNA processing is known to occur in a number of organisms (Evguenieva-Hackenberg, 2005), and the nature and extent of this in different organisms may alter the interactions of the

rRNA with the probes. Such interactions may also be altered by differences in probe length, and degradation of the rRNA in the samples.

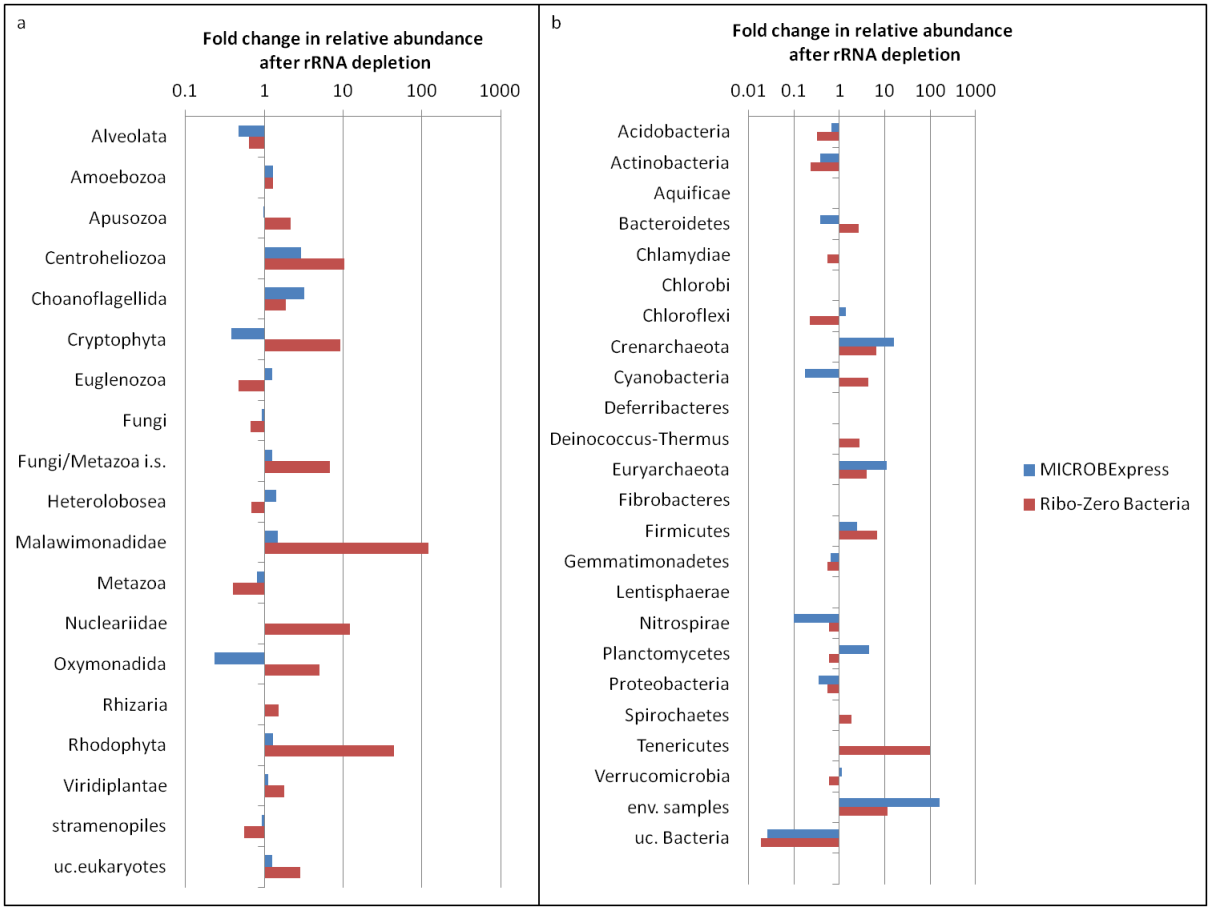


Figure 5.12: Fold changes in relative abundance of eukaryotic (a) and prokaryotic (b) phyla after treatment with either MICROBExpress or Ribo-Zero Bacteria. i.s. refers to incetae sedis, uc. refers to unclassified, and env. refers to environmental.

5.3.11 Identifying rRNA not removed by the Ribo-Zero Bacteria and Plant combination

The combination of plant and bacterial Ribo-Zero kits proved to be the most effective way of depleting rRNA (Figure 5.1), but it did not remove all rRNA. This is unlikely to be due to the saturation of the capture probes with rRNA due to the several fold excess of probes to sample. It is more likely that the probes in any given combination of kits are not capable of binding to rRNA sequences from particular organisms. When one fifth of the Ribo-Zero capture solution was derived from the plant root/seed kit the relative abundance of eukaryotic and archaeal sequences increased relative to when the bacteria kit alone was used, suggesting some of the bacterial removal had been sacrificed to remove eukaryotic

sequences. However, the two kits in combination were more effective than the Bacteria kit alone for reducing the total abundance of rRNA (Figure 5.1).

Ribosomal RNA sequences from Alveolata, Amoebozoa, and Fungi/Metazoa incertae sedis were effectively removed by the Ribo-Zero combination across all samples. While those of Apusozoa were only depleted in oat and pea rhizospheres, and unclassified Eukaryotes were only depleted by wheat and soil (Figure 5.13a). A number of eukaryotic sequences were enriched after treatment with the Ribo-Zero combination, including those of Centroheliozoa, Choanoflagellida, Heterolobosea, Rhizaria and Rhodophyta. Sequences from the Viridiplantae, the target of the plant kit, were slightly enriched in wheat and pea rhizospheres, but largely unchanged in the oat rhizosphere and soil. Their proportion in the un-treated sample is likely to be highly variable due to native plant RNA in soil and potential contamination with host plant material. These results suggest that the Ribo-Zero plant kit is effective at removing rRNA from a variety of, but not all, eukaryotes.

Cyanobacteria, Chloroflexi and Nitrospirae were among those prokaryotes successfully depleted. Some prokaryotes, including Bacteroidetes, Crenarchaeota, Euryarchaeota, Firmicutes, Spirochaetes and Tenericutes were resistant to depletion by the combination of the Bacteria and Plant kits (Figure 5.13b). These taxa also resisted depletion by the Bacteria kit alone (Figure 5.12b), and may prove a useful target for future improvements to the Ribo-Zero product line. Some changes in depletion were only observed in one environment, for example, Chlamydiae were enriched in the soil samples while Deinococcus-Thermus were enriched only in the pea rhizosphere (Figure 5.13b).

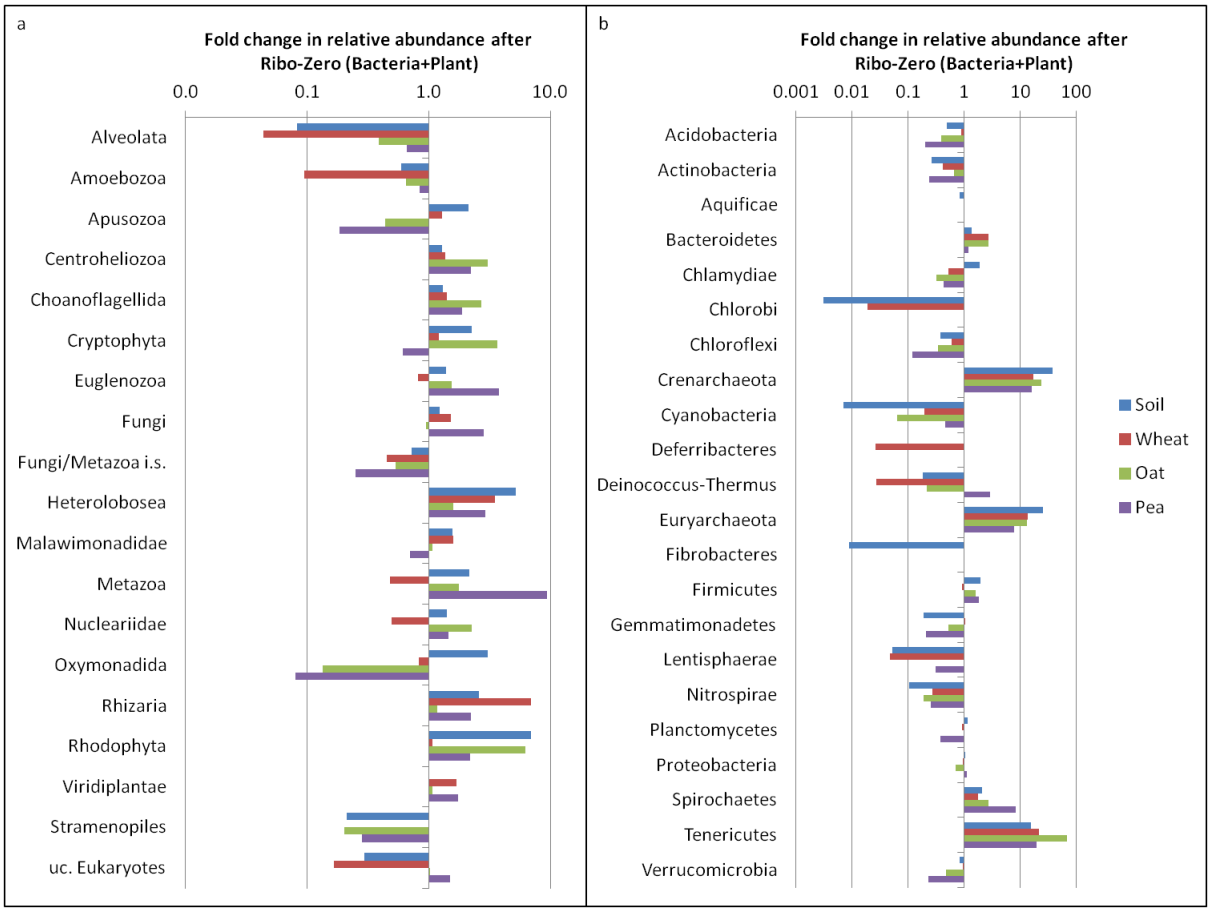


Figure 5.13: Fold changes in relative abundance of eukaryotic (a) and prokaryotic (b) phyla after treatment with either Ribo-Zero Bacteria + Plant. i.s. refers to incetae sedis, uc. refers to unclassified.

5.4 Conclusion

This work has demonstrated an effective strategy for the removal of the majority of rRNA from soil derived environmental samples. Subtractive hybridisation based approaches in the form of commercial kits, particularly Ribo-Zero, are strongly recommended due to their efficacy and straight-forward protocols. The comparisons of total and rRNA depleted samples have proved interesting, suggesting unexpected depletion of other eukaryotes by the Ribo-Zero Plant kit, and highlighted the taxa missed by Ribo-Zero Bacteria kits. They do have to be interpreted with some caution however, as the treated samples are not derived from the same un-treated samples, and only SSU rRNA has been analysed. It is likely that the community profile of treated samples would be quite different, but also more difficult

to interpret, if based on LSU rRNA, but again this would proved useful in determining which kits were best suited to different environments.

The results seen here highlight the importance of knowing the population structure of a microbial community before deciding on a method for depletion of its rRNA. If the relative proportions of different taxonomic groups (i.e. bacteria, archaea, eukaryotes) are known, an appropriate combination of rRNA depletion products can be used. In certain instances, the rRNA from specific taxa might also require identifying and removing, such as human or mouse sequences contaminating those of gut microbiomes. If metatranscriptomics was to be used to study microbial endophytes of plants, or information on the metabolism of the root was sought as well as that of its associated microbiome, the ratios of different kits may need to be adjusted to achieve sufficient removal. Combined with Illumina sequencing, the rRNA depletion approach demonstrated here will allow the metabolism of soil and rhizosphere microbes to be probed in unprecedented detail.

Chapter 6: Metabolic mapping of crop plant rhizosphere microbiomes

6.1 Introduction

A metatranscriptome is the total RNA pool of a group of organisms in a given environment at any given time. It provides a snap shot of community wide gene expression, revealing the activity of taxa and metabolic pathways. Any transcriptome is dominated by housekeeping transcripts, particularly ribosomal RNA (rRNA) which can make up around 97% of RNA species from soil samples for example. Removal of rRNA has been a significant challenge to metatranscriptomics until recently (Ciulla *et al.*, 2010). Early metatranscriptomic studies were limited to oceans (Gilbert *et al.*, 2008; Mason *et al.*, 2012; Ottesen *et al.*, 2011; Poretsky *et al.*, 2009; Shi *et al.*, 2009a; Shi *et al.*, 2010; Shi *et al.*, 2009b; Stewart *et al.*, 2010b), but more recently other environments have been probed using this technique, including lakes (Vila-Costa *et al.*, 2013), human (Gosalbes *et al.*, 2011; Ponten, 2011), mouse (Xiong *et al.*, 2012) termite (Raychoudhury *et al.*, 2011; Tartar *et al.*, 2009) and nematode (Bomar *et al.*, 2011) guts, deep sea hydrothermal vents (Lesniewski *et al.*, 2010; Lesniewski *et al.*, 2012) and soil (Bailly *et al.*, 2007; Damon *et al.*, 2012; de Menezes *et al.*, 2012; Urich *et al.*, 2008). The soil studies to date have been limited to fungi or other eukaryotes, due to the ease of mRNA enrichment. The exception to this is the study by Urich et al. (2008), which did not involve enrichment of mRNA, so little functional information was obtained.

Compared to marine and other generally oligotrophic environments, soils have higher microbial density, diversity and activity. Therefore, metatranscriptomes of soils have a larger and more heterologous proportion of rRNA. This has resulted in a significant challenge to deep sequencing of microbial metatranscriptomes from soil environments. Plants add another level of complexity to a soil microbiome, by selecting certain taxa and altering gene expression in their rhizospheres. Differential expression of particular functional marker genes in plant rhizospheres has been observed (Dandie *et al.*, 2011; Haichar *et al.*, 2012; Sharma *et al.*, 2005; Shrestha *et al.*, 2010), but the global metabolic profiles of soil and rhizosphere environments have yet to be studied in detail.

After removal of rRNA, transcripts encoding other housekeeping functions will still be dominant, so large amounts of sequence are required to obtain sufficient data on metabolic pathways and subtle changes that might play important roles in adaptation to different environments. Most studies to date have used Roche's 454 pyrosequencing

platform (Margulies *et al.*, 2005), which can generate up to 1 million reads in one sequencing run. Mean read lengths can now reach 1000bp, though with earlier chemistries this was lower (400bp and 700bp). Illumina sequencing (Bentley *et al.*, 2008), which generates several billion reads of between 100bp and 250bp, has the potential to allow metatranscriptomes to be studied in unprecedented detail. It has recently been used in high-throughput analysis of PCR amplified 16S rRNA genes (Caporaso *et al.*, 2012; Degnan and Ochman, 2012)

One limitation of many sequence-based analyses of microbial communities is that only relative comparisons can be made between samples due to unknown sequencing depth. A technique developed by Moran and colleagues to quantify sequencing depth using an RNA internal standard (RIS) showed that the ocean water they sampled contained 7.8×10^8 mRNA transcripts per gram (Gifford *et al.*, 2011). A known sequencing depth allows normalisation of transcripts across samples, indicates how much more sequencing is required to capture the entire diversity of a population, and allows calculation of transcripts per unit mass of substrate sampled. The rhizosphere is considered to be a more active region than bulk soil (Grayston *et al.*, 1997; Vale *et al.*, 2005) as demonstrated by enzyme assays, but this has not been shown at the level of global transcription.

To date it has been difficult to compare different metatranscriptomic studies due to differences in target environments and organisms, extraction techniques, RNA processing, sequencing and bioinformatic analyses. Additionally, conclusions have usually been drawn from single or a few samples, and thus have lacked statistical validation. Due to the financial constraints of high-throughput sequencing, a balance between sequence depth and biological replication is required. Is it more informative to have 10 Gb of sequence from one sample, 2 Gb from 5 samples from the same environment, or 1 Gb from 10 samples from two environments? The answer to this depends on the goal of the study, the resources available and the nature of the samples. Here, the use of Illumina sequencing and biological replication (n= 4 – 5) allowed comparative metatranscriptomic analysis of rRNA depleted RNA, resulting in quantitative metabolic mapping of soil, and the rhizospheres of wheat, oat and pea.

6.2 Materials and methods

Seeds were planted in Bawburgh farm soil (Baw3) and grown for 5 weeks before rhizosphere soil was harvested (see 2.2.2). Nucleic acids (DNA and RNA) were extracted from wheat, oat and pea rhizospheres, plus unplanted controls, using the PowerSoil RNA isolation and DNA elution accessory kits (see 2.3.1). Each sample was spiked with a known amount of the RIS prior to extraction of nucleic acids (see2.3.9). Five DNA samples from each environment were pooled, and one pooled sample from each environment was submitted to TGAC (2.5.1) for sequencing on the Illumina HiSeq (one lane each). Five RNA samples from each environment were treated with RiboZero Bacteria and Plant seed/root kits in a 4:1 ratio (see 2.4.5). Success of rRNA depletion was determined using qRT-PCR (2.3.7) before samples were submitted to TGAC for sequencing on the Illumina HiSeq (two samples per lane) (see 2.5.1).

All samples were de-multiplexed and quality filtered as standard. A 4 million read subset of each DNA sample was uploaded into MG-RAST (Meyer *et al.*, 2008) and analysed using default parameters. The RNA data were analysed by two different approaches, first at The European Bioinformatics Institute (EBI, Hinxton, UK) and then at TGAC (Figure 6.1) (see 2.5.3). The computing analysis at TGAC was largely performed by Mark Alston as part of the collaborative Capacity and Capability Challenge project (CCC_II_10). Residual rRNA sequences were removed from the RNA sample *in silico* using SortMeRNA (Kopylova *et al.*, 2012) and the number of copies of RIS recovered was determined using USEARCH (Edgar, 2010). Sequencing depth and transcriptional activity per gram of soil were then calculated (see 2.5.4). Non-rRNA reads were filtered using Sickle (https://github.com/najoshi/sickle) then analysed using Metaphyler to determine taxonomic composition. A subset of the data (25 million reads based on the lowest read count sample) were analysed using rapsearch2 (Zhao *et al.*, 2012), a reduced alphabet BLAST-like algorithm, against the non-redundant nucleotide collection at the National Centre for Biotechnology Information (NCBI). Output files were uploaded into MEGAN (Huson *et al.*, 2007) to visualise and compare hits by their SEED (Overbeek *et al.*, 2005) and KEGG (Kanehisa and Goto, 2000) assignments. Pair-wise comparisons were made between each plant rhizosphere and soil using t-tests with a 95% confidence interval. Multiple comparisons were statistically tested using analysis of variance (ANOVA). In addition, all samples in full were uploaded to MG-RAST (Meyer *et al.*, 2008) and analysed using default parameters. Multidimensional scaling analysis was performed in PRIMER6. Data were normalised to a percentage then square root

transformed before a Bray-Curtis similarity matrix was generated and used to plot data on x and y axis to generate the plot. The abundance of KEGG orthologs (KO) classes derived from rapsearch2 and MEGAN were mapped to KO pathways using STAMP (Parks and Beiko, 2010).

At EBI, a subset of reads (mean 92 million) were analysed using the EBI Metagenomics Portal courtesy of Peter Sterk. (Figure 6.1). SeqPrep was used to merge mate pairs and perform additional quality filtering. If reads did not overlap, both reads were used in the analysis. Further filtering, including a 100 bp cut-off was applied using Trimmomatic. Residual rRNA sequences were removed from the RNA sample *in silico* using rRNASelector (Lee *et al.*, 2011). Non-rRNA reads were analysed by InterProScan 5 (Quevillon *et al.*, 2005; Zdobnov and Apweiler, 2001) to generate and InterPro and Gene Ontology (GO) assignments from the UniProt database. Pair-wise comparisons were made between each plant rhizosphere and soil using the t-test with 95% confidence interval.

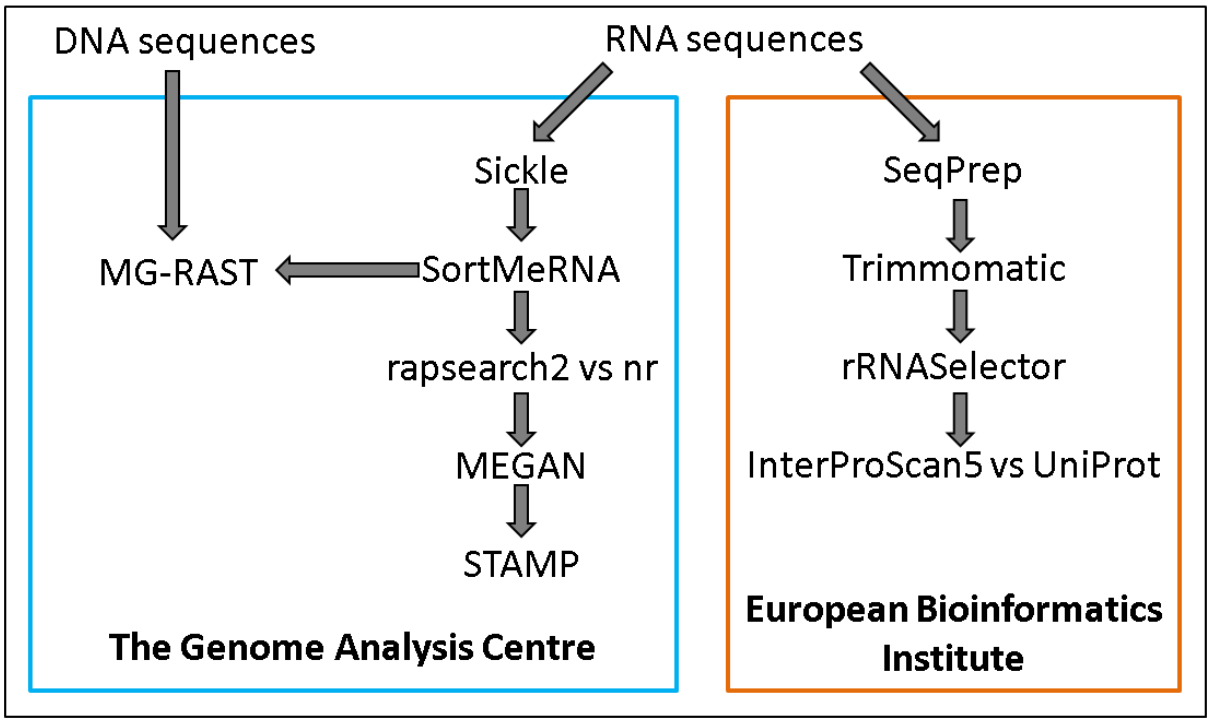


Figure 6.1: Workflow of bioinformatic analyses performed on DNA and RNA sequences for metabolic mapping.

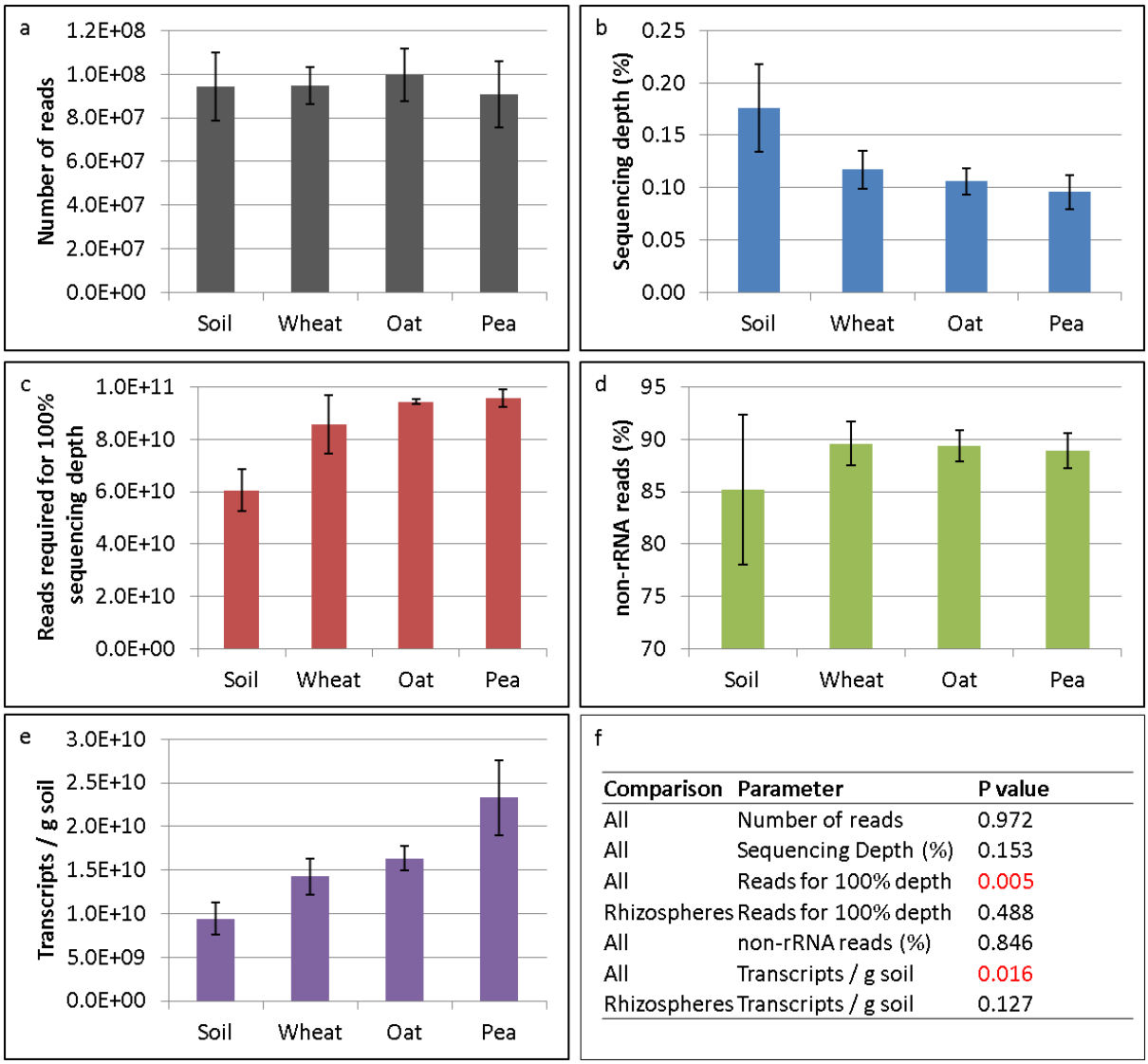
6.3 Results and discussion

6.3.1 Sequencing summary, calculation of sequencing depth and transcriptional activity

The number of pairs of reads per sample ranged from 31 million to 133 million, with a mean of 95 million. Variation in read number was independent of the origin of the sample (P=0.972) (Figures 6.2a and f). Based on the recovery of the RIS sequence, sequencing depth was highest in soil and lower in the rhizospheres, though there were no statistically significant differences between environments (P=0.153) (Figures 6.2b and f). Average sequencing depth was 0.12%, indicating that 12 out of every 10,000 transcripts were sequenced. This is a conservative estimate because the identity cut-off used to identify RIS sequences was 1. In the marine study in which the RIS approach was first demonstrated, sequencing depth was 0.000009% and 0.000015% for the two samples analysed (Gifford *et al.*, 2011). By comparison the current study sequenced between 8000 and 13000fold deeper. Despite the unprecedented level of sequencing depth attained here, additional sequencing effort, approaching 100 billion reads, would have been required to sequence all the transcripts in the soil and rhizosphere environments (Figure 6.2c). This highlights the richness of soil as a microbial habitat. This amount of sequence is prohibitively expensive at present, particularly when comparisons between different environments, requiring

biological replicates, are desired. In the near future however, novel sequence platforms and chemistries offering ultra-high-throughput (Eid *et al.*, 2009; Rothberg *et al.*, 2011) might make such studies feasible.

All samples were successfully depleted of rRNA, to around 10%, with the exception of one soil sample, which retained 43% of its rRNA (Figure 6.2d, see 5.3.9). Overall there were no differences between the level of rRNA depletion (Figure 6.2f), suggesting the chosen approach using bacterial and plant Ribo-Zero kits was suiTable 6.for both soil and rhizosphere environments. Taking into account the recovery of the RIS and the number of non-rRNA reads sequenced, the global transcriptional activity of each environment was calculated. Rhizospheres had a higher transcriptional activity than to soil (Figure 6.2e) and this was statistically significant (P=0.016). However there were no differences between rhizosphere samples (P=0.127) (Figure 6.2f). One gram of unplanted soil contained 9.41 billion transcripts, while one gram of seawater was shown to contain 0.78 billion transcripts (Gifford *et al.*, 2011). Soil is a densely populated and generally nutrient rich environment compared to oligotrophic oceans, so the 12fold higher level of transcriptional activity is not surprising, however this is the first time this quantitative comparison has been made. Wheat increased the transcriptional activity of soil 1.5fold, though this difference was not statistically significant (P=0.119). Oat and pea however increased the transcriptional activity in soil by 1.7 and 2.5fold respectively, and these differences were statistically significant (P=0.017 and 0.018 respectively). Interestingly, the pattern here is similar to that of the relative abundance of eukaryotes in these plant rhizospheres (Turner *et al.*, 2013). In fact there was reasonable correlation between the two (R^2=0.7799). These data indicate the rhizospheres of wheat, oat and pea have transcriptional activities 18, 21 and 30fold higher than an oligotrophic ocean environment respectively. It would be interesting to compare these transcriptional activities to other terrestrial environments and see what effect plant type, soil and other biotic or abiotic factors have on different levels of transcription in the soil. The importance of the addition of RNA standards in metatranscriptomic studies has recently been highlighted (Moran *et al.*, 2013), so these comparisons will be able to be made in the near future.



Comparison	Parameter	P value
All	Number of reads	0.972
All	Sequencing Depth (%)	0.153
All	Reads for 100% depth	0.005
Rhizospheres	Reads for 100% depth	0.488
All	non-rRNA reads (%)	0.846
All	Transcripts / g soil	0.016
Rhizospheres	Transcripts / g soil	0.127

Figure 6.2: Summary of (a) read numbers, (b) sequencing depth, (c) further sequencing effort requirements, (d) proportion of non-rRNA reads, (e) transcriptional activity per gram of soil. Statistical comparisons (f) were made using analysis of variance (ANOVA). Value are means of biological replicates ±1SEM, where n = 5 for soil, oat and pea rhizospheres, and 4 for the wheat rhizosphere.

Based on the rapsearch2 and MEGAN analyses, the proportion of reads able to be assigned to a known protein-coding genes was low (Figure 6.3a). For soil around 1.5% of reads were assigned a function, while this was slightly higher for the rhizosphere samples at almost 3% in the pea rhizosphere. The assignment rate was slightly better for the KEGG database than the SEED database but there was considerable variation within the rhizosphere samples in particular. Based on the InterProScan analysis (Figure 6.3b) assignment rate was higher

than rapsearch2 on reads before extensive quality filtering and increased on average over 2 fold after quality filtering by SeqPrep and Trimmomatic. This indicates that the rapsearch2 analysis would have benefited from further quality filtering prior to homology searching. This would have not resulted in more reads being assigned functions but would have improved the efficiency of the analysis. These data reinforce the importance of quality filtering of Illumina sequence reads (Bokulich *et al.*, 2013) and highlight that the majority of microbial functional diversity remains completely unknown, a limitation with all such high-throughput sequencing studies. Because of the differences in analyses, a number of other factors could have contributed to the differences in assignment rate. The merging of paired reads was performed by EBI but not by TGAC, this would have increased the length of some of the transcripts which would improve the assignment rate. Additionally the EBI analysis imposed a length cut-off of 100 bp, so even a single ambiguous nucleotide in a read would render it excluded from downstream analysis. This would have not allowed easy detection of many sRNAs. The reduced alphabet protein matrix used by rapsearch2 might have reduced its ability to assign some reads, though it performs analyses faster than InterProScan. Computational time is an important consideration for dealing with such large datasets. It would be interesting to input the EBI filtered reads into rapsearch2 to see to what extent the assignment rate improves. Importantly, the nature of the reads that did not pass EBI quality filters needs to be determined. Were they simply low quality, or do they actually match to anything? This would determine whether or not these filtering steps led to bias in assignment rates towards known proteins. Despite the different inputs and assignment rates for the different approaches, the actual number of reads able to be assigned was similar (≈0.75 million), indicating similar outcomes of the two approaches.

In pure cultures of well characterised organisms the assignment rates of sequenced transcriptomes to the genome are very high, for example over 97% in *E. coli* (Giannoukos *et al.*, 2012). In the marine environment, ≈10% of non-rRNA reads could be assigned to a protein in the SEED database from samples between 25 m and 125 m depth. Interestingly, the assignment rate dropped to 4.4% at 500 m (Shi *et al.*, 2010) indicating that less well studied environments, such as the deep sea, contain higher proportions of novel, and thus currently unassignable transcripts. Both programmes used here searched against protein databases, so any non-coding transcripts such as ncRNAs would likely be translated but not find a match. Furthermore, the more sequences obtained will inevitably reduce the proportion of those sequences that can be assigned. This is because as greater sequencing depth is achieved, the detection of rare and likely uncharacterised sequences will increase.

Genuinely novel sequences may represent transcripts from novel organisms or novel transcripts from known organisms. There are several ways in which these rates of assignment could be improved in future. Most fundamentally, novel microbes need to be isolated and studied in pure culture. Metatranscriptomics can actually aid in the culturing of such organisms by providing information on what conditions are like in the microbes natural habitat (Bomar *et al.*, 2011). Exploring the responses of well-characterised organisms to their natural environments could also help fill the gap in functional knowledge. An important limitation here is that despite the many genome sequences available in databases, essentially core genomes, a huge and unknown diversity of genetic information is available in the pan-genomes of different strains of the same species. This could contribute to the inability to assign protein coding reads reads at a high rate. Further genome sequencing of many environmental isolates should improve this in future. With such a large proportion, in fact the vast majority of reads unassigned here, the robustness of the metabolic mapping analysis is severely reduced (Huang *et al.* 2013).

Although reads were assigned to the KEGG, SEED (both via rapsearch2) and UniProt (via InterProScan) databases, some reads may have only been assigned by one of the programs used. Both tools could be used to slightly increase the number of assignable reads and provide validation if the same read was assigned the same function twice. The disadvantage of this is that analyses would take longer and inconsistencies in assignments might arise. Further improvements in read length and assembly tools will also improve the assignment rates for such studies in future.

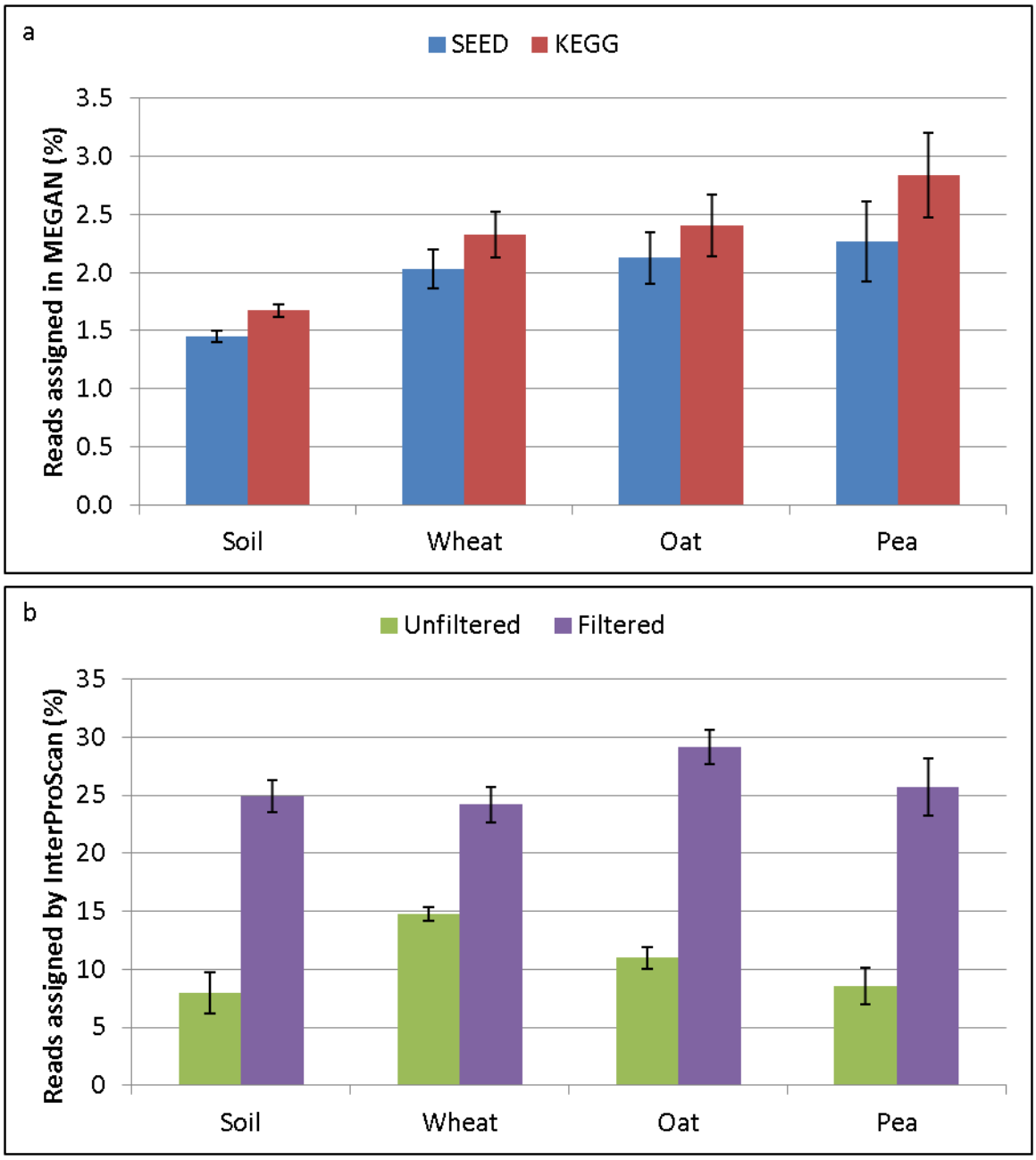


Figure 6.3: Proportions of input reads assigned by rapsearch2-MEGAN (a) and InterProScan (b). Values are means of biological replicates ±1SEM, where n = 5 for soil, oat and pea rhizospheres, and 4 for the wheat rhizosphere. In (a)the total reads were 25 million for each sample, while the unfiltered data in (b) was from a average 5.5 million read subset and the data in filtered by SeqPrep and Trimmomatic was from an average of 2.6 million reads due to discarding of low quality reads.

6.3.2 Analysis of metagenomic DNA and comparison with RNA using MG-RAST

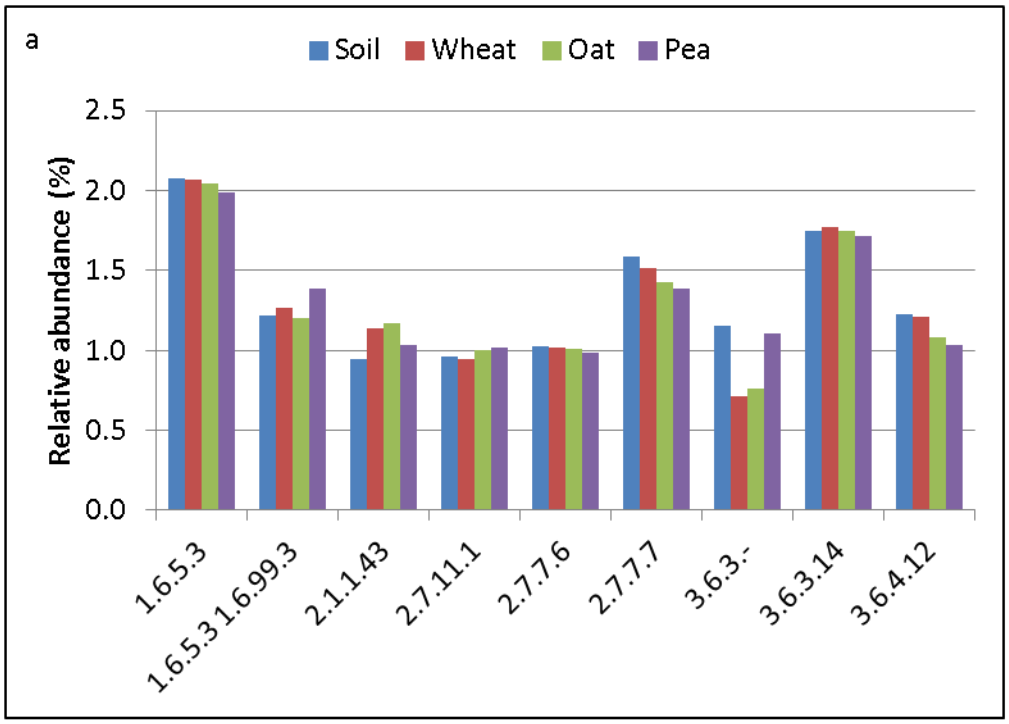
The metagenome samples generated on average over 79 million pairs of reads each (Table 6.1). Less than 0.1% of reads failed quality filters, and the frequency of rRNA genes was low (0.21%) as expected. Using all databases at MG-RAST's disposal, around 43% of sequences could be annotated as known protein, with an additional 50% as unknown protein, while the remaining 6.4% were classed as unknown. For the individual databases such as SEED and KEGG, this was lower, indicating that multiple databases can improve assignment rates. The 43% assigned to known proteins was higher than the best assignment rate of almost 30% for RNA (Figure 6.3b). A recent marine study showed that assignment rates for DNA were almost 3 fold higher than those for RNA (Shi *et al.*, 2010). This could be due to transcriptomes having higher numbers of poorly characterised or non-translated sequences such as small non-coding RNAs (ncRNAs) than metagenomes. The splicing of mRNA carried out by eukaryotes might also contribute to this, where certain regions of transcripts do not have corresponding genomic sequences. The results of such effects have been observed in human cell lines (Ameur *et al.*, 2011; Kapranov *et al.*, 2010) but would be both more extensive and complicated in soils where there are a huge diversity of mostly poorly characterised eukaryotes. The effect of this could be further exacerbated by the short reads of Illumina sequencing. Large populations of uncharacterised RNA viruses would also reduce the assignment rate of only RNA.

Analysis of metagenomic DNA samples with MG-RAST detected 960 classes of KEGG orthologs (KO) for soil, 953 for the wheat rhizosphere, 951 for the oat rhizosphere and 985 for the pea rhizosphere. There were 803 KO classes that were detected in all four samples and the nine most highly abundant KO classes (at least 1% in at least one sample, Figure 6.4a) represented around 12% of the total detected in all four environments. These included housekeeping genes involved in transcription (DNA and RNA polymerases), redox homeostasis (NADH:ubiquinone reductase and dehydrogenase) and energy metabolism (phosphate anhydride hydrolase and ATP synthase) (Figure 6.4b). The remaining 88% was made up of 794 KO classes.

Table 6.1: Summary of metagenomic DNA as analysed by MG-RAST based on a subset of 4 million reads. Annotated and Unknown protein assignments are based on all databases utilised by MG-RAST while the KEGG and SEED assignments are specific to those databases.

Information	Soil	Wheat	Oat	Pea
Total sequences (pairs of reads)	64156233	73210555	87541019	92177163
Failed QC (%)	0.09	0.11	0.08	0.11
rRNA gene (%)	0.34	0.27	0.12	0.10
Annotated protein (%)	40.46	42.55	41.57	47.17
Unknown protein (%)	53.08	50.12	51.89	46.30
Unknown (%)	6.03	6.96	6.34	6.32
KEGG hits (%)	34.77	36.76	36.33	41.44
SEED hits (%)	31.14	32.72	32.38	36.94

a



b

EC number	Enzyme name
1.6.5.3	NADH:ubiquinone reductase
1.6.99.3	NADH dehydrogenase
2.1.1.43	Histone-lysine N-methyltransferase
2.7.11.1	Non-specific serine/threonine protein kinase
2.7.7.6	DNA directed RNA polymerase
2.7.7.7	DNA directed DNA polymerase
3.6.3.-	Hydrolase acting on phosphorus-containing anhydrides
3.6.3.14	ATP synthase
3.6.4.12	DNA helicase

Figure 6.4: Relative abundance of assigned reads (a) and function (b) of the most highly abundant KO classes (at least 1% in at least one sample) from MG-RAST analysis of metagenomic DNA.

There were some KO classes that were differentially abundant in the rhizospheres compared to soil (Table 6.2). Chondroitin 4-sulfotransferase and carbon monoxide dehydrogenase were at least 2fold more abundant in all rhizosphere than in soil. Interestingly, chondodroitin 4-sulfotranferease is involved in formation of sulphated proteoglycans in animal connective tissue. The increased abundance of its coding gene might represent the increased populations of small metazoan such as nematodes in the rhizospheres. The increased abundance of carbon monoxide dehydrogenase genes suggest selection of taxa capable of utilising one carbon compounds (Bartholomew and Alexander, 1979). Plant derived methanol, another one carbon compound, is well established as contributing to microbial colonisation of plants (Delmotte *et al.*, 2009; Galbally and Kirstine, 2002; Knief *et al.*, 2012). Carbon monoxide has been shown to be produced in plant photosynthetic tissues (Wilks, 1959), but it's influence on the microbiome has not been studied. Tagatose-bisphosphate aldolase, which is involved in galactose metabolism, and MAPKKK, involved in eukaryotic cell signalling cascades in response to stresses, were more abundance in wheat and pea rhizospheres. There was an increased in abundance of CDP-paratose synthase, a component of some bacterial antigenic polysaccaraides, in the oat rhizosphere. Pea increased the abundance of L-arabinonate dehydratase, suggesting arabinoate as an important carbon source for microbes colonising the pea rhizosphere. The KO classes less abundant in the rhizospheres compared to soil included a monosaccharide-transporting ATPase and aminodeoxychorismate synthase, suggesting that utilisation of monosaccahrides and biosynthesis of aromatic amino acids are not as important for life in rhizosphere as they are in bulk soil. Presumably there are preferred carbon sources, such as organic acids, and aromatic amino acids provided by the plant so *de novo* synthesis is unnecessary. Because these values are based on signal biological samples they need to be interpreted with caution.

Table 6.2: Differentially (≥2 fold difference) abundant KO classes between soil and the rhizospheres of wheat, oat and pea. Data are based on MG-RAST analysis of metagenomic DNA of single biological samples.

Comparison	EC number	Fold difference	Function
Wheat vs soil	4.1.2.40	19.24	Tagatose-bisphosphate aldolase
	2.8.2.5	8.04	Chondroitin 4-sulfotransferase
	2.7.11.25	4.88	MAPKKK
	2.1.1.86	3.82	Tetrahydromethanopterin S-methyltransferase
	1.2.99.2	3.66	Carbon-monoxide dehydrogenase
	5.5.1.9	0.45	Cycloeucalenol cycloisomerase
	2.7.-.-	0.39	Transferring phosphorus-containing groups
	4.2.1.30	0.29	Glycerol dehydratase
	3.6.3.17	0.29	Monosaccharide-transporting ATPase
	2.6.1.85	0.26	Aminodeoxychorismate synthase
	4.2.1.59	0.18	3-hydroxyacyl-[acyl-carrier-protein] dehydratase
	4.1.99.-	0.18	Other Carbon-Carbon Lyases
Oat vs soil	2.8.2.5	7.96	Chondroitin 4-sulfotransferase
	2.4.1.175 2.4.1.226	5.42	N-acetylgalactosaminyl-proteoglycan 3-beta-glucuronosyltransferase
	1.2.99.2	3.33	Carbon-monoxide dehydrogenase
	1.1.1.342	2.07	CDP-paratose synthase;
	2.4.1.147	0.47	Acetylgalactosaminyl-*O*-glycosyl-glycoprotein β-1,3-*N*-acetylglucosaminyltransferase
	4.2.1.30	0.43	Glycerol dehydratase
	2.8.3.12	0.40	Glutaconate CoA-transferase
	2.7.-.-	0.33	Transferring phosphorus-containing groups
	2.8.2.4	0.30	Estrone sulfotransferase
	1.3.7.8	0.29	Benzoyl-CoA reductase
	2.6.1.85	0.28	Aminodeoxychorismate synthase
	3.6.3.17	0.23	Monosaccharide-transporting ATPase
	5.4.99.5	0.17	Chorismate mutase
	2.7.11.22	0.15	Cyclin-dependent kinase

	4.1.2.40	22.84	Tagatose-bisphosphate aldolase
	2.8.2.5	10.93	Chondroitin 4-sulfotransferase
	2.5.1.-	7.04	2,5-dichlorohydroquinone reductive dechlorinase
	2.7.11.25	6.22	MAPKKK
	1.2.99.2	4.30	Carbon-monoxide dehydrogenase
	1.14.-.-	3.28	Unspecific monooxygenase
	2.3.1.15	3.17	Glycerol-3-phosphate 1-O-acyltransferase
Pea vs soil	4.2.1.25	2.00	L-arabinonate dehydratase
	2.7.-.-	0.40	Transferring phosphorus-containing groups
	2.1.1.86	0.34	Tetrahydromethanopterin S-methyltransferase
	3.6.3.17	0.30	Monosaccharide-transporting ATPase
	2.6.1.85	0.30	Aminodeoxychorismate synthase
	5.4.99.5	0.20	Chorismate mutase
	2.7.11.22	0.2	Cyclin-dependent kinase

Multidimensional scaling was used to visualise the functional potential of soil and the rhizospheres of wheat, oat and pea based on the KEGG assignments of DNA and RNA as analysed by MG-RAST (Figure 6.5). All the DNA samples clustered closely together, suggesting that the effect of plants on the functional potential of the rhizosphere microbiome was somewhat limited. All the RNA samples were distinct from their respective DNA samples indicating differences in active and potential functions of microbiomes. The three rhizosphere RNA samples clustered closely together, indicating similarity of the active functions of their microbiomes. The soil RNA sample was distinct from the rhizosphere RNA samples, indicating that the active microbial processes occurring in an unplanted soil are completely different to those influenced by a plant. While these observations are interesting they are based on single DNA samples from pooled biological replicates, so they cannot be validated statistically.

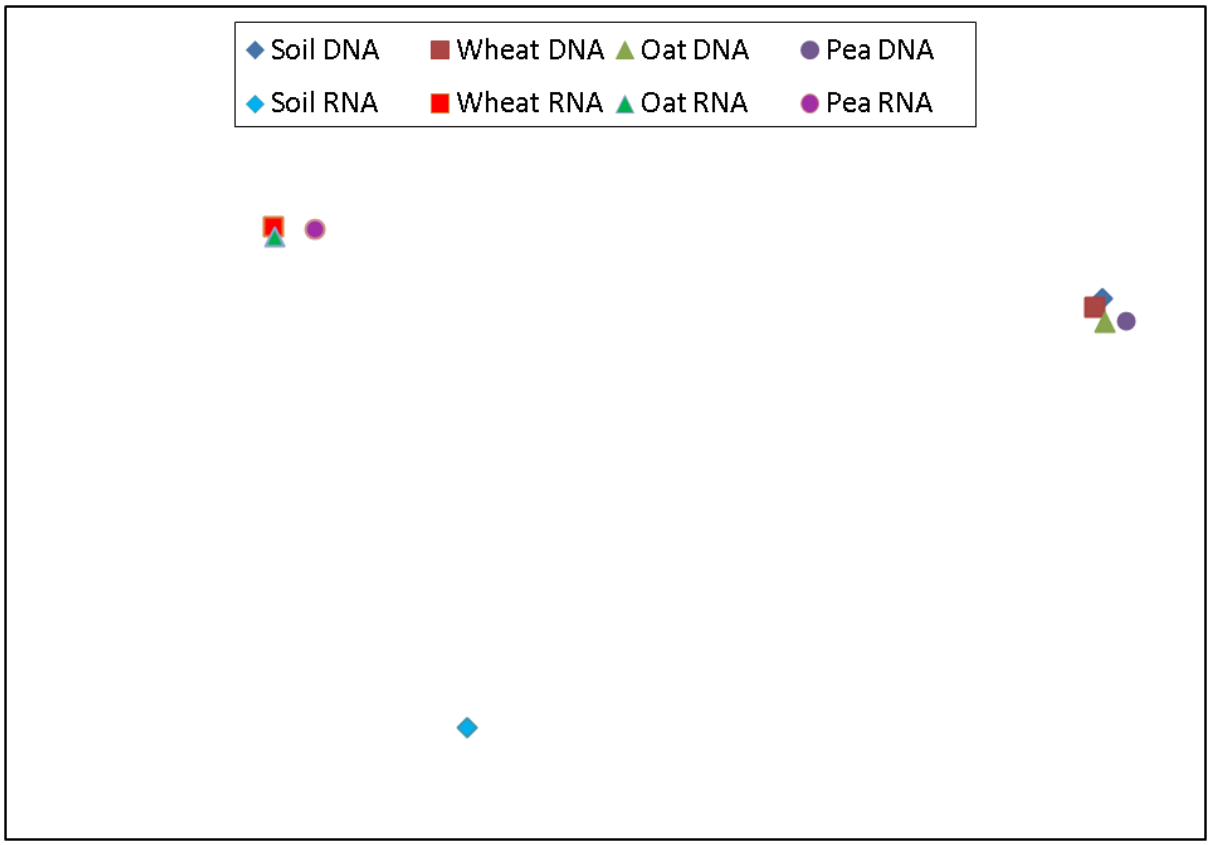
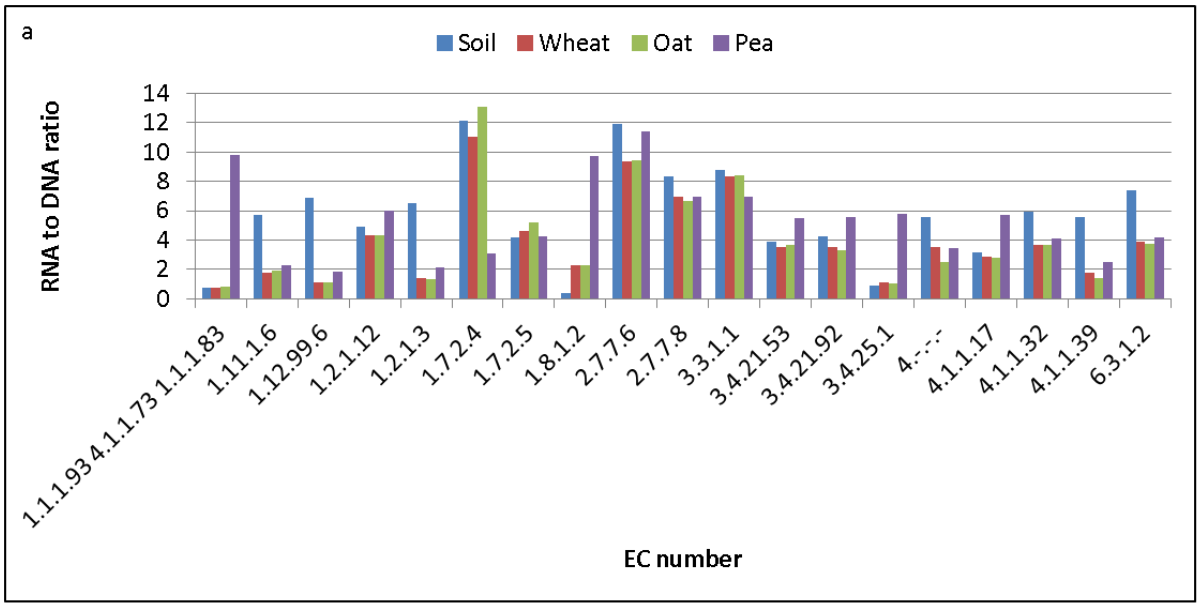


Figure 6.5: Multidimensional scaling representation of functional potential in soil and the rhizospheres of wheat, oat and pea. Data were generated in PRIMER6 and plotted in Excel. Data are based on totals of biological replicates for RNA, where n=5 for soil, oat and pea, and n=4 for wheat, and single samples for DNA, so no error bars are shown. The plots are based on relative abundance of KO classes analysed in MG-RAST. The x,y coordinates were generated and exported by PRIMER6 and scales, although arbitrary, were standardised across plots. The closer two points are together, the more similar their communities are.

In the metatranscriptomics dataset, the abundance of transcripts could have been influenced by their gene copy number. For example, housekeeping genes are typically highly expressed but they are also possessed by the majority of microbes, so are highly abundant, whereas low abundance genes could be highly expressed under certain conditions. By normalising transcripts to genes, the accuracy of metabolic profiles can be imprrved. Here the abundance of KO classes in the RNA samples was divided by the abundance in the respective DNA sample. There were some KO classes present either only in the DNA or only in the RNA datasets, which therefore could not be analysed here.

Of the 172 genes detected in all environments, 56, 34, 34 and 67 were more than 2fold more abundant in the RNA pools of soil, wheat oat and pea rhizospheres respectively, than in their DNA pools. There were 58, 37, 32 and 23 genes more than 2fold more abundant in the DNA pools of soil, wheat, oat and pea rhizospheres respectively, than in their RNA pools. Some of the genes with the highest expression ratios ratio (RNA:DNA, at least 5 fold in at least one environment) are shown in Figure 6.6. DNA directed RNA polymerase and polyribonucleotide nucleotidyltranferase, key enzymes involved in transcription, had high ratios in all samples, as did adenosylhomocystinase, which is involved in biosynthesis of adenosine. This indicates that as well as being widespread, these housekeeping functions are highly active in soil and the rhizosphere. The expression ratio of tartrate metabolising enzymes was particularly high in the pea rhizosphere. Interestingly, the uptake of tatrate was up-regulated in *R. leguminosarum* during colonisation of the pea rhizosphere, but the metabolism of tatrate was only up-regulated in the alfalfa rhizosphere (Ramachandran *et al.*, 2011). The expression ratio of catalase was much higher in soil than in the rhizospheres. Catalase is important for dealing with oxidative stress associated with aerobic metabolism and for coping with the reactive oxygen species (ROS) burst of the plant immune system. Therefore it might be expected that the expression ratio of catalase would be higher in the rhizospheres than in soil. However, the observation here might indicate that organisms possessing perhaps multiple genes encoding catalases would be more likely to survive in the rhizosphere, this would increase the abundance of catalase gene in the DNA pool, reducing the expression ratio. The expression ratios of a number of key nutrient cycling enzymes were increased. These included sulphite reductase, several peptidases, nitrous oxide reductase and nitric oxide reductase, indicating these as important functions in soil and rhizosphere environments. The latter two are of particular interest given their role in denitrification which is thought to be enhanced in the presence of sugars, organic and amino acids that are exuded by plants into the rhizosphere (Henry *et al.*, 2008; Philippot *et al.*, 2009).

a



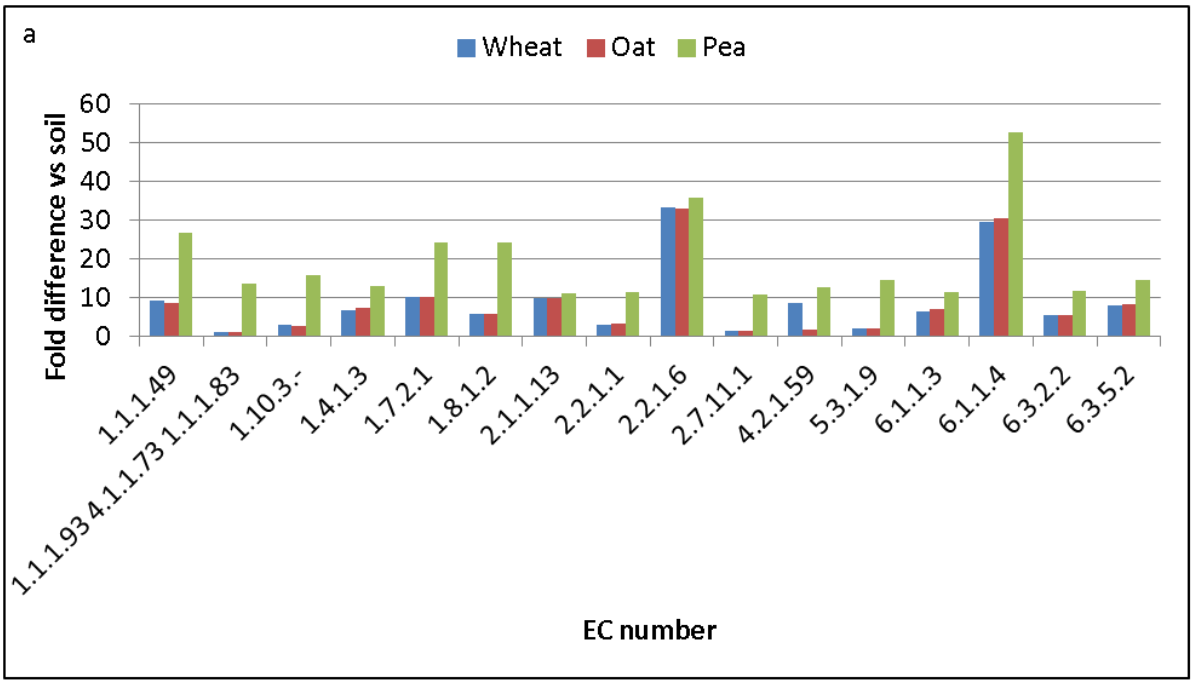
b

EC	Function
1.1.1.93 4.1.1.73 1.1.1.83	Tartrate dehydrogenase / decarboxylase / D-malic enzyme
1.11.1.6	Catalase
1.12.99.6	Hydrogenase
1.2.1.12	Glyceraldehyde-3-phosphate dehydrogenase
1.2.1.3	Aldehyde dehydrogenase
1.7.2.4	Nitrous oxide reducatase
1.7.2.5	Nitric oxide reductase
1.8.1.2	Sulfite reductase
2.7.7.6	DNA directed RNA polymerase
2.7.7.8	Polyribonucleotide nucleotidyltransferase
3.3.1.1	Adenosylhomocysteinase
3.4.21.53	Endopeptidase La
3.4.21.92	Endopeptidase Clp
3.4.25.1	Proteasome endopeptidase complex
4.-.-.-	Lyases
4.1.1.17	Orthinine decarboxylase
4.1.1.32	Phosphoenolpyruvate carboxykinase
4.1.1.39	Ribulose-bisphosphate carboxylase
6.3.1.2	Glutamine synthetase

Figure 6.6: (a) Expression ratios, i.e. the ratio of relative abundance in RNA to relative abundance in DNA) (≥5 fold) and (b) functions of KO classes determined by MG-RAST analysis of DNA and RNA against the KEGG database.

It is likely that there would have been differences in expression ratios between different environments for some genes. An enzyme encoded by a gene that has similar abundance across all samples might be highly expressed in only one environment, or a gene might be expressed differently in each environment but proportionally to its abundance. By comparing expression ratios of each of the three rhizospheres with that of soil some insight into this dynamic was provided (Figure 6.7). The expression ratio of acetolactate synthase was over 30 fold higher in all rhizospheres compared to soil. This enzymes catalyses the first step in the biosynthesis of branched chain amino acids valine, leucine and isoleucine. The expression ratio of leucine-tRNA ligase was also much higher in all rhizosphere, particularly the pea rhizosphere than soil, suggesting a particularly requirement for this amino acid during rhizosphere colonisation. While leucine is incorporated into many proteins, those containing leucine-rich repeats (LRRs) have a higher requirement. Proteins with this motif are often involved in protein-protein interactions (Kobe and Kajava, 2001) which might be important during colonisation of plant roots. Differences in expression ratios compared to soil were similar for wheat and oat rhizospheres, with the exception of 3-hydroxyacyl-[acyl-carrier-protein] dehydratase, involved in fatty acid biosynthesis, which was higher in the wheat rhizosphere than in that of oat. The pea rhizosphere had a number of expression ratios different to those found in soil. These included nitrite and sulphite reductases, glucose 6 phosphate isomerase and dehydrogenase, and glutamate-cysteine, which is involved in glutathione biosynthesis. These observations provide some insight into the different results produced between using DNA, RNA or both to determine changes in functional potential and activity in the rhizosphere. However, the DNA and RNA samples were not isolated from the same biological samples, and replicates were not available for the DNA samples. Therefore future steps should involve fully integrated metatranscriptomic and metagenomic analysis of microbial community functions.

a



b

EC number	Enzyme name
1.1.1.49	Glucose-6-phosphate dehydrogenase
1.1.1.93 4.1.1.73 1.1.1.83	Tartrate dehydrogenase / decarboxylase / D-malic enzyme
1.10.3.-	Oxidoreductase
1.4.1.3	Glutamate dehydrogenase
1.7.2.1	Nitrite reductase
1.8.1.2	Sulfite reductase
2.1.1.13	Methionine synthase
2.2.1.1	Transketolase
2.2.1.6	Acetolactate synthase
2.7.11.1	Non-specific serine/threonine protein kinase
4.2.1.59	3-hydroxyacyl-[acyl-carrier-protein] dehydratase.
5.3.1.9	Glucose-6-phosphate isomerase
6.1.1.3	Threonine-tRNA ligase
6.1.1.4	Leucine-tRNA ligase
6.3.2.2	Glutamate-cysteine ligase
6.3.5.2	GMP synthase (glutamine-hydrolysing)

Figure 6.7: (a) Fold difference in expression ratio and (b) function of the most highly different expression ratios (≥10 fold different in at least one rhizosphere sample vs soil) of KO classes based on MG-RAST analysis of DNA and RNA against the KEGG database.

6.3.3 Summary of rapseach2 and InterProScan analyses of RNA

The analysis by InterProScan provided abundance counts of Gene Ontology (GO) categories and InterPro assignment, while the rapsearch2 analysis provided abundance counts of both SEED and KEGG categories. One set of assignments from each analysis, GO and SEED respectively, were chosen for more detailed examination over the alternatives because they provided more easily interpretable data. A total of 1222 GO categories were significantly ($P \leq 0.05$) up-regulated in the rhizospheres (Figure 6.8a) compared to soil but only 7.7% were shared between all rhizospheres. Around 15%, 20% and 18% were specifically selected by wheat, oat and pea respectively. Wheat and oat together selected 13% and oat and pea together selected 37%, but wheat and pea together only shared 2.3%. Of the 1076 SEED categories up-regulated in the rhizospheres (Figure 6.8a) only 8.8% were shared between all rhizospheres, some of which are shown in Figure 6.8b. The wheat specifically induced 17.1% functions while this rose to 20.3 % and 23.2% for pea and oat respectively. There was considerable overlap (17.5%) in induced functions between the cereals, and also between oat and pea (10.2%), but much less between wheat and pea (2.9%).

The two analysis methods generally agreed that the changes in activity of different functions in the rhizospheres were plant species specific, though this specificity was observed to be stronger from the rapsearch2 analysis. A core set of functions was induced by all plants but this was a low proportion of the total. There was considerable overlap in induced functions between the cereals (wheat and oat) and also by oat and pea, with wheat having the lowest number of specific up-regulated functions from both analyses. Data tables for each pair-wise comparison to soil for the hits to the SEED database can be found in Appendix Tables A25 through to A31). To avoid repetition in describing transcriptional changes, only the assignments to the SEED database were chosen to be discussed further. Interpreting the information from them was more intuitive than from the GO categories, was consistent with the analysis of taxonomic composition in Chapter 4 (4.3.6 through to 4.3.13).

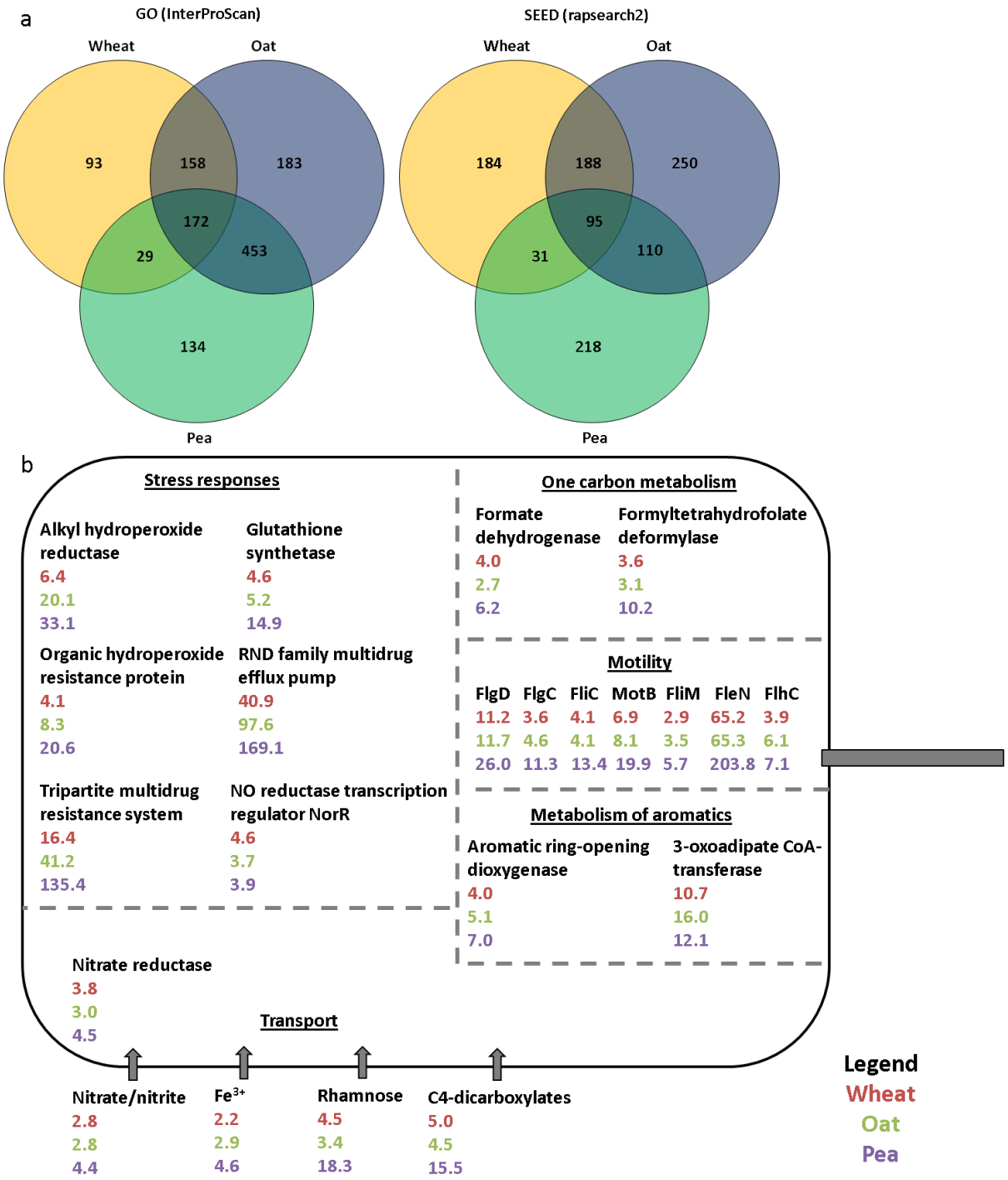


Figure 6.8: (a) Significantly (P≤0.05) up-regulated functional categories in the wheat, oat and pea rhizospheres compared to soil based on transcripts per gram of soil as determined by InterProScan (GO) and rapsearch2 (SEED). (b) Some general transcriptional responses to the rhizosphere (wheat, oat and pea) based on the hits to the SEED database as determined by rapsearch2. Values are fold differences in abundance of transcripts per gram relative to unplanted soil.

Multidimensional scaling was used to visualise the overall expressed functions in soil and the rhizospheres of wheat, oat and pea based on abundance of SEED categories determined by rapsearch2. Plots were generated with and without a standardisation step to generate relative and quantitative plots respectively (Figure 6.9). Both plots generated highly comparable results, with distinct expression profiles for each environment. Pea and soil were the most dissimilar, while wheat and oat were more similar to each other than to either soil or pea.

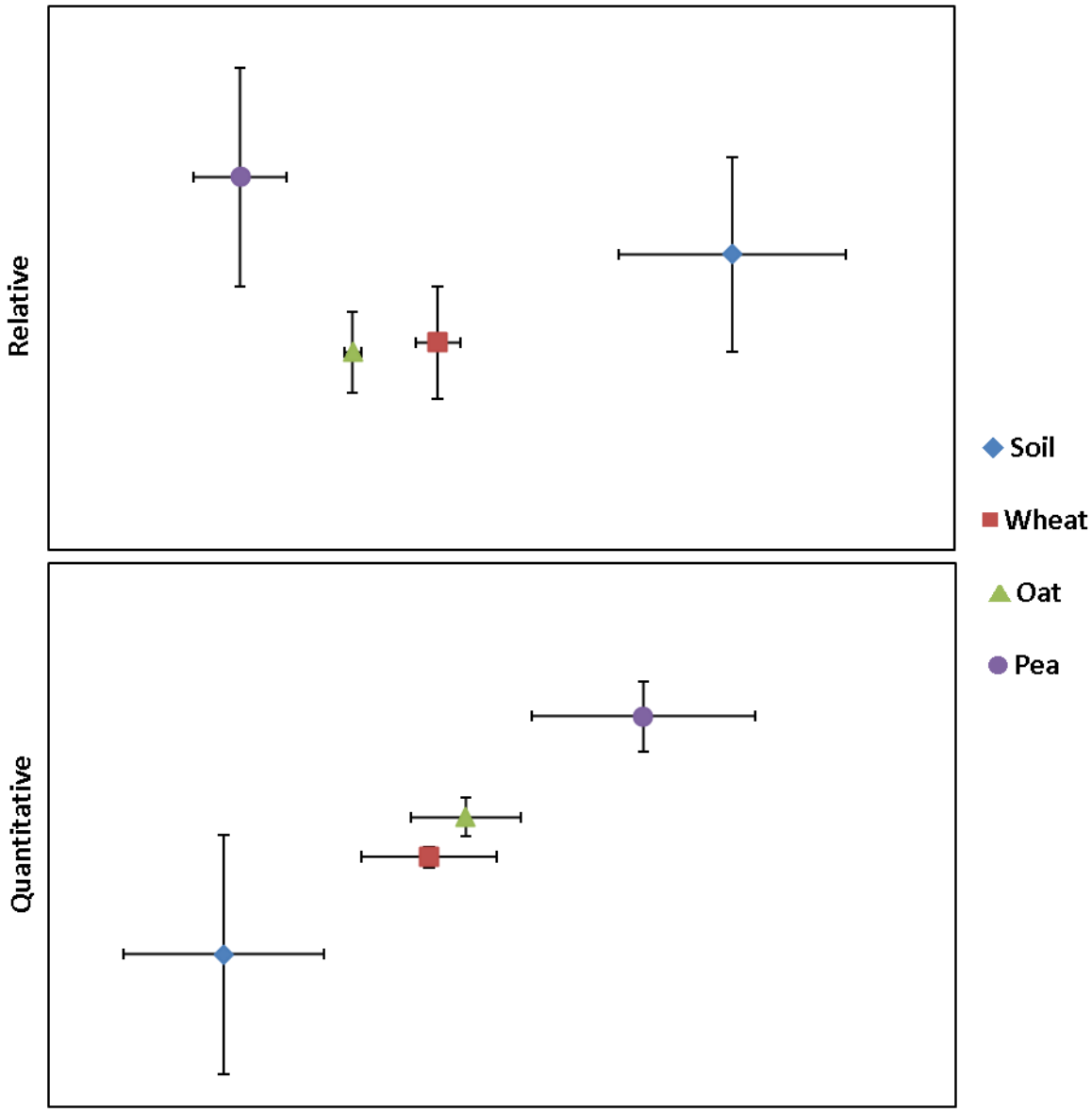


Figure 6.9: Multidimensional scaling representation of expressed functions structure in soil and the rhizospheres of wheat, oat and pea. Data were generated in PRIMER6 and plotted in Excel. Data are means (±1SEM) of biological replicates, where n=5 for soil, oat and pea.

The plots are based on both quantitative and relative abundance of all SEED categories analysed by rapsearch in MEGAN. The x,y coordinates were generated and exported by PRIMER6 and scales, although arbitrary, were standardised across plots. The closer two points are together, the more similar their communities are.

6.3.4 General transcriptional responses to the rhizosphere as determined by rapsearch2

Based on pair-wise comparisons of transcripts assigned to different SEED categories per gram of soil, it was possible to determine general microbial responses to the rhizospheres of wheat, oat and pea. Functions induced by all three plants were considered general rhizosphere responses, some of which are shown in Figure 6.8b with a full list in Appendix Table A25. All plants induced the uptake of C4-dicarboxylates, including fumarate, L-malate and succinate in their rhizospheres. Dicarboxylates are commonly found in root exudates of plants (Vranova *et al.*, 2013) and their uptake was induced in *R. leguminosarum* during colonisation of several plant rhizospheres (Ramachandran *et al.*, 2011). Uptake of nucleosides and rhamnose, a common sugar in plant glycosides (Wu and Prior, 2005) was also induced. Additionally, metabolism of rhamnose has been shown to be important for competitive nodulation in *Rhizobium spp*. (Oresnik *et al.*, 1998). Both ferric and ferrous iron, nitrate, nitrite, phosphonates and potassium were also taken up. Iron is usually a limiting nutrient, and improved microbial acquisition of iron enhances rhizosphere colonisation (Loper and Henkels, 1999). Nitrate and nitrite are soluble sources of fixed nitrogen and can act as alternative electron acceptors to oxygen. Some bacteria, such as the Methylophilales, are able to couple denitrification of nitrate to oxidation of methanol (Kalyuhznaya *et al.*, 2009), a one-carbon compound produced by degradation of plant pectin (Galbally and Kirstine, 2002). Phosphonates are organic sources of phosphorous that can be derived from cell membranes, while regulation of potassium uptake is critical for osmotic homeostasis which may be disrupted during colonisation of the rhizosphere due to high concentrations of water soluble compounds exuded by the plant.

A number of functions were strongly induced (≥10 fold) in all rhizospheres. These included housekeeping functions such as NADH dehydrogenase activity, responsible for maintaining redox homeostasis, as well as a number of growth related functions such as UDP-N-acetylglucosamine-N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase , involved in bacterial cell wall peptidoglycan biosynthesis

and ribosomal large subunit pseudouridine synthase C, involved in initiation of translation. Stress response related functions were also strongly induced, including a copper chaperone, two multidrug efflux systems (RND and tripartite), a hypothetical protein with a DnaJ-like domain and Paraquat-inducible protein B. DnaJ-domains are found in heat shock proteins such as Hsp40 and are known to interact with Hsp70 (Qiu *et al.*, 2006). Paraquat, a potent herbicide, is a superoxide generating agent and as such induces oxidative stress in organisms exposed to it (Koh and Roe, 1995). Paraquat was not applied to the plants during growth, and this microbial response is more likely a mechanism for dealing with the plant immune system's ROS burst. Other strongly induced functions included HPr activity, a key regulator of central carbon metabolism in Gram positive bacteria (Hanson *et al.*, 2002), 3-oxoadipate CoA-transferase, which is involved in benzoate metabolism, a Pirin-related protein co-expressed with the regulon of the iron chelator pyoverdine, and two genes involved in motility, flagellar synthesis regulator FleN and flagellar basal-body rod modification protein FlgD. Benozates are widespread in plants and are precursors of important regulatory molecules and hormones, including salicylic acid, the plant wounding response hormone (Wildermuth, 2006). Motility is required for colonisation of the rhizosphere but must be tightly controlled to allow relocation or attachment to plant surfaces if desired.

Less strongly (>1, <10 fold) induced functions included resistance to stresses via glutaredoxin and glutathione, alkyl hydroperoxide reductase, organic hydroperoxide resistance protein, metal dependent beta-lactamases, DNA mismatch repair protein MutL and cyanophycin synthase. Cyanophycin is a carbon and nitrogen storage molecule produced by cyanobacteria and some others under sulphur or phosphorus limitation (Krehenbrink *et al.*, 2002; Ziegler *et al.*, 1998). Motilty related functions induced included flagellar motor rotation protein MotB, flagellar biosynthesis protein FliC transcriptional activator FlhC, basal-body rod protein FlgC and motor switch protein FliM. Metabolism of aromatics and one carbon compounds was observed by an aromatic ring-opening dioxygenase, and a tungsten containing formate dehydrogenase and formyltetrahydrofolate deformylase respectively. Aromatic compounds and methanol, a one-carbon compound, are abundantly produced by plants (Galbally and Kirstine, 2002; Stafford, 1974). Nitrogen metabolism related functions included the anaerobic nitric oxide reductase transcription regulator NorR and a respiratory nitrate reductase, suggesting enhanced nitrogen cycling in the rhizosphere. Enhanced denitrification has been observed in rhizospehres compared to bulk soil REFS

6.3.5 Transcriptional responses to the wheat and oat rhizospheres as determined by rapsearch2

Wheat and oat rhizospheres induced a number of functions (Appendix Table A26), including transport of sugars such as the cell wall precursors D-xylose and L-arabinose as well as maltose, a breakdown product of starch. Transport of glycine betaine, which can be produced by plants in response to salt stress (Storey and Jones, 1975), was also induced. Glycine betaine has been detected in wheat root exudates (Fan *et al.*, 2001). Phospholipids and lipopolysaccharides were also taken up, suggesting scavenging of cell membrane components from lysed bacteria.

Strongly (≥10 fold) induced functions related to stress responses included glutathione S-transferase, the programmed cell death toxin YdcE and a hypothetical protein in a Rubrerythrin cluster. Programmed cell death could have been initiated by some bacteria unable to compete effectively in the rhizosphere, allowing other microbes to utilise their nutrients. Rubreythrins are proteins that protect anerobic microbes from oxidative stress (Sztukowska *et al.*, 2002). Metabolism of aromatics by trans-cinnamate 4-monooxygenase, and other carbon sources by oxaloacetate decarboxylase, galactokinase, L-arbinose isomerise, sucrose-6-phosphate hydrolase and D-theronine aldolase was observed. Other induced functions included ATPase activity for assembly of type IV secretion complex, the cell division initiation protein DivIVA and a 3',5'-cyclic-nucleotide phosphodiesterase. Type IV secretion systems are involved in conjugation and are also used by plant pathogens such as *Agrobacterium* spp. (Christie *et al.*, 2005), while cyclic nucleotides are important signalling molecules in bacteria(Camilli and Bassler, 2006).

Less strongly (>1, <10 fold) induced functions included a number of motility related proteins such as, the RNA polymerase sigma factor for the flagellar operon (RpoF), flagellar L-ring protein FlgH, hook-associated protein FlgK, P-ring protein FlgI,switch protein FliN and others (FliP, FliF, FlaA, FlaB,FlhA), as well as the chemotaxis regulators CheB and CheW, and phosphatase CheZ, a serine chemoreceptor and chemotaxis regulator signal transducer CheY. Metabolism of a number of aromatics, including protocatechuate, aromatic amino acids, 4-hydroxybenzoate, 1,4-dihydroxy-2-naphthoate and biphenyl-2,3-diol, as well as 4-hydroxyproline and glycolate was observed. Nitrogen metabolism functions included the NnrS protein involved in response to NO, nitric oxide reductase, and an assimilatory nitrate reductase. A number of sulphur metabolism functions were induced including sulfur oxidation proteins SoxX and SoxA, sulphate reduction associated complexes, sulfide

dehydrogenase, sulfite reduction-associated complex and an alkanesulfonate utilization operon regulator, suggesting enhanced sulphur cycling in cereal rhizospheres. A family of universal stress proteins were induced as was a thiol peroxidase, and one carbon metabolism related functions such as an NAD-dependent formate dehydrogenase and enolase, which is involved in the serine-glyoxalate cycle. Many functions involved in secretion and attachment were induced including Type II/IV secretion system secretin, pilus assembly protein, pilin Flp, type IV pilin proteins PilABCOM, and a type I secretion outer membrane protein. In addition, biosynthesis of phenazine, a precursor of many antifungal metabolites (McDonald *et al.*, 2001; Turner and Messenger, 1986), was induced.

6.3.6 Transcriptional responses to the wheat and pea rhizospheres as determined by rapsearch2

No transporters were induced in the wheat and pea rhizospheres that were not also induced in the oat rhizosphere. All induced functions can be found in Appendix Table A27.Functions strongly (≥10 fold) induced in the wheat and pea rhizospheres included the stage IV sporulation protein A, a transcriptional regulator (CytR) that represses biofilm formation in *Vibrio cholerae* (Haugo and Watnick, 2002), 2-keto-4-pentenoate hydratase which is involved in metabolism of aromatic compounds, L-allo-threonine aldolase, and a Rossmann fold nucleotide-binding protein Smf possibly involved in DNA uptake. This suggests that naturally competent bacteria are active in the rhizosphere and may be acquiring DNA from other rhizosphere colonisers in an attempt to adapt the environment. Less strongly (>1, <10 fold) induced functions included metabolism of alginate, aromatic amino acids, threonine, isoleucine and valine. Additionally, glycogen synthase, a carbon storage molecule, and the chemotaxis protein methyltransferase CheR were also induced.

6.3.7 Transcriptional responses to the oat and pea rhizospheres as determined by rapsearch2

Oat and pea induced a number of functions (Appendix Table A28), including transport of D-galactonate, an organic acid derivative of galactose, and fructose, commonly found in plant root exudates (Vranova *et al.*, 2013), was observed. Thiamin, a cofactor for enzymes involved in sugar and amino acid metabolism, and glutamine, a common amino acid in root exudates (Vranova *et al.*, 2013) were taken up. Transport of inorganic phosphate, a key

limiting nutrient, and cobalt, a metal cofactor for enzymes including thiocyanate hydrolase and methionine synthase were also induced.

Functions strongly (≥10 fold) induced in the oat and pea rhizopsheres included a eukaryotic translation initiation factor, the nitrite-sensitive transcriptional repressor NsrR, the transcriptional regulator of proline metabolism PutR, RNA polymerase alternative sigma factor H, and AlgT, which is required for alginate production (Hershberger *et al.*, 1995), flagellar repression (Tart *et al.*, 2005) and *in planta* fitness of *Pseudomonas* spp. (Schenk *et al.*, 2008). Metabolism of carbon compounds by 1-phosphofructokinase, an NADH-dependent butanol dehydrogenase and glycine oxidase was induced. Potential stress response related functions included induction of the RND family multidrug efflux membrane fusion protein MexC, Cu(I)-responsive transcriptional regulator and decaprenyl diphosphate synthase, important for cell wall biosynthesis in mycobacteria (Kaur *et al.*, 2004).

Less strongly (>1, <10 fold) induced functions included metabolism of threonine, benzoylformate, and also arylsulphates by 4-hydroxybenzoyl-CoA thioesterase and 2-hydroxychromene-2-carboxylate isomerase both are known to exist in pseudomonads (Benning *et al.*, 1998; Kuhm *et al.*, 1993) and hint at the structure of plant derived arylsulphates, which remain largely unidentified. Interestingly, expression of some arylsulphatases was increased in *R. leguminosarum* during colonisation of the rhizospheres of sugar beet, alfalfa and sugarbeet (Ramachandran *et al.*, 2011). Alternative arylsulphatases were up-regulated during symbiosis with pea (Karunakaran *et al.*, 2009), suggesting the plant arylsulphates are different depending on whether they are intracellular or exported, or that they change during nodule development. Other sulphur metabolic functions induced included sulfite reductase and sulfate adenylyltransferase. Oligopeptidase was induced, suggesting break down of proteins by bacteria, as well as beta-glucanase, which break down polymers such as cellulose, 2-hydroxy-3-oxopropionate reductase involved in dicarboxylate and glyoxalate metabolism. A number of stress response related functions were induced, including heat shock sigma factor RpoH, heat shock protein GrpE and protease HslV, cold-shock DEAD-box protein A, thioredoxin and thioredoxin reductase, glutathione S-transferases, alkyl and organic hyroperoxide reductases, thiol peroxidase, arsenate reductase, metallo-beta-lactamase, and RND and tripartite drug efflux systems. Additional induced functions included ferrochelatase, a fumarate and nitrate reduction regulatory protein, a twin-arginine translocation protein

TatA and the one carbon metabolism related functions carbon monoxide dehydrogenase and formate dehydrogenase.

6.3.8 Transcriptional responses specific to the wheat rhizosphere as determined by rapsearch2

Wheat specifically induced a number of functions (Appendix Table A29), including transporters for sugar and sugar derivatives fucose, beta-xyloside, and hexuronates. Fucose represented 3% of recovered carbohydrate from wheat root exudates (Moody *et al.*, 1988). Xylosides are glycoside derivatives of xylose, an abundant sugar in plants and a precursor to the cell wall component hemicelluloses. Hexuronates are organic acid oxidation products of sugars, and so are likely to occur in the rhizosphere where both sugars and reactive oxygen species are present. They could also be produced by other plants or microbe mediated oxidation. Induction of sulfonate and taurine transport was observed. Both are organic sulphur sources, and while taurine is primarily found in animal tissues, is has been detected in plants (Lahdesmaki, 1986). Transport of a number of organic nitrogen sources was observed, including choline, urea, cyanate, spermidine and putrescine. Choline is produced by plants in response to salt stress (Storey and Jones, 1975) and it is possible that bacteria were using it to protect themselves from osmotic stress in the rhizosphere or also as a carbon or nitrogen source. Urea can be produced by metabolism of arginine in plants (Witte, 2011), or it may be derived from nitrogenous waste from small animals such as nematodes. Cyanate could have resulted from oxidation of hydrogen cyanide that is known to be produced by some biocontrol strains of *Pseudomonas* (Haas and Keel, 2003). Spermidine and putrescine are polyamines known to regulate plant growth and inhibit nitric oxide synthase, and protect plants from oxidative stress respectively (Flores and Galston, 1982). In bacteria, they are important for growth, siderophore biosynthesis and stress tolerance (Wortham *et al.*, 2007). Transport of hemine, pyrimidiine, and N-acetylneuraminate were also induced. Hemine is an iron containing porphyrin, while pyrimidine is a precursor of nucleotides. Scavenging such compounds would be easier than synthesising them *de novo*, and may provide a selective advantage in the competitive rhizosphere environment. Interestingly, N-acetylneuraminate is a signalling molecule on mammalian cell membranes that can act as receptor for virus entry and also as a carbon and nitrogen source for bacteria (Vimr *et al.*, 2004). Induction of bicarbonate transport might represent a response to acidic conditions in the rhizosphere or carbon fixation by

autotrophs. Benzoate has been shown to reverse drought stress in wheat (Beltrano *et al.*, 1999), and its uptake microbes was induced here.

Strongly (≥10 fold) induced functions specific to the wheat rhizosphere included metabolism of aromatics by enzymes such as p-cumic aldehyde dehydrogenase, phenylacetaldehyde dehydrogenase and vanillin dehydrogenase, as well as metabolism of one carbon compounds via the transcriptional regulator of formaldehyde assimilation (HxlR) and formate dehydrogenase H. Other induced responses included those to DNA damage by MutS, oxidative stress by a probable monothiol glutaredoxin and transcriptional regulator SoxR, metal homeostasis by a putative silver efflux pump, copper sensory histidine kinase, and a putative heme iron utilization protein. Regulation of catabolite repression of carbon metabolism and oxygen sensing during nitrogen fixation were induced by the two-component system response regulators CreC and FixJ respectively. Metabolism of carbon compounds such as alpha-galactosides and D-mannonate, and nitrogen compound ethanolamine was also strongly induced. Ethanolamine is found in cell membranes of plants (Rontein *et al.*, 2001) and its metabolism by bacterial might indicate the breakdown of plant cells in the rhizosphere. Motility, flagellar proteins FlbB and FliS, and chemotaxis protein CheX were strongly induced, as was a xanthan biosynthesis glycosyltransferase which might play a role in attachment to roots or plant drought tolerance (Carminati and Vetterlein, 2013). Induction of Ni/Fe-hydrogenase is indicative of oxidation of hydrogen, a biproduct of nitrogen fixation (Hunt and Layzell, 1993). Hydrogen oxidation is known contribute to fertilisation of soils that have had legumes growing in them (Dong *et al.*, 2003; Maimaiti *et al.*, 2007), but nitrogen fixation and hydrogen release by free living diazotrophs might produce the same effects, albeit less strong, in non-legume rhizospheres. There was also induction of methylglyoxal synthase, which suggesting an excess of sugar phosphates available to the microbes (Weber *et al.*, 2005).

Other functions less strongly (>1, <10 fold) induced included additional motility related functions such as flagellar hook-length control protein FliK, regulatory protein FleQ and hook protein FlgE, and stress responses by superoxide dismutase, DNA repair proteins MutS and RadC, cobalt-zinc-cadmium resistance protein CzcD, and RsbR, the positive regulator of sigma-B which controls the general stress response in *Bacillus subtilis* (Akbar *et al.*, 1997). A number of functions related to aromatic (including aromatic amino acids), sugar (uronic) acid, one-carbon compound, seven-carbon sugar, hydrogen and propionate metabolism were also observed. Seven carbon sugars are rare in nature but have been

detected in some plants (MacLeod *et al.*, 2001). Sarcosine oxidase, which demethylates sarcosine to glycine, was induced. In plants, sarcosine derivatives can protect from heavy metal toxicity (Kovacs *et al.*, 2005) and may have been produced by wheat in response to these. A heavy metal resistance protein was also induced in the microbial population. There was evidence of enhanced nitrogen cycling in the wheat rhizosphere, with induction of the nitrogenase transcriptional regulator NifA, molybdenum cofactor biosynthesis, glutamine synthetase, glutamate synthase and dehydrogenase, glutamate-1-semialdehyde dehydrogenase, as well as denitrification related proteins, specifically nitrite reductase, nitric oxide reductase and nitrous oxide reductase. Other induced functions included a sulphur oxidation, dimethylsuphide reductase, polysulfide reductase, and a sulphite reduction-associated complex as well as an archaeal glycogen debranching enzyme, pyoverdine synthetase PvdD and type III secretion system related functions. Glycogen is a carbon storage molecule in animals, pyoverdine is an iron chelator, while type III secretion systems are important virulence factors of bacteria(Hueck, 1998).

6.3.9 Transcriptional responses specific to the oat rhizosphere as determined by rapsearch2

Oat specifically induced a number of functions (Appendix Table A30), including the uptake of sugars and related compounds, such as galactose and methyl galactosides, glucose and ribose. Galactose, glucose and ribose are frequently detected in plant root exudates (Vranova *et al.*, 2013), and galactose has been shown to inhibit auxin induced growth in oat (Yamamoto, 1987), which may be a counter response to microbial auxin production, bringing it back under plant control. Organic nitrogen sources arginine and orthinine were taken up, as was magnesium.

Strongly induced (≥10 fold) functions specific to the oat rhizosphere included housekeeping proteins such as the eukaryotic translation initiation factor 4B. Developmental changes included induction of two sporulation related proteins and aerial mycelium formation biosynthesis protein BldG. In *Steptomyces* spp. BldG indirectly activates transcription of genes involved in antibiotic production(Bibb *et al.*, 2000). These observations suggest that oat stimulates the production of antibiotics in some Actinomycetes. Metabolism of carbon sources such as glucans, glucitol, glycerol, alcohols, lactate, gluconate and glycolate, and nitrogen sources such as arginine, proline and serine, were induced. Metabolism of aromatics such as hydroquinones and halobenzoates was also induced as was PcaR, a

regulator of aromatic acid metabolism (Parales and Harwood, 1993). Induction of motility related flagellar hook-associated protein 3 and methyl-accepting chemotaxis protein III for ribose and galactose was also observed. A number of stress responses were induced including tetrathionate reductase, biosynthesis of mycothiol, a periplasmic thiol:disulfide oxidoreductase, nitrous oxide reductase and nitric oxide dioxygenase. Tetrathionate can be produced by the reaction of ROS with thiosulphate, and is used as an alternative electron acceptor by bacteria such as *Salmonella enterica* serotype Typhimurium during colonisation of inflamed gut mucosa (Winter *et al.*, 2010). A similar response may be involved in colonisation of the oat rhizosphere by certain bacteria. Mycothiol is a specific protective thiol of the Actinobacteria (Newton *et al.*, 2008), and is also involved in one carbon, alcohol and polyol metabolism via a mycothiol-dependent formaldehyde dehydrogenase (Norin *et al.*, 1997). Other strongly induced functions included organo- and alkanesulphonate metabolism, as well as potential attachment related pili and alginate biosynthesis.

Other functions less strongly (>1, <10 fold) induced included metabolism of aromatics, including phenylalanine, protocatecuate and p-hydroxybenzoate, as well as arginine, sorbitol, glucans and hydrogen. A number of stress response related functions including heat shock proteins GroES and GroEL, a DNA recombination protein, a drug/metabolite (DMT) transporter and acriflavin resistance protein were also induced. Motilty related proteins FliL, flagellar basal-body rod proteins FlgF and FlgB, as well an aerotaxis sensor receptor were also induced. Phosphate, nickel, copper and iron responsive transcriptional regulators, pathogenicity islands and related functions such as pilus regulation and biosynthesis were also induced. Other induced functions included nitrilotriacetate monooxygenase and a chitin binding protein. Nitrilotriacetate is a metal chelator which some bacteria are capable of acquiring metals from and degrading (Firestone and Tiedje, 1975), while the induction of a chitin binding protein suggests increased fungal presence in the oat rhizosphere.

6.3.10 Transcriptional responses specific to the pea rhizosphere as determined by rapsearch2

Pea specifically induced a number of functions (Appendix Table A31), including transport of amino acids serine, alanine, glycine and histidine, and also glucoronate. These were found to represent 14.9, 8.1, 12.8% of amino acids in pea root exudate respectively, although

histidine was not detected in one study (Moody *et al.*, 1988). Another study detected them at 9.8, 8.1, 7.8 and 7.0% of pea root exudates respectively (Gaworzewska and Carlile, 1982). Glucoronate was found to comprise 7.5% of carbohydrate in pea root exudates but was much lower in other plants (Moody *et al.*, 1988). These compounds were also detected in pea root exudate by high throughput metabolomics (Poole lab, unpublished data). Sialic acid and inostiol transport were induced, both are involved in eukaryotic cell signalling, and inositol uptake was induced in *R. leguminosarum* during colonisation of pea, alfalfa and sugar beet rhizospheres (Ramachandran *et al.*, 2011). Uptake of organic alkanesulphonates and inorganic sulphate were induced. Organic sulphur sources represent 95% of sulphur in soils, so the latter may have been generated from the former by the action of exported sulphatases which were induced particularly strongly by pea in *R. leguminosarum* (Ramachandran *et al.*, 2011). Riboflavin and tricarballylate transport was also induced. Riboflavin is involved in a number of metabolic pathways and its uptake may represent generally heighted requirement for key co-factors due to higher activity in the pea rhizosphere. Tricaballylate is a potent inhibitor of the Kreb's cycle, specifically aconitase, and is found in mycotoxins produced by *Fusarium* spp. infecting plants. It is able to be detoxified and utilised by some bacteria (Lewis *et al.*, 2004). The pea rhizosphere induced a number of metal uptake systems including those for tungstate, manganese, nickel, zinc, all of which are important cofactors in certain enzymes. For example, tungsten can substitute for molybdenum in aldehyde oxidoreducatases, manganese is a co-factor of arabinose and xylose isomerases, nickel is found in urease and hydrogenase, and zinc is found in a variety of enzymes, including alcohol dehydrogenase and phosphatases.

A number of cell wall biosynthesis related functions were strongly (≥10 fold) induced in the pea rhizosphere. These included alanine racemase, which converts L-alanine to the D-alanine that is found in bacterial cell walls, and a proposed peptidoglycan lipid II flippase, an essential component of bacterial cell walls (Ruiz, 2008). Chemotaxis and motility related functions included an aspartate chemoreceptor protein, flagellar hook-associated protein FliD and flagellar biosynthesis protein FlgN, but also the negative regulator of flagellin synthesis FlgM, suggesting that some bacteria were down regulating motility possibly during the process of attachment to roots. Other potential attachment related functions induced included pilus assembly protein CpaD and alginate biosynthesis protein Alg8. Three eukaryotic translation initiation factors were also induced, suggesting increased translation by these organisms in the pea rhizosphere. Carbon metabolism related functions included sucrose phosphorylase, co-enzyme PQQ synthesis protein F, succinate dehydrogenase,

glucarate dehydratase, L-rhamnose isomerise, malonate decarboxylase. Coenzyme PQQ is a component of a number of enzymes, including glucose and methanol dehydrogenases (Duine *et al.*, 1980; Vanschie *et al.*, 1987), reinforcing the importance of C1 metabolism in the pea rhizosphere (Ramachandran *et al.*, 2011). Nitrogen metabolism functions included aromatic amino acid aminotransferase, nitrite reductase and nitric oxide reductase activation protein NorD transcriptional regulator NnrR. Stress response related functions included the glutathione producing hydroxyacylglutathione hydrolase, the glutathione-regulated potassium-efflux system protein KefC and the outer membrane component of a tripartite multidrug resistance system. Other induced functions included iron acquisition via bacterioferritin-associated ferredoxin, a periplasmic protein involved in high-affinity ferric iron transport and a non-heme iron-containing ferritin, as well as hydrogen oxidation by coenzyme F420-reducing hydrogenase of methanogenic Archaea. This is interesting given the likely presence of both methanol and hydrogen in the pea rhizosphere (Galbally and Kirstine, 2002; Hunt and Layzell, 1993), both potential substrates of methanogenesis. *Methanosarcina barkeri*, a methanogen with the ability to utilise both of these substrates was enriched in the pea rhizosphere in a recent study (Turner *et al.*, 2013).

Other functions less strongly (>1, <10 fold) induced included the stress related functions chlorite dismutase, outer membrane stress sensor protease DegS, glutathione reductase and glutathione S-transferase, DNA repair protein RecN and a carbazol degradation cluster. Metabolism of polyols, aromatics such as catechol and vanillin, and one carbon compounds via the serine glyoxalate cycle was observed, as was the acquisition of cobalt via an aerobic cobaltochelatase. Sulphonate monoxygenation as well as both sulphite oxidation and reduction were also observed, suggesting enhanced sulphur cycling in the pea rhizosphere.

6.3.11 Differentially expressed metabolic pathways in the rhizospheres as determined by STAMP

To investigate the differential expression of metabolic pathways in the wheat, oat and pea rhizospheres compared to soil, the hits to KO classes determined by rapsearch2 were mapped to KO pathways using STAMP (Parks and Beiko, 2010). Based on relative abundance of transcripts, there were no pathways significantly induced in the wheat rhizosphere. There were however four pathways that were repressed (Figure 6.10a), these were for valine, leucine, and isoleucine degradation, arginine and proline metabolism,

nitrotoluene degradation and the peroxisome (Table 6.3). In the oat rhizosphere, several pathways were induced (Figure 6.10b), including ethylbenzene degradation, diterpenoid biosynthesis, plant hormone signal, plant circadian rhythm and epithelial cell signalling in *Helicobacter pylori* infection (Table 6.4). Ethybenzene is not known to be a constituent of oat root exudates, but plants are known to produce a variety of aromatic compounds (Stafford, 1974) and a number of bacterial isolates from plant rhizospheres have demonstrated the ability to degrade ethylbenzene (Djokic *et al.*, 2011; Kim *et al.*, 2008). A number of these induced pathways are likely derived from the host plant, for example diterpenoids are involved in biosynthesis of phytoalexins in rice (Wilderman *et al.*, 2004). The induction of a response to *H. pylori* infection suggests that some other Epsilonproteobacteria might be pathogens of plants or other eukaryotes such as nematodes in the rhizosphere. But presumably these would have to a have similar immune systems to that of humans to induce such a response. One of the pathways repressed in the oat rhizosphere was carbon fixation by prokaryotic organisms (Table 6.3), indicating that either obligate autotrophs were selected against and or facultative autotrophs switched to a heterotrophic mode of metabolism presumably utilising plant carbon instead. The pea rhizosphere induced more pathways than oat (Figure 6.11). These included galactose metabolism, atrazine degradation, meiosis in yeast, amoebiasis and a number of human disease pathways (Table 6.4). Galactose is one of the dominant sugars in pea root exudates (Knee *et al.*, 2001) and its metabolism was induced in *R. leguminosarum* during colonisation of the rhizospheres of pea and also alfalfa (Ramachandran *et al.*, 2011). Induction of yeast meiosis suggests increased sexual reproduction in fungi, while induction of the disease pathway for amoebiasis might indicate a response of the plant to protozoan infections. The induction of human disease related pathways, particularly in the pea rhizosphere, highlights the biases in functional databases towards human sequences. These processes were clearly not occurring in soil or the rhizosphere, but their observed induction is an indication of the high level of annotation of these pathways and the overlapping nature of different functions. That is, a function involved in the metabolism of a compound might be a normal cellular process for a microbe but ony occurs in human cells in a disease state. It would be useful to exclude human disease related pathways from future analyses.

Nitrogen metabolism, as well as that of histindine, tyrosine and fatty acids (Table 6.3), was repressed in the pea rhizosphere. This is somewhat surprising given that pea, a legume, forms nitrogen-fixing symbiosis with rhizobia.

Some pathways were induced by both oat and pea, including oxidative phosphorylation, brassinosteroid biosynthesis, flagellar assembly, RNA degradation, protein export, bacterial secretion system (Table 6.4). Brassinosteroids are involved in inversely regulating the plant immune system and plant growth in response to MAMPs such as bacterial flagellin (Lin *et al.*, 2013). The induction of flagellar assembly seen here, required for microbial translocation from soil to the rhizosphere, might also have contributed to the elicitation of the plant immune response, including brassinosteroid biosynthesis. The induction of oxidative phosphorylation is consistent with the increased metabolic activity of the oat and pea rhizospheres (Figure 6.2e), while induction of RNA degradation reflects responses to the dynamic nature of the rhizosphere environment. The induction of protein export and bacterial secretion systems indicates that some microbes were entering plant tissues (the endosphere) to become either endophytic PGPRs or pathogens. A number of pathways were repressed in the rhizospheres of all three plants, these included degradation of valine, leucine, isoleucine, arginine and proline (Table 6.3). Suggesting that amino acids were less important as carbon and nitrogen sources in the rhizosphere than in they were in soil. This may be due to the increased abundance of sugars, organic acids, and alternative nitrogen sources such as nitrate and ammonium in the rhizospheres. Peroxisome function was repressed in the wheat and oat rhizospheres while oat and pea repressed glycolysis, metabolism of pyruvate, glyoxylate, dicarboxylates, butanoate, propanoate, methane and also ABC transporters (Table 6.2). These observations are surprising as dicarboxylates for example are abundant in root exudates (Vranova *et al.*, 2013) and their metabolism has been shown to contribute to rhizosphere colonisation in *R. leguminosarum* (Ramachandran *et al.*, 2011). The repression of ABC transports is interesting. While some specific substrates secreted by the plant might induce their own uptake by microbes, the high abundance of others might reduce the need for expression of some genes involved in their uptake. This may have contributed to the overall repression of ABC transporters in the oat and pea rhizospheres, and suggests that soil is an oligotrophic environment where expression of transport systems is high. This fits with the fact that the majority of carbon in soils is locked up in recalcitrant forms such a lignins and other polymers, which are only slowly broken down by consortia of microbes, releasing little bioavailable carbon at a time.

All the changes in expression of different metabolic pathways seen here, while statistically significant, were very subtle and occurred in pathways that were of low relative abundance overall. Within these pathways, different individual functions would have been both induced or repressed, some of these very strongly, but when viewed collectively as a

pathway they might have levelled out. Additionally, the comparisons were limited to relative abundance, while the rhizospheres are known to be more transcriptionally active than soil (Figure 6.2e). Taking this into account would have resulted in an increase in the number of induced pathways and a reduction in the number of repressed pathways. This effect would be strongest for pea, and weakest for wheat, which would further exemplify the similarity between soil and the rhizosphere of wheat, and the strong effect of pea on the rhizosphere microbiome.

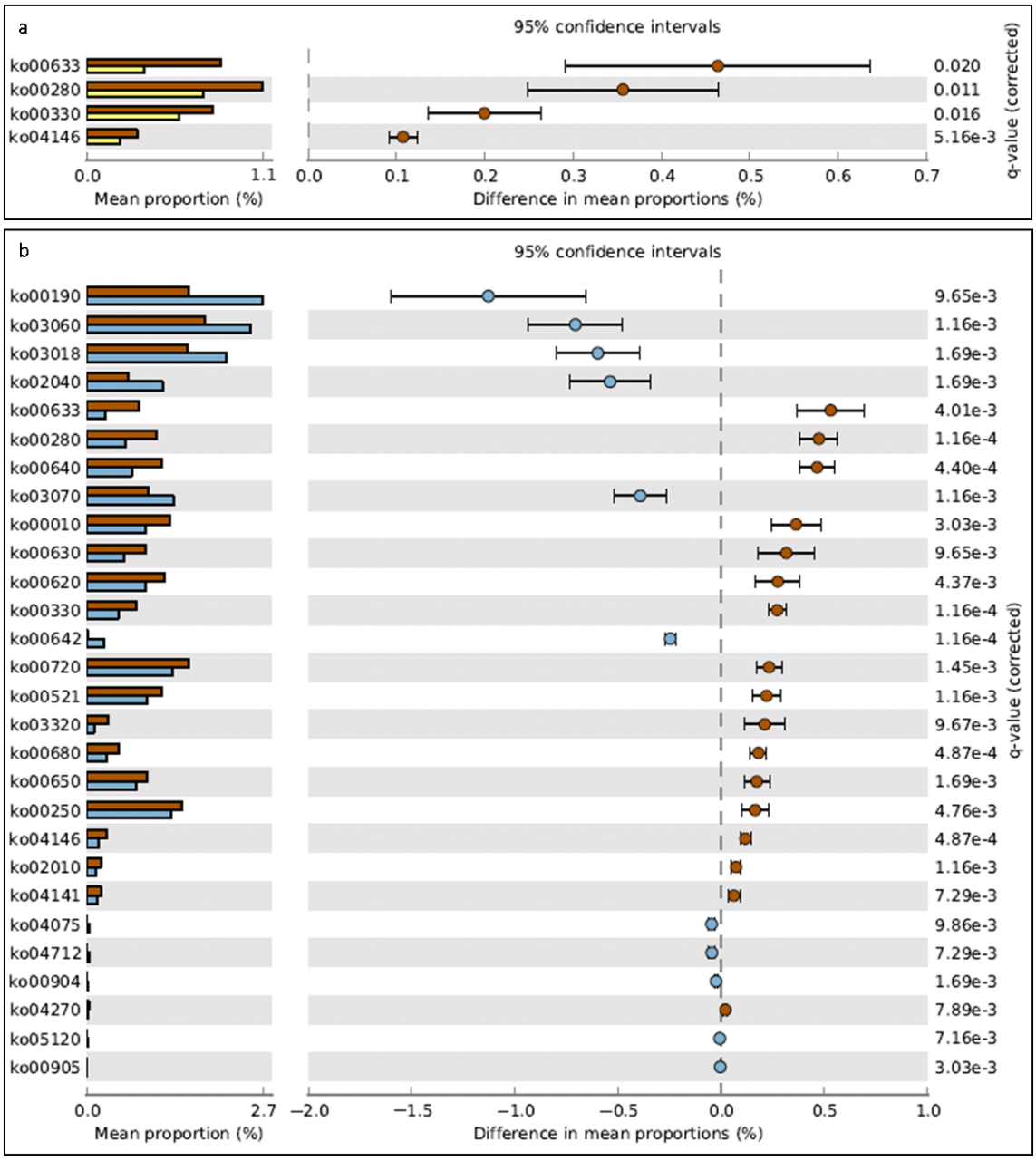


Figure 6.10: STAMP output of KO pathways differentially active in the rhizospheres of wheat (a) and oat (b) compared to soil, based on relative abundance. Soil is represented by

brown, while wheat and oat are represented by yellow and blue respectively. Differences in proportions are means (± range) and the q-values are a measurement of significance.

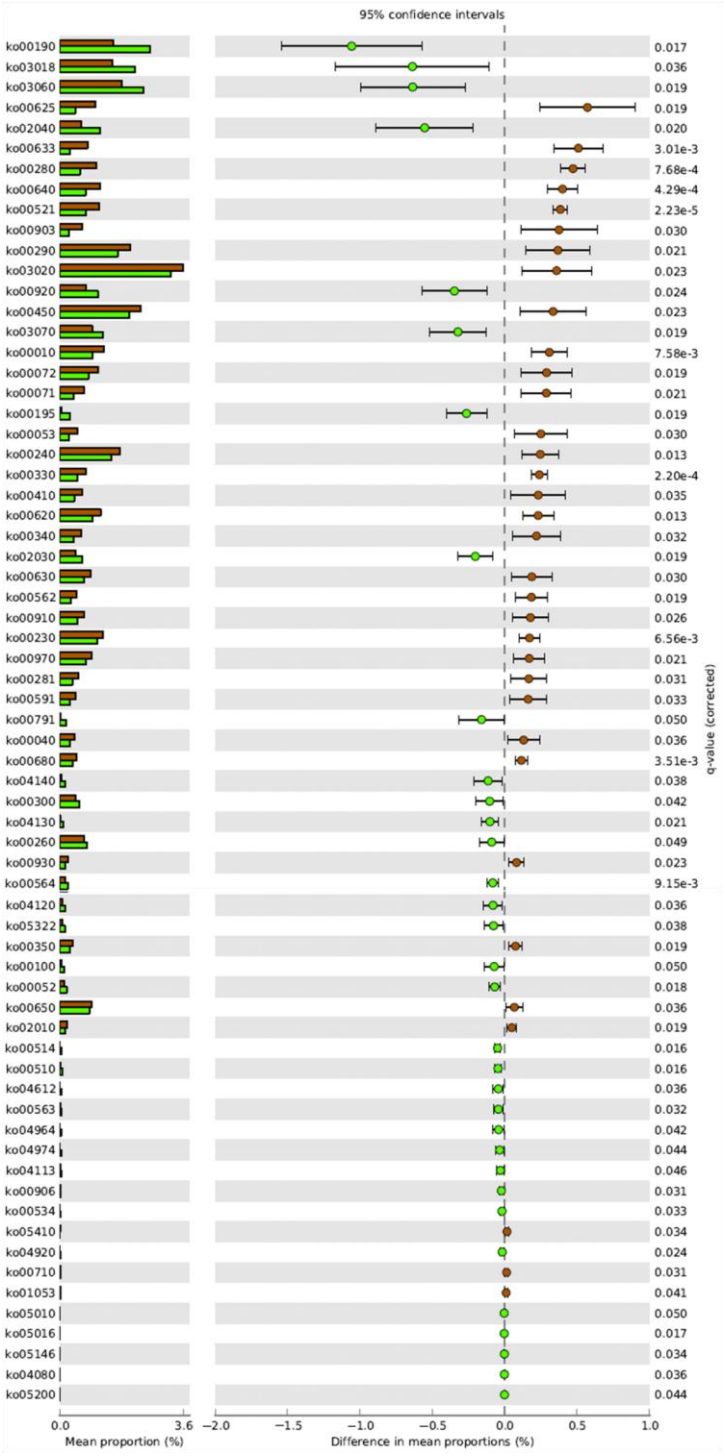


Figure 6.11: STAMP output of KO pathways differentially active in the rhizospheres of pea compared to soil, based on relative abundance. Soil is represented by brown, while pea is represented by green. Differences in mean proportions are means (± range) and the q-values are a measurement of significance.

Table 6.3: Key to the KO pathways with reduced activity in the rhizosphere derived from Figures 6.10 and 6.11.

Repressed in the rhizosphere		
KO	**Pathway**	**Rhizospheres**
ko00010	Glycolysis	Oat and Pea
ko00040	Pentose and glucuronate interconversions	Pea
ko00053	Ascorbate and aldarate metabolism	Pea
ko00071	Fatty acid metabolism	Pea
ko00072	Synthesis and degradation of ketone bodies	Pea
ko00230	Purine metabolism	Pea
ko00240	Pyrimidine metabolism	Pea
ko00250	Alanine, aspartate and glutamate metabolism	Oat
KO00280	Valine, leucine, isoleucine, degradation	Wheat, Oat and Pea
ko00281	Geraniol degradation	Pea
ko00290	Valine, leucine and isoleucine biosynthesis	Pea
ko00330	Arginine and proline metabolism	Wheat Oat and Pea
ko00340	Histidine metabolism	Pea
ko00350	Tyrosine metabolism	Pea
ko00410	beta-Alanine metabolism	Pea
ko00450	Selenocompound metabolism	Pea
ko00521	Streptomycin biosynthesis	Oat and Pea
ko00562	Inositol phosphate metabolism	Pea
ko00591	Linoleic acid metabolism	Pea
ko00620	Pyruvate metabolism	Oat and Pea
ko00625	Chloroalkane and chloroalkene degradation	Pea
ko00630	Glyoxylate and dicarboxylate metabolism	Oat and Pea
ko00633	Nitrotoluene degradation	Wheat, Oat and Pea

ko00640	Propanoate metabolism	Oat and Pea
ko00650	Butanoate metabolism	Oat and Pea
ko00680	Methane metabolism	Oat and Pea
ko00710	Carbon fixation in photosynthetic organisms	Pea
ko00720	Carbon fixation pathways in prokaryotes	Oat
ko00903	Limonene and pinene degradation	Pea
ko00910	Nitrogen metabolism	Pea
ko00930	Caprolactam degradation	Pea
ko00970	Aminoacyl-tRNA biosynthesis	Pea
ko01053	Biosynthesis of siderophore group nonribosomal peptides	Pea
ko02010	ABC transporters	Oat and Pea
ko03020	RNA polymerase	Pea
ko03320	Peroxisome proliferator-activated receptor signalling pathway	Oat
ko04141	Protein processing in endoplasmic reticulum	Oat
ko04146	Peroxisome	Wheat and Oat
ko04270	Vascular smooth muscle contraction	Oat
ko05410	Hypertrophic cardiomyopathy (HCM)	Pea

Table 6.4: Key to the KO pathways with enhanced activity in the rhizosphere derived from Figures 6.10 and 6.11.

Induced in the rhizosphere		
KO	**Pathway**	**Rhizospheres**
ko00052	Galactose metabolism	Pea
ko00100	Steroid biosynthesis	Pea
ko00190	Oxidative phosphorylation	Oat and Pea
ko00195	Photosynthesis	Pea
ko00260	Glycine, serine and threonine metabolism	Pea

ko00300	Lysine biosynthesis	Pea
ko00510	N-Glycan biosynthesis	Pea
ko00514	Other types of O-glycan biosynthesis	Pea
ko00534	Glycosaminoglycan biosynthesis - heparan sulfate / heparin	Pea
ko00563	Glycosylphosphatidylinositol(GPI)-anchor biosynthesis	Pea
ko00564	Glycerophospholipid metabolism	Pea
ko00642	Ethylbenzene degradation	Oat
ko00791	Atrazine degradation	Pea
ko00904	Diterpenoid biosynthesis	Oat
ko00905	Brassinosteroid biosynthesis	Oat and Pea
ko00906	Carotenoid biosynthesis	Pea
ko00920	Sulfur metabolism	Pea
ko02030	Bacterial chemotaxis	Pea
ko02040	Flagellar assembly	Oat and Pea
ko03018	RNA degradation	Oat and Pea
ko03060	Protein export	Oat and Pea
ko03070	Bacterial secretion system	Oat and Pea
ko04075	Plant hormone signal transduction	Oat
ko04080	Neuroactive ligand-receptor interaction	Pea
ko04113	Meiosis - yeast	Pea
ko04120	Ubiquitin mediated proteolysis	Pea
ko04130	SNARE interactions in vesicular transport	Pea
ko04140	Regulation of autophagy	Pea
ko04612	Antigen processing and presentation	Pea
ko04712	Circadian rhythm - plant	Oat
ko04920	Adipocytokine signaling pathway	Pea
ko04964	Proximal tubule bicarbonate reclamation	Pea

ko04974	Protein digestion and absorption	Pea
ko05010	Alzheimer's disease	Pea
ko05016	Huntington's disease	Pea
ko05120	Epithelial cell signaling in Helicobacter pylori infection	Oat
ko05146	Amoebiasis	Pea
ko05200	Pathways in cancer	Pea
ko05322	Systemic lupus erythematosus	Pea

6.4 Conclusion

The metabolism of microbiomes from soil and wheat, oat and pea rhizospheres was extensively analysed by two different approaches, based on different algorithms. Rapsearch2, uses an algorithm similar to BLASTx (Altschul *et al.*, 1990; Gish and States, 1993), while InterProScan uses hidden Markov models to identify homology. The databases that were used to search reads against were also different. Rapsearch2 searched against the NCBI non-redundant nucleotide database while InterProScan searched against UniProt. Despite these differences, there was general consensus that a number of microbial functions were induced specifically by one plant, by two plants, or by all of the three plants tested. There can be a high level of confidence in those functions shown to be induced by both methods. For those where only detected by one method, the question arises whether those reads were not assigned by the alternative methods or were they assigned to something else. Investigating this on a read for read basis would be time consuming, but would prove useful for validation and interpretation of such data in future.

In all samples the sequences analysed contained those derived from plants. This would potentially make conclusions difficult to infer. However, closer inspection of plant derived sequences showed the majority of them to be related to housekeeping functions such as transcription, translation and cell division. This meant that unfortunately there was insufficient sampling to determined plant root cell responses to the different rhizosphere microbiomes. However, it did mean that the conclusions drawn about metabolism from the total analysed dataset could be confidently associated with the microbiomes and would not be particularly influenced by plant material. Further analyses in future however will

completely remove plant sequences from the dataset. The effects of the presence of plant sequences on interpretation of the data can then be quantitatively assessed.

The analyses of such a large dataset here can be considered preliminary. A number of bioinformatics analysis tools are available for metagenomic and metatranscriptomic data sets. As data output from sequencing platforms increases and cost per base pair decreases, the bioinformatic effort required to determine the biological significance of data increases. Conventional tools such as BLAST are becoming obsolete, at least in their traditional forms. New analysis tools are often tested with simulated and real datasets, where the latter have been previous analysed with conventional tools. These datasets are typically small however, so for a dataset as large as the one generated here, it was important to test different analysis methods to determine which produced the most informative results with the least computational and human time and effort. The three analysis tools used here, rapsearch2 in conjunction with MEGAN, InterProScan as part of the EBI metagenomic pipeline, and MG-RAST, were useful for different elements of the analysis. Rapsearch2 and MEGAN allowed robust comparisons of taxonomic and functional compositions, with data easily extractable. The use of InterProScan reinforced the importance of strict quality filtering of Illumina sequencing data prior to homology searching. As it utilised hidden Markov models, it also allowed a comparison with the fundamentally different algorithm used in rapsearch2, and produced higher assignment rates than the latter. MG-RAST allowed numerous databases and parameters to be used to analyse taxonomic and functional information, revealing that multiple databases can improve assignments rates compared to single databases. It also allowed separation of plant derived sequences, something that was not straightforward with alternative methods. A combination of these tools will continue to be used to obtain the maximum information from this dataset

The overall effects of plants on the metabolism of the rhizosphere microbiomes were highly plant specific, with distinct representation on multidimensional scaling plots and also by the large proportions of specifically induced functions, some very strongly (>10fold), by each plant. For some functions, a few of the involved genes were up-regulated in each of the rhizospheres, and some by two or all three plants. This suggested that while many functions could be considered general to the rhizosphere, the way in which microbes in different rhizospheres carried out those functions was not the same. When differentially expressed pathways were analysed, the differences between soil and the rhizospheres were subtle. It is likely that within these pathways, some individual functions were induced

while others were repressed in response to the rhizosphere, giving the overall impression of little difference in expression of the pathway. For example if the pea rhizosphere had a increase in abundance of transcripts involved in nitrogen fixation and the soil had a higher proportion of transcripts involved in denitrification, there would be little observed difference between the two environments with regard to nitrogen metabolism. Additionally, many enzymes can catalyse reactions and their reverse reactions depending on the stoichiometry of the products and substrates. This makes interpretation of the functional role of such enzymes complicated *in situ*. The analysis of pathways was based on relative abundance, but further analyses of the data will take into account the differences in transcriptional activity between soil and the rhizospheres, which would increase the differences between the rhizospheres and soil and provide a more accurate representation. This may then reverse the observations that were unexpected, such as the repression of dicarboxylate metabolism. Importantly, direct multiple comparisons between plants will also be made rather than simply seeing what differences they shared with each other compared to soil.

The addition of an internal standard to the RNA allowed calculation of sequencing depth and absolute number of mRNA transcript per gram of soil. This is the first time these have been quantitatively measured with sufficient biological replication to allow statistical validation of comparisons. The rhizosphere of wheat was slightly more active than soil, though this difference was not statistically significant. The oat and pea rhizospheres were significantly more active than soil, highlighting a fundamental difference between unplanted soil and the rhizosphere, and consistent with past measurements of enzymatic activity and colorimetric respiratory assays (Grayston *et al.*, 1997; Vale *et al.*, 2005). This information allowed quantitative estimates of different transcript abundances to be made. All plants induced chemotaxis and motility related functions, cycling of important nutrients such as nitrogen, sulphur, and phosphorous, as well as numerous uptake systems and or metabolism of plant polymers, aromatics, sugars, organic acids, amino acids, one carbon compounds, hydrogen and metals. Resistance to stresses, particularly oxidative stress, were also induced. However the specific compounds or responses were highly plant species specific.

The analyses performed here focused on individual proteins and also the pathways which they contributed too, producing different results. While the overall sequencing depth was

determined, the extent to which specific proteins and pathways were sampled remains unknown. This will be addressed by rarefaction analysis in future. The heightened activity of the rhizospheres compared to soil would have effectively up regulated the majority of transcript abundances in the rhizospheres, meaning that very few functions were relatively repressed. Functions interpreted as being important for rhizosphere colonisation, particularly when shared by all three plants, may actually have been a result of the generally increased metabolic activity in those environments, for example a number of housekeeping functions were also induced in the rhizospheres. The MG-RAST and STAMP analyses gave some insight into important changes in relative abundance. Further to this, future analysis might look at the expression of specific metabolic genes relative to one or multiple housekeeping genes. It is important to consider however that levels of mRNA do no correlate well with levels of protein. So the induction of transcription of a particular gene does not necessarily correspond to higher amounts of its coded protein, or indeed the activity of that protein. Firstly, post-transciptional regulation and the instability of mRNA might mean than many transcripts are not translated. In addition, levels of substrates, products and inhibitors will affect the actitivies of functional proteins.

The amount of sequence generated here coupled with biological replication, the combination of metatranscriptomic and metagenomic data, the use of multiple analysis tools, and the relative and quantitative interpretations of data, allowed the metabolism of the soil and rhizosphere microbiomes to be studied in unprecedented detail. Together they revealed broad changes and both expected and novel transcriptional adaptations of microbes to a plant-associated lifestyle.

Chapter 7: General discussion and future perspectives

In this study, comparative metatranscriptomics was used to determine the taxonomic composition of active microbes as well as the metabolic process carried out by them in unplanted soil, and also in the rhizospheres of wheat, oat and pea, derived from the same soil. In addition a direct comparison was made between the taxonomic compositions of active microbes in the rhizosphere of a wild-type oat with that of an oat mutant (*sad1*) deficient in production of antifungal avenacins.

Taxonomic compositions based on rRNA allowed determination of relative abundance of all active microbes, with the exception of viruses, based on the abundance of SSU rRNA. For the first time this provided a proportional representation of prokaryotes and eukaryotes in the rhizosphere environment, revealing the similarity between the wheat rhizosphere and unplanted soil. Additionally, eukaryotes were shown to be over fivefold more abundant in the rhizospheres of oat and pea rhizospheres than in soil. Nematodes and bacterivorous protozoa were enriched in all rhizospheres, while the pea rhizosphere was highly enriched for fungi. This highlighted the complexity and diversity of plant microbiomes and highlighted the advantages of a global approach to studying them, as opposed to targeting a group of bacteria, for example, with PCR. It was not possible to determine the contribution of seed transmissible microbes to the rhizosphere microbial communities. It would be interesting to crush surface sterilised plant seeds and determine the presence and nature of any microbes residing within. Their presence could then be compared to those microbes selected in respective rhizospheres. This would be particualry important if endophytes were to be examined. In previous work, no microbial RNA was able to be isolated from the gnotobioitc rhizosphere of vermiculite grown plants (Philip Poole lab data), suggesting seed transmission is perhaps quite limited in the chosen plants.

The pea rhizosphere was highly distinct from soil and the other rhizosphere microbiomes, and the oat and wheat rhizospheres were not particularly similar despite the close relationship of the host plants. In future, this approach could be used to study the microbiomes of an unlimited variety plants including wild and elite varieties of both legumes and cereals, and even several closely related cultivars of the same species, to investigate links between host genetics, evolutionary history and microbiomes. These observations were validated using qPCR with the same plants grown in a different soil. However, both were farm soils, so it would be interesting to see the result of growing the plants in a variety of soils of distinct types. Soil provides the basal microbiome on which the

plant selects, but the native microbes and the availability of nutrients in soils might affect how and to and what extent a plant can manipulate the microbiome. This might be quite difficult to infer as differences in soil, both biotic and abiotic, would likely affect the growth and development of the plants which would in turn affect the microbiome, but would be interesting nonetheless.

The wheat variety used here was an elite line, while those of oat and pea, while domesticated, had been bred much less extensively. This raises two possibilities for why the rhizosphere of this wheat line was so similar to soil. Was it an effect of ploidy level? The wheat used was a hexaploid, while oat and pea were diploids. Or potentially of great significance, was the repression of eukaryotes related to domestication? These questions are to be answered in a study of over 20 lines of modern hexaploid wheat and its wild ancestors. Because elite varieties of wheat have been developed for greatest yield in agricultural settings, another interesting experiment would be to see how elite varieties of wheat perform in nutrient poor soils. Have they lost the ability to recruit plant growth promoting microbes? Would their wild ancestors perform better under such circumstances? Such experiments might aid in the breeding of plant for different conditions, either those anticipated in future due to climate change or currently experienced in developing countries.

Avenacins had little effect on the prokaryotic community of oat, but the eukaryotic community was strongly altered in the *sad1* mutant. This suggested that avenacins have a broader role than protecting from fungal pathogens, possibly through their actions on sterols in eukaryotic cell membranes. This demonstrated the sensitivity of metatranscriptomics to detect the effects of single mutations in host plant on the rhizosphere microbiome. In future this could be applied to complement the genetic finger printing and 16S PCR based techniques that have recently determined the differences in rhizosphere microbiomes of *A. thaliana* mutants with defective immune systems (Tkacz *et al.*, 2013b) and lacking glucosinolates (Tkacz *et al.*, 2013a), and also nodulation and mycorrhization defective mutants of *M. truncatula*. There are a number of other well characterised plant mutants which would be of interest, including strigolactone deficient *P. sativum* (Morris *et al.*, 2001) and commercial varieties engineered to resist diseases such as *Bt* maize.

The rRNA based community analysis could be optimised in a number of ways. The use of SSU rRNA proved effective for prokaryotes, and allowed straightforward comparisons with

published 16S PCR based studies. The dominance of Proteobacteria, the high abundance of other phyla, such as Actinobacteria and Firmicutes, and the selection of particular plant associated groups was consistent with published studies (Bulgarelli *et al.*, 2013). For eukaryotes however a large proportion of reads were assigned at high taxonomic levels. This was largely due to the relative similarities of eukaryotic SSU sequences and the less extensive databases. Using LSU sequences to determine their taxonomic compositions might allow improved specificity of read assignments, although the databases are less extensive and the LCA analysis parameters in MEGAN will require optimising. Utilising ITS sequences would not be realistic due to their hypervariability and poor coverage in databases and the fact that they too require targeting to particular groups of organisms, i.e. fungi. Optimising the analysis of eukaryotic sequences might allow more information to be gained on taxa that are known to be important for functioning of the rhizosphere but were not well detected, such as the mycorrhzial fungi. Further additions to the rRNA based analysis could involve spiking with an internal RNA standard, which would allow quantitative measurements of abundance of rRNA sequences per gram of soil as demonstrated in the metabolic analysis. There was a strong correlation between transcriptional activity determined by the metabolic analysis and the relative abundance of eukaryotes based on rRNA, but a direct comparison using the same samples would be needed to confirm this. Such a correlation, if confirmed, might be an indication of increased rhizodeposition by the plants which could contribute to higher proportions of rhizosphere eukaryotes. Alternatively the increased activity could be a result of the increased abundance of eukaryotes, which with larger cells, might contribute more rRNA transcripts to the sample.

To allow the study of microbial metabolism in the rhizosphere, an rRNA depletion procedure was developed for complex samples. Guided by the taxonomic composition determined by the analysis of rRNA, a combination of two subtractive hybridisation based kits was used to deplete prokaryotic and eukaryotic rRNA. By validating the success of depletion with qPCR prior to sequencing, efforts and resources would have been saved in the event of insufficient depletion determined post-sequencing. The procedure developed here is appropriate for metatranscriptomics of any complex community, and its reliance on commercial kits means it is widely available. The analyses of rRNA sequences in rRNA depleted samples revealed particular microbes that were resistant to depletion, and could potentially lead to optimisation of the kits in future. The importance of removing as much rRNA as possible, to obtain maximum numbers of useful reads, was reinforced by the

abundance of housekeeping transcripts, such as ribosomal proteins, in the remaining sequences and the fact that the majority of mRNA sequences could not be assigned to known proteins.

The analysis of protein coding genes provided an additional view of the taxonomic compositions, which largely agreed with the rRNA analysis. The large amount of sequencing used for this revealed many differentially abundant taxa between soil and the rhizospheres. Importantly, very few taxa were quantitatively depleted in the rhizospheres compared to soil, suggesting that plants don't so much deplete certain taxa but just select some taxa more than others. Plants selected a number of taxonomic groups known to possess metabolic traits potentially important for rhizosphere colonisation, such as plant growth promotion or pathogenesis. Such traits included cellulose and other plant polymer degradation, nitrogen fixation, hydrogen oxidation, methylotrophy and antibiotic production. Functional analysis revealed these traits to be more highly expressed in the rhizospheres than soil. Microbes also induced genes involved in chemotaxis, motility, attachment, pathogenesis, responses to oxidative stress, cycling of nitrogen and sulphur, acquisition of phosphorous, iron and other metals, and metabolism of a wide range of sugars, carboxylic acids, amino acids and aromatic compounds that were specific to different plant rhizospheres.

Relevant to all the studies was the importance of biological replication which is often lacking in metagenomic and metatranscriptomic analyses of microbiomes. While statistically significant differences were determined, there was large variability within samples from the same environment, unsurprising giving the complex and dynamic nature of the samples. A few samples had to be discarded due to the presence of contaminating sequences, which reduced the statistical power of comparisons. Together these reinforce the need for high numbers of biological replicates to overcome intra-sample variability and potential loss of samples. Such replication is now more feasible with platforms such as the Illumina MiSeq, although financial constraints associated with library construction still apply. In addition of biological replication, the inclusion of technical replicates would allow variability associated with library construction and sequencing to be determined. Also, to reduce variability due to abiotic condtions at the time of soil sampling, soil could be harvested from the same site at different times of the year. Performing this full factorial replication would attribute more selection power to the plants, and increase the robustness and reproduciablity of future work.

A number of the experimental aims of the project were completed while others are ongoing. The global differences in rhizosphere microbiome structure between different crop plants were determined, revealing changes in relative and total abundances and activity of bacteria, fungi and other microbes, such as nematodes and protozoa in the different rhizospheres compared to soil. A number of metabolic processes were observed to have enhanced activity in the rhizospheres compared to soil. Additionally the global level of transcription in the rhizospheres was generally higher than in soil. Microbial metabolism of a variety compounds was observed, with some common to all rhizospheres while others specific to one plant species. Many of the compounds that were sensed, transported and metabolised more so in the rhizospheres than bulk soil had been documented in plants, including from recent high-throughput metabolomic analysis of pea root exudates (Poole lab, unpublished data). It would be interesting to determine the compositions of root exudates from other plants in this way, to see if the presence particular compounds correlated with the presence or metabolism of different microbes associated with them. Directly comparing the exudates from anexically and soil grown plants would shed light on the changes in root metabolism during microbial interactions. There were many metabolic processes that were enhanced in all rhizospheres but by different means, i.e. different by enzymes carrying out similar processes, which may reflect different taxa taking part in those biochemical transformations. The preliminary functional analysis was based on differential total abundance of individual SEED assignments, and a relative comparison of expression of KO pathways. Continued analysis of metabolic pathways will take into account total abundances of transcripts, allowing quantitative analysis. Determining which microbes were contributed to each metabolic pathway will be determined in the near future.

Profiling microbial communities with metatranscriptomics allowed comparison of relative and absolute abundance of microbes and their metabolism, from multiple samples, across all domains of life, without PCR bias. This revealed profound differences in the taxonomic composition and metabolic functions of rhizosphere microbiomes between crop plants and soil. With further optimisation it will continue to be a powerful tool in microbial ecology, advancing our understanding of the complex interactions between plants and their microbiomes in the rhizosphere.

Bibliography

Adesemoye AO, Torbert HA, Kloepper JW (2009). Plant Growth-Promoting Rhizobacteria Allow Reduced Application Rates of Chemical Fertilizers. *Microbial Ecology* **58:** 921-929.

Akbar S, Kang CM, Gaidenko TA, Price CW (1997). Modulator protein RsbR regulates environmental signalling in the general stress pathway of Bacillus subtilis. *Molecular Microbiology* **24:** 567-578.

Akiyama K, Hayashi H (2006). Strigolactones: Chemical signals for fungal symbionts and parasitic weeds in plant roots. *Annals of Botany* **97:** 925-931.

Alon S, Vigneault F, Eminaga S, Christodoulou DC, Seidman JG, Church GM *et al* (2011). Barcoding bias in high-throughput multiplex sequencing of miRNA. *Genome Research* **21:** 1506-1511.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215:** 403-410.

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L *et al* (2011). Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nature Structural & Molecular Biology* **18:** 1435-1440.

Andrews JH (1992). Biological-control in the phyllosphere. *Annual Review of Phytopathology* **30:** 603-635.

Armah CN, Mackie AR, Roy C, Price K, Osbourn AE, Bowyer P *et al* (1999). The membrane-permeabilizing effect of avenacin A-1 involves the reorganization of bilayer cholesterol. *Biophysical Journal* **76:** 281-290.

Atamna-Ismaeel N, Finkel OM, Glaser F, Sharon I, Schneider R, Post AF *et al* (2012). Microbial rhodopsins on leaf surfaces of terrestrial plants. *Environmental Microbiology* **14:** 140-146.

Badri DV, Chaparro JM, Zhang R, Shen Q, Vivanco JM (2013a). Application of Natural Blends of Phytochemicals Derived from the Root Exudates of Arabidopsis to the Soil Reveal That Phenolic-related Compounds Predominantly Modulate the Soil Microbiome. *The Journal of biological chemistry* **288:** 4502-4512.

Badri DV, Zolla G, Bakker MG, Manter DK, Vivanco JM (2013b). Potential impact of soil microbiomes on the leaf metabolome and on herbivore feeding behavior. *The New phytologist* **198:** 264-273.

Bailly J, Fraissinet-Tachet L, Verner MC, Debaud JC, Lemaire M, Wesolowski-Louvel M *et al* (2007). Soil eukaryotic functional diversity, a metatranscriptomic approach. *Isme Journal* **1:** 632-642.

Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM (2006). The role of root exudates in rhizosphere interations with plants and other organisms. *Annual Review of Plant Biology* **57:** 233-266.

Bakker MG, Manter DK, Sheflin AM, Weir TL, Vivanco JM (2012). Harnessing the rhizosphere microbiome through plant breeding and agricultural management. *Plant and Soil* **360:** 1-13.

Barea JM, Pozo MJ, Azcon R, Azcon-Aguilar C (2005). Microbial co-operation in the rhizosphere. *Journal of Experimental Botany* **56:** 1761-1778.

Bari R, Jones J (2009). Role of plant hormones in plant defence responses. *Plant Molecular Biology* **69:** 473-488.

Bartholomew GW, Alexander M (1979). Microbial-metabolism of carbon-monoxide in culture and in soil. *Applied and Environmental Microbiology* **37:** 932-937.

Bednarek P (2012). Chemical warfare or modulators of defence responses - the function of secondary metabolites in plant immunity. *Current Opinion in Plant Biology* **15:** 407-414.

Bednarek P, Osbourn A (2009). Plant-Microbe Interactions: Chemical Diversity in Plant Defense. *Science* **324:** 746-748.

Behie SW, Zelisko PM, Bidochka MJ (2012). Endophytic Insect-Parasitic Fungi Translocate Nitrogen Directly from Insects to Plants. *Science* **336:** 1576-1577.

Bell SG, Wong LL (2007). P450 enzymes from the bacterium *Novosphingobium aromaticivorans*. *Biochemical and Biophysical Research Communications* **360:** 666-672.

Beltrano J, Ronco MG, Montaldi ER (1999). Drought stress syndrome in wheat is provoked by ethylene evolution imbalance and reversed by rewatering, aminoethoxyvinylglycine, or sodium benzoate. *Journal of Plant Growth Regulation* **18:** 59-64.

Benning MM, Wesenberg G, Liu RQ, Taylor KL, Dunaway-Mariano D, Holden HM (1998). The three-dimensional structure of 4-hydroxybenzoyl-CoA thioesterase from *Pseudomonas* sp. strain CBS-3. *Journal of Biological Chemistry* **273:** 33572-33579.

Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG *et al* (2008). Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456:** 53-59.

Berendsen RL, Pieterse CMJ, Bakker P (2012). The rhizosphere microbiome and plant health. *Trends in Plant Science* **17:** 478-486.

Berry AM, Harriott OT, Moreau RA, Osman SF, Benson DR, Jones AD (1993). Hopanoid lipids compose the frankia vesicle envelope, presumptive barrier of oxygen diffusion to nitrogenase. *Proceedings of the National Academy of Sciences of the United States of America* **90:** 6091-6094.

Bertin C, Yang XH, Weston LA (2003). The role of root exudates and allelochemicals in the rhizosphere. *Plant and Soil* **256:** 67-83.

Bibb MJ, Molle V, Buttner MJ (2000). sigma(BldN), an extracytoplasmic function RNA polymerase sigma factor required for aerial mycelium formation in *Streptomyces coelicolor* A3(2). *Journal of Bacteriology* **182:** 4606-4616.

Bitas V, Kim H-S, Bennett JW, Kang S (2013). Sniffing on microbes: diverse roles of microbial volatile organic compounds in plant health. *Molecular plant-microbe interactions : MPMI* **26:** 835-843.

Bittel P, Robatzek S (2007). Microbe-associated molecular patterns (MAMPs) probe plant immunity. *Current Opinion in Plant Biology* **10:** 335-341.

Bloemberg GV, Lugtenberg BJJ (2001). Molecular basis of plant growth promotion and biocontrol by rhizobacteria. *Current Opinion in Plant Biology* **4:** 343-350.

Bodenhausen N, Horton MW, Bergelson J (2013). Bacterial Communities Associated with the Leaves and the Roots of Arabidopsis thaliana. *PLoS One* **8:** e56329.

Bokulich NA, Subramanian S, Faith JJ, Gevers D, Gordon JI, Knight R *et al* (2013). Quality-filtering vastly improves diversity estimates from Illumina amplicon sequencing. *Nature Methods* **10:** 57-59.

Bomar L, Maltz M, Colston S, Graf J (2011). Directed Culturing of Microorganisms Using Metatranscriptomics. *Mbio* **2:** e00012-11.

Bonfante P (2010). Plant–Fungal Interactions in Mycorrhizas. *eLS*. John Wiley & Sons, Ltd.

Botero LM, D'Imperio S, Burr M, McDermott TR, Young M, Hassett DJ (2005). Poly(A) polymerase modification and reverse transcriptase PCR amplification of environmental RNA. *Applied and Environmental Microbiology* **71:** 1267-1275.

Bouarab K, Melton R, Peart J, Baulcombe D, Osbourn A (2002). A saponin-detoxifying enzyme mediates suppression of plant defences. *Nature* **418:** 889-892.

Bouffaud ML, Kyselkova M, Gouesnard B, Grundmann G, Muller D, Moenne-Loccoz Y (2012). Is diversification history of maize influencing selection of soil bacteria by roots? *Molecular Ecology* **21:** 195-206.

Bowyer P, Clarke BR, Lunness P, Daniels MJ, Osbourn AE (1995). Host-range of a plant-pathogenic fungus determined by a saponin detoxifying enzyme. *Science* **267:** 371-374.

Brady A, Salzberg SL (2009). Phymm and PhymmBL: metagenomic phylogenetic classification with interpolated Markov models. *Nature Methods* **6:** 673-676.

Brandt LJ (2013). American Journal of Gastroenterology Lecture: Intestinal Microbiota and the Role of Fecal Microbiota Transplant (FMT) in Treatment of *C. difficile* Infection. *The American journal of gastroenterology* **108:** 177-85.

Bremer C, Braker G, Matthies D, Beierkuhnlein C, Conrad R (2009). Plant presence and species combination, but not diversity, influence denitrifier activity and the composition of nirK-type denitrifier communities in grassland soil. *Fems Microbiology Ecology* **70:** 377-387.

Bressan M, Roncato M-A, Bellvert F, Comte G, Haichar FEZ, Achouak W *et al* (2009). Exogenous glucosinolate produced by *Arabidopsis thaliana* has an impact on microbes in the rhizosphere and plant roots. *Isme J* **3:** 1243-1257.

Broeckling CD, Broz AK, Bergelson J, Manter DK, Vivanco JM (2008). Root exudates regulate soil fungal community composition and diversty. *Applied and Environmental Microbiology* **74:** 738-744.

Broekaert WF, Terras FRG, Cammue BPA, Osborn RW (1995). Plant defensins - novel antimicrobial peptides as components of the host-defense system. *Plant Physiology* **108:** 1353-1358.

Bulgarelli D, Rott M, Schlaeppi K, van Themaat EVL, Ahmadinejad N, Assenza F *et al* (2012). Revealing structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488:** 91-95.

Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P (2013). Structure and Functions of the Bacterial Microbiota of Plants. *Annual Review of Plant Biology* **64:** 807-838.

Burd GI, Dixon DG, Glick BR (2000). Plant growth-promoting bacteria that decrease heavy metal toxicity in plants. *Canadian Journal of Microbiology* **46:** 237-245.

Camilli A, Bassler BL (2006). Bacterial small-molecule signaling pathways. *Science* **311:** 1113-1116.

Caporaso JG, Lauber CL, Walters WA, Berg-Lyons D, Huntley J, Fierer N *et al* (2012). Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *Isme Journal* **6:** 1621-1624.

Carminati A, Vetterlein D (2013). Plasticity of rhizosphere hydraulic properties as a key for efficient utilization of scarce resources. *Annals of Botany* **112:** 277-290.

Carter JP, Spink J, Cannon PF, Daniels MJ, Osbourn AE (1999). Isolation, characterization, and avenacin sensitivity of a diverse collection of cereal-root-colonizing fungi. *Applied and Environmental Microbiology* **65:** 3364-3372.

Castaldini M, Turrini A, Sbrana C, Benedetti A, Marchionni M, Mocali S *et al* (2005). Impact of *Bt* corn on rhizospheric and on beneficial mycorrhizal symbiosis and soil eubacterial communities iosis in experimental microcosms. *Applied and Environmental Microbiology* **71:** 6719-6729.

Cavaglieri L, Orlando J, Etcheverry M (2009). Rhizosphere microbial community structure at different maize plant growth stages and root locations. *Microbiological Research* **164:** 391-399.

Chaparro JM, Badri DV, Bakker MG, Sugiyama A, Manter DK, Vivanco JM (2013). Root exudation of phytochemicals in *Arabidopsis* follows specific patterns that are developmentally programmed and correlate with soil microbial functions. *PLoS One* **8:** e55731.

Chen WM, de Faria SM, Straliotto R, Pitard RM, Simoes-Araujo JL, Chou JF *et al* (2005). Proof that *Burkholderia* strains form effective symbioses with legumes: a study of novel

mimosa-nodulating strains from South America. *Applied and Environmental Microbiology* **71:** 7461-7471.

Cheung PY, Kinkle BK (2001). *Mycobacterium* diversity and pyrene mineralization in petroleum-contaminated soils. *Applied and Environmental Microbiology* **67:** 2222-2229.

Chin KJ, Janssen PH (2002). Propionate formation by Opitutus terrae in pure culture and in mixed culture with a hydrogenotrophic methanogen and implications for carbon fluxes in anoxic rice paddy soil. *Applied and Environmental Microbiology* **68:** 2089-2092.

Chow V, Nong G, St John FJ, Rice JD, Dickstein E, Chertkov O *et al* (2012). Complete genome sequence of *Paenibacillus* sp. strain JDR-2. *Standards in Genomic Sciences* **6:** 1-10.

Christie PJ, Atmakuri K, Krishnamoorthy V, Jakubowski S, Cascales E (2005). Biogenesis, architecture, and function of bacterial type IV secretion systems. *Annual Review of Microbiology* **59:** 451-485.

Chung EJ, Lim HK, Kim JC, Choi GJ, Park EJ, Lee MH *et al* (2008). Forest soil metagenome gene cluster involved in antifungal activity expression in *Escherichia coli*. *Applied and Environmental Microbiology* **74:** 723-730.

Ciulla D, Giannoukos G, Earl A, Feldgarden M, Gevers D, Levin J *et al* (2010). Evaluation of bacterial ribosomal RNA (rRNA) depletion methods for sequencing microbial community transcriptomes. *Genome Biology* **11:** 1.

Clarke KR (1993). Nonparametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* **18:** 117-143.

Cole JR, Wang Q, Cardenas E, Fish J, Chai B, Farris RJ *et al* (2009). The Ribosomal Database Project: improved alignments and new tools for rRNA analysis. *Nucleic Acids Research* **37:** D141-D145.

Combes-Meynet E, Pothier JF, Moenne-Loccoz Y, Prigent-Combaret C (2011). The *Pseudomonas* Secondary Metabolite 2,4-Diacetylphloroglucinol Is a Signal Inducing Rhizoplane Expression of *Azospirillum* Genes Involved in Plant-Growth Promotion. *Molecular Plant-Microbe Interactions* **24:** 271-284.

Conrad R, Erkel C, Liesack W (2006). Rice Cluster I methanogens, an important group of Archaea producing greenhouse gas in soil. *Current Opinion in Biotechnology* **17:** 262-267.

Costa R, Gotz M, Mrotzek N, Lottmann J, Berg G, Smalla K (2006). Effects of site and plant species on rhizosphere community structure as revealed by molecular analysis of microbial guilds. *Fems Microbiology Ecology* **56:** 236-249.

Cotta SR, Dias ACF, Marriel IE, Gomes EA, van Elsas JD, Seldin L (2013). Temporal dynamics of microbial communities in the rhizosphere of two genetically modified (GM) maize hybrids in tropical agrosystems. *Antonie Van Leeuwenhoek International Journal of General and Molecular Microbiology* **103:** 589-601.

Crombie WML, Crombie L (1986). Distribution of avenacins a-1, a-2, b-1 and b-2 in oat roots - their fungicidal activity towards take-all fungus. *Phytochemistry* **25:** 2069-2073.

Croswell A, Amir E, Teggatz P, Barman M, Salzman NH (2009). Prolonged Impact of Antibiotics on Intestinal Microbial Ecology and Susceptibility to Enteric *Salmonella* Infection. *Infection and Immunity* **77:** 2741-2753.

da Rocha UN, Andreote FD, de Azevedo JL, van Elsas JD, van Overbeek LS (2010). Cultivation of hitherto-uncultured bacteria belonging to the Verrucomicrobia subdivision 1 from the potato (*Solanum tuberosum* L.) rhizosphere. *Journal of Soils and Sediments* **10:** 326-339.

da Rocha UN, van Overbeek L, van Elsas JD (2009). Exploration of hitherto-uncultured bacteria from the rhizosphere. *Fems Microbiology Ecology* **69:** 313-328.

Dakora FD, Phillips DA (2002). Root exudates as mediators of mineral acquisition in low-nutrient environments. *Plant and Soil* **245:** 35-47.

Damiani I, Baldacci-Cresp F, Hopkins J, Andrio E, Balzergue S, Lecomte P *et al* (2012). Plant genes involved in harbouring symbiotic rhizobia or pathogenic nematodes. *New Phytologist* **194:** 511-522.

Damon C, Lehembre F, Oger-Desfeux C, Luis P, Ranger J, Fraissinet-Tachet L *et al* (2012). Metatranscriptomics Reveals the Diversity of Genes Expressed by Eukaryotes in Forest Soils. *PLoS One* **7:** e28967.

Dandie CE, Wertz S, Leclair CL, Goyer C, Burton DL, Patten CL *et al* (2011). Abundance, diversity and functional gene expression of denitrifier communities in adjacent riparian and agricultural zones. *Fems Microbiology Ecology* **77:** 69-82.

Darvill AG, Albersheim P (1984). Phytoalexins and their elicitors - a defense against microbial infection in plants. *Annual Review of Plant Physiology and Plant Molecular Biology* **35:** 243-275.

Davis KER, Sangwan P, Janssen PH (2011). Acidobacteria, Rubrobacteridae and Chloroflexi are abundant among very slow-growing and mini-colony-forming soil bacteria. *Environmental Microbiology* **13:** 798-805.

De Luca G, Barakat M, Ortet P, Fochesato S, Jourlin-Castelli C, Ansaldi M *et al* (2011). The Cyst-Dividing Bacterium *Ramlibacter tataouinensis* TTB310 Genome Reveals a Well-Stocked Toolbox for Adaptation to a Desert Environment. *PLoS One* **6:** e23784.

de Menezes A, Clipson N, Doyle E (2012). Comparative metatranscriptomics reveals widespread community responses during phenanthrene degradation in soil. *Environmental Microbiology* **14:** 2577-2588.

de Weert S, Vermeiren H, Mulders IHM, Kuiper I, Hendrickx N, Bloemberg GV *et al* (2002). Flagella-driven chemotaxis towards exudate components is an important trait for tomato root colonization by *Pseudomonas fluorescens*. *Molecular Plant-Microbe Interactions* **15:** 1173-1180.

Deacon JW, Mitchell RT (1985). Toxicity of oat roots, oat root extracts, and saponins to zoospores of *Pythium* spp and other fungi. *Transactions of the British Mycological Society* **84:** 479-487.

DeAngelis KM, Brodie EL, DeSantis TZ, Andersen GL, Lindow SE, Firestone MK (2009). Selective progressive response of soil microbial community to wild oat roots. *Isme Journal* **3:** 168-178.

Degnan PH, Ochman H (2012). Illumina-based analysis of microbial community diversity. *Isme Journal* **6:** 183-194.

Delmotte N, Knief C, Chaffron S, Innerebner G, Roschitzki B, Schlapbach R *et al* (2009). Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. *Proceedings of the National Academy of Sciences of the United States of America* **106:** 16428-16433.

Dennis PG, Miller AJ, Hirsch PR (2010). Are root exudates more important than other sources of rhizodeposits in structuring rhizosphere bacterial communities? *Fems Microbiology Ecology* **72:** 313-327.

Desmond E, Gribaldo S (2009). Phylogenomics of Sterol Synthesis: Insights into the Origin, Evolution, and Diversity of a Key Eukaryotic Feature. *Genome Biology and Evolution* **1:** 364-381.

Dijksterhuis J, Sanders M, Gorris LGM, Smid EJ (1999). Antibiosis plays a role in the context of direct interaction during antagonism of *Paenibacillus polymyxa* towards *Fusarium oxysporum*. *Journal of Applied Microbiology* **86:** 13-21.

Djokic L, Narancic T, Nikodinovic-Runic J, Savic M, Vasiljevic B (2011). Isolation and characterization of four novel Gram-positive bacteria associated with the rhizosphere of two endemorelict plants capable of degrading a broad range of aromatic substrates. *Applied Microbiology and Biotechnology* **91:** 1227-1238.

Dohrmann AB, Kuting M, Junemann S, Jaenicke S, Schluter A, Tebbe CC (2013). Importance of rare taxa for bacterial diversity in the rhizosphere of Bt- and conventional maize varieties. *Isme Journal* **7:** 37-49.

Donato JJ, Moe LA, Converse BJ, Smart KD, Berklein FC, McManus PS *et al* (2010). Metagenomic Analysis of Apple Orchard Soil Reveals Antibiotic Resistance Genes Encoding Predicted Bifunctional Proteins. *Applied and Environmental Microbiology* **76:** 4396-4401.

Dong Z, Wu L, Kettlewell B, Caldwell CD, Layzell DB (2003). Hydrogen fertilization of soils - is this a benefit of legumes in rotation? *Plant Cell and Environment* **26:** 1875-1879.

Doornbos RF, Geraats BPJ, Kuramae EE, Van Loon LC, Bakker P (2011). Effects of Jasmonic Acid, Ethylene, and Salicylic Acid Signaling on the Rhizosphere Bacterial Community of *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **24:** 395-407.

Dou DL, Zhou JM (2012). Phytopathogen Effectors Subverting Host Immunity: Different Foes, Similar Battleground. *Cell Host & Microbe* **12:** 484-495.

Dray S, Dufour AB (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22:** 1-20.

Duine JA, Frank J, Verwiel PEJ (1980). Structure and activity of prosthetic group of methanol dehydrogenase. *European Journal of Biochemistry* **108:** 187-192.

Duran N, Justo GZ, Ferreira CV, Melon PS, Cordi L, Martins D (2007). Violacein: properties and biological activities. *Biotechnology and Applied Biochemistry* **48:** 127-133.

Edgar RC (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26:** 2460-2461.

Edwards RA, Rohwer F (2005). Viral metagenomics. *Nature Reviews Microbiology* **3:** 504-510.

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G *et al* (2009). Real-Time DNA Sequencing from Single Polymerase Molecules. *Science* **323:** 133-138.

Eilers KG, Lauber CL, Knight R, Fierer N (2010). Shifts in bacterial community structure associated with inputs of low molecular weight carbon compounds to soil. *Soil Biology & Biochemistry* **42:** 896-903.

Eshinimaev BT, Medvedkova KA, Khmelenina VN, Suzina NE, Osipov GA, Lysenko AM *et al* (2004). New thermophilic methanotrophs of the genus *Methylocaldum*. *Mikrobiologiya* **73:** 530-539.

Evguenieva-Hackenberg E (2005). Bacterial ribosomal RNA in pieces. *Molecular Microbiology* **57:** 318-325.

Fan TWM, Lane AN, Shenker M, Bartley JP, Crowley D, Higashi RM (2001). Comprehensive chemical profiling of gramineous plant root exudates using high-resolution NMR and MS. *Phytochemistry* **57:** 209-221.

Fang WG, St Leger RJ (2010). Mrt, a Gene Unique to Fungi, Encodes an Oligosaccharide Transporter and Facilitates Rhizosphere Competency in *Metarhizium robertsii*. *Plant Physiology* **154:** 1549-1557.

Farinati S, DalCorso G, Panigati M, Furini A (2011). Interaction between selected bacterial strains and *Arabidopsis halleri* modulates shoot proteome and cadmium and zinc accumulation. *Journal of Experimental Botany* **62:** 3433-3447.

Farrell M, Hill PW, Farrar J, DeLuca TH, Roberts P, Kielland K *et al* (2013). Oligopeptides Represent a Preferred Source of Organic N Uptake: A Global Phenomenon? *Ecosystems* **16:** 133-145.

Farrelly V, Rainey FA, Stackebrandt E (1995). Effect of genome size and rrn gene copy number on PCR amplification of 16S ribosomal-RNA genes from a mixture of bacterial species. *Applied and Environmental Microbiology* **61:** 2798-2801.

Fdez-Gubieda ML, Muela A, Alonso J, Garcia-Prieto A, Olivi L, Fernandez-Pacheco R *et al* (2013). Magnetite Biomineralization in *Magnetospirillum gryphiswaldense*: Time-Resolved Magnetic and Structural Studies. *Acs Nano* **7:** 3297-3305.

Feil H, Feil WS, Chain P, Larimer F, DiBartolo G, Copeland A *et al* (2005). Comparison of the complete genome sequences of *Pseudomonas syringae* pv. syringae B728a and pv. tomato DC3000. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 11064-11069.

Firestone MK, Tiedje JM (1975). Biodegradation of metal-nitrilotriacetate complexes by a *Pseudomonas* species - mechanism of reaction. *Applied Microbiology* **29:** 758-764.

Fisher MM, Triplett EW (1999). Automated approach for ribosomal intergenic spacer analysis of microbial diversity and its application to freshwater bacterial communities. *Applied and Environmental Microbiology* **65:** 4630-4636.

Flores HE, Galston AW (1982). Polyamines and plant stress - activation of putrescine biosynthesis by osmotic shock. *Science* **217:** 1259-1261.

Folman LB, Postma J, van Veen JA (2003). Characterisation of *Lysobacter enzymogenes* (Christensen and Cook 1978) strain 3.1T8, a powerful antagonist of fungal diseases of cucumber. *Microbiological Research* **158:** 107-115.

Frias-Lopez J, Shi Y, Tyson GW, Coleman ML, Schuster SC, Chisholm SW *et al* (2008). Microbial community gene expression in ocean surface waters. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 3805-3810.

Friedrich MW (2005). Methyl-coenzyme M reductase genes: Unique functional markers for methanogenic and anaerobic methane-oxidizing Archaea. *Environmental Microbiology*. **397:** 428-442.

Galbally IE, Kirstine W (2002). The production of methanol by flowering plants and the global cycle of methanol. *Journal of Atmospheric Chemistry* **43:** 195-229.

Gan HM, Chew TH, Hudson AO, Savka MA (2012). Genome Sequence of *Novosphingobium* sp strain Rr 2-17, a nopaline crown gall-associated bacterium isolated from *Vitis vinifera* L. grapevine. *Journal of Bacteriology* **194:** 5137-5138.

Garbeva P, van Elsas JD, van Veen JA (2008). Rhizosphere microbial community and its response to plant species and soil history. *Plant and Soil* **302:** 19-32.

Garcia-Martinez J, Acinas SG, Anton AI, Rodriguez-Valera F (1999). Use of the 16S-23S ribosomal genes spacer region in studies of prokaryotic diversity. *Journal of Microbiological Methods* **36:** 55-64.

Garrity GM, Brown A, Vickers RM (1980). *Tatlockia* and *Fluoribacter* - 2 new genera of organisms resembling *Legionella pneumophila*. *International Journal of Systematic Bacteriology* **30:** 609-614.

Gaspar YM, Nam J, Schultz CJ, Lee LY, Gilson PR, Gelvin SB *et al* (2004). Characterization of the *Arabidopsis* lysine-rich arabinogalactan-protein AtAGP17 mutant (rat1) that results in a decreased efficiency of *Agrobacterium* transformation. *Plant Physiology* **135:** 2162-2171.

Gaworzewska ET, Carlile MJ (1982). Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. *Journal of General Microbiology* **128:** 1179-1188.

George E, Roemheld V, Marschner H (1994). Contribution of mycorrhizal fungi to micronutrient uptake by plants. *Biochemistry of metal micronutrients in the rhizosphere*. CRC Press, Inc., Boca Raton, Florida, USA; CRC Press, London, England, UK. pp 93-109.

George IF, Hartmann M, Liles MR, Agathos SN (2011). Recovery of As-Yet-Uncultured Soil Acidobacteria on Dilute Solid Media. *Applied and Environmental Microbiology* **77:** 8184-8188.

Gerlach W, Junemann S, Tille F, Goesmann A, Stoye J (2009). WebCARMA: a web application for the functional and taxonomic classification of unassembled metagenomic reads. *Bmc Bioinformatics* **10:** 430.

Ghosh S, Ghosh P, Maiti TK (2011). Production and metabolism of indole acetic acid (IAA) by root nodule bacteria (Rhizobium): A Review. *Journal of Pure and Applied Microbiology* **5:** 523-540.

Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ *et al* (2012). Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biology* **13:** R23.

Gifford SM, Sharma S, Rinta-Kanto JM, Moran MA (2011). Quantitative analysis of a deeply sequenced marine microbial metatranscriptome. *Isme Journal* **5:** 461-472.

Gilbert JA, Field D, Huang Y, Edwards R, Li W, Gilna P *et al* (2008). Detection of Large Numbers of Novel Sequences in the Metatranscriptomes of Complex Marine Microbial Communities. *PLoS One* **3:** e3042.

Gilbert JA, Meyer F, Jansson J, Gordon J, Pace N, Tiedje J *et al* (2010). The Earth Microbiome Project: Meeting report of the "1st EMP meeting on sample selection and acquisition" at Argonne National Laboratory October 6(th) 2010. *Standards in Genomic Sciences* **3:** 249-253.

Gillespie DE, Brady SF, Bettermann AD, Cianciotto NP, Liles MR, Rondon MR *et al* (2002). Isolation of antibiotics turbomycin A and B from a metagenomic library of soil microbial DNA. *Applied and Environmental Microbiology* **68:** 4301-4306.

Gish W, States DJ (1993). Identification of protein coding regions by database similarity search. *Nature Genetics* **3:** 266-272.

Glick BR (2005). Modulation of plant ethylene levels by the bacterial enzyme ACC deaminase. *Fems Microbiology Letters* **251:** 1-7.

Gomez-Alvarez V, Teal TK, Schmidt TM (2009). Systematic artifacts in metagenomes from complex microbial communities. *Isme Journal* **3:** 1314-1317.

Gosalbes MJ, Durban A, Pignatelli M, Abellan JJ, Jimenez-Hernandez N, Perez-Cobas AE *et al* (2011). Metatranscriptomic approach to analyze the functional human gut microbiota. *PLoS One* **6:** e17447.

Gottel NR, Castro HF, Kerley M, Yang ZM, Pelletier DA, Podar M *et al* (2011). Distinct Microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. *Applied and Environmental Microbiology* **77:** 5934-5944.

Grayston SJ, Vaughan D, Jones D (1997). Rhizosphere carbon flow in trees, in comparison with annual plants: The importance of root exudation and its impact on microbial activity and nutrient availability. *Applied Soil Ecology* **5:** 29-56.

Greenblum S, Turnbaugh PJ, Borenstein E (2012). Metagenomic systems biology of the human gut microbiome reveals topological shifts associated with obesity and inflammatory bowel disease. *Proceedings of the National Academy of Sciences of the United States of America* **109:** 594-599.

Gurtler JB, Douds DD, Dirks BP, Quinlan JJ, Nicholson AM, Phillips JG *et al* (2013). *Salmonella* and *Escherichia coli* O157:H7 survival in soil and translocation into leeks (*Allium porrum*) as influenced by an arbuscular mycorrhizal fungus (*Glomus intraradices*). *Applied and Environmental Microbiology* **79:** 1813-1820.

Gutierrez-Manero FJ, Ramos-Solano B, Probanza A, Mehouachi J, Tadeo FR, Talon M (2001). The plant-growth-promoting rhizobacteria *Bacillus pumilus* and *Bacillus licheniformis* produce high amounts of physiologically active gibberellins. *Physiologia Plantarum* **111:** 206-211.

Haas D, Keel C (2003). Regulation of antibiotic production in root-colonizing *Pseudomonas* spp. and relevance for biological control of plant disease. *Annual Review of Phytopathology* **41:** 117-153.

Haichar FE, Marol C, Berge O, Rangel-Castro JI, Prosser JI, Balesdent J *et al* (2008). Plant host habitat and root exudates shape soil bacterial community structure. *Isme Journal* **2:** 1221-1230.

Haichar FE, Roncato MA, Achouak W (2012). Stable isotope probing of bacterial community structure and gene expression in the rhizosphere of Arabidopsis thaliana. *Fems Microbiology Ecology* **81:** 291-302.

Haichar FEZ, Achouak W, Christen R, Heulin T, Marol C, Marais MF *et al* (2007). Identification of cellulolytic bacteria in soil by stable isotope probing. *Environmental Microbiology* **9:** 625-634.

Hamady M, Walker JJ, Harris JK, Gold NJ, Knight R (2008). Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* **5:** 235-237.

Han JI, Choi HK, Lee SW, Orwin PM, Kim J, Laroe SL *et al* (2011). Complete genome sequence of the metabolically versatile plant growth-promoting endophyte *Variovorax paradoxus* S110. *Journal of Bacteriology* **193:** 1183-1190.

Handelsman J (2004). Metagenomics: Application of genomics to uncultured microorganisms. *Microbiology and Molecular Biology Reviews* **68:** 669-685.

Hanson AD, Gage DA (1996). 3-dimethylsulfoniopropionate biosynthesis and use by flowering plants. *Biological and Environmental Chemistry of Dmsp and Related Sulfonium Compounds*. pp 75-86.

Hanson KG, Steinhauer K, Reizer J, Hillen W, Stulke J (2002). HPr kinase/phosphatase of *Bacillus subtilis*: expression of the gene and effects of mutations on enzyme activity, growth and carbon catabolite repression. *Microbiology-Sgm* **148:** 1805-1811.

Haralampidis K, Bryan G, Qi X, Papadopoulou K, Bakht S, Melton R *et al* (2001). A new class of oxidosqualene cyclases directs synthesis of antimicrobial phytoprotectants in monocots. *Proceedings of the National Academy of Sciences of the United States of America* **98:** 13431-13436.

Harrington AA, Kallio RE (1960). Oxidation of methanol and formaldehyde by *Pseudomonas methanica*. *Canadian Journal of Microbiology* **6:** 1-7.

Hassan S, Mathesius U (2012). The role of flavonoids in root-rhizosphere signalling: opportunities and challenges for improving plant-microbe interactions. *Journal of Experimental Botany* **63:** 3429-3444.

Haugo AJ, Watnick PI (2002). Vibrio cholerae CytR is a repressor of biofilm development. *Molecular Microbiology* **45:** 471-483.

He SM, Wurtzel O, Singh K, Froula JL, Yilmaz S, Tringe SG *et al* (2010). Validation of two ribosomal RNA removal methods for microbial metatranscriptomics. *Nature Methods* **7:** 807-812.

Heckman DS, Geiser DM, Eidell BR, Stauffer RL, Kardos NL, Hedges SB (2001). Molecular evidence for the early colonization of land by fungi and plants. *Science* **293:** 1129-1133.

Hein JW, Wolfe GV, Blee KA (2008). Comparison of rhizosphere bacterial communities in *Arabidopsis thaliana* mutants for systemic acquired resistance. *Microbial Ecology* **55:** 333-343.

Henry S, Texier S, Hallet S, Bru D, Dambreville C, Cheneby D *et al* (2008). Disentangling the rhizosphere effect on nitrate reducers and denitrifiers: insight into the role of root exudates. *Environmental Microbiology* **10:** 3082-3092.

Hershberger CD, Ye RW, Parsek MR, Xie ZD, Chakrabarty AM (1995). The algT (algU) gene of P*seudomonas aeruginosa*, a key regulator involved in alginate biosynthesis, encodes an alternative sigma-factor (sigma(e)). *Proceedings of the National Academy of Sciences of the United States of America* **92:** 7941-7945.

Heung LJ, Luberto C, Del Poeta M (2006). Role of sphingolipids in microbial pathogenesis. *Infection and Immunity* **74:** 28-39.

Hewson I, Poretsky RS, Dyhrman ST, Zielinski B, White AE, Tripp HJ *et al* (2009). Microbial community gene expression within colonies of the diazotroph, *Trichodesmium*, from the Southwest Pacific Ocean. *Isme Journal* **3:** 1286-1300.

Hodge A, Fitter AH (2010). Substantial nitrogen acquisition by arbuscular mycorrhizal fungi from organic material has implications for N cycling. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 13754-13759.

Holmes AJ, Roslev P, McDonald IR, Iversen N, Henriksen K, Murrell JC (1999). Characterization of methanotrophic bacterial populations in soils showing atmospheric methane uptake. *Applied and Environmental Microbiology* **65:** 3312-3318.

Holmstrom C, James S, Neilan BA, White DC, Kjelleberg S (1998). *Pseudoalteromonas tunicata* sp. nov., a bacterium that produces antifouling agents. *International Journal of Systematic Bacteriology* **48:** 1205-1212.

Hong SH, Bunge J, Leslin C, Jeon S, Epstein SS (2009). Polymerase chain reaction primers miss half of rRNA microbial diversity. *Isme Journal* **3:** 1365-1373.

Houlden A, Timms-Wilson TM, Day MJ, Bailey MJ (2008). Influence of plant developmental stage on microbial community structure and activity in the rhizosphere of three field crops. *Fems Microbiology Ecology* **65:** 193-201.

Hueck CJ (1998). Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiology and Molecular Biology Reviews* **62:** 379-443.

Hunt S, Layzell DB (1993). Gas-exchange of legume nodules and the regulation of nitrogenase activity. *Annual Review of Plant Physiology and Plant Molecular Biology* **44:** 483-511.

Hurek T, Handley LL, Reinhold-Hurek B, Piche Y (2002). *Azoarcus* grass endophytes contribute fixed nitrogen to the plant in an unculturable state. *Molecular Plant-Microbe Interactions* **15:** 233-242.

Huson DH, Auch AF, Qi J, Schuster SC (2007). MEGAN analysis of metagenomic data. *Genome Research* **17:** 377-386.

Hutchison CA, Venter JC (2006). Single-cell genomics. *Nature Biotechnology* **24:** 657-658.

Huang K, Brady A, Mahurkar A, White O, Gevers D, Huttenhower C, Segata N (2013). MetaRef: A pan-genomic datbase for comparative and community microbial genomics. Nucleic Acids Research 1-8.

Inceoglu O, Abu Al-Soud W, Salles JF, Semenov AV, van Elsas JD (2011). Comparative analysis of bacterial communities in a potato field as determined by pyrosequencing. *PLoS One* **6:** e23321.

Iniguez AL, Dong YM, Triplett EW (2004). Nitrogen fixation in wheat provided by *Klebsiella pneumoniae* 342. *Molecular Plant-Microbe Interactions* **17:** 1078-1085.

Janssen PH (2006). Identifying the dominant soil bacterial taxa in libraries of 16S rRNA and 16S rRNA genes. *Applied and Environmental Microbiology* **72:** 1719-1728.

Jiang CJ, Ma GF, Li SX, Hu TT, Che ZQ, Shen PH *et al* (2009). Characterization of a novel beta-glucosidase-like activity from a soil metagenome. *Journal of Microbiology* **47:** 542-548.

Jones DJ, Dangl JL (2006). The plant immune system. *Nature* **444:** 323-329.

Kalyuhznaya MG, Martens-Habbena W, Wang TS, Hackett M, Stolyar SM, Stahl DA *et al* (2009). Methylophilaceae link methanol oxidation to denitrification in freshwater lake sediment as suggested by stable isotope probing and pure culture analysis. *Environmental Microbiology Reports* **1:** 385-392.

Kanehisa M, Goto S (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **28:** 27-30.

Kannenberg EL, Poralla K (1999). Hopanoid biosynthesis and function in bacteria. *Naturwissenschaften* **86:** 168-176.

Kapranov P, St Laurent G, Raz T, Ozsolak F, Reynolds CP, Sorensen PHB *et al* (2010). The majority of total nuclear-encoded non-ribosomal RNA in a human cell is 'dark matter' un-annotated RNA. *BMC Biology* **8:** 149.

Karunakaran R, Ramachandran VK, Seaman JC, East AK, Mouhsine B, Mauchline TH *et al* (2009). Transcriptomic analysis of *Rhizobium leguminosarum* biovar viciae in symbiosis with host plants *Pisum sativum* and *Vicia cracca*. *Journal of Bacteriology* **191:** 4002-4014.

Kaur D, Brennan PJ, Crick DC (2004). Decaprenyl diphosphate synthesis in *Mycobacterium tuberculosis*. *Journal of Bacteriology* **186:** 7564-7570.

Kembel SW, Wu M, Eisen JA, Green JL (2012). Incorporating 16S gene copy number information improves estimates of microbial diversity and abundance. *PLoS Computational Biology* **8:** e1002743.

Kent AD, Triplett EW (2002). Microbial communities and their interactions in soil and rhizosphere ecosystems. *Annual Review of Microbiology* **56:** 211-236.

Kim BR, Nam HY, Kim SU, Kim SI, Chang YJ (2003). Normalization of reverse transcription quantitative-PCR with housekeeping genes in rice. *Biotechnology Letters* **25:** 1869-1872.

Kim J-S, Wang N (2009). Characterization of copy numbers of 16S rDNA and 16S rRNA of Candidatus *Liberibacter asiaticus* and the implication in detection in planta using quantitative PCR. *BMC research notes* **2:** 37.

Kim JM, Le NT, Chung BS, Park JH, Bae JW, Madsen EL *et al* (2008). Influence of soil components on the biodegradation of benzene, toluene, ethylbenzene, and o-, m-, and p-xylenes by the newly isolated bacterium *Pseudoxanthomonas spadix* BD-a59. *Applied and Environmental Microbiology* **74:** 7313-7320.

Klappenbach JA, Dunbar JM, Schmidt TM (2000). rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology* **66:** 1328-1333.

Klappenbach JA, Saxman PR, Cole JR, Schmidt TM (2001). rrndb: the Ribosomal RNA Operon Copy Number Database. *Nucleic Acids Research* **29:** 181-184.

Klein E, Ofek M, Katan J, Minz D, Gamliel A (2013). Soil suppressiveness to *Fusarium* disease: shifts in root microbiome associated with reduction of pathogen root colonization. *Phytopathology* **103:** 23-33.

Knee EM, Gong FC, Gao MS, Teplitski M, Jones AR, Foxworthy A *et al* (2001). Root mucilage from pea and its utilization by rhizosphere bacteria as a sole carbon source. *Molecular Plant-Microbe Interactions* **14:** 775-784.

Knief C, Delmotte N, Chaffron S, Stark M, Innerebner G, Wassmann R *et al* (2012). Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. *Isme Journal* **6:** 1378-1390.

Kniskern JM, Traw MB, Bergelson J (2007). Salicylic acid and jasmonic acid signaling defense pathways reduce natural bacterial diversity on *Arabidopsis thaliana*. *Molecular Plant-Microbe Interactions* **20:** 1512-1522.

Ko J, Lee Y, Park I, Cho B (2001). Identification of a structural motif of 23S rRNA interacting with 5S rRNA. *Febs Letters* **508:** 300-304.

Kobe B, Kajava AV (2001). The leucine-rich repeat as a protein recognition motif. *Current Opinion in Structural Biology* **11:** 725-732.

Koh YS, Roe JH (1995). Isolation of a novel paraquat-inducible (pqi) gene regulated by the soxRS locus in *Escherichia coli*. *Journal of Bacteriology* **177:** 2673-2678.

Kolb S (2009). Aerobic methanol-oxidizing Bacteria in soil. *Fems Microbiology Letters* **300:** 1-10.

Kopylova E, Noe L, Touzet H (2012). SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. *Bioinformatics* **28:** 3211-3217.

Kovacs S, Gaspar L, Cseh E, Kropf K, Sarvari E (2005). Protective effects of phosphonomethyl-sarcosine against the copper and cadmium induced inhibition of leaf development in poplar. *Acta Biologica Szegediensis* **49:** 61-63.

Krehenbrink M, Oppermann-Sanio FB, Steinbuchel A (2002). Evaluation of non-cyanobacterial genome sequences for occurrence of genes encoding proteins homologous to cyanophycin synthetase and cloning of an active cyanophycin synthetase from *Acinetobacter* sp strain DSM 587. *Archives of Microbiology* **177:** 371-380.

Krejcik Z, Denger K, Weinitschke S, Hollemeyer K, Paces V, Cook AM *et al* (2008). Sulfoacetate released during the assimilation of taurine-nitrogen by *Neptuniibacter caesariensis*: purification of sulfoacetaldehyde dehydrogenase. *Archives of Microbiology* **190:** 159-168.

Kuhm AE, Knackmuss H-J, Stolz A (1993). 2-Hydroxychromene-2-carboxylate isomerase from bacteria that degrade naphthalenesulfonates. *Biodegradation* **4:** 155-162.

Kuske CR, Ticknor LO, Miller ME, Dunbar JM, Davis JA, Barns SM *et al* (2002). Comparison of soil bacterial communities in rhizospheres of three plant species and the interspaces in an arid grassland. *Applied and Environmental Microbiology* **68:** 1854-1863.

Lahdesmaki P (1986). Determination of taurine and other acidic amino-acids in plants. *Phytochemistry* **25:** 2409-2411.

Lang E, Lapidus A, Chertkov O, Brettin T, Detter JC, Han C *et al* (2009). Complete genome sequence of *Dyadobacter fermentans* type strain (NS114(T)). *Standards in Genomic Sciences* **1:** 133-140.

Lebuhn M, Heulin T, Hartmann A (1997). Production of auxin and other indolic and phenolic compounds by *Paenibacillus polymyxa* strains isolated from different proximity to plant roots. *Fems Microbiology Ecology* **22:** 325-334.

Lee JH, Yi H, Chun J (2011). rRNASelector: A computer program for selecting ribosomal RNA encoding sequences from metagenomic and metatranscriptomic shotgun libraries. *Journal of Microbiology* **49:** 689-691.

Lee MH, Hong KS, Malhotra S, Park JH, Hwang EC, Choi HK *et al* (2010). A new esterase EstD2 isolated from plant rhizosphere soil metagenome. *Applied Microbiology and Biotechnology* **88:** 1125-1134.

Leininger S, Urich T, Schloter M, Schwark L, Qi J, Nicol GW *et al* (2006). Archaea predominate among ammonia-oxidizing prokaryotes in soils. *Nature* **442:** 806-809.

Lemanceau P, Bauer P, Kraemer S, Briat JF (2009). Iron dynamics in the rhizosphere as a case study for analyzing interactions between soils, plants and microbes. *Plant and Soil* **321:** 513-535.

Lemcke RM, Burrows MR (1980). Sterol requirement for the growth of *Treponema hyodysenteriae*. *Journal of General Microbiology* **116:** 539-543.

Lesniewski RA, Anantharaman K, Dick GJ (2010). Metatranscriptomic insights into the geomicrobiology of deep-sea hydrothermal plumes. *Geochimica Et Cosmochimica Acta* **74:** A583-A583.

Lesniewski RA, Jain S, Anantharaman K, Schloss PD, Dick GJ (2012). The metatranscriptome of a deep-sea hydrothermal plume is dominated by water column methanotrophs and lithotrophs. *Isme Journal* **6:** 2257-2268.

Lewis JA, Horswill AR, Schwem BE, Escalante-Semerena JC (2004). The tricarballylate utilization (tcuRABC) genes of *Salmonella enterica* serovar Typhimurium LT2. *Journal of Bacteriology* **186:** 1629-1637.

Lillis L, Doyle E, Clipson N (2009). Comparison of DNA- and RNA-based bacterial community structures in soil exposed to 2,4-dichlorophenol. *Journal of Applied Microbiology* **107:** 1883-1893.

Lin WW, Lu DP, Gao XQ, Jiang S, Ma XY, Wang ZH *et al* (2013). Inverse modulation of plant immune and brassinosteroid signaling pathways by the receptor-like cytoplasmic kinase BIK1. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 12114-12119.

Lindow SE (1996). Role of immigration and other processes in determining epiphytic bacterial populations - Implications for disease management. *Aerial Plant Surface Microbiology*. Plenum Press Div Plenum Publishing Corp: New York. 155-168.

Lindow SE, Brandl MT (2003). Microbiology of the phyllosphere. *Applied and Environmental Microbiology* **69:** 1875-1883.

Liu B, Gibbons T, Ghodsi M, Treangen T, Pop M (2010). Accurate and fast estimation of taxonomic profiles from metagenomic shotgun sequences. *BMC Genomics* **12:** S4.

Liu JW, Yap WH, Thanabalu T, Porter AG (1996). Efficient synthesis of mosquitocidal toxins in *Asticcacaulis excentricus* demonstrates potential of Gram-negative bacteria in mosquito control. *Nature Biotechnology* **14:** 343-347.

Liu WT, Marsh TL, Cheng H, Forney LJ (1997). Characterization of microbial diversity by determining terminal restriction fragment length polymorphisms of genes encoding 16S rRNA. *Applied and Environmental Microbiology* **63:** 4516-4522.

Loper JE, Henkels MD (1999). Utilization of heterologous siderophores enhances levels of iron available to *Pseudomonas putida* in the rhizosphere. *Applied and Environmental Microbiology* **65:** 5357-5363.

Lu YH, Rosencrantz D, Liesack W, Conrad R (2006). Structure and activity of bacterial community inhabiting rice roots and the rhizosphere. *Environmental Microbiology* **8:** 1351-1360.

Lugtenberg B, Kamilova F (2009). Plant-Growth-Promoting Rhizobacteria. *Annual Review of Microbiology*. Annual Reviews **63:** 541-556.

Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S *et al* (2012). Defining the core *Arabidopsis thaliana* root microbiome. *Nature* **488:** 86-90.

MacLeod JK, Flanigan IL, Williams JF, Collins JG (2001). Mass spectrometric studies of the path of carbon in photosynthesis: positional isotopic analysis of C-13-labelled C-4 to C-7 sugar phosphates. *Journal of Mass Spectrometry* **36:** 500-508.

Maimaiti J, Zhang Y, Yang J, Cen YP, Layzell DB, Peoples M *et al* (2007). Isolation and characterization of hydrogen-oxidizing bacteria induced following exposure of soil to hydrogen gas and their impact on plant growth. *Environmental Microbiology* **9:** 435-444.

Maizel JV, Mitchell HK, Burkhardt HJ (1964). Avenacin antimicrobial substance isolated from *Avena sativa* .1. Isolation and antimicrobial activity. *Biochemistry* **3:** 424-426.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA *et al* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437:** 376-380.

Mark GL, Dow JM, Kiely PD, Higgins H, Haynes J, Baysse C *et al* (2005). Transcriptome profiling of bacterial responses to root exudates identifies genes involved in microbe-plant interactions. *Proceedings of the National Academy of Sciences of the United States of America* **102:** 17454-17459.

Marschner (1995). *Mineral nutrition in higher plants*, 2nd edn. Academic Press: London.

Martinez I, Lattimer JM, Hubach KL, Case JA, Yang J, Weber CG *et al* (2012). Gut microbiome composition is linked to whole grain-induced immunological improvements. *Isme Journal* **7:** 269-280.

Mason OU, Hazen TC, Borglin S, Chain PSG, Dubinsky EA, Fortney JL *et al* (2012). Metagenome, metatranscriptome and single-cell sequencing reveal microbial response to Deepwater Horizon oil spill. *Isme Journal* **6:** 1715-1727.

Matilla MA, Espinosa-Urgel M, Rodriguez-Herva JJ, Ramos JL, Ramos-Gonzalez MI (2007). Genomic analysis reveals the major driving forces of bacterial life in the rhizosphere. *Genome Biology* **8:** R179.

Mattison RG, Harayama S (2001). The predatory soil flagellate *Heteromita globosa* stimulates toluene biodegradation by a *Pseudomonas* sp. *Fems Microbiology Letters* **194:** 39-45.

Matz C, Deines P, Boenigk J, Arndt H, Eberl L, Kjelleberg S *et al* (2004). Impact of violacein-producing bacteria on survival and feeding of bacterivorous nanoflagellates. *Applied and Environmental Microbiology* **70:** 1593-1599.

McCarren J, Becker JW, Repeta DJ, Shi YM, Young CR, Malmstrom RR *et al* (2010). Microbial community transcriptomes reveal microbes and metabolic pathways associated with dissolved organic matter turnover in the sea. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 16420-16427.

McDonald M, Mavrodi DV, Thomashow LS, Floss HG (2001). Phenazine biosynthesis in *Pseudomonas fluorescens*: Branchpoint from the primary shikimate biosynthetic pathway and role of phenazine-1,6-dicarboxylic acid. *Journal of the American Chemical Society* **123:** 9459-9460.

McGrath KC, Thomas-Hall SR, Cheng CT, Leo L, Alexa A, Schmidt S *et al* (2008). Isolation and analysis of mRNA from environmental microbial communities. *Journal of Microbiological Methods* **75:** 172-176.

McLean JS, Lombardo MJ, Badger JH, Edlund A, Novotny M, Yee-Greenbaum J *et al* (2013). Candidate phylum TM6 genome recovered from a hospital sink biofilm provides genomic insights into this uncultivated phylum. *Proceedings of the National Academy of Sciences of the United States of America* **110:** E2390-E2399.

Melotto M, Underwood W, Koczan J, Nomura K, He SY (2006). Plant stomata function in innate immunity against bacterial invasion. *Cell* **126:** 969-980.

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JHM *et al* (2011). Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332:** 1097-1100.

Mengoni A, Tatti E, Decorosi F, Viti C, Bazzicalupo M, Giovannetti L (2005). Comparison of 16S rRNA and 16S rDNA T-RFLP approaches to study bacterial communities in soil microcosms treated with chromate as perturbing agent. *Microbial Ecology* **50:** 375-384.

Meyer F, Paarmann D, D'Souza M, Olson R, Glass EM, Kubal M *et al* (2008). The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinformatics* **9:** 8.

Meyer SLF, Halbrendt JM, Carta LK, Skantar AM, Liu T, Abdelnabby HME *et al* (2009). Toxicity of 2,4-diacetylphloroglucinol (DAPG) to plant-parasitic and bacterial-feeding nematodes. *Journal of Nematology* **41:** 274-280.

Micallef SA, Shiaris MP, Colon-Carmona A (2009). Influence of *Arabidopsis thaliana* accessions on rhizobacterial communities and natural variation in root exudates. *Journal of Experimental Botany* **60:** 1729-1742.

Mirete S, de Figueras CG, Gonzalez-Pastor JE (2007). Novel nickel resistance genes from the rhizosphere metagenome of plants adapted to acid mine drainage. *Applied and Environmental Microbiology* **73:** 6001-6011.

Mitova M, Popov S, De Rosa S (2004). Cyclic peptides from a *Ruegeria* strain of bacteria associated with the sponge Suberites domuncula. *Journal of Natural Products* **67:** 1178-1181.

Mo B, Lian B (2011). Interactions between *Bacillus mucilaginosus* and silicate minerals (weathered adamellite and feldspar): Weathering rate, products, and reaction mechanisms. *Chinese Journal of Geochemistry* **30:** 187-192.

Moody SF, Clarke AE, Bacic A (1988). Structural analysis of secreted slime from wheat and cowpea roots. *Phytochemistry* **27:** 2857-2861.

Moran MA, Satinsky B, Gifford SM, Luo HW, Rivers A, Chan LK *et al* (2013). Sizing up metatranscriptomics. *Isme Journal* **7:** 237-243.

Morita M, Shibuya M, Kushiro T, Masuda K, Ebizuka Y (2000). Molecular cloning and functional expression of triterpene synthases from pea (*Pisum sativum*) - New alpha-amyrin-producing enzyme is a multifunctional triterpene synthase. *European Journal of Biochemistry* **267:** 3453-3460.

Morris SE, Turnbull CGN, Murfet IC, Beveridge CA (2001). Mutational analysis of branching in pea. Evidence that Rms1 and Rms5 regulate the same novel signal. *Plant Physiology* **126:** 1205-1213.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H (1986). Specific enzymatic amplification of DNA *in vitro* - the polymerase chain-reaction. *Cold Spring Harbor Symposia on Quantitative Biology* **51:** 263-273.

Muyzer G, Dewaal EC, Uitterlinden AG (1993). Profiling of complex microbial-populations by denaturing gradient gel-electrophoresis analysis of polymerase chain reaction-amplified genes-coding for 16S ribosomal-RNA. *Applied and Environmental Microbiology* **59:** 695-700.

Narberhaus F, Vogel J (2009). Regulatory RNAs in prokaryotes: here, there and everywhere. *Molecular Microbiology* **74:** 261-269.

Nawrocki EP, Kolbe DL, Eddy SR (2009). Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25:** 1335-1337.

Neal AL, Ahmad S, Gordon-Weeks R, Ton J (2012). Benzoxazinoids in root exudates of maize attract *Pseudomonas putida* to the rhizosphere. *PLoS One* **7:** e35498.

Neidhardt FC, Umbarger H (1996). Chemical Composition of *Escherichia coli*. In: Ingraham JL (ed). *Escherichia coli and Salmonella typhimurium* ASM Press: Washington D.C. p 13.

Newton GL, Buchmeier N, Fahey RC (2008). Biosynthesis and functions of mycothiol, the unique protective thiol of Actinobacteria. *Microbiology and Molecular Biology Reviews* **72:** 471-494.

Nguyen HTT, Le VQ, Hansen AA, Nielsen JL, Nielsen PH (2011). High diversity and abundance of putative polyphosphate-accumulating *Tetrasphaera*-related bacteria in activated sludge systems. *Fems Microbiology Ecology* **76:** 256-67.

Norin A, VanOphem PW, Piersma SR, Persson B, Duine JA, Jornvall H (1997). Mycothiol-dependent formaldehyde dehydrogenase, a prokaryotic medium-chain dehydrogenase/reductase, phylogenetically links different eukaryotic alcohol dehydrogenases - Primary structure, conformational modelling and functional correlations. *European Journal of Biochemistry* **248:** 282-289.

Nuccio EE, Hodge A, Pett-Ridge J, Herman DJ, Weber PK, Firestone MK (2013). An arbuscular mycorrhizal fungus significantly modifies the soil bacterial community and nitrogen cycling during litter decomposition. *Environmental Microbiology* **15:** 1870-1881.

Ofek M, Hadar Y, Minz D (2012). Ecology of Root Colonizing *Massilia* (Oxalobacteraceae). *PLoS One* **7:** e40117.

Ohana P, Delmer DP, Carlson RW, Glushka J, Azadi P, Bacic T *et al* (1998). Identification of a novel triterpenoid saponin from *Pisum sativum* as a specific inhibitor of the diguanylate cyclase of *Acetobacter xylinum*. *Plant and Cell Physiology* **39:** 144-152.

Oldroyd GED, Murray JD, Poole PS, Downie JA (2011). The rules of engagement in the Legume-Rhizobial symbiosis. *Annual Review Genetics, Vol 45*. Annual Reviews **45:** 119-144.

Oresnik IJ, Pacarynuk LA, O'Brien SAP, Yost CK, Hynes MF (1998). Plasmid-encoded catabolic genes in *Rhizobium leguminosarum* bv. trifolii: Evidence for a plant-inducible rhamnose locus involved in competition for nodulation. *Molecular Plant-Microbe Interactions* **11:** 1175-1185.

Osbourn A, Bowyer P, Lunness P, Clarke B, Daniels M (1995). Fungal pathogens of oat roots and tomato leaves employ closely related enzymes to detoxify different host plant saponins. *Molecular Plant-Microbe Interactions* **8:** 971-978.

Osbourn AE, Clarke BR, Dow JM, Daniels MJ (1991). Partial characterization of avenacinase from *Gaeumannomyces graminis* var avenae. *Physiological and Molecular Plant Pathology* **38:** 301-312.

Osbourn AE, Clarke BR, Lunness P, Scott PR, Daniels MJ (1994). An oat species lacking avenacin is susceptible to infection by *Gaeumannomyces graminis* var tritici. *Physiological and Molecular Plant Pathology* **45:** 457-467.

Ottesen EA, Marin R, Preston CM, Young CR, Ryan JP, Scholin CA *et al* (2011). Metatranscriptomic analysis of autonomously collected and preserved marine bacterioplankton. *Isme Journal* **5:** 1881-1895.

Overbeek R, Begley T, Butler RM, Choudhuri JV, Chuang HY, Cohoon M *et al* (2005). The subsystems approach to genome annotation and its use in the project to annotate 1000 genomes. *Nucleic Acids Research* **33:** 5691-5702.

Pagani H, Parenti F (1978). *Kineosporia*, a new genus of order Actinomycetales. *International Journal of Systematic Bacteriology* **28:** 401-406.

Pantanella F, Berlutti F, Passariello C, Sarli S, Morea C, Schippa S (2007). Violacein and biofilm production in *Janthinobacterium lividum*. *Journal of Applied Microbiology* **102:** 992-999.

Papadopoulou K, Melton RE, Leggett M, Daniels MJ, Osbourn AE (1999). Compromised disease resistance in saponin-deficient plants. *Proceedings of the National Academy of Sciences of the United States of America* **96:** 12923-12928.

Parales RE, Harwood CS (1993). Regulation of the pcaIJ genes for aromatic acid degradation in *Pseudomonas putida*. *Journal of Bacteriology* **175:** 5829-5838.

Parks DH, Beiko RG (2010). Identifying biologically relevant differences between metagenomic communities. *Bioinformatics* **26:** 715-721.

Paulsen IT, Press CM, Ravel J, Kobayashi DY, Myers GSA, Mavrodi DV *et al* (2005). Complete genome sequence of the plant commensal *Pseudomonas fluorescens* Pf-5. *Nature Biotechnology* **23:** 873-878.

Paungfoo-Lonhienne C, Rentsch D, Robatzek S, Webb RI, Sagulenko E, Nasholm T *et al* (2010). Turning the table: plants consume microbes as a source of nutrients. *PLoS One* **5:** e11915.

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL *et al* (2013). Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 6548-6553.

Peleg AY, Seifert H, Paterson DL (2008). Acinetobacter baumannii: Emergence of a successful pathogen. *Clinical Microbiology Reviews* **21:** 538-582.

Perrin DR, Bottomley W (1961). Pisatin - antifungal substance from *Pisum sativum* L. *Nature* **191:** 76-77.

Philippot L, Hallin S, Borjesson G, Baggs EM (2009). Biochemical cycling in the rhizosphere having an impact on global change. *Plant and Soil* **321:** 61-81.

Picardeau M, Bulach DM, Bouchier C, Zuerner RL, Zidane N, Wilson PJ *et al* (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of leptospira and the pathogenesis of leptospirosis. *PLoS One* **3:** e1607.

Pinhassi J, Bowman JP, Nedashkovskaya OI, Lekunberri I, Gomez-Consarnau L, Pedros-Alio C (2006). *Leeuwenhoekiella blandensis* sp nov., a genome-sequenced marine member of the family Flavobacteriaceae. *International Journal of Systematic and Evolutionary Microbiology* **56:** 1489-1493.

Pinto AJ, Raskin L (2012). PCR biases distort bacterial and archaeal community structure in pyrosequencing datasets. *PLoS One* **7:** e43093.

Ponten TS (2011). Metatranscriptomics of the human gut microbiome. *Genome Biology* **12:** 3-3.

Poralla K, Muth G, Hartner T (2000). Hopanoids are formed during transition from substrate to aerial hyphae in Streptomyces coelicolor A3(2). *Fems Microbiology Letters* **189:** 93-95.

Poretsky RS, Hewson I, Sun SL, Allen AE, Zehr JP, Moran MA (2009). Comparative day/night metatranscriptomic analysis of microbial communities in the North Pacific subtropical gyre. *Environmental Microbiology* **11:** 1358-1375.

Postma J, Schilder MT, Bloem J, van Leeumen-Haagsma WK (2008). Soil suppressiveness and functional diversity of the soil microflora in organic farming systems. *Soil Biology & Biochemistry* **40:** 2394-2406.

Pradella S, Hans A, Sproer C, Reichenbach H, Gerth K, Beyer S (2002). Characterisation, genome size and genetic manipulation of the myxobacterium *Sorangium cellulosum* So ce56. *Archives of Microbiology* **178:** 484-492.

Prokopowich CD, Gregory TR, Crease TJ (2003). The correlation between rDNA copy number and genome size in eukaryotes. *Genome* **46:** 48-50.

Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig WG, Peplies J *et al* (2007). SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Research* **35:** 7188-7196.

Qin B, Eagles J, Mellon FA, Mylona P, Pena-Rodriguez L, Osbourn AE (2010). High throughput screening of mutants of oat that are defective in triterpene synthesis. *Phytochemistry* **71:** 1245-1252.

Qiu XB, Shao YM, Miao S, Wang L (2006). The diversity of the DnaJ/Hsp40 family, the crucial partners for Hsp70 chaperones. *Cellular and Molecular Life Sciences* **63:** 2560-2570.

Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R *et al* (2005). InterProScan: protein domains identifier. *Nucleic Acids Research* **33:** W116-W120.

Raaijmakers JM, Weller DM (1998). Natural plant protection by 2,4-diacetylphloroglucinol - Producing *Pseudomonas* spp. in take-all decline soils. *Molecular Plant-Microbe Interactions* **11:** 144-152.

Radajewski S, Ineson P, Parekh NR, Murrell JC (2000). Stable-isotope probing as a tool in microbial ecology. *Nature* **403:** 646-649.

Ramachandran VK, East AK, Ramakrishnan K, Downie JA, Poole PS (2011). Adaptation of *Rhizobium leguminosarum* to pea, alfalfa and sugar beet rhizospheres investigated by comparative transcriptomics. *Genome Biology* **12:** R106.

Ransom-Jones E, Jones DL, McCarthy AJ, McDonald JE (2012). The Fibrobacteres: an important phylum of cellulose-degrading bacteria. *Microbial Ecology* **63:** 267-281.

Rausch C, Daram P, Brunner S, Jansa J, Laloi M, Leggewie G *et al* (2001). A phosphate transporter expressed in arbuscule-containing cells in potato. *Nature* **414:** 462-466.

Raychoudhury R, Sen R, Cai Y, Sun Y, Lietze VU, Boucias DG *et al* (2011). Comparative metatranscriptomic signatures of wood and paper feeding in the gut of the termite *Reticulitermes flavipes* (Isoptera: Rhinotermitidae). *Insect Molecular Biology* **22:** 155-171.

Rezzonico F, Binder C, Defago G, Moenne-Loccoz Y (2005). The type III secretion system of biocontrol *Pseudomonas fluorescens* KD targets the phytopathogenic chromista *Pythium ultimum* and promotes cucumber protection. *Molecular Plant-Microbe Interactions* **18:** 991-1001.

Rhew RC, Ostergaard L, Saltzman ES, Yanofsky MF (2003). Genetic control of methyl halide production in *Arabidopsis*. *Current Biology* **13:** 1809-1813.

Rodriguez H, Fraga R (1999). Phosphate solubilizing bacteria and their role in plant growth promotion. *Biotechnology Advances* **17:** 319-339.

Rodriguez H, Fraga R, Gonzalez T, Bashan Y (2006). Genetics of phosphate solubilization and its potential applications for improving plant growth-promoting bacteria. *Plant and Soil* **287:** 15-21.

Roesch LF, Fulthorpe RR, Riva A, Casella G, Hadwin AKM, Kent AD *et al* (2007). Pyrosequencing enumerates and contrasts soil microbial diversity. *Isme Journal* **1:** 283-290.

Rokhbakhsh-Zamin F, Sachdev D, Kazemi-Pour N, Engineer A, Pardesi KR, Zinjarde S *et al* (2011). Characterization of plant-growth-promoting traits of *Acinetobacter* species isolated from rhizosphere of *Pennisetum glaucum*. *Journal of Microbiology and Biotechnology* **21:** 556-566.

Rondon MR, August PR, Bettermann AD, Brady SF, Grossman TH, Liles MR *et al* (2000). Cloning the soil metagenome: a strategy for accessing the genetic and functional diversity of uncultured microorganisms. *Applied and Environmental Microbiology* **66:** 2541-2547.

Rontein D, Nishida I, Tashiro G, Yoshioka K, Wu WI, Voelker DR *et al* (2001). Plants synthesize ethanolamine by direct decarboxylation of serine using a pyridoxal phosphate enzyme. *Journal of Biological Chemistry* **276:** 35523-35529.

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 Journal* **3:** 675-684.

Rosenzweig N, Tiedje JM, Quensen JF, Meng QX, Hao JJJ (2012). Microbial communities associated with potato common scab-suppressive soil determined by pyrosequencing analyses. *Plant Disease* **96:** 718-725.

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M *et al* (2011). An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475:** 348-352.

Roux M, Schwessinger B, Albrecht C, Chinchilla D, Jones A, Holton N *et al* (2011). The *Arabidopsis* leucine-rich repeat receptor-like kinases BAK1/SERK3 and BKK1/SERK4 are required for innate immunity to hemibiotrophic and biotrophic pathogens. *Plant Cell* **23:** 2440-2455.

Ruiz N (2008). Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America* **105:** 15553-15557.

Ryals JA, Neuenschwander UH, Willits MG, Molina A, Steiner HY, Hunt MD (1996). Systemic acquired resistance. *Plant Cell* **8:** 1809-1819.

Ryan RP, Monchy S, Cardinale M, Taghavi S, Crossman L, Avison MB *et al* (2009). The versatility and adaptation of bacteria from the genus Stenotrophomonas. *Nature Reviews Microbiology* **7:** 514-525.

Sang MK, Kim KD (2012). The volatile-producing *Flavobacterium johnsoniae* strain GSE09 shows biocontrol activity against *Phytophthora capsici* in pepper. *Journal of Applied Microbiology* **113:** 383-398.

Sanger F, Nicklen S, Coulson AR (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America* **74:** 5463-5467.

Sangkhobol V, Skerman VBD (1981). *Chitinophaga*, a new genus of chitinolytic myxobacteria. *International Journal of Systematic Bacteriology* **31:** 285-293.

Saracchi M, Quaroni S, Corbetta C (2004). Variability on *Catellatospora* spp. strains isolated from plant materials. *Annals of Microbiology* **54:** 13-24.

Sass A, Marchbank A, Tullis E, LiPuma JJ, Mahenthiralingam E (2011). Spontaneous and evolutionary changes in the antibiotic resistance of *Burkholderia cenocepacia* observed by global gene expression analysis. *BMC Genomics* **12:** 373.

Schenk A, Weingart H, Ullrich MS (2008). The alternative sigma factor algt, but not alginate synthesis, promotes in planta multiplication of *Pseudomonas syringae* pv. Glycinea. *Microbiology-Sgm* **154:** 413-421.

Schmerk CL, Bernards MA, Valvano MA (2011). Hopanoid production is required forlow-pH tolerance, antimicrobial resistance, and motility in *Burkholderia cenocepacia*. *Journal of Bacteriology* **193:** 6712-6723.

Schmieder R, Lim YW, Edwards R (2012). Identification and removal of ribosomal RNA sequences from metatranscriptomes. *Bioinformatics* **28:** 433-435.

Segata N, Waldron L, Ballarini A, Narasimhan V, Jousson O, Huttenhower C (2012). Metagenomic microbial community profiling using unique clade-specific marker genes. *Nature Methods* **9:** 811-814.

Seipke RF, Loria R (2009). Hopanoids Are Not Essential for Growth of *Streptomyces scabies* 87-22. *Journal of Bacteriology* **191:** 5216-5223.

Semple KT, Cain RB (1996). Biodegradation of phenols by the alga *Ochromonas danica*. *Applied and Environmental Microbiology* **62:** 1265-1273.

Sessitsch A, Coenye T, Sturz AV, Vandamme P, Barka EA, Salles JF *et al* (2005). *Burkholderia phytofirmans* sp. nov., a novel plant-associated bacterium with plant-beneficial properties. *International Journal of Systematic and Evolutionary Microbiology* **55:** 1187-1192.

Sessitsch A, Reiter B, Pfeifer U, Wilhelm E (2002). Cultivation-independent population analysis of bacterial endophytes in three potato varieties based on eubacterial and Actinomycetes-specific PCR of 16S rRNA genes. *Fems Microbiology Ecology* **39:** 23-32.

Seymour M, Kirkegaard JA, Peoples MB, White PF, French RJ (2012). Break-crop benefits to wheat in Western Australia - insights from over three decades of research. *Crop & Pasture Science* **63:** 1-16.

Shagina I, Bogdanova E, Mamedov IZ, Lebedev Y, Lukyanov S, Shagin D (2010). Normalization of genomic DNA using duplex-specific nuclease. *Biotechniques* **48:** 455-459.

Shakya M, Quince C, Campbell JH, Yang ZMK, Schadt CW, Podar M (2013). Comparative metagenomic and rRNA microbial diversity characterization using archaeal and bacterial synthetic communities. *Environmental Microbiology* **15:** 1882-1899.

Sharma S, Aneja MK, Mayer J, Munch JC, Schloter M (2005). Diversity of transcripts of nitrite reductase genes (nirK and nirS) in rhizospheres of grain legumes. *Applied and Environmental Microbiology* **71:** 2001-2007.

Shi SJ, Richardson AE, O'Callaghan M, DeAngelis KM, Jones EE, Stewart A *et al* (2011). Effects of selected root exudate components on soil bacterial communities. *Fems Microbiology Ecology* **77:** 600-610.

Shi Y, Tyson GW, DeLong EF (2009a). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-269.

Shi Y, Tyson GW, Eppley JM, DeLong EF (2010). Integrated metatranscriptomic and metagenomic analyses of stratified microbial assemblages in the open ocean. *Isme Journal* **5:** 999-1013.

Shi YM, Tyson GW, DeLong EF (2009b). Metatranscriptomics reveals unique microbial small RNAs in the ocean's water column. *Nature* **459:** 266-U154.

Shrestha M, Shrestha PM, Frenzel P, Conrad R (2010). Effect of nitrogen fertilization on methane oxidation, abundance, community structure, and gene expression of methanotrophs in the rice rhizosphere. *Isme Journal* **4:** 1545-1556.

Shrestha PM, Kube M, Reinhardt R, Liesack W (2009). Transcriptional activity of paddy soil bacterial communities. *Environmental Microbiology* **11:** 960-970.

Singh BK, Bardgett RD, Smith P, Reay DS (2010). Microorganisms and climate change: terrestrial feedbacks and mitigation options. *Nature Reviews Microbiology* **8:** 779-790.

Smith MD, Angus BJ, Wuthiekanun V, White NJ (1997). Arabinose assimilation defines a nonvirulent biotype of *Burkholderia pseudomalle*i. *Infection and Immunity* **65:** 4319-4321.

Smith SE, Smith FA (2011). Roles of arbuscular mycorrhizas in plant nutrition and growth: new paradigms from cellular to ecosystem scales. *Annual Review of Plant Biology, Vol 62*. Annual Reviews **62:** 227-250.

Spoel SH, Dong XN (2012). How do plants achieve immunity? Defence without specialized immune cells. *Nature Reviews Immunology* **12:** 89-100.

Spor A, Koren O, Ley R (2011). Unravelling the effects of the environment and host genotype on the gut microbiome. *Nature Reviews Microbiology* **9:** 279-290.

Stacheter A, Noll M, Lee CK, Selzer M, Glowik B, Ebertsch L *et al* (2012). Methanol oxidation by temperate soils and environmental determinants of associated methylotrophs. *Isme Journal* **7:** 1051-1064.

Stafford HA (1974). METABOLISM OF AROMATIC-COMPOUNDS. *Annual Review of Plant Physiology and Plant Molecular Biology* **25:** 459-486.

Staley JT, Konopka A (1985). Measurement of *in situ* activities of non-photosynthetic microorganisms in aquatic and terrestrial habitats. *Annual Review of Microbiology* **39:** 321-346.

Stewart EJ (2012). Growing unculturable bacteria. *Journal of Bacteriology* **194:** 4151-4160.

Stewart FJ, Ottesen EA, DeLong EF (2010a). Development and quantitative analyses of a universal rRNA-subtraction protocol for microbial metatranscriptomics. *Isme Journal* **4:** 896-907.

Stewart FJ, Ulloa O, DeLong EF (2010b). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14:** 23-40.

Stewart FJ, Ulloa O, DeLong EF (2012). Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* **14:** 23-40.

Storey R, Jones RGW (1975). Betaine and choline levels in plants and their relationship to NaCl stress. *Plant Science Letters* **4:** 161-168.

Stracke S, Kistner C, Yoshida S, Mulder L, Sato S, Kaneko T *et al* (2002). A plant receptor-like kinase required for both bacterial and fungal symbiosis. *Nature* **417:** 959-962.

Stursova M, Zifcakova L, Leigh MB, Burgess R, Baldrian P (2012). Cellulose utilization in forest litter and soil: identification of bacterial and fungal decomposers. *Fems Microbiology Ecology* **80:** 735-746.

Sztukowska M, Bugno M, Potempa J, Travis J, Kurtz DM (2002). Role of rubrerythrin in the oxidative stress response of *Porphyromonas gingivalis*. *Molecular Microbiology* **44:** 479-488.

Takasaki K, Miura T, Kanno M, Tamaki H, Hanada S, Kamagata Y *et al* (2013). Discovery of glycoside hydrolase enzymes in an avicel adapted forest soil fungal community by a metatranscriptomic approach. *PLoS One* **8**.

Tart AH, Wolfgang MC, Wozniak DJ (2005). The alternative sigma factor AlgT represses *Pseudomonas aeruginosa* flagellum biosynthesis by inhibiting expression of fleQ. *Journal of Bacteriology* **187:** 7955-7962.

Tartar A, Wheeler MM, Zhou X, Coy MR, Boucias DG, Scharf ME (2009). Parallel metatranscriptome analyses of host and symbiont gene expression in the gut of the termite *Reticulitermes flavipes*. *Biotechnology for Biofuels* **2:** 25.

Teixeira L, Peixoto RS, Cury JC, Sul WJ, Pellizari VH, Tiedje J *et al* (2010). Bacterial diversity in rhizosphere soil from Antarctic vascular plants of Admiralty Bay, maritime Antarctica. *Isme Journal* **4:** 989-1001.

Temperton B, Field D, Oliver A, Tiwari B, Muhling M, Joint I *et al* (2009). Bias in assessments of marine microbial biodiversity in fosmid libraries as evaluated by pyrosequencing. *Isme Journal* **3:** 792-796.

Tett AJ, Turner TR, Poole PS (2012). Genomics and the rhizosphere. *eLS*. John Wiley & Sons, Ltd.

Thomas T, Evans FF, Schleheck D, Mai-Prochnow A, Burke C, Penesyan A *et al* (2008). Analysis of the *Pseudoalteromonas tunicata* genome reveals properties of a surface-associated life style in the marine environment. *PLoS One* **3:** e3252.

Timmusk S, Wagner EGH (1999). The plant-growth-promoting rhizobacterium *Paenibacillus polymyxa* induces changes in *Arabidopsis thaliana* gene expression: A possible connection between biotic and abiotic stress responses. *Molecular Plant-Microbe Interactions* **12:** 951-959.

Tkacz A, Chandra G, Fan J, Poole P (2013a). *Arabidopsis* glucosinolates alters the rhizosphere microbiome and suppress soil-borne *Cryptococcus spp. In preparation*.

Tkacz A, Cheema J, Chandra G, Boutrot F, Zipfel C, Grant A *et al* (2013b). Plant type, immunity and soil composition control the rhizosphere microbiome. *In preparation*.

Todd JD, Curson ARJ, Sullivan MJ, Kirkwood M, Johnston AWB (2012). The *Ruegeria pomeroyi* acuI gene has a role in DMSP catabolism and resembles yhdH of *E. coli* and other bacteria in conferring resistance to acrylate. *PLoS One* **7:** e35947.

Tokala RK, Strap JL, Jung CM, Crawford DL, Salove MH, Deobald LA *et al* (2002). Novel plant-microbe rhizosphere interaction involving *Streptomyces lydicus* WYEC108 and the pea plant (*Pisum sativum*). *Applied and Environmental Microbiology* **68:** 2161-2171.

Toljander JF, Lindahl BD, Paul LR, Elfstrand M, Finlay RD (2007). Influence of arbuscular mycorrhizal mycelial exudates on soil bacterial growth and community structure. *Fems Microbiology Ecology* **61:** 295-304.

Torsvik V, Goksoyr J, Daae FL (1990). High diversity in DNA of soil bacteria. *Applied and Environmental Microbiology* **56:** 782-787.

Torsvik V, Ovreas L (2002). Microbial diversity and function in soil: from genes to ecosystems. *Current Opinion in Microbiology* **5:** 240-245.

Toyama T, Furukawa T, Maeda N, Inoue D, Sei K, Mori K *et al* (2011). Accelerated biodegradation of pyrene and benzo[a]pyrene in the *Phragmites australis* rhizosphere by bacteria-root exudate interactions. *Water Research* **45:** 1629-1638.

Trojanowska MR, Osbourn AE, Daniels MJ, Threlfall DR (2000). Biosynthesis of avenacins and phytosterols in roots of *Avena sativa* cv. Image. *Phytochemistry* **54:** 153-164.

Tucker NP, Le Brun NE, Dixon R, Hutchings MI (2010). There's NO stopping NsrR, a global regulator of the bacterial NO stress response. *Trends in Microbiology* **18:** 149-156.

Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI (2007). The Human Microbiome Project. *Nature* **449:** 804-810.

Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI (2009). The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice. *Science Translational Medicine* **1:** 14.

Turner EM (1953). The nature of the resistance of oats to the take-all fungus. *Journal of Experimental Botany* **4:** 264-271.

Turner JM, Messenger AJ (1986). Occurrence, biochemistry and physiology of phenazine pigment production. *Advances in Microbial Physiology* **27:** 211-275.

Turner T, Ramakrishnan K, Walshaw J, Heavens D, Alston M, Swarbreck D *et al* (2013). Comparative metatranscriptomics reveals kingdom level changes in the rhizosphere microbiome of plants *Isme* Journal. *Epub ahead of print.*

Tyson GW, Chapman J, Hugenholtz P, Allen EE, Ram RJ, Richardson PM *et al* (2004). Community structure and metabolism through reconstruction of microbial genomes from the environment. *Nature* **428:** 37-43.

Ueda T, Suga Y, Yahiro N, Matsuguchi T (1995). Remarkable N_2-fixing bacterial diversity detected in rice roots by molecular evolutionary analysis of nifH gene-sequences. *Journal of Bacteriology* **177:** 1414-1417.

Urich T, Lanzen A, Qi J, Huson DH, Schleper C, Schuster SC (2008). Simultaneous assessment of soil microbial community structure and function through analysis of the meta-transcriptome. *PLoS One* **3:** e2527.

Uroz S, Buee M, Murat C, Frey-Klett P, Martin F (2010). Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environmental Microbiology Reports* **2:** 281-288.

Ursell LK, Knight R (2013). Xenobiotics and the human gut microbiome: metatranscriptomics reveal the active players. *Cell metabolism* **17:** 317-8.

Vale M, Nguyen C, Dambrine E, Dupouey JL (2005). Microbial activity in the rhizosphere soil of six herbaceous species cultivated in a greenhouse is correlated with shoot biomass and root C concentrations. *Soil Biology & Biochemistry* **37:** 2329-2333.

van Hannen EJ, van Agterveld MP, Gons HJ, Laanbroek HJ (1998). Revealing genetic diversity of eukaryotic microorganisms in aquatic environments by denaturing gradient gel electrophoresis. *Journal of Phycology* **34:** 206-213.

Vanschie BJ, Demooy OH, Linton JD, Vandijken JP, Kuenen JG (1987). PQQ-dependent production of gluconic acid by *Acinetobacter*, *Agrobacterium* and *Rhizobium* species. *Journal of General Microbiology* **133:** 867-875.

Venter JC, Remington K, Heidelberg JF, Halpern AL, Rusch D, Eisen JA *et al* (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science* **304:** 66-74.

Vila-Costa M, Sharma S, Moran MA, Casamayor EO (2013). Diel gene expression profiles of a phosphorus limited mountain lake using metatranscriptomics. *Environmental Microbiology* **15:** 1190-203.

Vimr ER, Kalivoda KA, Deszo EL, Steenbergen SM (2004). Diversity of microbial sialic acid metabolism. *Microbiology and Molecular Biology Reviews* **68:** 132-153.

Vincken JP, Heng L, de Groot A, Gruppen H (2007). Saponins, classification and occurrence in the plant kingdom. *Phytochemistry* **68:** 275-297.

Vokou D, Vareli K, Zarali E, Karamanoli K, Constantinidou HIA, Monokrousos N *et al* (2012). Exploring biodiversity in the bacterial community of the Mediterranean phyllosphere and its relationship with airborne bacteria. *Microbial Ecology* **64:** 714-724.

Vorholt JA (2012). Microbial life in the phyllosphere. *Nature Reviews Microbiology* **10:** 828-840.

Vranova V, Rejsek K, Skene KR, Janous D, Formanek P (2013). Methods of collection of plant root exudates in relation to plant metabolism and purpose: A review. *Journal of Plant Nutrition and Soil Science* **176:** 175-199.

Wang ET, Schornack S, Marsh JF, Gobbato E, Schwessinger B, Eastmond P *et al* (2012a). A common signaling process that promotes mycorrhizal and oomycete colonization of plants. *Current Biology* **22:** 2242-2246.

Wang KY, Shallcross DE (2000). Modelling terrestrial biogenic isoprene fluxes and their potential impact on global chemical species using a coupled LSM-CTM model. *Atmospheric Environment* **34:** 2909-2925.

Wang Y, Hayatsu M, Fujii T (2012b). Extraction of bacterial RNA from soil: challenges and Solutions. *Microbes and Environments* **27:** 111-121.

Weber J, Kayser A, Rinas U (2005). Metabolic flux analysis of *Escherichia coli* in glucose-limited continuous culture. II. Dynamic response to famine and feast, activation of the methylglyoxal pathway and oscillatory behaviour. *Microbiology-Sgm* **151:** 707-716.

Weber S, Stubner S, Conrad R (2001). Bacterial populations colonizing and degrading rice straw in anoxic paddy soil. *Applied and Environmental Microbiology* **67:** 1318-1327.

Weinberg Z, Perreault J, Meyer MM, Breaker RR (2009). Exceptional structured noncoding RNAs revealed by bacterial metagenome analysis. *Nature* **462:** 656-659.

Wendehenne D, Pugin A, Klessig DF, Durner J (2001). Nitric oxide: comparative synthesis and signaling in animal and plant cells. *Trends in Plant Science* **6:** 177-183.

Wilderman PR, Xu MM, Jin YH, Coates RM, Peters RJ (2004). Identification of syn-pimara-7,15-diene synthase reveals functional clustering of terpene synthases involved in rice phytoalexin/allelochemical biosynthesis. *Plant Physiology* **135:** 2098-2105.

Wildermuth MC (2006). Variations on a theme: synthesis and modification of plant benzoic acids. *Current Opinion in Plant Biology* **9:** 288-296.

Wilks SS (1959). Carbon monoxide in green plants. *Science* **129:** 964-966.

Williams SCP (2013). The other microbiome. *Proceedings of the National Academy of Sciences of the United States of America* **110:** 2682-2684.

Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW *et al* (2010). Gut inflammation provides a respiratory electron acceptor for *Salmonella*. *Nature* **467:** 426-429.

Witte CP (2011). Urea metabolism in plants. *Plant Science* **180:** 431-438.

Wittich RM, Wilkes H, Sinnwell V, Francke W, Fortnagel P (1992). Metabolism of dibenzo-para-dioxin by *Sphingomonas* sp strain-rw1. *Applied and Environmental Microbiology* **58:** 1005-1010.

Woese CR (1987). Bacterial evolution. *Microbiological Reviews* **51:** 221-271.

Woese CR, Kandler O, Wheelis ML (1990). Towards a natural system of organisms - proposal for the domains Archaea, Bacteria, and Eucarya. *Proceedings of the National Academy of Sciences of the United States of America* **87:** 4576-4579.

Wortham BW, Patel CN, Oliveira MA (2007). Polyamines in bacteria: Plelotropic effects yet specific mechanisms **603:** 106-115.

Wrage N, Velthof GL, van Beusichem ML, Oenema O (2001). Role of nitrifier denitrification in the production of nitrous oxide. *Soil Biology & Biochemistry* **33:** 1723-1732.

Wright SF, Upadhyaya A (1996). Extraction of an abundant and unusual protein from soil and comparison with hyphal protein of arbuscular mycorrhizal fungi. *Soil Science* **161:** 575-586.

Wu DQ, Li YQ, Xu YQ (2012). Comparative analysis of temperature-dependent transcriptome of *Pseudomonas aeruginosa* strains from rhizosphere and human habitats. *Applied Microbiology and Biotechnology* **96:** 1007-1019.

Wu M, Eisen JA (2008). A simple, fast, and accurate method of phylogenomic inference. *Genome Biology* **9:** R151.

Wu XL, Prior RL (2005). Identification and characterization of anthocyanins by high-performance liquid chromatography-electrospray ionization-tandem mass spectrometry in common foods in the United States: Vegetables, nuts, and grains. *Journal of Agricultural and Food Chemistry* **53:** 3101-3113.

Xie F, Williams A, Edwards A, Downie JA (2012). A plant arabinogalactan-like glycoprotein promotes a novel type of polar surface attachment by *Rhizobium leguminosarum*. *Molecular Plant-Microbe Interactions* **25:** 250-258.

Xiong X, Frank DN, Robertson CE, Hung SS, Markle J, Canty AJ *et al* (2012). Generation and analysis of a mouse intestinal metatranscriptome through Illumina based RNA-sequencing. *PLoS One* **7:** e36009.

Yamamoto R (1987). Effect of galactose on auxin-induced cell elongation in oat coleoptile segments in mannitol solutions. *Botanical Magazine-Tokyo* **100:** 43-49.

Yang RX, Luo CL, Chen YH, Wang GP, Xu Y, Shen ZG (2013). Copper-resistant bacteria enhance plant growth and copper phytoextraction. *International Journal of Phytoremediation* **15:** 573-584.

Yao J, Allen C (2006). Chemotaxis is required for virulence and competitive fitness of the bacterial wilt pathogen *Ralstonia solanacearum*. *Journal of Bacteriology* **188:** 3697-3708.

Yashiro E, Spear RN, McManus PS (2011). Culture-dependent and culture-independent assessment of bacteria in the apple phyllosphere. *Journal of Applied Microbiology* **110:** 1284-1296.

Yi H, Cho YJ, Won S, Lee JE, Yu HJ, Kim S *et al* (2011). Duplex-specific nuclease efficiently removes rRNA for prokaryotic RNA-seq. *Nucleic Acids Research* **39:** e40.

Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH *et al* (2006). The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology* **7:** 20.

Yu L, Nicolaisen M, Larsen J, Ravnskov S (2012). Molecular characterization of root-associated fungal communities in relation to health status of *Pisum sativum* using barcoded pyrosequencing. *Plant and Soil* **357:** 395-405.

Zdobnov EM, Apweiler R (2001). InterProScan - an integration platform for the signature-recognition methods in InterPro. *Bioinformatics* **17:** 847-848.

Zhao J, Hyman L, Moore C (1999). Formation of mRNA 3 ' ends in eukaryotes: Mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiology and Molecular Biology Reviews* **63:** 405-445.

Zhao YA, Tang HX, Ye YZ (2012). RAPSearch2: a fast and memory-efficient protein similarity search tool for next-generation sequencing data. *Bioinformatics* **28:** 125-126.

Zhou XG, Wu FZ (2012). p-Coumaric Acid Influenced Cucumber Rhizosphere Soil Microbial Communities and the Growth of Fusarium oxysporum f.sp cucumerinum Owen. *PLoS One* **7:** 11.

Zhulidov PA, Bogdanova EA, Shcheglov AS, Vagner LL, Khaspekov GL, Kozhemyako VB *et al* (2004). Simple cDNA normalization using kamchatka crab duplex-specific nuclease. *Nucleic Acids Research* **32:** 15.

Ziegler K, Diener A, Herpin C, Richter R, Deutzmann R, Lockau W (1998). Molecular characterization of cyanophycin synthetase, the enzyme catalyzing the biosynthesis of the cyanobacterial reserve material multi-L-arginyl-poly-L-aspartate (cyanophycin). *European Journal of Biochemistry* **254:** 154-159.

Appendix

Differentially abundant taxonomic groups from Chapter 3

Table A1: Prokaryotes differentially abundant in the wheat rhizosphere vs soil

Table A2: Prokaryotes differentially abundant in the oat rhizosphere vs soil

Table A3: Prokaryotes differentially abundant in the pea rhizosphere vs soil

Table A4: Eukaryotes differentially abundant in the wheat rhizosphere vs soil

Table A5: Eukaryotes differentially abundant in the oat rhizosphere vs soil

Table A6: Eukaryotes differentially abundant in the pea rhizosphere vs soil

Table A7: Prokaryotes differentially abundant in the sad1 oat rhizosphere vs soil

Table A8: Prokaryotes differentially abundant in the sad1 oat vs wt oat rhizosphere

Differentially abundant taxonomic groups from Chapter 4

Differentially abundant functional groups from Chapter 5

Other

Table A1: Prokaryotic taxonomic groups differentially abundant in the wheat rhizosphere compared to bulk soil (P≤0.05)

Taxonomic group	Soil mean (% of reads)	Wheat mean (% of reads)	P value	Fold change (vs soil)
Anabaena	0.03	0.01	0.003	0.18
Anabaena lemmermannii	0.03	0.01	0.002	0.17
Bacillaceae	0.82	1.22	0.014	1.48
Bacillales	2.81	3.88	0.009	1.38
Bacilli	2.88	3.98	0.007	1.38
Bacillus	0.55	0.88	0.002	1.59
Bacillus sp. IGCAR-1/07	0.19	0.33	0.018	1.75
Catellatospora	0.00	0.01	0.015	2.70
Cellvibrio	0.02	0.06	0.025	3.47
Chitinophaga	0.02	0.04	0.020	2.42
Clostridia	0.22	0.38	0.013	1.72
Clostridiaceae	0.01	0.03	0.045	3.09
Clostridiales	0.21	0.37	0.008	1.74
Dyadobacter	0.00	0.01	0.042	6.14
Fibrobacteraceae	0.00	0.02	0.007	5.90
Fibrobacterales	0.01	0.02	0.045	3.13
Fibrobacteres (class)	0.01	0.02	0.045	3.13
Firmicutes	3.30	4.65	0.006	1.41
Gemmatimonas	0.08	0.06	0.020	0.70
Kineosporiaceae	0.01	0.02	0.050	3.73
Kribbella	0.00	0.01	0.027	2.94
Lysinibacillus	0.13	0.21	0.021	1.67
Methylocaldum	0.01	0.03	0.049	3.78
Nitrospinaceae	0.09	0.11	0.013	1.30
Nostocaceae	0.06	0.03	0.037	0.54
Nostocales	0.08	0.04	0.007	0.52
Roseomonas	0.02	0.00	0.027	0.20

Sphingomonadaceae	0.31	0.16	0.007	0.52
Sphingomonadales	0.50	0.29	0.027	0.59
Terrimonas	0.01	0.03	0.040	2.15
Tetrasphaera	0.00	0.01	0.012	4.31
unclassified Clostridiales	0.01	0.02	0.034	2.35
unclassified Methylophilaceae	0.08	0.10	0.041	1.28
unclassified Verrucomicrobiaceae	0.03	0.07	0.012	2.49
Verrucomicrobia	0.87	1.22	0.010	1.41
Verrucomicrobiaceae	0.14	0.20	0.019	1.51
Verrucomicrobiae	0.14	0.22	0.017	1.55
Verrucomicrobiales	0.14	0.22	0.017	1.55
Verrucomicrobium	0.00	0.01	0.048	5.61
Xanthomonadales	0.53	0.43	0.014	0.82

Table A2: Prokaryotic taxonomic groups differentially abundant in the oat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (% of reads)**	**Oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Acetobacteraceae	0.14	0.07	0.049	0.50
Acidimicrobiales	1.36	0.88	0.002	0.65
Actinobacteria	8.52	5.86	0.006	0.69
Actinobacteria (class)	8.47	5.85	0.007	0.69
Actinomycetales	3.81	2.56	0.030	0.67
alpha proteobacterium FH6	0.06	0.01	0.005	0.20
Alphaproteobacteria	5.04	2.93	0.000	0.58
Anabaena	0.03	0.00	0.000	0.00
Anabaena lemmermannii	0.03	0.00	0.000	0.00
Bacillaceae	0.82	0.40	0.035	0.49
Bacillus pumilus	0.08	0.02	0.013	0.23
Bacillus subtilis	0.05	0.01	0.032	0.14

Bacteriovoracaceae	0.05	0.11	0.042	1.96
Blastococcus	0.06	0.02	0.028	0.39
Blastococcus saxobsidens	0.04	0.01	0.012	0.32
Burkholderiales Genera incertae sedis	0.36	0.66	0.001	1.84
Cellvibrio	0.02	0.09	0.002	4.82
Chloroflexaceae	0.04	0.06	0.030	1.67
Comamonadaceae	0.15	0.38	0.000	2.42
Euryarchaeota	0.06	0.10	0.056	1.67
Fibrobacterales	0.01	0.03	0.026	4.28
Fibrobacteres	0.01	0.03	0.028	4.08
Fibrobacteres (class)	0.01	0.03	0.026	4.28
Flavobacteriaceae	0.13	0.24	0.000	1.91
Gemmatimonas	0.08	0.03	0.001	0.39
Geodermatophilaceae	0.22	0.11	0.017	0.51
Herpetosiphonaceae	0.01	0.03	0.009	3.31
Herpetosiphonales	0.01	0.03	0.024	3.31
Iamibacter	0.16	0.08	0.002	0.50
Kineosporiaceae	0.01	0.02	0.047	3.04
Lysinibacillus	0.13	0.04	0.003	0.32
Lysinibacillus boronitolerans	0.12	0.03	0.001	0.27
Methylocystaceae	0.04	0.02	0.022	0.55
Methylophilaceae	0.08	0.21	0.000	2.56
Methylophilales	0.08	0.21	0.001	2.56
Nannocystis	0.17	0.09	0.048	0.54
Nostocaceae	0.06	0.01	0.006	0.13
Nostocales	0.08	0.01	0.000	0.11
Peptostreptococcaceae	0.00	0.03	0.001	8.94
Pseudomonadaceae	0.30	0.46	0.004	1.53
Rhizobiales	1.47	0.75	0.009	0.51
Rhodobacteraceae	0.45	0.25	0.015	0.56
Rhodobacterales	0.47	0.26	0.006	0.56
Solirubrobacter	0.25	0.10	0.047	0.39

Sphingomonas	0.10	0.05	0.046	0.53
unclassified Actinobacteria	0.13	0.07	0.015	0.56
unclassified Actinobacteridae	0.01	0.29	0.000	29.77
unclassified Actinomycetales	0.11	0.05	0.017	0.44
unclassified Alphaproteobacteria (miscellaneous)	0.06	0.01	0.010	0.23
unclassified Deltaproteobacteria	2.34	1.88	0.029	0.80
unclassified Intrasporangiaceae	0.02	0.00	0.008	0.11
unclassified Methylophilaceae	0.08	0.20	0.003	2.59
Verrucomicrobia	0.87	1.40	0.005	1.62
Verrucomicrobiaceae	0.14	0.20	0.024	1.48

Table A3: Prokaryotic taxonomic groups differentially abundant in the pea rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (% of reads)**	**Pea mean (% of reads)**	**P value**	**Fold change (vs soil)**
Acidimicrobiaceae	0.76	0.49	0.038	0.65
Acidimicrobiales	1.36	0.83	0.016	0.61
alpha proteobacterium FH6	0.06	0.02	0.022	0.36
Alphaproteobacteria	5.04	3.99	0.010	0.79
Amaricoccus	0.25	0.13	0.007	0.51
Anabaena	0.03	0.01	0.002	0.22
Anabaena lemmermannii	0.03	0.01	0.002	0.22
Aquabacterium	0.13	0.09	0.009	0.69
Azospira	0.01	0.02	0.016	2.48
Bacillaceae	0.82	0.38	0.014	0.46
Bacillales	2.81	1.97	0.039	0.70
Bacilli	2.88	2.06	0.040	0.71
Bacillus	0.55	0.26	0.003	0.46
Bacillus pumilus	0.08	0.03	0.019	0.34
Bacillus sp. IGCAR-1/07	0.19	0.08	0.008	0.44

Bdellovibrionaceae	0.29	0.11	0.012	0.38
Bdellovibrionales	0.34	0.15	0.015	0.44
Beijerinckiaceae	0.01	0.03	0.050	2.60
Blastococcus	0.06	0.02	0.040	0.40
Blastococcus saxobsidens	0.04	0.01	0.006	0.27
Burkholderiales Genera incertae sedis	0.36	0.22	0.045	0.61
candidate division OP10	0.19	0.11	0.047	0.56
Caulobacteraceae	0.04	0.08	0.042	2.05
Caulobacterales	0.04	0.08	0.042	2.05
Chlorobi	0.03	0.01	0.027	0.38
Chlorobia	0.03	0.01	0.027	0.38
Chlorobiales	0.03	0.01	0.027	0.38
Chryseobacterium sp. HX31	0.00	0.01	0.045	81.02
Conexibacteraceae	0.21	0.12	0.049	0.55
Cystobacteraceae	0.00	0.01	0.046	3.02
Desulfobacterales	0.10	0.04	0.018	0.45
Dyadobacter	0.00	0.10	0.038	58.00
Dyadobacter koreensis	0.00	0.03	0.006	84.58
Flavobacterium	0.06	0.32	0.047	5.02
Flavobacterium johnsoniae	0.00	0.01	0.025	Absent in soil
Flavobacterium sp. DB3.3-15	0.00	0.01	0.029	Absent in soil
Gemmatimonadaceae	0.46	0.18	0.040	0.40
Gemmatimonadales	0.46	0.18	0.040	0.40
Gemmatimonadetes	0.52	0.20	0.027	0.38
Gemmatimonadetes (class)	0.46	0.18	0.040	0.40
Gemmatimonas	0.08	0.03	0.007	0.33
Geodermatophilaceae	0.22	0.09	0.012	0.43
Haliangiaceae	0.70	0.22	0.029	0.31
Haliangium	0.70	0.22	0.029	0.31
Herpetosiphon	0.00	0.02	0.002	5.84
Herpetosiphonaceae	0.01	0.04	0.004	3.76

Herpetosiphonales	0.01	0.04	0.004	3.76
Iamibacter	0.16	0.09	0.007	0.53
Kineosporia	0.01	0.01	0.038	2.32
Kineosporiaceae	0.01	0.01	0.037	2.25
Kribbella	0.00	0.02	0.044	5.60
Leptothrix	0.06	0.02	0.007	0.36
Lysinibacillus	0.13	0.04	0.003	0.30
Lysinibacillus boronitolerans	0.12	0.03	0.003	0.27
Massilia	0.00	0.01	0.006	34.56
Methanosarcina	0.00	0.01	0.045	10.44
Methanosarcina barkeri str. Fusaro	0.00	0.01	0.034	14.46
Microbacteriaceae	0.24	0.40	0.010	1.68
Myxococcaceae	0.05	0.03	0.028	0.61
Myxococcales	5.29	2.14	0.040	0.41
Nannocystaceae	0.19	0.08	0.018	0.42
Nannocystis	0.17	0.07	0.024	0.42
Nitrosospira	0.02	0.00	0.011	0.10
Nitrospinaceae	0.09	0.04	0.022	0.49
Nocardiaceae	0.03	0.01	0.040	0.50
Nostocaceae	0.06	0.02	0.007	0.29
Nostocales	0.08	0.02	0.001	0.25
Oxalobacteraceae	0.07	0.24	0.024	3.51
Pedomicrobium	0.09	0.03	0.030	0.34
Peptostreptococcaceae	0.00	0.02	0.014	5.97
Planococcaceae	0.18	0.06	0.025	0.34
Prosthecobacter	0.00	0.01	0.037	2.85
Pseudoxanthomonas	0.01	0.03	0.018	5.64
Rhodobacteraceae	0.45	0.30	0.015	0.66
Rhodobacterales	0.47	0.32	0.031	0.67
Ruminococcaceae	0.13	0.08	0.039	0.63
Sinobacteraceae	0.19	0.09	0.007	0.47
Sphingobacteriaceae	0.06	0.13	0.034	2.34

Stenotrophomonas	0.00	0.01	0.009	4.04
Streptomyces	0.03	0.22	0.045	7.61
unclassified Myxococcales	0.53	0.25	0.040	0.47
unclassified Oxalobacteraceae	0.01	0.06	0.010	4.37
unclassified Sorangiineae	0.49	0.25	0.024	0.50
uncultured bacterium #0319-7B4	0.02	0.01	0.034	0.34
Variovorax	0.01	0.02	0.021	2.50
Verrucomicrobiaceae	0.14	0.21	0.017	1.55
Verrucomicrobiae	0.14	0.22	0.015	1.55
Verrucomicrobiales	0.14	0.22	0.015	1.55

Table A4: Eukaryotic taxonomic groups differentially abundant in the wheat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (% of reads)**	**Wheat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Acrobeloides	0.00	0.05	0.025	12.02
Acrobeloides maximus	0.00	0.05	0.025	12.02
Bilateria	0.15	0.32	0.019	2.14
Bodonidae	0.03	0.05	0.046	1.97
Cephalobidae	0.00	0.06	0.027	12.68
Cephaloboidea	0.00	0.06	0.027	12.68
Cercomonadida	0.02	0.05	0.023	2.47
Cercomonadidae	0.01	0.04	0.019	2.77
Chaetothyriomycetidae	0.00	0.01	0.000	6.46
Chromadorea	0.03	0.19	0.001	5.81
Criconematoidea	0.00	0.03	0.006	16.29
Euglenozoa	0.04	0.11	0.018	2.47
Eumetazoa	0.16	0.33	0.029	2.04
Eurotiomycetes	0.00	0.01	0.004	6.14
Kinetoplastida	0.03	0.07	0.015	2.23

Metazoa	0.17	0.34	0.039	1.94
Nematoda	0.03	0.19	0.001	5.75
Paratylenchidae	0.00	0.02	0.000	24.45
Paratylenchus	0.00	0.02	0.000	24.45
Paratylenchus dianthus	0.00	0.02	0.000	24.45
Pseudocoelomata	0.03	0.20	0.001	5.58
Rhabditida	0.01	0.07	0.044	10.47
Tylenchina	0.02	0.10	0.048	4.53
Tylenchulidae	0.00	0.02	0.000	24.45

Table A5: Eukaryotic taxonomic groups differentially abundant in the oat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (% of reads)**	**Oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Agaricomycetes	0.02	0.10	0.038	4.28
Agaricomycotina	0.03	0.11	0.042	3.61
Ascomycota	0.27	0.57	0.044	2.16
Belonolaimidae	0.00	0.01	0.050	Absent in soil
Bodonidae	0.03	0.16	0.020	6.22
Cercomonadida	0.02	0.11	0.001	5.73
Cercomonadidae	0.01	0.08	0.009	5.76
Cercomonas	0.01	0.03	0.048	3.79
Cercozoa	0.13	0.45	0.018	3.46
Chaetothyriomycetidae	0.00	0.02	0.010	10.61
Cnidaria	0.00	0.02	0.049	10.43
Criconematoidea	0.00	0.05	0.001	26.36
Dikarya	0.32	0.72	0.043	2.23
Diplonemida	0.00	0.03	0.044	20.97
Endopterygota	0.00	0.06	0.039	30.63
Euglenida	0.01	0.08	0.001	8.61

Euglenozoa	0.04	0.33	0.003	7.32
Fungi	0.67	2.04	0.030	3.06
Glaeseria	0.02	0.09	0.025	5.78
Glaeseria mira	0.02	0.09	0.025	5.78
Heteromitidae	0.00	0.02	0.000	9.22
Kinetoplastida	0.03	0.19	0.013	6.37
Leptomyxa	0.00	0.03	0.007	7.48
Leptomyxa reticulata	0.00	0.03	0.007	7.48
Leptomyxida	0.02	0.08	0.036	3.30
Leptomyxidae	0.00	0.03	0.007	7.48
Lobosea sp. Mb_5C	0.01	0.03	0.042	4.66
Paratylenchidae	0.00	0.03	0.000	36.58
Paratylenchus	0.00	0.03	0.000	36.58
Paratylenchus dianthus	0.00	0.03	0.000	36.58
Rhizaria	0.16	0.49	0.024	3.06
Sphenomonadales	0.01	0.07	0.007	12.24
Sphenomonadidae	0.00	0.04	0.018	16.39
Stichotrichia	0.01	0.03	0.032	4.59
Tylenchorhynchus	0.00	0.01	0.050	Absent in soil
Tylenchulidae	0.00	0.04	0.002	47.46
unclassified Lobosea	0.01	0.03	0.043	4.39

Table A6: Eukaryotic taxonomic groups differentially abundant in the pea rhizosphere compared to bulk soil (P≤0.05)

Taxonomic group	**Soil mean (% of reads)**	**Pea mean (% of reads)**	**P value**	**Fold change (vs soil)**
Ascomycota	0.27	5.18	0.028	19.52
Basidiomycota	0.04	0.11	0.036	2.71
Bilateria	0.15	2.15	0.038	14.16
Bodonidae	0.03	0.10	0.027	3.76
Bresslaua	0.00	0.02	0.023	5.42

Bresslaua vorax	0.00	0.02	0.023	5.42
Cercomonadida	0.02	0.07	0.005	3.47
Cercomonadidae	0.01	0.06	0.005	4.02
Cercomonas	0.01	0.03	0.035	2.93
Cercomonas edax	0.00	0.01	0.009	3.82
Cercozoa	0.13	0.26	0.018	2.01
Chaetothyriales	0.00	0.03	0.010	18.85
Chaetothyriomycetidae	0.00	0.03	0.029	21.67
Chromadorea	0.03	1.88	0.042	58.04
Chytridiaceae	0.29	1.21	0.004	4.20
Chytridiales	0.29	1.22	0.005	4.18
Chytridiomycetes	0.30	1.24	0.005	4.15
Chytridiomycota	0.30	1.25	0.004	4.16
Chytriomyces	0.29	1.21	0.004	4.22
Chytriomyces spinosus	0.29	1.21	0.004	4.22
Colpodea	0.01	0.04	0.005	5.66
Colpodida	0.01	0.03	0.005	5.72
Colpodidae	0.00	0.02	0.023	5.42
Criconematoidea	0.00	0.10	0.002	55.84
Cyrtolophosidida	0.00	0.01	0.012	5.51
Dikarya	0.32	5.45	0.026	16.81
Dimastigella	0.01	0.03	0.039	6.15
Dimastigella trypaniformis	0.01	0.03	0.039	6.15
Eumetazoa	0.16	2.16	0.037	13.40
Eurotiomycetes	0.00	0.05	0.041	23.16
Exophiala	0.00	0.02	0.012	18.86
Exophiala oligosperma	0.00	0.02	0.012	18.86
Fungi	0.67	7.15	0.020	10.72
Fungi/Metazoa group	0.89	9.64	0.022	10.79
Fusarium	0.01	0.32	0.012	30.97
Fusarium oxysporum	0.01	0.32	0.012	30.97
Fusarium oxysporum f. sp. lycopersici	0.01	0.32	0.012	30.97

Fusarium oxysporum f. sp. lycopersici 4286	0.01	0.32	0.012	30.97
Fusarium oxysporum species complex	0.01	0.32	0.012	30.97
Glomeraceae	0.00	0.04	0.000	10.09
Glomerales	0.00	0.04	0.000	10.09
Glomeromycetes	0.01	0.05	0.001	9.70
Glomeromycota	0.01	0.05	0.001	9.70
Glomus	0.00	0.04	0.001	10.18
Herpotrichiellaceae	0.00	0.02	0.007	17.90
Hyperamoeba	0.00	0.02	0.043	4.39
Hypocreales	0.02	1.78	0.017	73.49
Hypocreomycetidae	0.03	1.82	0.016	67.21
Kinetoplastida	0.03	0.11	0.024	3.57
Labyrinthulida	0.00	0.01	0.011	3.97
Meloidogyne	0.02	1.13	0.040	63.19
Meloidogyne incognita	0.02	1.13	0.040	63.19
Meloidogynidae	0.02	1.13	0.040	63.19
Meloidogyninae	0.02	1.13	0.040	63.19
Metazoa	0.17	2.18	0.036	12.57
mitosporic Ascomycota	0.01	0.07	0.007	9.53
mitosporic Herpotrichiellaceae	0.00	0.02	0.012	18.86
mitosporic Hypocreales	0.01	0.33	0.013	31.22
Mycetozoa	0.01	0.03	0.046	3.95
Nematoda	0.03	1.89	0.041	57.65
Neobodo	0.00	0.02	0.007	4.29
Neobodo designis	0.00	0.02	0.007	4.29
Paratylenchidae	0.00	0.06	0.007	78.59
Paratylenchus	0.00	0.06	0.007	78.59
Paratylenchus dianthus	0.00	0.06	0.007	78.59
Pezizales	0.00	0.01	0.005	12.64
Pezizomycetes	0.00	0.01	0.005	12.64
Pezizomycotina	0.19	3.91	0.027	20.85
Pseudocoelomata	0.03	1.91	0.041	54.55

Sordariomycetes	0.11	2.94	0.029	26.33
Tetracladium	0.01	0.07	0.012	9.13
Tylenchida	0.02	1.33	0.035	54.46
Tylenchina	0.02	1.25	0.034	55.06
Tylenchoidea	0.02	1.15	0.041	55.61
Tylenchulidae	0.00	0.06	0.008	79.70

Table A7: Prokaryotic taxonomic groups differentially abundant in the *sad1* oat rhizosphere compared to bulk soil (P≤0.05)

Taxonomic group	**Soil mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Acidimicrobiaceae	0.76	0.44	0.00	0.58
Acidimicrobiales	1.36	0.68	0.00	0.50
Acidovorax	0.01	0.03	0.02	3.89
Actinobacteria	8.52	4.25	0.00	0.50
Actinobacteria (class)	8.47	4.22	0.00	0.50
Actinomycetales	3.81	1.78	0.00	0.47
alpha proteobacterium FH6	0.06	0.02	0.01	0.38
Alphaproteobacteria	5.04	2.95	0.00	0.58
Amaricoccus	0.25	0.11	0.00	0.44
Amaricoccus tamworthensis	0.08	0.03	0.01	0.35
Amaricoccus veronensis	0.05	0.02	0.03	0.43
Bacillaceae	0.82	0.35	0.01	0.42
Bacillales	2.81	1.80	0.05	0.64
Bacilli	2.88	1.88	0.05	0.65
Bacillus	0.55	0.24	0.01	0.44
Bacillus pumilus	0.08	0.02	0.01	0.23
Bacillus sp. IGCAR-1/07	0.19	0.08	0.02	0.43
Bacteriovoracaceae	0.05	0.08	0.01	1.48
Bacteriovorax	0.00	0.01	0.00	5.44

Bacteroidetes	1.86	2.20	0.01	1.18
Bdellovibrio	0.03	0.05	0.00	1.97
Blastococcus	0.06	0.01	0.00	0.19
candidate division OD1	0.01	0.03	0.03	2.88
candidate division OP10	0.19	0.10	0.01	0.55
candidate division WS3	0.19	0.13	0.05	0.67
Candidatus Alysiosphaera	0.30	0.15	0.04	0.49
Chitinimonas	0.00	0.02	0.02	50.19
Chloroflexaceae	0.04	0.05	0.02	1.38
Chloroflexi (class)	0.05	0.09	0.02	1.69
Comamonadaceae	0.15	0.38	0.00	2.47
Comamonas	0.01	0.01	0.04	2.26
Conexibacteraceae	0.21	0.07	0.01	0.35
Cryomorphaceae	0.01	0.03	0.01	2.81
Cytophaga	0.00	0.01	0.03	5.99
Desulfobacterales	0.10	0.05	0.02	0.49
Euryarchaeota	0.06	0.11	0.04	1.79
Gemmatimonas	0.08	0.04	0.00	0.52
Geodermatophilaceae	0.22	0.06	0.00	0.26
Holophagae	0.09	0.04	0.00	0.51
Hyphomicrobiaceae	0.21	0.16	0.02	0.76
Iamibacter	0.16	0.07	0.00	0.44
Leadbetterella	0.00	0.01	0.02	Absent in soil
Lysinibacillus	0.13	0.04	0.00	0.31
Lysinibacillus boronitolerans	0.12	0.04	0.00	0.30
Lysobacter	0.04	0.03	0.04	0.68
Methanosarcinaceae	0.00	0.01	0.04	6.81
Methanosarcinales	0.00	0.01	0.04	6.20
Methylocystaceae	0.04	0.02	0.05	0.63
Methylophilaceae	0.08	0.16	0.02	1.89
Methylophilales	0.08	0.16	0.02	1.89
Micrococcaceae	0.07	0.03	0.01	0.45

Micromonosporaceae	0.16	0.06	0.04	0.39
Nitrospinaceae	0.09	0.04	0.01	0.43
Nitrospira (class)	0.34	0.19	0.04	0.57
Nitrospirae	0.34	0.19	0.04	0.57
Nitrospirales	0.34	0.19	0.04	0.57
Nocardiaceae	0.03	0.01	0.01	0.46
Nocardioidaceae	0.48	0.23	0.01	0.48
Nocardioides	0.20	0.09	0.03	0.42
Opitutaceae	0.11	0.21	0.03	1.94
Opitutae	0.11	0.22	0.03	1.94
Opitutales	0.11	0.21	0.03	1.94
Opitutus	0.09	0.20	0.00	2.29
Phyllobacteriaceae	0.03	0.02	0.01	0.51
Planococcaceae	0.18	0.07	0.02	0.37
Rhizobiales	1.47	0.79	0.01	0.54
Rhodobacteraceae	0.45	0.23	0.00	0.52
Rhodobacterales	0.47	0.25	0.00	0.54
Rhodospirillaceae	0.13	0.08	0.04	0.57
Rhodospirillales	0.50	0.27	0.03	0.54
Roseiflexus	0.03	0.04	0.05	1.28
Rubrobacterales	0.72	0.22	0.01	0.31
Sinobacteraceae	0.19	0.12	0.02	0.62
Solirubrobacter	0.25	0.06	0.02	0.23
Solirubrobacteraceae	0.30	0.08	0.02	0.25
Sorangium cellulosum 'So ce 56'	0.00	0.01	0.01	15.14
unclassified Actinobacteria	0.13	0.07	0.02	0.58
unclassified Actinomycetales	0.11	0.06	0.05	0.55
unclassified Alphaproteobacteria	0.53	0.31	0.02	0.60
unclassified Alphaproteobacteria (miscellaneous)	0.06	0.03	0.02	0.41
unclassified Bacillales	0.62	0.84	0.02	1.35
unclassified Comamonadaceae	0.04	0.06	0.05	1.69
unclassified Methylophilaceae	0.08	0.15	0.03	1.92

unclassified Rhodospirillaceae	0.03	0.02	0.01	0.48
unclassified Rhodospirillales	0.05	0.03	0.02	0.58
unclassified Rubrobacteridae	0.04	0.02	0.01	0.39
unclassified Verrucomicrobiaceae	0.03	0.06	0.02	2.20
Verrucomicrobia	0.87	1.48	0.00	1.71
Verrucomicrobiaceae	0.14	0.24	0.00	1.79
Verrucomicrobiae	0.14	0.25	0.00	1.80
Verrucomicrobiales	0.14	0.25	0.00	1.80

Table A8: Prokaryotic taxonomic groups differentially abundant in the *sad1* oat rhizosphere compared to the *wt* oat rhizosphere (P≤0.05).

Taxonomic group	***wt* oat mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs *wt* oat)**
candidate division OD1	0.01	0.03	0.007	2.70
Cyanobacteria	0.47	2.52	0.050	5.37
Cytophaga	0.00	0.01	0.044	5.81
Desulfurellaceae	0.02	0.01	0.008	0.50
Gemmata-like str. JW3-8s0	0.01	0.00	0.034	0.13
Geodermatophilaceae	0.11	0.06	0.021	0.51
Haliea	0.00	0.01	0.008	20.79
Kineosporiaceae	0.02	0.01	0.037	0.33
Leptospira	0.00	0.03	0.011	17.96
Leptospiraceae	0.00	0.04	0.019	8.98
Nitrospinaceae	0.08	0.04	0.007	0.46
Parachlamydiaceae	0.03	0.02	0.027	0.58
Sinobacteraceae	0.19	0.12	0.019	0.62
Spirochaetes	0.01	0.05	0.003	3.72
Spirochaetes (class)	0.01	0.05	0.012	3.57
Streptomycetaceae	0.08	0.03	0.029	0.40
Streptosporangiaceae	0.03	0.01	0.001	0.25

Thiotrichales	0.00	0.01	0.002	8.09
unclassified Bacillales	0.65	0.84	0.027	1.30
unclassified Intrasporangiaceae	0.00	0.02	0.034	6.19

Table A9: Eukaryotic taxonomic groups differentially abundant in the *sad1* oat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs soil)**
Adineta	0.00	0.01	0.034	5.72
Adineta vaga	0.00	0.01	0.034	5.72
Adinetida	0.00	0.01	0.034	5.72
Adinetidae	0.00	0.01	0.034	5.72
Agaricomycetes	0.02	0.14	0.036	5.89
Agaricomycetes incertae sedis	0.02	0.09	0.036	5.11
Agaricomycetidae	0.00	0.01	0.021	10.32
Agaricomycotina	0.03	0.15	0.039	4.94
Alveolata	0.27	1.30	0.002	4.82
Amoebozoa	0.50	1.40	0.014	2.80
Anaplectus	0.00	0.02	0.047	Absent in soil
Anaplectus sp. PDL-2005	0.00	0.02	0.047	Absent in soil
Apicomplexa	0.15	0.49	0.001	3.21
Aplanochytrium	0.00	0.01	0.044	Absent in soil
Arachnula	0.01	0.02	0.042	2.16
Arachnula sp. ATCC 50593	0.01	0.02	0.042	2.16
Araeolaimida	0.00	0.03	0.006	Absent in soil
Ascomycota	0.27	0.58	0.026	2.19
Auriculariales	0.02	0.08	0.021	4.44
Basidiomycota	0.04	0.16	0.042	3.94
Bodo	0.00	0.02	0.013	48.30
Bodonidae	0.03	0.13	0.000	5.05
Boletales	0.00	0.01	0.008	13.05

Cercomonadida	0.02	0.12	0.000	5.96
Cercomonadida environmental sample	0.00	0.01	0.007	6.69
Cercomonadidae	0.01	0.09	0.000	6.44
Cercomonas	0.01	0.03	0.009	3.84
Cercomonas longicauda	0.00	0.01	0.001	4.64
Cercozoa	0.13	0.55	0.003	4.27
Chaetothyriomycetidae	0.00	0.01	0.036	8.53
Choanoflagellida	0.00	0.02	0.039	18.97
Chromulinales	0.00	0.03	0.004	45.68
Chrysophyceae	0.00	0.04	0.006	27.00
Chytridiaceae	0.29	1.39	0.001	4.80
Chytridiales	0.29	1.40	0.001	4.78
Chytridiomycetes	0.30	1.41	0.001	4.73
Chytridiomycota	0.30	1.42	0.001	4.72
Chytriomyces	0.29	1.39	0.001	4.82
Chytriomyces spinosus	0.29	1.38	0.001	4.82
Ciliophora	0.08	0.63	0.007	7.84
Cnidaria	0.00	0.01	0.008	8.77
Coccidia	0.15	0.48	0.001	3.19
Codonosigidae	0.00	0.01	0.043	43.17
Cryptophyta	0.00	0.02	0.046	10.66
Cyrtolophosidida	0.00	0.02	0.048	9.56
Cyrtolophosididae	0.00	0.02	0.043	Absent in soil
Dactylopodida	0.01	0.14	0.005	23.06
Dermamoeba	0.01	0.05	0.018	3.46
Dermamoeba algensis	0.01	0.05	0.018	3.46
Dikarya	0.32	0.78	0.012	2.40
Dimastigella	0.01	0.02	0.031	3.98
Dimastigella trypaniformis	0.01	0.02	0.031	3.98
Dinophyceae	0.00	0.01	0.000	7.80
Diplonemida	0.00	0.03	0.005	17.40
Eimeriidae	0.15	0.48	0.001	3.21

Eimeriorina	0.15	0.48	0.001	3.20
Endopterygota	0.00	0.02	0.048	7.30
environmental samples++++++++++++	0.00	0.05	0.008	79.16
environmental samples++++++++++++++++++++++++	0.00	0.01	0.007	6.69
Eucoccidiorida	0.15	0.48	0.001	3.19
Euglenida	0.01	0.18	0.002	18.15
Euglenozoa	0.04	0.39	0.001	8.87
Euglypha	0.00	0.03	0.013	10.65
Euglypha tuberculata	0.00	0.02	0.014	8.80
Euglyphidae	0.01	0.07	0.022	12.42
Eurotiomycetes	0.00	0.02	0.024	7.32
Filamoeba	0.00	0.01	0.001	4.98
Filamoeba nolandi	0.00	0.01	0.001	4.98
Flabellinea	0.06	0.35	0.005	6.27
Flabellulidae	0.02	0.13	0.006	6.53
Foraminifera	0.02	0.07	0.050	3.21
Fungi	0.67	2.52	0.001	3.78
Fungi/Metazoa group	0.89	7.22	0.021	8.07
Glaeseria	0.02	0.12	0.009	8.11
Glaeseria mira	0.02	0.12	0.009	8.11
Glomeraceae	0.00	0.17	0.022	38.63
Glomerales	0.00	0.17	0.022	38.63
Glomeromycetes	0.01	0.19	0.020	34.98
Glomeromycota	0.01	0.19	0.020	34.98
Glomus	0.00	0.15	0.024	43.48
Glomus eburneum	0.00	0.02	0.005	25.34
Glomus etunicatum	0.00	0.01	0.004	13.47
Haptoria	0.00	0.02	0.027	11.76
Haptorida	0.00	0.01	0.032	6.94
Heteromitidae	0.00	0.01	0.045	4.65
Intramacronucleata	0.08	0.61	0.008	7.74
Kinetoplastida	0.03	0.18	0.000	5.79

Korotnevella	0.00	0.02	0.040	5.96
Labyrinthulida	0.00	0.03	0.009	8.36
Leptomyxa	0.00	0.10	0.006	28.91
Leptomyxa reticulata	0.00	0.10	0.006	28.91
Leptomyxida	0.02	0.23	0.005	9.74
Leptomyxidae	0.00	0.10	0.006	28.91
Lobosea sp. Mb_5C	0.01	0.04	0.034	6.83
Mayorella	0.00	0.11	0.005	213.18
Mayorella sp. JJP-2003	0.00	0.11	0.005	213.18
Mucoromycotina	0.00	0.01	0.002	9.54
Mycetozoa	0.01	0.02	0.047	2.51
Neobodo	0.00	0.03	0.017	6.88
Neobodo designis	0.00	0.03	0.017	6.88
Nolandella	0.00	0.01	0.033	3.52
Nuclearia	0.00	0.03	0.038	30.54
Nucleariidae	0.00	0.03	0.025	33.63
Oligohymenophorea	0.04	0.27	0.001	6.59
Orchitophryidae	0.02	0.06	0.028	3.13
Oxytrichidae	0.00	0.02	0.017	5.71
Paraflabellula	0.01	0.07	0.008	6.59
Paraflabellula hoguae	0.01	0.07	0.008	6.59
Paramoebidae	0.00	0.13	0.006	30.94
Paraphysomonadaceae	0.00	0.03	0.013	84.24
Paraphysomonas	0.00	0.03	0.013	84.24
Paraphysomonas foraminifera	0.00	0.01	0.045	47.24
Pessonella	0.00	0.01	0.006	40.59
Pessonella sp. PRA-29	0.00	0.01	0.006	40.59
Petalomonas	0.00	0.05	0.008	23.76
Petalomonas cantuscygni	0.00	0.05	0.008	23.76
Pezizales	0.00	0.02	0.007	20.37
Pezizomycetes	0.00	0.02	0.007	20.37
Philasterida	0.03	0.09	0.032	3.11

Plasmodiophorida	0.00	0.01	0.039	5.42
Plasmodiophoridae	0.00	0.01	0.039	5.42
Plectidae	0.00	0.03	0.006	Absent in soil
Plectoidea	0.00	0.03	0.006	Absent in soil
Prorodontida	0.00	0.05	0.007	79.90
Prorodontidae environmental sample	0.00	0.05	0.008	79.16
Prostomatea	0.00	0.05	0.007	79.90
Rhizaria	0.16	0.68	0.001	4.26
Rhynchomonas	0.00	0.03	0.016	8.78
Rhynchomonas nasuta	0.00	0.03	0.016	8.78
Rotaliina	0.00	0.01	0.031	18.35
Schizopyrenida	0.01	0.03	0.028	3.69
Scuticociliatia	0.04	0.14	0.008	3.75
Sphenomonadales	0.01	0.16	0.001	28.69
Sphenomonadidae	0.00	0.10	0.003	38.79
Spirotrichea	0.01	0.05	0.000	6.82
Stichotrichia	0.01	0.04	0.000	6.26
Stichotrichida	0.00	0.02	0.011	6.41
Thaumatomonadida	0.02	0.05	0.044	2.88
Thaumatomonas	0.00	0.02	0.006	8.08
Thaumatomonas seravini	0.00	0.02	0.006	8.08
Thraustochytriidae	0.00	0.02	0.012	11.01
Tracheleuglypha	0.00	0.04	0.038	14.50
Tracheleuglypha dentata	0.00	0.04	0.038	14.50
unclassified eukaryotes	0.01	0.03	0.016	2.77
unclassified Lobosea	0.01	0.05	0.028	6.48
unclassified Vannella	0.00	0.02	0.032	50.15
Vahlkampfia	0.00	0.01	0.025	4.10
Vahlkampfiidae	0.01	0.03	0.015	4.80
Vampyrellidae	0.01	0.02	0.042	2.16
Vannella	0.00	0.04	0.010	112.45
Vannellidae	0.01	0.11	0.027	8.72

Table A10: Eukaryotic taxonomic groups differentially abundant in the *sad1* oat rhizosphere compared to the *wt* oat rhizosphere (P≤0.05)

Taxonomic group	***wt* oat mean (% of reads)**	***sad1* oat mean (% of reads)**	**P value**	**Fold change (vs *wt* oat)**
Alveolata	0.70	1.30	0.047	1.85
Amoebozoa	0.81	1.40	0.035	1.72
Apicomplexa	0.26	0.49	0.003	1.93
Aplanochytrium	0.00	0.01	0.044	Absent in *wt* oat
Cercomonadida environmental sample	0.00	0.01	0.006	11.79
Coccidia	0.25	0.48	0.004	1.91
Criconematoidea	0.05	0.02	0.028	0.37
Dactylopodida	0.02	0.14	0.010	6.65
Dermamoeba	0.01	0.05	0.009	4.51
Dermamoeba algensis	0.01	0.05	0.009	4.51
Dicondylia	0.10	0.05	0.028	0.46
Eimeriidae	0.25	0.48	0.004	1.90
Eimeriorina	0.25	0.48	0.004	1.90
environmental samples	0.00	0.01	0.006	11.79
Eucoccidiorida	0.25	0.48	0.004	1.90
Euglenida	0.08	0.18	0.035	2.11
Flabellinea	0.12	0.35	0.019	2.85
Flabellulidae	0.05	0.13	0.021	2.51
Foraminifera	0.01	0.07	0.012	5.23
Heteromitidae	0.02	0.01	0.025	0.50
Hexapoda	0.11	0.05	0.030	0.50
Insecta	0.10	0.05	0.028	0.46
Korotnevella	0.00	0.02	0.021	10.68
Korotnevella stella	0.00	0.01	0.019	7.26
Leptomyxa	0.03	0.10	0.023	3.86
Leptomyxa reticulata	0.03	0.10	0.023	3.86
Leptomyxida	0.08	0.23	0.018	2.95

Leptomyxidae	0.03	0.10	0.023	3.86
Mandibulata	0.12	0.06	0.043	0.53
Mayorella	0.01	0.11	0.009	10.90
Mayorella sp. JJP-2003	0.01	0.11	0.009	10.90
Mucoromycotina	0.00	0.01	0.012	5.94
Neoptera	0.10	0.05	0.028	0.46
Nuclearia	0.00	0.03	0.049	11.13
Nucleariidae	0.00	0.03	0.042	7.84
Pancrustacea	0.11	0.06	0.041	0.56
Paraflabellula	0.03	0.07	0.029	2.58
Paraflabellula hoguae	0.03	0.07	0.029	2.58
Paramoebidae	0.01	0.13	0.008	10.95
Paratylenchidae	0.03	0.01	0.003	0.30
Paratylenchus	0.03	0.01	0.003	0.30
Paratylenchus dianthus	0.03	0.01	0.003	0.30
Pessonella	0.00	0.01	0.009	15.12
Pessonella sp. PRA-29	0.00	0.01	0.009	15.12
Pezizales	0.00	0.02	0.020	5.66
Pezizomycetes	0.00	0.02	0.020	5.66
Pterygota	0.10	0.05	0.028	0.46
Rotaliina	0.00	0.01	0.038	12.13
Sphenomonadales	0.07	0.16	0.030	2.34
Thaumatomastigidae	0.01	0.03	0.026	3.58
Thaumatomonadida	0.01	0.05	0.019	4.45
Thaumatomonas	0.00	0.02	0.005	14.80
Thaumatomonas seravini	0.00	0.02	0.005	14.80
Thraustochytriidae	0.01	0.02	0.046	3.68
Tylenchulidae	0.04	0.01	0.007	0.24

Table A11: Prokaryotic taxonomic groups selected only in the wheat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Wheat mean (transcripts / g)	P value	Fold change (vs soil)
Acidiphilium	2125	4132	0.036	1.94
Acidiphilium cryptum	560	1740	0.032	3.10
Acidiphilium cryptum JF-5	560	1740	0.032	3.10
Aerococcaceae	91	478	0.007	5.26
Aerococcus	7	278	0.008	41.88
Akkermansia	3597	9256	0.006	2.57
Akkermansia muciniphila	3597	9256	0.006	2.57
Akkermansia muciniphila ATCC BAA-835	3597	9256	0.006	2.57
Alistipes	1792	4950	0.021	2.76
Alteromonas	186	1038	0.016	5.58
Anaeromyxobacter dehalogenans	2552	4728	0.046	1.85
Bacillus halodurans	201	1213	0.002	6.04
Bacillus halodurans C-125	201	1213	0.002	6.04
Bacillus sp. m3-13	186	908	0.006	4.88
Burkholderia ambifaria	673	1354	0.013	2.01
Capnocytophaga sputigena	186	712	0.008	3.83
Capnocytophaga sputigena Capno	186	712	0.008	3.83
Chlorobi	14552	28108	0.049	1.93
Chlorobia	14495	27845	0.050	1.92
Chlorobiaceae	14495	27845	0.050	1.92
Chlorobiales	14495	27845	0.050	1.92
Chlorobium	4647	7972	0.046	1.72
Clostridium hathewayi	0	419	0.044	Absent in soil

Colwellia	461	1663	0.002	3.61
Colwellia psychrerythraea	461	1663	0.002	3.61
Colwellia psychrerythraea 34H	461	1663	0.002	3.61
Colwelliaceae	461	1663	0.002	3.61
Comamonas testosteroni KF-1	1712	3576	0.004	2.09
Coprothermobacter	384	914	0.024	2.38
Coprothermobacter proteolyticus	384	914	0.024	2.38
Coprothermobacter proteolyticus DSM 5265	384	914	0.024	2.38
Deinococcus deserti	722	1473	0.030	2.04
Deinococcus deserti VCD115	722	1473	0.030	2.04
Desulfatibacillum	2111	3564	0.032	1.69
Desulfatibacillum alkenivorans	2111	3564	0.032	1.69
Desulfatibacillum alkenivorans AK-01	2111	3564	0.032	1.69
Desulfitobacterium hafniense	40	491	0.048	12.34
Edwardsiella	170	881	0.039	5.19
Fibrobacter	3306	9296	0.027	2.81
Fibrobacter succinogenes	3306	9296	0.027	2.81
Fibrobacter succinogenes subsp. succinogenes	3306	9296	0.027	2.81
Fibrobacter succinogenes subsp. succinogenes S85	3306	9296	0.027	2.81
Fibrobacteraceae	3306	9296	0.027	2.81
Fibrobacterales	3306	9296	0.027	2.81
Fibrobacteres	3306	9296	0.027	2.81
Fibrobacteres (class)	3306	9296	0.027	2.81
Flavobacteria bacterium BBFL7	1053	2752	0.029	2.61
Flavobacteria bacterium MS024-3C	358	1960	0.000	5.48
Flavobacteriales bacterium HTCC2170	1981	4877	0.033	2.46
Francisella	195	1438	0.001	7.38

Francisellaceae	195	1438	0.001	7.38
Gallionellaceae	382	1328	0.027	3.48
gamma proteobacterium IMCC2047	176	675	0.009	3.84
gamma proteobacterium NOR5-3	1225	2481	0.026	2.02
Geobacter sp. M21	196	1307	0.001	6.66
Haemophilus parasuis	0	663	0.003	Absent in soil
Haemophilus parasuis 29755	0	368	0.046	Absent in soil
Herbaspirillum seropedicae	16	506	0.007	31.78
Herpetosiphon	16563	33914	0.021	2.05
Herpetosiphon aurantiacus	16563	33914	0.021	2.05
Herpetosiphon aurantiacus ATCC 23779	16563	33914	0.021	2.05
Herpetosiphonaceae	16563	33914	0.021	2.05
Herpetosiphonales	16563	33914	0.021	2.05
Hydrogenivirga	457	1641	0.011	3.59
Hydrogenivirga sp. 128-5-R1-1	457	1641	0.011	3.59
Kordia	2043	4689	0.042	2.30
Kordia algicida	2043	4689	0.042	2.30
Kordia algicida OT-1	2043	4689	0.042	2.30
Leptospira biflexa serovar Patoc strain 'Patoc 1 (Paris)'	1180	2654	0.033	2.25
Leptospira borgpetersenii	0	1139	0.015	Absent in soil
Leptospira borgpetersenii serovar Hardjo-bovis	0	1139	0.015	Absent in soil
Leptospira borgpetersenii serovar Hardjo-bovis L550	0	1035	0.018	Absent in soil
Leptospiraceae	5105	10451	0.038	2.05
Maricaulis	944	2879	0.001	3.05
Maricaulis maris	944	2879	0.001	3.05
Maricaulis maris MCS10	944	2879	0.001	3.05
Marinitoga	115	625	0.015	5.45

Marinitoga piezophila	115	625	0.015	5.45
Marinitoga piezophila KA3	115	625	0.015	5.45
Meiothermus silvanus	1911	3983	0.011	2.08
Meiothermus silvanus DSM 9946	1911	3983	0.011	2.08
Myxococcales	574926	1430795	0.038	2.49
Oceanimonas	100	1281	0.010	12.81
Oceanimonas sp. GK1	100	1281	0.010	12.81
Opitutaceae	202442	502335	0.011	2.48
Opitutaceae bacterium TAV2	5727	15190	0.010	2.65
Opitutae	207965	521905	0.009	2.51
Opitutales	202442	502335	0.011	2.48
Opitutus	105625	230619	0.029	2.18
Opitutus terrae	105601	229703	0.030	2.18
Opitutus terrae PB90-1	105601	229703	0.030	2.18
Pectobacterium	58	1083	0.022	18.54
Pelotomaculum	513	1165	0.014	2.27
Pelotomaculum thermopropionicum	513	1165	0.014	2.27
Pelotomaculum thermopropionicum SI	513	1165	0.014	2.27
Polyangiaceae	173791	732915	0.012	4.22
Prevotella ruminicola	139	1636	0.031	11.75
Prevotella ruminicola 23	139	1636	0.031	11.75
Prochlorales	754	2270	0.018	3.01
Prochlorococcaceae	754	2255	0.019	2.99
Prochlorococcus	754	2255	0.019	2.99
Prochlorococcus marinus	754	2165	0.016	2.87
Prosthecobacter	42	623	0.003	14.94
Prosthecochloris	262	1776	0.032	6.78

Prosthecochloris aestuarii	262	1776	0.032	6.78
Prosthecochloris aestuarii DSM 271	262	1776	0.032	6.78
Protochlamydia	688	2417	0.036	3.51
Protochlamydia amoebophila	688	2314	0.041	3.36
Protochlamydia amoebophila UWE25	688	2314	0.041	3.36
Sagittula	578	1372	0.022	2.38
Sagittula stellata	578	1372	0.022	2.38
Sagittula stellata E-37	578	1372	0.022	2.38
Saprospiraceae	1086	2631	0.010	2.42
Sorangiineae	173791	732915	0.012	4.22
Sorangium	173388	732550	0.012	4.22
Sorangium cellulosum	173388	732550	0.012	4.22
Sorangium cellulosum 'So ce 56'	172917	729448	0.012	4.22
Spirosoma	22778	46230	0.039	2.03
Spirosoma linguale	22778	46230	0.039	2.03
Spirosoma linguale DSM 74	22778	46230	0.039	2.03
Stenotrophomonas sp. SKA14	460	1509	0.045	3.28
Streptomyces sp. SPB74	224	900	0.046	4.01
Succinivibrionaceae	25	355	0.007	14.18
Sulcia	275	1003	0.040	3.65
Sulcia muelleri	275	1003	0.040	3.65
Synechococcus	8171	13372	0.040	1.64
Synechococcus sp. JA-2-3B'a(2-13)	351	1002	0.002	2.85
Synechococcus sp. JA-3-3Ab	142	992	0.001	6.99
Thermodesulfobiaceae	422	914	0.022	2.16
Thermotoga lettingae	127	526	0.010	4.15
Thermotoga lettingae TMO	127	526	0.010	4.15

Thiobacillus intermedius K12	2380	4748	0.021	1.99
Thiomonas	2394	4834	0.017	2.02
Thiomonas intermedia	2380	4748	0.021	1.99
unclassified Enterobacteriaceae	177	573	0.013	3.24
uncultured marine bacterium Ant29B7	42	576	0.045	13.81
uncultured Sphingobacteria bacterium	821	3273	0.020	3.99
Veillonella	729	2312	0.016	3.17
Verminephrobacter sp. At4	1971	3586	0.017	1.82
Verrucomicrobiaceae	42023	100425	0.015	2.39
Verrucomicrobium	35521	84570	0.020	2.38
Verrucomicrobium spinosum	35521	84570	0.020	2.38
Verrucomicrobium spinosum DSM 4136	35521	84570	0.020	2.38

Table A12: Prokaryotic taxonomic groups selected only in the oat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Oat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Actinomyces	1653	4316	0.019	2.61
Actinomycetaceae	2552	5358	0.022	2.10
Actinomycineae	2552	5358	0.022	2.10
Alkaliphilus	572	2418	0.019	4.22
Arthrobacter sp. K-1	0	2686	0.012	Absent in soil
Bacillus cereus NC7401	1630	4364	0.023	2.68
Bacillus pumilus	163	1856	0.036	11.40
Bacillus sp. B14905	38	499	0.018	13.17
Bacteriovorax	119	751	0.043	6.32

Betaproteobacteria	3393688	10864639	0.021	3.20
Brucella	35817	91141	0.048	2.54
Brucella melitensis	30079	81366	0.036	2.71
Brucella melitensis 16M	30057	81350	0.036	2.71
Burkholderia pseudomallei BCC215	0	769	0.008	Absent in soil
Burkholderiales	2367306	8011024	0.018	3.38
Burkholderiales Genera incertae sedis	103921	348445	0.017	3.35
Clostridiaceae	44298	302162	0.013	6.82
Clostridium	41207	288606	0.014	7.00
Clostridium acetobutylicum	254	2856	0.029	11.23
Clostridium acetobutylicum ATCC 824	234	2416	0.042	10.30
Clostridium beijerinckii	714	17726	0.026	24.84
Clostridium beijerinckii NCIMB 8052	714	17709	0.026	24.82
Clostridium botulinum B	0	844	0.030	Absent in soil
Clostridium botulinum B str. Eklund 17B	0	827	0.032	Absent in soil
Clostridium botulinum E	0	392	0.040	Absent in soil
Clostridium butyricum	204	4449	0.025	21.84
Clostridium butyricum 5521	204	4325	0.032	21.24
Clostridium carboxidivorans	838	14616	0.006	17.44
Clostridium carboxidivorans P7	838	14616	0.006	17.44
Clostridium cellulovorans	154	2291	0.016	14.83
Clostridium cellulovorans 743B	154	2291	0.016	14.83
Clostridium perfringens	432	1584	0.008	3.66
Collimonas	100	812	0.017	8.15
Enterobacter sakazakii ATCC BAA-894	0	408	0.041	Absent in soil
Enterococcus	394	1270	0.024	3.23
Erythrobacter sp. NAP1	2615	5350	0.028	2.05

Flavobacteria	493585	1560235	0.031	3.16
Flavobacteriaceae	231421	785663	0.039	3.39
Flavobacteriales	434802	1345943	0.036	3.10
Glaciibacter	67	812	0.016	12.17
Glaciibacter superstes	67	812	0.016	12.17
Hyphomonadaceae	13670	36330	0.013	2.66
Leeuwenhoekiella	1714	5219	0.019	3.04
Leeuwenhoekiella blandensis	1706	5207	0.020	3.05
Leeuwenhoekiella blandensis MED217	1706	5207	0.020	3.05
Legionella	22566	78851	0.005	3.49
Legionella pneumophila	2877	5912	0.037	2.06
Legionella taurinensis	18962	72028	0.006	3.80
Lysobacter	5338	8275	0.050	1.55
Lysobacter enzymogenes	5014	8116	0.034	1.62
Mycobacterium tuberculosis H37Ra	3370	17551	0.000	5.21
Novosphingobium sp. Rr 2-17	4392	13399	0.003	3.05
Oxalobacteraceae	378554	1160016	0.031	3.06
Paenibacillus polymyxa	1038	3539	0.019	3.41
Paenibacillus polymyxa E681	550	2673	0.026	4.86
Pedobacter sp. BAL39	19473	71816	0.005	3.69
Phenylobacterium	79295	175314	0.022	2.21
Phenylobacterium zucineum	79076	175297	0.022	2.22
Phenylobacterium zucineum HLK1	79076	175297	0.022	2.22
Photobacterium profundum	177	1050	0.004	5.93
Polaromonas naphthalenivorans	14854	38821	0.043	2.61
Polaromonas naphthalenivorans CJ2	14854	38821	0.043	2.61
Polaromonas sp. JS666	32653	90902	0.044	2.78

Propionibacterium acnes	41610	112970	0.049	2.71
Propionibacterium acnes KPA171202	17126	80661	0.004	4.71
Pseudomonas fluorescens Pf0-1	3911	26302	0.045	6.73
Pseudomonas putida F1	151	1927	0.049	12.72
Pseudomonas savastanoi	65	858	0.040	13.23
Pseudomonas sp. ND6	81	1058	0.027	13.06
Pseudomonas syringae pv. phaseolicola	0	858	0.027	Absent in soil
Pseudomonas syringae pv. phaseolicola 1448A	0	858	0.027	Absent in soil
Pseudomonas syringae pv. tomato str. DC3000	275	2147	0.038	7.79
Ramlibacter	43449	89505	0.034	2.06
Ramlibacter tataouinensis	43449	89505	0.034	2.06
Ramlibacter tataouinensis TTB310	43449	89505	0.034	2.06
Rhodobacterales bacterium HTCC2255	210	704	0.008	3.35
Saccharomonospora viridis	431	1866	0.011	4.33
Saccharomonospora viridis DSM 43017	431	1866	0.011	4.33
Sphingobacteria	2920665	6878587	0.048	2.36
Sphingobacteriales	2906474	6833650	0.049	2.35
Sphingobacterium	14219	43037	0.029	3.03
Sphingobacterium spiritivorum	9231	25848	0.042	2.80
Sphingomonadaceae	362816	812037	0.038	2.24
Sphingomonadales	591139	1334638	0.035	2.26
Streptomyces griseoflavus	798	9272	0.017	11.62
Streptomyces griseoflavus Tu4000	798	9272	0.017	11.62
uncultured soil bacterium	6117	17610	0.023	2.88
Vibrio	7621	14593	0.035	1.91
Vibrio vulnificus	4107	10551	0.014	2.57
Vibrio vulnificus CMCP6	3939	10397	0.010	2.64

Vibrionaceae	8924	17232	0.019	1.93
Xanthomonas	133040	456440	0.000	3.43
Xanthomonas albilineans	23251	116516	0.000	5.01
Xanthomonas albilineans GPE PC73	23244	116516	0.000	5.01
Xanthomonas axonopodis	116	591	0.046	5.07
Xanthomonas axonopodis pv. citri	116	529	0.050	4.54
Xanthomonas axonopodis pv. citri str. 306	116	529	0.050	4.54
Xanthomonas campestris pv. campestris	611	1988	0.049	3.25
Xanthomonas campestris pv. campestris str. B100	68	1019	0.012	14.95
Xanthomonas oryzae	3027	5984	0.006	1.98
Xanthomonas oryzae pv. oryzae	1515	3392	0.026	2.24
Xanthomonas oryzae pv. oryzae KACC10331	1346	2922	0.042	2.17
Xanthomonas oryzae pv. oryzicola	524	1852	0.004	3.53
Xanthomonas oryzae pv. oryzicola BLS256	524	1852	0.004	3.53

Table A13: Prokaryotic taxonomic groups selected only in the pea rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Pea mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Arthrobacter chlorophenolicus	1808	5752	0.007	3.18
Arthrobacter chlorophenolicus A6	1808	5752	0.007	3.18
Bifidobacterium	846	2024	0.017	2.39
Bradyrhizobiaceae	996110	4445480	0.029	4.46
Bradyrhizobium	329276	1376183	0.037	4.18
Bradyrhizobium japonicum	32116	125777	0.039	3.92
Bradyrhizobium japonicum USDA 110	20746	84733	0.024	4.08

Bradyrhizobium sp. BTAi1	5145	18336	0.029	3.56
Bradyrhizobium sp. ORS285	1834	6094	0.032	3.32
Burkholderia dolosa	216	924	0.033	4.28
Burkholderia dolosa AUO158	216	924	0.033	4.28
Burkholderia sp. H160	2562	8106	0.005	3.16
Burkholderia thailandensis MSMB43	146	2836	0.028	19.44
Chryseobacterium	6399	47391	0.030	7.41
Chryseobacterium gleum	5811	46978	0.029	8.08
Chryseobacterium gleum ATCC 35910	5811	46978	0.029	8.08
Citrobacter koseri	214	3733	0.029	17.44
Citrobacter koseri ATCC BAA-895	214	3733	0.029	17.44
Clostridium phytofermentans	881	2252	0.042	2.56
Clostridium phytofermentans ISDg	881	2252	0.042	2.56
Enterobacteriaceae	93868	1879574	0.026	20.02
Enterobacteriales	93868	1879574	0.026	20.02
Gluconobacter	623	1769	0.047	2.84
Haemophilus	1147	4459	0.017	3.89
Idiomarina	2147	24359	0.037	11.35
Idiomarina baltica	101	23760	0.029	234.27
Idiomarina baltica OS145	101	23760	0.029	234.27
Idiomarinaceae	2147	24359	0.037	11.35
Klebsiella pneumoniae	114	1785	0.048	15.63
Labrenzia alexandrii	740	2577	0.048	3.48
Labrenzia alexandrii DFL-11	740	2577	0.048	3.48
Magnetospirillum	18331	55912	0.033	3.05
Magnetospirillum magnetotacticum	1712	3643	0.015	2.13
Magnetospirillum magnetotacticum MS-1	1712	3643	0.015	2.13

Methylophilaceae	16638	152706	0.035	9.18
Methylophilales	17178	163296	0.035	9.51
Methylotenera	2768	49273	0.049	17.80
Mycobacterium avium	66251	372637	0.034	5.62
Mycobacterium avium complex (MAC)	68053	375444	0.035	5.52
Mycobacterium avium subsp. paratuberculosis	65772	369629	0.035	5.62
Mycobacterium avium subsp. paratuberculosis K-10	65734	369629	0.035	5.62
Mycobacterium gastri	348	1758	0.005	5.06
Mycobacterium intracellulare	405	1779	0.011	4.39
Mycobacterium smegmatis	9234	36747	0.017	3.98
Mycobacterium smegmatis str. MC2 155	9234	36747	0.017	3.98
Mycobacterium tuberculosis	77078	382650	0.020	4.96
Mycobacterium tuberculosis complex	78971	387912	0.019	4.91
Nitrobacter	40697	105531	0.049	2.59
Pantoea	1108	414943	0.024	374.46
Pantoea sp. aB	42	6916	0.010	165.75
Photorhabdus luminescens	935	3820	0.019	4.08
Photorhabdus luminescens subsp. laumondii	935	3820	0.019	4.08
Photorhabdus luminescens subsp. laumondii TTO1	935	3820	0.019	4.08
Pseudomonas	286093	4609770	0.049	16.11
Pseudomonas aeruginosa	3354	16663	0.028	4.97
Pseudomonas aeruginosa PA7	372	7759	0.044	20.86
Pseudomonas fluorescens SBW25	4276	45216	0.039	10.57
Pseudomonas putida	3435	17702	0.035	5.15
Pseudomonas putida group	3791	17702	0.040	4.67
Pseudomonas syringae group genomosp. 3	973	4870	0.050	5.01
Pseudomonas syringae pv. tomato	950	4834	0.047	5.09

Rahnella	148	1144	0.016	7.73
Renibacterium	671	3906	0.019	5.82
Renibacterium salmoninarum	671	3906	0.019	5.82
Renibacterium salmoninarum ATCC 33209	671	3906	0.019	5.82
Rhizobiaceae	152933	2566668	0.023	16.78
Rhizobiales	2254218	10545012	0.037	4.68
Rhizobium	49411	1305148	0.017	26.41
Rhizobium etli	8705	122077	0.022	14.02
Rhizobium etli CIAT 894	579	12554	0.034	21.68
Rhizobium etli Kim 5	230	9440	0.011	41.09
Rhizobium leguminosarum	19140	369614	0.024	19.31
Rhizobium leguminosarum bv. trifolii	3197	56680	0.022	17.73
Rhizobium leguminosarum bv. trifolii WSM1325	1775	38575	0.031	21.74
Rhizobium leguminosarum bv. trifolii WSM2304	1343	10315	0.012	7.68
Rhizobium/Agrobacterium group	83910	1961821	0.020	23.38
Rhodococcus	54840	162353	0.049	2.96
Rhodopseudomonas palustris BisB18	6372	26423	0.016	4.15
Rothia	1963	6152	0.045	3.13
Rothia dentocariosa	632	2358	0.019	3.73
Rothia mucilaginosa	456	2060	0.027	4.51
Sanguibacter	1050	2904	0.030	2.77
Sanguibacter keddieii	1050	2904	0.030	2.77
Sanguibacter keddieii DSM 10542	1050	2904	0.030	2.77
Sanguibacteraceae	1050	2904	0.030	2.77
Streptomyces ambofaciens	119	2462	0.008	20.69
Streptomyces ambofaciens ATCC 23877	102	2376	0.011	23.22
Streptomyces filamentosus	129	2932	0.005	22.75

Streptomyces ghanaensis	1628	18006	0.024	11.06
Streptomyces ghanaensis ATCC 14672	1628	18006	0.024	11.06
Streptomyces lividans	140	11775	0.047	83.99
Streptomyces lividans TK24	140	11737	0.048	83.72
Streptomyces roseosporus NRRL 15998	42	1571	0.014	37.66
unclassified Bradyrhizobiaceae	148	1347	0.001	9.11
Variovorax	53340	1559691	0.021	29.24
Variovorax paradoxus	53126	1556289	0.022	29.29
Variovorax paradoxus S110	25656	464220	0.019	18.09
Yersinia mollaretii	83	1585	0.020	19.05
Yersinia mollaretii ATCC 43969	83	1585	0.020	19.05

Table A14: Prokaryotic taxonomic groups selected only in the wheat and oat rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Wheat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**	**Oat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Acidovorax avenae	731	1582	0.019	2.16	1788	0.009	2.45
Acidovorax avenae subsp. citrulli	731	1582	0.019	2.16	1760	0.011	2.41
Acidovorax avenae subsp. citrulli AAC00-1	643	1582	0.004	2.46	1760	0.003	2.74
Aeromonadaceae	1420	3492	0.034	2.46	3537	0.040	2.49
Aeromonadales	1550	3920	0.019	2.53	3758	0.027	2.42
Asticcacaulis	7353	66639	0.006	9.06	42900	0.004	5.83
Bordetella avium	387	1291	0.011	3.34	1385	0.013	3.58
Bordetella avium 197N	387	1291	0.011	3.34	1385	0.013	3.58
Brevundimonas	6183	25567	0.001	4.14	31717	0.001	5.13
Brevundimonas sp. BAL3	1359	7324	0.001	5.39	7462	0.003	5.49

Burkholderiales bacterium 1_1_47	182	1037	0.036	5.71	1087	0.006	5.99
candidate division TM7 single-cell isolate TM7c	180	2212	0.006	12.28	3409	0.000	18.92
Caulobacter	57068	239981	0.013	4.21	599853	0.034	10.51
Caulobacter segnis	4139	14205	0.009	3.43	22092	0.036	5.34
Caulobacter segnis ATCC 21756	4139	14205	0.009	3.43	22092	0.036	5.34
Caulobacter sp. K31	22204	80785	0.040	3.64	218056	0.038	9.82
Caulobacteraceae	255099	815206	0.007	3.20	1364636	0.015	5.35
Caulobacterales	255099	815206	0.007	3.20	1364636	0.015	5.35
Chlorobium ferrooxidans	472	1511	0.025	3.20	1434	0.033	3.04
Chlorobium ferrooxidans DSM 13031	472	1511	0.025	3.20	1434	0.033	3.04
Chloroherpeton	2369	5256	0.001	2.22	4472	0.039	1.89
Chloroherpeton thalassium	2369	5256	0.001	2.22	4472	0.039	1.89
Chloroherpeton thalassium ATCC 35110	2369	5256	0.001	2.22	4472	0.039	1.89
Cryomorphaceae	688	3714	0.013	5.40	1608	0.046	2.34
Dokdonia	301	1347	0.020	4.48	1544	0.010	5.13
Dokdonia donghaensis	294	1316	0.023	4.47	1352	0.014	4.60
Dokdonia donghaensis MED134	294	1316	0.023	4.47	1352	0.014	4.60
Dyadobacter	45604	194022	0.013	4.25	219547	0.002	4.81
Dyadobacter fermentans	45604	194022	0.013	4.25	219547	0.002	4.81
Dyadobacter fermentans DSM 18053	45604	194022	0.013	4.25	219547	0.002	4.81
Enterococcaceae	536	1728	0.040	3.23	1597	0.035	2.98
Erythrobacter	9342	19586	0.048	2.10	19705	0.016	2.11
Erythrobacter sp. SD-21	1965	5361	0.013	2.73	6498	0.002	3.31
Erythrobacteraceae	9342	19605	0.048	2.10	19743	0.016	2.11
Flavobacteria bacterium MS024-2A	793	2996	0.003	3.78	1708	0.010	2.15
Herbaspirillum	93	555	0.025	5.95	899	0.038	9.65
Kingella	165	1040	0.002	6.30	708	0.018	4.29

Leptospira biflexa	1440	3294	0.005	2.29	2916	0.038	2.03
Leptospira biflexa serovar Patoc	1440	3213	0.009	2.23	2916	0.038	2.03
Methylibium	25224	59368	0.023	2.35	93545	0.007	3.71
Methylibium petroleiphilum	25224	59368	0.023	2.35	93545	0.007	3.71
Methylibium petroleiphilum PM1	25224	59368	0.023	2.35	93545	0.007	3.71
Novosphingobium	22604	45506	0.041	2.01	87547	0.002	3.87
Pedobacter	54572	286677	0.017	5.25	354466	0.000	6.50
Pedobacter heparinus	17790	74227	0.028	4.17	91421	0.001	5.14
Pedobacter heparinus DSM 2366	17790	74227	0.028	4.17	91421	0.001	5.14
Photobacterium	494	1489	0.002	3.02	1769	0.035	3.58
Polaribacter	602	3842	0.014	6.39	2084	0.007	3.46
Polaribacter irgensii	580	3823	0.014	6.59	2044	0.008	3.52
Polaribacter irgensii 23-P	580	3823	0.014	6.59	2044	0.008	3.52
Polaromonas	58810	129712	0.038	2.21	189521	0.027	3.22
Pseudoalteromonas tunicata	145	820	0.003	5.66	640	0.050	4.42
Pseudoalteromonas tunicata D2	145	820	0.003	5.66	640	0.050	4.42
Pseudomonas syringae group genomosp. 2	104	704	0.002	6.78	1592	0.037	15.33
Ralstonia solanacearum	1731	4595	0.000	2.65	3961	0.016	2.29
Ralstonia solanacearum GMI1000	135	991	0.009	7.36	967	0.016	7.18
Rheinheimera	465	2194	0.009	4.72	1242	0.024	2.67
Rheinheimera sp. E407-8	304	1998	0.007	6.57	1198	0.011	3.94
Rhodoferax	8731	19711	0.027	2.26	20701	0.010	2.37
Rhodoferax ferrireducens	8731	19711	0.027	2.26	20701	0.010	2.37
Rhodoferax ferrireducens T118	8731	19711	0.027	2.26	20701	0.010	2.37
Rikenellaceae	1820	5313	0.023	2.92	4452	0.043	2.45
Rubrivivax	20609	54166	0.009	2.63	77522	0.006	3.76
Sphingopyxis	15937	54778	0.019	3.44	41754	0.002	2.62

Sphingopyxis alaskensis	14479	51280	0.022	3.54	38071	0.002	2.63
Sphingopyxis alaskensis RB2256	14479	51280	0.022	3.54	38071	0.002	2.63
Taylorella	63	739	0.033	11.76	682	0.029	10.85
unclassified Burkholderiales	179229	504346	0.014	2.81	819285	0.004	4.57
unclassified Burkholderiales (miscellaneous)	214	1037	0.041	4.84	1087	0.006	5.08
unclassified Flavobacteria	2442	7708	0.001	3.16	5732	0.029	2.35
uncultured bacterium BLR10	979	7237	0.001	7.39	9640	0.020	9.84
uncultured bacterium BLR19	592	3846	0.001	6.49	3307	0.001	5.58
unidentified eubacterium SCB49	694	1804	0.008	2.60	1811	0.006	2.61
Xanthomonas campestris	1377	3118	0.013	2.27	3922	0.013	2.85

Table A15: Prokaryotic taxonomic groups selected only in the wheat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Wheat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**	**Pea mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Anabaena	820	2754	0.032	3.36	4850	0.040	5.92
Azotobacter	1246	3246	0.031	2.61	4018	0.032	3.22
Azotobacter group	1246	3246	0.031	2.61	4018	0.032	3.22
Azotobacter vinelandii	1239	2874	0.041	2.32	3969	0.034	3.20
Azotobacter vinelandii DJ	1239	2874	0.041	2.32	3969	0.034	3.20
Bacillus mucilaginosus	1202	3580	0.017	2.98	6672	0.008	5.55
Janthinobacterium lividum	1380	3614	0.034	2.62	10936	0.024	7.93
Neptuniibacter	288	1793	0.001	6.21	2245	0.048	7.78
Neptuniibacter caesariensis	288	1755	0.000	6.09	2245	0.048	7.78
uncultured bacterium pEAF66	7865	23924	0.012	3.04	46199	0.034	5.87

Table A16: Prokaryotic taxonomic groups selected only in the oat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Oat mean (transcripts / g)	P value	Fold change (vs soil)	Pea mean (transcripts / g)	P value	Fold change (vs soil)
Azospirillum sp. B510	15809	42908	0.033	2.71	88189	0.033	5.58
Clostridium botulinum	870	6183	0.011	7.10	2774	0.026	3.19
Comamonadaceae	631581	2326156	0.042	3.68	3961200	0.042	6.27
Coxiella burnetii RSA 334	154	2305	0.002	14.97	3720	0.000	24.16
Fluoribacter dumoffii	30	279	0.023	9.26	495	0.005	16.42
Hyphomonas	9022	31633	0.004	3.51	37261	0.049	4.13
Hyphomonas neptunium	9022	31633	0.004	3.51	37261	0.049	4.13
Hyphomonas neptunium ATCC 15444	9022	31633	0.004	3.51	37261	0.049	4.13
Legionellaceae	28560	87328	0.008	3.06	205369	0.026	7.19
Legionellales	35679	96134	0.014	2.69	219933	0.023	6.16
Magnetospirillum gryphiswaldense	7761	20801	0.014	2.68	33025	0.010	4.26
Magnetospirillum gryphiswaldense MSR-1	7696	20736	0.014	2.69	32915	0.010	4.28
Mycobacterium tuberculosis T17	22109	77432	0.018	3.50	144691	0.012	6.54
Paenibacillus sp. JDR-2	1299	15031	0.023	11.57	44319	0.036	34.13
Rhizobium etli 8C-3	0	3660	0.000	#DIV/0!	23369	0.021	Absent in soil
Rhizobium etli IE4771	312	1576	0.050	5.06	4963	0.012	15.93
Rhodococcus erythropolis	6294	21858	0.043	3.47	27833	0.047	4.42
Rhodococcus erythropolis PR4	483	2113	0.010	4.37	3207	0.015	6.64
Sphingomonas wittichii	14160	30226	0.025	2.13	64520	0.019	4.56
Sphingomonas wittichii RW1	14160	30226	0.025	2.13	64520	0.019	4.56
uncultured bacterium BLR18	416	1619	0.022	3.89	2880	0.045	6.93

Table A17: Prokaryotic taxonomic groups selected in wheat, oat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Wheat mean (transcripts / g)	P value	Fold change (vs soil)	Oat mean (transcripts / g)	P value	Fold change (vs soil)	Pea mean (transcripts / g)	P value	Fold change (vs soil)
Burkholderia phytofirmans	514	1929	0.042	3.76	1323	0.029	2.58	15364	0.046	29.91
Burkholderia thailandensis	528	2850	0.008	5.40	1496	0.007	2.83	5750	0.030	10.89
Fluoribacter	60	380	0.016	6.32	279	0.043	4.64	602	0.004	10.02
Janthinobacterium	15826	38926	0.001	2.46	50247	0.011	3.18	122980	0.046	7.77
Janthinobacterium sp. Marseille	14178	35311	0.001	2.49	45479	0.009	3.21	111400	0.049	7.86
Novosphingobium aromaticivorans	5743	12242	0.012	2.13	19119	0.004	3.33	28070	0.043	4.89
Ruegeria	524	1140	0.047	2.17	1042	0.045	1.99	2069	0.003	3.95
Sphingobacteriaceae	202182	936609	0.009	4.63	2478376	0.009	12.26	7451854	0.039	36.86

Table A18: Eukaryotic taxonomic groups selected only in the wheat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Wheat mean (transcripts / g)	P value	Fold change (vs soil)
Appendicularia	1827	3513	0.029	1.92
Aspergillus oryzae RIB40	0	930	0.043	Absent in soil
Cavosteliaceae	0	261	0.022	Absent in soil
Cercomonadida	1126	2560	0.049	2.27
Cryptosporidium muris	157	692	0.033	4.42
Cryptosporidium muris RN66	157	692	0.033	4.42
Culex	430	1496	0.018	3.48
Culex pipiens complex	430	1496	0.018	3.48

Culex quinquefasciatus	380	1496	0.016	3.93
Culicinae	1122	4159	0.023	3.71
Glomus mosseae	0	592	0.029	Absent in soil
Ochromonadaceae	883	3177	0.022	3.60
Ochromonadales	883	3177	0.022	3.60
Ochromonas	802	3177	0.021	3.96
Ochromonas danica	778	3139	0.024	4.03
Oikopleura	1827	3513	0.029	1.92
Oikopleura dioica	1827	3513	0.029	1.92
Oikopleuridae	1827	3513	0.029	1.92
Thraustochytriidae	427	1458	0.032	3.42
Trichostrongyloidea	49	480	0.004	9.88
Verticillium albo-atrum	134	750	0.031	5.62
Verticillium albo-atrum VaMs.102	134	750	0.031	5.62

Table A19: Eukaryotic taxonomic groups selected only in the oat rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Oat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Alternaria	113	782	0.029	6.93
Aspergillus	2597	9576	0.003	3.69
Aspergillus clavatus	137	941	0.007	6.87
Aspergillus clavatus NRRL 1	137	941	0.007	6.87
Aspergillus nidulans FGSC A4	382	997	0.017	2.61
Aspergillus terreus	690	1764	0.001	2.56
Aspergillus terreus NIH2624	690	1764	0.001	2.56

Caenorhabditis briggsae	196	2030	0.020	10.34
Emericella	382	997	0.017	2.61
Emericella nidulans	382	997	0.017	2.61
Eurotiomycetes	55337	289197	0.002	5.23
Fusarium oxysporum	42	801	0.007	18.88
Fusarium oxysporum species complex	42	801	0.007	18.88
mitosporic Emericella nidulans	382	997	0.017	2.61
mitosporic Pleosporaceae	113	816	0.028	7.24

Table A20: Eukaryotic taxonomic groups selected only in the pea rhizosphere compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Pea mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Acari	3786	15376	0.028	4.06
Ajellomyces	1195	12130	0.039	10.15
Ajellomycetaceae	1195	12130	0.039	10.15
Arachnida	4206	17587	0.024	4.18
Bilateria	582859	3327231	0.038	5.71
Bos	211	1095	0.047	5.20
Bos taurus	211	1095	0.047	5.20
Bovidae	224	1095	0.050	4.89
Bovinae	211	1095	0.047	5.20
Cetartiodactyla	781	4027	0.027	5.16
Chelicerata	4236	17839	0.022	4.21
Chytridiomycota	1700	9177	0.047	5.40
Colletotrichum	1623	6834	0.025	4.21

Colletotrichum higginsianum	1615	6557	0.018	4.06
Eukaryota	7360771	90916217	0.013	12.35
Eumalacostraca	672	3682	0.024	5.48
Eumetazoa	623679	3514239	0.038	5.63
Formicidae	553	6817	0.029	12.34
Heterodera	0	2166	0.007	Absent in soil
Heterodera glycines	0	2092	0.006	Absent in soil
Ixodes	1043	6582	0.022	6.31
Ixodes scapularis	1043	6533	0.023	6.27
Ixodida	1506	8285	0.018	5.50
Ixodidae	1192	7659	0.017	6.43
Ixodinae	1043	6582	0.022	6.31
Ixodoidea	1506	8285	0.018	5.50
Lasiosphaeriaceae	4000	12164	0.035	3.04
Malacostraca	672	3775	0.016	5.62
melanogaster subgroup	142	4136	0.039	29.20
Metazoa	700696	3764752	0.038	5.37
mitosporic Phyllachoraceae	1623	6834	0.025	4.21
mitosporic Tricholomataceae	662	2136	0.049	3.23
Moniliophthora	662	2136	0.049	3.23
Moniliophthora perniciosa	552	2100	0.035	3.81
Moniliophthora perniciosa FA553	552	2100	0.035	3.81
Naegleria	1364	4224	0.036	3.10
Neurospora	388	2873	0.040	7.41
Orthopteroidea	237	1729	0.020	7.29
Parasitiformes	3578	14499	0.033	4.05
Pecora	224	1095	0.050	4.89

Pezizales	7	232	0.011	35.00
Pezizomycetes	7	232	0.011	35.00
Phyllachoraceae	1636	6888	0.024	4.21
Podospora	3991	12146	0.035	3.04
Podospora anserina	3954	12121	0.036	3.07
Podospora anserina DSM 980	3635	11176	0.050	3.07
Raphidophyceae	42	814	0.034	19.60
Ruminantia	224	1095	0.050	4.89
Sarcoptiformes	42	834	0.046	19.99
Sclerotiniaceae	7554	24480	0.021	3.24
Sordariaceae	1993	6837	0.040	3.43
Trypanosoma cruzi strain CL Brener	757	2998	0.004	3.96
Tuber	7	141	0.015	21.20
Tuberaceae	7	141	0.015	21.20
Tylenchina	292	6139	0.024	21.02
Tylenchoidea	292	6139	0.024	21.02

Table A21: Eukaryotic taxonomic groups selected only in the wheat and oat rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Wheat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**	**Oat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Eurotiales	22242	277367	0.033	12.47	209775	0.001	9.43
Eurotiomycetidae	40939	347647	0.043	8.49	263532	0.001	6.44
mitosporic Trichocomaceae	11678	184268	0.033	15.78	135140	0.001	11.57
Panagrolaimoidea	33	1789	0.026	53.95	845	0.013	25.49
Penicillium	8349	144775	0.035	17.34	103091	0.000	12.35

Penicillium chrysogenum	7836	140902	0.035	17.98	99562	0.000	12.71
Penicillium chrysogenum complex	7836	140902	0.035	17.98	99562	0.000	12.71
Penicillium chrysogenum Wisconsin 54-1255	7836	140902	0.035	17.98	99534	0.000	12.70
Pyrenomonas	154	1875	0.003	12.18	2103	0.031	13.66
Pyrenomonas helgolandii	73	1860	0.002	25.49	2103	0.026	28.83
Trichocomaceae	22235	277150	0.033	12.46	209540	0.001	9.42

Table A22: Eukaryotic taxonomic groups selected only in the wheat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Wheat mean (transcripts / g)	P value	Fold change (vs soil)	Pea mean (transcripts / g)	P value	Fold change (vs soil)
Heteromita	393	2185	0.013	5.55	6415	0.036	16.31
Heteromita sp. PRA-74	393	2169	0.015	5.52	6415	0.036	16.31
Heteromitidae	473	2385	0.006	5.04	6524	0.038	13.79
Schizotrypanum	803	1967	0.044	2.45	3035	0.006	3.78
Trypanosoma cruzi	803	1967	0.044	2.45	3035	0.006	3.78

Table A23: Eukaryotic taxonomic groups selected only in the oat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	Soil mean (transcripts / g)	Oat mean (transcripts / g)	P value	Fold change (vs soil)	Pea mean (transcripts / g)	P value	Fold change (vs soil)
Aphelenchoididae	144	892	0.035	6.19	4185	0.039	29.05
Aphelenchoidoidea	144	892	0.035	6.19	4185	0.039	29.05
Bursaphelenchus	144	767	0.036	5.33	4185	0.039	29.05
Bursaphelenchus xylophilus	136	736	0.037	5.42	3429	0.042	25.27
Caenorhabditis brenneri	459	1940	0.049	4.23	7769	0.029	16.94

Heteroderidae	0	1483	0.033	#DIV/0!	2389	0.008	Absent in soil
Heteroderinae	0	1483	0.033	#DIV/0!	2389	0.008	Absent in soil
Tylenchida	605	4037	0.017	6.68	11768	0.023	19.46
Vespoidea	553	1003	0.038	1.82	6817	0.029	12.34

Table A24: Eukaryotic taxonomic groups selected in wheat, oat and pea rhizospheres compared to bulk soil (P≤0.05).

Taxonomic group	**Soil mean (transcripts / g)**	**Wheat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**	**Oat mean (transcripts / g)**	**P value**	**Fold change (vs soil)**	**Pea mean (transcripts / g)**	**P value**	**Fold change (vs soil)**
Euglenozoa	13567	51310	0.000	3.78	32849	0.036	2.42	72299	0.027	5.33
Kinetoplastida	11824	47241	0.000	4.00	30221	0.034	2.56	61199	0.016	5.18
Trypanosomatidae	11058	44407	0.000	4.02	28898	0.035	2.61	59385	0.017	5.37
Trypanosoma	6283	28538	0.000	4.54	21460	0.021	3.42	40907	0.011	6.51

Table A25: Functions induced in all three (wheat, oat and pea) rhizopspheres compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Wheat mean transcripts /g**	**P value**	**Fold change (vs soil)**	**Oat mean transcripts /g**	**P value**	**Fold change (vs soil)**	**Pea mean transcripts /g**	**P value**	**Fold change (vs soil)**
2,3,4,5-tetrahydropyridine-2,6-dicarboxylate N-succinyltransferase (EC 2.3.1.117)	1887	9547	0.000	5.1	9428	0.046	5.0	20163	0.031	10.7
3-oxoacyl-[acyl-carrier-protein] synthase, KASI (EC 2.3.1.41)	2883	9064	0.050	3.1	8069	0.033	2.8	21357	0.048	7.4
3-oxoadipate CoA-transferase subunit B (EC 2.8.3.6)	151	1618	0.024	10.7	2414	0.001	16.0	1822	0.032	12.1
6,7-dimethyl-8-ribityllumazine	1367	5524	0.013	4.0	8135	0.003	6.0	21642	0.045	15.8

synthase (EC 2.5.1.9)										
Acetyl-CoA synthetase (ADP-forming) alpha and beta chains, putative	1189	5267	0.028	4.4	4497	0.007	3.8	4902	0.015	4.1
Acetylglutamate kinase (EC 2.7.2.8)	3006	9540	0.018	3.2	10080	0.025	3.4	25917	0.023	8.6
Aerobic C4-dicarboxylate transporter for fumarate, L-malate, D-malate, succunate	1805	8945	0.005	5.0	8035	0.004	4.5	28038	0.047	15.5
Alkyl hydroperoxide reductase protein F (EC 1.6.4.-)	839	5351	0.017	6.4	16853	0.000	20.1	27767	0.021	33.1
Alpha-1,2-mannosidase	778	4301	0.004	5.5	4479	0.006	5.8	11833	0.013	15.2
Alpha-L-arabinofuranosidase II precursor (EC 3.2.1.55)	206	805	0.014	3.9	649	0.044	3.1	1976	0.002	9.6
Arginine-tRNA-protein transferase (EC 2.3.2.8)	662	2369	0.002	3.6	2754	0.001	4.2	3673	0.023	5.5
Aromatic ring-opening dioxygenase	441	1770	0.047	4.0	2233	0.005	5.1	3091	0.000	7.0
Beta-hexosaminidase (EC 3.2.1.52)	1046	5021	0.004	4.8	2850	0.027	2.7	8521	0.045	8.1
Carboxynorspermidine decarboxylase, putative (EC 4.1.1.-)	142	832	0.031	5.9	2887	0.023	20.4	4577	0.004	32.3
Chaperone protein HscB	268	2064	0.008	7.7	2979	0.011	11.1	7098	0.025	26.5
Copper chaperone	65	700	0.016	10.7	2118	0.001	32.5	7827	0.031	120.1
Copper-translocating P-type ATPase (EC 3.6.3.4)	775	3346	0.028	4.3	5303	0.042	6.8	10811	0.006	13.9
Cyanophycin synthase (EC 6.3.2.29)(EC 6.3.2.30)	2334	8021	0.023	3.4	11873	0.018	5.1	16511	0.038	7.1
Cysteine synthase A (EC 2.5.1.47)	50	774	0.019	15.5	745	0.008	14.9	415	0.014	8.3
Cytochrome c4	1236	3202	0.020	2.6	4585	0.003	3.7	6980	0.040	5.6
Cytochrome O ubiquinol oxidase subunit I (EC 1.10.3.-)	4610	18358	0.022	4.0	46938	0.003	10.2	177077	0.048	38.4
Cytochrome O ubiquinol oxidase subunit II (EC 1.10.3.-)	989	5747	0.013	5.8	16448	0.001	16.6	56736	0.020	57.4

Cytochrome O ubiquinol oxidase subunit IV (EC 1.10.3.-)	390	1651	0.039	4.2	4852	0.003	12.4	16161	0.033	41.5
Detection	4498	12805	0.025	2.8	19258	0.030	4.3	38202	0.018	8.5
Diaminopimelate epimerase (EC 5.1.1.7)	708	3818	0.014	5.4	3776	0.018	5.3	11620	0.027	16.4
Dihydrolipoamide dehydrogenase of pyruvate dehydrogenase complex (EC 1.8.1.4)	4553	23744	0.000	5.2	23627	0.001	5.2	56873	0.020	12.5
DNA mismatch repair protein MutL	2473	8416	0.022	3.4	8973	0.005	3.6	21102	0.050	8.5
Enoyl-CoA hydratase (EC 4.2.1.17)	4547	12326	0.041	2.7	13069	0.042	2.9	26561	0.044	5.8
Exodeoxyribonuclease III (EC 3.1.11.2)	2954	12264	0.027	4.2	11924	0.016	4.0	22313	0.035	7.6
Ferrous iron transport protein B	5263	11604	0.007	2.2	15024	0.020	2.9	24251	0.047	4.6
FIG003437: hypothetical with DnaJ-like domain	107	1241	0.012	11.6	2145	0.003	20.1	6448	0.020	60.3
FKBP-type peptidyl-prolyl cis-trans isomerase slpA (EC 5.2.1.8)	73	433	0.014	6.0	1107	0.021	15.2	2219	0.008	30.5
Flagellar basal-body rod modification protein FlgD	337	3783	0.005	11.2	3936	0.014	11.7	8759	0.031	26.0
Flagellar basal-body rod protein FlgC	2372	8435	0.008	3.6	11014	0.001	4.6	26690	0.037	11.3
Flagellar biosynthesis protein FliC	11194	45865	0.008	4.1	45465	0.003	4.1	150446	0.029	13.4
Flagellar motor rotation protein MotB	1198	8322	0.000	6.9	9762	0.003	8.1	23893	0.034	19.9
Flagellar motor switch protein FliM	5382	15697	0.010	2.9	18920	0.007	3.5	30652	0.024	5.7
Flagellar synthesis regulator FleN	15	977	0.023	65.2	979	0.005	65.3	3053	0.015	203.8
Flagellar transcriptional activator FlhC	2282	8813	0.010	3.9	13890	0.001	6.1	16118	0.018	7.1
Formyltetrahydrofolate deformylase (EC 3.5.1.10)	1364	4894	0.014	3.6	4247	0.017	3.1	13955	0.038	10.2
Functional role page for Anaerobic	284	1300	0.018	4.6	1050	0.021	3.7	1104	0.050	3.9

nitric oxide reductase transcription regulator NorR										
Glutamate--cysteine ligase (EC 6.3.2.2)	1522	6688	0.013	4.4	8829	0.003	5.8	15386	0.020	10.1
Glutamyl-tRNA reductase (EC 1.2.1.70)	4937	11002	0.044	2.2	16600	0.013	3.4	20157	0.020	4.1
Glutaredoxin 3	399	2911	0.020	7.3	5365	0.000	13.4	9873	0.012	24.7
Glutaredoxin-related protein	3683	14417	0.014	3.9	21152	0.002	5.7	45438	0.023	12.3
glutaryl-Coa dehydrogenase	11049	25826	0.028	2.3	28790	0.046	2.6	71913	0.043	6.5
Glutathione synthetase (EC 6.3.2.3)	628	2876	0.001	4.6	3255	0.006	5.2	9377	0.008	14.9
Guanylate kinase (EC 2.7.4.8)	1174	5812	0.011	4.9	7566	0.043	6.4	14185	0.028	12.1
Heme A synthase, cytochrome oxidase biogenesis protein Cox15-CtaA	289	1939	0.000	6.7	1804	0.013	6.2	3948	0.016	13.7
HflC protein	1553	5643	0.012	3.6	8395	0.002	5.4	19150	0.033	12.3
Histidinol dehydrogenase (EC 1.1.1.23)	2326	6372	0.034	2.7	4932	0.035	2.1	16176	0.028	7.0
HPr kinase/phosphorylase (EC 2.7.1.-) (EC 2.7.4.-)	226	3399	0.000	15.0	4295	0.000	19.0	4436	0.006	19.6
Large-conductance mechanosensitive channel	836	2768	0.000	3.3	2433	0.008	2.9	6573	0.042	7.9
L-asparaginase (EC 3.5.1.1)	997	3181	0.007	3.2	5147	0.010	5.2	7996	0.031	8.0
Lipid-A-disaccharide synthase (EC 2.4.1.182)	556	2557	0.040	4.6	2810	0.003	5.1	9463	0.050	17.0
Membrane fusion component of tripartite multidrug resistance system	250	4101	0.004	16.4	10281	0.022	41.2	33814	0.036	135.4
Membrane fusion protein of RND family multidrug efflux pump	15	613	0.003	40.9	1461	0.022	97.6	2532	0.024	169.1
Metal-dependent hydrolases of the beta-lactamase superfamily I; PhnP	335	1885	0.048	5.6	1378	0.046	4.1	2868	0.032	8.6

protein										
Molybdenum cofactor biosynthesis protein MoaE	304	1718	0.012	5.6	1490	0.001	4.9	4790	0.038	15.7
Murein-DD-endopeptidase (EC 3.4.99.-)	370	3190	0.000	8.6	8845	0.005	23.9	7365	0.020	19.9
Na+/H+ antiporter NhaA type	266	1879	0.027	7.1	1255	0.005	4.7	1548	0.046	5.8
NADH dehydrogenase subunit 1	11016	46298	0.041	4.2	129635	0.031	11.8	196552	0.040	17.8
NADH dehydrogenase subunit 2	12498	69788	0.040	5.6	188260	0.033	15.1	217728	0.002	17.4
NADH dehydrogenase subunit 4L	546	3938	0.001	7.2	11344	0.023	20.8	31825	0.018	58.3
NADH dehydrogenase subunit 6	1760	22771	0.029	12.9	98252	0.014	55.8	170689	0.002	97.0
NADH-ubiquinone oxidoreductase chain N (EC 1.6.5.3)	1496	8840	0.000	5.9	9938	0.001	6.6	22076	0.031	14.8
NADP-specific glutamate dehydrogenase (EC 1.4.1.4)	3744	13965	0.009	3.7	12279	0.029	3.3	37249	0.046	9.9
Nitrate/nitrite transporter	6390	17808	0.012	2.8	17657	0.050	2.8	28269	0.041	4.4
Octaprenyl-diphosphate synthase (EC 2.5.1.-)	158	1549	0.011	9.8	2896	0.001	18.4	5547	0.029	35.2
Organic hydroperoxide resistance protein	2156	8797	0.024	4.1	17797	0.001	8.3	44359	0.006	20.6
Paraquat-inducible protein B	0	230	0.000	Absent in soil	364	0.028	Absent in soil	1750	0.001	Absent in soil
Peptidyl-prolyl cis-trans isomerase ppiB (EC 5.2.1.8)	3051	7305	0.009	2.4	12902	0.013	4.2	27975	0.032	9.2
Periplasmic thiol:disulfide interchange protein DsbA	1835	7485	0.000	4.1	5102	0.008	2.8	12265	0.041	6.7
Phosphoribosyl-AMP cyclohydrolase (EC 3.5.4.19)	532	3224	0.011	6.1	3167	0.004	5.9	7579	0.004	14.2
Phosphoribosylanthranilate isomerase (EC 5.3.1.24)	360	1323	0.032	3.7	1473	0.038	4.1	1954	0.006	5.4
Phosphoribosylformylglycinamidine cyclo-ligase (EC 6.3.3.1)	5100	16434	0.045	3.2	18666	0.042	3.7	29367	0.038	5.8
Phosphoserine aminotransferase	7735	28604	0.033	3.7	35645	0.038	4.6	85104	0.021	11.0

(EC 2.6.1.52)										
Pirin-related protein, coexpressed with pyoverdine biosynthesis regulon	77	1105	0.021	14.4	2740	0.001	35.7	4659	0.021	60.6
Poly(A) polymerase (EC 2.7.7.19)	835	2430	0.021	2.9	3142	0.005	3.8	8882	0.035	10.6
Predicted L-rhamnose ABC transporter, ATP-binding component	224	1013	0.048	4.5	768	0.015	3.4	4101	0.019	18.3
Protein-L-isoaspartate O-methyltransferase (EC 2.1.1.77)	871	3147	0.005	3.6	2796	0.013	3.2	8832	0.035	10.1
Respiratory nitrate reductase alpha chain (EC 1.7.99.4)	11293	42661	0.010	3.8	33411	0.026	3.0	51291	0.025	4.5
Ribonucleotide reductase of class Ia (aerobic), beta subunit (EC 1.17.4.1)	11542	35441	0.001	3.1	40436	0.013	3.5	89077	0.046	7.7
Ribosomal large subunit pseudouridine synthase C (EC 4.2.1.70)	317	3219	0.029	10.2	3294	0.004	10.4	6171	0.005	19.5
Ribosomal RNA small subunit methyltransferase B (EC 2.1.1.-)	322	1538	0.020	4.8	1916	0.004	5.9	2356	0.020	7.3
Septum site-determining protein MinC	187	921	0.004	4.9	1087	0.000	5.8	1641	0.021	8.8
Signal peptidase I	7796	24102	0.048	3.1	25782	0.020	3.3	47492	0.017	6.1
Sporulation_gene_orphans	295	2383	0.018	8.1	2544	0.021	8.6	6555	0.000	22.2
Thiazole biosynthesis protein ThiH	134	1624	0.002	12.1	968	0.006	7.2	1476	0.002	11.0
Thiol:disulfide interchange protein DsbC	620	2450	0.024	4.0	5846	0.014	9.4	8214	0.035	13.2
TldD family protein, Actinobacterial subgroup	1158	5909	0.006	5.1	3597	0.038	3.1	7557	0.026	6.5
TPR repeat containing exported protein	163	2138	0.045	13.1	2467	0.003	15.1	10928	0.013	66.8
tungsten-containing formate	1755	7030	0.047	4.0	4705	0.022	2.7	10945	0.004	6.2

dehydrogenase alpha subunit										
UDP-N-acetylglucosamine--N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase (EC 2.4.1.227)	57	1049	0.049	18.3	1099	0.004	19.1	4877	0.036	84.9

Table A26: Functions induced only in wheat and oat rhizopspheres compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Wheat mean transcripts /g**	**P value**	**Fold change (vs soil)**	**Oat mean transcripts /g**	**P value**	**Fold change (vs soil)**
(3R)-hydroxymyristoyl-[acyl carrier protein] dehydratase (EC 4.2.1.-)	1268	6894	0.003	5.4	9089	0.016	7.2
1,4-dihydroxy-2-naphthoate octaprenyltransferase (EC 2.5.1.-)	94	628	0.043	6.7	935	0.008	9.9
2-polyprenylphenol hydroxylase and related flavodoxin oxidoreductases	641	2128	0.022	3.3	1686	0.029	2.6
3',5'-cyclic-nucleotide phosphodiesterase (EC 3.1.4.17)	38	791	0.030	21.1	398	0.046	10.6
3-carboxy-cis,cis-muconate cycloisomerase (EC 5.5.1.2)	82	799	0.008	9.7	1100	0.001	13.4
3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (EC 3.1.3.45)	32	786	0.038	24.8	777	0.019	24.5
3-deoxy-D-manno-octulosonic-acid transferase (EC 2.-.-.-)	78	392	0.047	5.0	631	0.022	8.1
4-hydroxybenzoate polyprenyltransferase (EC 2.5.1.-)	323	2154	0.001	6.7	2174	0.013	6.7
4-hydroxyproline epimerase (EC 5.1.1.8)	113	952	0.005	8.4	826	0.009	7.3
4-keto-6-deoxy-N-Acetyl-D-hexosaminyl-(Lipid carrier) aminotransferase	1176	2649	0.026	2.3	3012	0.048	2.6
5-formyltetrahydrofolate cyclo-ligase (EC 6.3.3.2)	189	1194	0.000	6.3	1060	0.002	5.6
5'-nucleotidase (EC 3.1.3.5)	632	3765	0.040	6.0	2445	0.002	3.9
5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase.	133	772	0.011	5.8	627	0.021	4.7
6-phosphofructokinase (EC 2.7.1.11)	4573	11320	0.003	2.5	12277	0.022	2.7
Acetyl-coenzyme A carboxyl transferase beta chain (EC 6.4.1.2)	8277	20741	0.006	2.5	24963	0.016	3.0
ADP-heptose--lipooligosaccharide heptosyltransferase II (EC 2.4.1.-)	102	727	0.013	7.1	800	0.005	7.9
Alkanesulfonate utilization operon LysR-family regulator CbI	188	1098	0.017	5.8	2017	0.000	10.7
Alkylphosphonate utilization operon protein PhnA	402	2714	0.040	6.7	2479	0.005	6.2
Alpha-L-fucosidase (EC 3.2.1.51)	980	3889	0.000	4.0	3070	0.031	3.1
Alpha-xylosidase (EC 3.2.1.-)	251	2088	0.006	8.3	2798	0.028	11.1

Aromatic-amino-acid aminotransferase (EC 2.6.1.57)	1214	8452	0.001	7.0	12451	0.003	10.3
Assimilatory nitrate reductase large subunit (EC:1.7.99.4)	2037	9005	0.000	4.4	6547	0.019	3.2
ATP synthase B chain (EC 3.6.3.14)	5570	14450	0.023	2.6	25326	0.022	4.5
ATP synthase delta chain (EC 3.6.3.14)	3126	11274	0.025	3.6	21021	0.022	6.7
ATPase provides energy for both assembly of type IV secretion complex and secretion of T-DNA complex (VirB11)	8	140	0.033	16.8	397	0.033	47.5
ATP-dependent DNA helicase pcrA (EC 3.6.1.-)	62	535	0.009	8.7	593	0.015	9.6
Benzoyl-CoA oxygenase component A	274	733	0.041	2.7	3670	0.008	13.4
Beta-carotene ketolase (EC 1.14.-.-)	2364	7091	0.039	3.0	4747	0.029	2.0
Beta-ketoadipyl CoA thiolase (EC 2.3.1.-)	105	785	0.003	7.5	1065	0.035	10.1
Biphenyl-2,3-diol 1,2-dioxygenase (EC 1.13.11.39)	654	3577	0.026	5.5	8180	0.002	12.5
CBSS-562.2.peg.5158_SK3_including	205	1133	0.007	5.5	981	0.001	4.8
Cell division initiation protein DivIVA	57	623	0.022	10.8	617	0.008	10.8
Cell division protein FtsZ (EC 3.4.24.-)	8949	23172	0.018	2.6	27332	0.015	3.1
Chemotaxis regulator - transmits chemoreceptor signals to flagelllar motor components CheY	11080	26784	0.018	2.4	40601	0.002	3.7
Chemotaxis response - phosphatase CheZ	1653	7575	0.000	4.6	11364	0.003	6.9
Chemotaxis response regulator protein-glutamate methylesterase CheB (EC 3.1.1.61)	1106	7612	0.007	6.9	7933	0.023	7.2
Circadian input kinase A	963	4269	0.005	4.4	4365	0.008	4.5
Cobalt-precorrin-8x methylmutase (EC 5.4.1.2)	8	183	0.042	21.9	387	0.018	46.4
COG3178: Predicted phosphotransferase related to Ser/Thr protein kinases	81	1307	0.025	16.1	1093	0.010	13.5
Copper-containing nitrite reductase (EC 1.7.2.1)	9310	31716	0.011	3.4	33930	0.008	3.6
Coproporphyrinogen III oxidase, aerobic (EC 1.3.3.3)	1890	6440	0.000	3.4	7489	0.024	4.0
Cystathionine gamma-lyase (EC 4.4.1.1)	3059	9323	0.004	3.0	10042	0.007	3.3
Cytochrome b6-f complex iron-sulfur subunit PetC1 (Rieske iron sulfur protein EC 1.10.99.1)	13	200	0.017	15.1	232	0.024	17.5
Cytochrome c-552 precursor	75	1131	0.007	15.0	2830	0.001	37.5
Cytochrome c553	372	1822	0.001	4.9	2588	0.020	7.0
Cytochrome c-type biogenesis protein Ccs1/ResB	181	1700	0.010	9.4	2278	0.001	12.6
Cytochrome c-type biogenesis protein ResA	97	708	0.021	7.3	2341	0.030	24.0
Cytochrome oxidase biogenesis protein Cox11-CtaG, copper delivery to Cox1	1207	4931	0.034	4.1	6183	0.005	5.1

D(-)-3-hydroxybutyrate oligomer hydrolase (EC 3.1.1.22)	29	238	0.047	8.1	188	0.013	6.5
DNA recombination-dependent growth factor C	205	1133	0.007	5.5	981	0.001	4.8
DNA topoisomerase III (EC 5.99.1.2)	8839	19751	0.003	2.2	20794	0.013	2.4
DNA-3-methyladenine glycosylase II (EC 3.2.2.21)	541	2360	0.012	4.4	1957	0.010	3.6
DNA-binding heavy metal response regulator	218	1131	0.028	5.2	1468	0.028	6.7
DNA-binding protein HU-alpha	1198	3025	0.018	2.5	3377	0.008	2.8
D-xylose proton-symporter XylE	94	850	0.025	9.0	869	0.012	9.2
Electron transport complex protein RnfC	0	257	0.022	Absent in soil	292	0.038	Absent in soil
Enolase (EC 4.2.1.11)	27189	65755	0.035	2.4	65206	0.047	2.4
Excinuclease ABC subunit A, dimeric form	2774	7547	0.033	2.7	8238	0.005	3.0
Export ABC transporter ATP-binding protein	605	2486	0.001	4.1	2080	0.032	3.4
FIG134348: essential endopeptidase	1030	4738	0.044	4.6	7158	0.022	6.9
Flagellar biosynthesis protein FlhA	6166	16513	0.020	2.7	19552	0.020	3.2
Flagellar biosynthesis protein FliP	1545	5503	0.003	3.6	6344	0.024	4.1
Flagellar hook-associated protein FlgK	478	2391	0.018	5.0	3431	0.012	7.2
Flagellar L-ring protein FlgH	475	3975	0.002	8.4	4677	0.020	9.8
Flagellar motor switch protein FliN	1669	6162	0.007	3.7	7581	0.006	4.5
Flagellar M-ring protein FliF	1795	6268	0.005	3.5	9643	0.021	5.4
Flagellar P-ring protein FlgI	1299	5917	0.015	4.6	6985	0.044	5.4
Flagellin protein FlaA	25929	81816	0.005	3.2	171622	0.048	6.6
Flagellin protein FlaB	16445	47338	0.036	2.9	48004	0.027	2.9
Flavodoxin	582	2011	0.021	3.5	3040	0.023	5.2
Flavohaemoglobin	20082	46230	0.035	2.3	65784	0.030	3.3
Flp pilus assembly protein, pilin Flp	118	615	0.015	5.2	1102	0.004	9.3
Galactokinase (EC 2.7.1.6)	28	860	0.040	30.4	630	0.032	22.3
Galactose-1-phosphate uridylyltransferase (EC 2.7.7.10)	554	2056	0.001	3.7	2991	0.024	5.4
GatB protein	242	1790	0.011	7.4	1500	0.027	6.2
Glutamine amidotransferase, class-II	1023	4421	0.032	4.3	4441	0.020	4.3
Glutathione S-transferase, unnamed subgroup 2 (EC 2.5.1.18)	0	173	0.050	Absent in soil	206	0.040	Absent in soil
Glycerol-3-phosphate dehydrogenase [NAD(P)+] (EC 1.1.1.94)	1025	3336	0.044	3.3	2953	0.019	2.9
Glycine dehydrogenase [decarboxylating] (glycine cleavage system P2 protein) (EC 1.4.4.2)	2736	7007	0.006	2.6	5525	0.030	2.0

Glycolate dehydrogenase (EC 1.1.99.14), iron-sulfur subunit GlcF	1067	2416	0.042	2.3	2373	0.045	2.2
Gram-Positive cell wall components	551	2841	0.044	5.2	3503	0.043	6.4
Group 2 RNA polymerase sigma factor	99	850	0.019	8.6	907	0.002	9.1
Homolog of fucose/glucose/galactose permeases	28	173	0.014	6.1	265	0.020	9.4
Hypothetical protein i Rubrerythrin cluster	98	1194	0.030	12.2	1213	0.000	12.3
Hypothetical transmembrane protein coupled to NADH-ubiquinone oxidoreductase chain 5 homolog	492	1359	0.044	2.8	3367	0.047	6.8
Intracellular PHB depolymerase (EC 3.1.1.-)	3635	11100	0.004	3.1	14640	0.000	4.0
Iron-sulfur cluster regulator IscR	1524	10519	0.012	6.9	31994	0.006	21.0
Isocitrate dehydrogenase [NAD] (EC 1.1.1.41)	1874	6691	0.022	3.6	6136	0.003	3.3
Kynurenine 3-monooxygenase (EC 1.14.13.9)	333	890	0.008	2.7	1436	0.016	4.3
L-arabinose isomerase (EC 5.3.1.4)	152	3590	0.015	23.6	2438	0.013	16.0
Lipid carrier : UDP-N-acetylgalactosaminyltransferase (EC 2.4.1.-)	330	1648	0.033	5.0	1302	0.014	3.9
low-specificity D-threonine aldolase	14	193	0.006	13.6	388	0.044	27.3
LSU ribosomal protein L18p (L5e)	31043	84474	0.035	2.7	131255	0.012	4.2
LSU ribosomal protein L22p (L17e)	66107	170721	0.043	2.6	271670	0.013	4.1
LSU ribosomal protein L25p	37206	87649	0.046	2.4	124986	0.016	3.4
LSU ribosomal protein L29p (L35e)	37888	112997	0.046	3.0	164525	0.016	4.3
LSU ribosomal protein L4p (L1e)	91350	234310	0.029	2.6	340742	0.024	3.7
Malate dehydrogenase (EC 1.1.1.37)	26092	58007	0.009	2.2	69691	0.020	2.7
Mannose-1-phosphate guanylyltransferase (GDP) (EC 2.7.7.22)	1781	4616	0.001	2.6	5429	0.035	3.0
Methyl-accepting chemotaxis protein I (serine chemoreceptor protein)	3668	11646	0.010	3.2	16707	0.001	4.6
Monofunctional biosynthetic peptidoglycan transglycosylase (EC 2.4.2.-)	589	2055	0.011	3.5	3465	0.000	5.9
MSHA biogenesis protein MshL	204	696	0.021	3.4	1091	0.039	5.3
Muconate_lactonizing_enzyme_family	285	1503	0.017	5.3	1019	0.031	3.6
N-acetylglucosamine-regulated TonB-dependent outer membrane receptor	8	344	0.016	41.3	542	0.022	65.0
N-acetylglutamate synthase (EC 2.3.1.1)	585	1899	0.049	3.2	2927	0.013	5.0
N-acetylmuramic acid 6-phosphate etherase (EC 4.2.-.-)	500	1884	0.003	3.8	1857	0.032	3.7
N-acylamino acid racemase	23	289	0.009	12.7	490	0.032	21.5
NAD-dependent formate dehydrogenase beta subunit	1080	3016	0.018	2.8	2331	0.033	2.2
Nitrate ABC transporter, ATP-binding protein	191	3541	0.011	18.5	2966	0.002	15.5
Nitric-oxide reductase (EC 1.7.99.7), quinol-dependent	19351	42498	0.038	2.2	54818	0.031	2.8

NnrS protein involved in response to NO	88	859	0.001	9.7	1229	0.014	13.9
OpcA, an allosteric effector of glucose-6-phosphate dehydrogenase, actinobacterial	53	578	0.007	11.0	713	0.020	13.5
Oxaloacetate decarboxylase alpha chain (EC 4.1.1.3)	7	229	0.005	34.6	1059	0.007	159.7
Peptide transport system permease protein sapC (TC 3.A.1.5.5)	459	1244	0.016	2.7	1946	0.029	4.2
Phenazine biosynthesis protein PhzF	596	2091	0.030	3.5	1506	0.036	2.5
Phosphoenolpyruvate synthase (EC 2.7.9.2)	13530	33273	0.020	2.5	36467	0.030	2.7
Phospholipid-lipopolysaccharide ABC transporter	145	565	0.050	3.9	615	0.025	4.2
Phosphomethylpyrimidine kinase (EC 2.7.4.7)	325	1731	0.000	5.3	1421	0.018	4.4
Phospho-N-acetylmuramoyl-pentapeptide-transferase (EC 2.7.8.13)	2259	8735	0.019	3.9	5933	0.023	2.6
Phosphonate ABC transporter phosphate-binding periplasmic component (TC 3.A.1.9.1)	243	1142	0.032	4.7	1426	0.003	5.9
Phosphoribosylamine--glycine ligase (EC 6.3.4.13)	3432	10203	0.043	3.0	11195	0.018	3.3
Phosphoribosylformimino-5-aminoimidazole carboxamide ribotide isomerase (EC 5.3.1.16)	1250	4687	0.008	3.7	6424	0.006	5.1
Phosphoribosylglycinamide formyltransferase (EC 2.1.2.2)	1063	2779	0.020	2.6	3021	0.007	2.8
Phosphoribosylglycinamide formyltransferase 2 (EC 2.1.2.-)	689	1919	0.025	2.8	2528	0.003	3.7
p-hydroxycinnamoyl CoA hydratase/lyase	102	528	0.033	5.2	403	0.020	4.0
Pole remodelling regulatory diguanylate cyclase	558	2407	0.011	4.3	2693	0.009	4.8
Polyhydroxyalkanoate granule-associated protein PhaF	1071	3476	0.016	3.2	6196	0.000	5.8
Positive regulator of CheA protein activity (CheW)	3992	11073	0.030	2.8	15230	0.039	3.8
Predicted Lactate-responsive regulator, IclR family	71	292	0.038	4.1	1046	0.020	14.6
Predicted maltose transporter MalT	231	1996	0.003	8.7	961	0.034	4.2
Primosomal replication protein N	2388	7841	0.027	3.3	9968	0.031	4.2
probable cytochrome oxidase (cbb3-type)	7	515	0.020	77.6	140	0.041	21.1
Probable valine-pyruvate aminotransferase (EC 2.6.1.66)	347	1800	0.001	5.2	2720	0.010	7.8
Programmed cell death toxin YdcE	48	508	0.039	10.5	753	0.029	15.6
Protein of unknown function Smg	281	1135	0.005	4.0	2090	0.001	7.4
Protocatechuate 4,5-dioxygenase alpha chain (EC 1.13.11.8)	130	1288	0.000	9.9	2082	0.002	16.0
Protocatechuate 4,5-dioxygenase beta chain (EC 1.13.11.8)	844	3937	0.007	4.7	5147	0.043	6.1
Pro-zeta-carotene desaturase, prolycopene producing (EC 1.-.-.-)	0	721	0.022	Absent in soil	2255	0.029	Absent in soil
Putative diheme cytochrome c-553	614	2686	0.003	4.4	3389	0.004	5.5

Quinolinate phosphoribosyltransferase [decarboxylating] (EC 2.4.2.19)	1130	2613	0.033	2.3	2885	0.046	2.6
Radical SAM domain heme biosynthesis protein	23	1882	0.016	80.7	3174	0.039	136.1
Rhamnogalacturonides degradation protein RhiN	115	1545	0.006	13.4	814	0.022	7.1
Ribulokinase (EC 2.7.1.16)	332	2384	0.003	7.2	1321	0.009	4.0
RNA polymerase sigma factor for flagellar operon	7667	15959	0.016	2.1	18954	0.050	2.5
S-adenosyl-methyltransferase MraW (EC 2.1.1.-)	740	3828	0.003	5.2	4159	0.004	5.6
Serine-protein kinase rsbW (EC 2.7.11.1)	146	1311	0.017	9.0	1330	0.021	9.1
Serine-pyruvate aminotransferase/archaeal aspartate aminotransferase	243	978	0.018	4.0	744	0.020	3.1
Signal recognition particle receptor protein FtsY (=alpha subunit) (TC 3.A.5.1.1)	4670	12781	0.021	2.7	15895	0.005	3.4
Spermidine synthase (EC 2.5.1.16)	448	1914	0.042	4.3	2339	0.008	5.2
SSU ribosomal protein S5p (S2e)	78280	193006	0.032	2.5	286325	0.017	3.7
SSU ribosomal protein S9p (S16e)	34166	87686	0.035	2.6	124745	0.020	3.7
Succinate dehydrogenase cytochrome b-556 subunit	974	6156	0.033	6.3	7886	0.006	8.1
Succinate dehydrogenase flavoprotein subunit (EC 1.3.99.1)	45305	106230	0.033	2.3	104011	0.029	2.3
Succinyl-CoA ligase [ADP-forming] alpha chain (EC 6.2.1.5)	26339	56468	0.045	2.1	67851	0.039	2.6
Sucrose-6-phosphate hydrolase (EC 3.2.1.26)	400	6919	0.015	17.3	5863	0.000	14.7
Sulfate_reduction-associated_complexes	416	1675	0.000	4.0	2088	0.002	5.0
Sulfide dehydrogenase [flavocytochrome C] flavoprotein chain precursor (EC 1.8.2.-)	235	1623	0.018	6.9	1529	0.014	6.5
Sulfite reduction-associated complex DsrMKJOP iron-sulfur protein DsrO (=HmeA)	147	799	0.044	5.4	1359	0.001	9.3
sulfur oxidation protein SoxA	107	561	0.030	5.3	680	0.003	6.4
Sulfur oxidation protein SoxX	40	361	0.008	9.1	252	0.037	6.3
Sulfur_oxidation	5933	10591	0.023	1.8	10133	0.047	1.7
Teichoic acid export ATP-binding protein TagH (EC 3.6.3.40)	109	435	0.042	4.0	447	0.020	4.1
Thiamine-monophosphate kinase (EC 2.7.4.16)	716	2563	0.039	3.6	2414	0.013	3.4
Thiol peroxidase, Tpx-type (EC 1.11.1.15)	173	1577	0.006	9.1	1064	0.010	6.1
Toxins and superantigens	605	2486	0.001	4.1	2080	0.032	3.4
TPR domain protein, putative component of TonB system	1163	8191	0.010	7.0	2665	0.049	2.3
trans-cinnamate 4-monooxygenase	0	2412	0.003	Absent in soil	6832	0.000	Absent in soil
Transcription accessory protein (S1 RNA-binding domain)	1267	4395	0.031	3.5	3464	0.022	2.7
Transketolase, C-terminal section (EC 2.2.1.1)	7195	17295	0.050	2.4	15132	0.048	2.1

tRNA(Ile)-lysidine synthetase	79	702	0.031	8.9	836	0.028	10.6
tRNA-guanine transglycosylase (EC 2.4.2.29)	3221	13344	0.011	4.1	11563	0.008	3.6
Type I secretion outer membrane protein, TolC precursor	628	1853	0.002	3.0	2773	0.009	4.4
Type II/IV secretion system protein TadC, associated with Flp pilus assembly	717	2431	0.011	3.4	2668	0.014	3.7
Type II/IV secretion system secretin RcpA/CpaC, associated with Flp pilus assembly	399	2331	0.017	5.8	3058	0.025	7.7
Type IV fimbrial assembly protein PilC	5869	12836	0.048	2.2	13492	0.026	2.3
Type IV fimbrial assembly, ATPase PilB	13250	33158	0.012	2.5	27727	0.006	2.1
Type IV pilin PilA	492	2518	0.001	5.1	2275	0.000	4.6
Type IV pilus biogenesis protein PilM	3360	11875	0.030	3.5	8841	0.029	2.6
Type IV pilus biogenesis protein PilO	1844	4603	0.013	2.5	3715	0.024	2.0
Tyrosine-protein kinase Wzc (EC 2.7.10.2)	241	1331	0.001	5.5	3128	0.010	13.0
UDP-N-acetylglucosamine 2-epimerase (EC 5.1.3.14)	2485	7985	0.017	3.2	8430	0.016	3.4
Undecaprenyl-phosphate N-acetylglucosaminyl 1-phosphate transferase (EC 2.7.8.-)	334	1219	0.032	3.6	1273	0.013	3.8
Universal stress protein family	875	2847	0.004	3.3	3716	0.032	4.2
Universal_stress_protein_family	949	3358	0.013	3.5	4361	0.017	4.6
UPF0246 protein YaaA	283	1061	0.013	3.7	1255	0.013	4.4
Xanthine-guanine phosphoribosyltransferase (EC 2.4.2.22)	378	2035	0.001	5.4	2393	0.042	6.3
YaaA	283	1079	0.015	3.8	1289	0.010	4.6
YgjD_and_YeaZ	1238	6029	0.042	4.9	7978	0.018	6.4

Table A27: Functions induced only in wheat and pea rhizopspheres compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Wheat mean transcripts /g**	**P value**	**Fold change (vs soil)**	**Pea mean transcripts /g**	**P value**	**Fold change (vs soil)**
2-keto-3-deoxygluconate permease (KDG permease)	0	350	0.028	Absent in soil	428	0.048	Absent in soil
2-keto-4-pentenoate hydratase (EC 4.2.1.-)	7	286	0.002	43.2	660	0.006	99.5
3-deoxy-manno-octulosonate cytidylyltransferase (EC 2.7.7.38)	128	1501	0.000	11.7	1759	0.042	13.8
4-hydroxy-2-oxovalerate aldolase (EC 4.1.3.-)	191	1024	0.000	5.4	1593	0.007	8.4

Chemotaxis protein methyltransferase CheR (EC 2.1.1.80)	5994	13733	0.018	2.3	27965	0.027	4.7
Cyclic AMP receptor protein	287	1283	0.002	4.5	967	0.014	3.4
Cytochrome c-type protein NapC	334	1969	0.036	5.9	2575	0.020	7.7
CytR_regulation	287	1457	0.003	5.1	1167	0.002	4.1
DNA ligase (EC 6.5.1.2)	1952	7714	0.016	4.0	11753	0.036	6.0
DNA replication protein DnaC	72	1100	0.011	15.2	5279	0.031	73.1
Gamma-glutamyltranspeptidase (EC 2.3.2.2)	3794	9020	0.050	2.4	13033	0.037	3.4
Glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21)	576	3047	0.003	5.3	5260	0.046	9.1
GMP reductase	76	547	0.022	7.2	1254	0.000	16.5
L-allo-threonine aldolase	70	977	0.004	14.0	2014	0.026	28.8
LysR family transcriptional regulator YbhD	7	176	0.000	26.5	424	0.038	63.9
mandelate racemase/muconate lactonizing enzyme family protein	529	2728	0.005	5.2	7494	0.012	14.2
Molybdopterin-guanine dinucleotide biosynthesis protein MobB	106	438	0.004	4.1	382	0.008	3.6
NADH-ubiquinone oxidoreductase chain J (EC 1.6.5.3)	873	2924	0.013	3.4	9580	0.036	11.0
Nucleoside 5-triphosphatase RdgB (dHAPTP, dITP, XTP-specific) (EC 3.6.1.15)	370	1179	0.019	3.2	2312	0.016	6.3
Phytoene dehydrogenase (EC 1.14.99.-)	264	2277	0.000	8.6	2189	0.049	8.3
Possible hypoxanthine oxidase XdhD (EC 1.-.-.-)	495	1409	0.036	2.8	4235	0.011	8.6
Probable poly(beta-D-mannuronate) O-acetylase (EC 2.3.1.-)	2013	6294	0.001	3.1	6861	0.030	3.4
Ribosome small subunit-stimulated GTPase EngC	332	2396	0.030	7.2	5267	0.016	15.9
Rossmann fold nucleotide-binding protein Smf possibly involved in DNA uptake	46	542	0.006	11.9	894	0.024	19.6
Serine phosphatase RsbU, regulator of sigma subunit	363	2624	0.014	7.2	10697	0.036	29.5
Stage IV sporulation protein A	0	614	0.027	Absent in soil	2951	0.019	Absent in soil
Threonine dehydratase, catabolic (EC 4.3.1.19)	476	1382	0.000	2.9	2949	0.014	6.2
Transcriptional (co)regulator CytR	0	174	0.033	Absent in soil	200	0.040	Absent in soil
Transcriptional regulator BkdR of isoleucine and valine catabolism operon	309	881	0.011	2.9	2996	0.040	9.7
Transcriptional regulator, Crp/Fnr family	16001	33060	0.039	2.1	79877	0.028	5.0
UDP-N-acetylmuramate--alanine ligase (EC 6.3.2.8)	2664	5406	0.048	2.0	11320	0.048	4.2

Table A28: Functions induced only in oat and pea rhizopspheres compared to soil (P≤0.05).

Assignment in the SEED database	Soil mean transcripts / g	Oat mean transcripts /g	P value	Fold change (vs soil)	Pea mean transcripts /g	P value	Fold change (vs soil)
1-phosphofructokinase (EC 2.7.1.56)	0	1025	0.016	Absent in soil	1462	0.022	Absent in soil
2-hydroxy-3-oxopropionate reductase (EC 1.1.1.60)	1388	4181	0.033	3.0	6270	0.039	4.5
2-hydroxychromene-2-carboxylate isomerase	229	1460	0.017	6.4	3378	0.005	14.8
2-Oxobutyrate dehydrogenase E1 (EC:1.2.4.1)	14	132	0.030	9.3	729	0.035	51.3
4-hydroxybenzoyl-CoA thioesterase family active site	471	3188	0.018	6.8	3952	0.042	8.4
50S ribosomal subunit maturation GTPase RbgA (B. subtilis YlqF)	40	965	0.003	24.3	1474	0.029	37.0
5-keto-D-gluconate 5-reductase (EC 1.1.1.69)	134	471	0.028	3.5	3015	0.004	22.5
5-Methyltetrahydrofolate--homocysteine methyltransferase	7457	19138	0.011	2.6	31290	0.017	4.2
ADP-ribose pyrophosphatase (EC 3.6.1.13)	350	827	0.032	2.4	3418	0.023	9.8
Alkyl hydroperoxide reductase protein C (EC 1.6.4.-)	23132	203319	0.002	8.8	343326	0.042	14.8
Amidophosphoribosyltransferase (EC 2.4.2.14)	30779	79158	0.042	2.6	149632	0.050	4.9
Arsenate reductase (EC 1.20.4.1)	1512	4564	0.041	3.0	13301	0.026	8.8
ATP phosphoribosyltransferase catalytic subunit (EC 2.4.2.17)	818	4235	0.013	5.2	5202	0.023	6.4
ATP-dependent Clp protease adaptor protein ClpS	13521	36569	0.028	2.7	93174	0.033	6.9
ATP-dependent protease HslV (EC 3.4.25.-)	925	6591	0.017	7.1	15952	0.015	17.3

ATP-dependent RNA helicase RhlB	1220	2400	0.049	2.0	3861	0.005	3.2
Bacterial cytostatics, differentiation factors and antibiotics	9246	23239	0.013	2.5	46357	0.042	5.0
Benzoylformate decarboxylase (EC 4.1.1.7)	606	1088	0.027	1.8	1799	0.003	3.0
Beta-glucanase precursor (EC 3.2.1.73)	213	1163	0.048	5.5	3540	0.019	16.6
Carbon monoxide dehydrogenase F protein	168	822	0.023	4.9	2640	0.007	15.7
CBSS-261594.1.peg.2640	1536	6342	0.024	4.1	12475	0.001	8.1
Cell division protein MraZ	1052	4533	0.002	4.3	16368	0.027	15.6
COG0779: clustered with transcription termination protein NusA	2437	12788	0.006	5.2	32059	0.023	13.2
Cold-shock DEAD-box protein A	15114	88724	0.024	5.9	177825	0.044	11.8
Cu(I)-responsive transcriptional regulator	103	1212	0.024	11.8	1409	0.017	13.7
Cysteine synthase B (EC 2.5.1.47)	2858	15920	0.020	5.6	35467	0.027	12.4
Cytochrome O ubiquinol oxidase subunit III (EC 1.10.3.-)	1193	9402	0.002	7.9	29951	0.040	25.1
Cytochrome oxidase biogenesis protein Sco1/SenC/PrrC, putative copper metallochaperone	667	4417	0.033	6.6	7917	0.024	11.9
decaprenyl diphosphate synthase	13	246	0.043	18.5	212	0.027	16.0
D-galactonate transporter	59	628	0.002	10.7	2604	0.014	44.4
D-glucarate permease	31	116	0.022	3.7	968	0.031	31.2
Dihydroneopterin aldolase (EC 4.1.2.25)	83	1800	0.014	21.8	1753	0.017	21.2
DNA polymerase I (EC 2.7.7.7)	4101	12907	0.004	3.1	31583	0.037	7.7
Enoyl-[acyl-carrier-protein] reductase [NADPH] (EC 1.3.1.10)	238	2221	0.001	9.3	2664	0.002	11.2
Eukaryotic translation initiation factor 5	98	3263	0.000	33.2	8970	0.000	91.3

Ferrochelatase, protoheme ferro-lyase (EC 4.99.1.1)	752	5479	0.000	7.3	8227	0.010	10.9
Formate dehydrogenase chain D (EC 1.2.1.2)	282	1271	0.015	4.5	3602	0.007	12.8
Fructose ABC transporter, ATP-binding component FrcA	1044	5999	0.045	5.7	3955	0.012	3.8
Fructose ABC transporter, substrate-binding component FrcB	1070	14347	0.014	13.4	15483	0.024	14.5
Fumarate and nitrate reduction regulatory protein	824	4047	0.049	4.9	12255	0.002	14.9
Glutamine amidotransferase chain of NAD synthetase	4735	9914	0.047	2.1	16416	0.038	3.5
Glutathione S-transferase (EC 2.5.1.18)	5933	17680	0.032	3.0	49142	0.029	8.3
Glutathione S-transferase, unnamed subgroup (EC 2.5.1.18)	220	852	0.048	3.9	3100	0.044	14.1
Glycine oxidase ThiO (EC 1.4.3.19)	22	553	0.028	25.6	699	0.025	32.3
GTP-binding protein HflX	6335	28568	0.003	4.5	64052	0.029	10.1
Heat shock protein GrpE	2299	14235	0.034	6.2	33066	0.019	14.4
Hemoglobin-like protein HbO	204	1590	0.018	7.8	2457	0.027	12.0
hypothetical protein that often co-occurs with aconitase	340	1350	0.019	4.0	6462	0.003	19.0
Inner membrane component of tripartite multidrug resistance system	2726	16836	0.020	6.2	48649	0.016	17.8
Iron-regulated protein A precursor	164	756	0.001	4.6	2858	0.046	17.4
LSU ribosomal protein L31p	75628	175532	0.027	2.3	281106	0.047	3.7
L-threonine 3-dehydrogenase (EC 1.1.1.103)	2918	6620	0.021	2.3	22837	0.040	7.8
Magnesium and cobalt efflux protein CorC	2392	10933	0.040	4.6	25443	0.045	10.6
Mannose-1-phosphate guanylyltransferase (EC 2.7.7.13)	8	411	0.018	49.3	777	0.027	93.1

Membrane-bound lytic murein transglycosylase B precursor (EC 3.2.1.-)	126	705	0.038	5.6	1549	0.007	12.3
Metallo-beta-lactamase superfamily protein PA0057	322	3039	0.012	9.4	2591	0.000	8.1
Multidrug efflux RND membrane fusion protein MexC	0	678	0.041	Absent in soil	834	0.034	Absent in soil
N-Acetylneuraminate cytidylyltransferase (EC 2.7.7.43)	61	746	0.009	12.2	1487	0.022	24.4
NADH dehydrogenase (EC 1.6.99.3)	3772	32913	0.017	8.7	78299	0.042	20.8
NADH dehydrogenase subunit 3	2470	23584	0.011	9.5	36742	0.007	14.9
NADH dehydrogenase subunit 5	35399	399604	0.035	11.3	768871	0.033	21.7
NADH-dependent butanol dehydrogenase A (EC 1.1.1.-)	46	1210	0.025	26.6	4633	0.023	101.7
Nitrite-sensitive transcriptional repressor NsrR	0	646	0.007	Absent in soil	1678	0.004	Absent in soil
Oligopeptidase A (EC 3.4.24.70)	1670	11404	0.004	6.8	14021	0.025	8.4
Organic hydroperoxide resistance transcriptional regulator	1318	7396	0.002	5.6	8927	0.006	6.8
PA0057_cluster	322	3471	0.017	10.8	3212	0.001	10.0
Peptidyl-prolyl cis-trans isomerase ppiD (EC 5.2.1.8)	543	5304	0.025	9.8	14346	0.047	26.4
Phosphoadenylyl-sulfate reductase [thioredoxin] (EC 1.8.4.8)	988	4035	0.001	4.1	12111	0.021	12.3
Phosphocarrier protein of PTS system	324	3100	0.000	9.6	7259	0.046	22.4
Phosphoribosylaminoimidazole-succinocarboxamide synthase (EC 6.3.2.6)	11300	33313	0.027	2.9	68743	0.050	6.1
Phytoene desaturase, pro-zeta-carotene producing (EC 1.-.-.-)	68	530	0.016	7.8	757	0.025	11.1
Predicted transcriptional regulator of N-Acetylglucosamine utilization,	42	743	0.023	17.9	876	0.006	21.1

GntR family							
Preprotein translocase subunit SecG (TC 3.A.5.1.1)	1342	8415	0.020	6.3	26871	0.028	20.0
Probable acyl-[acyl-carrier protein] desaturase DESA1 (Acyl-[ACP] desaturase) (Stearoyl-ACP desaturase) (Protein DES) (EC 1.14.19.2)	2236	6184	0.048	2.8	12480	0.003	5.6
Programmed Cell Death and Toxin-antitoxin Systems	2236	13543	0.024	6.1	40156	0.021	18.0
programmed frameshift-containing	62	1035	0.043	16.8	2152	0.001	34.8
Protein AraJ precursor	20	611	0.030	30.7	810	0.046	40.7
Protein containing plastocyanin/azurin family domain	17	182	0.047	10.9	275	0.034	16.5
Protein of unknown function DUF156	1432	5504	0.045	3.8	12093	0.001	8.4
Proton/glutamate symport protein	233	5358	0.006	23.0	6310	0.000	27.0
PutR, transcriptional activator of PutA and PutP	0	363	0.030	Absent in soil	357	0.041	Absent in soil
Pyruvate carboxyl transferase (EC 6.4.1.1)	3117	10346	0.048	3.3	13040	0.013	4.2
Queuosine biosynthesis QueD, PTPS-I	1642	5296	0.049	3.2	7742	0.050	4.7
Quinone oxidoreductase (EC 1.6.5.5)	2876	8770	0.008	3.0	15588	0.027	5.4
Rare lipoprotein A precursor	685	1849	0.026	2.7	6986	0.024	10.2
Riboflavin synthase alpha chain (EC 2.5.1.9)	491	3742	0.006	7.6	15486	0.049	31.5
Ribonucleotide reduction protein NrdI	101	1014	0.033	10.0	3016	0.016	29.7
Ribosomal large subunit pseudouridine synthase D (EC 4.2.1.70)	2854	8954	0.026	3.1	13810	0.018	4.8

Ribosomal protein L11 methyltransferase (EC 2.1.1.-)	270	3112	0.001	11.5	5055	0.008	18.7
Ribosome hibernation protein YhbH	4396	13121	0.041	3.0	32439	0.047	7.4
RNA polymerase sigma factor RpoH	37132	121542	0.027	3.3	292946	0.019	7.9
RNA polymerase sigma-H factor AlgT	13	2227	0.003	167.9	2839	0.049	214.0
RND efflux system, membrane fusion protein CmeA	539	3830	0.020	7.1	6715	0.003	12.5
Secretion	10922	27756	0.048	2.5	57770	0.042	5.3
Serine acetyltransferase (EC 2.3.1.30)	2867	14417	0.023	5.0	37793	0.025	13.2
Serine/threonine protein kinase PrkC, regulator of stationary phase	436	3190	0.016	7.3	3563	0.028	8.2
Sua5 YciO YrdC YwlC family protein	1220	3866	0.037	3.2	8268	0.005	6.8
Sulfate adenylyltransferase subunit 1 (EC 2.7.7.4)	15527	44296	0.022	2.9	89818	0.042	5.8
Sulfate adenylyltransferase subunit 2 (EC 2.7.7.4)	22180	63246	0.017	2.9	150899	0.041	6.8
Sulfite reductase [NADPH] hemoprotein beta-component (EC 1.8.1.2)	18461	59130	0.022	3.2	179560	0.042	9.7
Thiamin ABC transporter, substrate-binding component	23	347	0.025	14.8	1164	0.045	49.6
Thiol peroxidase, Bcp-type (EC 1.11.1.15)	9786	35696	0.027	3.6	56239	0.048	5.7
Thioredoxin	81214	232349	0.045	2.9	651096	0.045	8.0
Thioredoxin reductase (EC 1.8.1.9)	37345	99311	0.043	2.7	237561	0.031	6.4
Triosephosphate isomerase (EC 5.3.1.1)	3324	11693	0.025	3.5	24017	0.014	7.2
Twin-arginine translocation protein TatA	3300	14005	0.006	4.2	22473	0.034	6.8
UDP-3-O-[3-hydroxymyristoyl] N-	3215	12424	0.004	3.9	43692	0.012	13.6

acetylglucosamine deacetylase (EC 3.5.1.-)							
UPF0269 protein yggX	922	3553	0.001	3.9	9044	0.018	9.8
Xanthine dehydrogenase, iron-sulfur cluster and FAD-binding subunit A (1.17.1.4)	1060	4784	0.020	4.5	3194	0.017	3.0
Zinc uptake regulation protein ZUR	229	2727	0.017	11.9	4457	0.007	19.5

Table A29: Functions induced only in the wheat rhizopsphere compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Wheat mean transcripts /g**	**P value**	**Fold change (vs soil)**
2-cys peroxiredoxin BAS1, chloroplast precursor	0	90	0.042	Absent in soil
3-dehydro-L-gulonate 2-dehydrogenase (EC 1.1.1.130)	0	90	0.042	Absent in soil
Biotin carboxyl carrier protein of methylcrotonyl-CoA carboxylase	0	103	0.042	Absent in soil
Capsular polysaccharide biosynthesis/export periplasmic protein WcbA	0	92	0.037	Absent in soil
Colicin I receptor precursor	0	173	0.004	Absent in soil
C-terminal binding protein 2	0	70	0.048	Absent in soil
Glycine betaine ABC transport system, permease/glycine betaine-binding protein OpuABC	0	147	0.025	Absent in soil
Heptaprenyl diphosphate synthase component II (EC 2.5.1.30)	0	351	0.031	Absent in soil
Kef-type transport system 2 (probable substrate potassium), subunit 2	0	127	0.012	Absent in soil
Outer membrane receptor for ferric coprogen and ferric-rhodotorulic acid	0	241	0.021	Absent in soil
p-cumic aldehyde dehydrogenase (CymC) [EC:1.2.1.3]	0	103	0.042	Absent in soil
Predicted L-arabinose ABC transport system, permease protein 2	0	157	0.038	Absent in soil
Probable monothiol glutaredoxin GrlA	0	109	0.050	Absent in soil
Putative inner membrane protein YjeT (clustered with HflC)	0	68	0.034	Absent in soil
Putative mobilization protein BF0133	0	106	0.021	Absent in soil
Pyrimidine ABC transporter, transmembrane component 2	0	90	0.042	Absent in soil
Transcriptional regulator HxlR, formaldehyde assimilation	0	190	0.022	Absent in soil
transcriptional regulator SoxR	0	165	0.034	Absent in soil
Transcriptional repressor of aga operon	0	276	0.025	Absent in soil

Two-component system response regulator CreC	0	90	0.042	Absent in soil
Xanthan biosynthesis glycosyltransferase GumD	0	292	0.028	Absent in soil
Zn-dependent hydrolase, RNA-metabolising	0	166	0.035	Absent in soil
Predicted beta-xyloside ABC transporter, substrate-binding component	7	733	0.028	110.5
Stage V sporulation protein E	7	484	0.021	73.0
Flagellar protein FlbB	8	420	0.033	50.3
Alpha-galactosidase precursor (EC 3.2.1.22)	53	2546	0.024	47.8
Phosphonate ABC transporter permease protein phnE (TC 3.A.1.9.1)	8	347	0.021	41.6
Putative silver efflux pump	7	258	0.030	38.9
CAMP phosphodiesterases class-II:Metallo-beta-lactamase superfamily	8	271	0.007	32.5
Lactyl (2) diphospho-(5')guanosine:7,8-didemethyl-8-hydroxy-5-deazariboflavin 2-phospho-L-lactate transferase	15	470	0.019	31.4
N-Acetyl-D-glucosamine permease, possible	15	430	0.001	28.7
HemX protein, negative effector of steady-state concentration of glutamyl-tRNA reductase	8	211	0.000	25.3
Ethanolamine utilization polyhedral-body-like protein EutN	16	383	0.001	23.6
Putative hemine transporter ATP-binding subunit	7	143	0.004	21.6
Carotenoid cis-trans isomerase (EC 5.2.-.-)	37	727	0.000	19.9
Methyl-directed repair DNA adenine methylase (EC 2.1.1.72)	15	297	0.009	19.8
Formate dehydrogenase H (EC 1.2.1.2)	8	165	0.043	19.8
Two-component nitrogen fixation transcriptional regulator FixJ	30	516	0.003	17.2
D-glycero-D-manno-heptose 1,7-bisphosphate phosphatase (EC 3.1.1.-)	30	484	0.028	16.2
Chemotaxis protein CheX	32	499	0.033	15.4
Flagellar biosynthesis protein FliS	44	618	0.026	13.9
2-amino-4-hydroxy-6-hydroxymethyldihydropteridine pyrophosphokinase (EC 2.7.6.3)	36	496	0.030	13.8
Putative heme iron utilization protein	46	620	0.015	13.3
Quinone-reactive Ni/Fe-hydrogenase large chain (EC 1.12.5.1)	20	262	0.000	13.2
SSU ribosomal protein S13e (S15p)	8	107	0.004	12.8
Phenylacetaldehyde dehydrogenase (EC 1.2.1.39)	31	390	0.025	12.6
Nitrate ABC transporter, permease protein	268	3302	0.001	12.3
TRAP-type transport system, small permease component, predicted N-acetylneuraminate transporter	110	1313	0.049	11.9
Methylglyoxal synthase (EC 4.2.3.3)	191	2236	0.017	11.7
vanillin dehydrogenase	17	189	0.016	11.3
GTP-binding protein related to HflX	68	755	0.047	11.1

PTS system, mannitol-specific IIC component (EC 2.7.1.69)	27	293	0.045	11.0
tRNA pseudouridine 13 synthase (EC 4.2.1.-)	29	322	0.026	11.0
RNA pseudouridylate synthase, group 1	57	617	0.042	10.9
Copper sensory histidine kinase CusS	20	216	0.007	10.8
D-mannonate oxidoreductase (EC 1.1.1.57)	113	1208	0.018	10.7
MutS-related protein, family 1	17	173	0.037	10.4
Cyanate ABC transporter, permease protein	27	255	0.017	9.6
Anaerobic nitric oxide reductase flavorubredoxin	82	786	0.021	9.6
Archease	29	274	0.008	9.4
Urea ABC transporter, permease protein UrtB	255	2298	0.050	9.0
salicylate esterase	80	712	0.010	8.9
Cytochrome c nitrite reductase, small subunit NrfH	150	1320	0.024	8.8
O-succinylbenzoic acid--CoA ligase (EC 6.2.1.26)	100	868	0.049	8.7
FIG138315: Putative alpha helix protein	172	1446	0.037	8.4
Possible fucose ABC transporter, ATP-binding component	32	251	0.026	7.9
Hydroxymethylglutaryl-CoA reductase (EC 1.1.1.34)	312	2332	0.021	7.5
Transcriptional regulatory protein PhoP	69	518	0.003	7.5
TcuA: flavoprotein used to oxidize tricarballylate to cis-aconitate	141	1039	0.023	7.4
Putative malate dehydrogenase (EC 1.1.1.37), similar to archaeal MJ1425	275	2003	0.048	7.3
TldD-domain protein	82	555	0.026	6.8
Cephalosporin hydroxylase	149	981	0.002	6.6
Urea carboxylase-related aminomethyltransferase (EC 2.1.2.10)	282	1830	0.046	6.5
Nitric-oxide reductase subunit B (EC 1.7.99.7)	385	2481	0.028	6.4
Type IV prepilin peptidase TadV/CpaA	38	246	0.037	6.4
Shufflon-specific DNA recombinase	66	425	0.001	6.4
Hydrogenase transcriptional regulatory protein hoxA	56	361	0.014	6.4
Pyoverdine sidechain non-ribosomal peptide synthetase PvdD	140	825	0.007	5.9
secreted alkaline phosphatase	177	1026	0.023	5.8
N-methylhydantoinase (ATP-hydrolyzing) (EC 3.5.2.14)	52	297	0.007	5.7
UDP-GlcNAc-specific C4,6 dehydratase/C5 epimerase (PseB, first step of pseudaminic acid biosynthesis)	111	629	0.049	5.7
Mannose-6-phosphate isomerase (EC 5.3.1.8)	323	1818	0.032	5.6
NAD-dependent formate dehydrogenase delta subunit	28	158	0.033	5.6

Superoxide dismutase [Cu-Zn] precursor (EC 1.15.1.1)	283	1581	0.023	5.6
Hypothetical membrane protein, possible involvement in cytochrome functioning/assembly	381	2124	0.031	5.6
Choline_Transport	74	414	0.007	5.6
Acetoacetyl-CoA synthetase [leucine] (EC 6.2.1.16)	180	994	0.016	5.5
L-ribulose-5-phosphate 4-epimerase (EC 5.1.3.4)	310	1700	0.012	5.5
Glucosamine-6-phosphate deaminase [isomerizing], alternative (EC 3.5.99.6)	50	272	0.043	5.5
DNA-directed RNA polymerase gamma subunit (EC 2.7.7.6)	211	1142	0.010	5.4
Urea carboxylase-related ABC transporter, periplasmic substrate-binding protein	207	1093	0.018	5.3
Mutator mutT protein (7,8-dihydro-8-oxoguanine-triphosphatase) (EC 3.6.1.-)	155	808	0.041	5.2
Glutamine synthetase family protein in hypothetical Actinobacterial gene cluster	323	1675	0.007	5.2
Urea carboxylase-related ABC transporter, ATPase protein	321	1642	0.004	5.1
Glyceraldehyde-3-phosphate dehydrogenase (EC 1.2.1.12) (GAPDH)	178	901	0.021	5.1
Spermidine Putrescine ABC transporter permease component potC (TC_3.A.1.11.1)	398	2001	0.002	5.0
Polysulfide reductase, subunit B, putative	131	652	0.001	5.0
ABC-type protease exporter, ATP-binding component PrtD/AprD	209	1040	0.007	5.0
Type III secretion system related	552	2719	0.000	4.9
Acetaldehyde dehydrogenase, ethanolamine utilization cluster	145	710	0.044	4.9
Glutaminase (EC 3.5.1.2)	335	1606	0.003	4.8
COG0840: Methyl-accepting chemotaxis protein	546	2607	0.000	4.8
O-succinylbenzoate-CoA synthase (EC 4.2.1.-)	143	680	0.037	4.8
4,4'-diapolycopene oxidase	38	182	0.039	4.7
Cyanate ABC transporter, ATP-binding protein	92	428	0.006	4.6
Trehalase (EC 3.2.1.28)	38	176	0.003	4.6
Sulfur oxidation protein SoxB	247	1117	0.002	4.5
Ribosomal protein S6 glutaminyl transferase	244	1105	0.033	4.5
Dihydroorotate dehydrogenase, catalytic subunit (EC 1.3.3.1)	138	613	0.006	4.4
Methylthioribulose-1-phosphate dehydratase (EC 4.2.1.109)	82	364	0.009	4.4
Alpha-glucosidase, family 31 of glycosyl hydrolases, COG1501	526	2297	0.007	4.4
Benzoate transport, ATP binding protein	62	268	0.025	4.3
Ni,Fe-hydrogenase I small subunit	155	668	0.037	4.3
Sulfur oxidation protein SoxY	103	436	0.019	4.3
2-keto-3-deoxy-L-fuconate dehydrogenase	303	1278	0.001	4.2

Pyruvate:ferredoxin oxidoreductase, alpha subunit (EC 1.2.7.1)	684	2804	0.033	4.1
Alpha-N-arabinofuranosidase (EC 3.2.1.55)	857	3470	0.044	4.1
Taurine transport ATP-binding protein TauB	222	896	0.023	4.0
Na+/H+-dicarboxylate symporters	1110	4481	0.010	4.0
AA3-600 quinol oxidase subunit I	110	436	0.013	4.0
Ribonuclease HI (EC 3.1.26.4)	649	2565	0.026	3.9
Signal transduction histidine kinase HoxJ (hydrogenase regulation)	279	1060	0.020	3.8
Xanthosine phosphorylase (EC 2.4.2.1)	163	608	0.041	3.7
4-coumarate--CoA ligase 1 (EC 6.2.1.12)	109	404	0.016	3.7
Anaerobic dimethyl sulfoxide reductase chain A (EC 1.8.99.-)	399	1469	0.047	3.7
Alfa-L-rhamnosidase (EC 3.2.1.40)	105	387	0.031	3.7
Endonuclease V (EC 3.1.21.7)	194	704	0.036	3.6
Phosphoribosylaminoimidazole carboxylase catalytic subunit (EC 4.1.1.21)	1109	4003	0.001	3.6
DNA repair protein RadC	763	2711	0.046	3.6
ammonium/methylammonium permease	288	1019	0.014	3.5
ABC-type nitrate/sulfonate/bicarbonate transport system, ATPase component	929	3241	0.023	3.5
3-hydroxyanthranilate 3,4-dioxygenase (EC 1.13.11.6)	591	2044	0.050	3.5
Cytochrome c552 precursor (EC 1.7.2.2)	1643	5673	0.010	3.5
7,8-didemethyl-8-hydroxy-5-deazariboflavin synthase subunit 1	118	405	0.023	3.4
Glycerol-3-phosphate acyltransferase (EC 2.3.1.15)	335	1146	0.004	3.4
Quinate/shikimate dehydrogenase [Pyrroloquinoline-quinone] (EC 1.1.99.25)	193	661	0.004	3.4
Hexuronate transporter	1006	3433	0.005	3.4
Manganese uptake regulation protein MUR	157	521	0.050	3.3
Predicted nucleoside ABC transporter, ATP-binding component	369	1208	0.031	3.3
Acyl-[acyl-carrier-protein]--UDP-N-acetylglucosamine O-acyltransferase (EC 2.3.1.129)	3209	10402	0.007	3.2
DNA mismatch repair protein MutS	5400	17460	0.047	3.2
Sulfite reduction-associated complex DsrMKJOP protein DsrP (= HmeB)	246	789	0.025	3.2
Aromatic amino acids and derivatives.1	825	2601	0.012	3.2
L-arabonate dehydratase (EC 4.2.1.25)	1739	5411	0.028	3.1
D,D-heptose 7-phosphate kinase	306	948	0.001	3.1
NADH:ubiquinone oxidoreductase 17.2 kD subunit	459	1418	0.039	3.1
Cytochrome c-type biogenesis protein DsbD, protein-disulfide reductase (EC 1.8.1.8)	1631	5036	0.037	3.1

Cobalt-zinc-cadmium resistance protein CzcD	410	1254	0.037	3.1
Xanthosine permease	520	1591	0.019	3.1
Pyruvate dehydrogenase E1 component beta subunit (EC 1.2.4.1)	9519	28951	0.029	3.0
Sarcosine oxidase beta subunit (EC 1.5.3.1)	1193	3587	0.047	3.0
Adenylylsulfate reductase alpha-subunit (EC 1.8.99.2)	441	1325	0.018	3.0
Formate dehydrogenase N alpha subunit (EC 1.2.1.2); selenocysteine-containing	421	1236	0.037	2.9
Ferric siderophore transport system, biopolymer transport protein ExbB	1058	3054	0.028	2.9
Octanoate-[acyl-carrier-protein]-protein-N-octanoyltransferase	1578	4485	0.019	2.8
Flagellar hook protein FlgE	6541	18202	0.039	2.8
Nitrogenase (molybdenum-iron)-specific transcriptional regulator NifA	1420	3897	0.001	2.7
Ribosomal-protein-S5p-alanine acetyltransferase	128	349	0.037	2.7
Nitrous oxide reductase maturation protein NosF (ATPase)	293	787	0.021	2.7
Nicotinamidase (EC 3.5.1.19)	728	1950	0.041	2.7
Flagellar hook-length control protein FliK	1517	4010	0.009	2.6
Dihydrofolate reductase (EC 1.5.1.3)	1252	3261	0.041	2.6
Dihydropyrimidine dehydrogenase [NADP+] (EC 1.3.1.2)	814	2097	0.037	2.6
Molybdenum cofactor biosynthesis protein MoaB	273	675	0.016	2.5
Flagellar regulatory protein FleQ	2828	6975	0.012	2.5
Asparagine synthetase [glutamine-hydrolyzing] (EC 6.3.5.4) AsnH	1147	2813	0.030	2.5
Nitrite reductase [NAD(P)H] large subunit (EC 1.7.1.4)	7975	19408	0.028	2.4
Propionate catabolism operon regulatory protein PrpR	253	614	0.003	2.4
Glutamate-1-semialdehyde aminotransferase (EC 5.4.3.8)	5336	12877	0.036	2.4
Pyruvate-flavodoxin oxidoreductase (EC 1.2.7.-)	18931	45626	0.020	2.4
Uronate isomerase (EC 5.3.1.12)	1876	4509	0.008	2.4
Methylmalonyl-CoA decarboxylase, alpha chain (EC 4.1.1.41)	595	1396	0.027	2.3
Glutamate synthase [NADPH] small chain (EC 1.4.1.13)	24914	55197	0.024	2.2
ABC transporter involved in cytochrome c biogenesis, CcmB subunit	391	854	0.032	2.2
Denitrification	59039	127861	0.048	2.2
Putative glycogen debranching enzyme, archaeal type, TIGR01561	458	963	0.002	2.1
L-arabinose-specific 1-epimerase (mutarotase)	229	474	0.028	2.1
Dihydropyrimidinase (EC 3.5.2.2)	2370	4908	0.039	2.1
2-dehydro-3-deoxygluconate kinase (EC 2.7.1.45)	1236	2555	0.036	2.1

RsbR, positive regulator of sigma-B	2067	4193	0.018	2.0
Propionate--CoA ligase (EC 6.2.1.17)	1579	3138	0.008	2.0
NAD-specific glutamate dehydrogenase (EC 1.4.1.2)	3449	6517	0.038	1.9
Pyrophosphate-energized proton pump (EC 3.6.1.1)	29281	51566	0.048	1.8
Propionyl-CoA carboxylase carboxyl transferase subunit (EC 6.4.1.3)	5565	8265	0.035	1.5

Table A30: Functions induced only in the oat rhizopsphere compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Oat mean transcripts /g**	**P value**	**Fold change (vs soil)**
3-hydroxyacyl-CoA dehydrogenase (PaaH) (EC 1.1.1.157)	0	491	0.021	Absent in soil
Arginine N-succinyltransferase, alpha subunit (EC 2.3.1.109)	0	106	0.012	Absent in soil
ATP-dependent protease LonB-like Type I	0	777	0.015	Absent in soil
Blr3520 protein homolog, hypothetical protein	0	45	0.043	Absent in soil
Choline binding protein A	0	3359	0.041	Absent in soil
Cyanobacteria-specific RpoD-like sigma factor, type-16	0	45	0.043	Absent in soil
Deoxyribonuclease TatD	0	212	0.049	Absent in soil
Eukaryotic translation initiation factor 4B	0	406	0.027	Absent in soil
Glucitol operon GutQ protein	0	45	0.043	Absent in soil
Glycerol-1-phosphate dehydrogenase [NAD(P)] (EC 1.1.1.261)	0	147	0.011	Absent in soil
Inducers of aerial mycelium formation biosynthesis protein BldG	0	342	0.019	Absent in soil
Inner membrane protein forms channel for type IV secretion of T-DNA complex (VirB3)	0	73	0.011	Absent in soil
Lacto-N-Biose_I_and_Galacto-N-Biose_Metabolic_Pathway	0	114	0.008	Absent in soil
Menaquinone-cytochrome C oxidoreductase, cytochrome C subunit	0	193	0.027	Absent in soil
Menaquinone-cytochrome C reductase iron-sulfur subunit	0	199	0.028	Absent in soil
Methyl-accepting chemotaxis protein III (ribose and galactose chemoreceptor protein)	0	106	0.012	Absent in soil
Pca regulon regulatory protein PcaR	0	145	0.025	Absent in soil
Predicted glucose transporter in maltodextrin utilization gene cluster	0	766	0.043	Absent in soil
Protein YigP (COG3165) clustered with ubiquinone biosynthetic genes	0	298	0.020	Absent in soil
PTS system, mannose-specific IIA component (EC 2.7.1.69)	0	185	0.034	Absent in soil
Putative CDP-glycosylpolyol phosphate:glycosylpolyol glycosylpolyolphosphotransferase	0	195	0.010	Absent in soil

Putative membrane protein YfcA	0	166	0.026	Absent in soil
Similar to phosphoglycolate phosphatase, clustered with ubiquinone biosynthesis SAM-dependent O-methyltransferase	0	333	0.007	Absent in soil
Similar to TadZ/CpaE, associated with Flp pilus assembly	0	104	0.008	Absent in soil
Stage 0 sporulation two-component response regulator (Spo0A)	0	432	0.039	Absent in soil
Stage V sporulation protein B	0	219	0.010	Absent in soil
Substrate-specific component RibU of riboflavin ECF transporter	0	137	0.047	Absent in soil
Tetrathionate reductase subunit A	0	315	0.033	Absent in soil
Transcriptional regulator in cluster with Zn-dependent hydrolase	0	375	0.039	Absent in soil
Predicted cell-wall-anchored protein SasA (LPXTG motif)	23	2138	0.040	91.7
Guanine-hypoxanthine permease	8	667	0.002	88.0
ATPase component of general energizing module of ECF transporters	25	1881	0.036	75.1
Chlorohydroquinone/hydroquinone 1,2-dioxygenase	161	10122	0.005	63.0
Alcohol dehydrogenase (EC 1.1.1.1); Acetaldehyde dehydrogenase (EC 1.2.1.10)	123	5292	0.034	43.1
Two-component response regulator SA14-24	27	1049	0.027	39.5
Transcription repressor of multidrug efflux pump acrAB operon, TetR (AcrR) family	8	251	0.019	33.2
Quinate permease	81	2345	0.025	28.8
Dihydroneopterin triphosphate pyrophosphohydolase type 2	28	706	0.044	25.0
Phosphate-specific outer membrane porin OprP	38	928	0.003	24.5
Glycerol dehydrogenase (EC 1.1.1.6)	7	162	0.010	24.4
Ortho-halobenzoate 1,2-dioxygenase beta-ISP protein OhbA	35	804	0.004	23.0
Periplasmic thiol:disulfide oxidoreductase DsbB, required for DsbA reoxidation	37	801	0.033	21.9
Flavohemoprotein (Hemoglobin-like protein) (Flavohemoglobin) (Nitric oxide dioxygenase) (EC 1.14.12.17)	512	10772	0.050	21.0
Thiamin-regulated outer membrane receptor Omr1	46	922	0.019	20.2
ABC exporter for hemopore HasA, ATP-binding component HasD	48	954	0.021	19.8
Nitrous oxide reductase maturation transmembrane protein NosY	7	129	0.029	19.5
Flagellar hook-associated protein 3	15	295	0.037	19.5
4-diphosphocytidyl-2-C-methyl-D-erythritol kinase (EC 2.7.1.148)	208	3925	0.000	18.9
Hypothetical, related to broad specificity phosphatases COG0406	48	837	0.042	17.6
L-lactate dehydrogenase (EC 1.1.1.27)	59	956	0.011	16.3
L-serine dehydratase, alpha subunit (EC 4.3.1.17)	23	364	0.036	16.0
IMP dehydrogenase related 1 (EC 1.1.1.205)	27	411	0.026	15.5
Glycolate dehydrogenase (EC 1.1.99.14), FAD-binding subunit GlcE	13	198	0.000	14.9

Phosphogluconate dehydratase (EC 4.2.1.12)	1602	23430	0.023	14.6
Galactose/methyl galactoside ABC transport system, ATP-binding protein MglA (EC 3.6.3.17)	38	554	0.040	14.6
4-amino-4-deoxy-L-arabinose transferase and related glycosyltransferases of PMT family	15	218	0.028	14.6
Glucan 1,6-alpha-glucosidase (EC 3.2.1.70)	50	730	0.045	14.6
Putative reductase (alkanesulfonate metabolism)	13	191	0.032	14.4
Isopentenyl-diphosphate delta-isomerase (EC 5.3.3.2)	107	1539	0.007	14.4
Sigma factor RpoE negative regulatory protein RseA	83	1164	0.019	14.1
Type cbb3 cytochrome oxidase biogenesis protein CcoS, involved in heme b insertion	22	301	0.000	13.9
ATP synthase protein I	311	4249	0.043	13.7
alginate biosynthesis protein AlgJ	16	210	0.014	13.0
Organosulfonate utilization protein SsuF	20	247	0.021	12.4
Maltose O-acetyltransferase (EC 2.3.1.79)	94	1106	0.036	11.8
Glycosyltransferase MshA involved in mycothiol biosynthesis (EC 2.4.1.-)	136	1571	0.011	11.6
PTS system, N-acetylglucosamine-specific IIA component (EC 2.7.1.69)	30	295	0.020	9.7
N-acetylmannosaminyltransferase (EC 2.4.1.187)	58	540	0.010	9.2
ATP-dependent nuclease, subunit B	38	340	0.024	8.9
Phosphoenolpyruvate-protein phosphotransferase of PTS system (EC 2.7.3.9)	1360	12029	0.008	8.8
Pyrophosphate--fructose 6-phosphate 1-phosphotransferase, beta subunit (EC 2.7.1.90)	131	1146	0.023	8.7
Arginine/ornithine antiporter ArcD	85	727	0.018	8.5
Ribonucleotide reductase of class Ib (aerobic), beta subunit (EC 1.17.4.1)	332	2829	0.008	8.5
Hypothetical flavoprotein YqcA (clustered with tRNA pseudouridine synthase C)	55	460	0.014	8.4
Predicted nucleoside ABC transporter, permease 1 component	74	607	0.017	8.2
Peptidoglycan-associated lipoprotein precursor	1516	12257	0.036	8.1
Protocatechuate 3,4-dioxygenase alpha chain (EC 1.13.11.3)	137	1087	0.026	7.9
Predicted L-arabinose ABC transport system, ATP-binding protein	32	248	0.031	7.8
16S rRNA processing protein RimM	2562	19763	0.018	7.7
Aerotaxis sensor receptor protein	1612	12072	0.032	7.5
Transcriptional repressor of PutA and PutP	2205	16253	0.012	7.4
5-carboxymethyl-2-hydroxymuconate delta-isomerase (EC 5.3.3.10)	142	1046	0.013	7.3
Maltose operon transcriptional repressor MalR, LacI family	46	335	0.035	7.3
cyclolysin secretion ATP-binding protein	17	122	0.019	7.3
FIG005453: Putative DeoR-family transcriptional regulator	70	508	0.008	7.3

DNA polymerase II (EC 2.7.7.7)	48	352	0.043	7.3
Predicted rhamnose oligosaccharide ABC transport system, substrate-binding component	97	701	0.018	7.2
Putative coproporphyrinogen III oxidase of BS HemN-type, oxygen-independent (EC 1.3.99.22), in heat shock gene cluster	86	621	0.009	7.2
Flagellar biosynthesis protein FliL	361	2511	0.014	7.0
Transcriptional regulatory protein RtcR	20	138	0.033	7.0
FIG146085: 3'-to-5' oligoribonuclease A, Bacillus type	73	503	0.026	6.9
internalin, putative	126	870	0.037	6.9
Eukaryotic translation initiation factor 3 110 kDa subunit	387	2662	0.031	6.9
Glycine betaine transporter OpuD	207	1422	0.043	6.9
Cytochrome d ubiquinol oxidase subunit I (EC 1.10.3.-)	2861	19373	0.033	6.8
Ribitol 2-dehydrogenase (EC 1.1.1.56)	68	458	0.000	6.7
Copper resistance protein D	32	207	0.045	6.5
Adenosylcobinamide-phosphate synthase	63	409	0.046	6.4
Diaminobutyrate-pyruvate aminotransferase (EC 2.6.1.46)	433	2755	0.030	6.4
Binding-protein-dependent transport systems inner membrane component:ATP/GTP-binding site motif A (P-loop) :TrkA-N:Potassium e	27	168	0.049	6.3
Thiamin-phosphate pyrophosphorylase (EC 2.5.1.3)	52	322	0.000	6.2
POTASSIUM/PROTON ANTIPORTER ROSB	60	372	0.002	6.2
Nitrilotriacetate monooxygenase component B (EC 1.14.13.-)	121	745	0.050	6.1
Protocatechuate 3,4-dioxygenase beta chain (EC 1.13.11.3)	1466	8964	0.046	6.1
Hydrogenase-4 component E (EC 1.-.-.-)	37	225	0.018	6.1
Similar to eukaryotic Peptidyl prolyl 4-hydroxylase, alpha subunit (EC 1.14.11.2)	169	1017	0.041	6.0
Chitin binding protein	35	210	0.015	6.0
Ribose ABC transport system, periplasmic ribose-binding protein RbsB (TC 3.A.1.2.1)	343	1991	0.022	5.8
Alpha-galactosidase (EC 3.2.1.22)	649	3625	0.017	5.6
D-amino acid dehydrogenase small subunit (EC 1.4.99.1)	2942	15990	0.007	5.4
Histidinol-phosphate aminotransferase (EC 2.6.1.9)	675	3601	0.004	5.3
Copper-sensing two-component system response regulator CpxR	121	630	0.042	5.2
Ribosome protection-type tetracycline resistance related proteins	20	103	0.049	5.2
Flp pilus assembly protein TadB	490	2483	0.015	5.1
Two-component sensor PilS	81	411	0.001	5.0
Phosphate regulon sensor protein PhoR (EC 2.7.3.-)	337	1639	0.020	4.9

DNA-directed RNA polymerase delta (= beta'') subunit (EC 2.7.7.6)	169	816	0.011	4.8
Heat-inducible transcription repressor HrcA	2107	9994	0.042	4.7
Exodeoxyribonuclease VII small subunit (EC 3.1.11.6)	284	1325	0.008	4.7
Cobyric acid synthase	128	592	0.044	4.6
Recombination inhibitory protein MutS2	803	3675	0.045	4.6
Type IV pilus biogenesis protein PilP	1183	5343	0.000	4.5
Proteasome subunit alpha (EC 3.4.25.1), bacterial	494	2229	0.030	4.5
Sugar Phosphotransferase Systems, PTS	3469	15247	0.019	4.4
Beta-ketoadipyl CoA thiolase (EC 2.3.1.9)	240	1050	0.003	4.4
Cytochrome d ubiquinol oxidase subunit II (EC 1.10.3.-)	1084	4687	0.031	4.3
Excinuclease ABC subunit A paralog in greater Bacteroides group	3652	15576	0.020	4.3
Predicted transcriptional regulator of sulfate adenylyltransferase, Rrf2 family	73	310	0.033	4.2
Cytochrome c oxidase subunit CcoQ (EC 1.9.3.1)	75	317	0.021	4.2
Chaperone protein HtpG	24231	102089	0.011	4.2
Uroporphyrinogen-III methyltransferase (EC 2.1.1.107)	841	3515	0.010	4.2
Ferredoxin-dependent glutamate synthase (EC 1.4.7.1)	3319	13880	0.028	4.2
DNA recombination protein RmuC	674	2783	0.001	4.1
Heat shock protein 60 family co-chaperone GroES	40914	166942	0.021	4.1
LSU ribosomal protein L24p (L26e)	53844	218629	0.016	4.1
N-succinyl-L,L-diaminopimelate desuccinylase (EC 3.5.1.18)	1089	4388	0.005	4.0
Heat shock protein 60 family chaperone GroEL	277911	1108161	0.015	4.0
Inactive homolog of metal-dependent proteases, putative molecular chaperone	207	820	0.009	4.0
Fe-S oxidoreductase-like protein in Rubrerythrin cluster	1219	4733	0.004	3.9
Iron-dependent repressor IdeR/DtxR	594	2294	0.049	3.9
Dimethyladenosine transferase (EC 2.1.1.-)	1110	4278	0.023	3.9
Malate:quinone oxidoreductase (EC 1.1.99.16)	2998	11458	0.006	3.8
N-acetylornithine carbamoyltransferase (EC 2.1.3.9)	593	2260	0.001	3.8
LSU ribosomal protein L23p (L23Ae)	60835	231177	0.023	3.8
Flagellar basal-body rod protein FlgF	1001	3797	0.024	3.8
probable iron binding protein from the HesB_IscA_SufA family in Nif operon	172	650	0.043	3.8
Protein of unknown function DUF81	793	2973	0.012	3.8
CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase (EC 2.7.8.5)	1369	5047	0.018	3.7

Flagellar basal-body rod protein FlgB	1815	6672	0.047	3.7
Guanine deaminase (EC 3.5.4.3)	841	3090	0.020	3.7
Chromosome (plasmid) partitioning protein ParA	3697	13425	0.016	3.6
Putative preQ0 transporter	261	938	0.000	3.6
LSU ribosomal protein L30p (L7e)	20450	73024	0.032	3.6
Agmatine deiminase (EC 3.5.3.12)	1614	5744	0.012	3.6
Chaperone protein DnaK	310308	1097733	0.029	3.5
Biosynthesis of phenylpropanoids	2410	8486	0.012	3.5
Anthranilate phosphoribosyltransferase (EC 2.4.2.18)	1012	3558	0.008	3.5
Sorbitol dehydrogenase (EC 1.1.1.14)	1511	5301	0.050	3.5
Putative glucanase glgE (EC 3.2.1.-)	1010	3540	0.016	3.5
Benzoyl-CoA oxygenase component B	4644	16166	0.013	3.5
Permease of the drug/metabolite transporter (DMT) superfamily	1074	3693	0.048	3.4
Pathogenicity islands	408160	1399743	0.022	3.4
Nickel responsive regulator NikR	66	225	0.042	3.4
Cysteine desulfurase (EC 2.8.1.7), IscS subfamily	22929	77702	0.032	3.4
Cysteine desulfurase (EC 2.8.1.7)	33849	113845	0.018	3.4
Predicted uronate isomerase TM0442	210	697	0.030	3.3
Cytochrome c551/c552	2681	8881	0.007	3.3
Protein folding	880790	2910167	0.029	3.3
UDP-N-acetylglucosamine 1-carboxyvinyltransferase (EC 2.5.1.7)	9697	31905	0.046	3.3
Two-component sensor histidine kinase PleC	658	2165	0.025	3.3
Ferric iron ABC transporter, ATP-binding protein	964	3164	0.017	3.3
Thioredoxin-disulfide_reductase	115068	367370	0.015	3.2
ATP-dependent RNA helicase NGO0650	8156	25960	0.017	3.2
L-pipecolate dehydrogenase (EC 1.5.99.3)	426	1348	0.017	3.2
LSU ribosomal protein L17p	67164	212412	0.031	3.2
ATP-dependent DNA helicase RecQ	20616	64613	0.007	3.1
Heat_shock_dnaK_gene_cluster_extended	476201	1491798	0.035	3.1
Phosphate transport regulator (distant homolog of PhoU)	2434	7526	0.039	3.1
tRNA (Guanine37-N1) -methyltransferase (EC 2.1.1.31)	17319	52418	0.005	3.0
Bacterioferritin	11950	35975	0.022	3.0

Iron-sulfur cluster assembly scaffold protein IscU	22025	66101	0.018	3.0
LSU ribosomal protein L1p (L10Ae)	137592	412751	0.037	3.0
Ferric uptake regulation protein FUR	3801	11402	0.022	3.0
LSU ribosomal protein L21p	76204	228563	0.035	3.0
Aminopeptidase C (EC 3.4.22.40)	757	2261	0.050	3.0
Putative sugar nucleotidyltransferase	446	1332	0.021	3.0
Queuosine Biosynthesis QueC ATPase	1444	4282	0.033	3.0
LSU ribosomal protein L20p	78362	227579	0.027	2.9
Dihydrolipoamide acetyltransferase component of pyruvate dehydrogenase complex (EC 2.3.1.12)	13760	39947	0.022	2.9
Naphthoate synthase (EC 4.1.3.36)	1051	3041	0.022	2.9
LSU ribosomal protein L6p (L9e)	138595	400508	0.042	2.9
Electron transport complex protein RnfB	599	1727	0.042	2.9
LSU ribosomal protein L34p	6508	18652	0.013	2.9
LSU ribosomal protein L5p (L11e)	227679	649310	0.040	2.9
ATP-dependent RNA helicase Bcep18194_A5658	3947	11249	0.021	2.9
Acriflavin resistance protein	57383	162150	0.038	2.8
LSU ribosomal protein L3p (L3e)	194517	548884	0.047	2.8
LSU ribosomal protein L9p	49431	139245	0.023	2.8
Electron transfer flavoprotein, alpha subunit	13199	37139	0.032	2.8
Porphobilinogen synthase (EC 4.2.1.24)	6895	19358	0.045	2.8
FIG003492: Threonine dehydrogenase and related Zn-dependent dehydrogenases	253	710	0.030	2.8
Nucleoside diphosphate kinase (EC 2.7.4.6)	30466	84885	0.040	2.8
Transcriptional regulator, IclR family	2646	7364	0.016	2.8
Lipopolysaccharide ABC transporter, ATP-binding protein LptB	1783	4931	0.037	2.8
Dihydrolipoamide dehydrogenase of 2-oxoglutarate dehydrogenase (EC 1.8.1.4)	15823	43571	0.037	2.8
LSU ribosomal protein L16p (L10e)	116752	321229	0.035	2.8
LSU ribosomal protein L14p (L23e)	149013	409006	0.047	2.7
Cyanophycin synthase II	1793	4916	0.039	2.7
Magnesium and cobalt transport protein CorA	2807	7688	0.049	2.7
SSU ribosomal protein S19p (S15e)	141847	387033	0.043	2.7
Pyridoxal kinase (EC 2.7.1.35)	198	540	0.045	2.7
ATP synthase gamma chain (EC 3.6.3.14)	42485	115536	0.033	2.7

Kup system potassium uptake protein	7236	19515	0.039	2.7
Lipoate synthase	19523	52542	0.016	2.7
Aconitate hydratase 2 (EC 4.2.1.3)	11055	29562	0.008	2.7
ATP-dependent RNA helicase RhlE	26922	71670	0.023	2.7
SSU ribosomal protein S3p (S3e)	197920	524491	0.032	2.7
LSU ribosomal protein L35p	92932	246126	0.041	2.6
Putative cytochrome C-type biogenesis protein	638	1689	0.004	2.6
Phenylalanine-4-hydroxylase (EC 1.14.16.1)	2449	6476	0.003	2.6
Lipoic acid	22730	59997	0.019	2.6
Succinyl-CoA ligase [ADP-forming] beta chain (EC 6.2.1.5)	30373	79827	0.025	2.6
Protein acetyltransferase	435	1141	0.048	2.6
Predicted L-lactate dehydrogenase, Iron-sulfur cluster-binding subunit YkgF	2392	6239	0.011	2.6
Adenylosuccinate synthetase (EC 6.3.4.4)	13549	35073	0.039	2.6
ATP synthase epsilon chain (EC 3.6.3.14)	8714	22192	0.048	2.5
Integration host factor beta subunit	12949	32933	0.032	2.5
Butyryl-CoA dehydrogenase (EC 1.3.99.2)	2050	5155	0.038	2.5
Segregation and condensation protein A	2945	7331	0.020	2.5
GTPase and tRNA-U34 5-formylation enzyme TrmE	1763	4355	0.046	2.5
Clustering-based subsystems.1	79322	195700	0.044	2.5
Electron transfer flavoprotein-ubiquinone oxidoreductase (EC 1.5.5.1)	8106	19935	0.033	2.5
P-hydroxybenzoate hydroxylase (EC 1.14.13.2)	795	1936	0.035	2.4
ATP-dependent DNA helicase UvrD/PcrA	11389	27375	0.050	2.4
Pterin-4-alpha-carbinolamine dehydratase (EC 4.2.1.96)	1339	3171	0.019	2.4
3-oxoacyl-[acyl-carrier-protein] synthase, KASIII (EC 2.3.1.41)	14405	33638	0.045	2.3
GTP pyrophosphokinase (EC 2.7.6.5), (p)ppGpp synthetase II	12852	30008	0.033	2.3
Acetoacetyl-CoA reductase (EC 1.1.1.36)	9943	23142	0.047	2.3
Pyruvate kinase (EC 2.7.1.40)	8839	20364	0.048	2.3
Dihydrolipoamide dehydrogenase (EC 1.8.1.4)	6154	14086	0.033	2.3
MiaB family protein, possibly involved in tRNA or rRNA modification	9301	21285	0.027	2.3
Biogenesis_of_c-type_cytochromes	24434	53958	0.047	2.2
Aspartate carbamoyltransferase (EC 2.1.3.2)	6024	13185	0.039	2.2
Phosphoenolpyruvate-dihydroxyacetone phosphotransferase (EC 2.7.1.121), dihydroxyacetone binding subunit DhaK	1515	3208	0.048	2.1

Topoisomerase IV subunit A (EC 5.99.1.-)	8500	17861	0.049	2.1
MSHA biogenesis protein MshE	2246	4620	0.041	2.1
NADH-ubiquinone oxidoreductase chain C (EC 1.6.5.3)	11092	22632	0.033	2.0
Rod shape-determining protein MreB	37384	75166	0.043	2.0
Dihydroxy-acid dehydratase (EC 4.2.1.9)	23858	45170	0.047	1.9
Helicase PriA essential for oriC/DnaA-independent DNA replication	1176	2008	0.025	1.7

Table A31: Functions induced only in the pea rhizopsphere compared to soil (P≤0.05).

Assignment in the SEED database	**Soil mean transcripts / g**	**Pea mean transcripts /g**	**P value**	**Fold change (vs soil)**
2-dehydro-3-deoxyphosphogalactonate aldolase (EC 4.1.2.21)	0	267	0.006	Absent in soil
Bacterioferritin-associated ferredoxin	0	849	0.045	Absent in soil
CblX, a non-orthologous displasment for Alpha-ribazole-5'-phosphate phosphatase	0	250	0.022	Absent in soil
Coenzyme F420-reducing hydrogenase, beta subunit	0	365	0.044	Absent in soil
DNA polymerase III chi subunit (EC 2.7.7.7)	0	981	0.032	Absent in soil
ElaA protein	0	293	0.047	Absent in soil
Erythritol transcriptional regulator EryD	0	575	0.031	Absent in soil
Ferredoxin-NADP(+) reductase (EC 1.18.1.2)	0	473	0.024	Absent in soil
Flagellar biosynthesis protein FlgN	0	1077	0.024	Absent in soil
Hexuronate utilization operon transcriptional repressor ExuR	0	178	0.024	Absent in soil
L-2,4-diaminobutyric acid acetyltransferase (EC 2.3.1.-)	0	407	0.016	Absent in soil
Methyl-accepting chemotaxis protein II (aspartate chemoreceptor protein)	0	323	0.032	Absent in soil
Nudix-related transcriptional regulator NrtR	0	285	0.024	Absent in soil
Phosphonates transport ATP-binding protein PhnL	0	97	0.045	Absent in soil
Phosphoribosylaminoimidazolecarboxamide formyltransferase (EC 2.1.2.3)	0	141	0.011	Absent in soil
Predicted L-rhamnose ABC transporter, transmembrane component 2	0	328	0.050	Absent in soil
RNA 3'-terminal phosphate cyclase (EC 6.5.1.4)	0	300	0.029	Absent in soil
Succinate dehydrogenase cytochrome b560 subunit	0	383	0.042	Absent in soil
Transcriptional regulator of rhamnose utilization, DeoR family	0	1194	0.040	Absent in soil
UDP-glucose:(heptosyl) LPS alpha1,3-glucosyltransferase WaaG (EC 2.4.1.-)	0	551	0.037	Absent in soil

5-dehydro-4-deoxyglucarate dehydratase (EC 4.2.1.41)	8	7402	0.034	886.9
Eukaryotic translation initiation factor 4E	8	2580	0.023	309.1
Periplasmic protein p19 involved in high-affinity Fe2+ transport	7	1851	0.020	279.1
Nitric oxide -responding transcriptional regulator NnrR (Crp/Fnr family)	8	1869	0.013	223.9
Phosphoethanolamine transferase EptA specific for the 1 phosphate group of core-lipid A	7	1046	0.011	157.7
Outer membrane component of tripartite multidrug resistance system	8	1031	0.015	136.1
Two-component system histidine kinase	8	1109	0.019	132.9
Negative regulator of flagellin synthesis	17	2009	0.044	120.3
Nonheme iron-containing ferritin	22	2402	0.023	111.1
LptA, protein essential for LPS transport across the periplasm	39	3158	0.046	81.7
Alkanesulfonates-binding protein	201	14574	0.025	72.5
Glutathione-regulated potassium-efflux system protein KefC	23	1457	0.025	62.4
Cytochrome c-type heme lyase subunit nrfE, nitrite reductase complex assembly	13	802	0.021	60.4
Glutamine-dependent 2-keto-4-methylthiobutyrate transaminase	17	934	0.014	55.9
PhnJ protein	20	959	0.018	48.2
Transcriptional regulator NanR	45	1959	0.037	43.6
Flagellar hook-associated protein FliD	418	16118	0.045	38.6
Alanine racemase, catabolic (EC 5.1.1.1)	30	1148	0.050	38.3
Uncharacterized MobA-related protein	20	719	0.003	36.2
Plasmid replication protein RepA	68	2405	0.029	35.2
ABC-type tungstate transport system, periplasmic binding protein	70	2417	0.030	34.5
LysR family transcriptional regulator YeiE	32	1052	0.019	32.5
Sialic acid transporter (permease) NanT	16	512	0.014	31.6
Zinc ABC transporter, periplasmic-binding protein ZnuA	138	4293	0.035	31.2
Riboflavin transporter PnuX	53	1638	0.039	30.7
Nitric oxide reductase activation protein NorD	22	653	0.008	30.0
Predicted signal transduction protein	90	2513	0.035	28.1
Pyridoxine biosynthesis glutamine amidotransferase, glutaminase subunit (EC 2.4.2.-)	69	1824	0.045	26.4
Apolipoprotein N-acyltransferase (EC 2.3.1.-)	36	929	0.003	25.9
Lipopolysaccharide heptosyltransferase I (EC 2.4.1.-)	37	924	0.049	25.3
Predicted D-glucarate or D-galactorate regulator, GntR family	30	748	0.034	25.0
Rrf2 family transcriptional regulator	141	3399	0.008	24.1

D-erythrose-4-phosphate dehydrogenase (EC 1.2.1.72)	15	350	0.043	23.4
Eukaryotic translation initiation factor 2 beta subunit	125	2840	0.003	22.8
TolA protein	481	10945	0.006	22.8
6-phospho-beta-glucosidase (EC 3.2.1.86)	47	1035	0.005	22.2
ECF_class_transporters	184	4058	0.050	22.0
Sucrose phosphorylase (EC 2.4.1.7)	13	285	0.016	21.5
Coenzyme PQQ synthesis protein F (EC 3.4.99.-)	15	322	0.035	21.5
Negative regulator of flagellin synthesis FlgM	144	3079	0.034	21.4
membrane c-type cytochrome cy	226	4111	0.009	18.2
TcuC: integral membrane protein used to transport tricarballylate across the cell membrane	38	689	0.031	18.0
Potassium-transporting ATPase B chain (EC 3.6.3.12) (TC 3.A.3.7.1)	1875	32965	0.037	17.6
Cell division topological specificity factor MinE	233	4079	0.046	17.5
Eukaryotic translation initiation factor 3 39 kDa subunit	105	1781	0.018	16.9
ATP-dependent RNA helicase SrmB	191	2908	0.047	15.2
putative Adenosine kinase (EC 2.7.1.20)	153	2322	0.017	15.2
Glucarate dehydratase (EC 4.2.1.40)	575	8674	0.000	15.1
Ribonuclease E inhibitor RraA	160	2381	0.001	14.9
Predicted NAD regulator in Alphaproteobacteria	20	285	0.018	14.3
Predicted L-rhamnose isomerase RhaI (EC 5.3.1.14)	1738	24808	0.041	14.3
Proposed peptidoglycan lipid II flippase MurJ	340	4703	0.031	13.8
Potassium-transporting ATPase A chain (EC 3.6.3.12) (TC 3.A.3.7.1)	1779	23320	0.032	13.1
Methylated-DNA--protein-cysteine methyltransferase (EC 2.1.1.63)	332	3942	0.036	11.9
Phosphoribosylformylglycinamidine synthase, PurS subunit (EC 6.3.5.3)	263	3131	0.018	11.9
RNA polymerase sigma factor SigB	4357	50724	0.043	11.6
2-C-methyl-D-erythritol 4-phosphate cytidylyltransferase (EC 2.7.7.60)	197	2282	0.050	11.6
Histidine ABC transporter, ATP-binding protein HisP (TC 3.A.1.3.1)	307	3526	0.010	11.5
Menaquinone via futalosine step 1	119	1353	0.015	11.4
Uncharacterized protein, similar to the N-terminal domain of Lon protease	347	3884	0.019	11.2
Glycerol-3-phosphate regulon repressor, DeoR family	141	1583	0.003	11.2
putative Cytochrome bd2, subunit I	1221	13626	0.014	11.2
SAM-dependent methyltransferase 2, in cluster with Hydroxyacylglutathione hydrolase (EC 3.1.2.6)	96	1055	0.012	10.9
Alginate biosynthesis protein Alg8	40	427	0.044	10.7

Pyrrolidone-carboxylate peptidase (EC 3.4.19.3)	16	169	0.047	10.6
Translation initiation factor SUI1	371	3881	0.022	10.5
Malonate decarboxylase alpha subunit	463	4838	0.028	10.5
Hypothetical protein DUF454	22	226	0.024	10.4
Biosynthetic Aromatic amino acid aminotransferase beta (EC 2.6.1.57)	630	6555	0.005	10.4
Alkanesulfonates transport system permease protein	441	4576	0.039	10.4
Flp pilus assembly protein CpaD	31	315	0.036	10.2
Transport ATP-binding protein CydC	42	412	0.031	9.9
Hypothetical protein Q, similar to Chlorite dismutase	327	3213	0.035	9.8
Inositol transport system sugar-binding protein	354	3464	0.002	9.8
Putative permease often clustered with de novo purine synthesis	199	1920	0.047	9.7
Ribosomal RNA small subunit methyltransferase E (EC 2.1.1.-)	323	3114	0.033	9.6
Potassium-transporting ATPase C chain (EC 3.6.3.12) (TC 3.A.3.7.1)	168	1600	0.001	9.5
sulfonate monooxygenase	485	4516	0.003	9.3
Outer membrane stress sensor protease DegS	1244	11400	0.040	9.2
Eukaryotic translation initiation factor 5A	3895	35403	0.038	9.1
Catechol 1,2-dioxygenase 1 (EC 1.13.11.1)	53	475	0.030	9.0
PTS system, glucitol/sorbitol-specific IIB component and second of two IIC components (EC 2.7.1.69)	27	238	0.020	9.0
Cystathionine beta-lyase (EC 4.4.1.8)	739	6562	0.039	8.9
Pantothenate kinase (EC 2.7.1.33)	459	4060	0.049	8.8
Tol biopolymer transport system, TolR protein	507	4455	0.050	8.8
D-beta-hydroxybutyrate permease	198	1713	0.033	8.7
DNA-binding protein Fis	919	7915	0.026	8.6
L-alanine-DL-glutamate epimerase	120	1027	0.044	8.6
ATP synthase B' chain (EC 3.6.3.14)	403	3388	0.020	8.4
Transport_of_Nickel_and_Cobalt	381	3162	0.022	8.3
Tetraacyldisaccharide 4'-kinase (EC 2.7.1.130)	66	544	0.049	8.3
Manganese ABC transporter, ATP-binding protein SitB	175	1447	0.035	8.3
Arsenical pump-driving ATPase (EC 3.6.3.16)	732	6045	0.031	8.3
Glucokinase (EC 2.7.1.2)	397	3254	0.027	8.2
GTP cyclohydrolase II (EC 3.5.4.25)	1002	7848	0.011	7.8
NAD-dependent formate dehydrogenase gamma subunit	199	1547	0.001	7.8

Ribonucleotide reductase of class Ib (aerobic), alpha subunit (EC 1.17.4.1)	1597	12420	0.002	7.8
Phosphocarrier protein kinase/phosphorylase, nitrogen regulation associated	798	6196	0.033	7.8
O-acetylhomoserine sulfhydrylase (EC 2.5.1.49)	1755	13380	0.002	7.6
Copper metallochaperone, bacterial analog of Cox17 protein	99	754	0.017	7.6
Phosphatidylserine decarboxylase (EC 4.1.1.65)	1419	10430	0.001	7.4
response regulator in two-component regulatory system with PhoQ	1324	9612	0.003	7.3
Phytoene synthase (EC 2.5.1.32)	759	5413	0.020	7.1
Glutathione reductase (EC 1.8.1.7)	1411	10065	0.031	7.1
Adenosine (5')-pentaphospho-(5'')-adenosine pyrophosphohydrolase (EC 3.6.1.-)	3597	25520	0.038	7.1
Selenoproteins	94345	669236	0.046	7.1
Ccs1/ResB-related putative cytochrome C-type biogenesis protein	238	1684	0.018	7.1
Flagellar basal-body P-ring formation protein FlgA	98	694	0.018	7.1
Glutathione S-transferase family protein	1883	12962	0.036	6.9
RNA polymerase sporulation specific sigma factor SigH	259	1764	0.034	6.8
Erythritol phosphate dehydrogenase EryB	88	594	0.044	6.8
N-formylglutamate deformylase (EC 3.5.1.68)	271	1805	0.033	6.7
N-acetylglucosamine-6-phosphate deacetylase (EC 3.5.1.25)	295	1922	0.009	6.5
2',3'-cyclic-nucleotide 2'-phosphodiesterase (EC 3.1.4.16)	331	2132	0.001	6.4
Eukaryotic peptide chain release factor GTP-binding subunit	643	4129	0.002	6.4
Putative two-component sensor histidine kinase	899	5765	0.036	6.4
Arsenical resistance operon repressor	858	5479	0.002	6.4
Endonuclease IV (EC 3.1.21.2)	223	1404	0.011	6.3
Leucyl/phenylalanyl-tRNA--protein transferase (EC 2.3.2.6)	594	3697	0.037	6.2
4-hydroxy-3-methylbut-2-enyl diphosphate reductase (EC 1.17.1.2)	8068	48567	0.013	6.0
DNA repair protein RecN	1652	9707	0.033	5.9
Aspartokinase (EC 2.7.2.4)	17682	102525	0.046	5.8
Probable VANILLIN dehydrogenase oxidoreductase protein (EC 1.-.-.-)	177	1022	0.036	5.8
Phosphate ABC transporter, periplasmic phosphate-binding protein PstS (TC 3.A.1.7.1)	13345	75811	0.050	5.7
Predicted glycogen synthase, ADP-glucose transglucosylase (EC 2.4.1.21), Actinobacterial type	330	1869	0.001	5.7
Foldase protein PrsA precursor (EC 5.2.1.8)	908	5112	0.018	5.6
Flavodoxin reductases (ferredoxin-NADPH reductases) family 1	4949	27778	0.041	5.6
D-serine/D-alanine/glycine transporter	456	2553	0.006	5.6

Sulfate and thiosulfate binding protein CysP	17067	94874	0.046	5.6
Biogenesis_of_cytochrome_c_oxidases	23244	128652	0.045	5.5
Multiple polyol-specific dehydrogenase (EC 1.1.1.-)	306	1636	0.013	5.3
Gluconate transporter family protein	351	1870	0.022	5.3
Ribosomal subunit interface protein	5120	27212	0.003	5.3
Cytochrome oxidase biogenesis protein Surf1, facilitates heme A insertion	396	2076	0.027	5.2
Glutamate-ammonia-ligase adenylyltransferase (EC 2.7.7.42)	773	4016	0.049	5.2
tRNA-specific adenosine-34 deaminase (EC 3.5.4.-)	815	4206	0.049	5.2
Phosphate starvation-inducible protein PhoH, predicted ATPase	12767	65749	0.036	5.2
Fructose-bisphosphate aldolase class II (EC 4.1.2.13)	20047	102773	0.037	5.1
Carboxynorspermidine dehydrogenase, putative (EC 1.1.1.-)	1691	8513	0.030	5.0
3-oxoacyl-[acyl-carrier protein] reductase (EC 1.1.1.100)	15843	78975	0.023	5.0
Pyruvate dehydrogenase E1 component alpha subunit (EC 1.2.4.1)	12981	64653	0.024	5.0
Omega-amino acid--pyruvate aminotransferase (EC 2.6.1.18)	6126	30458	0.038	5.0
Twin-arginine translocation protein TatB	2242	11027	0.022	4.9
Putative deoxyribonuclease YcfH	1367	6712	0.044	4.9
Adenylosuccinate lyase (EC 4.3.2.2)	9702	47538	0.047	4.9
Lactate 2-monooxygenase (EC 1.13.12.4)	58	285	0.017	4.9
Ribonucleotide reductase transcriptional regulator NrdR	1793	8724	0.010	4.9
Cell division protein FtsI [Peptidoglycan synthetase] (EC 2.4.1.129)	2126	10320	0.003	4.9
GTP cyclohydrolase I (EC 3.5.4.16) type 1	11500	55690	0.026	4.8
Glucose-6-phosphate isomerase (EC 5.3.1.9)	11171	53450	0.040	4.8
Galactofuranosyl transferase (EC 2.-.-.-)	141	668	0.028	4.7
Beta-ureidopropionase (EC 3.5.1.6)	905	4294	0.036	4.7
Dihydroxyacetone kinase family protein	373	1748	0.030	4.7
SOS-response repressor and protease LexA (EC 3.4.21.88)	3802	17647	0.040	4.6
carbazol_degradation_cluster	970	4499	0.050	4.6
Sulfate transport system permease protein CysW	2298	10630	0.045	4.6
photosystem I P700 chlorophyll a apoprotein subunit Ib (PsaB)	282	1295	0.043	4.6
Pyrimidines	6367	29150	0.038	4.6
Oxidase	687	3104	0.014	4.5
Plant Octadecanoids	687	3104	0.014	4.5

Adenylate cyclase (EC 4.6.1.1)	6725	30112	0.046	4.5
Radical SAM family enzyme, similar to coproporphyrinogen III oxidase, oxygen-independent, clustered with nucleoside-triphosphatase RdgB	1236	5526	0.030	4.5
Cytidine deaminase (EC 3.5.4.5)	434	1935	0.013	4.5
Potassium efflux system KefA protein / Small-conductance mechanosensitive channel	1711	7617	0.046	4.5
Acetyl-CoA acetyltransferase (EC 2.3.1.9)	10612	47135	0.038	4.4
Methionine aminotransferase, PLP-dependent	273	1209	0.028	4.4
ZZ_gjo_need_homes	28960	127840	0.040	4.4
Beta-lactamase class C and other penicillin binding proteins	1257	5433	0.028	4.3
Miscellaneous	41236	175763	0.037	4.3
Sulfate transport system permease protein CysT	2540	10389	0.030	4.1
NAD kinase (EC 2.7.1.23)	1659	6735	0.024	4.1
Recombination protein RecR	4729	18906	0.048	4.0
Menaquinone via futalosine step 3	5543	21958	0.033	4.0
NADH-ubiquinone oxidoreductase chain D (EC 1.6.5.3)	68591	268844	0.038	3.9
5-methyltetrahydrofolate--homocysteine methyltransferase (EC 2.1.1.13)	30761	119044	0.049	3.9
Cytochrome c oxidase polypeptide II (EC 1.9.3.1)	19575	75556	0.045	3.9
Periplasmic aromatic aldehyde oxidoreductase, FAD binding subunit YagS	834	3157	0.041	3.8
Glutamine, glutamate, aspartate, asparagine; ammonia assimilation	32601	122063	0.049	3.7
Aspartate ammonia-lyase (EC 4.3.1.1)	11712	43797	0.022	3.7
Ribosomal-protein-S18p-alanine acetyltransferase (EC 2.3.1.-)	421	1534	0.049	3.6
polyhydroxyalkanoate synthesis repressor PhaR	3133	11407	0.034	3.6
DinG family ATP-dependent helicase CPE1197	200	724	0.005	3.6
POSSIBLE ACYL-[ACYL-CARRIER PROTEIN] DESATURASE DESA2 (ACYL-[ACP] DESATURASE) (STEAROYL-ACP DESATURASE)	417	1506	0.037	3.6
Adenosine deaminase (EC 3.5.4.4)	1979	7128	0.015	3.6
Sulfite reductase [NADPH] flavoprotein alpha-component (EC 1.8.1.2)	1087	3845	0.033	3.5
YbbK	11670	40746	0.035	3.5
DNA topoisomerase IB (poxvirus type) (EC 5.99.1.2)	1191	4122	0.004	3.5
Excinuclease ABC subunit C	8543	29545	0.036	3.5
Vanillate O-demethylase oxygenase subunit (EC 1.14.13.82)	858	2911	0.007	3.4
3-isopropylmalate dehydratase small subunit (EC 4.2.1.33)	12533	42117	0.035	3.4
Putative stomatin/prohibitin-family membrane protease subunit aq_911	3052	9626	0.026	3.2

Neopullulanase (EC 3.2.1.135)	670	2032	0.032	3.0
Ku domain protein	4762	13766	0.008	2.9
Aerobic cobaltochelatase cobS subunit (EC 6.6.1.2)	11334	32444	0.049	2.9
Glutamine synthetase (EC 6.3.1.2)	594	1622	0.045	2.7
Predicted nucleoside ABC transporter, substrate-binding component	1167	3052	0.036	2.6
Sulfite oxidase	2523	6379	0.023	2.5
Serine--glyoxylate aminotransferase (EC 2.6.1.45)	3999	9837	0.047	2.5

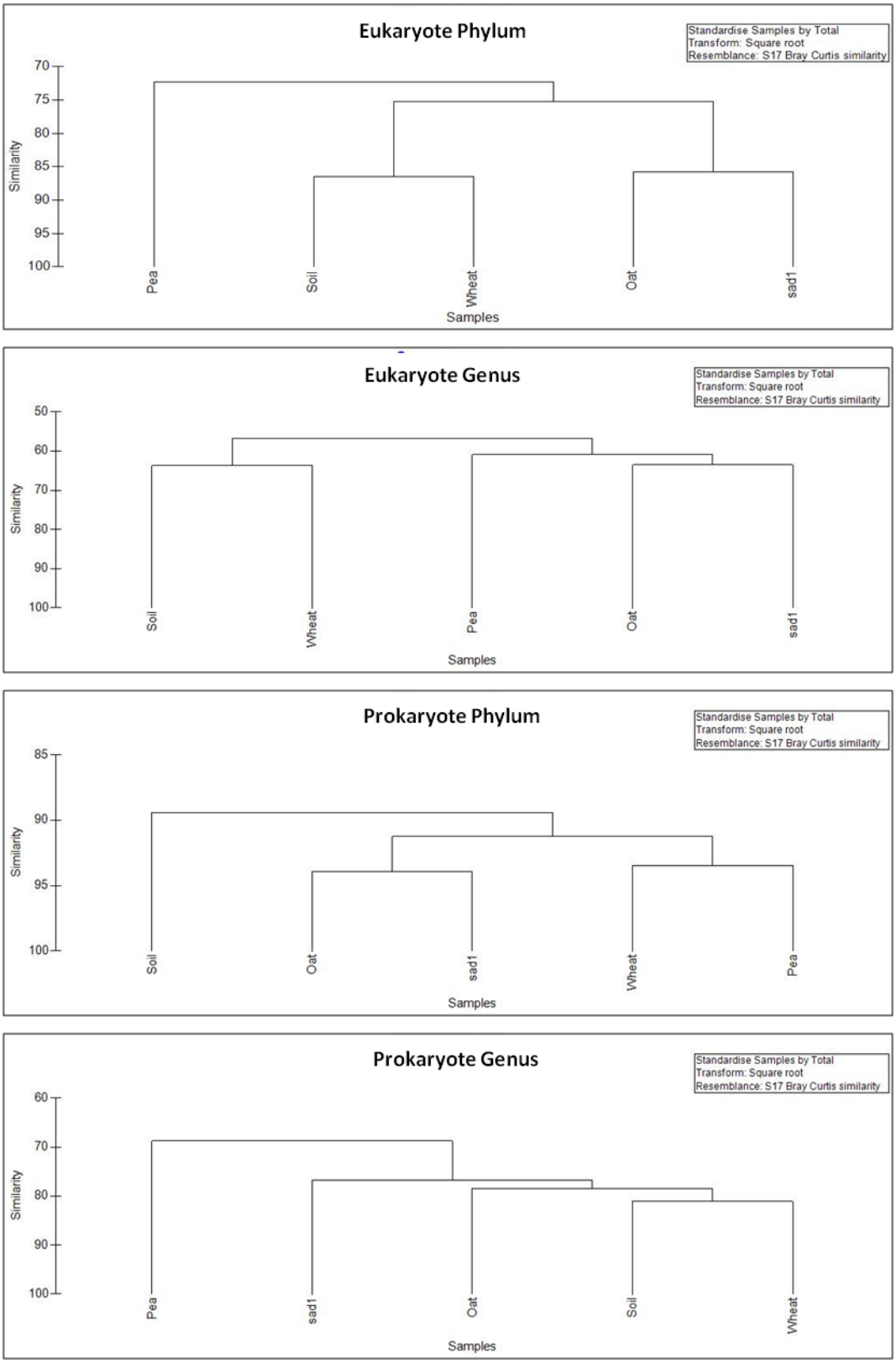
Table A32: List of abbreviations used throughout the theis, either initially defined in text or common abbreviations.

Abbreviation	Meaning	Abbreviation	Meaning
A_{260nm}	Absorbance at 260 nm	LSU	Large subunit
ABC	ATP-binding cassette	MAMP	Microbe-associated molecular pattern
ACC	Aminoacyclopropane-1-carboxylic acid	MAPKKK	Mitogen-activated protein kinase kinase kinase
AGP	Arabinogalactan protein	MDS	Multidimensional scaling
AMF	Arbuscular mycorrhizal fungi	MID	Multiplex identifier
ARISA	Automated ribsosomal intergenic spacer analysis	MMLV	Murine leukemia virus
ATP	Adenosine triphosphate	mRNA	Messenger RNA
ATPase	Adenosine triphosphatase	N	Normality
BAC	Bactrial artificial chromosome	NAD+	Nicotinamide adenine dinucleotide , oxidised
bp	Base pairs	NADH	Nicotinamide adenine dinucleotide, reduced
Bt	*Bacillus thuringiensis*	NCBI	National Centre for Biotechnology Information
bv	Biovar	ncRNA	Non-coding RNA
C1	One carbon	NO	Nitric oxide
C4	Four carbon	OTU	Operational taxonomic unit
CCC	Capacity and Capability Challenge	PAMP	Pathogen-associated molecular pattern
cDNA	copy DNA	PCA	Principle component analysis
CoA	Co-enzyme A	PCR	Polymerase chain reaction
Ct	Threshold cycle	PGPR	Plant growth promoting rhizobacteria
DAPG	2,4-Diacetylphloroglucinol	PQQ	Pyrroloquinoline quinone
DGGE	Denaturing gradient gel electrophoresis	PTI	PAMP triggered immunity

DMSP	Dimethylsulphionoproprionate	qPCR	quantitative PCR
DMT	Drug/metabolite transporter	qRT-PCR	quantitative reverse transcription PCR
DNA	Deoxyribonucleic acid	rDNA	Ribosomal DNA
DNase	Deoxyribonuclease	RIS	Ribonucleic acid internal standard
dsDNA	Double stranded DNA	RNA	Ribonucleic acid
DSN	Duplex specifc nuclease	RNase	Ribonuclease
EC	Enzyme commission	RND	Resistance/nodulaiton/division class
EDTA	Ethytlene diamine tetraacetic acid	ROS	Reactive oxygen species
Ef-Tu	Elongation factor thermo unstable	rpm	Revolutions per minute
ETI	Effector triggered imuunity	rRNA	Ribosomal RNA
FAM	6-carboxyfluoroscein	SAR	Systemic acquired resistance
Gb	Gigabase pairs	SEM	Standard error of the mean
GB	Gigabytes	shc	Squalene hopene cyclase
gDNA	genomic DNA	sRNA	Small RNA
GMP	Guanosine monophosphate	SSC	Saline sodium citrate
GO	Gene Ontology	SSU	Small subunit
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	T_a	Annealing temperature
IAA	Indole-3-acetic acid	TBE	Tris-EDTA boric acid
ISR	Induced systemic resistance	TE	Tris-EDTA
ITS	Internally transcribed spacer	TRFLP	Terminal restriction fragment length polymorphism
KO	KEGG ortholog	Tris-HCl	tris(hydroxymethyl)aminomethane hydrochloride
LB	Luria-Bertani	tRNA	Transfer ribonucleic acid
LCA	Lowest common ancestor	va	Variety
LIB	Library	WTA	Whole transcriptome amplification
LRR	Leucine rich repeat		

Table A33: Soil chemical analyses perfomed by Macualay Soils (Aberdeen, UK).

Soil Code	Sample Location	Date	pH	Nitrate (mg/kg NO^{3-})	Phosphorus (mg/kg P)	Potassium (mg/kg K)	Magnesium (mg/kg Mg)	Organic Matter (%)
JIC1	John Innes	March 2009	7.49	149	171.4	184.4	60.3	2.3
Baw3	Bawburgh	July 2012	7.53	28.69	136.8	167.4	63.55	2.79
Baw4	Bawburgh	August 2012	7.51	39.18	135.2	183.9	39.88	2.62



Supplementary Figure A1: Bray-Curtis similarity tree based on read counts used in Figure 3.7. Data were normalised and square root transformed and trees generated using Primer E.