The microbiology of lean and obese soil

Frances Patricia Jones

September 2016

Department of Geography and Environmental Science, University of Reading, UK

Department of AgroEcology, Rothamsted Research, UK

Submitted in partial fulfilment of the requirement for the degree of Doctor of Philosophy

Abstract

The bacterial genus *Bradyrhizobium* is biologically important within soils, with different representatives found to perform a range of functions including nitrogen fixation through symbioses, photosynthesis and denitrification. The Highfield experiment at Rothamsted provides an opportunity to study the impact of plants on microbial communities as it has three long-term contrasting regimes; permanent grassland, arable and bare fallow (devoid of plants). The bare fallow plots have a significant reduction in soil carbon and microbial biomass. *Bradyrhizobium* has been shown by metagenomic studies on soil to be one of the most abundant and active groups including in bare fallow soil indicating that some phenotypes are adapted to survive in the absence of plants.

A culture collection was created with isolates obtained from contrasting soil types from Highfield in addition to woodland soil, gorse (*Ulex europeaus*) and broom (*Cytisus scoparius*) root nodules. The collection's phylogeny has been explored by sequencing housekeeping genes to determine whether soil treatment affects the core genome. One grassland and one bare fallow isolate had their genome sequenced and differences have been assessed to establish their potential for a range of functions and to direct future experiments. The functional diversity of the collection has been investigated using carbon metabolism assays to identify key substrates and determine whether the isolates group according to soil treatment. Symbiosis capacity and role in nitrogen cycling has been examined using nodulation tests, anaerobic growth on nitrate and nitrous oxide production and reduction through denitrification.

A high level of diversity can be seen throughout the collection with differences being linked to niche adaptation. Understanding more about *Bradyrhizobium* could give clues on how above ground management impacts a key group within the soil community. Furthermore, the first assembled genomes of two non-symbiotic *Bradyrhizobium* strains isolated from soil provide an important resource for microbiology and soil ecology.

Table of contents

List of tables

List of figures

Abbreviation list

2,4-D - 2, 4-dichlorophenoxyacetic acid

ANOVA - Analysis of variance

Ar - Argon

ATP - Adenosine triphosphate

BLAST - Basic Local Alignment Search Tool

bv - Biovar

bp - Base pairs

C - Carbon

C_2H_2 - Acetylene

CH_4 - Methane

Cu - Copper

cv - Cultivar

df - Degrees of freedom

ERIC - Enterobacterial repetitive intergenic consensus

EtBr - Ethidium bromide

GLM - Generalised linear model

kb - Kilobase

Kbp - Kilobase pairs

KNO_3 - Potassium nitrate

iTOL - Interactive tree of life

ITS - Internal transcribed spacer

MAG - Modified arabinose gluconate

Mbp - Megabase pairs

MLSA - Multilocus sequence analysis

N - Nitrogen

N_2 - Dinitrogen gas

NCBI - National Center for Biotechnology Information

N_2O - Nitrous oxide

NH_4^+ - Ammonium

NH_3 - Ammonia

NO - Nitric oxide

NO_3^- - Nitrate

NO_2^- - Nitrite

OD - Optical density

PC - Principal component

PCA - Principal components analysis

PGPB - Plant growth promoting bacteria

ppm - Parts per million

qPCR - Quantitative real time polymerase chain reaction

RDP - Ribosomal Database Project

rRNA - Ribosomal RNA

rpm - Revolutions per minute

sdH_2O - Sterile distilled water

SNP - Single nucleotide polymorphism

sv - Serovar

TBE - Tris/Borate/EDTA buffer

tRNA - Transfer RNA

UK - United Kingdom

USDA - United States Department of Agriculture

xg - Relative centrifugal force

YEM – Yeast extract mannitol medium

Statement of original authorship

Declaration

I confirm that this is my own work and the use of all material from other sources has been properly and fully acknowledged.

Frances Patricia Jones

Acknowledgements

I would like to take this opportunity to thank the large number of people who have helped me throughout my PhD. Firstly, I would like to thank my supervisors for their guidance and support over the past four years; Prof Penny Hirsch and Ian Clark at Rothamsted Research and Dr Liz Shaw and Prof Martin Woodward at the University of Reading. I was very grateful to have a supervisory team who were always willing to answer questions and able to offer a wealth of both theoretical and practical help and knowledge. I would also like to thank Dr Rob King at Rothamsted Research for his help and guidance through my adventure into the world of bioinformatics and genome assembly. Also, Dr Steve Hanley for the Ion Torrent sequencing he performed. I would also like to thank the statisticians at Rothamsted Research who have helped me analyse my data; Suzanne Clark, Dr Andrew Mead and Dr Kirsty Hassall. I would like to thank the various past and present members of the Molecular Microbial Ecology group at Rothamsted Research for making the last four years so enjoyable.

I would like to thank the Biotechnology and Biosciences Research Council (BBSRC), the University of Reading and the Lawes Agricultural Trust for funding this work. In addition, to the Federation of European Microbiological Societies and the Society for Applied Microbiology for travel grants which funded international travel to conferences to allow me to present this work. I would also like to thank the members of the BBSRC Doctoral Training Partnership in Food Security for making this new scheme so worthwhile. I would like to thank all of the people I was fortunate to work with at the Royal Institution during my three-month professional industrial placement. I was astounded by the enthusiasm and passion each and every one of the people I met had for science communication and the Royal Institution. The placement allowed me to experience the world of science communication and has changed the way I think about my own research and how I can communicate this to the public and other scientists. The skills, knowledge and experiences I gained from spending three months at the Royal Institution will be invaluable in my future career and I can't thank them enough.

I would like to thank my husband, Tarran. Without years of his unfaltering love, support and encouragement, my PhD experience would have been a very different one. I would also like to thank my mum, dad and my sister, Lizzie for their support and being the listening ear when science doesn't always go to plan! I would also like to thank all of the friends I have made throughout my PhD and for making Rothamsted Research such an amazing place to work.

Dedication

For Gran

1 Introduction

1.1 Soil microbial communities

Soil is the most diverse ecosystem on Earth with up to a billion individual bacterial cells and 18,000 genomes per gram and is vital for human and plant health (Reynolds et al., 2003, Delmont et al., 2012, Nacke et al., 2014, Mellado-Vazquez et al., 2016). The large abundance of organisms and the heterogeneity of the soil limits diversity analyses of the soil environment (Hirsch et al., 2010). Traditionally, only bacteria which could be cultured in the laboratory were studied but with the development of molecular techniques, the estimates of soil diversity increased rapidly (Shade et al., 2012). It is now widely accepted that when dealing with soil communities, both culturable and unculturable organisms should be assessed (Nichols, 2007, Hirsch et al., 2010, Delmont et al., 2012, Shade et al., 2012, Lakshmanan et al., 2014). A large proportion of soil organisms, thought to be between 95% and 99%, are considered to be unculturable due to specific requirements for growth including sensitivity to oxygen levels, temperature, pH or the presence or absence of particular nutrients (Janssen et al., 2002, Sait et al., 2002, da Rocha et al., 2009, Vartoukian et al., 2010, Delmont et al., 2012). Culturability has also been shown to decrease with increasing depth of soil sampling from 19% in the first 2 cm to 2.4% for between 8cm and 10 cm depth (Sait et al., 2002). This has led to the development of metagenomic studies as this method does not require any form of culturing. The DNA can be extracted directly from the soil and sequenced allowing for a snapshot in time of the species present and their abundances (Rappe and Giovannoni, 2003, da Rocha et al., 2009, Hirsch et al., 2010). Despite metagenomics allowing for the assessment of unculturable bacteria, rarer species will not be included in the metagenomic datasets as their DNA will be at much lower abundance relative to other DNA from more abundant species (Caporaso et al., 2011, Lynch and Neufeld, 2015). Caporaso et al. (2011) showed that only species occurring more than 10,000 times were considered reliable for diversity analyses in the metagenome of synthetic communities with a sequencing depth of 3.1 million reads per sample.

Isolation of novel members of the soil community allow for experiments to be carried out to understand what processes these isolates are involved in and give clues about rarer members of the community often missed by culture independent approaches (Janssen et al., 2002, Davis et al., 2011, Shade et al., 2012). Davis et al. (2011) showed isolation conditions including the length of incubation of isolation plates, are critical when isolating novel organisms including members of Acidobacteria and Chloroflexi, with a greater abundance of novel cultures visible after more than 12 weeks of incubation. It has also been shown that culturable bacteria from a range of

phylogenetic groups have variable incubation times. Alpha-Proteobacteria and Firmicutes are often considered slower growers whereas Actinobacteria for example *Arthrobacter* are faster growers (Mitsui et al., 1997). The variability in growth needs to be considered when looking into culture-dependent community data as faster growers will form a large proportion of the biomass.

Soil microorganisms show high heterogeneity and are often grouped according to taxonomy, lifestyle, location or particular pathways they are involved in for example carbon or nitrogen metabolism. They are often described as copiotrophs, organisms that exploit labile carbon, or oligotrophs, organisms capable of growing at very low levels of nutrient availability, sometimes decomposing recalcitrant organic carbon and plant matter (Shade et al., 2012, Cesarz et al., 2013, Mellado-Vazquez et al., 2016). A third group, the autotrophs do not require complex nutrients and obtain energy by photosynthesis or chemical reactions. Copiotrophs are able to use a range of nutrients when available and usually have faster growth rates. In contrast, oligotrophs are organisms which thrive in nutrient poor environments and have evolved mechanisms to cope with a range of environmental conditions for example, slow growth and higher substrate affinities (Koch, 2001, Fierer et al., 2007, Mellado-Vazquez et al., 2016). Carbon is often the limiting factor for microbial growth but nitrogen and phosphorus limitation also play an important role (Alden et al., 2001, Ekblad and Nordgren, 2002, Ilstedt and Singh, 2005). Both copiotrophs and oligotrophs can survive in nutrient poor conditions but, with the exception of copiotrophs with dormancy strategies such as sporulation, it is only oligotrophs that are able to survive in these conditions over a long period of time (Koch, 2001, Fierer et al., 2007). This is not the case when in a nutrient rich environment; oligotrophs are unable to compete and survive in these environments (Koch, 2001). The rhizosphere, rich in available carbon, is thought to be a hotspot for copiotrophic bacteria (Bird et al., 2011, Mellado-Vazquez et al., 2016).

Soil bacteria sequenced have shown to have large genome sizes and typical genomes are between 6.5 Mb and 9 Mb with nitrogen-fixing bradyrhizobia typically being 9Mb whereas the minimum genome size for *Escherichia coli* is 4 Mb (Perna et al., 2001, Kaneko et al., 2002, Gonzalez et al., 2006, Young et al., 2006, Kaneko et al., 2011). The bacterial genome is often described as consisting of two parts; a core genome and an accessory genome. The core genome contains genes which are highly conserved required for basic cellular metabolism and growth or "housekeeping", and are present in all bacteria. The accessory genome is more variable, often transferred between related lines, and contains functions which could reflect the environmental conditions (Young et al., 2006, Sugawara et al., 2013, Mauchline et al., 2014, Remigi et al., 2016).

Both soil biomass and metabolic activity is dominated by both bacteria and fungi (Alden et al., 2001). A huge range of processes occur in the soil and a large number of these are controlled by

the microbial community (Bastida et al., 2013). These include large parts of the nitrogen cycle, carbon cycling, degradative functions, decomposition, plant growth promotion and plant and animal pathogenesis (Ilstedt and Singh, 2005, Delmont et al., 2012, Lakshmanan et al., 2014, Nacke et al., 2014). Understanding productivity and ways to improve agricultural sustainability is underpinned by studying these soil processes and the ways in which microbial communities adapt and are influenced by a range of abiotic and biotic factors (Delmont et al., 2012, Yang et al., 2013, Lakshmanan et al., 2014, Nacke et al., 2014, Lareen et al., 2016).

1.2 Abiotic and biotic effects on soil microbial communities

Microbial community composition and abundance can be altered by a range of both abiotic and biotic factors. Abiotic factors include pH, moisture and carbon to nitrogen ratio and biotic factors include presence of plants and their root exudates and soil meso- and microfauna (Lakshmanan et al., 2014, Sagova-Mareckova et al., 2015, Kumar et al., 2016). Microbial communities have relatively rapid generation times compared to meso- and microfauna and so are able to quickly adapt to utilise substrates available to them. Therefore, they can be heavily influenced by their environment and populations are determined by energy and nutrient availability (Joergensen and Scheu, 1999, Porter and Rice, 2013, Nunan et al., 2015).

Microbial communities in soil have major impacts on soil structure, organic matter decomposition and nutrient cycling including the nitrogen and carbon cycle (Garbeva et al., 2004, Hogberg and Read, 2006, Sagova-Mareckova et al., 2015). In turn, the soil structure itself can impact the microbial community as it is an important feature in the functioning of the soil ecosystem with impacts on water retention, crop yield, distribution of organic matter and nutrient recycling and the decline in soil structure is used as a measure of degradation in agricultural soils (Golchin et al., 1994, Bronick and Lal, 2005). Pollution can also affect microbial populations by acting as inhibitors or stimulants for certain groups within the soil or providing an energy source for those capable of metabolising the pollutant itself (Czechowska and van der Meer, 2012). Soil pH has been shown to affect both fungal and bacterial communities with bacteria being most sensitive (Bastida et al., 2013, Cesarz et al., 2013, Sagova-Mareckova et al., 2015). For example, populations of *Rhizobium* have been shown to increase in abundance and an increase in nodulation was noted when lime was applied with only a slight increase in soil pH (Andrade et al., 2002). The mineralisation of organic matter in soil is closely linked with the presence of particular groups of bacteria. Increased rates of carbon mineralisation were noted when the populations of Acidobacteria declined and beta-Proteobacteria and Bacteroidetes increased (Fierer et al., 2007).

Plants influence the physical structure, chemical composition and hydrological properties of the soil through root growth, senescence, release of exudates resulting in alterations in carbon, nitrogen and water availability (Kowalchuk et al., 2002, Lakshmanan et al., 2014, Bell et al., 2015, Sagova-Mareckova et al., 2015). The alterations in the chemical and physical properties alter the structure of the root microbiome (Chen et al., 2015, Lareen et al., 2016). Exopolysaccharides and biofilms produced by soil microorganisms in addition to organic matter, clay and root exudates bind soil particles into aggregates increasing the aggregate stability in the rhizosphere compared to the bulk soil (Golchin et al., 1994, Bronick and Lal, 2005, Bird et al., 2011, Gottel et al., 2011, Puga-Freitas and Blouin, 2015). Long-term aggregate stability is linked to recalcitrant carbon and the presence of metal ions and is often used as an indicator for structure that will support plant growth (Bronick and Lal, 2005).

Bare fallow soil is soil which contains no plant inputs and is regularly tilled to maintain plant-free soil (Hirsch et al., 2009). Carbon present in long-term bare fallow soils, deprived of plant inputs, is relatively stable and the microbial communities present are also stable showing no diversity changes with continuation of bare fallow conditions as they are adapted to metabolising simple substrates (Nunan et al., 2015). Bare fallow soil has been shown to contain 82% less organic carbon and arable soils contained 70% less organic carbon than grasslands (Hirsch et al., 2009). However, it has been shown by culturing communities in 96-well plates that bacteria are often able to metabolise substrates even when there have been many generations since exposure but to a lesser extent than those communities which have evolved in the presence of this substrate (Nunan et al., 2015). Rare microbial species are being revealed through DNA analysis and evidence from bacterioplankton in the Baltic Sea suggests that these rare members can become active and increase in abundance when organic carbon is added (Shade et al., 2012). The impact of bare fallow on soil microbial community composition is not clear. The species richness of microbial communities has been shown to be higher during crop years with differences in community composition being found between the first and second fallow years due to a legacy effect from the initial crop presence in the first fallow year (Castro et al., 2016). However, the presence of plants has also been shown to increase microbial biomass rather than microbial diversity (Hirsch et al., 2009, Chen et al., 2015). Microbial community composition is likely to depend on the soil, crop and climate.

Plant communities have been shown to alter microbial activity in soil with key bacterial groups for example *Burkholderia* and *Bradyrhizobium* exhibiting increased transcription in forest soils compared to grassland. Microbial community composition in forest soils is driven by intermediates and products of wood breakdown and activity is dominated by lignin and aromatic

compound degradation (Nacke et al., 2014). Diversity present in rainforests is particularly sensitive to land use change and this extends to below ground communities (Lima-Perim et al., 2016). Bacterial photosynthesis genes have been shown to have higher transcription rates in grasslands than in forests, thought to be due to increased exposure in grasslands to solar radiation (Nacke et al., 2014). Rainforest soils are highly diverse and the plant community composition is thought to be the biggest driver of this diversity (Lima-Perim et al., 2016). High plant diversity in grasslands has been associated with increased primary productivity thought to be due to the broad spectrum of resources and, as a result, is also associated with responses of soil microorganisms on plant growth (Latz et al., 2015, Sagova-Mareckova et al., 2015). Increased microbial activity in soil results in more nitrogen being made available to the plant and is predicted to lead to higher rates of photosynthesis where nitrogen is otherwise limited (Hamilton and Frank, 2001).

1.3 Plant-microbe interactions

Plant – soil microbe interactions have occurred since plants first began to colonise terrestrial environments around 700 million years ago and plants and microbes have co-evolved over this time (da Rocha et al., 2009, Puga-Freitas and Blouin, 2015, Kumar et al., 2016). Plant associations with fungi were established 400 million years ago and over 80% of plants can form associations with arbuscular mycorrhizal fungi (Fabre et al., 2015). There are three major classifications of plant-microbe interactions; saprophytic, pathogenic and beneficial (Lugtenberg et al., 2002, Morgan et al., 2005, van der Heijden et al., 2008, Kimbrel et al., 2013). These interactions are crucial for adaptation to the environment (Fabre et al., 2015, Kumar et al., 2016, Lareen et al., 2016). Microbes which decompose dead plant material are said to have a saprophytic interaction; this is an important stage in both nitrogen and carbon cycling within any ecosystem (Lugtenberg et al., 2002). Pathogenic interactions are those which the plant is infected with microbes which attack the stem, leaves or root which can result in impaired growth or tissue death (Lugtenberg et al., 2002, van der Heijden et al., 2008). Once a pathogenic interaction has been detected by the plant, a hypersensitive response (HR) is employed which results in localised tissue death to contain the infection and can prompt systemic acquired resistance (SAR) which provides non-specific resistance to pathogens (Baker et al., 1997). HR in particular can impact the rhizosphere microbial communities as plant tissue death would temporarily increase the quantity and composition of root exudates available to microbes (Compant et al., 2010). The root microbiome has major benefits for the plant including disease suppression, increased stress tolerance and can both directly and indirectly affect plant growth (Pieterse and Dicke, 2007, Lakshmanan et al.,

2014, Latz et al., 2015). For example, pseudomonads have been shown to produce a number of antibiotics which can protect the plant against plant pathogens and these mechanisms can respond to both biotic and abiotic cues (Latz et al., 2015).

The rhizosphere, the area immediately around the root, is the location of key biological processes in soil including nitrogen cycling, phosphorus solubilisation, disease suppression and the production of phytohormones by bacteria resulting in increased plant growth (da Rocha et al., 2009, Lakshmanan et al., 2014, Latz et al., 2015, Lareen et al., 2016). These processes are performed by the great abundance of microorganisms in the rhizosphere, which compete to utilise nutrients including carbon released from plant roots (DeAngelis et al., 2008, Compant et al., 2010, Bird et al., 2011, Gottel et al., 2011, Bell et al., 2015, Lareen et al., 2016). Plant roots release many organic compounds into their rhizosphere including monosaccharide sugars, phenols, lignin, hormones, proteins, organic acids and amino acids and the microbes in the soil are able to utilise a range of these to survive in the energy and nutrient limited conditions commonly found in soil (Steer and Harris, 2000, Bronick and Lal, 2005, Kardol et al., 2007, Bird et al., 2011, Lakshmanan et al., 2014, Kumar et al., 2016, Lareen et al., 2016). Root exudates from living plants are able to prime soil organic matter decomposition and this phenomenon can be monitored at an ecosystem carbon exchange level (Bird et al., 2011). Approximately 11% of all fixed carbon will be released as root exudates and this also accounts for 27% of all carbon which is allocated to the plant roots and half of the carbon released from microbial respiration is derived from recent photosynthate (Hogberg and Read, 2006, Balasooriya et al., 2013, Lakshmanan et al., 2014). Plant derived carbon is incorporated into the soil either as dead organic matter or root exudates (Cesarz et al., 2013).

The root microbiome is often considered to be an extension of the plant genome as plants are able to recruit members into their rhizosphere and benefit from associations with microbes and compounds produced from microbial processes (Lakshmanan et al., 2014, Lareen et al., 2016). Soil microorganisms interact with plants through hormone signalling pathways which can select for specific microbes and can therefore impact on levels of local nutrient cycling, be beneficial for example promoting plant growth or can be detrimental interaction by attracting plant pathogens (Puga-Freitas and Blouin, 2015, Kumar et al., 2016). The *Arabidopsis* root microbiome was shown to be altered due to the release of jasmonic acid and salicylic acid signalling pathways in response to both abiotic and biotic stress (Lakshmanan et al., 2014). Some bacteria are chemoattracted to root exudates by chemotaxis, controlled by the *cheA* gene, where cells will move towards the exudate to colonise the rhizosphere (Compant et al., 2010). Bacterial quorum sensing can lead to gene expression regulation in response to specific exudates and therefore is also vital to root and

rhizosphere colonisation (DeAngelis et al., 2008, Compant et al., 2010). There is often a distinction between rhizosphere and bulk soil organisms with the rhizosphere containing up to ten times the number of bacteria found in bulk soil and with higher levels of activity in the rhizosphere (Steer and Harris, 2000, Kowalchuk et al., 2002, DeAngelis et al., 2008). The rhizosphere is dominated by copiotrophic bacteria (many Gram-positive Proteobacteria such as *Pseudomonas*) which utilise simple carbon sources released by plants as root exudates. In contrast, bulk soil is often dominated by oligotrophs, slower growing organisms able to utilise more complex substrates (Bird et al., 2011, Mellado-Vazquez et al., 2016). Changes in bulk soil microbial communities are distinct from those of the rhizosphere: plant-soil-microbiome feedbacks shift the community structure and function (da Rocha et al., 2009, Bell et al., 2015). These shifts can also be caused by changes to pH and water availability which are influenced by the presence of plants creating niches within the rhizosphere for specific soil microbes (da Rocha et al., 2009, Lakshmanan et al., 2014, Bell et al., 2015, Chen et al., 2015, Sagova-Mareckova et al., 2015, Lareen et al., 2016).

Plant species, plant genotype and plant growth stage have been found to alter microbial communities and plants have evolved the ability to shape their root microbiome; however, the major drivers of this recruitment are not clear (Lakshmanan et al., 2014, Bell et al., 2015, Latz et al., 2015, Castro et al., 2016, Lareen et al., 2016). Exudate composition is specific to the plant species and variations can be observed along the root itself, the growth stage of the plant and plant stress levels in addition to soil type and nutrient availability; this can alter the microbial community composition and colonisation (Marschner et al., 2001, Garbeva et al., 2004, Compant et al., 2010, Lakshmanan et al., 2014, Zhang et al., 2014a, Kumar et al., 2016). For example, *Burkholderia cepacia* populations have been shown to decline as the age of the host plant increases (Garbeva et al., 2004). In addition, when chickpea (*Cicer arietinum*) is phosphorus limited, large amounts of organic acids are released to mobilise soil phosphorus and could result in selection of bacterial species tolerant to acidic conditions (Marschner et al., 2001). The plant developmental stage has also been shown to affect the root microbiome in a number of important crop plants: maize (*Pisum sativum*), wheat (*Triticum aestivum*) and sugar beet (*Beta vulgaris*) (Lakshmanan et al., 2014). Differences in the rhizosphere microbial community were noted between the seedling stage for *Arabidopsis* and three other developmental stages: vegetative, bolting and flowering (Chaparro et al., 2014). Differences in disease resistance associated with plant growth stages has been noted in bean, maize, soybean (*Glycine max*), cabbage (*Brassica oleracea*), cotton (*Gossypium* spp.) and *Arabidopsis* (Chaparro et al., 2014). The younger root hair zones release higher quantities of exudates when compared to mature root zones, correlating with higher nutrient uptake by the plant (DeAngelis et al., 2008, Libault et al.,

2010). The larger the plant diversity, the greater the level of heterogeneity seen in the soil due to the complex range of root exudates released into the rhizosphere (Kowalchuk et al., 2002).

Plant nutrient uptake reduces the levels of these nutrients remaining in the soil and can result in the production of enzymes by soil microorganisms to utilise organic matter (Bell et al., 2015, Castro et al., 2016). Weeds in agricultural systems have also been shown to influence the rhizosphere community as they provide different rhizosphere environments and so they are able to host bacteria and fungi distinct from the crop. Root exudates alter rates of both carbon and nitrogen cycling and this can in turn impact on the microorganisms present (Marschner et al., 2001, Bird et al., 2011). Nitrogen is often limiting in soil and so plant and soil microorganisms compete for available nitrogen however amino acids contained in plant exudates are often used as both a nitrogen and carbon source for soil microbes (Nordin et al., 2004, Zhang et al., 2014a). This competition for nitrogen is particularly evident in low productivity grasslands where both inorganic and organic nitrogen sources were quickly utilised by the microbial community (Bardgett et al., 2003). It has been argued that soil microbes are able to outcompete plants for nitrogen as they have a high growth rate, high substrate affinities and a high surface area to volume ratio and as a result, plants often use nitrogen after it has been recycled and mineralised by microorganisms (Nordin et al., 2004, Harrison et al., 2007, Kong et al., 2011). Rhizosphere microorganisms in return also play a role in plant nutrient uptake and aid abiotic stress tolerance and plant fitness (Porter and Rice, 2013, Kumar et al., 2016). Due to the varied ways the root microbiome can alter plant growth and health, it is clear that the root microbiome has a large role to play in global food security and sustainable intensification of agriculture.

1.4 Soil microbial ecology and agriculture

Global populations have risen steadily and the demand for food is also increasing leading to intensification of agriculture and a need to do this sustainably (Baker et al., 1997, Ali and McNear, 2014, Lakshmanan et al., 2014, Gupta et al., 2015). Approaches that minimise soil degradation and maximise soil productivity are now commonly included in environmental policy and it is estimated that soil degradation can cost up to £250m a year in England alone (Gregory et al., 2015). Soil degradation includes soil erosion, decline in soil organic matter, soil contamination and soil compaction and is seen as a key component to sustainable agriculture (Ferreira et al., 2000, Bastida et al., 2013, Gregory et al., 2015). Using soil microbiology to monitor key members of the microbial community has been suggested as a method of determining soil health (Yang et al., 2013). Land management, for example tilling, fertiliser application and crop rotation, affects the

composition and the community structure of microorganisms which are present within the soil by reducing organic matter content, altering pH, increasing pollution of groundwater and affecting rates of nutrient cycling (Garbeva et al., 2004, van Diepeningen et al., 2006, Kong et al., 2011, Zhalnina et al., 2013, Castro et al., 2016, Hollowell et al., 2016b, Lareen et al., 2016). Agricultural practices have major implications for soil microbial communities. The removal of plant cover by harvesting for example has been shown to influence both soil bacterial and fungal communities. Bacteria and both non-mycorrhizal and mycorrhizal fungi abundances in soil were reduced by 74%, 89% and 84% respectively when plants were removed; when replanted, abundances remained depressed at 64%, 74% and 71% respectively when compared to undisturbed grassland plots (Mikola et al., 2014).

Plant growth is usually limited by nitrogen, relying on inorganic forms of nitrogen for growth (Harrison et al., 2007, DeAngelis et al., 2008, Libault et al., 2010). The application of nitrogen fertiliser can decrease soil biomass and has been shown to be more influential on fungal communities than bacterial communities (Ekblad and Nordgren, 2002, Zhao et al., 2014). The implementation of organic farming methods has been shown to increase the diversity of bacteria, fungi, nematodes and other soil invertebrates compared to traditionally managed land involving cultivation of a monoculture crop (van Diepeningen et al., 2006). Incorporating straw into the soil alters the microbial communities present and can be an economical and environmental way of increasing soil nutrients (Peng et al., 2016). Arable fields, both monoculture and rotations, have been shown to have a lower microbial diversity index and biomass than permanent grasslands suggesting that either agricultural practices or having a monoculture for one to a number of years can impact the soil microbiology (van Elsas et al., 2002).

Land management has been targeted as a way to increase the productivity and sustainability of agriculture for example, no-till and use of grasslands (Attard et al., 2011, Kong et al., 2011). Conventional tillage can lead to losses of organic carbon and erosion whereas no-till allows for direct sowing through the stubble protecting the soil from erosion, improving soil structure and increasing microbial populations (Attard et al., 2011, Yang et al., 2013, Sun et al., 2014). Tilling disrupts the soil structure and this has important implications for water retention, plant growth, fertiliser retention and can lead to increased pollution (Bronick and Lal, 2005, Kong et al., 2011). Nitrogen levels have been shown to not be significantly different when comparing differing tillage regimes (conventional, deep plough and minimum till) but were significantly different when compared to no tillage (Sun et al., 2014). Reducing tilling has also shown to increase retention of nitrogen, phosphorus, organic carbon (Yang et al., 2013). Microbial biomass was greatest in conventional till and deep ploughing however the community composition was most diverse with

no tillage (Sun et al., 2014). Tilling could act to homogenise the soil, disperse the crop residues and reduce the number of niches present and this could provide an explanation for both the reduced microbial diversity and the increased biomass.

Organisms involved in beneficial interactions are often exploited to be used as commercial inoculants as biofertilisers (Lugtenberg et al., 2002, Compant et al., 2010, Brown et al., 2012, Marcondes de Souza et al., 2012). Soil additives derived from microbes are the fastest growing sector of the agrichemical industry (Lugtenberg et al., 2002, Ali and McNear, 2014). Biostimulants are another option for increasing productivity of agriculture to meet the growing food demand. These are fertilisers which improve the rhizosphere by increasing abundance of beneficial microorganisms through direct application or stimulate nutrient mobilisation and disease suppression through application of microbial secondary metabolites to increase plant health and efficiency (Ali and McNear, 2014). Microbes produce a range of secondary metabolites in order to compete with other microorganisms and communication between bacterial cells requires signalling molecules which upon perception induce gene expression to alter activity (Lareen et al., 2016). For example, the interactions between rhizobia and arbuscular mycorrhizal fungi resulted in higher plant growth of *Amorpha canescens* than when either of these microorganisms were inoculated alone (Larimer et al., 2014). Oil seed rape (*Brassica napus*) has also been shown to have increased yields when inoculated with both *Azospirillum* spp. and *Azotobacter* spp. (Panke-Buisse et al., 2015). Compounds commonly found in root exudates have been shown to influence the microbial communities when applied directly to the soil without a plant being present and so the application of these compounds could result in an increase in beneficial interactions occurring in the soil (Lareen et al., 2016).

Plant growth promoting rhizobacteria (PGPR), also known as plant growth promoting bacteria (PGPB) can promote plant growth either directly or indirectly through disease suppression, nitrogen fixation, phosphorus solubilisation and the production of phytohormones (Morgan et al., 2005, Compant et al., 2010, Gottel et al., 2011, Zhang et al., 2014a). Some PGPB produce antibiotics, lytic enzymes and siderophores which can aid with colonisation and can reduce the abundance of phytopathogens (Compant et al., 2010). Bioinoculants have been suggested as a way of sustainably intensifying agriculture and controlling disease but introducing high numbers of particular organisms into the soil will disrupt natural microbial communities (Lakshmanan et al., 2014, Gupta et al., 2015, Lareen et al., 2016). Bioinoculants can also be affected by the presence of plants as they have the ability to shift the community and the bioinoculants could result in altered activity in the soil as a response to root exudates (Latz et al., 2015). Introduced, beneficial strains of rhizobia have been shown to be outcompeted by native populations as persistence

depends on a range of factors both abiotic and biotic (Minamisawa et al., 1992, Howieson et al., 2013, Narozna et al., 2015, Hollowell et al., 2016b). Successful colonisation of the root is fundamental to the success of the bioinoculant and chemotaxis towards the root exudates is vital for this (Zhang et al., 2014a).

1.5 The Fabaceae plant family

Fabaceae is the largest plant family and legumes comprise approximately 27% of global crops and form a symbiotic relationship with rhizobia through root nodule formation for symbiotic nitrogen fixation in order to overcome nitrogen limitation (Vauclare et al., 2013, Hollowell et al., 2016a, Remigi et al., 2016). Soybeans were first domesticated 4000 years ago and are an important source of protein across the world. They were domesticated from a limited gene pool and so have limited genetic variability (Tang et al., 2012, de Carvalho et al., 2013). Crop rotations incorporating legumes were first described by the Romans as a way to enhance productivity of non-legume crops (Hirsch et al., 2001). It has been estimated that rhizobia fix approximately 150 Tg of nitrogen annually which is the largest nitrogen input in all terrestrial ecosystems (Gaby and Buckley, 2012, Hollowell et al., 2016a). It is estimated that bacterial nitrogen fixation is worth $7 billion per year in Brazil (de Carvalho et al., 2013). Symbiotic nitrogen fixation is of economic relevance due to the environmental and financial costs associated with fertiliser inputs (Ferguson and Mathesius, 2014, Miri et al., 2016, Okazaki et al., 2016). *Bradyrhizobium* has been shown to persist in the soil for 20 years and still be capable of inducing nodulation for symbiotic nitrogen fixation even without the host plant being present for the majority of this time (Narozna et al., 2015).

1.6 Rhizobiales and the Bradyrhizobiaceae

Proteobacteria is a large phylum containing extensively studied genera, for example *Escherichia* and *Pseudomonas,* and this phylum contains a large class of bacteria, the alpha-Proteobacteria (Figure 1.1) (Rappe and Giovannoni, 2003, Gourion et al., 2009, Rivas et al., 2009, Okubo et al., 2012, Remigi et al., 2016). Within this class is the order Rhizobiales and a key characteristic of this order is the ability to form symbioses with legumes to allow them to thrive in areas of nitrogen limitation (van Rhijn and Vanderleyden, 1995, Giraud et al., 2007, Rivas et al., 2009, Libault et al., 2010, Okubo et al., 2012, Busset et al., 2016) . The ability to form symbioses with legumes is thought to have evolved through horizontal gene transfer as the genes involved in this process are usually located on symbiotic islands in bradyrhizobia (Kaneko et al., 2002, Rivas et al., 2009, Sachs et al., 2010, Stepkowski et al., 2011, Okubo et al., 2012, Tang et al., 2012) and on symbiotic plasmids in other rhizobial genera (van Rhijn and Vanderleyden, 1995). Phylogenies reconstructed using symbiotic genes differed to those constructed using housekeeping genes suggesting horizontal gene transfer of the genes for symbiosis (Stepkowski et al., 2011).

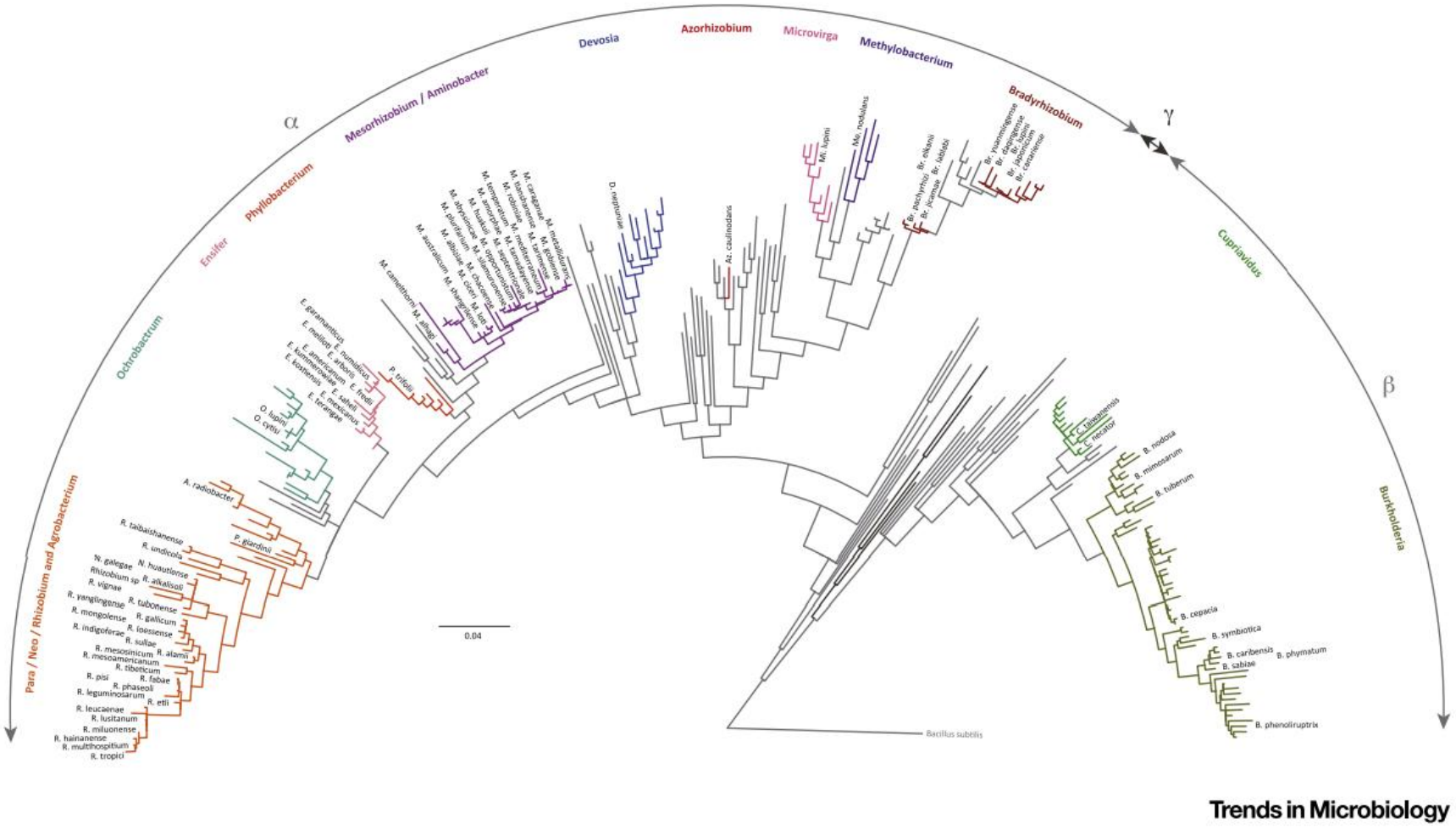


Figure 1.1: Phylogenetic tree showing selected Proteobacteria showing symbiotic rhizobia in a number of genera.
Rhizobiales are distinct from other common soil bacteria *Burkholderia* and *Cupravidus*. Rhizobiales contain a number of genera highlighted in different colours. Image from Remigi et al. (2016).

Bradyrhizobium is slow-growing and is biologically important within soils, with different representatives found to perform a wide range of biochemical functions including photosynthesis, nitrogen fixation through symbioses, denitrification and aromatic compound degradation (Fuhrmann, 1990, Streeter, 2003, Okubo et al., 2012, Kimbrel et al., 2013). The *Bradyrhizobium* genus was created due to the distinction between fast and slow growing rhizobia (Jarvis et al., 1982, Jordan, 1982, Lindström, 1989, Wang et al., 1998). Bradyrhizobia were first isolated from root nodules but as a result of 16S sequencing what we know about the genus has grown and includes both symbiotic and non-symbiotic variants (Jordan, 1982, Okubo et al., 2012, VanInsberghe et al., 2015, Jones et al., 2016). The role this genus plays within the nitrogen cycle gives agricultural significance in understanding the ecology of this group (Noel and Brill, 1980). Both agricultural and natural rhizobial populations are often dominated by a few genotypes often selected by the legume host (Hollowell et al., 2016b).

Bradyrhizobium is a successful genus of rhizobia found globally and in a range of soil and aquatic environments. Members of the genus have been found to nodulate crops and wild legumes and have both free-living and symbiotic lifestyles (Hollowell et al., 2016b). Type strains are required when establishing whether new species names are to be assigned to a group of novel isolates. As in order for taxa to be assigned to a type strain, additional information on the biology must be obtained in order to distinguish the strain from other type strains. For example, fingerprinting profiles, level of sequence similarity for a number of genes and DNA-DNA hybridisation (Whitman et al., 2015). According to the international bacteriological code of nomenclature, isolates are not considered a different species if they are identical to another type strain and the group of isolates with the same species name must be changed if the type strain is no longer within that group (Buchanan et al., 1948). *Bradyrhizobium* contains a number of species; 26 have been described so far including *B. japonicum* USDA 6 (T)*, B. elkanii* USDA 76 (T) and *B. canariense* BTA-1 (T) and is a highly diverse genus (Rivas et al., 2009, Saeki et al., 2010, Stepkowski et al., 2011, Batista et al., 2013). Phylogeny studies of strains isolated from the nodules of lupins (*Lupinus* spp.) and serradella (*Ornithopus* spp.) showed a high level of diversity when creating phylogenetic trees from housekeeping genes with *B. japonicum* and *B. canariense* being the dominant species present (Stepkowski et al., 2011).

Antibiotic resistance has also arisen in this genus to aid survival within soil and to allow for *Bradyrhizobium* to have competitive advantage over other members of the soil community. Resistance to a number of different antibiotics has been noted and this is thought to be through genes often contained on plasmids (Cole and Elkan, 1973, Kuykendall et al., 1988). *B. elkanii* in particular is resistant to a number of antibiotics including kanamycin, rifampicin, spectinomycin,

chloramphenicol, streptomycin and carbenicillin with *B. diazoefficiens* USDA 110 only being mildly resistant to rifampicin and spectinomycin (Soares and Pereira Passaglia, 2010). Heavy metal resistance has been reported in *B. japonicum* which aids the wide distribution of the genus globally (Tong and Sadowsky, 1994). Soil temperature could influence the ecological niche the nodulating strains of *Bradyrhizobium* fulfil as they have been shown to differ in their ecological niche from temperate to subtropical regions of Asia along an abiotic gradient (Saeki et al., 2010).

The degradation of organic compounds is an important function which both symbiotic and non-symbiotic strains are able to perform as it aids survival in agricultural environments where the application of organic compounds alters nutrient levels within the soil (Niewiadomska et al., 2012). 2,4-dichlorophenoxyacetic acid (2,4-D) is a herbicide that has become a model compound in the study of how microorganisms have evolved the capacity to able to degrade xenobiotic organic compounds within soil using *tfd* genes (Kamagata et al., 1997, Itoh et al., 2002, Niewiadomska et al., 2012). *Bradyrhizobium* spp. have also been shown to be able to utilise 2,4-D as a sole carbon and energy source (Itoh et al., 2000). The *tfd* genes encode the enzyme 2,4-D alpha-ketoglutarate dioxygenase which catalyses the first step of 2,4-D mineralisation and is usually present in both beta- and gamma-Proteobacteria. Also, *tfdA*-like genes are present in the alpha-Proteobacteria suggesting that, despite coding for a similar protein, these genes are evolutionally distinct from those found in the other classes within the phylum (Kamagata et al., 1997, Itoh et al., 2002, Kaneko et al., 2002). The bradyrhizobial strains RD5-C2, BTH, HWK12 and HW13 were found to harbour genes encoding for the same protein and were isolated from a pristine environment within Japan with no exposure to 2,4-D. This shows that the existence of these genes is independent to 2,4-D exposure suggesting involvement in soil processes in addition to herbicide degradation (Kamagata et al., 1997, Itoh et al., 2000, Itoh et al., 2002).

B. japonicum is an important member of the *Bradyrhizobium* genus as it forms a symbiotic relationship with agriculturally important species such as soybean (*G. max*), mungbean (*Vigna radiata*)*, Lotus* spp. and other legumes by forming nodules for nitrogen fixation on the roots in exchange for carbon (Fuhrmann, 1990, Kaneko et al., 2002, Giraud et al., 2007, Gourion et al., 2009, Libault et al., 2010, Sachs et al., 2010, Schirrmeister et al., 2011, Batista and Hungria, 2012, Khatoon et al., 2012, Batista et al., 2013). A range of *Bradyrhizobium* species can form nodules on soybean (*G. max*) but it is *B. japonicum, B. diazoefficiens* and *B. elkanii* which are most commonly used as commercial inoculants with *B. japonicum* SEMIA5080 and *B. elkanii* SEMIA587 being the most widely used in Brazil (Soares and Pereira Passaglia, 2010, de Campos et al., 2011, de Carvalho et al., 2013).

1.7 Nodulation

Plants able to form symbiotic associations with nitrogen fixing bacteria belong to the Rosid clade and legumes able to be nodulated by rhizobia belong to the Galegoid tribe (Cullimore et al., 2001, Fabre et al., 2015, Remigi et al., 2016). Evolution of nodulation ability began 100 million years ago and is thought to have a monophyletic origin (Fabre et al., 2015). The root nodule is thought to have evolved around 60 million years ago and is agronomically and ecologically important due to its involvement in nitrogen fixation (Cullimore et al., 2001, Ferguson and Mathesius, 2014, Miri et al., 2016, Okazaki et al., 2016). Rhizobia gain DNA forming part of the accessory genome through acquisition of symbiosis islands and plasmids which contain genes required for infecting the host. This enhances the fitness of the strain to benefit from the host environment (Hollowell et al., 2016a). The host range can be any number of legumes from an extensive range to a limited number of lines within one species (Cullimore et al., 2001). A number of non-legume crops associate with nitrogen-fixing bacteria that do not result in a nodule for example sugarcane (*Saccharum officinarum*), maize (*Zea mays*), soybean (*G. max*) and rice (*Oryza sativa*) (Hirsch et al., 2001).The plant genus *Parasponia* in the family Ulmaceae is the only non-legume known to form a nodule symbiosis with *Rhizobium* (van Rhijn and Vanderleyden, 1995, Op den Camp et al., 2012). The symbiotic relationship between *Bradyrhizobium* and its legume host is initiated by chemical exchange of flavonoids which are released by the plant (van Rhijn and Vanderleyden, 1995, Giraud et al., 2007, Khatoon et al., 2012, de Carvalho et al., 2013, Busset et al., 2016). Flavonoids induce expression of nodulation genes which encode enzymes required for Nod factor synthesis which feeds back to the plant to begin the nodule formation process (Cullimore et al., 2001, Okazaki et al., 2016) (Figure 1.2).

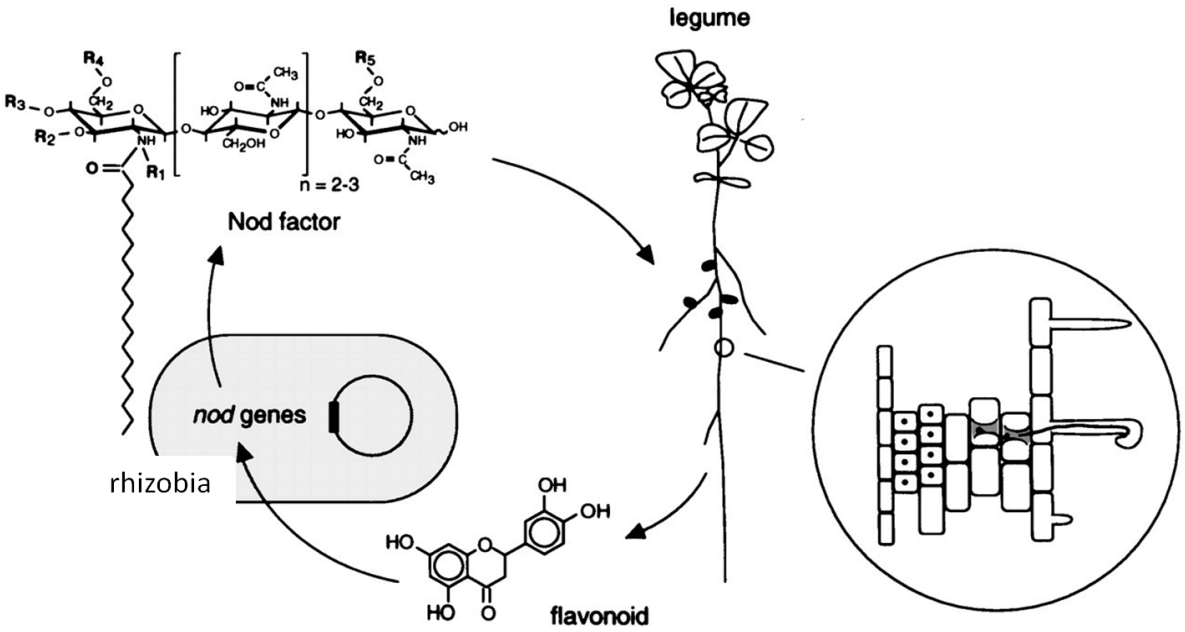


Figure 1.2: Summary of the nodulation process.
The diagram shows the release of flavonoids stimulating the expression of *nod* genes resulting in the production of a Nod factor required for the initiation of nodule organogenesis. Adapted from Schultze and Kondorosi (1998).

Flavonoids are released from all higher plants and are not unique to legumes and all contain two benzene rings and depending on the substitutions in the structure results in flavonols, flavones, flavanones and isoflavonoids (Hirsch et al., 2001, Morgan et al., 2005). Isoflavonoids are unique to legumes and include genistein and daidzein which are produced by soybeans and together with other non-isoflavones for example phenolics and are primary inducers of genes involved in nodulation in *B. japonicum* (Hirsch et al., 2001, Ferguson et al., 2010, Batista and Hungria, 2012, Fabre et al., 2015, Busset et al., 2016). The release of specific flavonoids allow for rhizobia to distinguish between legumes and find their host (Hirsch et al., 2001). The protein expression in *B. japonicum* in response to genistein is not limited to nodulation, a number of metabolic pathways can be triggered and one of these is the PhyR-sigma (EcfG) signalling cascade associated with a stress response notably carbon starvation (Gourion et al., 2009, Batista and Hungria, 2012). Flavonoids are perceived by rhizobia as aglycones and induce *nod* gene expression by producing a conformational change in the NodD protein enabling it to bind to the *nod* box located in the promoter region (Hirsch et al., 2001).

Type III secretion systems (T3SSs) are specific to Gram-negative bacteria and deliver proteins to the host cell (Baker et al., 1997, de Campos et al., 2011, Schirrmeister et al., 2011, Marcondes de Souza et al., 2012). The secretion system is made up of around 20 proteins many homologous to flagella assembly and is activated by *nodD1* in response to flavonoid detection (de Campos et al.,

2011). The secreted proteins, nodulation outer proteins (Nops), are host specific and can promote nodulation in some species and in some limit nodulation if recognised as a virulence factor (de Campos et al., 2011, Kimbrel et al., 2013). de Campos et al. (2011) showed that if the T3SS was mutated and therefore did not function correctly, the nodule formation process was delayed. As well as inducing *nod* gene expression, the release of flavonoids can also induce chemotaxis (Hirsch et al., 2001). However, this is not essential for nodulation only colonisation as non-motile mutants were still able to form root nodules (van Rhijn and Vanderleyden, 1995).

Once the flavonoids are detected by the bacteria, expression of the *nodABC* genes are induced and results in the synthesis of a Nod factor, which aids the nodulation process (Debelle et al., 2001, Gage, 2004, Giraud et al., 2007, Ferguson et al., 2010, Libault et al., 2010, Khatoon et al., 2012, de Carvalho et al., 2013, Busset et al., 2016). Nod factors were described in the 1990s as lipo-chitooligosaccharides - a lipid attached to the non-reducing end of a chain of N-acetyl-D-glucosamine molecules (Cullimore et al., 2001, Fabre et al., 2015). They are produced from a range of bacteria and all have the same core structure; four or five N-acetylglucosamines with a fatty acid molecule containing 16 or 18 carbon atoms attached to the non-reducing end (Cullimore et al., 2001, Oldroyd, 2013). Variations in chemical substitutions on the sugars within the Nod factor, variations on the acyl chain and the quantity produced is specific to the rhizobia producing them and influences host range (Cullimore et al., 2001, Oldroyd, 2013, Ormeno-Orrillo et al., 2013). Modifications can include acetate, sulphate or sugars (Hirsch et al., 2001, Ormeno-Orrillo et al., 2013). No correlations have been found between *nodA* sequence and bacterial taxonomy however correlations have been found between the sequence and the structural characteristics of the Nod factor (Debelle et al., 2001). Nod factor fucosylation by the *nodZ* gene is closely linked to the ability of *B. japonicum* to nodulate *Macroptilium atropurpureum* and *Vigna umbellata* but not linked to soybean nodulation (Ormeno-Orrillo et al., 2013).

Legumes produce a range of enzymes within the root which are able to degrade and cleave the lipo-chitooligosaccharides depending on the substitutions present on the sugars (Cullimore et al., 2001). Nod factors are detected by a LysM receptor-like kinase in the plant and form part of the specificity of the communication between the rhizobia and plant (Fabre et al., 2015, Miri et al., 2016, Okazaki et al., 2016). Nod factors trigger responses in the epidermis, cortex and vasculature of the legume, resulting in ion fluxes, increases in intracellular pH, root hair tip calcium ion accumulation, root hair deformation, formation of pre-infection threads, cell division resulting in the nodule primordium and gene expression (Cullimore et al., 2001, Hirsch et al., 2001). Nodule formation is the result of a signalling cascade beginning with the production of proteins required for calcium spiking (Fabre et al., 2015).

The plant then responds to these calcium influxes expressing a range of developmental pathways allowing for the rhizobia to enter the root through infection threads and the formation of the nodule primordium (Miri et al., 2016). Lipo-chitooligosaccharides are able to enter into membranes but are unable to enter into the plant cell without the plant recognising the molecule and having a specific transport mechanism (Cullimore et al., 2001). Microbial signals begin the process of mitotic divisions of plant cells which will become infected by the rhizobia (Geurts et al., 2016). Lipopolysaccharide comprises three parts and is the main component of the outer membrane of Gram-negative bacteria including rhizobia and is thought to be involved in the symbiotic association between legumes and rhizobia (Noh et al., 2015, Busset et al., 2016). The O-antigen part of the lipopolysaccharide is variable and this variability is thought to reduce the plant defence response during infection in symbiosis (Noh et al., 2015, Busset et al., 2016).

The root hair is the primary target for the rhizobia as it is a structure which is able to be modified easily by rearranging microtubules but they can invade through intercellular spaces in the epidermis (van Rhijn and Vanderleyden, 1995, Ferguson et al., 2010, de Carvalho et al., 2013, Fabre et al., 2015). These modifications allow for the uptake of the rhizobia by the root hair becoming less rigid (Geurts et al., 2016). Nod factor detection by the plant root hair cell results in root hair deformation through curling, calcium concentration changes and root cortex cell division (Halverson and Stacey, 1986, Gage, 2004, Libault et al., 2010). Once root hair curling has been initiated, the infection thread formation is stimulated which aids the most common infection route of the root used by rhizobia: entering root tissue through the root hair (Halverson and Stacey, 1986, Gage, 2004, Ferguson et al., 2010). Infection thread formation is a plant response to invasion of the plant cell by the rhizobia (van Rhijn and Vanderleyden, 1995). Folding of the cell wall in the centre of the curl or degradation of the cell wall leading to folding of the cell membrane followed by tip growth results in a tubular structure, the infection thread, growing into the root tissue (Gage, 2004, Ferguson et al., 2010). The infection thread is filled with bacteria which are dividing and the thread branches to provide a larger surface area for the bacteria to enter the nodule cells (Gage, 2004). Plasma membrane invaginations and subsequent uptake of the rhizobia results in the bacteria not being in contact with the cytoplasm but is encapsulated by the plant cell (Geurts et al., 2016). Both root hair curling and formation of infection threads require the bacteria to be present and in contact with the plant cell wall (Hirsch et al., 2001) (Figure 1.3).

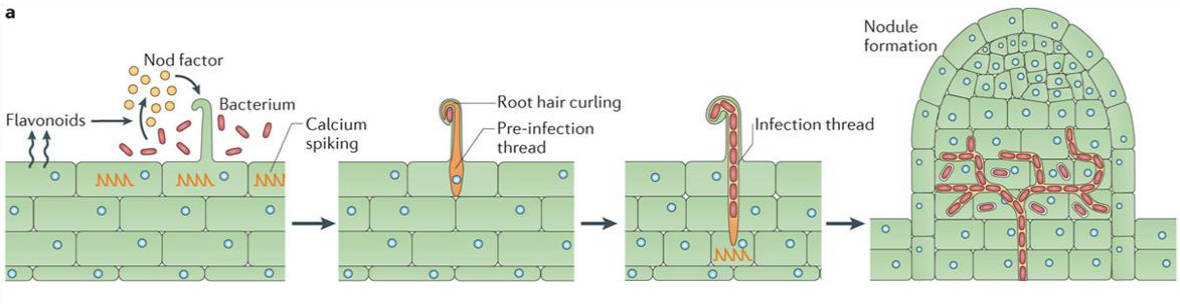


Figure 1.3: Summary of the response to Nod factor detection.
The process begins with root hair curling and the formation of the infection thread resulting in the nodule filled with differentiating bacteroids capable of nitrogen fixation. Adapted from Oldroyd (2013).

There is an alternative mode of rhizobial invasion, known as "crack invasion", which bypasses some of the molecular signalling described above. This mechanism may predate the evolution of infection threads and present in aquatic legumes such as *Aeschynomene* and legumes belonging to the Aeschynomeneae tribe including the agronomically-important peanut (*Arachis hypogea*): *Bradyrhizobium* sp. enter the root via cracks in the root epidermis to access the cortical cells (Sprent, 2007).

Bacteria reside in symbiosomes inside the nodule cells, separated from the plant cytoplasm by a peribacteroidal membrane. Here, the bacteria differentiate into bacteroids and produce proteins required for fixing atmospheric nitrogen into ammonium (van Rhijn and Vanderleyden, 1995, Gage, 2004, Vauclare et al., 2013, Delmotte et al., 2014, Miri et al., 2016). Oxygen is limited within the nodule and as a result bacterial metabolism is changed to cope with this (Vauclare et al., 2013, Delmotte et al., 2014). Once inside the nodule, osmotic stress is an important factor and the metabolism of trehalose acts as an osmoprotectant during colonisation and infection (Vauclare et al., 2013). The plant benefits from the association by more nitrogen being available and the rhizobia benefit from a source of fixed carbon to provide the energy needed for nitrogenase activity, a source of nitrogen though amino acid exudates and protection from the pathogens and fluctuating environmental conditions (Morgan et al., 2005, Oldroyd, 2013, Delmotte et al., 2014). Nodulation increases rhizobial abundance in soil: although bacteroids of some rhizobia are not viable after nodule senescence, the infection threads provide a protected niche from which undifferentiated cells are released (Hirsch, 1996).

Aeschynomene spp. are semi-aquatic plants and form a monophyletic clade containing two species; *Aeschynomene indica* and *A. evenia* which are important forage crops due to high protein

content (Arrighi et al., 2013, Delmotte et al., 2014, Okazaki et al., 2016). Root and stem nodules are formed by a range of *Bradyrhizobium* strains however these strains have been found to lack *nodABC* which are required to form Nod factors (Giraud et al., 2007, Fabre et al., 2015, Busset et al., 2016, Okazaki et al., 2016). Strains able to nodulate *A. evenia* have recently been suggested as a model legume for examining legume-rhizobia interactions as *nod*-independent nodulation is thought to be the ancestral state of the legume-rhizobia symbiosis (Arrighi et al., 2013, Fabre et al., 2015). Nod factor-independent symbiosis between *B. elkanii* and *G. max* has been shown to be dependent on the T3SS (Okazaki et al., 2016). *Aeschynomene*-nodulating *Bradyrhizobium* are also photosynthetic and the majority lack T3SS (Giraud et al., 2007, Okazaki et al., 2016). The photosynthetic ability of these strains are particularly important during stem nodulation as, rather than relying on the plant host, they can synthesise their own energy required for nitrogen fixation (Arrighi et al., 2013, Giraud et al., 2013). Photosynthesis genes have also been noted in *Bradyrhizobium* strain, ORS278, which is a symbiont of soybean as this strain contains both *nod* genes and photosynthesis genes. Other photosynthetic strains which lack nod genes are not able to form root nodules on soybeans (Giraud et al., 2013). Photosynthetic *Bradyrhizobium* are also able to fix nitrogen when free-living (Giraud et al., 2013). Both a *nod*-dependent and *nod*-independent method for root nodule formation is possible in photosynthetic *Bradyrhizobium* strains which also contain *nodABC* genes (Giraud et al., 2013).

Some isolates of *Bradyrhizobium* have been found to be non-symbiotic and have either never acquired, or lost the ability to form nodules. A reason for this could be because the genes required to form nodules are located on symbiosis islands and if lost then the strain will no longer be able to form nodules on the roots (Okubo et al., 2012, Ormeno-Orrillo et al., 2013). Alternatively, these genes may have never been acquired by the strains. The absence of nodulation ability has been noted in the strain *Bradyrhizobium* sp. S23321 isolated from paddy soil in Japan, in other agricultural soils and in at least one unmanaged soil system and are thought to be closely related to the nodulating strains (Sachs et al., 2010, Okubo et al., 2012, Jones et al., 2016). Some environments may drive selection against symbiotic strains; for example, the lack of suitable hosts may encourage alternative survival strategies. Also, if competitors or pathogens disturb beneficial interactions or if nitrogen is no longer limited, the benefits to the plant are inadequate to compensate the cost in forming the nodules which benefit only the bacterium in nitrogen rich environments.

1.8 The microbial nitrogen cycle

The microbial nitrogen cycle (Figure 1.4) is an important process within the soil and *Bradyrhizobium* has been previously shown to be involved in two major steps; nitrogen fixation and denitrification. Nitrogen fixation is the conversion of dinitrogen gas to ammonium and is a one step process. Denitrification is the conversion of nitrate into dinitrogen gas and this is a four step process with intermediates of nitrite, nitric oxide and nitrous oxide. Land use can impact on the composition of the microbial community and can alter abundances of key groups involved in nitrogen cycling within the soil. *Ca. Nitrososphaera*, an ammonia oxidising archaeon has been previously found to be positively correlated with agricultural land whereas *Bradyrhizobium* was found to be negatively correlated with agriculture (Zhalnina et al., 2013).

1.8.1 Nitrogen fixation

Nitrogen fixation is an ancient process and is a vital part of the nitrogen cycle (Figure 1.4) (Buckley et al., 2007, Terakado-Tonooka et al., 2013). Nitrogenase subunits are encoded by the *nifH, nifD* and *nifK* genes and catalyse the reaction of nitrogen into ammonium (Buckley et al., 2007, Gaby and Buckley, 2012, Terakado-Tonooka et al., 2013). Malate and mono- and dicarboxylic acids for example lactate and succinate are readily utilised by *Bradyrhizobium* as energy for nitrogenase activity (Terakado-Tonooka et al., 2013). The conversion of dinitrogen gas to ammonium is energetically expensive with each dinitrogen molecule converted requiring 16 ATP molecules (Meakin et al., 2006). The nitrogen fixation reaction occurs in parallel to the hydrolysis of ATP to provide energy for cleaving the N,N bond in atmospheric nitrogen (Lee et al., 2014).

1.8.2 Denitrification

Denitrification is a four step anaerobic process which converts NO_3^- to NO_2^-, NO_2^- to NO, NO to N_2O and ultimately N_2O to N_2 (Figure 1.4) (Kristjansson and Hollocher, 1980, Delgado et al., 2003, Bedmar et al., 2005, Falk et al., 2010). This has major implications for agriculture and global climate change as it results in the loss of nitrogen and the release of a potent greenhouse gas in nitrous oxide (Falk et al., 2010, Pastorelli et al., 2011, Zhu et al., 2012). Nitrogen removal through heterotrophic denitrification is an important step in the global nitrogen cycle carried out by many groups from all known bacterial phyla including *Bradyrhizobium* (Kaneko et al., 2002, Zhu et al., 2012). Denitrifiers are common among Proteobacteria and in particular the Alpha and Beta subclass (Pastorelli et al., 2011). Nitrate reductase is encoded by *narG*/*napA*, nitrite reductase by

nirK/*nirS*, nitric oxide reductase by *norB*/*norC* and nitrous oxide reductase by *nosZ* (Delgado et al., 2003, Bedmar et al., 2005).

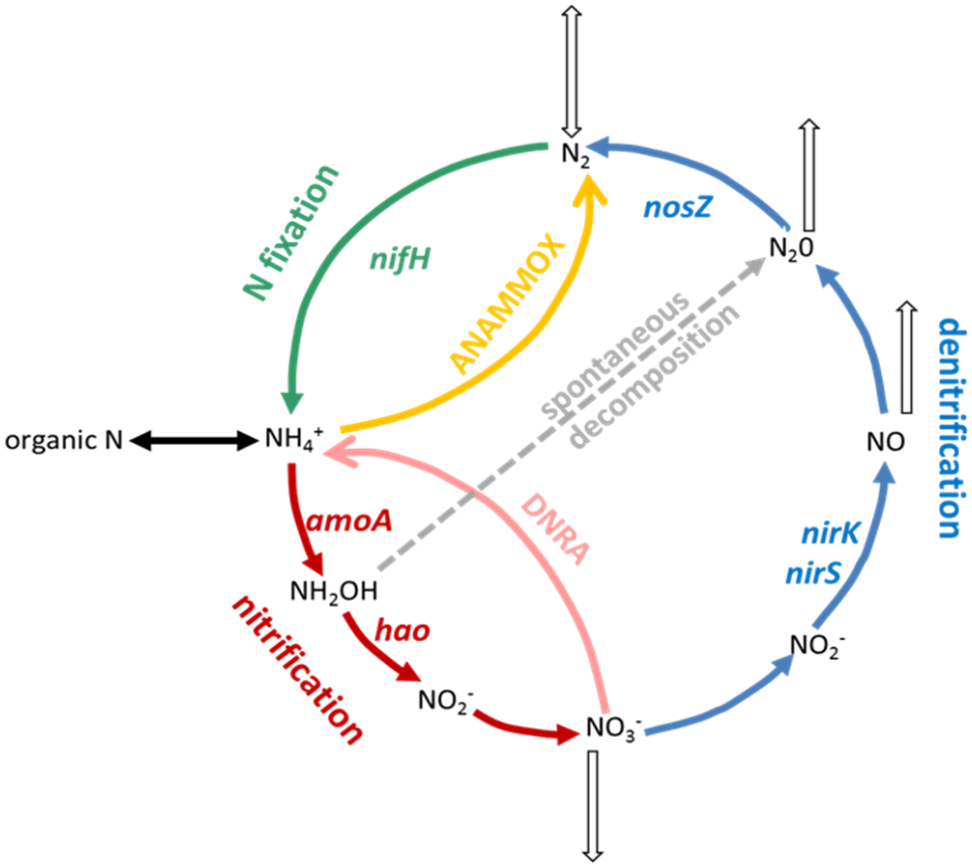


Figure 1.4: Summary of the microbial nitrogen cycle
The diagram shows the steps in the microbial nitrogen cycle: the conversion of nitrogen gas to ammonium through nitrogen fixation controlled by the *nifH* gene. The steps involved in nitrification controlled by the *amoA* and *hao* genes. It also shows the conversion of nitrate to nitrite, nitrite to nitric oxide (*nirK/S*), nitric oxide to nitrous oxide and nitrous oxide to nitrogen gas (*nosZ*) through denitrification.

1.9 The Highfield experiment and *Bradyrhizobium*

The Highfield experiment is a long-term field experiment at Rothamsted Research that has three treatment regimes; grassland (high carbon, "obese"), arable rotation and bare fallow (low carbon, "lean"), allowing for assessment of how land management affects the microbial communities. Highfield had been maintained as a grassland meadow for several hundred years until arable plots were established in 1949, planted with winter wheat (*Triticum aestivum*) whereas the grassland plots continued to be managed as traditional grassland. The bare fallow plots are maintained by tilling and have had no plant inputs apart from occasional weeds since 1959 (Hirsch et al., 2009).

In addition to the Highfield plots, a woodland area from the Broadbalk experiment was used for comparison of land use. This area of land has not been managed since 1868 when it was taken out of arable use to develop into natural woodland due to unmanaged succession. The Highfield and Broadbalk experiments are described in more detail in Chapter 3. The terms lean and obese draw comparisons with gut microbiology studies where differences in microbial communities have been noted between lean and obese mice. Bacteroidetes and Firmicutes abundances are strongly linked to obesity where obese mice guts contained fewer Bacteroidetes compared with lean mice guts (Ley et al., 2005). Obese soils are those with high levels of carbon, which is usually limiting in soil and lean soils are those with low levels of carbon from photosynthate impacting the microbial communities present.

Many active biological processes are dependent on carbon inputs from recent photosynthesis, exemplified by increased carbon dioxide released by soil respiration from the grassland plots, compared to the arable and bare fallow treatments at Highfield (Hogberg and Read, 2006, Hirsch et al., 2009). This long-term experiment has allowed the microorganisms to adapt to the change in the abiotic and biotic conditions which arise due to the lack of plant inputs allowing the study of changes displayed by microbial communities associated to the different land uses. Grasslands have high plant diversity and as a result higher root biomass resulting in a broader spectrum of root exudates released and a greater number of potential niches for soil bacteria (Balasooriya et al., 2013, Latz et al., 2015). Various groups of microorganisms including *Bradyrhizobium* have been shown to be associated with different land management regimes. It was found that the archaeon *Ca. Nitrososphaera* was positively correlated with agriculture and also soil pH and NH_3 concentrations and that *Bradyrhizobium* was negatively correlated with agriculture, soil pH and NH_3 concentrations (Zhalnina et al., 2013). Ferreira et al. (2000) showed that population of *Bradyrhizobium* increased when no-till or crop rotations involving soybeans were applied. It was suggested that using legumes as break crops or converting to no-till treatments could be used as a way of encouraging beneficial bacteria and in particular could increase the abundance of symbiotically competent *Bradyrhizobium* and promote their maintenance within the soil to increase yields of the host plant (Ferreira et al., 2000).

Metagenomic studies, where DNA Is extracted directly from soil organisms *in situ* and sequenced to identify groups and genes, have shown *Bradyrhizobium*-like species to be one of the most abundant groups on soil from Highfield, comprising approximately 0.1% of the bare-fallow and approximately 0.2% of the grassland sequences indicating it is an important group within this system. DNA was extracted from soil in October 2011 and sequenced before being analysed for species abundances (unpublished data from the Hirsch group at Rothamsted Research, Ian Clark

can be contacted for more information). *Bradyrhizobium* had also previously been shown to be highly represented in grassland soils from the Park Grass experiment, another long-term field experiment at Rothamsted Research (Delmont et al., 2012). RNA was also extracted from the Highfield plots in October 2011, in order to carry out metatranscriptome sequencing to compliment the metagenome and to measure transcriptional activity of the communities. *Bradyrhizobium* was shown in the metatranscriptome from the Highfield plots to be an active member of the community (unpublished data from the Hirsch group at Rothamsted Research, Ian Clark can be contacted for more information). It has also been suggested that *Bradyrhizobium* could be used as an indicator for perturbation arising from agriculture (Zhalnina et al., 2013). Understanding the mechanisms of *Bradyrhizobium* adaptation to independent living within the Highfield experiment could release information on the extent of the genetic potential of this globally important genus.

1.10 Project objectives

Bradyrhizobium is a well-known group and is ubiquitous, being found in a wide variety of soils across the world with different managements- nodules, arable soil, paddy field soil, forest soil (Delmont et al., 2012, Okubo et al., 2012, Zhalnina et al., 2013, VanInsberghe et al., 2015). The metatranscriptome from the Highfield experiment also shows that *Bradyrhizobium* is a very active member of the community. Studying *Bradyrhizobium* will help give an overview of the biology of this abundant and active group and give clues on how members of this group differ depending on the soil they are isolated from. The lack of known plant hosts for *Bradyrhizobium* in the Highfield soils indicated that the majority were most likely to be free-living, without any symbiotic phase.

1.10.1 Aims

1. To create a culture collection

The bulk of research on *Bradyrhizobium* has been focussed on symbiotic strains isolated from root nodules. The Highfield experiment identified that there was a high abundance of *Bradyrhizobium* present in the bare fallow soil and so an isolation method for free-living strains needed to be developed to form a culture collection for future experiments. Collecting isolates from the three permanent treatments from Highfield (bare fallow, arable and grassland), the woodland section of the Broadbalk experiment and also UK symbiotic strains from gorse (*Ulex europaeus*) and broom (*Cytisus scoparius*) along with reference strains allows for a comparison of isolates

subjected to a range of treatments to assess diversity and function. Gorse and broom were chosen as they were legumes known to be nodulated by *Bradyrhizobium* and were abundant close to the Rothamsted site providing a geographically relevant comparison.

2. To assess the diversity of the collection using housekeeping genes

The core genome includes functions which are essential to all bacteria and therefore are highly conserved. Assessing housekeeping genes located across the chromosome allows for an assessment of the level of diversity present in the culture collection and other strains in the database and will give an indication of how soil treatment affects this group of bacteria. It provides a deeper assessment of diversity than 16S rRNA gene sequences alone and helps determine whether soil treatment selects for particular strains of *Bradyrhizobium*.

3. To assess genome wide diversity with *de novo* genome sequencing and assembly

Exploring the whole genomes of two isolates from varying soil treatments allows for a broader assessment of the level of diversity and will also give indications on differences occurring in the accessory genome. The genome comparisons will be used to direct future experiments to assess whether differences noted in the genomes of these isolates can be seen across the culture collection as a whole and if they correlate with other studies from the literature.

4. To assess the carbon metabolism

The level of soil organic carbon differs dramatically between the plots on the Highfield experiment and so assessing the carbon metabolism ability of these isolates will determine whether soil treatment is a driver for the evolution of particular carbon metabolism pathways. It also allows an assessment of the variability seen in carbon metabolism and whether particular substrates are associated with any one isolate or any group of isolates.

5. To examine the nitrogen cycling capability

Nodulation for symbiotic nitrogen fixation is the process for which *Bradyrhizobium* is traditionally known, the assessment of this ability will determine whether these isolates are free-living or are able to form symbiotic associations with legumes. Denitrification is a process which has implications in agriculture and also impacts on global climate change. Assessing the denitrification ability will allow us to determine whether isolates from particular soil treatments group according to denitrification ability and whether the genes required for this pathway are expressed and functional.

2 General materials and methods

2.1 Culture medium

All isolates were cultured on modified arabinose gluconate (MAG) agar which is the recipe used at the USDA for culturing rhizobia and is published in van Berkum (1990) (P Elia 2012, pers. comm., 16 November). The MAG broth consisted of (per litre): HEPES (Sigma Aldrich, USA) 1.3 g, MES (Sigma Aldrich, USA) 1.1 g, yeast extract (Merck KGaA) 1.0 g, L-arabinose (Sigma Aldrich, USA) 1.0g, D-gluconic acid sodium salt (Sigma Aldrich, USA) 1.0 g, potassium dihydrogen orthophosphate (Fisher Chemicals, USA) 0.22 g, sodium sulphate (Fisons, UK) 0.25 g. The following volumes of stock solutions were also added to the medium (per litre): 2 ml ammonium chloride (Fisons, UK; 1.6 g L^{-1}), 1 ml iron (II) chloride (Sigma Aldrich, USA; 0.067 g L^{-1}), 1 ml calcium chloride dihydrate (Sigma Aldrich, USA; 0.15 g L^{-1}), 1 ml magnesium sulphate heptahydrate (Fisons, UK; 1.8 g L^{-1}). All chemicals were of analytical grade. The pH was adjusted to 6.6 using 5 M potassium hydroxide. For solid media, 1.8% w/v (18 g L^{-1}) bacto-agar (Becton, Dickinson & Company, USA) was added to the MAG broth. The medium was sterilised by autoclaving at 121^{o}C, at 100 kPa for 15 minutes and allowed to cool before pouring. Warm MAG agar (<50^{o}C) (approximately 20 ml) was poured in to Petri dishes (Thermo Fisher Scientific, USA) in a sterile flow hood and stored at 4^{o}C once set.

2.2 Growth conditions

All soil dilutions and confirmed *Bradyrhizobium* isolates were cultured on MAG agar plates. The plates were incubated at 28^{o}C for up to 3 weeks. All isolates which were grown in broth culture, were grown in MAG broth, incubated at 28^{o}C and shaken at 100 rpm for up to 2 weeks, until all cultures were opaque.

2.3 DNA extraction

For screening isolates, microLYSIS-Plus (Microzone Ltd, UK) was used as this is a quick and efficient method for extracting DNA and large quantities of isolates were expected to be screened. For each isolate, half a colony was added to 10 µl of microLYSIS-Plus solution. The microLYSIS conditions are a series of heating and cooling steps and were as follows: 65^{o}C for 15 minutes, 96^{o}C for 2 minutes, 65^{o}C for 4 minutes, 96^{o}C for 1 minute, 65^{o}C for 1 minute and finally 96^{o}C for 30 seconds and then held at 4^{o}C. After the microLYSIS cycle, the solution was spun at

maximum speed (14,000 x g for 30 seconds) so the supernatant containing the DNA could be obtained.

For sequencing, DNA was extracted using the GenElute Bacterial Genomic DNA kit (Sigma Aldrich Co.) using the manufacturer's instructions to obtain good quality DNA sample from broth cultures. The quantity and quality of eluted DNA was determined by its OD at 260 nm and the 260/280 nm ratio (1.8-2.0 indicating good quality DNA) using NanoDrop Microvolume UV Spectrophotometer (Thermo Fisher Scientific Inc). The quantity was also determined by Qubit Fluorometric Quantification (Life Technologies) using the manufacturer's instructions and stored at -20^{o}C.

2.4 Polymerase chain reaction (PCR) mixture

Unless otherwise stated, each PCR reaction mixture was 20 µl and consisted of 10x BioLine reaction buffer (2 µl), BioLine dNTP mix 25 mM of each nucleotide (0.4 µl), 50 mM $MgCl_2$ (1 µl), 1µl of each forward and reverse primers to be tested (2 pM/µl), BioLine DNA polymerase (0.2 µl), 1 µl of template DNA and 13.4 µl of nuclease-free water.

2.5 Agarose gel electrophoresis and PCR product purification

For PCR products which were between 200 and 3000 bp, the PCR product was examined on a 1.5% (w/v) agarose gel in 1 x Tris-borate-EDTA (TBE) stained with EtBr (0.2 µg ml^{-1}), with 1 x TBE as the running buffer. The bands were viewed under UV light. If the image showed a clean, single band the PCR product was purified directly. In the case of multiple bands, the correct size band was extracted using a scalpel. Both the extracted band contained in agarose gel and the direct PCR product was purified using the Wizard SV gel and PCR clean up kit (Promega Corp.) using the manufacturer's instructions. The DNA was quantified (1 µl) using the NanoDrop.

2.6 DNA sequencing and sequence analysis

Single read DNA sequencing was performed by MWG Eurofins using the value read service unless otherwise stated. Sequencing was carried out using the primers used in PCR amplification and the ab1 files were downloaded into Geneious (Biomatters Ltd). Once in Geneious, poor quality regions were removed (<75% HQ) (Figure 2.1). Both forward and reverse sequences (when available) were aligned using the MUSCLE alignment tool in Geneious using default options and 8 iterations (Edgar, 2004). The consensus sequence was then used to search either the NCBI or RDP databases using BLAST to determine the identity of the PCR product.

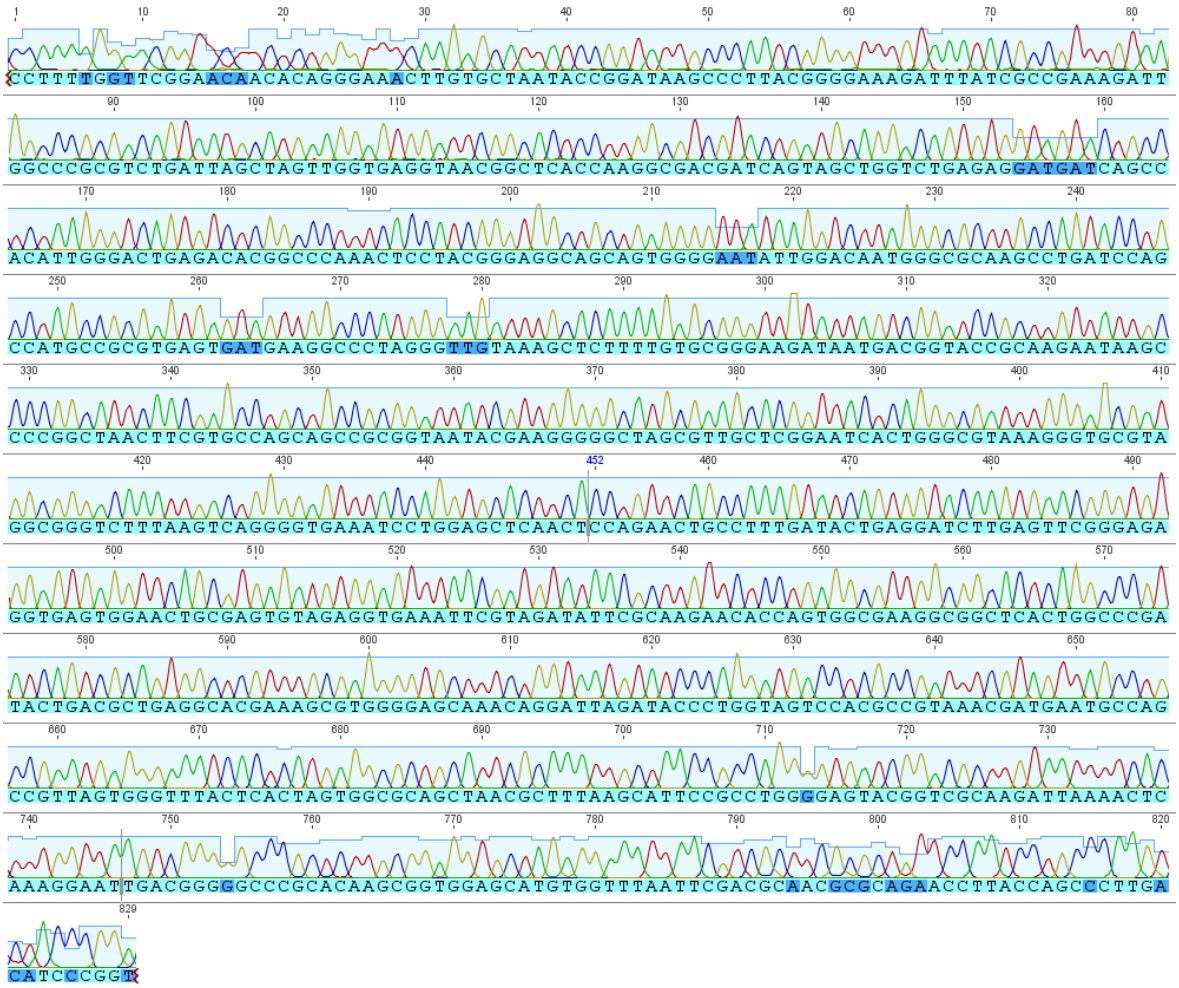


Figure 2.1: Example of a high quality sequence visualised in Geneious.

2.7 Storage of isolates

Isolates confirmed as *Bradyrhizobium* were stored on beads in a glycerol solution (Microbank, Pro-Lab Diagnostics Inc) at -80^{o}C. Colonies were placed into the bead glycerol solution and then were shaken to force the cells into the central cavity of the beads. The excess glycerol solution was then removed and the culture was placed immediately at -80^{o}C. Colonies were also stored on MAG agar slopes and stored at 4^{o}C.

3 Isolation & identification of *Bradyrhizobium* from soil and root nodules

3.1 Introduction

The importance of *Bradyrhizobium* was described in Chapter 1 and the aim of this thesis is to improve the understanding of the diversity and function of *Bradyrhizobium* in soils from long-term field experiments at Rothamsted Research, UK. The initial step of this process was to create a culture collection through isolating *Bradyrhizobium* from soils subjected to different treatments.

3.1.1 Isolation of bacteria from root nodules

The genus *Bradyrhizobium* was originally known for the ability to form nodules on the roots of legumes for symbiotic nitrogen fixation. Bacteria able to nodulate are found across the globe and are a diverse group which demonstrates their ability to survive in a range of climates and nodulate a range of legume species and are classified within the alpha-Proteobacteria and beta-Proteobacteria (Wang et al., 1998, Moulin et al., 2001, Saeki et al., 2010). Soybeans (*Glycine* spp.) are a major crop in America, Brazil and other tropical regions and are nodulated by numerous strains of *Bradyrhizobium*. Due to the economic value of soybeans, much of the literature is focussed on looking at its symbionts. In many previous studies, *Bradyrhizobium* species have been isolated from nodules (van Berkum and Fuhrmann, 2000, Giraud et al., 2007, Torres et al., 2015). The nodules are harvested, surface sterilised and then crushed before being plated onto agar to allow the bacterial cells that have not differentiated into nitrogen-fixing bacteroids to grow (Lange, 1961, Dupuy and Dreyfus, 1992, Weir et al., 2004). Some rhizobia, once differentiated lose the ability to grow however some bradyrhizobia are still viable. Isolation of bacteria from nodules is very selective however rhizobia have differing nodulation abilities and a wide range of legumes are able to form a symbiotic relationship with a number of rhizobial strains (Louvrier et al., 1995).

Another technique for isolating bacteria from the soil community is the trap plant method. To isolate and enumerate rhizobia in soil, a host plant is used and nodules counted, however; this assumes that there are rhizobia present in the soil which are able to form root nodules on the chosen host plant (Woomer et al., 1988, Garrity et al., 2005). A range of trap plants are used depending on the isolates of interest and include cowpea (*Vigna unguiculata*) (Guimaraes et al., 2012, Silva et al., 2014), alfalfa (*Medicago sativa* L.) (Del Papa et al., 1999) and common bean (*Phaseolus vulgaris*) (Roman-Ponce et al., 2016). The trap plant method would not be suitable for isolating free-living rhizobia from soil which do not form root nodules.

Enumeration of specific bacteria in soil is usually achieved through the most probable number technique first described by McCrady (McCrady, 1915). This is a method where dilution to extinction of bacterial cells occurs through several consecutive dilution plating steps that are replicated to provide statistical confidence (Tong and Sadowsky, 1994). This method is usually used to quantify bacteria in the given solution or environment (Cochran, 1950, Woomer et al., 1988, Tong and Sadowsky, 1994).

3.1.2 Isolation of bacteria from soil

Bacteria are frequently isolated from soil through plating serial dilutions of soil suspensions on appropriate media. The free-living strain *Bradyrhizobium* sp. S23321 and *Bradyrhizobium* sp. S58 were isolated in this way (Hattori and Hattori, 1980, Mitsui et al., 1997). In the case of S23321, serial dilutions of soil supernatant were plated onto a range of media: soil extract medium, nutrient broth and 100x diluted nutrient broth with the aim of assessing the paddy field soil community. Colonies forming in the first three days were marked as were those taking over 6 days to form allowing for distinction between fast and slow growing colonies (Gorlach et al., 1994, Mitsui et al., 1997). Serial dilution plating of soil supernatant is a common method to look at bacterial communities and plating onto agar with differing levels of nutrients enables the culturing of oligotrophic organisms which are typically slow growing in addition to fast growing organisms able to grow on high nutrient containing medium (Gorlach et al., 1994).

3.1.3 Medium design and growth conditions

Selective medium to isolate *Bradyrhizobium* from soil primarily relies on the use of heavy metals or antibiotics to which they are less sensitive than other soil bacteria and fungi (Tong and Sadowsky, 1994, Louvrier et al., 1995). Antibiotics used for selective isolation of rhizobia include cycloheximide, pentachloronitrobenzene, neomycin, sodium benzylpenicillin, kanamycin and streptomycin (Graham, 1969, van Berkum and Maier, 1988). However, antibiotic resistance often spans a wide range of taxa and strains within the same genus can often have varying levels of antibiotic resistance leading to sampling bias and selection of antibiotic resistant strains (Tong and Sadowsky, 1994).

Heavy metals are a common source of pollution in the environment and high concentrations are toxic to many microorganisms and so resistance to these metals evolves (Kinkle et al., 1987, Dean-Ross and Mills, 1989). This provides an opportunity to exploit heavy metal resistance and use this in creating a selective medium which can be more specific to the organisms in question.

Resistance to zinc and cobalt are used in a media presented by Tong and Sadowsky (1994). They developed a medium which was selective for *Bradyrhizobium* and growth of other *Rhizobium* spp. was inhibited. However, resistance to a range of heavy metals including lead, copper, aluminium, iron, nickel, cadmium and mercury have also been identified in *B. japonicum* with resistance being variable within this species (Kinkle et al., 1987). In addition, dyes that specifically stain or do not stain desired colonies for example Congo red can be used (Graham, 1969, Tong and Sadowsky, 1994, Louvrier et al., 1995). Use of antibiotics and heavy metals were not used as it has been reported that there are large variations within the same genus and even species in levels of antibiotic and heavy metal resistances (Tong and Sadowsky, 1994). These levels were not known for the population of *Bradyrhizobium* in the soils in question and so were not used as part of the isolation method.

3.1.4 Key characteristics of *Bradyrhizobium* for the isolation method

The *Rhizobium* genus was established in 1889 by Frank (Yoon et al., 2010). A distinction was made between fast and slow growing rhizobia in 1980 and the creation of the *Bradyrhizobium* genus was made primarily due to the time taken to grow on yeast mannitol agar (Jarvis et al., 1982, Jordan, 1982, Lindström, 1989, Wang et al., 1998). When describing the creation of the *Bradyrhizobium* genus in 1982, it was noted that the optimal growth temperature was 25^{o}C to 30^{o}C and produced colonies which were less than 1 mm in diameter on yeast mannitol agar after seven days (Jordan, 1982). *Bradyrhizobium* takes around seven days and other rhizobia being much faster growing (Fuhrmann, 1990, Streeter, 2003, Okubo et al., 2012). Therefore, all isolates forming within seven days of plating were considered unlikely to be *Bradyrhizobium*.

3.1.5 Culture medium for *Bradyrhizobium* isolation

The medium chosen for isolating and culturing *Bradyrhizobium* was modified arabinose gluconate (MAG) (Full recipe in General Methods 2.1). This is a medium which was first published in 1990 (van Berkum, 1990) and is a modified version of the arabinose gluconate medium published in 1987 and 1988 (Sadowsky et al., 1987, van Berkum and Maier, 1988). Historically, rhizobia have been grown on yeast mannitol agar (Lange, 1961, Graham, 1969, Jarvis et al., 1982). The MAG medium is used by the USDA to culture their rhizobia collection (P Elia 2012, pers. comm., 16 November). *Bradyrhizobium* are not able to grow on standard rich media like Luria-Bertani (LB) and so an alternative is required. *Bradyrhizobium* are able to grow on dilute medium, 1/10 TSA,

however MAG was chosen as it was considered to be more selective to rhizobia and therefore would facilitate the isolation of a large number of bradyrhizobia from soil. The MAG medium is a rich medium with D-gluconic acid, yeast extract and L-arabinose being the carbon sources. The yeast extract and ammonium chloride provide the nitrogen sources.

3.1.6 Use of 16S rRNA gene for identification

Traditionally, the ability to grow on different media, growth rate, growth inhibitors, optimal growth temperatures, morphology and carbohydrate metabolism including the form of the product in this process were used to determine the identity of the isolate in question (Graham, 1969, Jarvis et al., 1982, Jordan, 1982, Lane et al., 1985). However, with the development of molecular methods, the 16S ribosomal RNA gene sequence became a widely accepted method of identifying bacteria. The 16S rRNA gene along with other rRNA genes occurs in an operon, is ubiquitous across bacteria, and is highly conserved; sequence change is slow allowing for evolutionary relationships to be tested (Fox et al., 1980, Lane et al., 1985, Kundig et al., 1995). The 16S gene is approximately 1500 bp and regions along the gene change at different rates resulting in both conserved and variable regions that provide a way of assessing relatedness between bacterial groups (Woese, 1987, Yoon et al., 2010). Sequencing provides information on different isolates and can rapidly provide robust identification and allow for phylogenetic comparisons to be made with other isolates across the database (Woese, 1987, Edwards et al., 1989, Weisburg et al., 1991, Young et al., 1991). Primers should be designed to amplify both conserved and variable regions to allow distinction between isolates. A single 16S rRNA gene was noted in *B. japonicum* (Kundig et al., 1995) however some *Bradyrhizobium* strains have two copies (Klappenbach et al., 2001, Acinas et al., 2004).

3.1.7 Aims

In this chapter, I aim to:

- Develop a robust isolation method for *Bradyrhizobium* from soil
- Design *Bradyrhizobium* specific 16S rRNA gene primers to use in screening
- Create a culture collection of *Bradyrhizobium* isolates from a range of soils from Rothamsted sites and nodules from gorse *(Ulex europeaus)* and broom *(Cytisus scoparius)*
- Assess the 16S rRNA diversity of the culture collection

3.2 Materials and methods

3.2.1 Description of study sites

The soil samples used for isolating *Bradyrhizobium* were taken from two long-term experiments at Rothamsted Research, UK; the Highfield experiment and the Broadbalk experiment. The Highfield experiment has been managed as a traditional grassland, planted with a grass/clover mix since 1843. In 1949, parts of this grassland were converted into arable plots which are planted with winter wheat (*Triticum aestivum*). In 1959, bare fallow plots were established which are plots containing no plants and are maintained by tilling several times a year. This experiment allows for comparison between grassland, arable and bare fallow soils and to understand the influence that above ground management and vegetation has on the below ground communities (Hirsch et al., 2009). Soil samples were taken from plot 10 (permanent grassland), plot 8.1 (permanent bare fallow) and plot 14 (permanent arable) (Figure 3.1).

The Broadbalk experiment was established in 1843 and is continuous wheat or wheat rotation experiment with strips of the field receiving a range of fertiliser treatments comprising no additions, farmyard manure to various quantities of nitrogen, phosphorus, potassium, sodium and magnesium. In 1882, an area of the Broadbalk experiment was left unharvested and was no longer used for cultivation. This area, the Broadbalk wilderness, is approximately 0.2 ha with half of this being untouched woodland, a quarter is stubbed and a quarter of this area is regularly mowed. The untouched woodland area is dominated by ash, sycamore and hawthorn and is the area where the woodland soil samples were taken (Figure 3.2).

Figure 3.1: Aerial photograph and plan of the Highfield experiment plots.

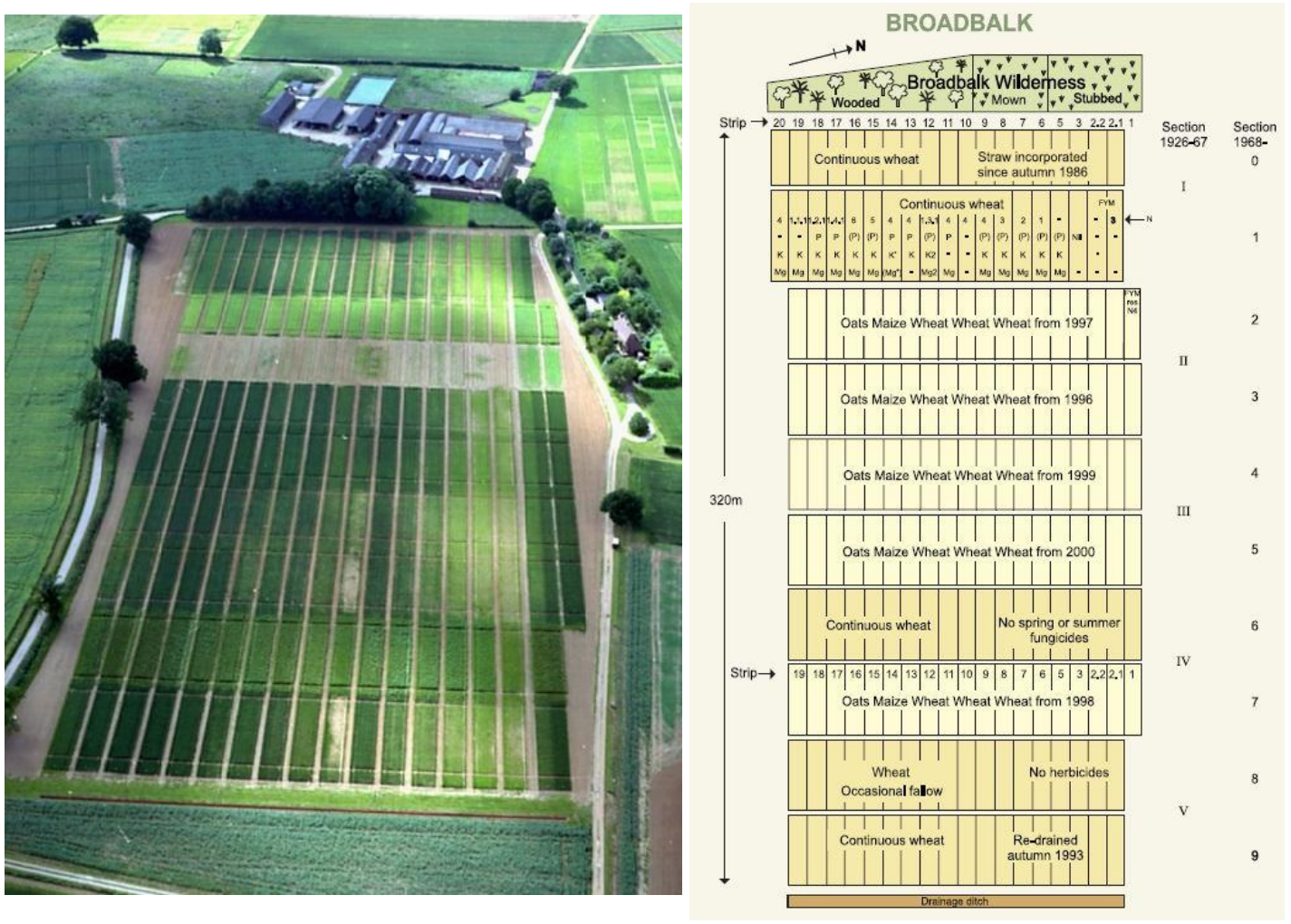


Figure 3.2: Aerial photograph and schematic diagram of the Broadbalk experiment.

3.2.2 Isolation of free-living *Bradyrhizobium* isolates

In order to isolate free-living *Bradyrhizobium* from the Highfield and Broadbalk soil, 10 cm depth soil cores were taken from the plots described in Section 3.2.1. All equipment was sterilised with 70% ethanol prior to sampling and between plots. Ten cores were taken across each plot, homogenised to avoid sampling bias and 10 g was taken from each plot to be resuspended and serially diluted.

The 10 g soil samples were diluted ten-fold using 10 ml sdH_2O and vortexed for 15 minutes to allow bacterial cells to be dislodged from the soil particles and to be brought into suspension. Serial dilutions were carried out to a dilution of 10^{-9} for each soil type and 100 µl was plated onto MAG agar plates (see General Methods 2.1 for recipe). The plates were incubated at 28^{o}C and all colonies appearing in the first 7 days were marked to be ignored as *Bradyrhizobium* typically takes 7 days to form visible colonies. All colonies appearing after 7 days were isolated, numbered and re-streaked onto fresh MAG agar plates. This process was repeated several times in order to generate a large number of isolates from all soil types to be screened using PCR and sequencing of the 16S rRNA gene.

3.2.3 Isolation of symbiotic *Bradyrhizobium* isolates

Gorse (*Ulex europeaus*) and broom (*Cytisus scoparius*) plants from Harpenden Common (close to Rothamsted Research, Harpenden, UK) were harvested and nodules on the roots were examined. Figure 3.3 shows the harvested gorse and broom plants and the presence of root nodules. Four nodules from each plant species were extracted and placed in 11% sodium hypochlorite solution for 3.5 minutes. The nodules were then transferred to sdH2O and rinsed for 1 minute, the water washing step was repeated six times with fresh sterile distilled water each time. The nodules were crushed with a pestle and mortar sterilised with 70% ethanol, plated onto MAG agar plates and incubated at 28^{o}C before being screened using PCR. Nodulation of soybean (*Glycine max*) is examined in Chapter 7.

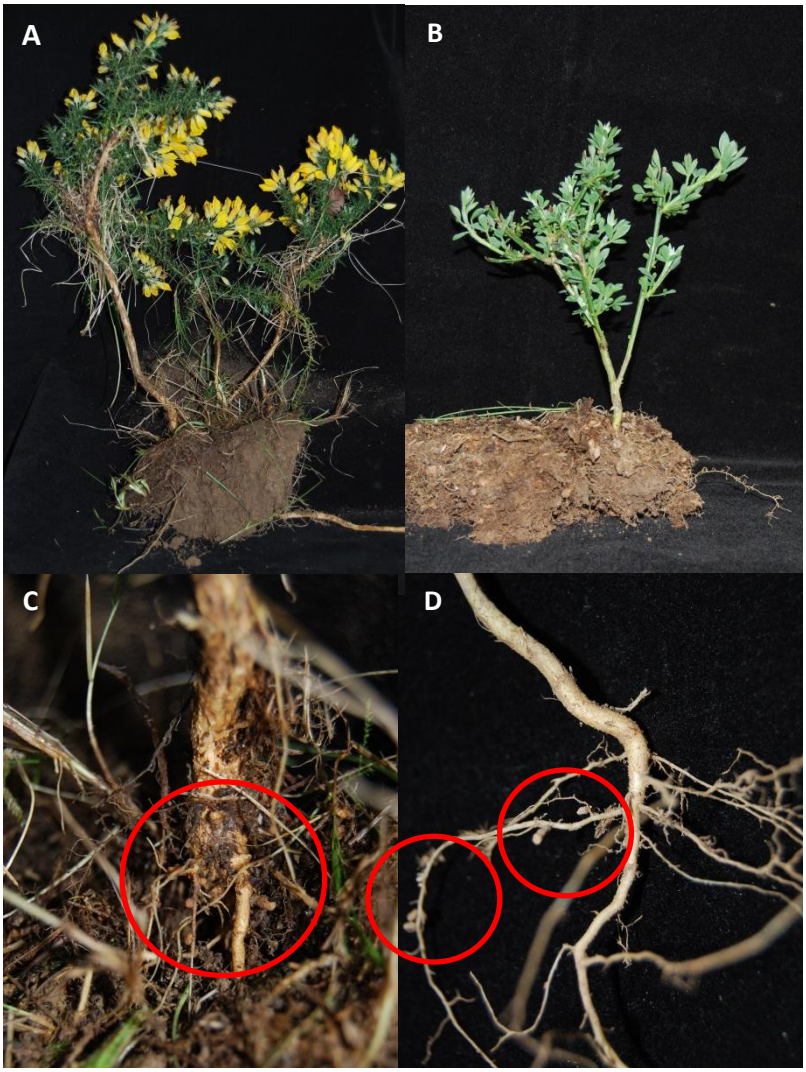


Figure 3.3: Gorse (*Ulex europaeus*) (A+C) and broom (*Cytisus scoparius*) (B+D) plants and root nodules (circled in red).

3.2.4 *Bradyrhizobium* specific 16S primer design

Primers specific to the genus *Bradyrhizobium* amplifying the 16S rRNA gene were designed using published 16S sequences for type strains in the family Bradyrhizobiaceae from the Ribosomal Database Project (RDP) (Database accessed 11/2012) (Accession numbers listed in Table 3.1) (Cole et al., 2014). *B. diazoefficiens* USDA 110 was not included in the primer design process as at the time of design, this was classified as *B. japonicum* and so *B. japonicum* USDA 6 sequence was used as the type strain. Regions of the gene which showed specificity to the *Bradyrhizobium* genus were selected and compared to all sequences in the database to check the specificity using the MUSCLE alignment tool in Geneious (v 7) (Edgar, 2004). Three regions were identified; two were used for forward primers and one for the reverse primer (See Figure 3.4). This gave two primer sets amplifying both a large and small fragment of the 16S gene; Bradj16S70F

(5’- GCGGGCGTAGCAATACGTCAGC-3’) and Bradj16S1430R (5’- GCCGGCTGCCTCCCTTGCGGGTTA-3’) amplifies a 1360bp region and Bradj16S1217F (5’- TGCTAAGGGGCGACCCTTCGC-3’) and Bradj16S1430R amplifies a 213bp product. Confirmation of their specificity was achieved *in vitro* by using NCBI blast and the RDP Seqmatch. Regions were identified which allowed for amplification of *Bradyrhizobium* but excluded other closely related strains, variable regions of the 16S rRNA gene were chosen for this purpose. This was to limit the number of colonies being sent for sequencing and to make the PCR screen more rigorous. Some sampling bias may be introduced by this screening method; choosing the variable regions in the 16S gene would result in some *Bradyrhizobium* strains being discarded if the 16S sequence was distinct from the sequences used in the primer design as the primers would not bind during the PCR.

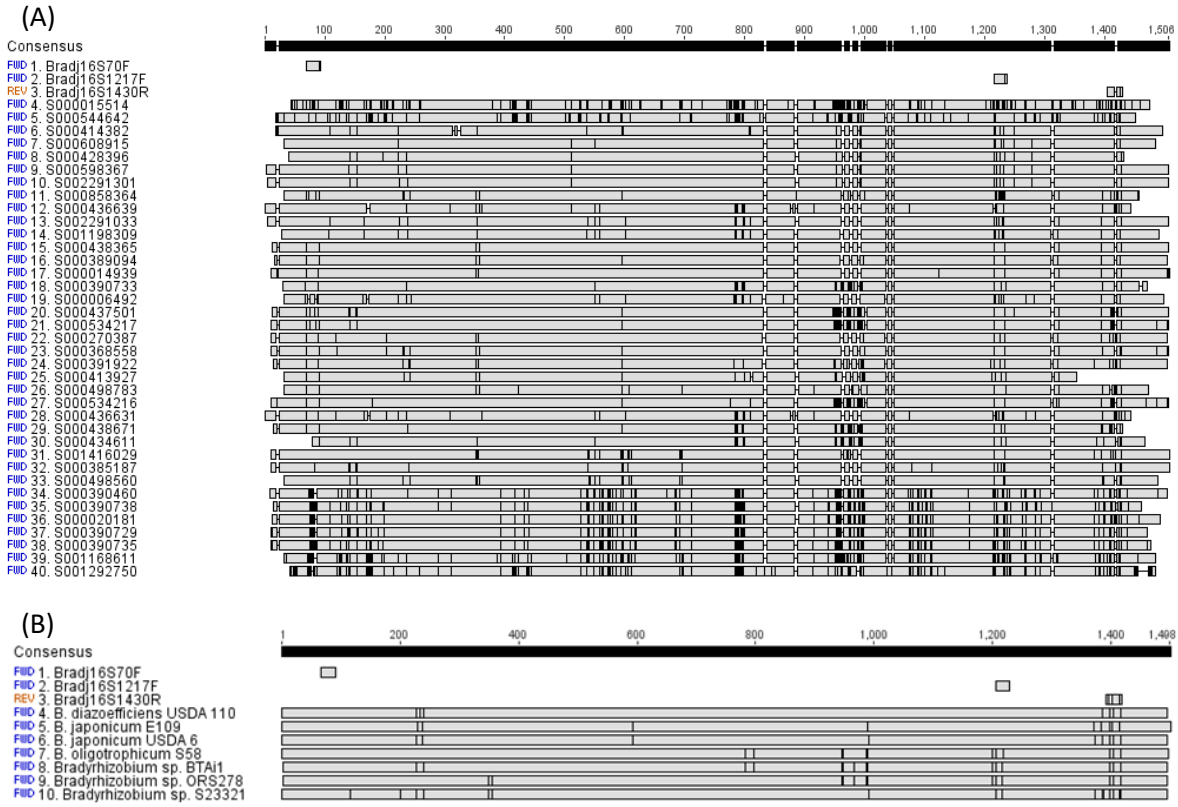


Figure 3.4: 16S sequence alignment showing the position of the designed *Bradyrhizobium*-specific primers Bradj16S70F, Bradj16S1217F and Bradj16S1430R.
(A) shows the alignment of all sequences downloaded from the RDP database.
(B) shows the position of the primers for reference *Bradyrhizobium* sequences from the NCBI database included in the sequence analysis.

Table 3.1: Accession numbers for sequences used to design *Bradyrhizobium* 16S primers

Locus ID	Sequence identity	Accession number
S000006492	*Pseudomonas carboxydohydrogena*	AB021393
S000014939	*Rhizobium lupini*	X87273
S000015514	*Blastobacter capsulatus*	X73042
S000020181	*Bosea thiooxidans*	AJ250796
S000270387	*Bradyrhizobium canariense*	AJ558025
S000368558	*Bradyrhizobium betae*	AY372184
S000385187	*Rhodopseudomonas rhenobacensis*	AB087719
S000389094	*Bradyrhizobium liaoningense*	AF208513
S000390460	*Bosea minatitlanensis*	AF273081
S000390729	*Bosea eneae*	AF288300
S000390733	*Afipia birgiae*	AF288304
S000390735	*Bosea vestrisii*	AF288306
S000390738	*Bosea massiliensis*	AF288309
S000391922	*Bradyrhizobium denitrificans*	AF338176
S000413927	*Bradyrhizobium oligotrophicum*	D78366
S000414382	*Nitrobacter hamburgensis*	L11663
S000428396	*Nitrobacter alkalicus*	AF069956
S000434611	*Afipia massiliensis*	AY029562
S000436631	*Afipia felis*	M65248
S000436639	*Afipia clevelandensis*	M69186
S000437501	*Bradyrhizobium elkanii*	U35000
S000438365	*Bradyrhizobium japonicum*	U69638
S000438671	*Afipia broomeae*	U87759
S000498560	*Rhodopseudomonas faecalis*	AF123085
S000498783	*Bradyrhizobium yuanmingense*	AF193818
S000534216	*Bradyrhizobium jicamae*	AY624134
S000534217	*Bradyrhizobium pachyrhizi*	AY624135
S000544642	*Rhodoblastus sphagnicola*	AM040096
S000598367	*Nitrobacter winogradskyi*	CP000115
S000608915	*Nitrobacter vulgaris*	AM114522
S000858364	*Bradyrhizobium iriomotense*	AB300992
S001168611	*Microvirga flocculans*	AB098515
S001198309	*Oligotropha carboxidovorans*	CP001196
S001292750	*Salinarimonas rosea*	EU878006
S001416029	*Rhodopseudomonas palustris*	AB498815
S002291033	*Oligotropha carboxidovorans*	CP001196
S002291301	*Nitrobacter winogradskyi*	CP000115

3.2.5 PCR optimisation

The PCR conditions were optimised using a gradient PCR program (Bio-Rad MJ Research PTC-200 thermal cycler), where the annealing temperature can be set to a gradient to determine the temperature at which the primers work most efficiently. The melting temperature (T_m) of the forward primers was 65.7°C (Bradj16S70F) and 65.8°C (Bradj16S1217F) and 71.3°C for the reverse primer (Bradj16S1430R). As a result, the gradient chosen to test was 60-68°C. Each PCR reaction mixture was as stated in General Methods 2.4. The positive control test template DNA was extracted from *B. japonicum* USDA 6. The PCR conditions were as follows: held at 95°C, followed by 30 cycles of 95°C for 15 seconds, the annealing gradient (60-68°C) for 15 seconds and 72°C for 1 minute. The product was held at 72°C for 10 minutes for full elongation of the product and held at 10°C until storage at -20°C. The PCR product was examined on a 1.5% (w/v) agarose gel as described in General Methods 2.5 using either a 100 bp or 1 kb ladder marker for comparison and showed a strong band across all temperatures (Figure 3.5). Negative controls were PCR reactions containing no DNA and were tested at the lowest, medium and highest annealing temperatures. A higher annealing temperature increases the specificity of the PCR reaction and so an annealing temperature of 68°C was used for these primer sets.

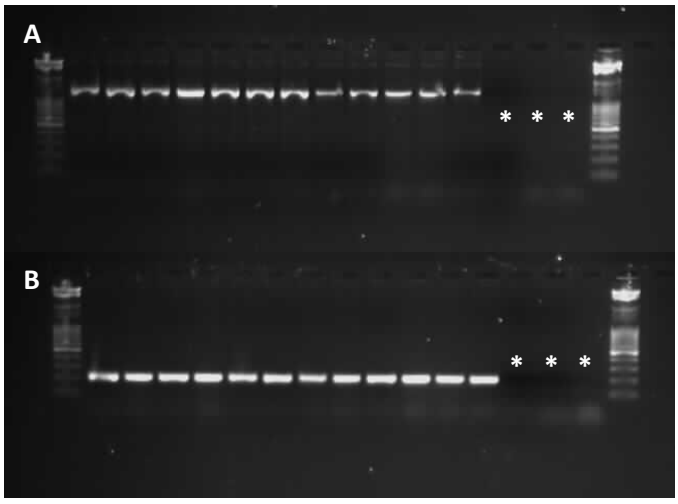


Figure 3.5: Gradient PCR agarose electrophoresis gel.
Annealing gradient from left-right: 60°C-68°C
* indicates a negative control (no DNA)
A: Bradj16S70F and Bradj16S1430R, 1360bp product.
B: Bradj16S1217F and Bradj1430R, 213bp product.
The first and last lanes are 100bp ladders

3.2.6 PCR amplification

For screening isolates, the 16S rRNA was amplified using the long fragment primer set (Bradj16S70F and Bradj16S1430R) to obtain a 1360 bp product. This primer set was chosen so that a nearly full length fragment could be sequenced. This would mean that subsequent sequence analysis would be more robust than if the 213 bp fragment was sequenced. The reaction mixture was as described in General Methods 2.4 and PCR conditions were as follows: held at 95^{o}C, followed by 30 cycles of 95^{o}C for 15 seconds, annealing temperature of 68^{o}C for 15 seconds and 72^{o}C for 1 minute. The product was held at 72^{o}C for 10 minutes for full elongation of the product and held at 10^{o}C until storage at -20^{o}C. The PCR products (5 µl) were examined on a 1.5% (w/v) agarose gel, purified and sent for sequencing as described in General Methods 2.5 and 2.6.

3.2.7 Reference isolates included in the culture collection

Two strains of *Bradyrhizobium*; *B. japonicum* USDA 6 and *Bradyrhizobium* sp. RD5-C2, were included in the culture collection as reference isolates as they have already been characterised. *B. japonicum* USDA 6 is a strain which was originally isolated from *Glycine max* nodules in Japan (van Berkum and Fuhrmann, 2000) and the full genome has been sequenced and is available in the NCBI database (Kaneko et al., 2011). A slope of this strain was provided by Patrick Elia from the USDA so that it could be included in this analysis. *Bradyrhizobium* sp. RD5-C2 was isolated from soil in Hawaii which was covered with lava before human contact and so is considered a pristine environment (Kamagata et al., 1997). This strain was originally isolated to look at 2,4-dichlorophenoxyacetic acid (2,4-D) degradation (Itoh et al., 2000, Itoh et al., 2002).

Other *Bradyrhizobium* strains which have complete genome sequences available were included in the sequence analysis: their 16S rRNA gene sequences were downloaded from the NCBI database. This includes *B. diazoefficiens* USDA 110 which was isolated from *G. max* nodules in the USA (van Berkum and Fuhrmann, 2000). Strains which induce stem and root nodules on *Aeschynomene*: *Bradyrhizobium* sp. ORS 278, *Bradyrhizobium* sp. BTAi1 and *B. oligotrophicum* S58 were isolated from *Aeschynomene* stem nodules in Africa (Giraud et al., 2007), *Aeschynomene* stem nodules in North America (Giraud et al., 2007) and serial dilution plating paddy field soil in Japan (Hattori and Hattori, 1980) respectively. Despite being isolated from soil, *B. oligotrophicum* S58 is able to induce nodules on *Aeschynomene* (Okubo et al., 2013). *B. japonicum* E109 was isolated from *G. max* nodules in Argentina. This was a re-isolation of *B. japonicum* USDA 6 from *G. max* nodules grown in Argentina (Torres et al., 2015). *Bradyrhizobium* sp. S23321 was isolated by serial dilution

plating paddy field soil from Japan (Mitsui et al., 1997). This is a free-living strain which is unable to form nodules on legume roots (Okubo et al., 2012).

3.2.8 Sequence analysis

Sanger sequencing of amplified regions from *Bradyrhizobium* isolates was carried out as described in General Methods 2.6. Sequences were imported in Geneious for editing and submitted into both the NCBI as described in General Methods 2.6 but restricting the BLAST search to type strains. The similarity in the sequence is expected to be 90-100%. All *Bradyrhizobium* isolates were compared to reference full length 16S rRNA sequences from the strains described in Section 3.2.7 which have complete genome sequences available. Accession numbers for the genomes for these strains are shown in Table 3.2, numbered 1-7. Type strains for *Bradyrhizobium* species identified from the BLAST searches were also included in the alignments (Accession numbers are listed in Table 3.2, numbered 9-17). All *Bradyrhizobium* isolates were aligned to each other and reference strains from the NCBI database using the MUSCLE alignment tool in Geneious using the default options and 8 iterations. This enables comparison of the sequences and indicates differences between them. A region of the alignment which was approximately 1191 bp was extracted so that the sequences aligned over an equal length. The extracted sequences were re-aligned using the same parameters so that bootstrapping could be used when creating the phylogenetic tree. A phylogeny using the 16S alignment was created using Geneious tree builder with Tamura-Nei as the genetic distance model, neighbour-joining as the tree building method, 1000 bootstraps and branches were collapsed with less than a 70% threshold bootstrap support. The outgroup was *Mesorhizobium loti* MAFF303099 (Accession number included in Table 3.2). Newick files were uploaded into iTOL so that clades could be highlighted depending on soil treatment. Sequences were uploaded to NCBI, accession numbers can be found in Appendix A.

Table 3.2: Accession numbers for sequences used in the 16S phylogeny analysis

	Isolate identification	Isolate name	Accession number
1	*Bradyrhizobium* sp.	S23321	NC_017082
2	*B. diazoefficiens*	USDA 110	NC_004463
3	*B. japonicum*	USDA 6	NC_017249
4	*B. japonicum*	E109	NZ_CP010313
5	*B. oligotrophicum*	S58	NC_020453
6	*Bradyrhizobium* sp.	ORS278	NC_009445
7	*Bradyrhizobium* sp.	BTAi1	NC_009485
8	*Mesorhizobium loti*	MAFF303099	NR_074162.1
9	*B. liaoningense*	2281	NR_114611.1
10	*B. neotropicale*	BR 10247	NR_133987.1
11	*B. erythrophlei*	CCBAU53325	NR_135877.1
12	*B. rifense*	CTAW71	NR_116361.1
13	*B. ottawaense*	OO99	NR_133988.1
14	*B. elkanii*	USDA 76	NR_117947.1
15	*B. cytisi*	CTAW11	NR_116360.1
16	*B. lupini*	USDA 3051	NR_134836.1
17	*B. canariense*	BTA-1	NR_042177.1

3.3 Results

3.3.1 Creation of a Rothamsted *Bradyrhizobium* culture collection

The culture collection consists of sequenced *Bradyrhizobium* isolates from the grassland (11 isolates), bare fallow (10 isolates), arable (10 isolates) and woodland (11 isolates) soils, 3 isolates from gorse nodules and 2 isolates from broom nodules. A total of 668 slow growing isolates were screened: 277 from grassland, 194 from bare fallow, 50 from arable and 147 from woodland. The 16S rRNA gene fragment of each of these isolates has been sequenced and confirmed as *Bradyrhizobium* (Figure 3.6 & Tables 3.3, 3.4, 3.5, 3.6 & 3.7). The positive controls were DNA from *B. japonicum* USDA 6 extracted using microLYSIS-Plus and the GenElute kit as described in General Methods 2.3. Negative controls were reactions containing no DNA, only microLYSIS-Plus solution and water.

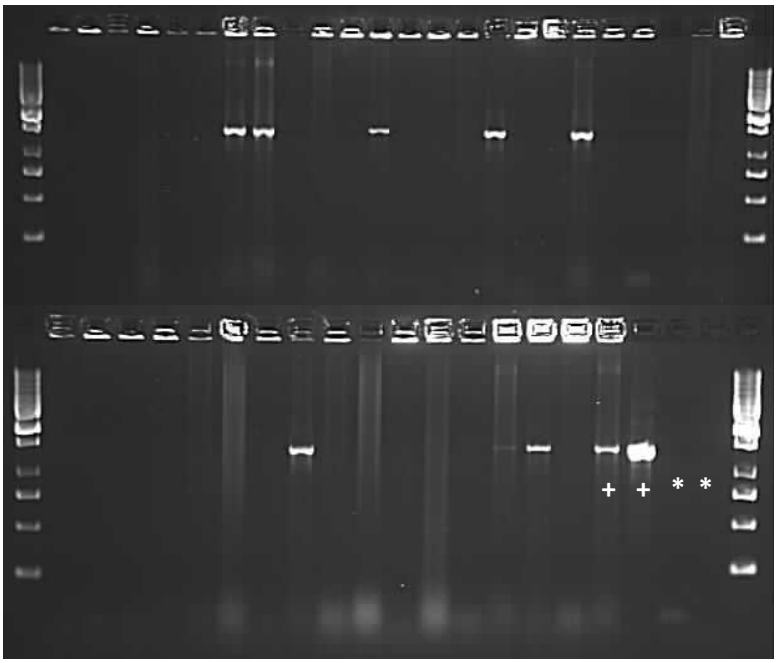


Figure 3.6: An example of an agarose electrophoresis gel used in screening possible isolates.
A band indicates a positive for *Bradyrhizobium*. The first and last lanes are 100bp ladders. + indicates a positive control.
* indicates a negative control (no DNA)

Table 3.3: NCBI BLAST results summary for the grassland isolates

Isolate number	Isolate	Treatment	Type strain best matches	Strain name of best match	Accession of best match	NCBI max score	% Identical Sites	% Pairwise Identity	Query coverage
1	G2	Grassland	*B. canariense*	BTA-1	NR_042177	2156.31	99.80%	99.80%	100.00%
2	G3	Grassland	*B. ottawaense*	OO99	NR_133988	2214.02	99.90%	99.90%	100.00%
		Grassland	*B. liaoningense*	2281	NR_114611	2214.02	99.90%	99.90%	100.00%
3	G5	Grassland	*B. erythrophlei*	CCBAU 53325	NR_135877	2150.9	99.70%	99.70%	100.00%
		Grassland	*B. elkanii*	SEMIA 6208	AY904773	2150.9	99.70%	99.70%	100.00%
4	G6	Grassland	*B. canariense*	BTA-1	NR_042177	2156.31	99.80%	99.80%	99.83%
5	G8	Grassland	*B. rifense*	CTAW71	NR_116361	2208.61	99.80%	99.80%	100.00%
6	G9	Grassland	*B. erythrophlei*	CCBAU 53325	NR_135877	2064.34	98.10%	98.10%	100.00%
		Grassland	*B. elkanii*	SEMIA 6208	AY904773	2064.34	98.10%	98.10%	100.00%
7	G15	Grassland	*B. ottawaense*	OO99	NR_133988	2195.99	99.90%	99.90%	100.00%
		Grassland	*B. liaoningense*	2281	NR_114611	2195.99	99.90%	99.90%	100.00%
8	G22	Grassland	*B. canariense*	BTA-1	NR_042177	2226.64	99.80%	99.80%	100.00%
9	G31	Grassland	*B. canariense*	BTA-1	NR_042177	2212.22	99.80%	99.80%	100.00%
10	G46	Grassland	*B. erythrophlei*	CCBAU 53325	NR_135877	2154.51	99.30%	99.30%	100.00%
		Grassland	*B. elkanii*	SEMIA 6208	AY904773	2154.51	99.30%	99.30%	100.00%
11	G54	Grassland	*B. canariense*	BTA-1	NR_042177	2165.33	99.80%	99.80%	100.00%

Table 3.4: NCBI BLAST results summary for the bare fallow isolates

Isolate number	Isolate	Treatment	Type strain best matches	Strain name of best match	Accession of best match	NCBI max score	% Identical Sites	% Pairwise Identity	Query coverage
12	BF2	Bare fallow	*B. canariense*	BTA-1	NR_042177	2156.31	99.80%	99.80%	100.00%
13	BF6	Bare fallow	*B. lupini*	USDA 3051	NR_134836	2183.36	100.00%	100.00%	100.00%
		Bare fallow	*B. canariense*	NBRC 103049	NR_114199	2183.36	100.00%	100.00%	100.00%
14	BF20	Bare fallow	*B. ottawaense*	OO99	NR_133988	2159.92	99.90%	99.90%	100.00%
		Bare fallow	*B. liaoningense*	2281	NR_114611	2159.92	99.90%	99.90%	100.00%
15	BF26	Bare fallow	*B. erythrophlei*	CCBAU 53325	NR_135877	2102.21	98.30%	98.30%	100.00%
		Bare fallow	*B. elkanii*	SEMIA 6208	AY904773	2102.21	98.30%	98.30%	100.00%
16	BF28	Bare fallow	*B. erythrophlei*	CCBAU 53325	NR_135877	2156.31	99.40%	99.40%	100.00%
		Bare fallow	*B. elkanii*	SEMIA 6208	AY904773	2156.31	99.40%	99.40%	100.00%
17	BF30	Bare fallow	*B. erythrophlei*	CCBAU 53325	NR_135877	2136.47	98.50%	98.50%	100.00%
		Bare fallow	*B. elkanii*	SEMIA 6208	AY904773	2136.47	98.50%	98.50%	100.00%
18	BF41	Bare fallow	*B. erythrophlei*	CCBAU 53325	NR_135877	2154.51	99.80%	99.80%	100.00%
		Bare fallow	*B. elkanii*	SEMIA 6208	AY904773	2154.51	99.80%	99.80%	100.00%
19	BF48	Bare fallow	*B. erythrophlei*	CCBAU 53325	NR_135877	2073.36	98.30%	98.30%	100.00%
		Bare fallow	*B. elkanii*	SEMIA 6208	AY904773	2073.36	98.30%	98.30%	100.00%
20	BF49	Bare fallow	*B. lupini*	USDA 3051	NR_134836	2219.43	100.00%	100.00%	100.00%
		Bare fallow	*B. canariense*	NBRC 103049	NR_114199	2219.43	100.00%	100.00%	100.00%
21	BF52	Bare fallow	*B. cytisi*	CTAW11	NR_116360	2181.56	99.80%	99.80%	100.00%

Table 3.5: NCBI BLAST results summary for the arable isolates

Isolate number	Isolate	Treatment	Type strain best matches	Strain name of best match	Accession of best match	NCBI max score	% Identical Sites	% Pairwise Identity	Query coverage
22	A1	Arable	*B. lupini*	USDA 3051	NR_134836	2237.46	100.00%	100.00%	100.00%
		Arable	*B. canariense*	NBRC 103049	NR_114199	2237.46	100.00%	100.00%	100.00%
23	A2	Arable	*B. lupini*	USDA 3051	NR_134836	2221.23	100.00%	100.00%	100.00%
		Arable	*B. canariense*	NBRC 103049	NR_114199	2221.23	100.00%	100.00%	100.00%
24	A3	Arable	*B. lupini*	USDA 3051	NR_134836	2219.43	100.00%	100.00%	100.00%
		Arable	*B. canariense*	NBRC 103049	NR_114199	2219.43	100.00%	100.00%	100.00%
25	A4	Arable	*B. canariense*	BTA-1	NR_042177	2159.92	99.80%	99.80%	100.00%
26	A5	Arable	*B. lupini*	USDA 3051	NR_134836	2165.33	100.00%	100.00%	100.00%
		Arable	*B. canariense*	NBRC 103049	NR_114199	2165.33	100.00%	100.00%	100.00%
27	A10	Arable	*B. cytisi*	CTAW11	NR_116360	2147.3	99.70%	99.70%	100.00%
28	A12	Arable	*B. cytisi*	CTAW11	NR_116360	2176.15	99.70%	99.70%	100.00%
29	A16	Arable	*B. erythrophlei*	CCBAU 53325	NR_135877	2208.61	99.70%	99.70%	100.00%
		Arable	*B. elkanii*	SEMIA 6208	AY904773	2208.61	99.70%	99.70%	100.00%
30	A19	Arable	*B. canariense*	BTA-1	NR_042177	2186.97	99.80%	99.80%	100.00%
31	A20	Arable	*B. canariense*	BTA-1	NR_042177	2156.31	99.80%	99.80%	100.00%

Table 3.6: NCBI BLAST results summary for the woodland isolates

Isolate number	Isolate	Treatment	Type strain best matches	Strain name of best match	Accession of best match	NCBI max score	% Identical Sites	% Pairwise Identity	Query coverage
32	W4	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2118.44	98.10%	98.10%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2118.44	98.10%	98.10%	100.00%
33	W5	Woodland	*B. japonicum*	USDA 6	NR_112552	2161.72	100.00%	100.00%	100.00%
34	W9	Woodland	*B. neotropicale*	BR 10247	NR_133987	2118.44	98.10%	98.10%	100.00%
35	W19	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2118.44	98.10%	98.10%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2118.44	98.10%	98.10%	100.00%
36	W22	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2150.9	99.70%	99.70%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2150.9	99.70%	99.70%	100.00%
37	W23	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2206.81	99.70%	99.70%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2206.81	99.70%	99.70%	100.00%
38	W24	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2116.64	98.10%	98.10%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2116.64	98.10%	98.10%	100.00%
39	W46	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2091.39	98.10%	98.10%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2091.39	98.10%	98.10%	100.00%
40	W49	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2203.2	99.80%	99.80%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2203.2	99.80%	99.80%	100.00%
41	W74	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2159.92	99.30%	99.30%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2159.92	99.30%	99.30%	100.00%
42	W91	Woodland	*B. erythrophlei*	CCBAU 53325	NR_135877	2064.34	98.10%	98.10%	100.00%
		Woodland	*B. elkanii*	SEMIA 6208	AY904773	2064.34	98.10%	98.10%	100.00%

Table 3.7: NCBI BLAST results summary for the symbiotic/reference isolates

Isolate number	Isolate	Treatment	Type strain best matches	Strain name of best match	Accession of best match	NCBI max score	% Identical Sites	% Pairwise Identity	Query coverage
44	GOR1	Symbiotic	*B. canariense*	BTA-1	NR_042177	2156.31	99.80%	99.80%	100.00%
45	GOR2	Symbiotic	*B. canariense*	BTA-1	NR_042177	2159.92	99.80%	99.80%	99.92%
46	GOR4	Symbiotic	*B. ottawaense*	OO99	NR_133988	2183.36	99.80%	99.80%	100.00%
	GOR4	Symbiotic	*B. liaoningense*	2281	NR_114611	2183.36	99.80%	99.80%	100.00%
47	BRM1	Symbiotic	*B. cytisi*	CTAW11	NR_116360	2147.3	99.70%	99.70%	100.00%
48	BRM2	Symbiotic	*B. rifense*	CTAW71	NR_116361	2140.08	99.70%	99.70%	99.92%

3.3.2 16S sequence analysis

When comparing the 16S sequences with other *Bradyrhizobium* strains from the database, there is a high level of similarity. There is 95.3% pairwise percent identity which is a measure of how similar the sequences are looking at each base. There is one region of the 16S amplicon at approximately 1000 bp which is highly variable (Figure 3.6). The pairwise identity for the treatments were 97.9% for bare fallow, 99.0% for arable, 98.4% for grassland, 98.7% for symbiotic isolates and 98.6% for woodland isolates. This suggests that the grassland and woodland isolates are more variable than the other treatments.

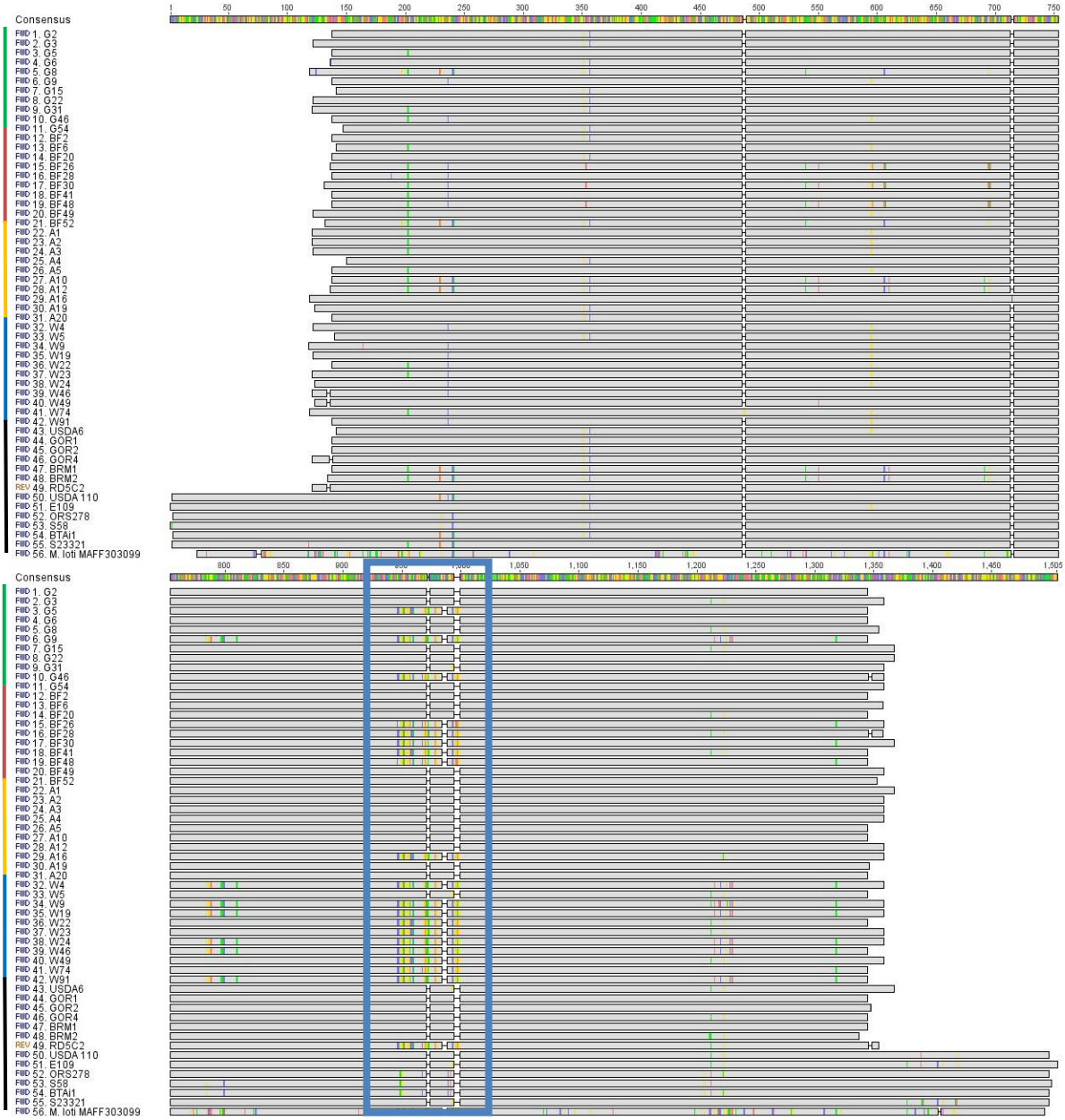


Figure 3.7: Sequence alignment comparing Highfield isolates to reference 16S sequences.
The coloured regions indicate nucleotide differences.
Key: Black = symbiotic/reference, green = grassland, red = bare fallow, orange = arable, blue = woodland

The 16S amplicon phylogeny (Figure 3.8) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. There is a large clade consisting of primarily woodland isolates (W4, W9, W19, W24, W46, W91 and G9) with a bootstrap of 100%. The broom isolates (BRM1 and BRM2) cluster with arable isolates A10 and A12, the bare fallow isolate BF52, the grassland isolate G8 and two type strains, *B. cytisi* CTAW11 and *B. rifense* CTAW71 with a score of 100% with BRM1, BRM2, A10 and A12 separating with a bootstrap of 86.8%. The bare fallow isolates, BF30, BF26 and BF48, cluster closely with a score of 73.1%. The woodland isolate W5 clusters closely with the two *B. japonicum* strains included in the analysis (E109 and USDA6T) with a bootstrap of 100%. There is a large clade with strains from a range of soil treatments and gorse and broom nodules and a number of type strains with a bootstrap score of 100%. The arable isolates, A1, A2, A3 and A5 and the bare fallow isolates BF6 and BF49 form a clade with the *B. lupini* USDA 3051 type strain with a bootstrap of 100%. The 16S phylogeny does not group the isolates according to soil treatment and so suggests that there are a range of species present in all of the soil treatments and that the culture collection contains a range of species. As a result, a high level of diversity is expected in subsequent experiments.

Figure 3.9 shows the same phylogeny has Figure 3.8 but colour coded according to top hit from the BLAST results. Some 16S sequences for the culture collection were matched to two different type strains however, many do not cluster with the corresponding type strain in the phylogeny (Figure 3.8 & 3.9). This suggests that despite the top hit being a particular type strain, more information is needed in order to assign species to these strains. It is also possible that there are a number of novel strains included in the culture collection. Figure 3.9 also shows that two type strains are very similar: *B. cytisi* and *B. rifense.*

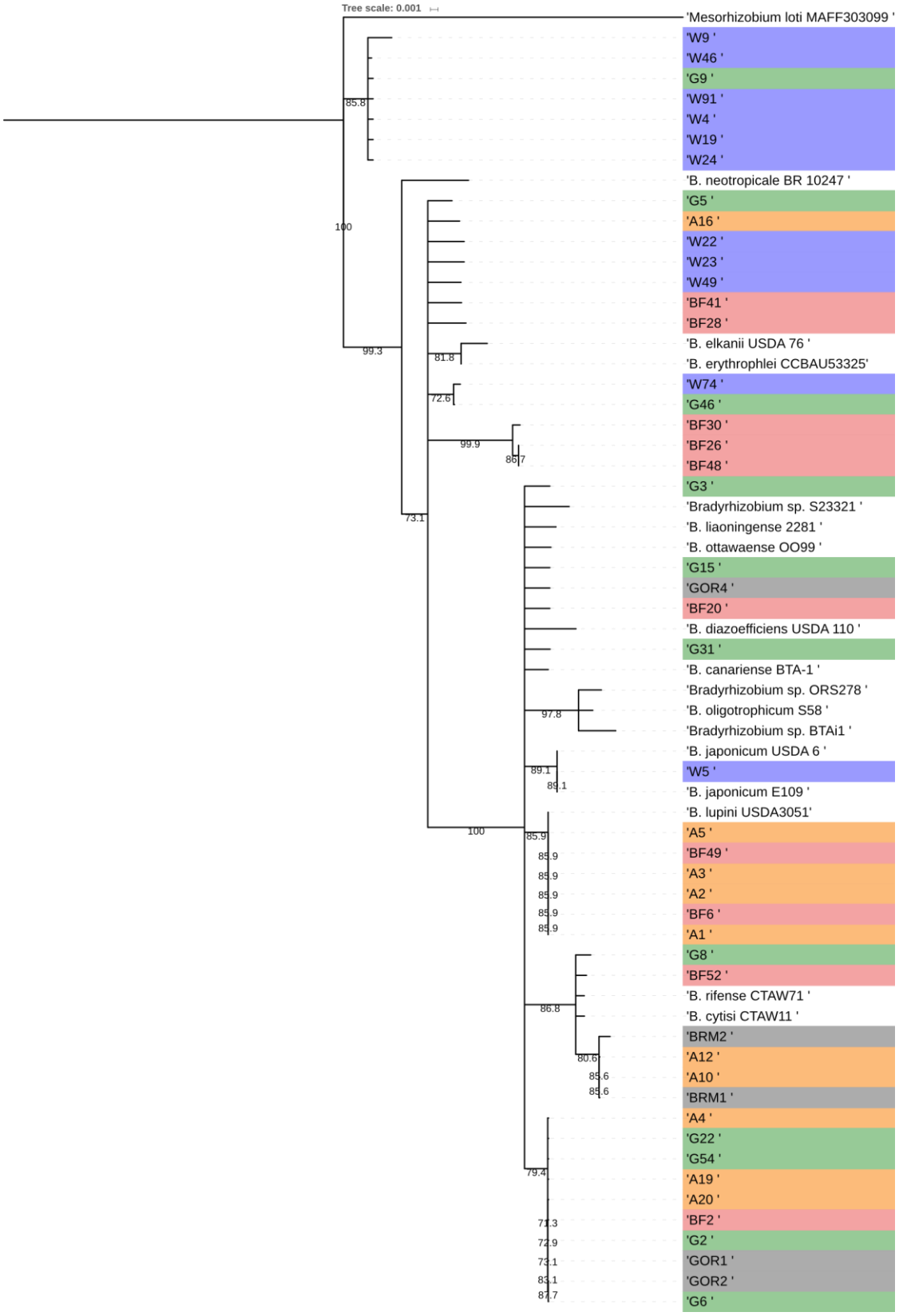


Figure 3.8: Phylogenetic tree based on 16S sequences using Neighbour-Joining and 1000 bootstraps.

Branch labels are the bootstrap values. Key: Grey = symbiotic, green = grassland, pink = bare fallow, orange = arable, blue = woodland.

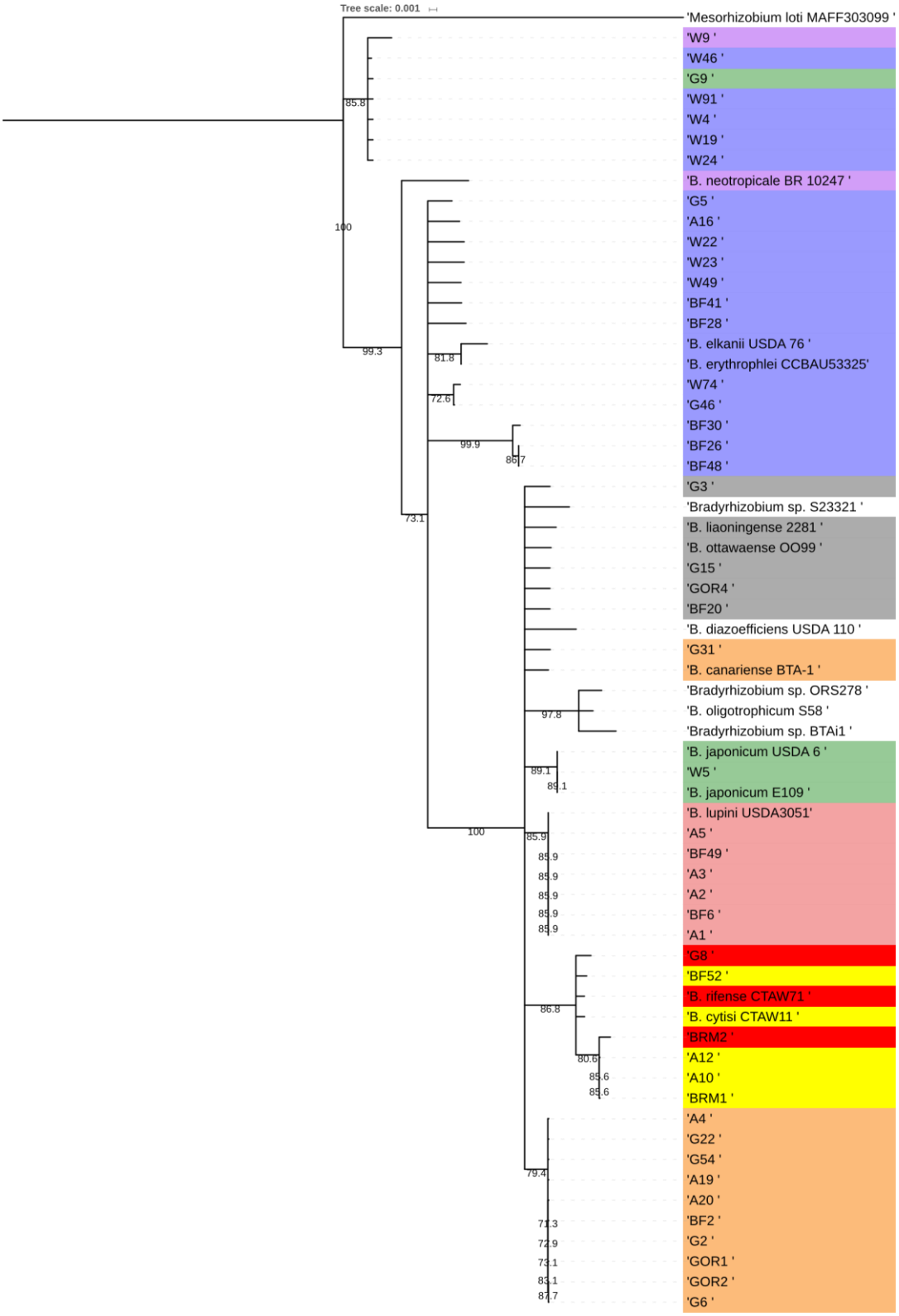


Figure 3.9: Phylogenetic tree based on 16S sequences using Neighbour-Joining and 1000 bootstraps colour coded according to top BLAST hits

Branch labels are the bootstrap values. Key: purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense,* yellow = *B. cytisi*, red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense.*

3.4 Discussion

3.4.1 Summary of the *Bradyrhizobium* culture collection

The isolation technique proved successful and a collection of 42 isolates from varying soil types at Rothamsted (and Harpenden Common) has been created. This is in addition to five isolates from gorse and broom root nodules. All 47 strains have a 16S rRNA gene amplicon with a >98% identity with a *Bradyrhizobium* type strain. The top BLAST hits for the culture collection comprise ten different *Bradyrhizobium* species: *B. canariense*, *B. ottawaense*, *B. liaoningense*, *B. cytisi*, *B. rifense*, *B. erythrophlei*, *B. elkanii*, *B. neotropicale*, *B. japonicum* and *B. lupini*. The 16S phylogeny did not place many of the isolates with the type strain for the top BLAST hit and therefore suggests that the species present in the culture collection could be novel species.

The 16S rRNA gene does not provide a large amount of variability as it is a highly conserved region. Generally, the isolates did not group according to origin however, as the 16S gene is highly conserved and sequence change is slow, this was a predicted result. A wider range of genes should be used to assess relatedness, housekeeping genes for example DNA gyrase (*gyrB*), ATP synthase (*atpD*) or recombinase A (*recA*), are usually used for this (Rivas et al., 2009) and will be examined in Chapter 4.

Creating a culture collection results in sampling bias and inferences of the community as a whole cannot be made from a culture collection. Using the media, incubation temperature, primer sequences used for screening and morphology observations all introduce bias into the isolation method and will reduce the level of diversity seen in the collection. Other members of the *Bradyrhizobium* community will not have been isolated using the set isolation method if there was variability In the 16S sequence, inability to grow at 28^{o}C or inability to grow on the MAG medium or the isolate morphology was different.

3.4.2 Isolation method of *Bradyrhizobium* from soil and nodules

The trap plant method is only suitable for isolating nodulating strains of rhizobia and so it was not selected for isolating *Bradyrhizobium* in this instance as free-living strains were the isolate targets. The trap plant method is also time consuming as it requires the growth of the host before harvesting of the nodules can take place. A semi-selective medium was used to isolate *Bradyrhizobium* from both nodules and soil.

The MAG agar was known to be suitable for growing *Bradyrhizobium* isolates and so formed the basis of the isolation method. Time to form colonies on agar is an established method of distinguishing between *Bradyrhizobium* and other rhizobia and was a simple method of monitoring isolates which were potentially *Bradyrhizobium*. Once a number of isolates had been confirmed as *Bradyrhizobium*, a typical morphology could be identified which increased the hit rate of the isolation method to 25-30%. This was white circular, raised colonies with a full edge and a smooth surface. Figure 3.10 shows *Bradyrhizobium* growing on a MAG agar plate.

Figure 3.10: *B. japonicum* USDA 6 growing on a MAG agar plate.

3.4.3 Importance of a culture collection

Numerous characteristics including both molecular and biochemical functions can be measured in the laboratory through the creation of a culture collection. It also allows for the comparison of isolates within the collection (Gorlach et al., 1994). Culture dependent methods are useful for studying the possession of key functions, for example, those involved in soil nutrient cycling. The culture collection will be used in subsequent experiments to assess the variability, both genetic and metabolic, and to determine whether the isolates group according to where they were isolated from. The collection will also be stored at Rothamsted to be able to be used in future experiments.

4 Multilocus sequence analysis of the *Bradyrhizobium* culture collection

4.1 Introduction

4.1.1 Diversity of *Bradyrhizobium*

Bradyrhizobium is a genus which is widely reported for being extremely heterogeneous with a high level of genetic diversity affecting morphology, symbiotic ability and physiology (Madrzak et al., 1995, Moulin et al., 2004, Giongo et al., 2008, Steenkamp et al., 2008, Menna and Hungria, 2011). The environmental conditions including chemical and physical soil conditions may have a relationship with the genetic diversity of indigenous rhizobia (Giongo et al., 2008). Free-living rhizobia have been found in a wide range of soils; in both bulk soil and the rhizosphere, and as a result considerable diversity is detected genetically and functionally (Tan et al., 2001). The diversity in *Bradyrhizobium* reflects the wide geography, ecosystems and associated legumes with which it is found associated (Vinuesa et al., 2008). In Chapter 3, a culture collection of free-living and symbiotic bradyrhizobia was created and the 16S rRNA gene sequence was used to assess the relatedness. The 16S rRNA gene can be useful to assess phylogeny as it is slow to evolve and is ubiquitous across bacteria however the diversity of the 16S rRNA gene is low which is why despite the large diversity found in *Bradyrhizobium*, there are relatively few species described (van Berkum et al., 2003, Acinas et al., 2004, Menna et al., 2009). The overall aim of this chapter is to assess the relatedness of strains within the culture collection and with sequences available in the database through DNA fingerprinting techniques and multilocus sequence analysis.

4.1.2 Enterobacterial Repetitive Intergenic Consensus DNA fingerprinting

Repetitive DNA sequences have been extensively studied in eukaryotes and in the early 1990s began to be studied in prokaryotes in order to assess relatedness between strains (Versalovic et al., 1991). Prokaryotes were originally thought to have fewer repeat sequences due to genome streamlining however numerous bacterial groups are found to contain relatively high levels of repetitive sequences (Hulton et al., 1991, Giovannoni et al., 2005). Short intergenic repeat sequences were identified in enteric bacteria and include two classes of conserved central inverted repeats; class I are repetitive extragenic palindromic elements (REPs) and class II are enterobacterial repetitive intergenic consensus sequences (ERIC sequences) (de Bruijn, 1992).

REP and ERIC sequences are found in Gram-negative enteric bacteria and the ERIC method was first developed in *Escherichia coli* and *Salmonella typhimurium* (Hulton et al., 1991, Versalovic et al., 1991). Both REP and ERIC sequences are highly conserved and allow for species to be distinguished from each other (Versalovic et al., 1991). The ERIC sequence is 126 bp and the sequence is not homologous to any known plasmid, insertion sequence, phage or eukaryote sequence (Hulton et al., 1991). These sequences are located in different positions depending on the species of bacteria and are sensitive enough to be used to distinguish different strains of *Mycobacterium tuberculosis* isolated from the same patient (Sechi et al., 1998). ERIC sequences have been previously studied in rhizobia and found to be both present and conserved (Laguerre et al., 1997, Giongo et al., 2008). Both REP and ERIC sequences were found to be present in rhizobial genera; *Rhizobium*, *Bradyrhizobium*, *Azorhizobium* and *Agrobacterium* (de Bruijn, 1992). Rhizobia have been shown to be highly diverse and have been studied extensively using DNA fingerprinting techniques which provide a way of assessing relatedness on a larger scale than the 16S rRNA gene alone (Vinuesa et al., 1998).

4.1.3 Multilocus diversity analyses

Bacterial genomes often have multiple copies of the rRNA operon with 40% of bacterial genomes thought to have one or two copies and variable copy numbers have been noted within species for example *Vibrio cholerae* and *Bacillus cereus* (Acinas et al., 2004). Some *Bradyrhizobium* strains have been shown to contain multiple rRNA copies and as it is highly conserved, sequence similarity is high and therefore distinction between species or strains can be problematic (Acinas et al., 2004, Rivas et al., 2009). In addition to the 16S rRNA gene, the internal transcribed spacer (ITS) region and the 23S rRNA gene are often sequenced to assess diversity (van Berkum et al., 2003). Restriction fragment length polymorphism of 16S rRNA gene and the ITS regions have also been used in rhizobia (Bala et al., 2003, Lafay and Burdon, 2007). The 16S-23S intergenic spacer region has been previously assessed and was shown to be more suitable for *Bradyrhizobium* than *Rhizobium* however, this region is not protein-encoding (Tan et al., 2001).

Multilocus sequence analysis (MLSA) of protein encoding genes was found to be a suitable method for assessing the relatedness of isolates and is able to be used to distinguish strains of the same genus or species (Vinuesa et al., 2008). The core genome has been described as a way of differentiating between bradyrhizobial isolates using marker genes for example ITS, *rrs*, *recA* and *glnII* (Steenkamp et al., 2008). Using housekeeping genes has been shown to be a good method for assessing phylogeny in *Bradyrhizobium*, especially when a number of genes are concatenated

(Steenkamp et al., 2008, Menna et al., 2009, Zhang et al., 2014b). The housekeeping genes, *atpD*, *glnII*, *recA*, *rpoB* and concatenated sequences have also previously been used in the assessment of 33 soybean-nodulating isolates and were able to split the isolates into species (Vinuesa et al., 2008). The diversity of *Bradyrhizobium* isolates from a range of host legumes were assessed using *atpD*, *recA*, *dnaK*, *gyrB* and *rpoB* (Rivas et al., 2009). In an assessment of the *Afipia* genus, it was found that *rpoB* was a good candidate gene for distinguishing between species within this genus (Khamis et al., 2003).

Genes involved in nitrogen fixation and nodulation have been previously used to assess diversity of *Bradyrhizobium* including *nodY/K*, *nodA*, *nodZ* and *nifH* in addition to housekeeping genes for example *glnII*, *dnaK* and *recA* (Menna and Hungria, 2011). Nodulation genes form part of the accessory genome with *nodABC* being single copy genes in rhizobia and so are good candidates for MLSA (Moulin et al., 2004, Steenkamp et al., 2008). Other nodulation genes, *nodY/K*, are only present in *Bradyrhizobium* and so are not able to be used in analyses where other rhizobia are included (Menna and Hungria, 2011). Moulin et al. (2004), used 16S rRNA gene and *dnaK* to identify taxonomy but used *nodABC* to assess the phylogeny of *Bradyrhizobium* isolates. Nodulation genes are not feasible for this analysis as the free-living strains in the *Bradyrhizobium* culture collection do not contain symbiotic genes (See Chapter 7 for more information).

4.1.4 Chosen housekeeping genes

The following genes have been chosen to sequence to assess the *Bradyrhizobium* culture collection. All five have been previously used to assess the genetic diversity of nodulating *Bradyrhizobium* and so were chosen as good candidates for assessing the diversity of free-living *Bradyrhizobium*.

1. ***gyrB***: DNA gyrase subunit B involved in genome replication (Rivas et al., 2009)
2. ***atpD***: β-subunit of the membrane ATP synthase (Menna et al., 2009, Rivas et al., 2009)
3. ***recA*:** DNA recombination and repair protein involved in recombination and enabling the exchange of DNA (Rivas et al., 2009)
4. ***dnaK*:** 70kDa chaperone protein. Homologous to the Hsp70 protein (Moulin et al., 2004, Rivas et al., 2009)
5. ***rpoB*:** β-subunit of RNA polymerase which is required for transcription (Rivas et al., 2009). This is a large gene approximately 4000 bp in length (Khamis et al., 2003).

4.1.5 Specific objectives

In this chapter, I aim to:

- Assess the relatedness of the culture collection through ERIC DNA fingerprinting
- Amplify and sequence housekeeping genes for all isolates of the *Bradyrhizobium* culture collection
- Assess the diversity of the culture collection using five different housekeeping genes to determine whether the housekeeping gene sequences group the isolates according to soil treatment
- Determine which of the five housekeeping genes sequenced are most informative of diversity within the culture collection
- Compare the housekeeping gene clades with 16S rRNA gene clades to assess whether soil treatment selects for a specific 16S phylotype

4.2 Materials and methods

4.2.1 Enterobacterial Repetitive Intergenic Consensus DNA fingerprinting

Enterobacterial Repetitive Intergenic Consensus (ERIC) PCR was carried out to generate DNA fingerprints in order to compare the isolates which were confirmed as *Bradyrhizobium.* Each PCR reaction mixture was 25 µl and consisted of 10x BioLine reaction buffer (2.5 µl), BioLine dNTP mix 10 mM each (0.5 µl), 50 mM $MgCl_2$ (1 µl), ERIC2 primer (5-AAGTAAGTGACTGGGGTGAGCG-3') 50 pM/µl (1 µl), R1CIRE primer (5'-CACTTAGGGGTCCTCGAATGTA-3') 50 pM/µl (1µl), BioLine DNA polymerase (0.25 µl), 1 µl of DNA extracted using GenElute bacterial genomic DNA kit (as described in General Methods 2.3) as template and 18 µl of nuclease-free water. Primer sequences are taken from Versalovic et al. (1991). The PCR conditions were as follows: held at 95^{o}C, followed by 30 cycles of 94^{o}C for 3 minutes, 52^{o}C for 1 minute and 72^{o}C for 3 minutes. The samples were next held at 72^{o}C for 5 minutes for elongation and held at 10^{o}C. The PCR products (5 µl) were examined on a 1.5% (w/v) agarose gel as described in General Methods 2.5. Products were run for 2 hours at 100V then 1.5 hours at 80V. A size marker of bacteriophage Lamda DNA digested with *Eco*RI and *Hin*dIII (Lamda/*Eco*RI/*Hin*dIII) (ThermoFisher Scientific Inc, USA) was used for comparison in addition to a 100 bp ladder (ThermoFisher Scientific Inc, USA) for smaller DNA fragments.

4.2.2 ERIC DNA fingerprinting cluster analysis

The ERIC profile for each isolate was converted into a binary matrix using CLIQS (Totallab, UK) where a band was given the value of 1 and the absence of this band was given the value of 0. When determining whether a band was present, minimum slope was set to 100, noise reduction to 10 and percentage maximum peak was set to 4. This binary matrix was then used to create a phylogeny using the *pvclust* package in R (v 3.2.2) where relationships between ERICS profiles could be assessed (Suzuki and Shimodaira, 2006). The UPGMA clustering method was used and 1000 bootstraps. The *P*-values provided are the bootstrap *P*-value (BP) and the approximately unbiased (AU) *P*-value. The BP value is the number of times the cluster is generated in the bootstraps and approximately unbiased probabilities were created through multiscale bootstrapping (Suzuki and Shimodaira, 2006). The AU value aims to reduce bias as the sequence length is varied to generate sets of bootstrap replicates and then generated bootstrap values for

each set. The AU probability is the change in bootstrap *P*-value for each replicate set (Shimodaira, 2002; Suzuki and Shimodaira, 2006).

4.2.3 Housekeeping genes PCR amplification

The PCR reaction mix was as described in General Methods 2.4. The PCR primers used are described in Table 4.1. The PCR conditions for each primer set are listed in Table 4.2, products were held at 10°c until storage at -20°c. The PCR products (5 µl) were examined on a 1.5% (w/v) agarose gel as described in General Methods 2.5. A 100 bp ladder was used as a size comparison (1 kb used for *rpoB*) (ThermoFisher Scientific Inc, USA). PCR amplicons were sequenced as described in General Methods 2.6.

Table 4.1: List of housekeeping genes used and primer information

Gene	Primer name	Primer sequence (5'-3')	Size (bp)	Reference
gyrB	gyrB343F	TTC GAC CAG AAY TCC TAY AAGG	675	Rivas et al. (2009)
	gyrB1043R	AGC TTG TCC TTS GTC TGCG		
atpD	atpD352F	GGC CGC ATC ATS AAC GTS ATC	487	Rivas et al. (2009)
	atpD871R	AGM GCC GAC ACT TCM GAR CC		
recA	recA63F	ATC GAG CGG TCG TTC GGC AAG GG	435	Rivas et al. (2009)
	recA504R	TTG CGC AGC GCC TGG CTC AT		
dnaK	dnaK1466F	AAG GAR CAN CAG ATC CGC ATC CA	286	Rivas et al. (2009)
	dnaK1777R	TAS ATS GCC TSR CCR AGC TTCAT		
rpoB	rpoB83F	CCT SAT CGA GGT TCA CAG AAG GC	1492	Rivas et al. (2009)
	rpoB1540R	AGC TGC GAG GAA CCG AAG		

Table 4.2: PCR conditions for housekeeping genes used

Primer set	1	2	3	4	5	Number of cycles steps 2-4
gyrB343F gyrB1043R	95°C 5 mins	94°C 2 mins (5 cycles) 30 secs (28 cycles)	57°C 2 mins (5 cycles) 1 min (28 cycles)	72°C 1.5 mins	72°C 5 mins	5 then 28
atpD352F atpD871R	95°C 5 mins	94°C 2 mins (2 cycles) 30 secs (30 cycles)	64.3°C 1 min (2 cycles) 1 min (30 cycles)	72°C 1 min	72°C 5 mins	2 then 30
recA63F recA504R	95°C 5 mins	94°C 45 secs	60°C 1 min	74°C 1.5 min	72°C 5 mins	32
dnaK1466F dnaK1777R	94°C 5 mins	94°C 1 min	62°C 1 min	72°C 40 secs	72°C 5 mins	35
rpoB83F rpoB1540R	95°C 5 mins	94°C 2 mins (3 cycles) 30 secs (30 cycles)	58.2°C 2 mins (3 cycles) 1 min (30 cycles)	72°C 1 min	72°C 5 mins	3 then 30

4.2.4 Housekeeping genes primer design and PCR optimisation

For some isolates and housekeeping genes, single, clear bands were could not be produced, probably due to primer mismatch, or sequences returned were poor quality. Primers were redesigned for *gyrB*, *recA*, *dnaK* and *rpoB* in order to obtain sequences for all isolates and all five housekeeping genes. Sequences for these genes were downloaded for reference strains with complete genomes from the NCBI database (Accession numbers listed in Chapter 3, Table 3.2, numbered 1-7). Existing primer sequences were aligned and new primers were designed which were specific to *Bradyrhizobium*. Redesigned primer sequences are listed in Table 4.3. The original primer sequence was used for *gyrB* reverse primer (listed in Table 4.1). A gradient PCR program was used to determine the optimal annealing temperature for the redesigned primers (Bio-Rad MJ Research PTC-200 thermal cycler). The primer melting temperature is listed in Table 4.3 and as a result of these temperatures a gradient to test all primer sets was chosen to be 60-67°C. Each

PCR reaction mixture was as described in General Methods 2.4 and the PCR conditions were as described in Chapter 3 (Section 3.2.6) with the final elongation step being 5 minutes. PCR products were examined using gel electrophoresis as described in General Methods 2.5.

Table 4.3: Redesigned primer sets for four housekeeping genes

Gene	Primer name	Primer sequence (5'-3')	Primer T_m (^oC)	Size (bp)
gyrB	gyrBF_redes	TTC GAC CAG AAY TCC TAC AAG G	58.8	600-700
recA	recAF_redes	ATC GAG CGB CAG TTY GGC AAG GG	66.3	386-437
	recAR_redes	TTG CGC ARC GCC TGR CTC AT	61.4	
dnaK	dnaKF_redes	AAG GAG CAG CAG ATC CGC ATC CA	64.2	248-350
	dnaKR_redes	TAC ATG GCC TCG CCG AGC TTC AT	64.2	
rpoB	rpoBF_redes	CAG GAG AAG GGS CTG AAG GC	63.5	1150-1385
	rpoBR_redes	AGC TGC GAS GAG CCG AAG	60.5	

4.2.5 PCR amplification and sequencing

A range of annealing temperatures was identified from the gradient PCR program and these temperatures were not the same for all isolates or each housekeeping gene. Annealing temperatures used to generate good quality sequences for each isolate and gene combination tested are listed in Table 4.4. The PCR reaction mixture was as described in General Methods 2.4 and the PCR conditions were as follows: held at 95^{o}C, followed by 30 cycles of 95^{o}C for 15 seconds, chosen annealing temperature for 15 seconds and 72^{o}C for 1 minute. The product was held at 72^{o}C for 5 minutes for elongation of the product and held at 10^{o}C until storage at -20^{o}C. PCR amplicons were examined using agarose gel electrophoresis and purified as described in General Methods 2.5. Sequencing was carried out by MWG Eurofins as described in General Methods 2.6.

Table 4.4: Isolates amplified using redesigned primers and annealing temperature used

Gene	Isolate	Annealing temperature used (°C)
gyrB	G9	60
	W4	62.8
	W24	62.8
recA	G22	60
	BF30	60
dnaK	G46	61.4
	BF28	62.8
	BF41	62.8
	A16	62.8
	W23	62.8
	W74	Never amplified
	W91	64
	GOR4	61
rpoB	BF30	62.8
	W5	62.8
	W74	62.8

4.2.6 Sequence analysis

Sequences were imported in Geneious for editing as described in General Methods 2.6. Reference sequences from the NCBI database were imported and used in the subsequent analysis (Accession numbers listed in Chapter 3, Table 3.2 numbered 1-7). Sequences for types strains identified in the 16S analysis (Section 3.3.1), accession numbers for the sequences used can be found in Table 4.5. Sequences for each gene were aligned using the MUSCLE alignment tool in Geneious using the default options and 8 iterations (Edgar, 2004). Aligned regions of the following sizes for each gene were extracted: *gyrB*: 541 bp, *atpD*: 425 bp, *recA*: 288 bp, *dnaK*: 301 bp, *rpoB*: 533 bp. These sequences were also concatenated in the following order; *gyrB*, *atpD*, *recA*, *dnaK* and *rpoB*. The concatenated sequences were aligned and the 2011 bp aligned region was extracted. The 16S sequences were not included in the concatenated loci phylogeny. Phylogenetic trees were created for each gene individually and for the concatenated loci using the Geneious tree builder with

Tamura-Nei as the genetic distance model, neighbour-joining as the clustering method, 1000 bootstraps and branches were with less than a 70% threshold bootstrap support. *Mesorhizobium loti* MAFF303099 was used as the outgroup (NC_002678.2). Newick files were uploaded into iTOL for visualisation. Sequences were uploaded to the NCBI database, accession numbers can be found in Appendix B.

Table 4.5: Accession numbers for housekeeping genes sequences for type strains used in analysis

Type strain	*gyrB*	*atpD*	*recA*	*dnaK*	*rpoB*
B. canariense LMG 22265	FM253220.1	LM253135.1	FM253177.1	LM253306.1	FM253263.1
B. ottawaense OO99	HQ873179.1	HQ455212.1	HQ587287.1	JF308816.1	HQ587518.1
B. liaoningense LMG 18230	FM253223.1	FM253137.1	FM253180.1	AY923041.1	FM253266.1
B. cytisi CTAW11	KF532653.1	LM994389.1	LM994316.1	LM994142.1	LM994166.1
B. rifense CTAW71	KF962699.1	LM994390.1	LM994317.1	KM994143.1	KF962715.1
B. elkanii USDA 76	AM418800.1	AM418751.1	AY591568.1	AY328392.1	LC167350.1
B. japonicum USDA 6	AP012206.1	AP012206.1	AP012206.1	AP012206.1	AP012206.1

4.3 Results

4.3.1 ERIC DNA fingerprinting profile

ERIC DNA fingerprinting allowed for a comparison of the *Bradyrhizobium* culture collection (Figure 4.1).

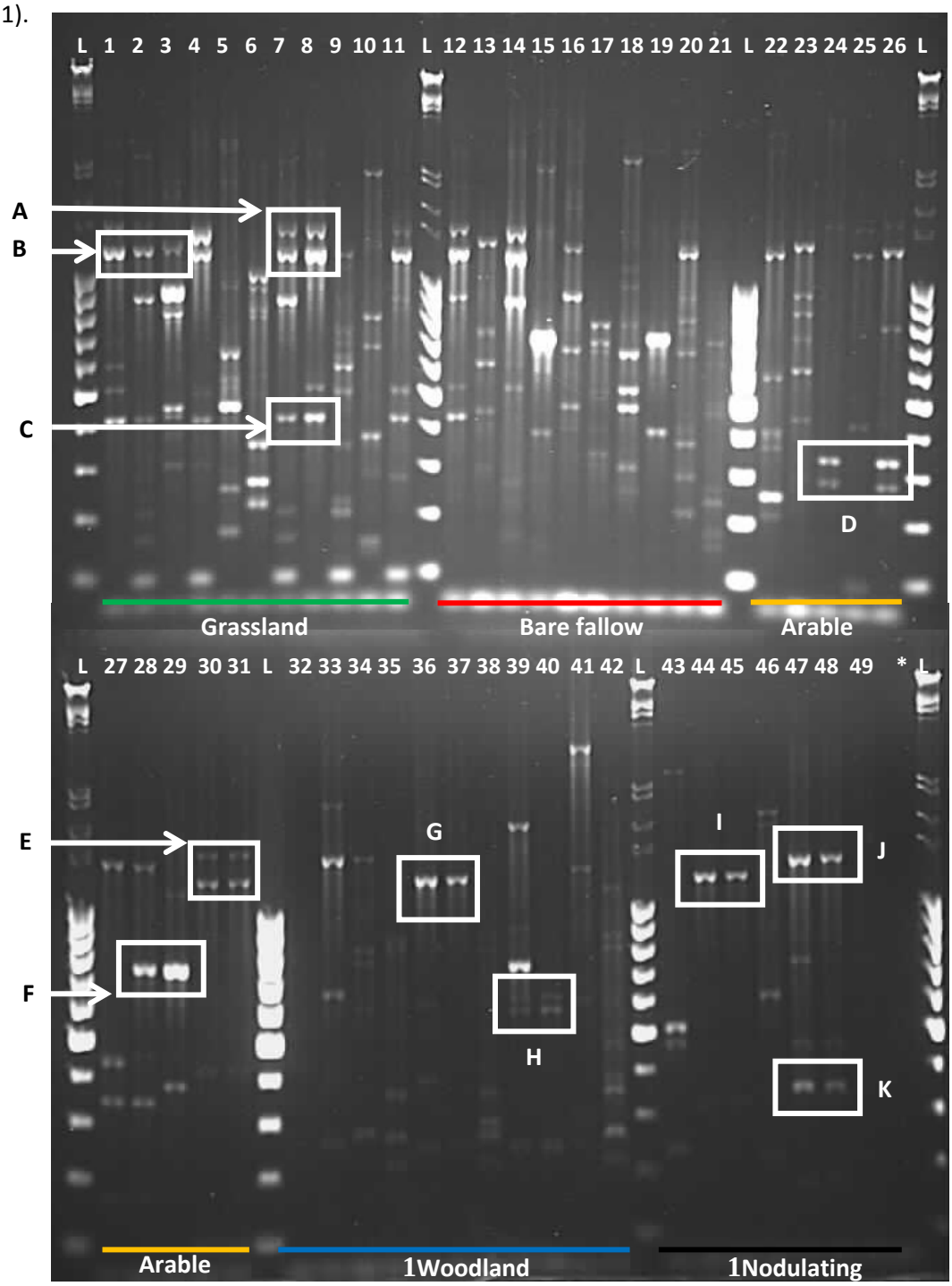


Figure 4.1: ERIC DNA fingerprinting agarose electrophoresis gel of the *Bradyrhizobium* culture collection.

L are the marker lanes; Lamda/*Eco*RI/*Hin*dIII and 100 bp ladder. * indicates a negative control. Numbered lanes correspond to the isolates listed in Appendix C.

The banding profiles shows that all *Bradyrhizobium* isolated from Highfield are different from the nodulating strain (USDA 6- Lane 43); however, there is a high level of diversity within the Highfield samples. The grassland isolates G15 and G22 share bands at the top and middle of their profile (Figure 4.1-A & Figure 4.1-C) suggesting that these isolates are similar. Similarly, grassland isolates G2, G3 and G5 share a cluster of bands at the top of the profile (Figure 4.1-B) suggesting similarity. The bare fallow isolates (Lanes 12-21) have the highest level of diversity with each isolate showing a different banding profile. Within the arable isolates, A2 and A5 share bands at the lower end of their profile (Figure 4.1-D), A19 and A20 share bands at the top of the profile (Figure 4.1-E) and A12 and A16 share bands in the middle of the profile (Figure 4.1-F). The woodland isolates W22 and W23 share a band at the top of their profiles (Figure 4.1-G) and W46 and W49 share a cluster of bands in the centre of their profiles (Figure 4.1-H). The nodulating strains give different profiles to the rest of the *Bradyrhizobium* isolates. The gorse isolates GOR1 and GOR2, share a band at the top of their profiles (Figure 4.1-I). GOR4 gives a different ERIC profile to the other gorse isolates suggesting that this isolate is distinct from the other isolates from gorse nodules. The broom isolates (BRM1 and BRM2) give very similar profiles, sharing bands at the top (Figure 4.1- J) and bottom (Figure 4.1-K) of their profiles.

4.3.2 ERIC cluster analysis

The ERIC sequence dendrogram (Figure 4.2) shows a high level of diversity across the culture collection. The isolates are not split into clusters according to soil treatment and so this suggests that there are different strains of *Bradyrhizobium* in all soil treatments. The broom isolates cluster closely with an approximately unbiased (AU) *P*-value of 91 which suggests that these isolates are similar. They are also part of a larger clade containing woodland isolates W5, W9, W19, W24, W74 and W91 with an AU *P*-value of 86. Symbiotic *B. japonicum* USDA 6 is distinct from all other strains in the culture collection with an AU *P*-value of 98. The gorse isolates GOR1 and GOR2 and the woodland isolates W22 and W23 are identical with an AU *P*-value of 100% and form part of a larger cluster with the arable isolates A4, A19 and A20 with an AU *P*-value of 91%. The woodland isolate W4 and the 2,4-D degrading strain *Bradyrhizobium* sp. RD5-C2 were not included in the dendrogram as the bands were not clear enough to be detected using the CLIQS software.

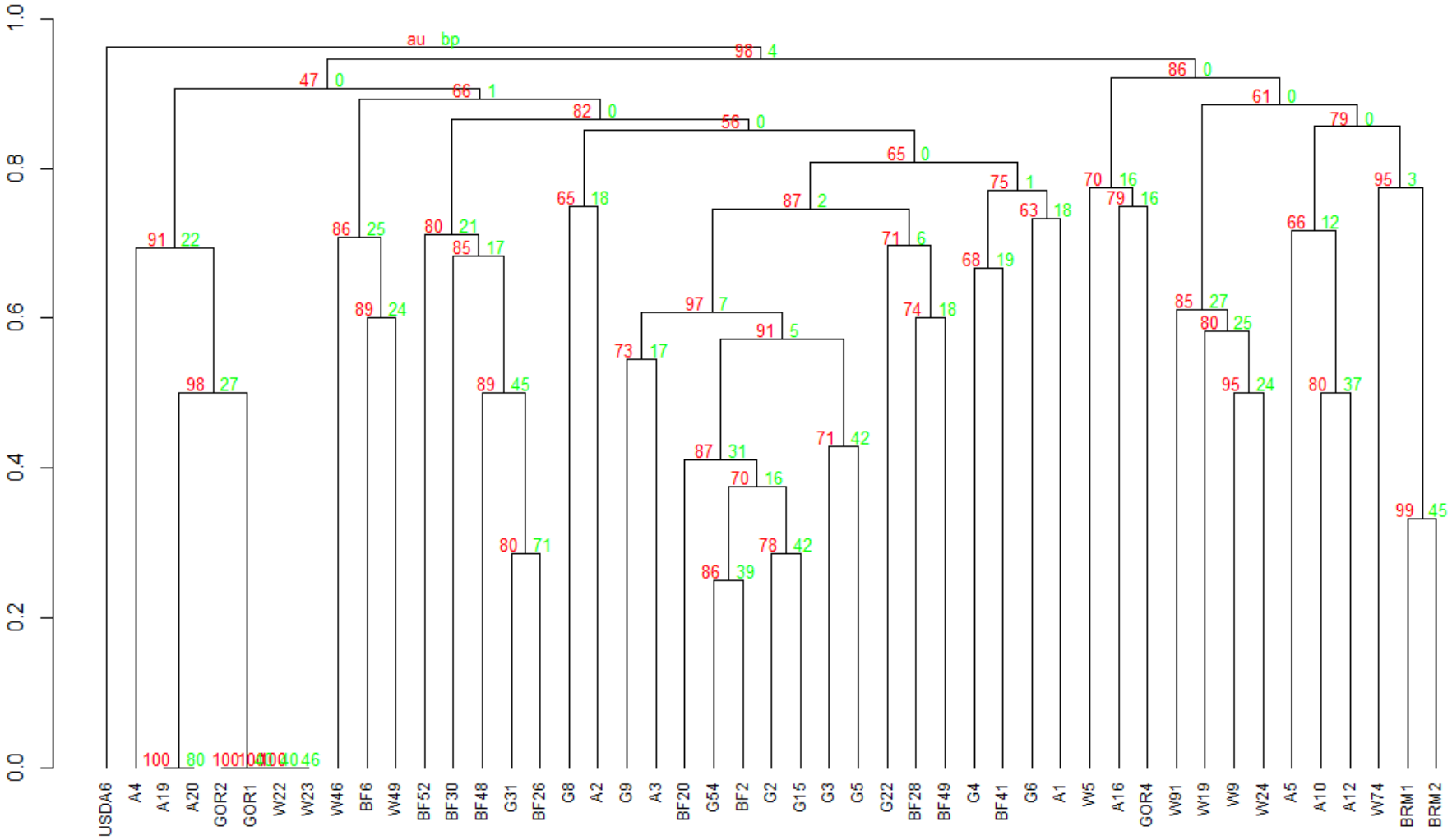


Figure 4.2: ERICS profile cluster analysis using UPGMA clustering method and 1000 bootstraps.
Values in red indicate approximately unbiased (AU) *P*-value generated by multiscale bootstrap resampling and in green, the bootstrap *P*-value (BP), the number of times the cluster is generated in the bootstraps.

4.3.3 Core genome phylogeny

4.3.3.1 DNA gyrase, *gyrB*

The *gyrB* phylogeny (Figure 4.3) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. The woodland isolates W5 and W9 form their own cluster away from both *M. loti* and the other *Bradyrhizobium* strains with a bootstrap of 100%. The woodland isolate W24 clusters closely with the two *B. japonicum* strains used in the analysis (E109 and USDA6T) with a bootstrap of 86.5%. The two broom isolates, BRM1 and BRM2 cluster closely with two arable isolates A10 and A12 with a bootstrap of 86.5%. The *B. canariense* LMG 22265 type strain clusters with four arable isolates (A1, A2, A3 and A5) and two bare fallow isolates (BF6 and BF49) with a bootstrap of 86.5%. All gorse isolates (GOR1, GOR2 and GOR4) group together with a bare fallow isolate BF20 with a bootstrap of 93.4%. The *B. elkanii* USDA 76 type strain clusters closely with *Bradyrhizobium* sp. RD5-C2 which is a strain isolated from soil in Japan with a bootstrap of 86.5%. One grassland isolate, G8 and one bare fallow isolate BF52 cluster closely with the *B. rifense* type strain CTAW71 with a bootstrap of 86.5%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.4 shows the same phylogeny as Figure 4.3 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. This shows that there is a large number groups of *B. erythrophlei*/*B. elkanii* strains separated from other strains in the analysis (blue). These are strains where the top hit could not be distinguished between *B. erythrophlei* and *B. elkanii.* The *B. rifense* and *B. cytisi* groups (red and yellow respectively) are placed closely in the phylogeny and suggest that differences in *gyrB* sequence for these strains are small. The isolates which cluster with the *B. canariense* type strain LMG 22265 (pink) are strains where the top BLAST hit for the 16S sequence were unable to be distinguished between *B. lupini* and *B. canariense*. Isolates where the top BLAST hit was *B. ottawaense*/*B. liaoningense* (grey) are not clustered with the type strains for these species showing differences in the *gyrB* sequence. Many strains where the top BLAST hit was *B. canariense* only are not clustered with this type strain.

Two isolates were removed from the analysis, G5 and W91, as the sequences were from contaminants.

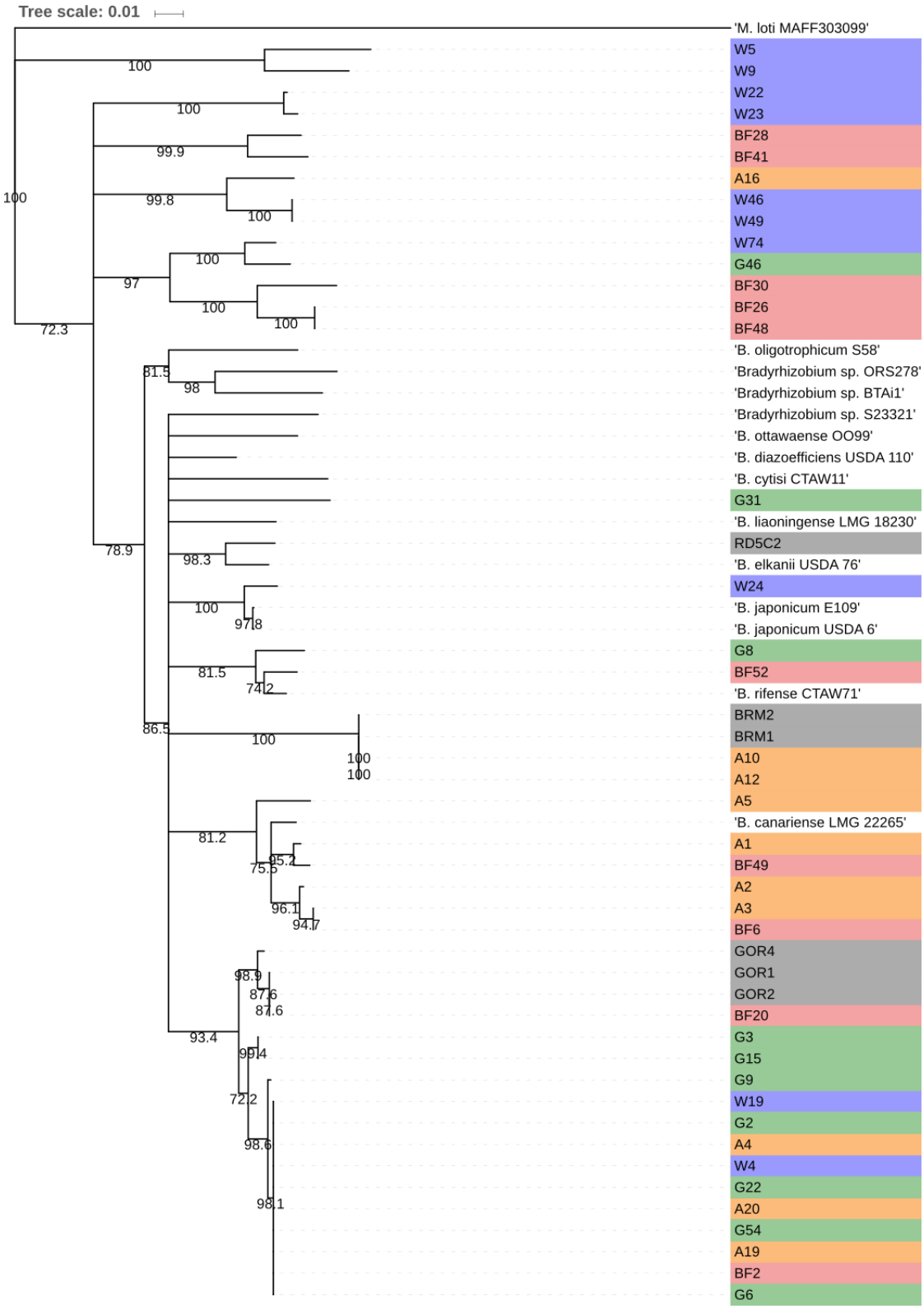


Figure 4.3: *gyrB* sequence phylogeny using Neighbour-Joining and 1000 bootstraps. Isolates are colour coded according to soil treatment: green = grassland, pink = bare fallow, orange = arable, blue = woodland and grey = reference and symbiotic strains.

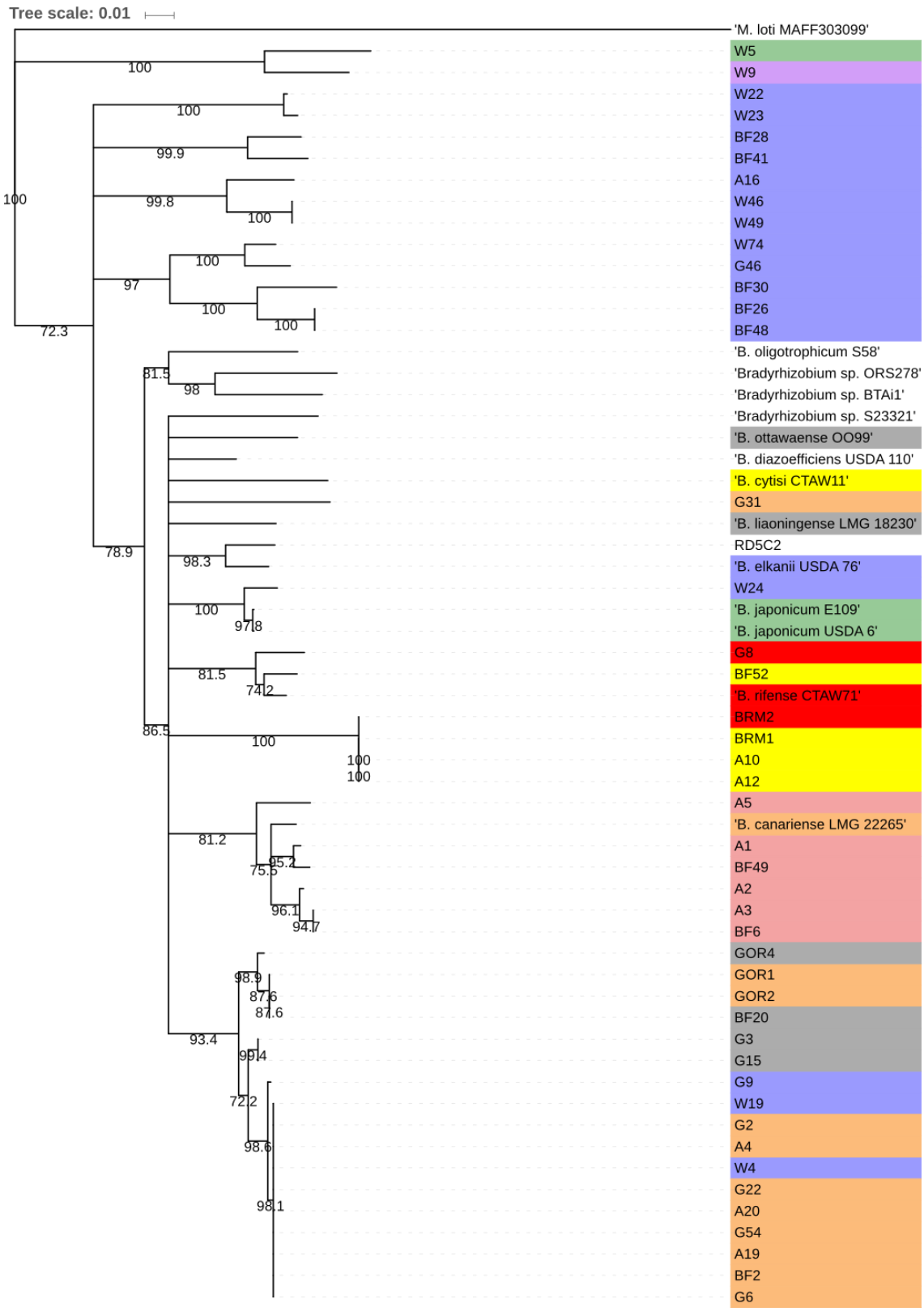


Figure 4.4: *gyrB* sequence phylogeny showing clades corresponding to 16S top BLAST hit. Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense*.

4.3.3.2 ATP synthase, *atpD*

The *atpD* phylogeny (Figure 4.5) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. The woodland isolates W4, W5, W9, W19, W24 and W91 along with the grassland isolate G9 form a separate cluster away from *M. loti* and the other *Bradyrhizobium* isolates with a bootstrap of 100% suggesting that this is a divergent group. The bare fallow isolate BF52 clusters closely with the *B. rifense* type strain CTAW71 with a bootstrap of 72.4%. Similar to the *gyrB* phylogeny, the broom isolates (BRM1 and BRM2) cluster closely with two arable isolates (A10 and A12) with a bootstrap of 72.4%. No isolates cluster with the two *B. japonicum* strains included in the analysis (E109 and USDA6T) with a bootstrap of 71.8%. The three gorse isolates (GOR1, GOR2 and GOR4) are part of a large group of isolates from grassland (G2, G3, G6, G15, G22, G31 and G54), bare fallow (BF2 and BF20) and arable (A4, A19 and A20) with a bootstrap of 71.8%. No type strains are in this cluster. The arable isolate A3 and the bare fallow isolate BF6 cluster with the *B. canariense* type strain LMG 22265 with a bootstrap of 86.9%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.6 shows the same phylogeny as Figure 4.5 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. This shows that there is a large number of *B. erythrophlei*/*B. elkanii* strains separated from other strains in the analysis (blue). These are strains where the top hit could not be distinguished between *B. erythrophlei* and *B. elkanii.* The divergent group of *Bradyrhizobium* isolates (W4, W5, W9, W19, W24, W91 and G9) is made up of isolates where the top BLAST hit was *B. neotropicale* (purple), *B. japonicum* (green) and *B. erythrophlei*/*B. elkanii* (blue). The isolates which cluster with the *B. canariense* type strain LMG 22265 (pink) were strains where the top BLAST hit for the 16S sequence were unable to be distinguished between *B. lupini* and *B. canariense*. Isolates where the top BLAST hit was *B. ottawaense*/*B. liaoningense* (grey) are not clustered with the type strains for these species showing differences in the *atpD* sequence. The large clade made up of grassland, arable, bare fallow and gorse isolates are strains where the top BLAST hit was *B. canariense* only (orange) or *B. ottawaense*/*B. liaoningense* (grey), however no type strains are in this cluster.

Two isolates were removed from the analysis, G5 and BF30, as the sequences were from contaminants.

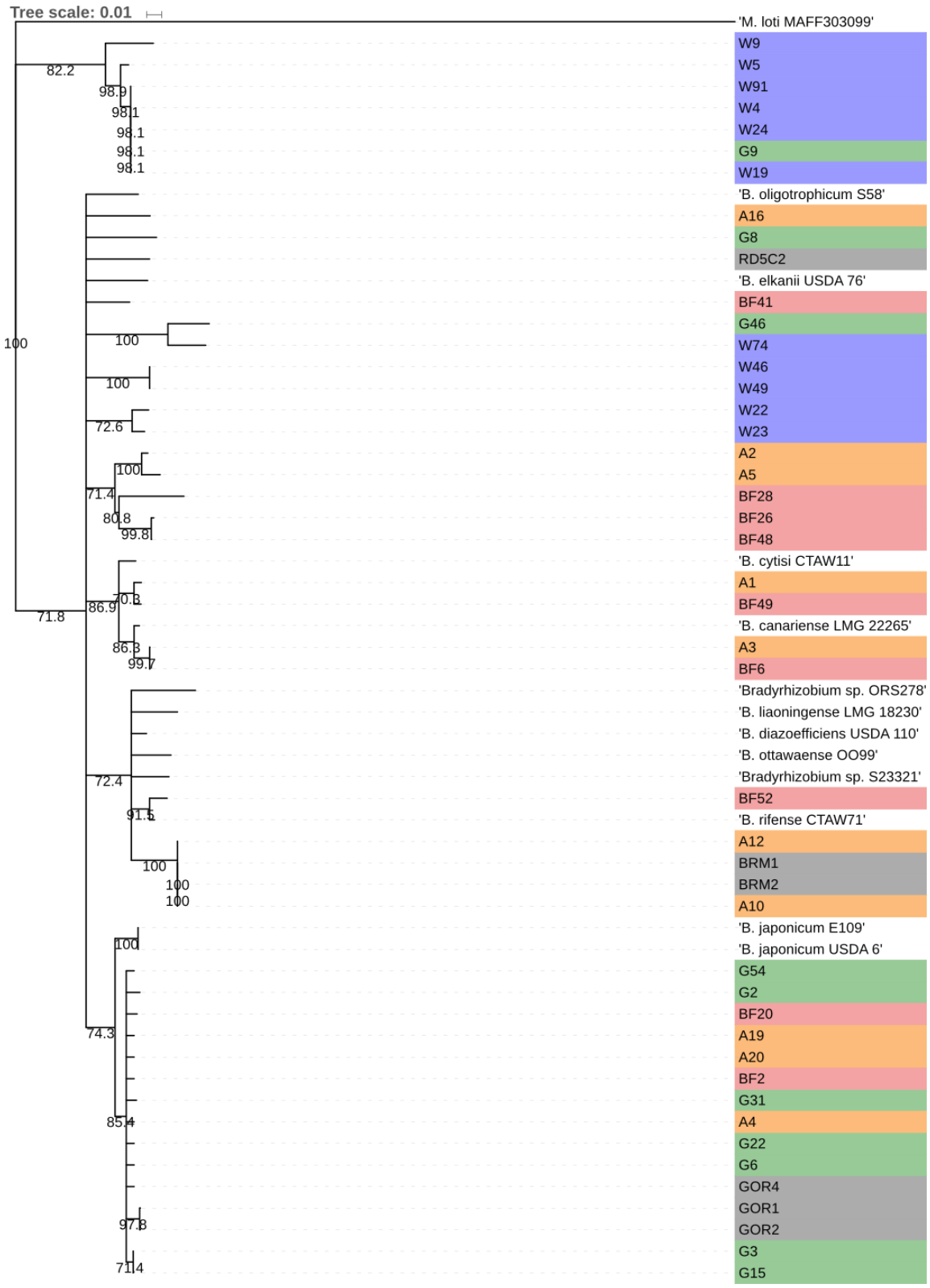


Figure 4.5: *atpD* sequence phylogeny using Neighbour-Joining and 1000 bootstraps. Isolates are colour coded according to soil treatment: green = grassland, pink = bare fallow, orange = arable, blue = woodland and grey = reference and symbiotic strains.

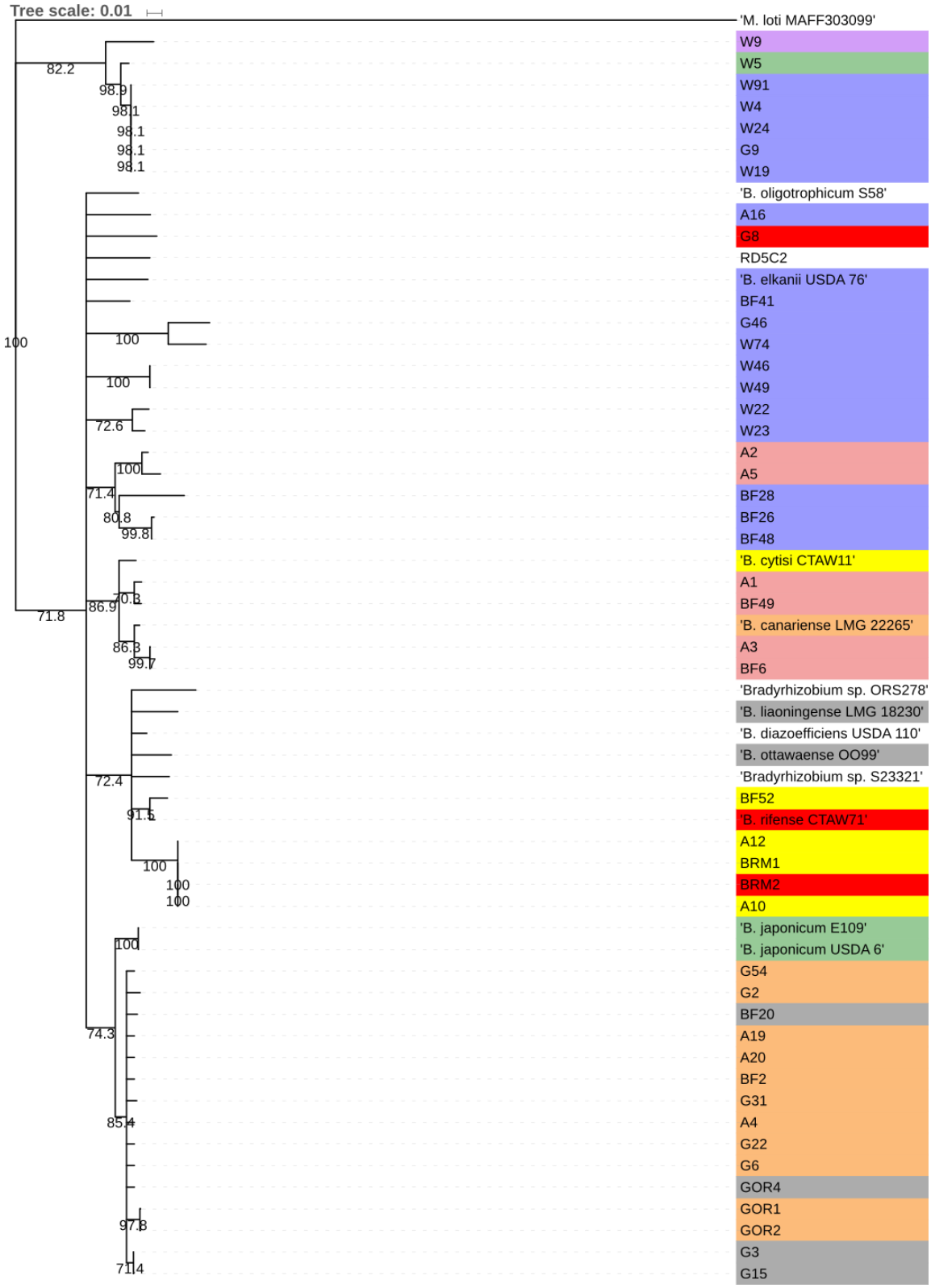


Figure 4.6: *atpD* sequence phylogeny showing clades corresponding to 16S top BLAST hit. Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense*.

4.3.3.3 Recombinase A, *recA*

The *recA* phylogeny (Figure 4.7) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100% however the divergence of *M. loti* is from the *Bradyrhizobium* isolates in the analysis is low. The resolution is poor for this gene as many of the isolates are forming individual clades or with a small number of isolates and the branching occurs at the base of the tree. The bare fallow isolate clusters closely with the *B. cytisi* type strain CTAW11 with a bootstrap of 100%. Similar to the *gyrB* and *atpD* phylogenies, the broom isolates (BRM1 and BRM2) cluster closely with two arable isolates A10 and A12 with a bootstrap of 100%. The divergent *Bradyrhizobium* clade consisting of woodland isolates (W4, W5, W9, W19, W24 and W91) and one grassland isolate (G9) which was highlighted in the *atpD* phylogeny is evident in the *recA* phylogeny with a bootstrap of 100%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.8 shows the same phylogeny as Figure 4.7 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. The divergent group of *Bradyrhizobium* isolates (W4, W5, W9, W19, W24, W91 and G9) is made up of isolates where the top BLAST hit was *B. neotropicale* (purple), *B. japonicum* (green) and *B. erythrophlei*/*B. elkanii* (blue). The latter are strains where the top hit could not be distinguished between *B. erythrophlei* and *B. elkanii.* Isolates where the top BLAST hit was *B. ottawaense*/*B. liaoningense* (grey) or *B. canariense* only (orange) are not clustered with the type strains for these species showing differences in the *recA* sequence however the two *B. japonicum* (green) strains included in the analysis (E109 and USDA6T) are part of this group.

Four isolates were removed from the analysis, G5, G46 and BF30, as the sequences were from contaminants and G22 as this sequence was low quality.

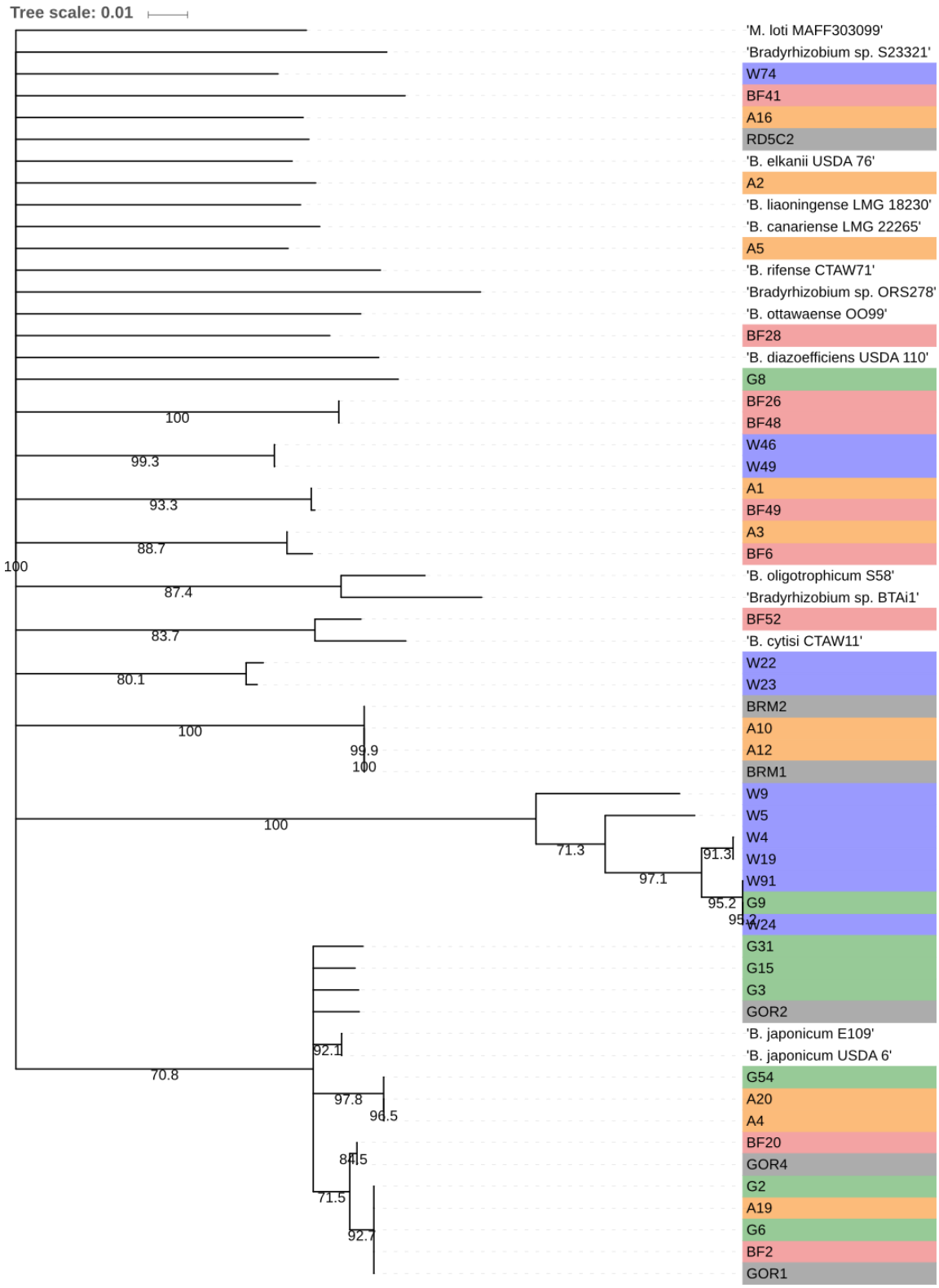


Figure 4.7: *recA* sequence phylogeny using Neighbour-Joining and 1000 bootstraps. Isolates are colour coded according to soil treatment: green = grassland, pink = bare fallow, orange = arable, blue = woodland and grey = reference and symbiotic strains.

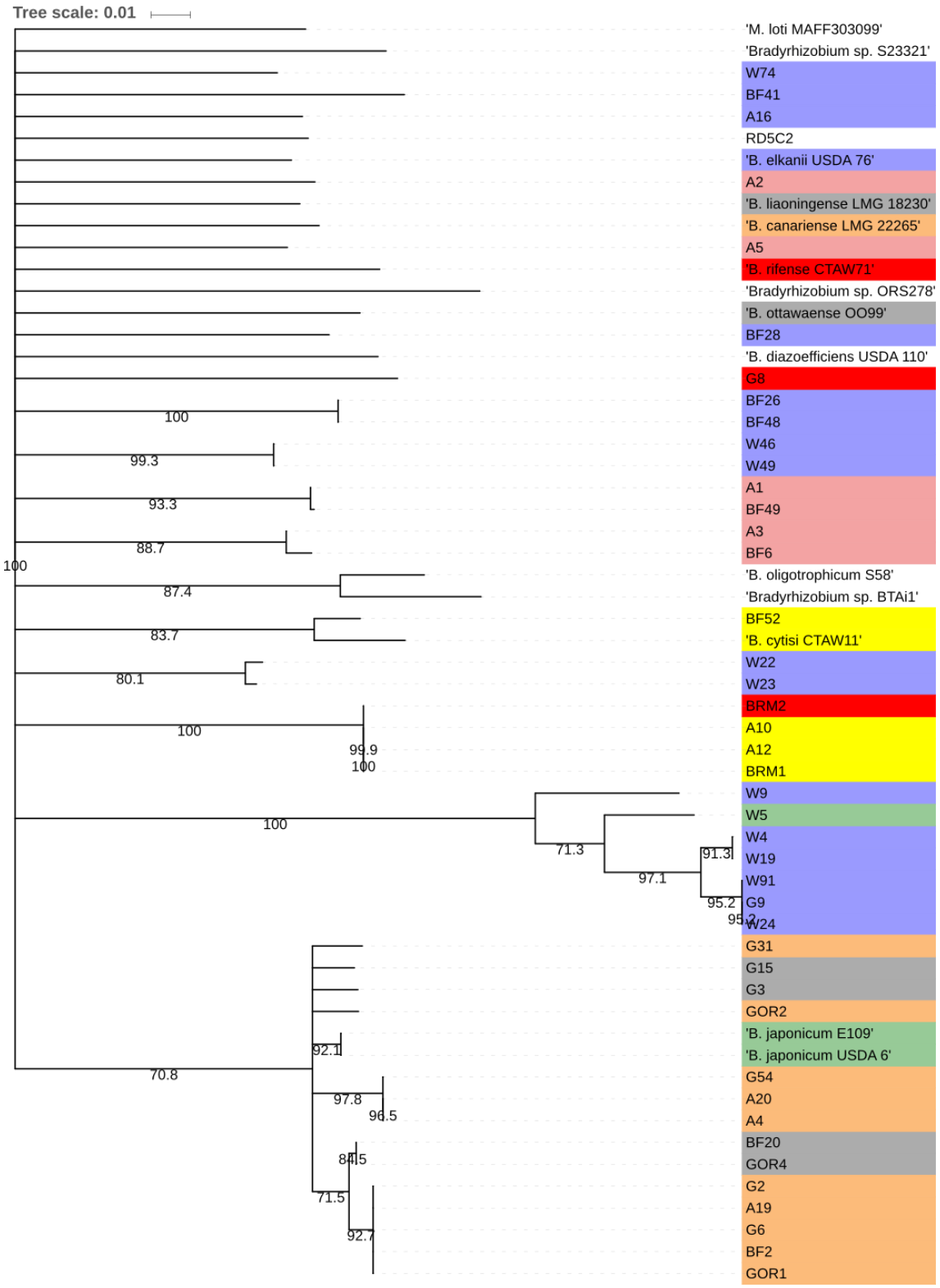


Figure 4.8: *recA* sequence phylogeny showing clades corresponding to 16S top BLAST hit. Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense*.

4.3.3.4 Chaperone protein, *dnaK*

The *dnaK* phylogeny (Figure 4.9) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. There is high divergence within *Bradyrhizobium* and the resolution is poor when compared to *gyrB* and *atpD*. Many of the isolates are forming individual clades or with a small number of isolates and the branching occurs at the base of the tree. The same divergent group formed for woodland and grassland isolates (W4, W5, W9, W19, W22, W24 and G9) as highlighted in the *atpD* and *recA* phylogenies is present in the *dnaK* phylogeny with a bootstrap of 100%. The *B. elkanii* type strain USDA 76 clusters with the *Bradyrhizobium* sp. RD5-C2 strain similar to the *gyrB* phylogeny with a bootstrap of 100%. The *B. cytisi* and *B. rifense* type strains CTAW11 and CTAW71 respectively, cluster with the broom isolates BRM1 and BRM2, arable isolates A10 and A12, a grassland isolate G8 and a bare fallow isolate BF52 with a bootstrap of 100%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.10 shows the same phylogeny as Figure 4.9 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. The divergent group of *Bradyrhizobium* isolates (W4, W5, W9, W19, W22, W24, W91 and G9) is made up of isolates where the top BLAST hit was *B. neotropicale* (purple), *B. japonicum* (green) and *B. erythrophlei*/*B. elkanii* (blue). The latter are strains where the top hit could not be distinguished between *B. erythrophlei* and *B. elkanii.* The clade containing the *B. cytisi* (yellow) and *B. rifense* (red) type strains CTAW11 and CTAW71 respectively is formed of isolates where the top BLAST hit was either *B. cytisi* or *B. rifense* showing that distinction between these species for *dnaK* is small.

Three isolates were removed from the analysis, G5, BF30 and GOR4, as the sequences were from contaminants. For one isolate, W74, *dnaK* was never able to be amplified.

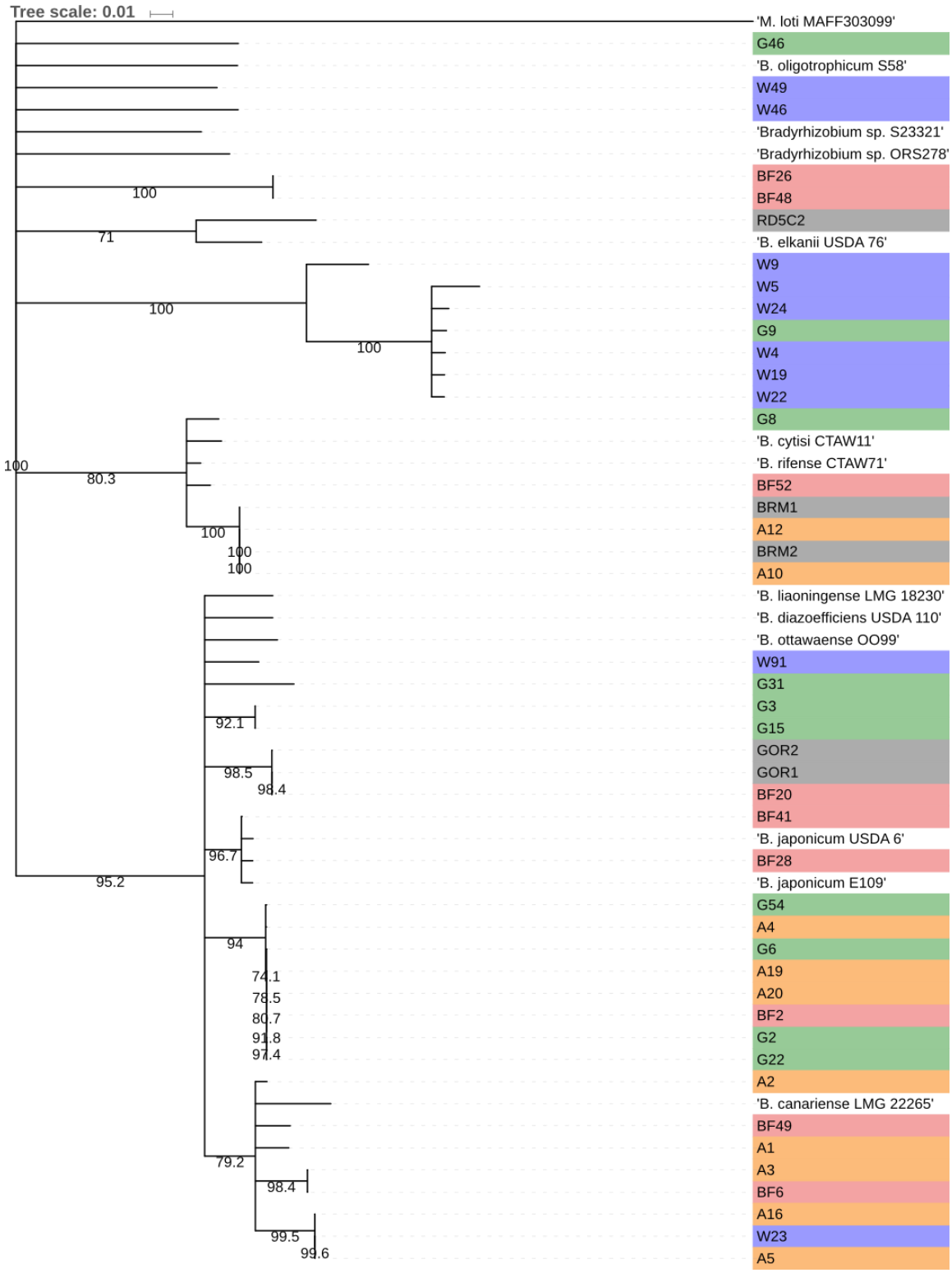


Figure 4.9: *dnaK* sequence phylogeny using Neighbour-Joining and 1000 bootstraps. Isolates are colour coded according to soil treatment: green = grassland, pink = bare fallow, orange = arable, blue = woodland and grey = reference and symbiotic strains.

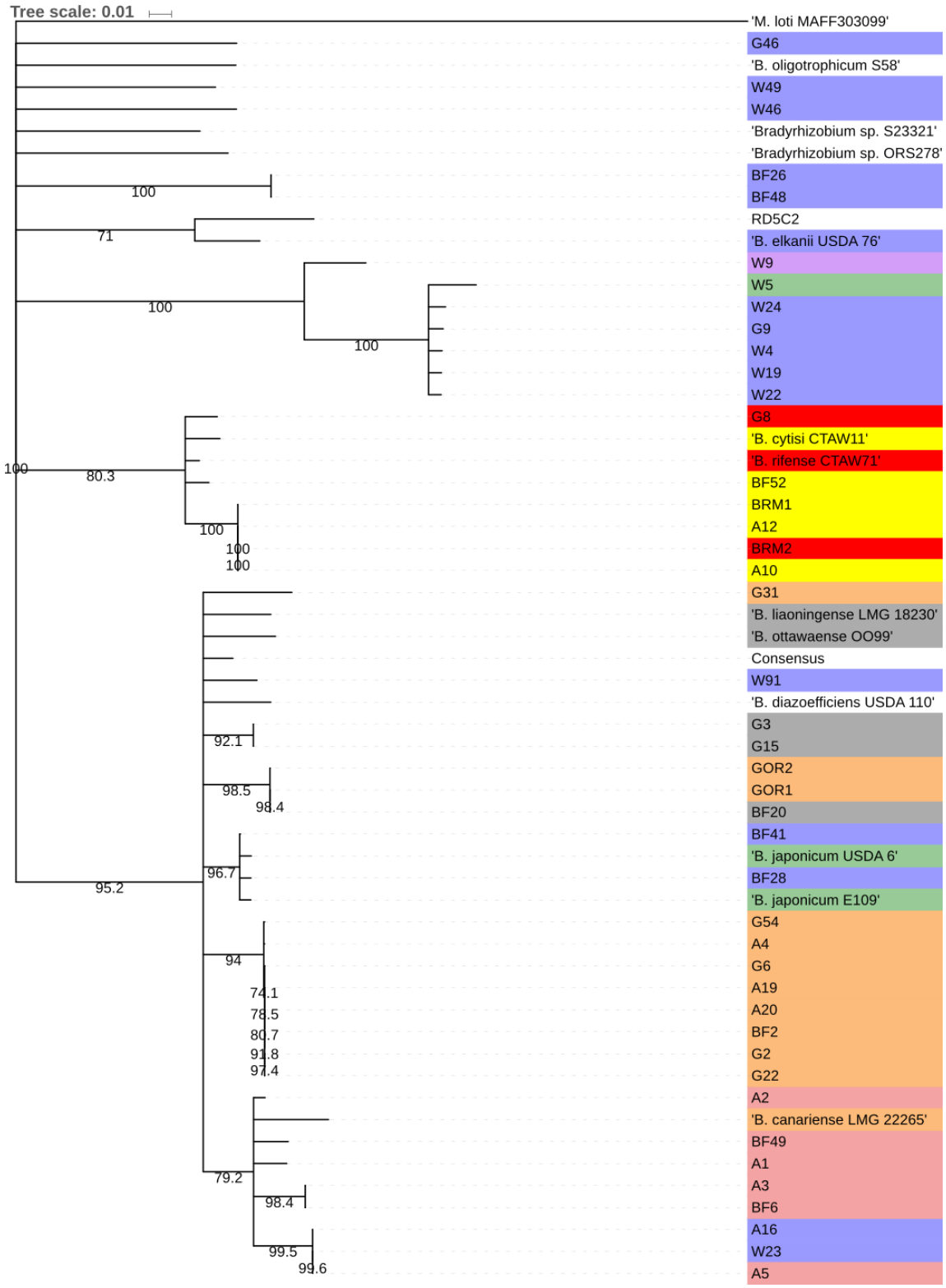


Figure 4.10: *dnaK* sequence phylogeny showing clades corresponding to 16S top BLAST hit. Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense*.

4.3.3.5 RNA polymerase, *rpoB*

The *rpoB* phylogeny (Figure 4.11) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. There are many isolates which form individual clades and branch at the bare of the tree but *M. loti* is the most divergent strain. The *B. cytisi* and *B. rifense* type strains CTAW11 and CTAW71 respectively, cluster with the grassland isolate G8 and the bare fallow isolate BF52 with a bootstrap of 100%. The *B. canariense* type strain LMG 22265 clusters with a number of arable isolates (A1, A2, A3 and A5) and two bare fallow isolates BF6 and BF49 with a bootstrap of 100%. Similar to the other phylogenies, the broom isolates (BRM1 and BRM2) cluster with two arable strains A10 and A12 with a bootstrap of 100%. The group of woodland and grassland isolates (W4, W9, W19, W24, W91 and G9) highlighted in previous phylogenies are the most divergent *Bradyrhizobium* cluster with a bootstrap of 100%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.12 shows the same phylogeny as Figure 4.11 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. The most divergent group of *Bradyrhizobium* isolates (W4, W9, W19, W24, W91 and G9) is made up of isolates where the top BLAST hit was *B. neotropicale* (purple) and *B. erythrophlei*/*B. elkanii* (blue). The latter are strains where the top hit could not be distinguished between *B. erythrophlei* and *B. elkanii.* The clade containing the *B. cytisi* and *B. rifense* type strains CTAW11 and CTAW71 respectively is formed of isolates where the top BLAST hit was either *B. cytisi* or *B. rifense* showing that distinction between these species for *rpoB* is small. The *B. canariense* type strain LMG 22265 clustered closely with strains where the top BLAST hit was unable to be distinguished between *B. lupini* and *B. canariense.* The *B. ottawaense* OO99 and *B. liaoningense* LMG 18230 type strains formed individual clades with a bootstrap of 100% and did not cluster with isolates where the top BLAST hits were identified as these species. The *B. elkanii* type strain USDA 76 clustered with a number of isolates from different soil treatments (BF28, BF41, W22, W23, W46, W49, A16 and *Bradyrhizobium* sp. RD5-C2) but where the top BLAST hits were unable to be distinguished from *B. erythrophlei* and *B. elkanii.* This suggests that the *rpoB* sequences for these strains is very similar to *B. elkanii.*

One isolate was removed from the analysis, G5, as the sequence was from a contaminant. For three isolates, BF30, W5 and W74, *rpoB* was never able to be amplified or good quality sequence obtained.

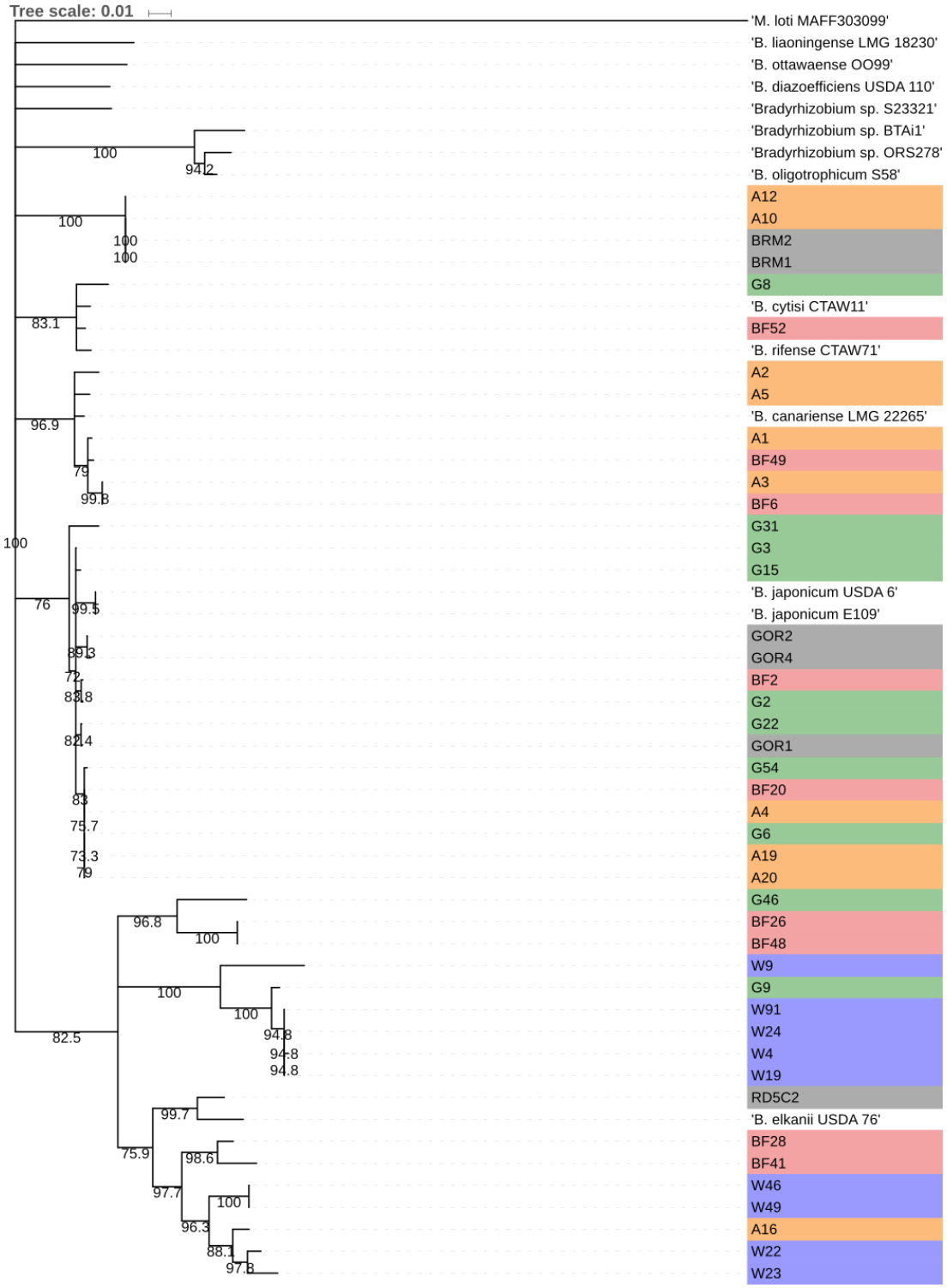


Figure 4.11: *rpoB* sequence phylogeny using Neighbour-Joining and 1000 bootstraps. Isolates are colour coded according to soil treatment: green = grassland, pink = bare fallow, orange = arable, blue = woodland and grey = reference and symbiotic strains.

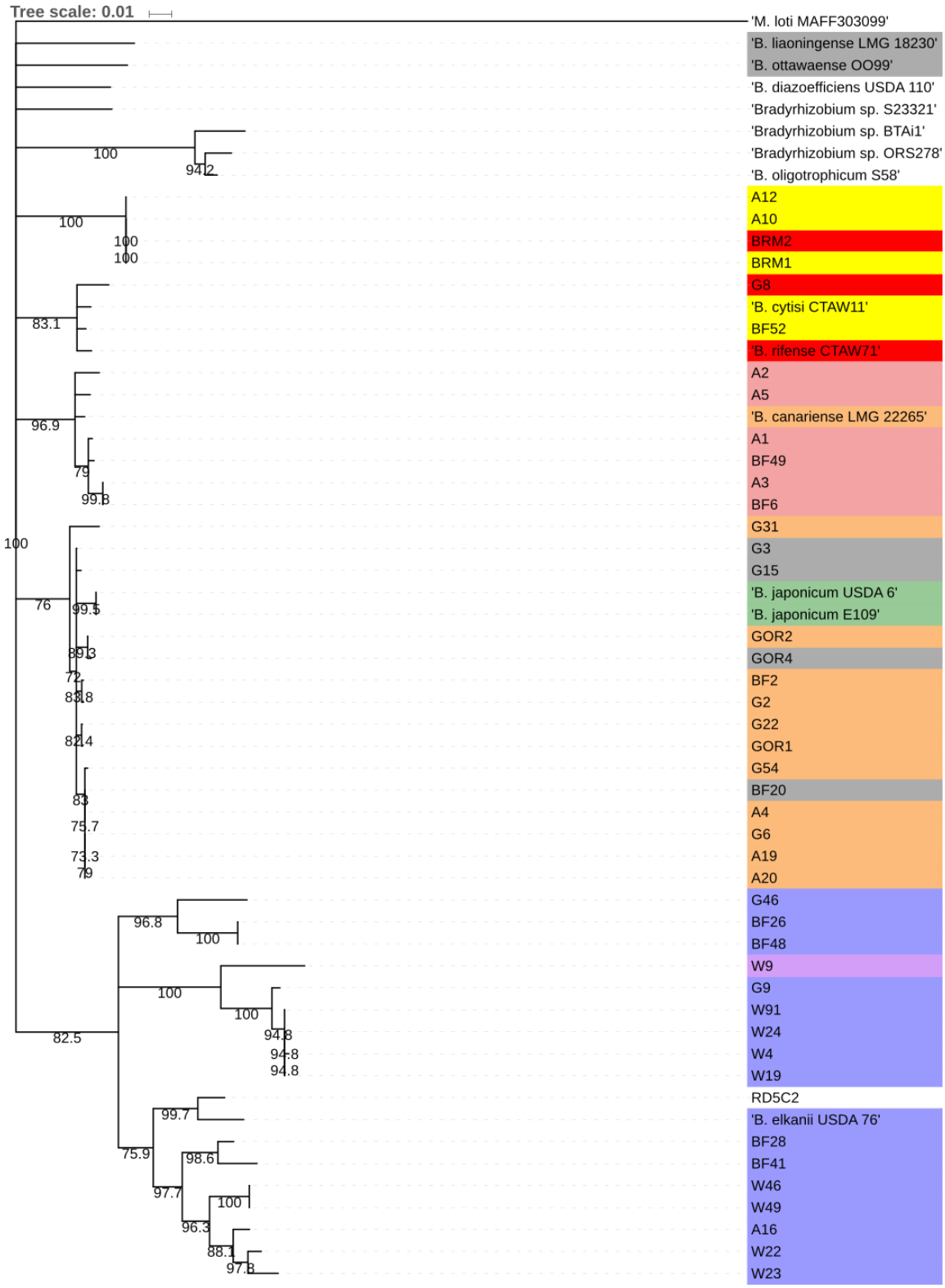


Figure 4.12: *rpoB* sequence phylogeny showing clades corresponding to 16S top BLAST hit. Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini*/*B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei*/*B. elkanii,* grey = *B. ottawaense*/*B. liaoningense*.

4.3.3.6 Concatenated loci

The concatenated loci phylogeny (Figure 4.13) was rooted using *Mesorhizobium loti* MAFF303099 as the outgroup with a bootstrap of 100%. The *B. elkanii* type strain USDA 76 clusters closely with the *Bradyrhizobium* sp. RD5-C2 strain which was isolated in Japan with a bootstrap of 72.6%. The most divergent *Bradyrhizobium* group is made up of woodland (W4, W9, W19 and W24) and grassland (G9) isolates with a bootstrap of 72.9%. The *B. cytisi* and *B. rifense* type strains CTAW11 and CTAW71 respectively, are clustered with the grassland isolate G8 and the bare fallow isolates BF52 with a bootstrap of 81.2%. The *B. canariense* type strain LMG 22265 clusters with two arable isolates (A1 and A3) and two bare fallow isolates (BF6 and BF49) with a bootstrap of 100%. The *B. liaoningense* type strain LMG 18230 and the *B. ottawaense* type strain OO99 cluster closely with a bootstrap of 99.5%. There is a group of woodland (W22, W23, W46 and W49), bare fallow (BF28 and BF41) and arable (A16) isolates which form a cluster with a bootstrap of 72.6%. Overall, there does not appear to be clustering according to which soil treatment the isolate were taken from.

Figure 4.14 shows the same phylogeny as Figure 4.13 but is colour coded according to top hit when the 16S sequence was submitted to the NCBI database. The most divergent group of *Bradyrhizobium* isolates (W4, W19, W24 and G9) is made up of isolates where the top BLAST hit was *B. erythrophlei*/*B. elkanii* (blue). W9 is related to this group but is more divergent, this was also the only isolate in this group where the top BLAST hit was *B. neotropicale.* There are no type strains in this cluster and they could represent a new *Bradyrhizobium* species. The clade containing the *B. cytisi* and *B. rifense* type strains CTAW11 and CTAW71 respectively is formed of isolates where the top BLAST hit was either *B. cytisi* or *B. rifense* showing that distinction between these species for all five housekeeping genes is small. The *B. canariense* type strain LMG 22265 clustered closely with strains where the top BLAST hit was unable to be distinguished between *B. lupini* and *B. canariense.* The *B. ottawaense* OO99 and *B. liaoningense* LMG 18230 type strains form a cluster with a bootstrap of 99.5% and did not cluster with isolates where the top BLAST hits were identified as these species. The *B. elkanii* type strain USDA 76 clustered with *Bradyrhizobium* sp. RD5-C2 but not with isolates where the top BLAST hits were unable to be distinguished from *B. erythrophlei* and *B. elkanii.* There is a large group of isolates (W22, W23, W46, W49, BF28, BF41 and A16) where the top BLAST hit was *B. erythrophlei*/*B. elkanii* however this group did not include any type strains and so this could represent a new *Bradyrhizobium* species. There is also a clade of a number of isolates with a bootstrap of 99.5% where the top BLAST hit for the 16S sequence was *B. canariense* or *B. ottawaense*/*B. liaoningense* however these were clustered most closely with the two *B. japonicum* strains included in the analysis, E109

and USDA6T. These are unlikely to represent a new *Bradyrhizobium* species due to the *B. japonicum* USDA 6 type strain clustering closely.

A number of isolates were not included in the concatenated loci phylogeny as sequences for all five genes were not obtained. These are: G5, G22, G46, BF30, W5, W74, W91 and GOR4.

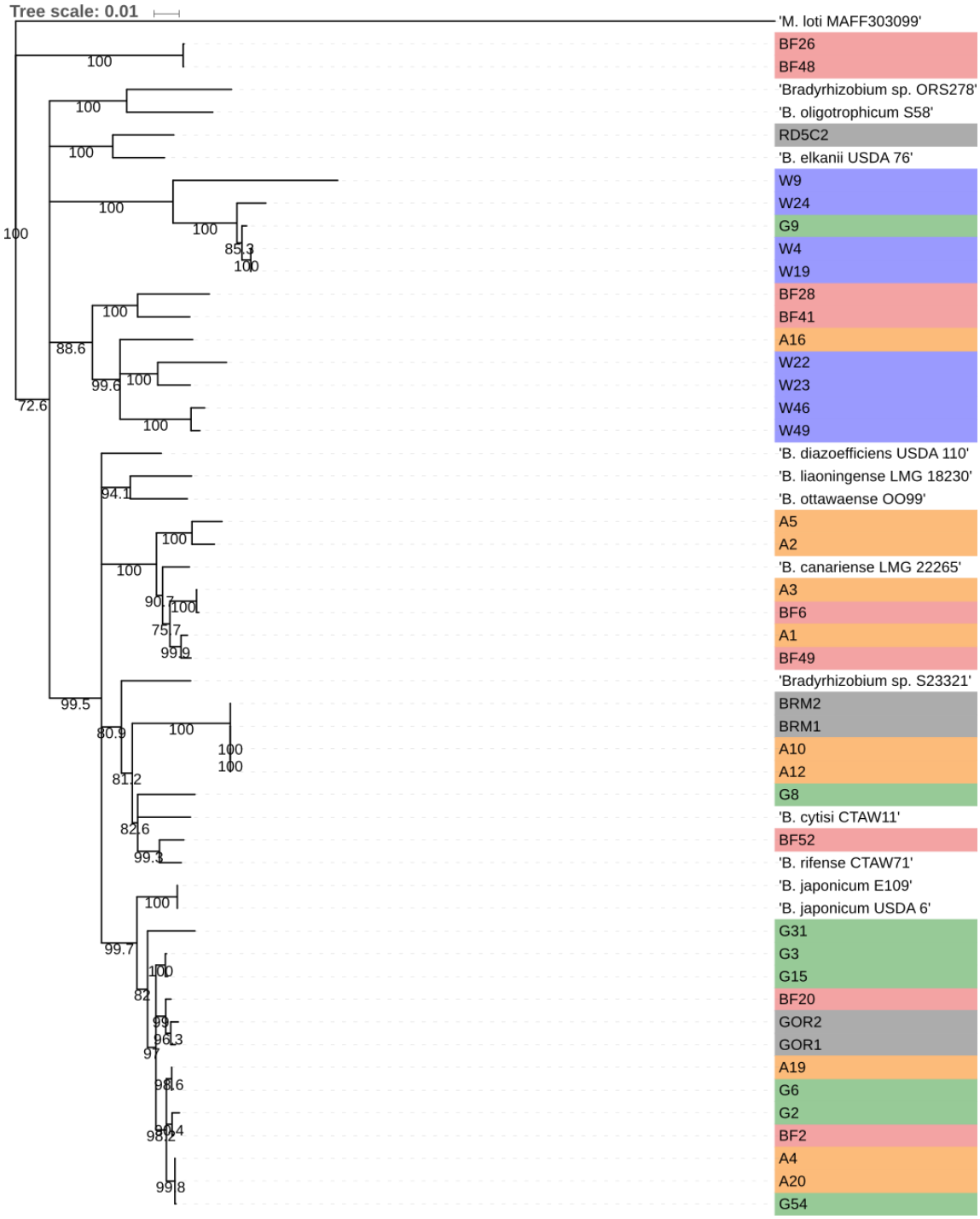


Figure 4.13: Concatenated loci phylogeny using Neighbour-Joining and 1000 bootstraps.

Isolates are colour coded by soil treatment: green = grassland, red = bare fallow, orange = arable, blue = woodland and grey=reference/symbiotic strains.

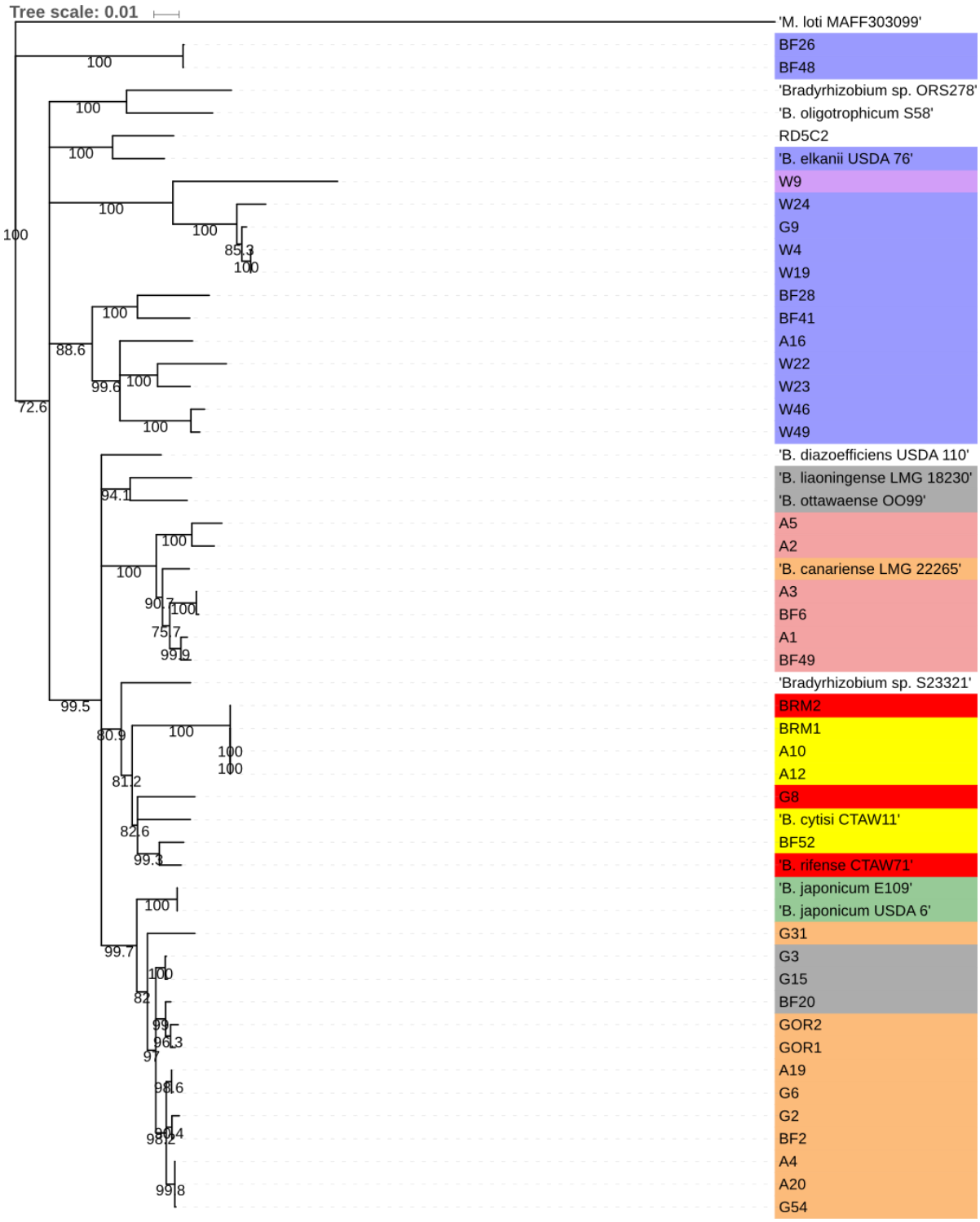


Figure 4.14: Concatenated loci phylogeny showing clades corresponding to 16S top BLAST hit.

Isolates are colour coded according to top hit when the 16S sequence was submitted to the NCBI database. Purple = *B. neotropicale,* green = *B. japonicum,* pink = *B. lupini/B. canariense,* orange = *B. canariense* only, yellow = *B. cytisi,* red = *B. rifense,* blue = *B. erythrophlei/B. elkanii,* grey = *B. ottawaense/B. liaoningense*.

4.4 Discussion

4.4.1 ERIC sequence DNA fingerprinting

The ERIC sequence profiles show that there is considerable diversity across the entire *Bradyrhizobium* culture collection. None of the isolates give the same profile which suggests that there is a high genetic diversity in this genus. This has been previously reported in *Bradyrhizobium* (Giongo et al., 2008). The cluster analysis did not group all isolates from the same soil treatment together which suggests that diversity is not dependent on soil treatment. The broom isolates were clustered together suggesting there is a lower diversity within broom nodules. Two gorse isolates are also clustered closely and in the same clade as the broom isolates. However, GOR4 was in a separate clade suggesting that there is still potential for a high level of diversity within the symbiotic isolates as seen with the free-living isolates. The ERIC dendrogram also showed that the *B. japonicum* USDA 6 type strain was distinct from all other strains in the Rothamsted culture collection and this could reflect the geographical variation between these isolates.

4.4.2 Limitations of ERIC sequence DNA fingerprinting

The ERIC sequence DNA fingerprinting technique requires the user to determine what resolution should be used to assess the bands. This can lead to overestimation of the number of bands in the profile and to an overestimation in the level of diversity. Band size is estimated using markers however, this could result in bands of different sizes being classed as the same reducing the level of diversity seen. Also, the ERIC-PCR amplicons are not sequenced and therefore it is not possible to include other strains from the database without having the DNA of the strain to carry out the ERIC-PCR. However, ERIC sequences represent both core and accessory genomic DNA so in theory allow a more comprehensive assessment of diversity than methods based on a single gene.

4.4.3 Core genome diversity

Assessing protein encoding genes is useful to give insights of recombination occurring within the population (Vinuesa et al., 2008). A high level of genetic diversity has also been noted in *atpD* gene sequence and it was concluded that this could have been due to recombination and this should be taken into account when looking at *atpD* diversity in *Bradyrhizobium* (Menna et al., 2009). The level of diversity for our culture collection was also found to be high and many clades containing single or pairs of isolates were generated. Previous studies have shown that the

Bradyrhizobium genus has a high level of diversity and estimates of this are increasing with each phylogenetic study taking place (Menna et al., 2009, Menna and Hungria, 2011). In a previous study into the diversity of *Bradyrhizobium*, the *dnaK* gene was found to be most informative of all five genes looked at (*gyrB*, *atpD*, *recA*, *dnaK* and *rpoB*) (Rivas et al., 2009). This is in contrast to our results showing that both *recA* and *dnaK* had low resolution and *gyrB*, *atpD* and *rpoB* were the most informative. Moulin et al. (2004) used *dnaK* along with 16S rRNA gene for assessing taxonomy and a low level of diversity has been seen in Zhang et al. (2014b). A high level of sequence diversity was seen in our *dnaK* phylogeny and there were fewer clades according to 16S top BLAST hit than *rpoB*. However, one woodland isolate W74 was unable to be amplified and could suggest this isolate has high variation in the sequence. The sequence similarity of *atpD* and *recA* have previously been reported to be high (Zhang et al., 2014b) which supports the results seen in the phylogenies of the *Bradyrhizobium* culture collection here. The phylogenies of *atpD*, *glnII*, *recA* and *rpoB* of symbiotic *Bradyrhizobium* have previously shown that they group isolates according to 16S rRNA gene sequence (Vinuesa et al., 2008). This was seen in the concatenated loci phylogeny where many isolates were in clades containing isolates with the same top BLAST hit from the 16S sequences. Differences between phylogenies generated from my culture collection and previously published work could be due to the focus on the literature being on symbiotic isolate collections. Recently, sequence analysis of two housekeeping genes (*glnII* and *recA*) showed that symbiotic and non-symbiotic *Bradyrhizobium* were diverse but contained only a few phylotypes (Hollowell et al., 2016a).

Bradyrhizobium is thought to be a monophyletic group and photosynthetic isolates have previously been found to be the most basal group suggesting that non-photosynthetic strains are descended from these photosynthetic individuals (Steenkamp et al., 2008). In contrast, the concatenated loci tree placed bare fallow isolates at the base. The high diversity of the housekeeping genes suggests the potential for there to be high diversity in other functions which are not part of the "core genome". The isolates do not group according to soil treatment but to 16S top BLAST hit when looking at the housekeeping genes. Any differences seen in *Bradyrhizobium* isolates from different soil treatments might only possible from functions which are not essential to the bacteria. Housekeeping genes give an insight into the level of diversity however; it is the functions linked to their niche which would give clues to how the environment influences bacterial communities.

5 Genome analysis of two free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils

This work was published in Jones et al. (2016) included in Appendix T.

5.1 Introduction

5.1.1 Importance of genome sequencing

Genome assembly can give an insight to bacterial evolution and into the environment in which the isolate is found, advancing both biomedical and environmental microbiology (Hernandez et al., 2008, Loman et al., 2012). The environment can drive genome streamlining and so the genomes can provide an insight into the processes needed for survival (Giovannoni et al., 2005). Sequencing genomes can give information about novel processes, for example those not traditionally thought to be associated with the strains in question, and link genotype and phenotype giving the ability to understand processes which are of interest to biotechnology (Tritt et al., 2012). The knowledge gained from the genomes can be used to formulate hypotheses, direct future experiments and research.

5.1.2 Nitrogen metabolism in *Bradyrhizobium*

The slow-growing bacterial genus *Bradyrhizobium* has been shown to be one of the most abundant groups in soil (Uroz et al., 2010, VanInsberghe et al., 2015) including soils sampled from long-term field experiments at Rothamsted Research, UK (Delmont et al., 2012, Zhalnina et al., 2013). A key characteristic of the order Rhizobiales including the genus *Bradyrhizobium* is the ability to form nitrogen-fixing symbioses with legumes to increase nitrogen availability to plants (van Rhijn and Vanderleyden, 1995, Giraud et al., 2007, Rivas et al., 2009, Okubo et al., 2012). This ability is thought to have evolved through horizontal gene transfer as the genes involved in this process are usually located on symbiosis islands on the chromosomes of bradyrhizobia (Rivas et al., 2009, Sachs et al., 2010, Okubo et al., 2012) or on symbiotic plasmids in many rhizobia (van Rhijn and Vanderleyden, 1995). Some isolates of *Bradyrhizobium* have been shown to be non-symbiotic and do not possess the ability to form nodules. The absence of nodulation ability has been noted in the strain *Bradyrhizobium* sp. S23321 isolated from paddy soil in Japan (Okubo et al., 2012) although nitrogen fixation (*nif*) genes were present. Recently, *Bradyrhizobium* ecotypes

from forest soils have been shown to lack both nodulation and nitrogen fixation genes (VanInsberghe et al., 2015).

5.1.3 Reference *Bradyrhizobium* genomes

Bradyrhizobium is studied extensively due to its symbiotic relationship with soybean, a major global crop, and consequently several genomes have been published. Currently, there are seven complete *Bradyrhizobium* genomes in the NCBI database. Six of these are symbiotic and are able to fix nitrogen and form root nodules on legumes (*B. diazoefficiens* USDA 110, *B. japonicum* USDA 6, *B. japonicum* E109, *Bradyrhizobium* sp. ORS278, *Bradyrhizobium* sp. BTAi1 and *B. oligotrophicum* S58) with ORS278, BTAi1 and S58 able to form both stem and root nodules on the aquatic legume *Aeschynomene* (Giraud et al., 2007, Okubo et al., 2013). The other bradyrhizobial genome (*Bradyrhizobium* sp. S23321) is free-living because it is unable to form nodules; however, it still contains the genes required for nitrogen fixation. More information on these strains can be found in Chapter 3 (3.2.7). Four genomes sequenced from North American forest soils were also missing nodulation and nitrogen fixation genes (*Bradyrhizobium* sp. LTSP849, *Bradyrhizobium* sp. LTSP857, *Bradyrhizobium* sp. LTSP885 and *Bradyrhizobium* sp. LTSPM299). These genomes were sequenced using shotgun sequencing of the soil community and were assembled to near completion (VanInsberghe et al., 2015). Due to the availability of a diverse array of genome reference sequences, *Bradyrhizobium* is an appropriate model to study other soil bacteria: understanding the mechanisms of *Bradyrhizobium* adaptation to independent living in agricultural soils under contrasting management may reveal the genetic potential of this globally important genus.

5.1.4 Aims

In this chapter, I aim to:

- Present the genome and gene annotations of two isolates from the culture collection
- Identify the level of genetic heterogeneity in the genomes and compare them with other reference genomes
- Mine the genomes to detect functional differences between them and other reference genomes

5.2 Materials and methods

5.2.1 Isolation and DNA extraction

The soil bradyrhizobia sequenced for this chapter were isolated as described in Chapter 3. Two isolates, one from grassland soil (G22) and one from bare fallow soil (BF49) were chosen at random from the generated *Bradyrhizobium* culture collection to have their genomes sequenced *de novo.* DNA was extracted from isolates grown in MAG broth incubated at 28^{o}C and shaken at 100 rpm using the GenElute Bacterial Genomic DNA kit (Sigma Aldrich, USA) as described in General Methods 2.1, 2.2 and 2.3. Extracted DNA was quantified using Qubit Fluorometric Quantification (Life Technologies, USA) as described in General Methods 2.3.

5.2.2 *De novo* genome sequencing

Two sequencing platforms were used; Ion Torrent PGM™ (Life Technologies, USA) (located at Rothamsted Research, UK) and Illumina HiSeq 2000 (Illumina Inc, USA) (located at BGI, the Beijing Genomics Institute, China). For Ion Torrent sequencing, two barcoded, unamplified sequencing libraries were constructed using an Ion Plus Fragment Library Kit (Life Technologies, USA) using 1µg template DNA. Libraries were pooled and sequence template generated using an Ion PGM™ Template OT2 400 Kit. Sequencing was performed using the Ion PGM™ Sequencing 400 Kit and an Ion 318™ Chip. The G22 sample generated 3,652,179 reads (read length mean 262 bp; read length mode 345 bp; total bp 958,495,622 bp) with coverage of approximately 105X. The BF49 sample generated 3,129,252 reads (read length mean 270 bp; read length mode 363 bp; total bp 845,983,526 bp) with coverage of 93X. The Ion Torrent library construction and sequencing were carried out by Dr Steve Hanley, Rothamsted Research. Illumina sequencing was performed by BGI and 40 ug of genomic DNA was used. A 6 kb mate library with 90 bp read length and 50X coverage was created using the Illumina HiSeq 2000 sequencing platform resulting in 2,997,699 paired reads for G22 and 3,026,005 paired reads for BF49.

5.2.3 *De novo* genome assembly and annotation

The raw sequence data was quality checked using FASTQC and was quality trimmed accordingly. The Ion Torrent data was quality trimmed using the Trimmomatic tool and a quality score of 20 as a baseline. The FASTQ quality trimmer on Illumina data using a baseline quality score of 25. The sequence data was assembled using SOAPdenovo2 assembler and gap closer (Luo et al., 2012)

using a maximum read length of 615 bp and the average insert size of 6 kb. A range of assemblies were produced using kmer values; 61-83, 81, 83, 85, 87, 89, 91. Example code can be found in Appendix D. The genomes were manually curated using Geneious (Biomatters Ltd v8.1.5). The Geneious mapping tool was used to extend and further scaffold the reference sequence against the set of scaffolds generated from a multi-kmer assembly 61-83. Gap closing was carried out by PCR: Primer3 in Geneious was used to design primers for these regions. The PCR reaction mixture was 50 µl and consisted of Q5 buffer (10 µl) (New England Biolabs, USA), BioLine dNTP mix 10 mM of each nucleotide (1 µl), 1 µl of each forward and reverse primers to be tested (100 pM/µl), Q5 high-fidelity DNA polymerase (0.5 µl) (New England Biolabs, USA), 10 µl GC enhancer (New England Biolabs, USA), 2 µl of template DNA (5 ng/µl) and 20.5 µl of nuclease-free water. DMSO (5µl in 50 µl reaction) or BSA (1µl in a 50 µl reaction) was added if products were not able to be amplified. The PCR conditions were as described in Chapter 3 (Section 3.2.6), altering the annealing temperature depending on the primer sequence (primer sequences can be found in Appendix E). The PCR products were viewed using agarose gel electrophoresis as described in General Methods 2.5 and sequencing was carried out by MWG Eurofins as described in General Methods 2.6. Any remaining contigs from the 61-83 kmer reference assembly over 500 bp were investigated while those under were discarded. The G22 assembly had 19 contigs over 500 bp unplaced within the chromosome, with some containing annotated genes. No contigs over 500 bp remained after the completion of the final BF49 reference and so we can be confident that all genes are present in the assembly. The genomes were uploaded into RAST (Rapid Annotation using Subsystem Technology) for annotation using the default parameters (Aziz et al., 2008, Overbeek et al., 2014).

5.2.4 Genome comparisons

Genome annotations were downloaded from RAST and examined manually using KEGG (Kanehisa and Goto, 2000, Kanehisa et al., 2016). OrthoVenn was used to assess gene orthology using an e-value of $1e^{-10}$ (Wang et al., 2015). BRIG (BLAST Ring Image Generator) was used to compare the genomes with other *Bradyrhizobium* isolates (NCBI blast version 2.2.31) (Alikhan et al., 2011). Large-Scale Genome Alignment Tool (LASTZ) was used to assess genome-wide sequence similarity in Geneious (Harris, 2007).

5.2.5 16S rRNA phylogeny analysis

The 16S rRNA gene sequences for G22 and BF49 were compared with other *Bradyrhizobium* sequences in the NCBI database. All *Bradyrhizobium* 16S sequences from the NCBI RefSeq database and the four *Bradyrhizobium* isolates from North American forest soils (VanInsberghe et al., 2015) were downloaded. These sequences were aligned using MUSCLE using 8 iterations (Edgar, 2004). The aligned region was extracted (1220 bp) and a phylogenetic tree was created using the neighbour joining clustering method with 1000 bootstraps using the Geneious tree builder tool. A 75% support threshold was used for drawing the phylogenetic tree. Accession numbers for the sequences used in this analysis can be found in Appendix F.

5.2.6 Core genome phylogeny

OrthoVenn was used to identify genes which were present in all 9 genomes and were considered the "core genome" (Wang et al., 2015). The inflation value was set to 1.5 and an e-value of $1e^{-10}$. These genes were uploaded into the call SNPs and infer phylogeny (CSI) tool (Kaas et al., 2014) hosted by the Center for Genomic Epidemiology (Epidemiology, 2011) using the default options using *B. diazoefficiens* USDA 110 as the reference sequence. Default options were as follows: minimum of 10x depth at SNP positions, 10% relative minimum depth at SNP positions, a minimum of 10 bp distance between SNPs, minimum SNP quality score of 30, minimum read mapping quality score of 25 and minimum z-score of 1.96. Altered FastTree was selected. The CSI tool works but aligning the genomes in question to the selected reference and calling SNPs across the whole genome. These SNPs are then concatenated, ignoring indels, and FastTree is used to generate maximum likelihood trees (Leekitcharoenphon et al, 2012). No outgroup was used as this would limit the number of homologous genes included in the analysis. The core genome phylogenetic tree (Newick file) was visualised in Geneious.

5.2.7 Availability of Data and Materials

The sequences have been deposited in the ENA database [Study ID: PRJEB10689, Sample ID G22 and unplaced contigs: ERS955657, G22 plasmid: ERS955536, Sample ID BF49: ERS954959]. The raw sequence data was also uploaded to ENA [G22: ERR1110561, BF49: ERR1110562] for Illumina and Ion Torrent [G22: ERR1110625, BF49: ERR1110597]. The isolates have been stored on beads as described in General Methods 2.7.

5.3 Results and Discussion

5.3.1 General genome description and comparisons

The chromosome of the grassland isolate G22 is 9,022,917 bp in size while the bare fallow isolate BF49 genome is 7,547,693 bp, constituting a 1.5 Mbp size difference in addition to a 364,482 bp plasmid in G22. The genome size for G22 is similar to nodulating strains *B. diazoefficiens* USDA 110, *B. japonicum* USDA 6, *B. japonicum* E109, *B. oligotrophicum* S58 and *Bradyrhizobium* sp. BTAi1, whereas for BF49 it is closer in size to the free-living strain *Bradyrhizobium* sp. S23321 and the photosynthetic, nodulating strain *Bradyrhizobium* sp. ORS278. Table 5.1 summarises the genome information of the two novel strains G22 and BF49 along with the other seven completed *Bradyrhizobium* genomes in the database. G22 had 19 contigs which could not be placed in the chromosome or plasmid. These were annotated and used in the subsequent analysis.

G22 shows 1356 more genes (8787) than BF49 (7431) and this rises to 1902 when including the genes contained on the plasmid (541). GC content is similar between the strains at 63.7% and 63.8% for G22 and BF49 respectively, consistent with other *Bradyrhizobium* strains listed in Table 5.1. The G22 plasmid has a GC content of 60.7%, identical to the plasmid of BTAi1. The G22 and BF49 genomes show high pairwise sequence identity in comparison to both USDA 110 (G22: 84.4%, BF49: 83.4%) and S23321 (G22: 83.4%, BF49: 83.2%) using LASTZ (Large-Scale Genome Alignment Tool).

Table 5.1: Summary of the seven published complete *Bradyrhizobium* genomes and the two novel strains G22 and BF49.

Taxonomy	Strain	Host	Origin	Genome size (bp)	GC content	Proteins	rRNA operon	tRNA	Gene	Accession number	Reference
Bradyrhizobium sp.	BF49	Free-living	UK	7,547,693	63.80%	7380	1	48	7431	ERS954959	(Jones et al., 2016)
Bradyrhizobium sp.	G22	Free-living	UK	9,022,917	63.70%	8653	1	49	8706	ERS955657	(Jones et al., 2016)
Bradyrhizobium sp.	G22 plasmid	Free-living	UK	364,482	60.70%	541	1	2	546	ERS955536	(Jones et al., 2016)
Bradyrhizobium sp.	G22 unplaced	Free-living	UK	68,438	62.10%	81	0	0	81	ERS955657	(Jones et al., 2016)
Bradyrhizobium sp.	S23321	Free-living	Japan	7,231,841	64.30%	6898	1	45	6951	NC_017082	(Okubo et al., 2012)
B. diazoefficiens	USDA 110	*Glycine max*	USA	9,105,828	64.10%	8317	1	50	8373	NC_004463	(Kaneko et al., 2002)
B. japonicum	USDA 6	*Glycine max*	Japan	9,207,384	63.70%	8829	2	51	8888	NC_017249	(Kaneko et al., 2011)
B. japonicum	E109	*Glycine max*	Argentina	9,224,208	63.70%	8233	2	54	8621	NZ_CP010313	(Torres et al., 2015)
B. oligotrophicum	S58	*Aeschynomene*	Japan	8,264,165	65.10%	7228	2	51	7285	NC_020453	(Okubo et al., 2013)
Bradyrhizobium sp.	ORS278	*Aeschynomene*	Africa	7,456,587	65.50%	6752	2	50	6818	NC_009445	(Giraud et al., 2007)
Bradyrhizobium sp.	BTAi1	*Aeschynomene*	N. America	8,264,687	64.90%	7394	2	52	7553	NC_009485	(Giraud et al., 2007)
Bradyrhizobium sp.	BTAi1 plasmid	*Aeschynomene*	N. America	228,826	60.70%	203	0	0	216	NC_009475	(Giraud et al., 2007)

5.3.2 Orthologous gene clusters and core genome phylogeny

G22 and BF49 were compared with the free-living isolate, S23321, and the symbiotic isolate, USDA 110 (Figure 5.1). This suggests that there is a core genome of 4562 genes which are present in all four genomes assessed. The 103 genes present only in the USDA 110 genome include those involved in nodulation and uptake hydrogenase. The 171 genes which are only present in the USDA 110 and S23321 genomes include nitrogen fixation genes. Only a small number of genes are unique to any one isolate (G22: 99, BF49: 90, S23321: 35, USDA 110: 103). OrthoVenn identified a core genome of 3442 homologous gene families across the nine complete genomes available in the database. The core genome SNP analysis (Figure 5.2) shows that G22 clusters closest with the *B. japonicum* strains; E109 and USDA 6. BF49 is separate from G22 being more basal in the phylogenetic tree. The closest relatives are the free-living strain S23321 and the soybean nodulating type-strain USDA 110. Two of the photosynthetic, *Aeschynomene*-nodulating isolates (BTAi1 and S58) cluster together.

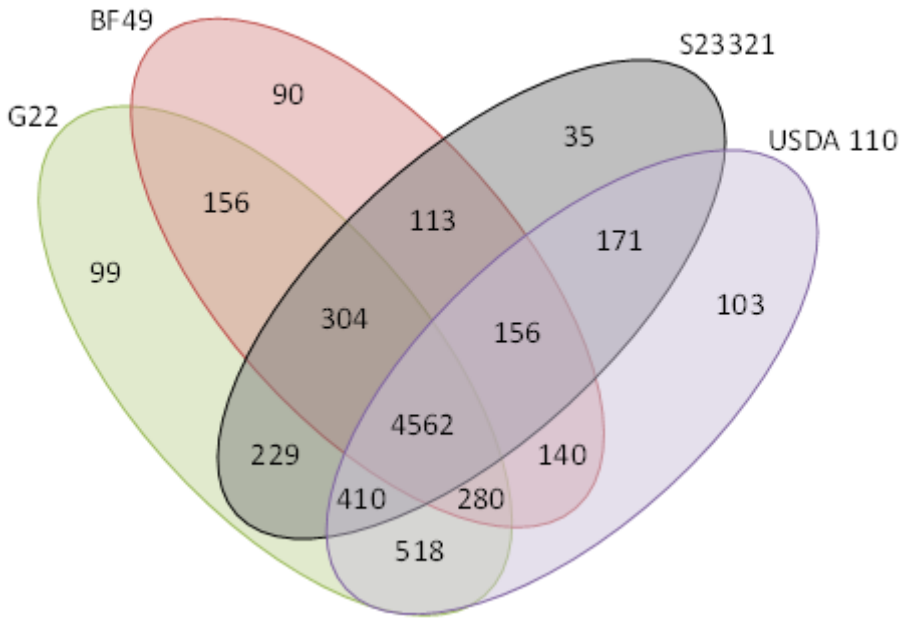


Figure 5.1: Venn diagram showing orthologous gene clusters for G22, BF49 and two published genomes, S23321 and USDA 110.
Output generated using data from OrthoVenn.

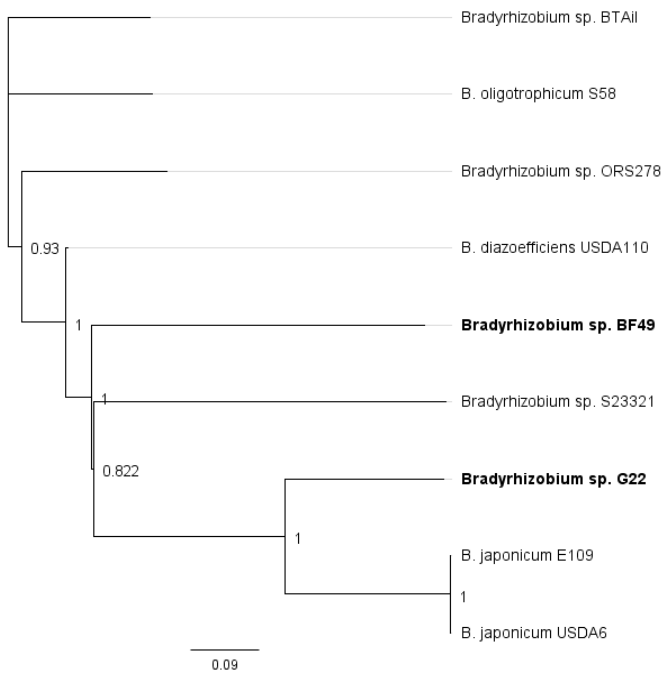


Figure 5.2: Phylogenetic tree for the 3442 homologous gene clusters in the core genome for the nine *Bradyrhizobium* genomes.
The node labels represent the certainty of that node in the phylogenetic tree where one is maximum certainty.

5.3.3 Isolate identification and 16S phylogeny

The 16S rRNA sequence from G22 shows 100% identity with *Bradyrhizobium* sp. VUPMI37 [Accession number: HG940535] and *Bradyrhizobium* sp. ICMP12674 [Accession number: AY491080] (Weir et al., 2004) from the NCBI and RDP databases respectively. *Bradyrhizobium* sp. VUPMI37 was originally isolated from *Vigna unguiculata* nodules in Spain and *Bradyrhizobium* sp. ICMP12674 was originally isolated from *Ulex europaeus* in New Zealand (Weir et al., 2004, Bejarano et al., 2014). The BF49 16S rRNA fragment shows 100% identity with *B. canariense* SEMIA928 from the NCBI database [Accession number: FJ390904] (Menna et al., 2009) and *B. canariense* LG-6 from the RDP database [Accession number: GU306140]. *B. canariense* SEMIA928 was originally isolated from *Lupinus* spp. in Australia and *B. canariense* LG-6 was originally isolated from *Lupinus angustifolius* root nodules in Spain (Menna and Hungria, 2011). The 16S phylogeny (Figure 5.3) clusters G22 with VUPMI37 and ICMP12674, which were the top hits from the NCBI and RDP databases. Similarly, BF49 clusters with the top hits from the NCBI and RDP databases (SEMIA928 and LG-6) in addition to *B. lupini* USDA 3051 which was the second hit from the NCBI database [Accession number: NR_134836]. This strain, isolated in the USA, was originally isolated from *Lupinus* and was reclassified from *Rhizobium lupini* to *B. lupini* in 2015 (Peix et al., 2015). Both G22 and BF49 are in the same clade as the free-living strain, S23321, and the soybean-

nodulating strains, E109, USDA 6 and USDA 110. Two of the four strains from North American forest soils, *Bradyrhizobium* sp. LTSP849 and LTSP857 (VanInsberghe et al., 2015) were also in this clade. The photosynthetic strains (ORS278, BTAi1 and S58) are in a separate clade including the remaining two strains from North American forest soils *Bradyrhizobium* sp. LTSP885 and LTSPM299. Sequences with identical sequence names with one having a suffix of "2", indicates where sequences are from the same strain but has numerous identical copies. For example, *B. elkanii* USDA 76 has two copies of 16S rRNA.

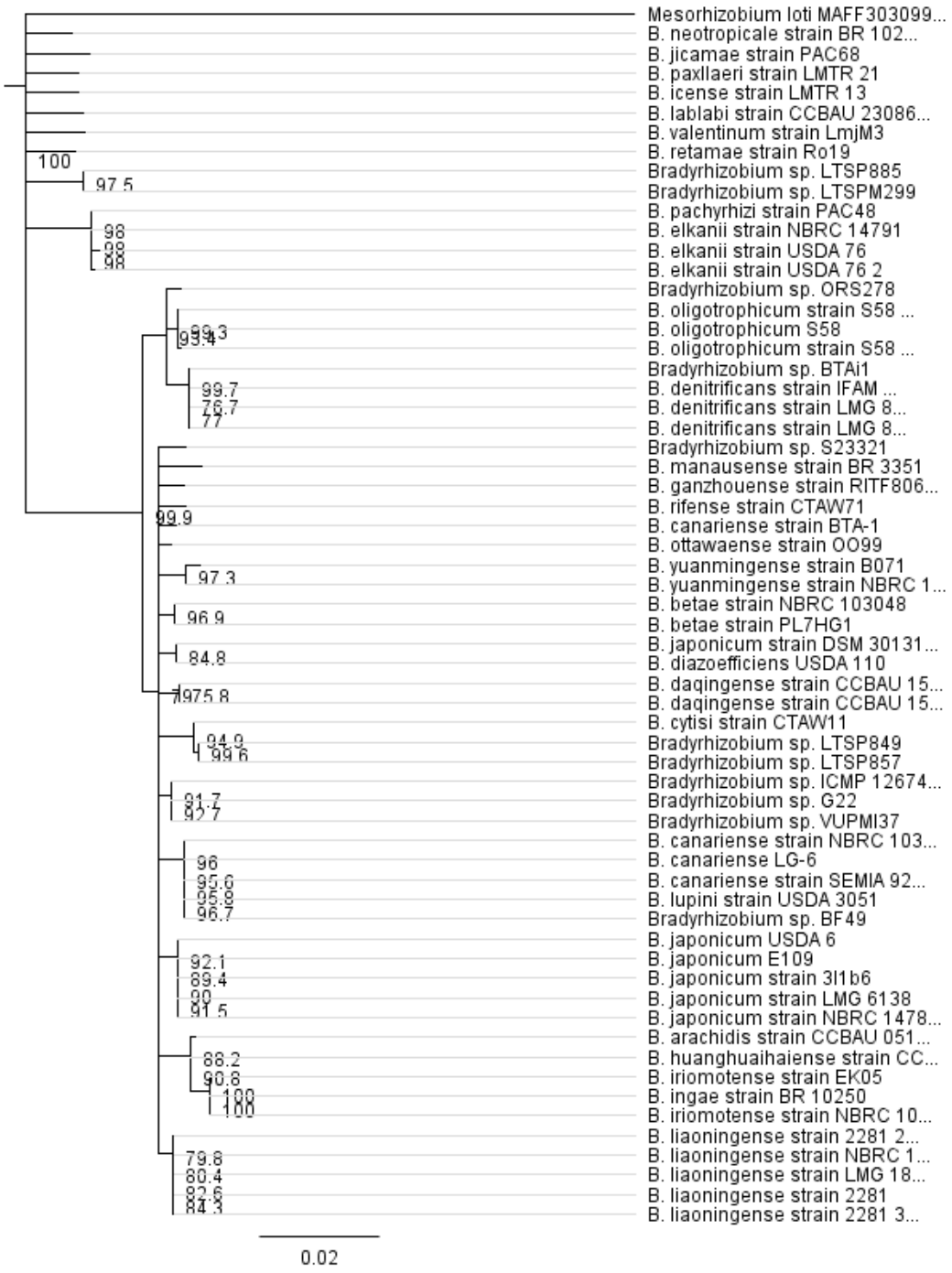


Figure 5.3: Phylogenetic tree for the 16S rRNA gene using Neighbour-Joining clustering method with 1000 bootstraps.

5.3.4 Genes involved in nitrogen fixation and nodulation

Nitrogen fixation and nodulation genes including *nifDKH, nodD* and *nodABC* are all absent from both G22 and BF49 (Table 5.2) and so we suggest that these isolates are not able to either form nodules or fix atmospheric nitrogen. The absence of both *nif* and *nod* genes is in contrast to other published, complete *Bradyrhizobium* genomes and indicates similarity with forest soil bacteria (LTSP849, LTSP857, LTSP885 and LTSPM299) (VanInsberghe et al., 2015). *B. diazoefficiens* USDA 110 (previously *B. japonicum* USDA 110) is a known symbiotic strain and contains all *nif* and *nod* genes listed. The seven complete *Bradyrhizobium* genomes which have previously been published either contain both *nif* and *nod* genes (the soybean-nodulating strains *B. diazoefficiens* USDA 110, *B. japonicum* USDA 6 and *B. japonicum* E109) or just *nif* genes (the *Aeschynomene*-nodulating strains *Bradyrhizobium* sp. BTAi1, *Bradyrhizobium* sp. S58 and *Bradyrhizobium* sp. ORS278 and the free-living strain *Bradyrhizobium* sp. S23321) (Figure 5.4). These strains use a *nod*-independent route for stem and root nodulation of *Aeschynomene* (Okubo et al., 2013). The FixLJ two component system is present in both the grassland and bare fallow isolate (Table 5.2). FixLJ acts in response to low oxygen conditions in soil and in root nodules and controls the expression of genes for both nitrogen fixation and denitrification (Ferrieres et al., 2004, Bobik et al., 2006). This two-component system has also been shown to regulate the response to nitric oxide in *Sinorhizobium meliloti* and to regulate a high proportion of genes induced by the presence of nitric oxide (Meilhoc et al., 2010). The presence of *fixLJ* in G22 and BF49 is consistent with other *Bradyrhizobium* isolates including all seven completed genomes.

5.3.5 Genes involved in denitrification

Both genomes encode a nitrate reductase, NapA/B, which catalyses the reduction of NO_3^- to NO_2^-; the first stage in denitrification (Delgado et al., 2003, Bedmar et al., 2005, Fernandez et al., 2008). This is common among *Bradyrhizobium* including all seven complete genomes (Table 5.2 & Figure 5.4) and the four genomes from North American forest soils (Table 5.2). In addition, all fully-sequenced isolates including both G22 and BF49 contain *nirK,* encoding a respiratory copper-containing nitrite reductase which is involved in the second stage of denitrification reducing NO_2^- to NO. The third stage of denitrification is the conversion of NO into N_2O, a potent greenhouse gas, and is catalysed by a nitric oxide reductase encoded by *norB/C* (Bedmar et al., 2005, Fernandez et al., 2008). The presence of a nitric oxide reductase gene has been noted in all previously published and complete *Bradyrhizobium* genomes in addition to G22 and BF49. The

denitrification pathway is not present in the four genomes of strains from North American forest soils.

The ability to convert the greenhouse gas nitrous oxide into environmentally benign nitrogen gas in the final stage of denitrification is an attribute with global importance (Richardson et al., 2009). Nitrous oxide reductase encoded by *nosZ* catalyses this process but it is not ubiquitous across *Bradyrhizobium* (Bedmar et al., 2005, Henry et al., 2006, Fernandez et al., 2008, Shiina et al., 2014). Of the seven published complete genomes, only *B. diazoefficiens* USDA 110, *Bradyrhizobium* sp. BTAi1 and *B. oligotrophicum* S58 contain the *nosZ* gene. It is absent from the grassland isolate, G22 but present in the bare fallow isolate, BF49. The presence of the gene shows a potential for BF49 to perform this function and increases the agricultural and environmental importance of this isolate and related phylotypes.

5.3.6 Uptake hydrogenase

The uptake of hydrogen is catalysed by uptake hydrogenase which is encoded by the *hup* genes (Zuber et al., 1986, van Berkum, 1987, Black et al., 1994). This process produces ATP which is used by nitrogen-fixing bacteria to mediate for energy lost through the nitrogen fixation process (Zuber et al., 1986, Black et al., 1994). The nickel-iron hydrogenase, encoded by *hupSL* (Black et al., 1994), is absent from both G22 and BF49 but is present in all of the symbiotic strains of *Bradyrhizobium* with complete genomes. These genes are also absent from LTSP849, LTSP857, LTSP885 and LTSPM299 (Table 5.2). The proteins encoded by *hupU* and *hupV* show high homology with the nickel-iron hydrogenase with *hupU* and *hupV* encoding the small and large subunits respectively and *hupV* is thought to be necessary for the activation of hydrogenase (Black et al., 1994). These genes are also absent from both G22 and BF49 but are present in symbiotic strains including USDA 110, BTAi1 and ORS278.

5.3.7 Photosynthesis and carbon fixation

Genes for photosynthesis are present in four of the published complete *Bradyrhizobium* genomes; S23321 (free-living), S58, BTAi1 and ORS278 (*Aeschynomene* host) (Figure 5.4-b). These include genes for bacteriochlorophyll (*bchCXYZ*/*FNBHLM*), carotenoids (*crtEF*), light harvesting polypeptides (*pucBAC*/*pufBA*) and reaction centre subunits (*puhA*/*pufLM*) (Gregor and Klug, 1999, Igarashi et al., 2001). They are not present in the soybean-nodulating isolates (USDA 110, USDA 6, and E109), G22 or BF49 genomes or the four isolates from forest soils (LTSP849, LTSP857, LTSP885 and LTSPM299) (Table 5.2). Many heterotrophic bacteria including *Bradyrhizobium* can fix carbon

using the Calvin-Benson-Bassham cycle (CBB cycle). The significance of this is unclear in most cases although the photosynthetic *Aeschynomene*-nodulating strain ORS278 has been shown to require an active CBB cycle for symbiotic nitrogen fixation (Gourion et al., 2011). The first stage of the CBB cycle is catalysed by Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo) which is present in both G22 and BF49 genomes, consistent with other *Bradyrhizobium* isolates (Table 5.2). Transketolase is an important enzyme in both the CBB cycle and pentose phosphate pathway, catalysing the interconversion of sugars (Horecker et al., 1956, Fullam et al., 2012). Genes for the transketolase enzyme are present in both G22 and BF49 consistent with all other complete *Bradyrhizobium* isolates which have been published.

Table 5.2: Summary of presence or absence of genes of interest between *Bradyrhizobium* genomes.

This compares the seven published complete *Bradyrhizobium* genomes, four isolates from North America forest soils and the two novel strains G22 and BF49 (+ = present, - = absent).

Function	Gene(s)	G22	BF49	S23321	ORS278	BTAi1	S58	USDA 110	USDA 6	E109	LTSP849	LTSP857	LTSP885	LTSPM299
Nitrogen fixation	*nifDKH*	-	-	+	+	+	+	+	+	+	-	-	-	-
Nitrogen fixation	*fixLJ*	+	+	+	+	+	+	+	+	+	-	-	-	-
Nodulation	*nodD*	-	-	-	-	-	-	+	+	+	-	-	-	-
Nodulation	*nodABC*	-	-	-	-	-	-	+	+	+	-	-	-	-
Denitrification	Nitrate reductase *napA/B*	+	+	+	+	+	+	+	+	+	+	+	+	+
Denitrification	Nitrite reductase *nirK*	+	+	+	+	+	+	+	+	+	-	-	-	-
Denitrification	Nitric oxide reductase *norB/C*	+	+	+	+	+	+	+	+	+	-	-	-	-
Denitrification	Nitrous oxide reductase *nosZ*	-	+	-	-	+	+	+	-	-	-	-	-	-
Uptake hydrogenase	*hup*	-	-	-	+	+	+	+	+	+	-	-	-	-
Photosynthesis	Bacteriochlorophyll *bchCXYZ/FNBHLM*	-	-	+	+	+	+	-	-	-	-	-	-	-
Photosynthesis	Carotenoids *crtEF*	-	-	+	+	+	+	-	-	-	-	-	-	-
Photosynthesis	Light harvesting complexes *pucBAC/pufBA*	-	-	+	+	+	+	-	-	-	-	-	-	-
Photosynthesis	Reaction centre subunits *puhA/pufLM*	-	-	+	+	+	+	-	-	-	-	-	-	-
Carbon fixation	RuBisCo	+	+	+	+	+	+	+	+	+	+	+	+	+

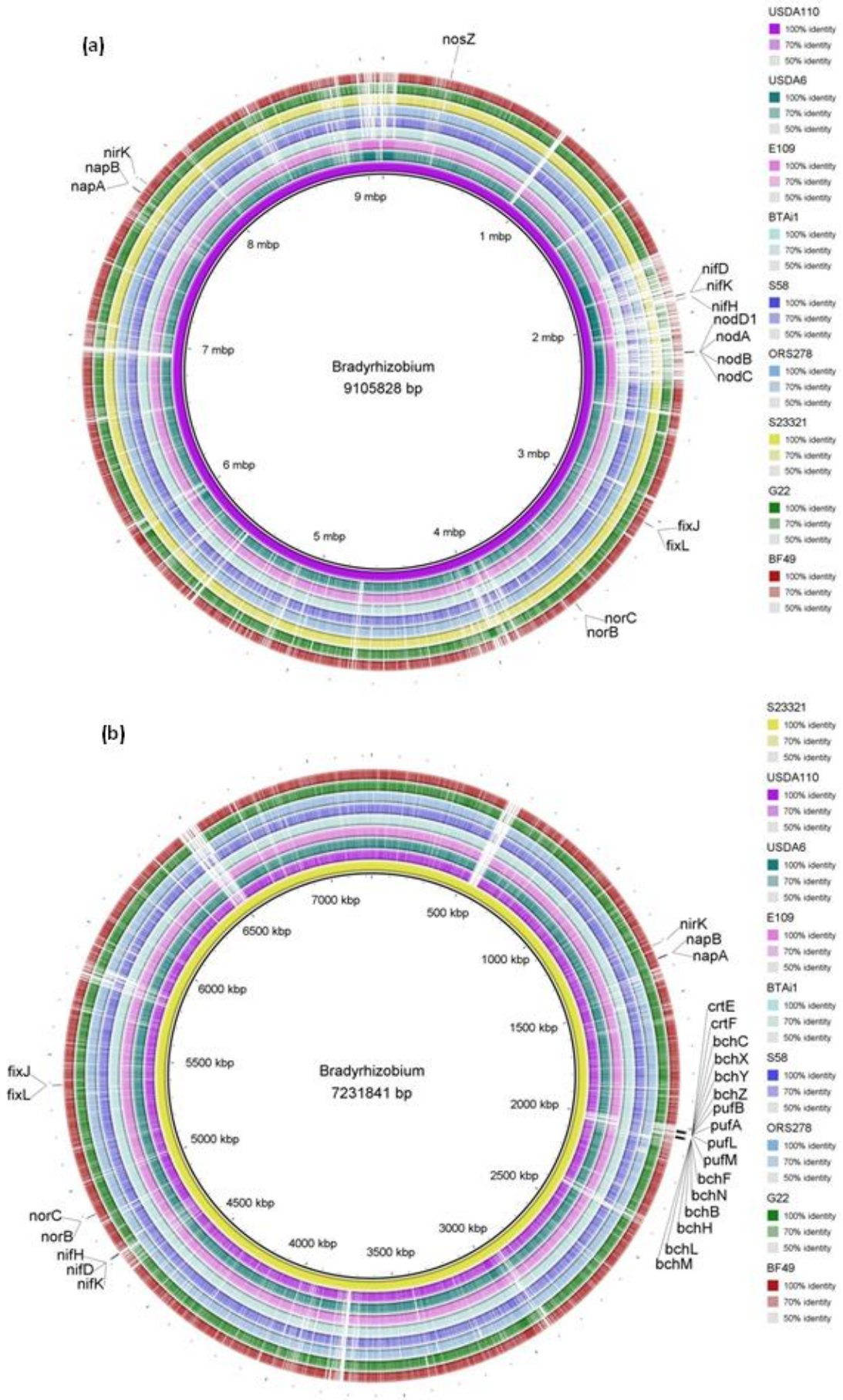


Figure 5.4: Whole chromosome comparisons showing positions of genes involved with nitrogen cycling, nodulation and photosynthesis on the reference genome sequence.

Reference sequence: USDA110 (a) and S23321 (b).

5.3.8 Carbon metabolism

Malonic acid can be found in plant tissues, including legumes being first characterised from alfalfa leaves in 1925 and has been found to be in very high concentrations in soybeans (Dimroth and Hilbi, 1997, Kim, 2002, Suvorova et al., 2012). This pathway was found to be closely associated with the symbiotic nitrogen fixation pathway in *Rhizobium leguminosarium* sv *trifolii* (Kim et al., 1993). Malonic acid is activated before being broken down into acetate and carbon dioxide through decarboxylation (Schmid et al., 1996). It is often converted to malonyl-CoA by a CoA transferase, which is present in G22 (Phosphoribosyl-dephospho-CoA transferase) and the malonyl-CoA is then decarboxylated by malonate decarboxylase (Schmid et al., 1996, Dimroth and Hilbi, 1997). Malonate decarboxylase has four subunits; alpha, beta, gamma and delta; genes for all of which are present in G22. The grassland isolate also contains two *mad* genes; *madL* and *madM.* These genes have previously been reported to be part of the malonate decarboxylase operon as transport proteins thought to be involved in malonate uptake (Dimroth and Hilbi, 1997). The grassland isolate G22 also contains a malonyl CoA acyl carrier protein transacylase and triphosphoribosyl-dephospho-CoA synthetase. All genes involved in malonate decarboxylation are absent in the bare fallow isolate BF49. Malonate transport and utilisation genes are also present in other *Bradyrhizobium* isolates including BTAi1, ORS278, USDA 110, S23321 and all four of the forest strains (LTSP849, LTSP857, LTSP885 and LTSPM299). This suggests that BF49 may have lost this pathway rather than G22 gaining these genes.

Amino sugars are often used as an energy source by bacteria as these sugars are able to provide both carbon and nitrogen (Plumbridge and Vimr, 1999). N-acetyl-glucosamine (GlcNAc) is an example of an abundant amino sugar found in nature as it is the main component of chitin (Yang et al., 2006). GlcNAc is converted into fructose 6-phosphate via a number of reactions including deacetylation and deamination and is transported by the phosphotransferase transport system (Alvarez-Anorve et al., 2005, Yang et al., 2006). Two GlcNAc ATP-binding cassette (ABC) transport permease proteins and sugar-binding proteins are present in the grassland isolate only and are absent from the bare fallow isolate. GlcNAc is phosphorylated during transport and produces N-acetylglucosamine-6-phosphate (GlcNAc6P) which is then acetylated by GlcNAc6P deacetylase encoded by *nagA* and is present in the grassland isolate but absent from BF49 (Alvarez-Anorve et al., 2005). This process produces glucosamine-6-phosphate (GlcN6P) which is deaminated via a GlcN6P deminase encoded by *nagB* to produce fructose-6-phosphate and ammonia (Alvarez-Anorve et al., 2005). GlcN6P deaminase is present in G22 but absent from BF49. The presence of

nagA and *nagB* is not common across the published *Bradyrhizobium* genomes which suggest that G22 may have acquired these genes from the environment.

Inositol is a sugar alcohol which is abundant in nature and has been shown to have three stereoisomers; *myo-, D-chiro-* and *scyllo-* (Yoshida et al., 1997, Kohler et al., 2010). Catabolism of the *myo*-inositol form has been studied in Firmicutes, Enterobacteriaceae and Rhizobiaceae, and is well characterised for *Bacillus subtilis*, *Klebsiella aerogenes* and *Enterobacter aerogenes* (Berman and Magasani, 1966, Yoshida et al., 1997, Kohler et al., 2010). The first stage of *myo*-inositol catabolism is catalysed by *myo*-inositol-2-dehydrogenase (Idh) which oxidises *myo*-inositol into 2-keto-*myo*-inositol (2KMI) (Yoshida et al., 1997, Kohler et al., 2010) and is present in the bare fallow isolate, BF49. This was initially shown in *B. subtilis* but this enzyme is thought to be involved in similar pathways in *Cryptococcus melibiosum* and *Rhizobium leguminosarum* bv *viciae* and it has also been shown that *Sinorhizobium meliloti* can use *myo*-inositol as a sole carbon source (Poole et al., 1994, Kohler et al., 2010). The bare fallow isolate also contained an epi-inositol hydrolase and inosose dehydratase which are both associated with inositol catabolism. Only the bare fallow isolate BF49 contained genes associated with inositol catabolism, they were absent from the grassland isolate. The inositol pathway is also present in other *Bradyrhizobium* isolates including the photosynthetic and symbiotic strains; BTAi1, OR278, USDA 6 and E109. The presence of this pathway in the bare fallow isolate and not in the grassland isolates is surprising as the ability to catabolise inositol is thought to be a competitive advantage within the rhizosphere as rhizopines are inositol derivatives and the catabolism of rhizopine plays a role in competition for nodule occupancy (Galbraith et al., 1998, Kohler et al., 2010).

Sugars undergo phosphorylation before metabolism resulting in sugar phosphates which are formed by kinases. However, in bacteria, phosphorylation of carbohydrates can occur by utilising phosphoenolpyruvate (PEP) which is part of the phosphoenolpyruvate phosphotransferase system (PTS) (Kundig et al., 1964, Gutknecht et al., 2001, Bachler et al., 2005). Dihydroxyacetone is a product of glycerol oxidation and is phosphorylated via dihydroxyacetone kinases (Bachler et al., 2005). The PTS kinases comprise three subunits; DhaK, DhaL and DhaM with DhaK containing the substrate binding site and DhaL containing ADP which acts as a co-enzyme (Gutknecht et al., 2001, Bachler et al., 2005) . The *dhaKLM* operon is present in the grassland isolate but absent from the bare fallow isolate. This is also absent from other *Bradyrhizobium* isolates which suggests that G22 gained this operon from the environment.

Methane (CH_4) is a greenhouse gas and is therefore of global significance. Methanotrophs are organisms which are able to utilise CH_4 and are classified into two types; type I (gamma-Proteobacteria) and type II (alpha-Proteobacteria) (Jhala et al., 2014). There are three enzymes

which are involved in the metabolism of CH_4; particulate methane monooxygenase (pMMO), soluble particulate methane monooxygenase (sMMO) and methanol dehydrogenase (MDH) (Murrell et al., 2000, Jhala et al., 2014). Although there are 13 recognised genera of methanotrophs, sMMO has been shown to be present and expressed in other genera including *Rhizobium* sp., *Bacillus megaterium* and *Paenibacillus illinoisensis* in the conversion of CH_4 to methanol (CH_3OH) (Jhala et al., 2014). Methane monooxygenase component A alpha and beta chains, component C and a regulatory protein are all present in only the grassland isolate G22. Putative methane monooxygenases have also been reported in. BTAi1, ORS278, E109 and S58.

5.3.9 Phosphorus utilisation

Alkylphosphonate is an organic phosphonate which is degraded by enzymes produced by the *phn* genes. This has been well studied in *Escherichia coli* with an operon of 17 genes being identified, *phnA* to *–Q* (Chen et al., 1990) which are also present in *Sinorhizobium meliloti* (Parker et al., 1999). The breakdown of phosphonates can be catalysed by a carbon-phosphorus (C-P) lyase and this is encoded by *phnG* to *–M;* all of which are present in both G22 and BF49. The genes *phnF* and *phnO* are thought to be regulatory proteins and *phnN* and *phnP* are accessory proteins; all are present in both genomes. Phosphonate ABC transporter proteins are produced by *phnCDE,* only *phnE1* and *phnE2* are present in both genomes, *phnCD* are absent. It is reported that for the uptake and breakdown of phosphonates, all genes in the operon from *phnC* to *–P* need to be present (Metcalf and Wanner, 1991). It is unclear whether these isolates are able to complete this process as only 12 of the 14 essential genes in the operon are present; additional experiments would be needed to determine this.

The Pho regulon is responsible for the uptake and transport of inorganic phosphorus. When phosphate levels are low, this is sensed by PhoR, a sensor kinase protein and activates PhoB, a response regulator which then binds to the Pho box of over 30 genes and regulates the induction of PstSCAB (Bardin and Finan, 1998, Baek and Lee, 2006, Santos-Beneit, 2015). The PstSCAB phosphate-specific transport system has been characterised in *E. coli, Acinetobacter johnsonii* and *Bacillus cereus* (Wanner and Chang, 1987, Bardin and Finan, 1998). This transport system is only present in the grassland isolate, G22. Both PhoR and PhoB are also only present in the grassland isolate. PhoU is also involved in sensing the levels of inorganic phosphorus which is only present in the grassland isolate (Bardin and Finan, 1998). Also in the Pho regulon are alkaline phosphatases which hydrolyse organic phosphorus and are encoded by *phoA/C/D* (Bradshaw et al., 1981, Apel et al., 2007, Fraser et al., 2015). These genes are common across numerous

bacteria including *Bradyrhizobium* and are present in both G22 and BF49. The alkaline phosphatases along with other phosphorus utilisation genes present in the Pho regulon are also present in the other published *Bradyrhizobium* genomes including USDA 110, USDA 6 and S23321. Polyphosphate kinase is highly conserved in bacteria and is present in both G22 and BF49 (Zhang et al., 2002). Low-affinity inorganic phosphate transporters, inorganic pyrophosphatases and phosphate starvation-induced protein PhoH are all present in both G22 and BF49.

5.3.10 Motility

The bacterial flagellum is usually made up of six components and these are composed of proteins which are encoded by a host of genes (Liu and Ochman, 2007). There are 24 core genes which are thought to be involved with flagella synthesis; filament formation (*fliC*), hook filament junction (*flgK, flgL*), hook protein (*flgE*), rod proteins (*flgB, flgC, flgG, flgF*), basal body proteins (*fliF, fliG, fliM, fliN*), motor proteins (*motA, motB*), export apparatus (*flhA, flhB, fliI, fliP, fliR, fliQ*), basal-body rod modification protein (*flgD*). There are three other flagella proteins which are considered core but aren't uniformly present in flagellated genomes; FliD, FliE and FliH (Kanehisa, 2000, Liu and Ochman, 2007). All are present in both G22 and BF49. Both genomes also contain other flagellar biosynthesis proteins; FlhF, FlhF and FliL. The presence of these genes supports the hypothesis that these isolates are motile. The presence of a flagellum is consistent with other *Bradyrhizobium* isolates which have been identified.

5.3.11 Chemotaxis

Chemotaxis is a process by which bacteria are able to sense the chemicals in their environment and move to more favourable conditions (Adler, 1966, Falke et al., 1997, Bren and Eisenbach, 2000). This process is due to a change in the direction of the flagellum rotation in response to signals which are received by the receptors from sensory proteins (Bren and Eisenbach, 2000). Chemotaxis receptors form a foundation to which the receptor supermolecular complex attaches (Falke et al., 1997). The receptor supermolecular complex is comprised of two proteins, CheW and CheA which are present in both G22 and BF49 (Falke et al., 1997, Bren and Eisenbach, 2000). CheA is a histidine kinase which is phosphorylated when a binding protein attaches to a receptor generating a signalling event which is transmitted across the bilayer to CheA which is bound to the receptor (Falke et al., 1997). The phosphate group which is bound to CheA is transferred to CheY or CheB which are response regulator proteins and all three proteins are present in G22 and BF49 (Falke et al., 1997, Bren and Eisenbach, 2000). The phosphorylated CheY protein then

interacts with FliM, a protein at the base of the flagella ultimately causing a change the rotational direction of the flagella via the FliM protein (Bren and Eisenbach, 2000). All proteins are present in both isolates.

5.3.12 Toxin, bacteriocin and antibiotic resistance genes

Heavy metals cobalt (Co^{2+}), zinc (Zn^{2+}) and cadmium (Cd^{2+}) resistance is regulated by the *czc* gene family. The genes *czcA, czcB* and *czcC* code for a membrane-bound protein which catalyses an efflux of Co^{2+}, Zn^{2+} and Cd^{2+} (Nies, 1992, Grosse et al., 1999). The gene *czcD* is thought to have a regulatory function but is not essential for the efflux of these metals (Grosse et al., 1999). These genes are organised in a four gene operon; *czcCBAD*, and these are all present in both G22 and BF49. The *czc* genes are not ubiquitous across *Bradyrhizobium* however they are present in S23321 and the isolates from *Aeschynomene* nodules; BTAi1, ORS278 and S58.

Arsenic resistance genes have been identified in numerous bacteria and on both chromosomes and plasmids including all of the seven published, complete *Bradyrhizobium* genomes (Chauhan et al., 2009). All genes involved in arsenic resistance which are present in G22 and BF49 are only present on the chromosome. Arsenic detoxification begins with the reduction of arsenate to arsenite by an arsenic reductase protein, ArsC which is present in both G22 and BF49. This is part of a larger operon including arsenic regulatory protein, ArsR and arsenic chaperones, ArsD, which only ArsR is present in both G22 and BF49 (Wu and Rosen, 1991, Chauhan et al., 2009). After the reduction step, the arsenite is transported out of the cell by an arsenite transporter (ArsB) which is present in both isolates. An arsenic pump driving ATPase (ArsA) is only present in BF49.

2,4-dichlorophenoxyacetic acid (2,4-D) is a model compound in the study of how microorganisms have evolved the capacity to be able to degrade xenobiotic organic compounds within soil using *tfd* genes (Kamagata et al., 1997, Itoh et al., 2002). The *tfdA* gene encodes the enzyme 2,4-D alpha-ketoglutarate dioxygenase which catalyses the first step of 2,4-D mineralisation and is present in both G22 and BF49 (Itoh et al., 2000, Zaprasis et al., 2010). The *tfdA* gene is a widespread gene and is present in all published and complete *Bradyrhizobium* genomes.

Colicins are bacteriocins produced by Gram-negative bacteria as a defence mechanism (Riley and Wertz, 2002, Gerard et al., 2005). Genes involved in colicin V and other bacteriocin production are present on both the G22 and BF49 chromosome including amidophosphoribosyltransferase, colicin V production proteins and dihydrofolate synthase. Colicin production is not unique to these isolates, with genes required for its production being present in USDA 6, E109 and S58. Colicin E2 is a DNA endonuclease produced by a broad range of bacteria being first characterised

in *E. coli* (Schaller and Nomura, 1976). The *cre* operon consists of four genes required for colicin E2 tolerance. The first is *creA* which encodes CreA and is present in both G22 and BF49. The second is CreB which encodes a response regulator and is missing from both isolates. The third, CreC, is a sensory kinase which forms part of the two component response system which is present in both isolates. The final gene in the operon is *creD* (also termed *cet*) encoding an inner membrane protein which is missing from both isolates (Cariss et al., 2010). Marinocine is a broad-spectrum antibacterial protein which is encoded by a lysine oxidase, LodA (Lucas-Elio et al., 2006). This is part of a two gene operon along with *lodB* coding for a dehydrogenase flavoprotein. Both *lodA* and *lodB* are present only in the grassland isolate, G22. These genes are not common across *Bradyrhizobium* and are absent from all other *Bradyrhizobium* genomes which have been published.

Beta-lactam antibiotics are a class of antibiotics which include cephalosporins and penicillin-based antibiotics including ampicillin and benzypenicillin (Livermore, 1995). Beta-lactamases are the enzymes which provide bacteria with resistance to these antibiotics. Class C beta-lactamases (*ampC*) originated in Gram-negative bacteria but are readily transferred via plasmids to other taxa (Doi and Paterson, 2007). This beta-lactamase is present in both the grassland and bare fallow isolates and other *Bradyrhizobium* strains including USDA 110 and E109. Some beta-lactamases require zinc ions to disrupt the beta-lactam ring through hydrolysis (Livermore, 1995). Metal-dependent hydrolases are present in both G22 and BF49.

Multidrug efflux proteins are common across bacteria including *Bradyrhizobium* and all seven of the complete *Bradyrhizobium* genomes contain some form of multidrug efflux pump or antimicrobial resistance. A homolog of YdhE/NorM multidrug efflux pump protein is present in both G22 and BF49. Long et al. (2008) showed that *E. coli* expressing YdhE/NorM had increased resistance to a variety of antimicrobials including fluoroquinolones. In addition, multidrug resistance transporters encoded by the *bcr* gene family, *cflA,* nodT, *cmaBA,* betalactamases and fluoroquinolone resistance proteins are all present in both G22 and BF49. The activity of the ABC-type exporter, MacB, is closely linked with two accessory proteins; the periplasmic fusion protein MacA and the outer membrane protein, TolC. Together MacAB and TolC produce an efflux pump (Lu and Zgurskaya, 2012). Both MacA and TolC are present in both isolates; however, MacB is only present in the bare fallow isolate, BF49. A resistance-nodulation-cell division (RND) efflux system is usually comprised of three proteins; CmeA (periplasm), CmeB (inner membrane) and CmeC (outer membrane) (Oh et al., 2014). However, only CmeA and CmeB are present in both G22 and BF49. An outer membrane lipoprotein involved in an RND efflux system, NodT is present in both isolates.

5.3.13 Plasmid

Plasmid replication genes *repABC* and *parAB* are only found on the G22 plasmid and not on either the G22 or BF49 chromosomes. The *trb* operon for conjugal transfer consists of 12 genes; *traI, trbBCDEJKLFGHI* (Li et al., 1999). From this operon, 10 out of the 12 genes are present on the G22 chromosome; only *trbH* and *traI* are missing. Conjugal transfer genes *trbBCDEFGIJL* are absent from BF49. Conjugative transfer DNA nicking endonuclease genes *traR*/*traO* are present only on the G22 plasmid. The other genes which were unique to the plasmid were two genes involved in purine utilisation (*yagS* and a putative xanthine transporter) and one for osmotic stress (*opgC*). The purine utilisation gene, *yagS,* encodes a periplasmic aromatic aldehyde oxidoreductase. YagS is usually part of an operon *yagTSRQ* which encodes a molybdenum-containing iron sulphur flavoprotein, where YagS is the FAD-containing subunit. The role proposed for this protein is detoxification of aromatic aldehydes (Xi et al., 2000, Neumann et al., 2009). The flavoprotein produced from *yagTSRQ* shows homology with xanthine dehydrogenase (Neumann et al., 2009). The osmotic stress gene, *opgC,* is involved in the synthesis of osmoregulated periplasmic glucans (OPGs). The exact role of OPGs is not understood however they have been shown to potentially play a role in the interaction between bacteria and their eukaryotic host (Roset et al., 2006, Kawaharada et al., 2010). The OpgC protein has been examined in *Rhodobacter sphaeroides* and was shown to encode a succinyltransferase homolog involved in the succinyl modification of OPGs (Roset et al., 2006, Kawaharada et al., 2010). The functions of other genes contained on the plasmid are also on the chromosome of both G22 and BF49 including genes for DNA ligases, DNA repair, cAMP signalling and RNA processing and modification.

5.4 Conclusions

5.4.1 Heterogeneity of the *Bradyrhizobium* genomes

Bradyrhizobium have been isolated from nodules of a range of host plants globally as previously described in the General Introduction and Chapter 3. The closest strains for the 16S sequence from G22 and BF49 are strains from Spain, New Zealand and Australia highlighting the cosmopolitan nature of the genus. The strains described here are the first to be isolated and genome-sequenced from European soil and are unique compared to other completed *Bradyrhizobium* genomes due to the absence of previously characterised genes and gene clusters for symbiosis, nitrogen fixation and photosynthesis. They are also distinct from the North American forest isolates as G22 and BF49 contain genes for denitrification. The presence or absence of the terminal denitrification gene, *nosZ,* may determine whether the end product of denitrification is the potent greenhouse gas, nitrous oxide or the less problematic nitrogen gas. However, the release of nitrous oxide is a loss of nitrogen from soil which is of major concern to farmers. The genome analyses indicate that the free-living soil *Bradyrhizobium* have the potential to carry out many degradative and transformative functions in soil; the marked differences between two isolates from comparable soils that have undergone different management indicates that they form part of an extremely heterogeneous group.

5.4.2 Ecological implications of the genome sequences

The genomes show evidence of potential ecological differences between the isolates including carbon metabolism pathways, nitrogen cycling and bacteriocin production. A number of genes involved in a range of carbon metabolism pathways are only present in the grassland isolate which could be a result of adaptation to its environment. The grassland plots contain a higher amount of organic carbon when compared to the bare fallow plots (Hirsch et al., 2009). Pathways involving malonic acid, N-acetyl-glucosamine, PTS kinases involved in phosphorylation of sugars and methane metabolism genes are only present in the G22 genome. In addition to differences in carbon metabolism, there is a key difference in genes associated with denitrification. The bare fallow isolate, BF49, contains the gene required for the final stage of denitrification; the conversion of nitrous oxide to nitrogen, where the grassland isolate was missing this gene. The potential for these isolates to be involved in the metabolism of methane and nitrous oxide have potential implications for global greenhouse gas emissions. The grassland isolate, G22, also contained genes required for the production of marinocine, a broad spectrum antibacterial

compound. These genes were missing from the bare fallow genome suggesting that the grassland isolate could require these genes to gain a competitive advantage over other members of the grassland microbial community.

5.4.3 Future exploitation of the *Bradyrhizobium* genomes

The genomes have highlighted that there are differences in presence or absence of genes involved in nitrogen and carbon cycling. In order to confirm that these genes are functional and are changing which pathways the isolates are involved in functional assays need to be carried out. Denitrification assays and carbon metabolism assays are the next steps to be taken to assess functional differences in the culture collection. Both carbon metabolism assays and nitrogen cycling are examined in this thesis in Chapter 6 and Chapter 7 respectively. The genomes could also be used to mine the metagenomes described in the General Introduction (Section 1.9) to determine whether patterns seen in the culture collection are representative of the *Bradyrhizobium* population as a whole. The genomes could also be used as reference genomes if more isolates from Highfield or Broadbalk are sequenced.

6 Carbon metabolism and growth assays of *Bradyrhizobium*

6.1 Introduction

6.1.1 Carbon in the environment

Organic carbon content is a fundamental influence on other soil properties (biotic and abiotic) and soil functioning (Joergensen and Scheu, 1999). Organic carbon content has been assessed under a range of treatments on the Highfield experiment (Study sites described in Chapter 1, Section 1.9 and Chapter 3, Section 3.2.1) and it was found that when compared to grassland, bare fallow soil contained 82% less organic carbon and arable soil contained 70% less (Hirsch et al., 2009). Organic carbon present in bare fallow soils is extremely stable and microbial communities are able to adapt to this for survival (Nunan et al., 2015). Maintaining plant-free soil through tilling releases large amounts of carbon and reducing tilling regimes in agriculture has been targeted as a way of reducing carbon loss from soils (Bird et al., 2011, Sun et al., 2014). The soil stores twice as much carbon as is in the atmosphere and three times as much as is stored in vegetation (Ilstedt and Singh, 2005, Guenet et al., 2011). Higher concentrations of carbon in arable soils compared to bare fallow soils has also led to higher mineralisation rates in arable soils (Guenet et al., 2011). Organic farming practices have been shown to sequester fourteen times the amount of carbon compared to conventional farming systems and systems which limit fertiliser inputs (Kong et al., 2011).

Plants are rarely carbon limited as they fix their own carbon through photosynthesis and most of this carbon is ultimately released via soil processes (Hogberg and Read, 2006, Balasooriya et al., 2013, Sagova-Mareckova et al., 2015). Approximately 11% of all carbon fixed by plants is released as root exudates (Lakshmanan et al., 2014). Plant roots release a range of compounds into the rhizosphere in combinations and proportions specific to the plant species and these include sugars, amino acids, phenolics and organic acids, in addition to releasing a number of other carbon sources through cell death and other rhizodeposits (Marschner et al., 2001, Lakshmanan et al., 2014, Zhang et al., 2014a, Kumar et al., 2016, Lareen et al., 2016). Heterotrophic soil microorganisms rely on this release of carbon from the plant roots and these carbon inputs from plants are vital for biological processes within soil (Hirsch et al., 2009, Mikola et al., 2014, Bell et al., 2015). Heterotrophic bacteria are important in cycling carbon and carbon is one of the main factors in soil which is limiting to microbial growth (Alden et al., 2001, Ekblad and Nordgren, 2002, Graneli et al., 2004, Ilstedt and Singh, 2005, Hogberg and Read, 2006). Gram-negative bacteria

usually dominate in the rhizosphere and utilise simple carbon substrates like root exudates (Balasooriya et al., 2013, Mellado-Vazquez et al., 2016). Carbon utilisation is often a way of classifying organisms (Koch, 2001). Copiotrophs have been shown to utilise more labile forms of carbon while oligotrophs are more likely to preferentially utilise more complex forms of carbon (Koch, 2001, Fierer et al., 2007). Looking at functional diversity with respect to C use is important to get a better understanding of the role of microbial communities in determining the dynamics of C turnover in soil (Garland and Mills, 1991, Preston-Mafham et al., 2002).

6.1.2 Biolog MicroPlates™

The Biolog™ system allows for phenotypic analysis of bacterial and fungal isolates by testing metabolism of a range of carbon substrates within a 96-well plate, producing a metabolic fingerprint of the isolates (Garland and Mills, 1991, Campbell et al., 1997). Each well contains a redox tetrazolium dye which is reduced during microbial metabolic activity, forming formazan. This creates a colour change and turns purple if the substrate is metabolised, to provide a colorimetric measure of metabolism (Guckert et al., 1996, Campbell et al., 1997, Preston-Mafham et al., 2002). The system allows for a rapid assessment of substrates able to be utilised and is used for identification of specific strains of bacteria. It is also often used to assess community composition and functional differences between microbial communities (Garland and Mills, 1991, Campbell et al., 1997).

There are a range of MicroPlates available tailored for both Gram-positive and Gram-negative isolates (GENIII), Gram-negatives only (GN2), Gram-positives only (GP2), yeasts (YT), filamentous fungi (FF) and anaerobic microorganisms (AN). All of these plates contain 95 carbon sources tailored to the microorganisms being tested. The Biolog EcoPlate™ is specifically designed for environmental community analysis and contains 31 carbon sources replicated three times on a single 96-well plate. There are also MT2 plates in which each well contains only the tetrazolium redox dye and a buffered medium and it is left for the user to add the substrates to be tested (Biolog Inc, USA).

GN2 plates were chosen as *Bradyrhizobium* is Gram-negative and they contained many of the substrates in the EcoPlates which are designed for environmental samples but allowed for three times as many carbon sources to be analysed to give a broader view of carbon metabolism.

6.1.3 Experiments using Biolog™

The Biolog system is commonly used for isolate characterisation as Biolog has a database for quick identification based on Biolog response and includes medical, veterinary and environmental microorganism profiles (Biolog Inc, USA). Community analysis using Biolog is widely used. For example, Biolog MicroPlates were used to distinguish between different grassland soils and it was found that the grassland community was able to readily metabolise carbohydrates, amino acids and carboxylic acids. The Biolog plates were sensitive enough to distinguish between soil samples which were all grasslands but had varying plant communities (Campbell et al., 1997). It has also been used to distinguish between aquatic, bulk soil and rhizosphere microbial communities and it was found that these samples had differing metabolic fingerprints and were able to be distinguished from their carbon metabolism profile (Garland and Mills, 1991). The rhizosphere contains a much larger quantity and diversity of organic carbon than bulk soil and as a result the average well colour development was significantly higher for rhizosphere samples than bulk soil. Zero tillage also was shown to increase the number of carbon sources able to be used compared to conventional till practices (Yang et al., 2013). Guenet et al. (2011) tested the ability of microbial communities from long-term bare fallow to degrade fresh organic matter. The bare fallow communities had retained the ability to degrade carbon sources in the Biolog EcoPlate™ and had a high diversity index. It was predicted that the metabolic capabilities of bare fallow communities have been maintained due to decomposition of microbial cells in the absence of C inputs (Guenet et al., 2011). Agricultural management impacts on the microbial community have also been assessed using Biolog. For example, the application of rice straw resulted in an increased metabolic potential of the microbial community (Peng et al., 2016). Also, long-term bare fallow and arable microbial communities have been compared using EcoPlates. This resulted in bare fallow communities preferentially metabolising simpler carbon sources which require less energy expenditure over more complex carbon sources. It was concluded that the bare fallow communities are better adapted to utilising these substrates compared to arable communities (Nunan et al., 2015).

Carbon metabolism of *Bradyrhizobium* has also been previously tested although much of this work has focussed on symbiotic isolates. Carbon metabolism assays of symbionts of *Vicia* spp*., Medicago sativa, Melilotus* sp., *G. max* and *Lotus corniculatus* were used to cluster the isolates and resulted in three clusters; *Rhizobium, Sinorhizobium* and *Mesorhizobium*/*Bradyrhizobium* (Zabaloy and Gomez, 2005). The carbon metabolism profiles have also been used to assess the similarity of *Acacia albida* symbionts and identified five clusters based on carbon usage (Dupuy et

al., 1994). Symbionts of *Acmispon strigosus* have also been characterised using Biolog and it was found that carbohydrates and carboxylic acids were readily utilised and that this could be a strategy for surviving in soil, being able to successfully compete within the rhizosphere and to cope with a carbon rich environment found inside the nodule (Hollowell et al., 2016b). Biolog plates were also used to determine the roles of isocitrate lyase (*aceA*) and nitrogenase (*nifD*/*nifH*) in chemoautotrophic growth in *B. japonicum* USDA 110 wild-type and mutant strains (Franck et al., 2008). To date, to our knowledge, a culture collection dominated by free-living, non-symbiotic, *Bradyrhizobium* strains has not been tested using Biolog MicroPlates.

6.1.4 Limitations of the Biolog™ system

A number of limitations are related to the Biolog MicroPlate system. For example, the majority of soil microorganisms are not able to be cultured and may not grow in the Biolog plates meaning that community that responds to the C sources presented in the Biolog plate may be not representative of the original soil community (Janssen et al., 2002, Preston-Mafham et al., 2002, da Rocha et al., 2009). The removal of soil particles can also lead to the death of particular members of the community and so alters the community composition. The rate of colour development also does not necessarily correspond to cell density as some members of the community may be more active than others or biomass could be higher in one soil compared to another. In addition, in a mixed species sample, interactions between members could lead to the tetrazolium dye being reduced due to the use of waste products rather than the substrate being tested (Preston-Mafham et al., 2002).

The Biolog MicroPlates allow for rapid assays of metabolism of a number of carbon sources. Although some microbial growth within the wells occurs, it is the colour change caused by the reduction of the tetrazolium redox dye which causes the increase in OD that is detected in the assay. Fewer problems are associated with assessing single species metabolism than communities (Preston-Mafham et al., 2002). *Bradyrhizobium* is able to be cultured in a 96-well plate and so the Biolog GN2 MicroPlates provided a sound method for assessing carbon metabolism capabilities of a range of isolates from varying soil treatments.

6.1.5 Analysis of Biolog™ data

Many previous studies have used Biolog MicroPlates to assess and compare community diversity requiring cell density correction before inoculation of the plates. Average well colour development (AWCD) is one way that the OD reading can be standardised by taking the average

OD of all 95 response wells (Garland and Mills, 1991, Guckert et al., 1996, Yang et al., 2013). One advantage of using AWCD is that different samples can be compared which have different inoculum density allowing for community comparisons. For this experiment, the cultures were diluted to the same cell density before inoculating the Biolog plates as only a single isolate was being tested in each plate and therefore, standardisation though the calculation of AWCD was not required. The Biolog output is a curve for each well which represents the colour development (caused by the formation of the reduced tetrazolium product) over time. The area under the curve has also been used with Biolog data as it has been argued that this takes into account lag phase, maximum OD and rate of usage and is a better method of analysis than taking individual time points (Guckert et al., 1996). However, two curves can give the same area under the curve where the shape of the curve is drastically different. This information could be masked if only considering the area under the curve. Multivariate analysis allows for the profiles of carbon metabolism to be assessed in addition to highlighting key substrates which make the largest contribution to the principal components through examining the loadings matrices.

6.1.6 *Bradyrhizobium* growth

Growth rates are often used to assess fitness of particular strains and phenotypic properties (Hall et al., 2014). A key characteristic of the *Bradyrhizobium* genus is slow growth, taking typically seven days to form visible colonies when grown on agar; this formed a large part of the isolation method discussed in Chapter 3 (Fuhrmann, 1990, Streeter, 2003, Okubo et al., 2012). We predicted that growth rate could influence the carbon metabolism response as slower growing isolates will have higher energy requirements than faster growing strains and will utilise the substrates to a higher level and therefore growth rate needed to be assessed. Conversely, the capacity to use certain substrates efficiently might influence the growth rate.

6.1.7 Aims

In this chapter, I aim to:

- Assess the carbon metabolism capability of a subset of the *Bradyrhizobium* culture collection
- Determine whether the isolates from a particular soil treatment cluster according to their carbon metabolism profile
- Assess growth rate differences of the entire *Bradyrhizobium* culture collection
- Verify if any patterns seen in the carbon metabolism profiles correspond with growth rate

6.2 Materials and methods

6.2.1 Growth assays of *Bradyrhizobium*

All isolates in the culture collection (Detailed in Chapter 3) were grown in MAG broth as described in General Methods 2.1 and 2.2 for 6 days. Each well of a 96-well plate contained 140 µl of MAG broth before 20 µl of bacterial culture was inoculated. Three replicate plates were set up and the position of the isolates was randomised for each plate to remove bias associated with well position. The optical density (OD) was read using a Varioscan SkanIt plate reader (Thermo Fisher Scientific Inc. , USA) at 590 nm and 25^{o}C every 24 hours for a total of 98 hours (0, 24, 48, 72 and 98 hours). The plates were incubated at 25^{o}C and shaken at 100 rpm.

6.2.2 Statistical analysis of growth assays

All data as normalised by subtracting the average OD of the blank (no bacteria) wells for each replicate from the inoculated wells. Growth rate was calculated by subtracting the OD at 0 hours from the OD at 72 hours and then dividing by 3 to give the increase in OD per day. By 98 hours, some wells had begun precipitating and so the 72-hour time point was used to calculate the growth rate. The GrowthRates program was used to identify maximum OD for each well (Hall et al., 2014) (Appendix G). The average of the blank wells was calculated to be 0.042 and this was set as the background reading when GrowthRates was run. The calculated rates and maximum OD were tested using ANOVA and Tukey's HSD. Fligner-Killeen tests were carried out to test for homogeneity of variances. Both individual isolates and isolates grouped by soil treatment were tested. To test effect of treatment, growth rates and maximum OD values were averaged for each isolate before grouping by treatment. This was to ensure that the data points used in the ANOVA were independent. Growth curves are provided in Appendix H. Tukey's HSD outputs are provided in Appendix I.

6.2.3 Carbon metabolism profiling using Biolog GN2 MicroPlates™

The ability of isolates to metabolise different carbon substrates was assessed using Biolog GN2 MicroPlates™ (Biolog Inc., USA), a full list of substrates is provided in Table 6.1. A selection of isolates from the culture collection chosen to represent each of the soils treatments; bare fallow (BF6, BF49 and BF41), grassland (G22, G15 and G54), arable (A2, A5 and A16) and woodland (W22, W49 and W24) and a symbiotically-effective type strain, *B. japonicum* USDA 6 were grown

in MAG broth for 7 days as described in General Methods 2.1 and 2.2. Cell density was estimated from 1 ml of culture stained with 0.05% methylene blue using a haemocytometer. The cultures were diluted to a cell density of 10^6 μl^{-1}. The diluted cultures were centrifuged at 14000 xg for 1 minute, the supernatant was removed and the cells re-suspended in sterile deionised water. Each well of a MicroPlate was inoculated with 140 µl (10^8 cells/ 140 µl) of bacterial culture including the water control well. Three replicate plates per isolate were used. The OD was read as described in Section 6.2.1 over a total of 98 hours. The plates were incubated at 25^{o}C and shaken at 100 rpm.

6.2.4 Statistical analysis of Biolog™ data

The Biolog data was analysed using principal components analysis (PCA) using the inbuilt function, *prcomp*, in R (version 3.2.2). Each substrate was considered a variable. The PCA was run for each time point and ignoring time as a factor and analysing the entire dataset at once. For visualisation of the PCA, biplots were drawn using PC1 and PC2 and the loadings matrices were extracted for each time point individually and all time points together (Appendix J for loadings matrices and Appendix L for PCA plots for 0, 24, 48 and 72 hours). The loadings matrices identified specific substrates which were associated with the isolates and also substrates making the greatest contribution to the principal component. The substrates were also grouped into carbohydrates, carboxylic acids, amines and amides, amino acids, polymers and miscellaneous according to guilds identified in previously published work (Preston-Mafham et al., 2002). A full list of the guilds can be found in Table 6.1. Biolog utilisation curves for all substrates can be found in Appendix K. Substrates not able to be utilised by any isolates were removed from the dataframe before PCA analysis. These substrates are summarised in Section 6.3.2.

Analysis of variance (ANOVA) was carried out to test for difference in the OD in the water well between the isolates tested. A significant difference indicates different OD values for each isolate in the absence of a carbon source and therefore can be considered background. The data was normalised by subtracting the water well at each time point before PCA was completed. There was a highly significant difference between the isolates in the water well and for all time points; 0 hours (ANOVA, F=39.95, df=38, P<0.01), 24 hours (ANOVA, F=40.24, df=38, P<0.01), 48 hours (ANOVA, F=45.44, df=38, P<0.01), 72 hours (ANOVA, F=45.91, df=38, P<0.01) and 98 hours (ANOVA, F=59.92, df=38, P<0.01). All data was normalised before analysis by subtracting the OD for the water well at each time point as a result.

The OD at 98 hours and the entire dataset, ignoring time as a factor were converted to binary matrices where a positive response was denoted 1 and a negative response was denoted 0 (Appendix M). The response was considered positive if it was twice the OD of the water well as described in Hollowell et al. (2016b). The binary matrix was used to determine counts of substrates used at 98 hours by each isolate and substrate guild. Differences in substrate utilisation between each guild and soil treatment were tested using generalised linear models and Poisson errors. The residual deviance was larger than the residual degrees of freedom showing evidence of overdispersion and so quasipoisson errors were used. Model simplification was carried out as treatment showed significant interaction with guild resulting in no interaction term in the model (Statistics outputs listed in Appendix N).

6.2.5 Cluster analysis

This binary matrix was then used to create a dendrogram using the *pvclust* package in R (v 3.2.2) where relationships between carbon metabolism profiles could be assessed (Suzuki and Shimodaira, 2006). The UPGMA clustering method was used with 1000 bootstraps. The *P*-values provided are approximately unbiased (AU) *P*-value and the bootstrap *P*-value (BP).

Table 6.1: List of carbon substrates and position in the Biolog GN2 MicroPlates™

C= carbohydrate, CA= carboxylic acid, AM = amine/amide, AA = amino acid, P = polymer, M = miscellaneous.

Well		Substrate	Guild	Well		Substrate	Guild
A1	1	Water	-	A7	49	N-acetyl-D-galactosamine	C
B1	2	i-Erythritol	C	B7	50	m-Inositol	C
C1	3	D-melibiose	C	C7	51	Sucrose	C
D1	4	Acetic acid	CA	D7	52	D-gluconic acid	CA
E1	5	p-Hydroxy phenylacetic acid	CA	E7	53	Malonic acid	CA
F1	6	Bromo succinic acid	M	F7	54	L-alanyl-glycine	AA
G1	7	L-histidine	AA	G7	55	L-pyroglutamic acid	AA
H1	8	Urocanic acid	M	H7	56	2-aminoethanol	AM
A2	9	alpha-Cyclodextrin	P	A8	57	N-acetyl-D-glucosamine	C
B2	10	D-fructose	C	B8	58	alpha-D-lactose	C
C2	11	beta-Methyl-D-glucoside	C	C8	59	D-trehalose	C
D2	12	Cis-aconitic acid	CA	D8	60	D-glucosaminic acid	CA
E2	13	Itaconic acid	CA	E8	61	Propionic acid	CA
F2	14	Succinamic acid	AM	F8	62	L-asparagine	AA
G2	15	Hydroxy-L-proline	AA	G8	63	D-serine	AA
H2	16	Inosine	M	H8	64	2,3-butanediol	M
A3	17	Dextrin	P	A9	65	Adonitol	C
B3	18	L-fucose	C	B9	66	Lactulose	C
C3	19	D-psicose	C	C9	67	Turanose	C
D3	20	Citric acid	CA	D9	68	D-glucuronic acid	CA
E3	21	alpha-Keto butyric acid	CA	E9	69	Quinic acid	CA
F3	22	Glucuronamide	AM	F9	70	L-aspartic acid	AA
G3	23	L-leucine	AA	G9	71	L-serine	AA
H3	24	Uridine	M	H9	72	Glycerol	M
A4	25	Glycogen	P	A10	73	L-arabinose	C
B4	26	D-galactose	C	B10	74	Maltose	C
C4	27	D-raffinose	C	C10	75	Xylitol	C
D4	28	Formic acid	CA	D10	76	alpha-Hydroxy butyric acid	CA
E4	29	alpha-Keto glutaric acid	CA	E10	77	D-saccharic acid	CA
F4	30	L-alaninamide	AM	F10	78	L-glutamic acid	AA
G4	31	L-ornithine	AA	G10	79	L-threonine	AA
H4	32	Thymidine	M	H10	80	D,L-alpha-glycerol phosphate	M
A5	33	Tween 40	P	A11	81	D-arabitol	C
B5	34	Gentiobiose	C	B11	82	D-mannitol	C
C5	35	L-rhamnose	C	C11	83	Methyl pyruvate	C
D5	36	D-galactonic acid lactone	CA	D11	84	beta-Hydroxy butyric acid	CA
E5	37	alpha-Keto valeric acid	CA	E11	85	Sebacic acid	CA
F5	38	D-alanine	AA	F11	86	Glycyl-L-aspartic acid	AA
G5	39	L-phenylalanine	AA	G11	87	D,L-carnitine	AA
H5	40	Phenyethylamine	AM	H11	88	Glucose-1-phosphate	M
A6	41	Tween 80	P	A12	89	D-cellobiose	C
B6	42	alpha-D-glucose	C	B12	90	D-mannose	C
C6	43	D-sorbitol	C	C12	91	Mono-methyl-succinate	M
D6	44	D-galacturonic acid	CA	D12	92	gamma-Hydroxy butyric acid	CA
E6	45	D,L-lactic acid	CA	E12	93	Succinic acid	CA
F6	46	L-alanine	AA	F12	94	Glycyl-L-glutamic acid	AA
G6	47	L-proline	AA	G12	95	gamma-Amino butyric acid	AA
H6	48	Putrescine	AM	H12	96	Glucose-6-phosphate	M

6.3 Results

6.3.1 Growth curve analysis

6.3.1.1 Growth rate

Bare fallow isolate BF28 and grassland isolate G46 were removed from the analysis as they showed no signs of growth. Bare fallow isolate BF26 and grassland isolate G31 were also removed as replicate 2 was set up in error. Growth curves for all isolates are included in Appendix H.

Once the growth rates had been calculated, differences were tested using ANOVA to test for a difference between treatments and individual isolates. The Fligner-Killeen test between growth rate and treatment showed that the variances were equal and so satisfies this assumption of ANOVA (Chi-squared=7.5323, df=4, *P*=0.1103). The difference between the growth rates for isolates from each soil treatment (grassland, bare fallow, arable, woodland and symbiotic/reference) were shown to be highly statistically significant (ANOVA, *F*=6.414, df=4, *P*<0.001) (Figure 6.1). Post-hoc Tukey's HSD comparisons revealed that the growth rate for woodland isolates was significantly lower than isolates from all other soil treatments (*P*<0.001). All other comparisons were not significant. Tukey's HSD results can be found in Appendix I.

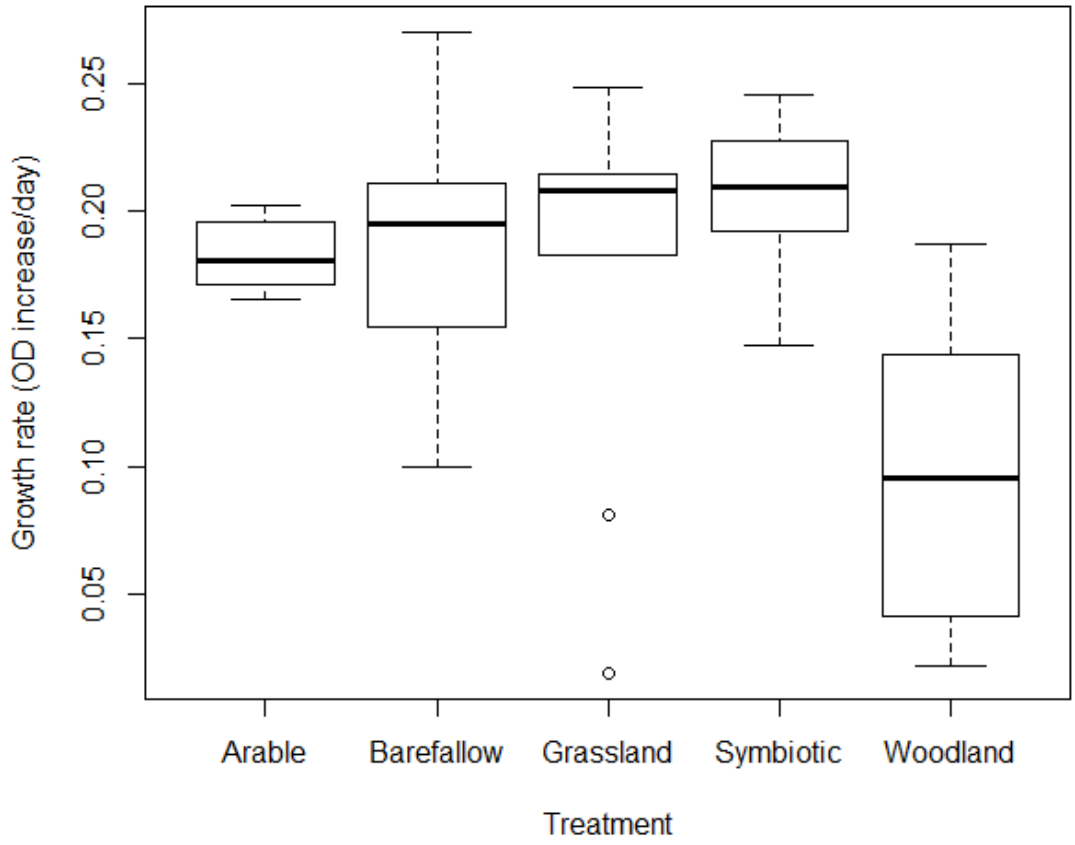


Figure 6.1: Boxplot showing the growth rate for each soil treatment.
The horizontal bar inside the boxes show the median and the whiskers show the minimum and maximum interquartile range.

The Fligner-Killeen test between growth rate and individual isolates showed that the variances were equal and so satisfies this assumption of ANOVA (Chi-squared=28.619, df=44, *P*=0.9648). There was also a significant difference for growth rate between the individual isolates (ANOVA, *F*=6.12, df=44, *P*<0.001) (Figure 6.2). However, with treatment groupings, the arable isolates had growth rates that did not differ from one another (*P*>0.05). The same pattern is seen with the symbiotic isolates. The bare fallow isolates are also consistent in growth rate with only BF52 being different from BF48 (Tukey's HSD, difference=0.17, *P*=0.009). The grassland isolates are more variable than the arable, bare fallow and symbiotic isolates for growth rate however the majority of these differences are due to G9 have an extremely low growth rate compared to other grassland isolates. G22 had a significantly slower growth rate than G8 (Tukey's HSD, difference=0.167, *P*=0.01) and G6 (Tukey's HSD, difference=0.164, *P*=0.02). The woodland isolates are most variable of all soil treatments when examining growth rate. W22 had a significantly higher growth rate when compared to W24 (Tukey's HSD, difference=0.16, *P*=0.03) and W24 had a significantly lower growth rate than W23 (Tukey's HSD, difference=0.152, *P*=0.046). W9 and W24 showed extremely low growth rates (Figure 6.2).

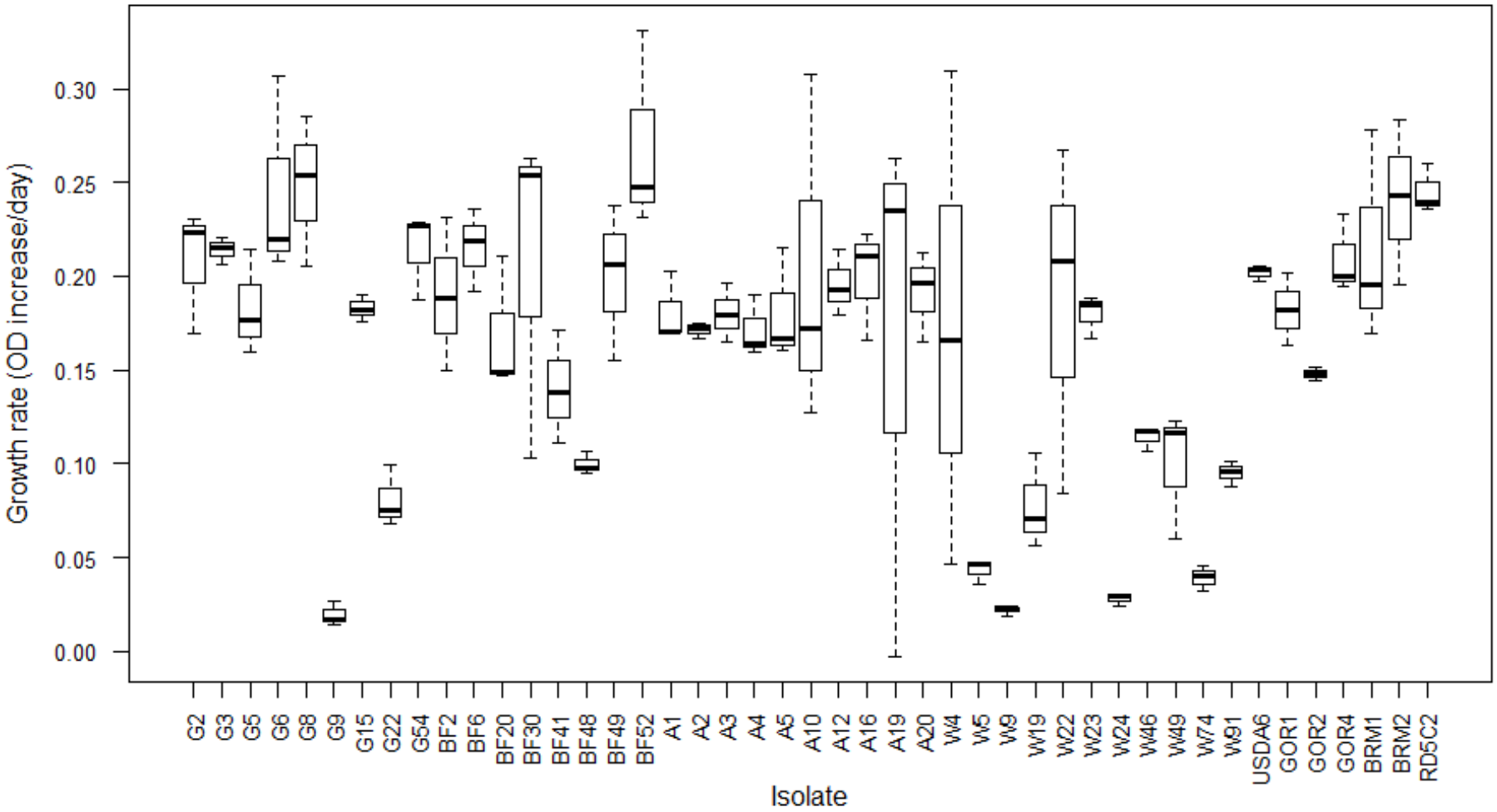


Figure 6.2: Boxplot showing the growth rate for each isolate.

The horizontal bar inside the boxes show the median and the whiskers show the minimum and maximum interquartile range.

6.3.1.2 Maximum OD

The Fligner-Killeen test between maximum OD and soil treatment showed that the variances were equal and so satisfies this assumption of ANOVA (Chi-squared=7.5975, df=4, *P*=0.1075).The maximum OD value reached within 98 hours shows a significant difference between the isolates according to soil treatments (ANOVA, *F*=6.432, df=4, *P*<0.001) (Figure 6.3). The woodland treatment was significantly different from grassland (Tukey's HSD, difference=-026, *P*=0.01), arable (Tukey's HSD, difference=-0.29, *P*<0.01), bare fallow (Tukey's HSD, difference=-0.31, *P*<0.01) and symbiotic treatments (Tukey's HSD, difference=-0.33 *P*<0.01). In all of these cases, the woodland treatment had significantly lower maximum OD. No other pairwise comparisons of soil treatment were significant. Full Tukey's HSD output is provided in Appendix I.

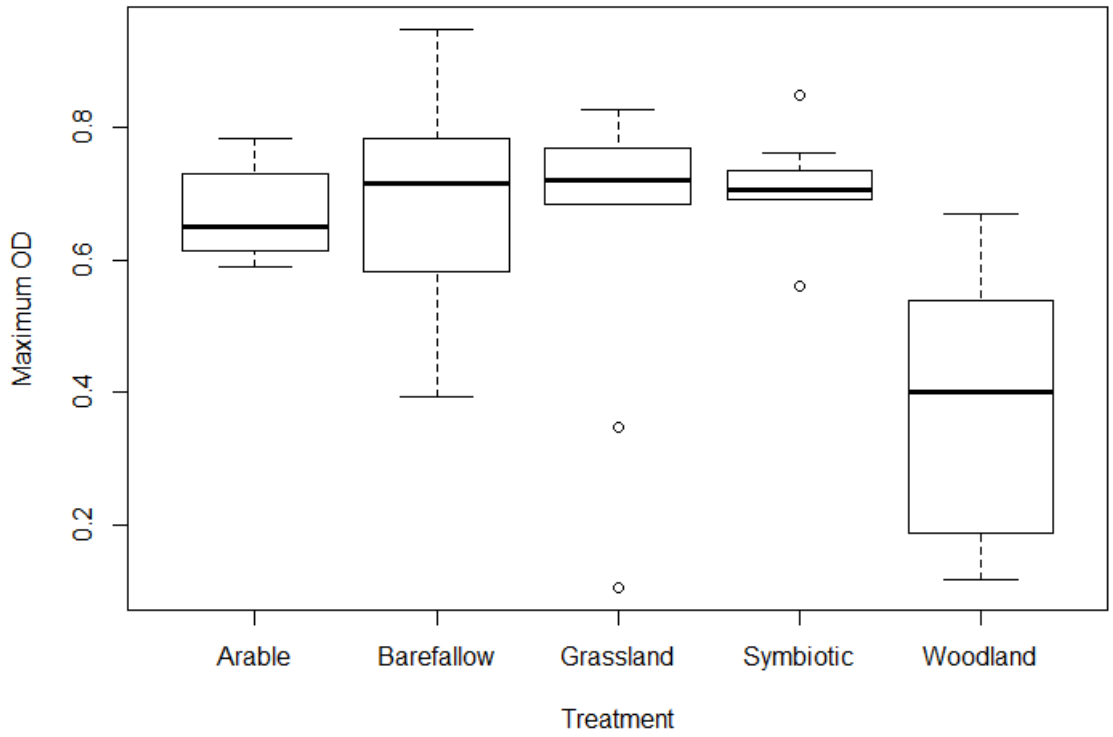


Figure 6.3: Boxplot showing the maximum OD for each soil treatment.
The horizontal bar inside the boxes show the median and the whiskers show the minimum and maximum interquartile range.

The Fligner-Killeen test between maximum OD and individual isolates showed that the variances were equal and so satisfies this assumption of ANOVA (Chi-squared=28.795, df=44, *P*=0.9628).There is also a significant difference in the maximum OD between the individual isolates (ANOVA, *F*=5.808, df=44, *P*<0.001) (Figure 6.4). The grassland isolate G9 had significantly lower OD than a number of other grassland isolates. No other pairwise comparisons between grassland isolates were significant. No significant differences were found within the arable isolates or within the symbiotic isolates. The woodland isolates were all statistically similar except for the maximum OD for W22 and W24 which was significantly different with W22 being higher than W24 (Tukey's HSD, difference=0.51, *P*=0.04).

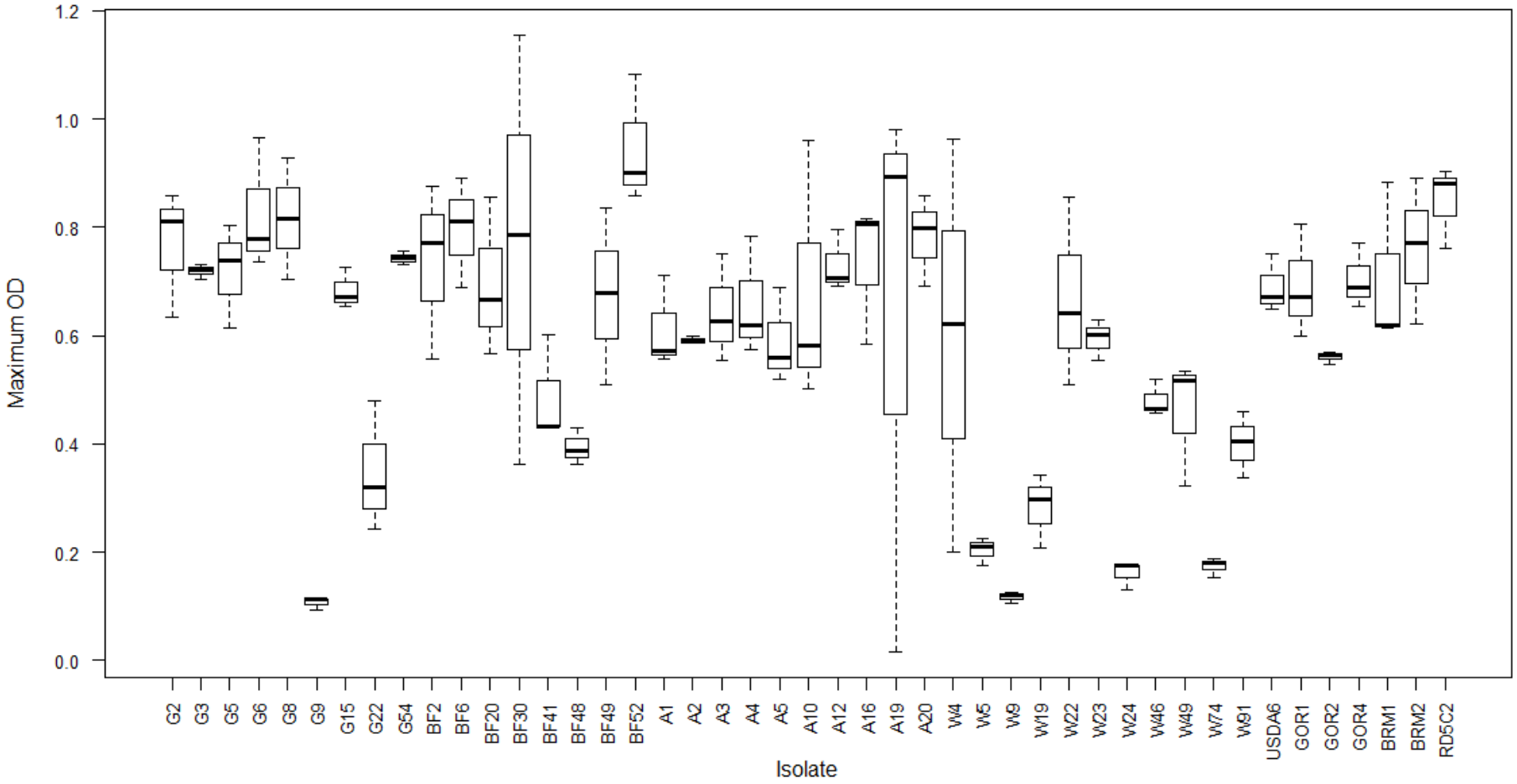


Figure 6.4: Boxplot showing the maximum OD for each isolate.

The horizontal bar inside the boxes show the median and the whiskers show the minimum and maximum interquartile range.

6.3.2 Carbon metabolism

Substrates which were universally not used across all isolates were removed from the PCA analysis. A substrate was considered not utilised if the OD at 98 hours was less than double the OD of the water well at 98 hours. These substrates are listed in Table 6.2.

Table 6.2: Summary of which substrates were able to be metabolised

Substrates not used by any isolate		Substrates used by at least one isolate	
• i-Erythritol	• D-trehalose	• Acetic acid	• D-sorbitol
• D-melibiose	• D-serine	• p-Hydroxy phenylacetic acid	• D-galacturonic acid
• L-histidine	• 2,3-butanediol	• Bromo succinic acid	• D,L-lactic acid
• Urocanic acid	• Lactulose	• D-fructose	• L-alanine
• α-Cyclodextrin	• Turanose	• Cis-aconitic acid	• L-proline
• β-Methyl-D-glucoside	• L-serine	• Itaconic acid	• D-gluconic acid
• Hydroxy-L-proline	• Xylitol	• Succinamic acid	• Malonic acid
• Inosine	• D,L-α-glycerol phosphate	• L-fucose	• L-alanyl-glycine
• Dextrin	• D,L-carnitine	• Citric acid	• L-pyroglutamic acid
• D-psicose	• Glucose-1-phosphate	• Glucuronamide	• D-glucosaminic acid
• α-Keto butyric acid	• D-cellobiose	• L-leucine	• Propionic acid
• Uridine	• ɣ-Amino butyric acid	• D-galactose	• L-asparagine
• Glycogen	• Glucose-6-phosphate	• Formic acid	• Adonitol
• D-raffinose	• N-acetyl-D-galactosamine	• α-Keto glutaric acid	• D-glucuronic acid
• L-ornithine	• m-Inositol	• L-alaninamide	• Quinic acid
• Thymidine	• Sucrose	• Tween 40	• L-aspartic acid
• Gentiobiose	• 2-aminoethanol	• L-rhamnose	• Glycerol
• Phenyethylamine	• N-acetyl-D-glucosamine	• D-galactonic acid lactone	• L-arabinose
• Putrescine	• α-D-lactose	• α-Keto valeric acid	• Maltose
		• D-alanine	• α-Hydroxy butyric acid
		• L-phenylalanine	• D-saccharic acid
		• Tween 80	• L-glutamic acid
		• α-D-glucose	• L-threonine
		• β-Hydroxy butyric acid	• D-arabitol
		• Sebacic acid	• D-mannitol
		• Glycyl-L-aspartic acid	• Methyl pyruvate
		• D-mannose	• ɣ-Hydroxy butyric acid
		• Mono-methyl-succinate	• Succinic acid
		• Glycyl-L-glutamic acid	

6.3.2.1 All time points

The principal components analysis (PCA) biplot (Figure 6.5-A) shows that there is a separation of the time points across PC1 and PC2 which account for 58.43% and 11.7% of the variation respectively. The first two principal components were visualised as together they accounted for 70.13% of the variation. The third PC accounted for 7.78%. Figure 6.5-B shows the PCA biplot colour coded according to soil treatment. The woodland isolates (blue) showed no separation across PC1 or PC2 and were clustered together. The bare fallow isolates (red) separated across PC1 but not PC2. The grassland isolates (green) clustered together and showed no separation across PC2 and small amounts across PC2. The arable isolates (orange) separated across PC2 and a small amount across PC1. The USDA 6 (purple) clustered separately from the grassland and bare fallow isolates across PC1 and PC2.

Figure 6.6 shows the 57 utilised substrates colour coded according to guild. More carbohydrates are associated with separation across PC1; this suggests that carbohydrates are more associated with the separation of the bare fallow isolates. Carboxylic acids are more associated with separation across PC2; this associates more closely with the separation of arable isolates and the symbiotic isolate USDA6 from the grassland and bare fallow isolates (Figure 6.5-A & Figure 6.6).

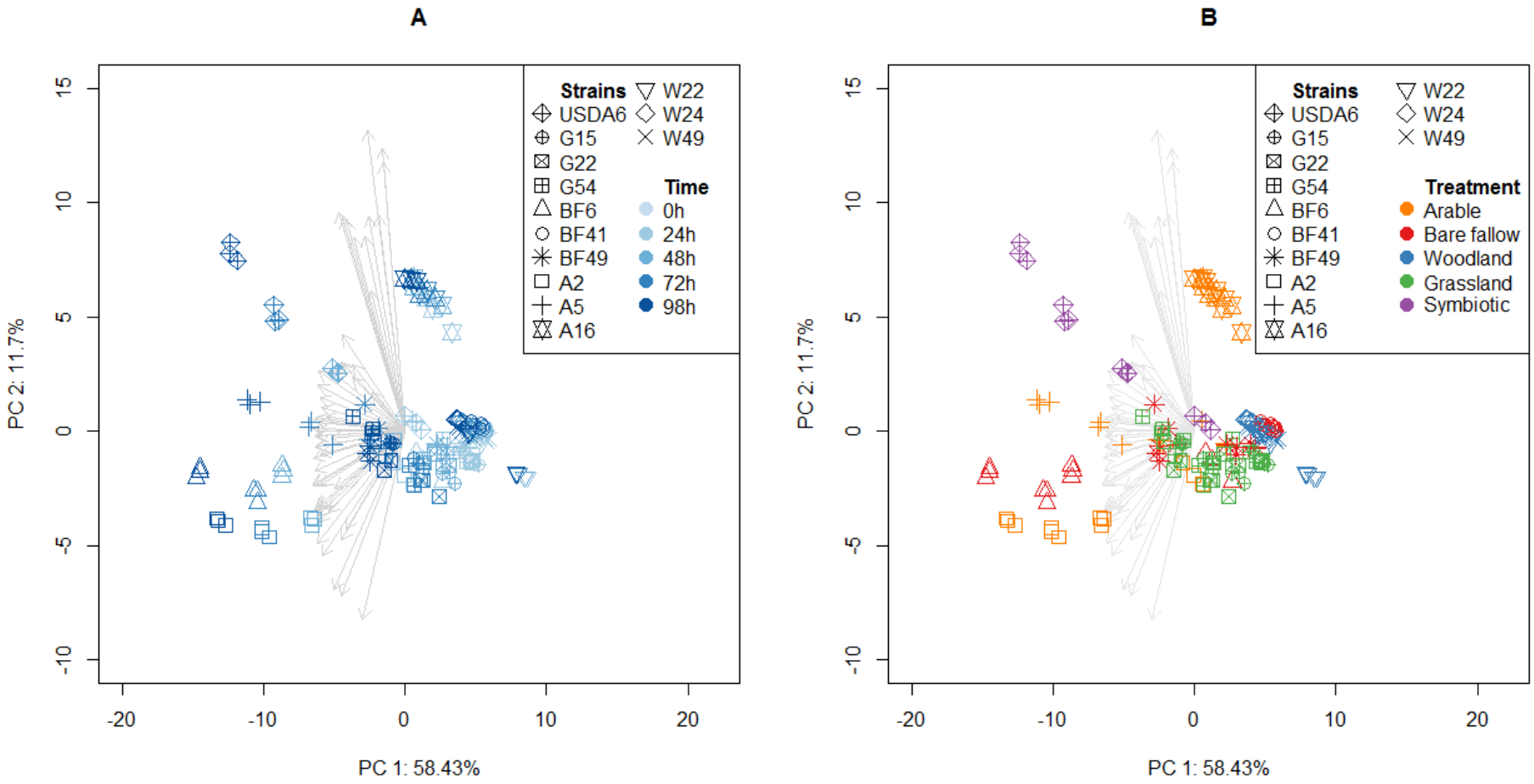


Figure 6.5: PCA biplots for all 95 substrates showing PC1 and PC1 accounting for 70.13% of variation.
A is colour coded according to time point and B is colour coded according to soil treatment.

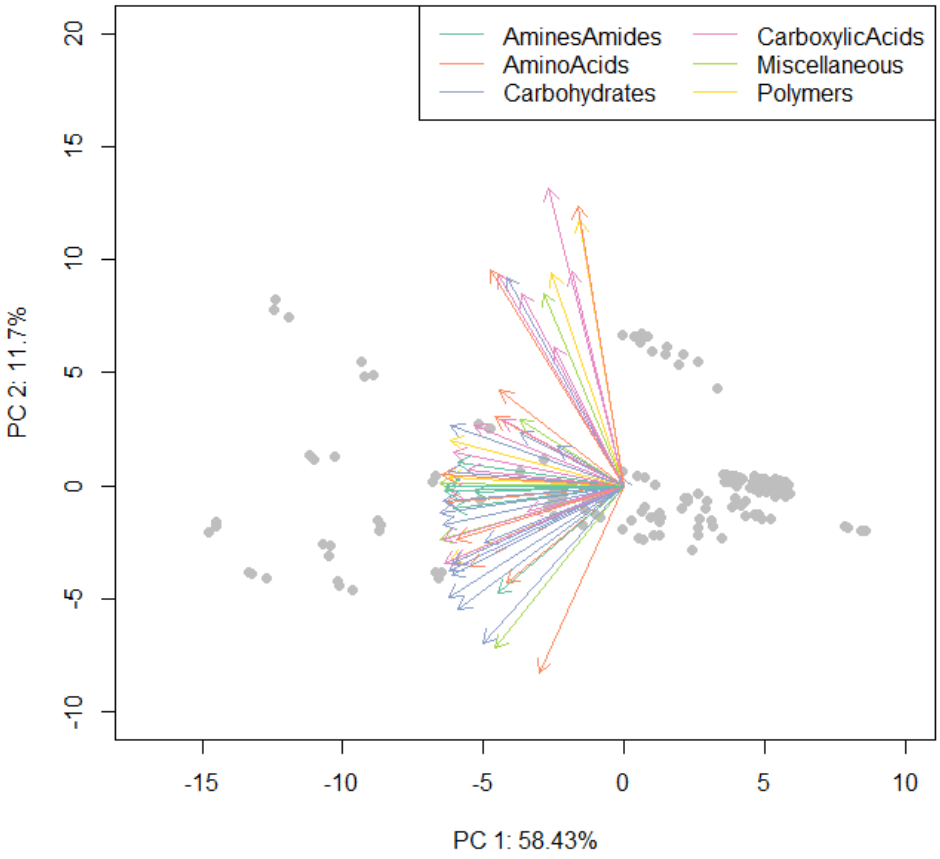


Figure 6.6: PCA biplot for all 95 substrates for PC1 and PC2 accounting for 70.13% of variation.
The arrows represent each substrate and are colour coded according to guild. More carbohydrates are associated with separation across PC1.

The highest loadings and therefore the substrates which make the largest contribution to PC1 were α-D-glucose, succinamic acid and methyl pyruvate. Alpha-D-glucose is a carbohydrate involved in a number of pathways including glycolysis, fructose and mannose metabolism, starch and sucrose metabolism and amino sugar and nucleotide sugar metabolism (Kanehisa and Goto, 2000, Kanehisa et al., 2016). All isolates were able to use the substrate except for the woodland isolates and BF41 which showed low levels of metabolism in most substrates (Figure 6.7-A). Succinamic acid is a dicarboxylic acid monoamide which is also known as 4-amino-4-oxobutanoic acid (Hastings et al., 2013). It is a metabolite in central metabolism. Figure 6.7-B shows the Biolog curve for succinamic acid. All isolates except BF41 and W49 were able to utilise this substrate. Methyl pyruvate is a carbohydrate which is also known as methyl 2-oxopropanoate (Hastings et al., 2013). It is involved in core metabolism and ATP production. All isolates were able to use this substrate (Figure 6.7-C). A summary of substrates with the highest loadings for PC1 and PC2 can be found in Table 6.3. Biolog utilisation graphs for all substrates can be found in Appendix K.

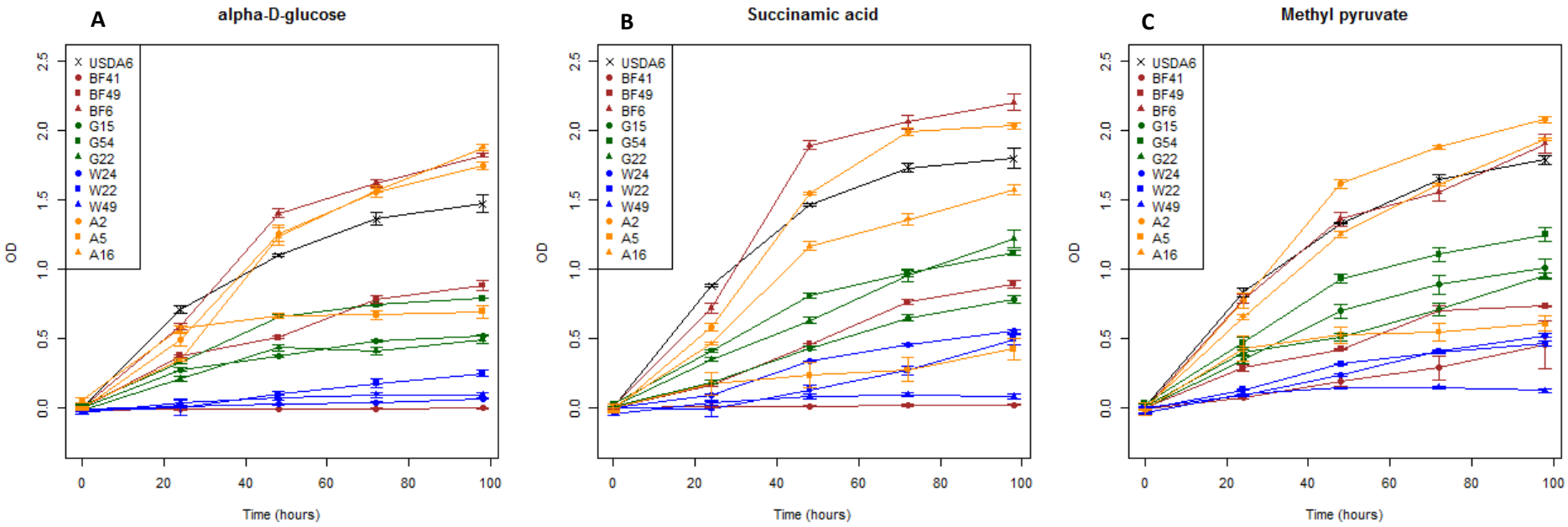


Figure 6.7: Biolog curves for substrates with the highest loadings for PC1 and all time points.
A: α-D-glucose had the highest loading for PC1, B: Succinamic acid had the second highest loading for PC1, C: Methyl pyruvate had the third highest loading for PC1. Error bars are standard error.

6.3.2.2 Time 98 hours

The principal components analysis (PCA) biplot (Figure 6.8-A) shows that there is a separation of the isolates across PC1 and PC2 which account for 59.46% and 11.87% of the variation respectively. The first two principal components were visualised as together they accounted for 71.33% of the variation. The third PC accounted for 7.24%. Figure 6.8-A shows the PCA biplot colour coded according to soil treatment. The bare fallow isolates are all separated from each other across PC1 but not PC2. The grassland isolates are clustered together and are not separated across PC1 or PC2. The arable isolates are separated from each other across PC1 and PC2. A2 and A5 are separated across PC2 but not PC1. The woodland isolates are not separated across PC1 or PC2. The grassland isolates cluster separately from the symbiotic isolate USDA 6 across both PC1 and PC2. USDA 6 is also distinct from A2 and A5 across PC2 and A16 across PC1. USDA 6 is also distinct from BF41 and BF49 across PC1 and BF6 across PC2.

Figure 6.8-B shows the 57 substrates colour coded according to category. Carboxylic acids are more associated with separation across PC2; the separation of A2 and A5, the grassland isolates and the separation of USDA 6 from BF6, A2 and A5. All other guilds are more likely to be associated with the separation of PC1; the separation of the bare fallow isolates and the separation of USDA 6 from grassland isolates, A16, BF41 and BF49.

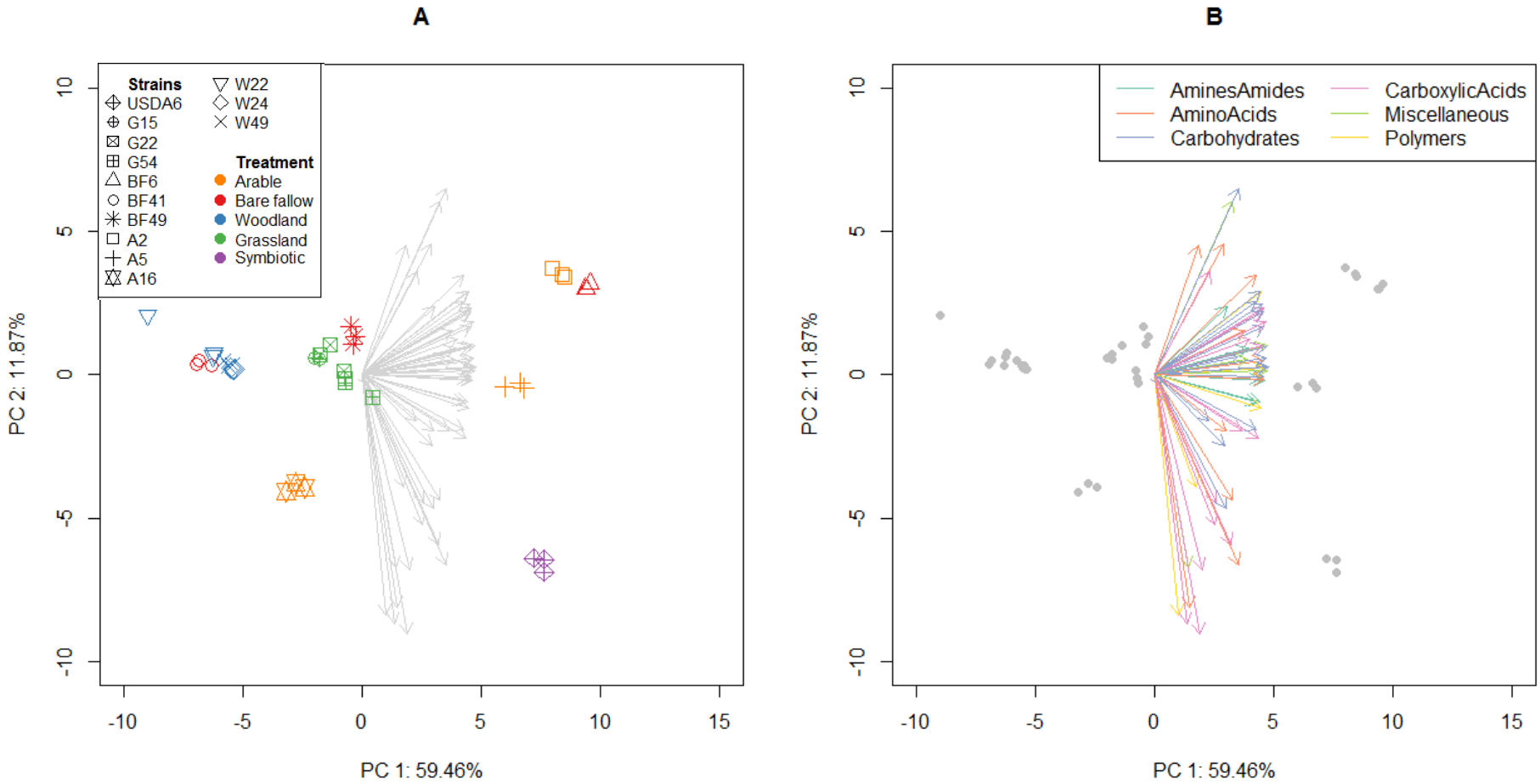


Figure 6.8: Principal components analysis biplot showing PC1 and PC2 accounting for 71.33% of the variation for time 98 hours.
A shows isolates colour coded according to soil treatment and B shows the 58 utilised substrates classified into 6 guilds.

The highest loadings and therefore the substrates which make the largest contribution to PC1 were L-arabinose, α-D-glucose and succinamic acid. L-arabinose is a monosaccharide and is an aldose sugar. It is involved in numerous pathways including ascorbate and aldarate metabolism and amino sugar and nucleotide sugar metabolism. In the ascorbate and aldarate metabolism pathway, L-arabinose is converted to L-arabino-1,4-lactone by L-arabinose 1-dehydrogenase (EC 1.1.1.46, EC 1.1.1.376) involved in central metabolism (Kanehisa and Goto, 2000, Kanehisa et al., 2016). All isolates except the woodland isolates and the bare fallow isolate BF41 were able to utilise this substrate (Figure 6.9). Alpha-D-glucose and succinamic acid are described in Section 6.3.2.1. A summary of substrates with the highest loadings for PC1 and PC2 can be found in Table 6.3. Biolog utilisation graphs for all substrates can be found in Appendix K.

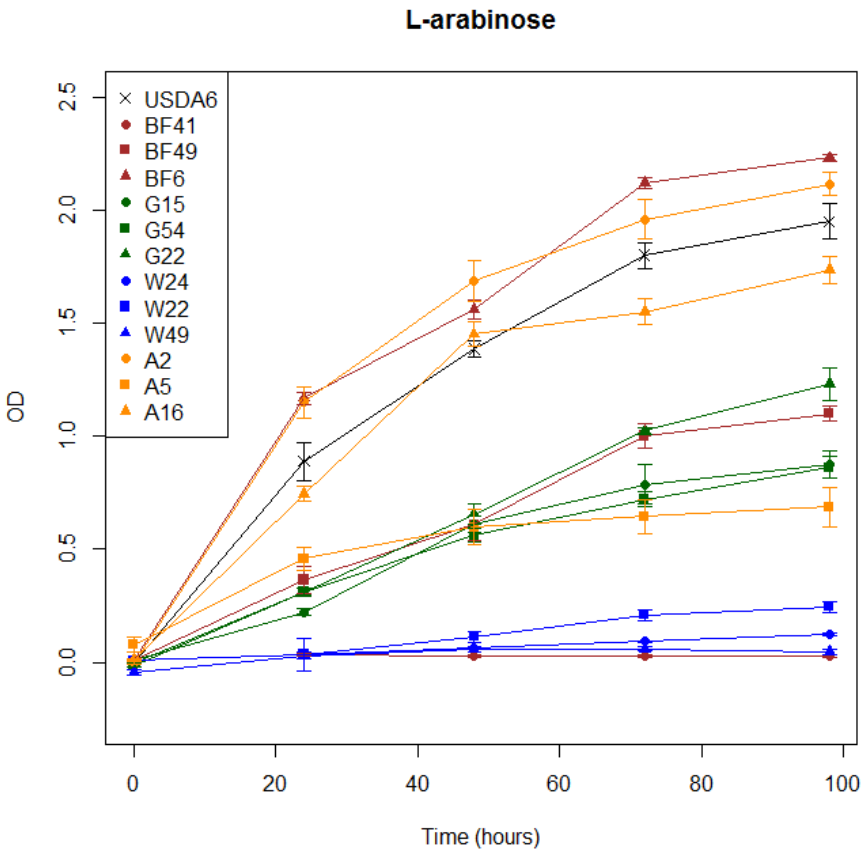


Figure 6.9: Biolog metabolism graph for L-arabinose which had the highest loading for PC1 at 98 hours.
Error bars are standard error.

Table 6.3: Summary of the top 5 highest loadings for PC1 and PC2 for each time point individually and all time points together.

Loading rank	Time points				
	24	48	72	98	All
1 (PC1)	Beta-hydroxy butyric acid	Succinamic acid	L-arabinose	L-arabinose	Alpha-D-glucose
2 (PC1)	D-gluconic acid	Bromo succinic acid	Alpha-D-glucose	Alpha-D-glucose	Succinamic acid
3 (PC1)	L-arabinose	Alpha-D-glucose	L-threonine	Succinamic acid	Methyl pyruvate
4 (PC1)	Methyl pyruvate	Succinic acid	D-sorbitol	Methyl pyruvate	L-arabinose
5 (PC1)	Bromo succinic acid	D-galactose	Propionic acid	D-mannitol	Mono-methyl-succinate
1 (PC2)	Tween 40	L-alanine	L-alanine	L-alanine	L-alanine
2 (PC2)	D-sorbitol	Tween 40	L-threonine	L-proline	L-threonine
3 (PC2)	L-alanine	L-threonine	D-sorbitol	D-sorbitol	D-sorbitol
4 (PC2)	L-threonine	D-sorbitol	Malonic acid	L-threonine	L-glutamic acid
5 (PC2)	L-asparagine	Propionic acid	L-glutamic acid	Glycyl-L-aspartic acid	L-proline

6.3.3 Carbon metabolism cluster analysis

Figure 6.10 shows the dendrogram for the time point 98 hours. The three grassland isolates cluster together with an AU score of 100%. G22 is more similar to A16 than G54 with an AU score of 94%; however, the grassland isolates form a distinct clade. The bare fallow isolate BF41 is separate from the other bare fallow isolates and is most similar to the woodland isolate W22 with an AU score of 95%. The bare fallow isolate BF6 is most similar to the arable isolate A2 (AU=80%) and is part of a larger clade containing A2, A5, BF49 and USDA 6 with an AU score of 99%. The arable isolate A16 is separate from the other arable isolates with an AU score of 99%.

Figure 6.11 shows the dendrogram for all of the time points showing a similar pattern to the dendrogram at time point 98 hours. The grassland isolates again cluster together and away from other isolates with an AU score of 100%. G15 and G22 are most similar with an AU score of 93%. The bare fallow isolates BF41 is separate from the other bare fallow isolates with a score of 98%. The bare fallow isolate BF6 is most similar to the arable isolate A2 with a score of 100% and forms part of a larger clade with BF49, USDA 6 and A5 with an AU score of 100%. The arable isolate A16 is again separated from the other arable isolates and is not clustered with other isolates with a score of 100% which again corresponds with the PCA biplot for time 0 hours (Appendix L). The woodland isolates do not form a single clade and are not clustered with any other isolate and are near the base of the tree (W49: 98%, W22: 64%, W24: 74%).

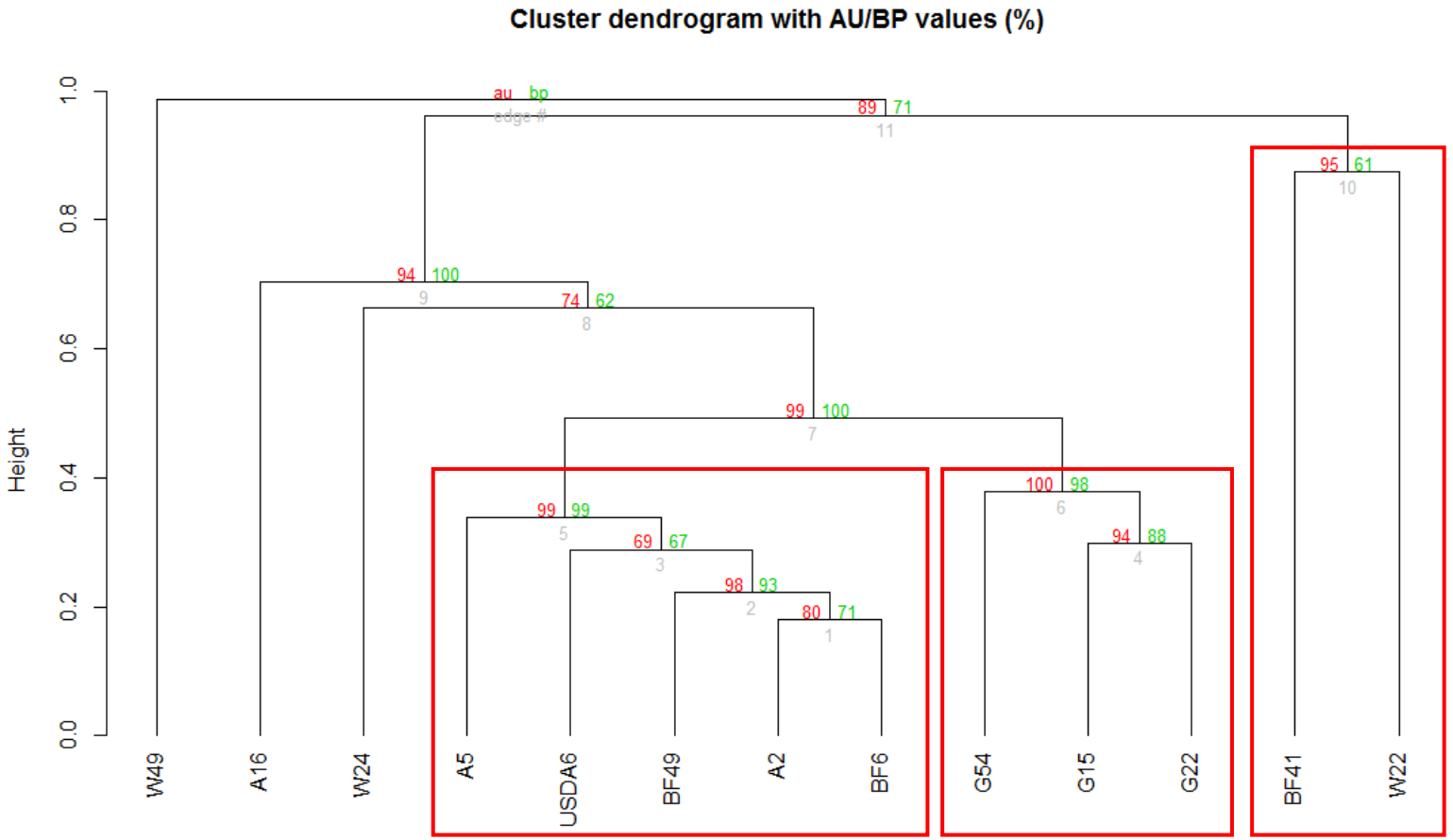


Figure 6.10: Carbon metabolism profile cluster analysis using UPGMA clustering method and 1000 bootstraps for the OD at 98 hours.

Values in red indicate approximately unbiased (AU) *P*-value and in green, the bootstrap *P*-value (BP). Grey numbers indicate the edge numbers. Red boxes are drawn around clades with an AU value of >95%.

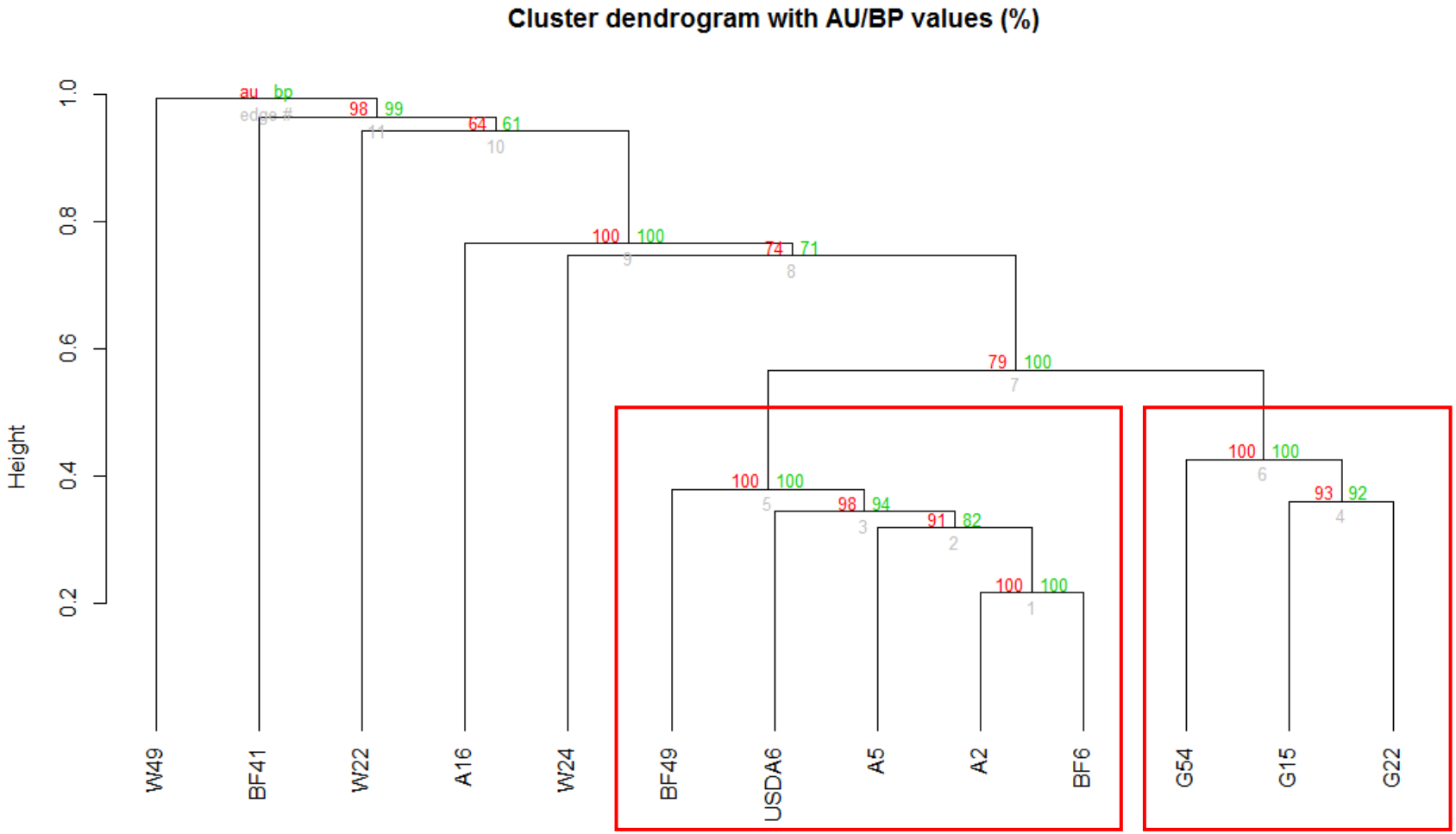


Figure 6.11: Carbon metabolism profile cluster analysis using UPGMA clustering method and 1000 bootstraps for all time points. Values in red indicate approximately unbiased (AU) *P*-value and in green, the bootstrap *P*-value (BP). Grey numbers indicate the edge numbers. Red boxes are drawn around clades with an AU value of >95%.

6.3.4 Substrate utilisation profile at 98 hours

Of the substrates able to be utilised (57), none are unique to the woodland, bare fallow, grassland or arable isolates. Three substrates are only able to be utilised by the symbiotic isolate USDA 6; L-phenylalanine, L-alanine and L-proline. These are all amino acid substrates. There were no substrates able to be utilised by every isolate assayed. There are 38 substrates which were not able to be used by any of the isolates (Listed in Table 6.2). The full binary matrix is provided in Appendix M. Figure 6.12 shows the number of substrates able to be utilised by the individual isolates. The three grassland isolates were able to use a similar number of substrates. The bare fallow and arable isolates are more variable in the extent they use the substrates. The woodland isolates were not able to use many substrates to the extent that OD increased to twice the value measured in the water-only control. This could be linked to growth rate discussed in Section 6.3.1 where the woodland isolates tended to have lower growth rates but reached significantly lower maximum OD. Only the arable isolate A5 was able to use a higher number of substrates that the symbiotic isolate USDA 6.

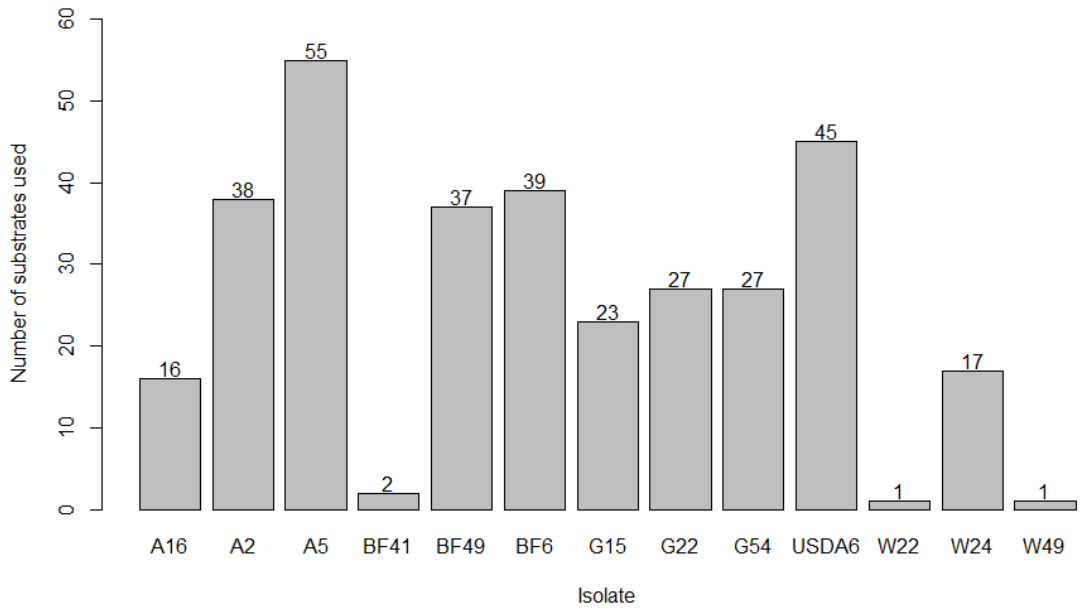


Figure 6.12: Total number of substrates able to be utilised at time 98 hours.
An OD of twice the water well was considered positive.

Woodland isolates had significantly reduced substrate usage (GLM, Estimate=-1.7284, *P*<0.001). Both carbohydrates (GLM, Estimate=1.6314, *P*<0.001) and carboxylic acids (GLM, Estimate=2.0441, *P*<0.001) were used significantly more than other guilds (Statistics output is listed in Appendix N). Figure 6.13 shows the number of substrates utilised by isolates from each soil treatment and the guild composition of these counts.

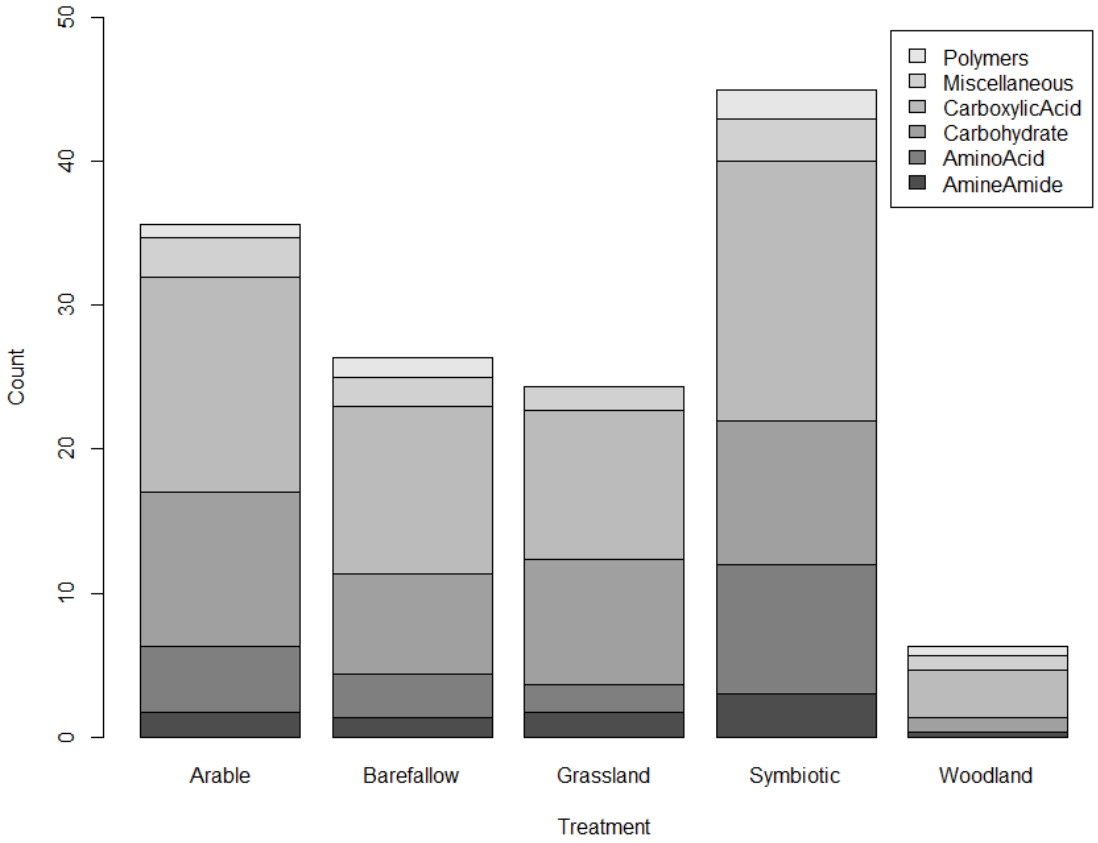


Figure 6.13: Stacked barplot showing number of substrates utilised for each guild and each soil treatment.

6.4 Discussion

6.4.1 Growth curve analysis of the *Bradyrhizobium* culture collection

Carbon is often the limiting factor for microbial growth however nitrogen and phosphorus limitation play an important role (Alden et al., 2001, Ekblad and Nordgren, 2002, Ilstedt and Singh, 2005, Demoling et al., 2007). A number of other factors including temperature, pH and water content also have implications on bacterial growth and activity within soil (Demoling et al., 2007). Soil pH has been shown to be particularly influential over bacterial growth with low pH leading to lower growth rates (Baath, 1996, Alden et al., 2001, Demoling et al., 2007). Strains with lower growth rates tend to have higher carbon requirements as more carbon is used for maintenance than growth and biomass production which can alter the C/N ratio within the environment. Faster growing strains are more likely to be limited by nitrogen or phosphorus due to lower carbon demands than slower growing isolates (Demoling et al., 2007). Diversity of growth characteristics has been described in *Escherichia coli* and it was reported than there was high variability in growth and extent of growth inhibition by lactoferrin between the strains and no clear patterns were found with respect to growth according to pathogenicity or pathogroups (Sekse et al., 2012). *E. coli* strains have also been shown to exhibit varying levels of biomass production and acetic acid accumulation (Luli and Strohl, 1990). Rhizobia have also shown to have variable growth rates and between strains isolated from nodules from the same host plant. Cowpea (*Vigna unguiculata*) symbionts from plants grown in different soil types showed fast growth rates despite usually being nodulated by *Bradyrhizobium* spp. where slow growth is characteristic. From over 500 nodules, the isolates grouped into 78 groups based on morphology and growth rate (Martins et al., 1997).

The growth rates analysis showed that the woodland isolates had significantly lower growth rates and reached a significantly lower maximum OD when compared to all other soil treatments. No differences were found for other pairwise comparisons for soil treatment. The maximum OD for each isolate was not significantly different within the arable, bare fallow or symbiotic isolates. The grassland isolate G9 was significantly lower than a number of other isolates but this appears to be an anomaly for grassland isolates as no other pairwise comparison between grassland isolates was significant. The MAG medium contains both gluconic acid and arabinose which are also included in the Biolog plates. The bare fallow isolate, BF41 and the three woodland strains tested were either unable to use both of these substrates at all or to a much lower OD than the other

isolates. This could be reflected in the low growth rates seen in the woodland isolates; that other strains are able to use a number of carbon sources in the medium and so growth rate is able to be higher. The arable and symbiotic isolates were the most consistent groups for growth rates; there were no significant differences between any of the arable or symbiotic isolates. The growth rates for bare fallow and grassland isolates were more variable than the other soil treatments and this could reflect the heterogeneous nutrient availability in these soil treatments associated with more opportunistic carbon scavenging in bare fallow and a diverse plant community releasing exudates in grassland.

Woodland isolates are more adapted to using complex substrates for growth and could explain the lower OD and growth rates for these isolates and the high variation observed in this soil treatment (Waldrop and Firestone, 2004). Carbon within woodland soils derived from plant litter is high in complex substrates for example lignin and tannins and these substrates are found at higher quantities in woodland soils than grassland soils (Deschaseaux and Ponge, 2001, Waldrop and Firestone, 2004). Communities with the ability to degrade these complex carbon sources have been found to be distinct between woodland and grassland soils. Groups able to degrade simpler carbon sources were not found to be different between these soils despite the over community being altered by above ground plant community (Waldrop and Firestone, 2004). The symbiotic isolates are adapted to living inside nodules with a supply of carbon, nitrogen and other micronutrients in the photosynthate which is a homogenous environment compared to soil (Delmotte et al., 2010). The bare fallow soil contains low levels of organic carbon and therefore bacteria isolated from this soil will be adapted to being opportunistic having to scavenge carbon when it becomes available (Hirsch et al., 2009). The majority of the grassland isolates had similar growth rates with G9 and G22 being slower growing than a number of the other grassland isolates. The grassland is a nutrient rich environment with a constant supply of carbon through rhizodeposits and exudates from plants (Hogberg and Read, 2006, Lakshmanan et al., 2014). However, differences between bare fallow and grassland soils for growth rate were not seen this suggests that both groups of isolates have retained pathways for metabolising a range of compounds contained within the MAG medium. This is similar to reported results where bare fallow communities grown in Biolog plates, had retained the ability to metabolise straw and cellulose despite fresh organic matter not being applied to the bare fallow soil for 80 years (Guenet et al., 2011).

6.4.2 Key substrates identified from Biolog™ assays

For PC1, many of the substrates with highest loadings across the time points were involved in central metabolism for example succinamic acid, L-arabinose, α-D-glucose and methyl pyruvate. This suggests that the growth rate within 98 hours of the isolates is an important factor for PC1 which describes the highest percentage of the variation. Alpha-D-glucose was a substrate which contributed most to PC1 for all time points together and had the second highest loading for the 98 hours' time point. It was also influential for 48 hours and 72 hours' time points. Succinamic acid was shown to have a high loading in all time points together, 48 hours and 98 hours. L-arabinose was a substrate contributed greatest to the variation in PC1 at 98 hours and was also influential for other time points. L-arabinose is in the culture medium used to isolate and maintain the culture collection and so it was predicted that the isolates would respond in a similar way to this substrate and for this substrate to have a small loading. Other carbon sources including D-gluconic acid and yeast extract, are in the MAG medium and these could be providing the carbon sources for energy and growth for isolates not able to metabolise L-arabinose. This means that the strains could still be isolated using the MAG medium but allowing for between strain variability when arabinose is supplied as the sole carbon source. Alternatively, faster growing isolates could be metabolising these substrates providing the variation described by PC1 at 98 hours.

For PC2, the substrates with the highest loadings included many amino acids for example L-alanine, L-threonine, L-proline and L-asparagine. The high loadings show that these substrates had the greatest influence over the variation seen in PC2. Both L-alanine and L-proline were substrates which were only able to be used by the symbiotic isolate *B. japonicum* USDA 6 and so we can conclude that these substrates were influential as they separate the free-living isolates from the symbiotic isolate. This PC accounted for a smaller amount of variation than PC1 and so is a distinguishable feature after growth rate.

6.4.3 Variability of carbon metabolism

The cluster analysis showed that the grassland isolates are much more consistent in their carbon metabolism profile compared to other treatments. The bare fallow and arable isolates were much more variable. The woodland isolates used a small number of substrates and this could reflect their natural environment. Woodland soils usually contain more complex carbon sources for example lignin and tannins derived from plant litter as previously described (Deschaseaux and Ponge, 2001, Waldrop and Firestone, 2004). The woodland isolates could have lost the ability to

metabolise more labile sources of carbon as faster growing groups within the soil could be metabolising these substrates in the native environment. These groups have been found to be similar to those found in grassland communities (Waldrop and Firestone, 2004). The three grassland isolates tested all metabolised a similar number of substrates (where the OD was double the OD of the water well). Both the arable and bare fallow isolates showed a variable number of substrates that they were able to utilise which could be linked to the lower amounts of organic carbon found in the Highfield plots and the heterogeneity seen in nutrient availability between the plots (Hirsch et al., 2009, Barre et al., 2010, Lakshmanan et al., 2014). The larger the plant diversity, for example the level of diversity seen in grassland compared to bare fallow or arable plots, the greater the level of heterogeneity seen in the soil due to the complex range of root exudates released into the rhizosphere (Kowalchuk et al., 2002).

Both carbohydrates and carboxylic acids were utilised more than the other guilds and this could reflect the types of carbon available in the soil. Rhizobia are known to have a high number of ABC transport systems and this is thought to enable scavenging of a range of nutrients even when they are at low concentrations (Prell and Poole, 2006). The symbiotic isolate USDA 6 utilised a larger number of substrates than all woodland, grassland and bare fallow isolates and a higher number than two of the arable isolates (A2 and A16). Only A5 was able to utilise more carbon sources than USDA 6. Three substrates are only able to be utilised by the symbiotic isolate USDA 6; L-phenylalanine, L-alanine and L-proline with L-alanine and L-proline substrates identified as particularly influential across PC2. Amino acids are nitrogen-containing molecules and while in the nodule, symbiotic rhizobia receive a supply of amino acids from the host plant (Lodwig et al., 2003). Legume root exudates contain a high number of amino acids, sugars and carboxylic acids. For example, white clover (*Trifolium repens* L.) exudates include serine, glycine and asparagine at high levels (Paynel et al., 2001). Positive chemotaxis has been shown *Rhizobium leguminosarum* towards edible pea roots (*Pisum sativum* L.) due to the release of a number of amino acids including alanine, asparagine, leucine and valine (Gaworzewska and Carlile, 1982). The ability to utilise amino acid substrates would allow for symbiotic strains to colonise the rhizosphere and find host plants. This could suggest why amino acids were influential in the separation of a symbiotic strain from free-living strains and why it was only the symbiotic isolate able to metabolise these substrates.

The carbon metabolism profiles showed there was high diversity in the number and level of carbon substrate utilisation. Woodland isolates were unable to use many of the substrates contained in the Biolog plate and this could reflect the forms of carbon found in their native environments. The growth rate of the woodland strains could also explain their low substrate

usage in the Biolog plates as this was significantly lower than the other soil treatments. The bare fallow isolates had variable carbon use profiles supporting the hypothesis that these isolates are opportunistic in their carbon metabolism in the environment. The growth rates were not significantly different between grassland, arable and bare fallow isolates and this suggests that the carbon metabolism profiles for these isolates are likely to reflect forms of carbon either previously or currently present in these soils.

7 Nitrogen cycling ability of *Bradyrhizobium*

7.1 Introduction

7.1.1 Nitrogen in the environment

Nitrogen is often limiting for plant growth in soil and so in agriculture, there is a reliance on the application of nitrogenous fertilisers to overcome this (Geurts et al., 2016). The application of fertilisers alters nitrogen cycling within soil and can lead to nitrate and nitrite pollution and eutrophication of rivers and emission of the potent greenhouse gas nitrous oxide. Managing nitrogen is one of the biggest challenges facing agriculture today (Kraft et al., 2011, Graf et al., 2014). Both nitrogen fixation and denitrification are important microbial processes in the nitrogen cycle. Nitrogen fixation increases the level of biologically useful form of nitrogen in the soil complementing other nitrogen inputs (fertilisers and deposition) and denitrification returns nitrogen into the atmosphere (Buckley et al., 2007, Falk et al., 2010, Pastorelli et al., 2011, Zhu et al., 2012). Nitrogen is lost from the soil through denitrification, anaerobic ammonium oxidation (anammox) and nitrification. Anammox is the process where NH_4^+ is oxidised and nitrite is reduced to produce dinitrogen gas and has been detected in both aquatic and terrestrial ecosystems (Long et al., 2013). Nitrification occurs in aerobic conditions and is the conversion of NH_4^+ which produces N_2O and nitrate by ammonium oxidising bacteria (Maag and Vinther, 1996, Khalil et al., 2004).

The loss of nitrogen from soil is particularly important to agriculture where nitrogenous fertilisers are a large expense to farmers and the production of N_2O, an intermediate of denitrification, contributes to global climate change (Maag and Vinther, 1996, Long et al., 2013). In addition, increased carbon dioxide levels in the atmosphere have been linked with increased uptake of nitrogen by soil microbes increasing the problem of nitrogen limitation for plants (Kelly et al., 2013). As described in the General Introduction (Section 1.3), plants are intimately linked to their rhizosphere and influence the soil communities through the release of root exudates and rhizodeposits (Kelly et al., 2013, Kumar et al., 2016). Recruitment of symbiotic nitrogen fixing bacteria is one mechanism used by plants to overcome nitrogen limitation often found in soil (Geurts et al., 2016). Plant and soil microorganisms compete for available nitrate, the terminal electron acceptor and substrate in denitrification, however as both carbon and nitrogen is taken by soil microbes, amino acids contained in plant exudates are often used as both a nitrogen and carbon source (Nordin et al., 2004, Zhang et al., 2014a).

7.1.2 Nitrogen fixation through symbiosis

Nitrogen fixation through symbioses is an important characteristic of the bacterial order Rhizobiales (Bedmar et al., 2005, Hollowell et al., 2016a, Hollowell et al., 2016b). Nitrogen fixation occurs in nodules on the roots of legumes, where bacteroids synthesise nitrogenase, encoded by the *nifH* gene, to convert nitrogen into ammonium which is utilised by plants (Mergel et al., 2001, Bedmar et al., 2005, Geurts et al., 2016, Miri et al., 2016). The nodulation process is discussed in detail in Section 1.7. The conversion of dinitrogen gas to ammonium is energetically expensive with each dinitrogen molecule converted requiring 16 ATP molecules (Meakin et al., 2006). Nitrogenase is denatured at high oxygen levels and so a diffusion barrier creates an oxygen limited environment within the nodule with the plant producing leghaemoglobin to transport oxygen to the bacteroids for respiration (Meakin et al., 2006).

7.1.3 Denitrification and *Bradyrhizobium*

As discussed in the General Introduction (Section 1.8.2), the denitrification process has important implications for agriculture including the loss of nitrogen from the soil and contributing to global climate change through nitrous oxide emissions (Mergel et al., 2001, Bedmar et al., 2005, Fernandez et al., 2008). In oxygen limited environments, for example in soil or inside the nodule, denitrifiers utilise NO_3^-, NO_2^-, NO and/or nitrous oxide as the final electron acceptor instead of oxygen (Vairinhos et al., 1989, Mahne and Tiedje, 1995, Braker and Tiedje, 2003, Sameshima-Saito et al., 2004, Bedmar et al., 2005). When oxygen is limited, nitrate can be used for growth by rhizobia and *Bradyrhizobium* is able to denitrify nitrate: the presence of oxygen inhibits denitrification (Vairinhos et al., 1989, Delgado et al., 2003, Bedmar et al., 2005, Delmotte et al., 2014). The conversion of NO_3^- into NO_2^- is catalysed by nitrate reductase and in *Bradyrhizobium* this enzyme is a periplasmic nitrate reductase, NapABC. The alternative version of the nitrate reductase is membrane bound and is encoded by the *nar* genes (Delgado et al., 2003, Bedmar et al., 2005, Monza et al., 2006, Kraft et al., 2011). In *B. japonicum* USDA 110, the *nap* genes are part of a six gene cluster; *napDABCDE* (Sameshima-Saito et al., 2006, Fernandez et al., 2008). This reaction also occurs in nitrate ammonification and assimilatory nitrate reduction (Vairinhos et al., 1989, Bedmar et al., 2005, Monza et al., 2006).

The second stage in denitrification, the reduction of NO_2^- to NO, is considered the first unique and definitive step of this process and the nitrite reductase gene has two forms; *nirK* and *nirS* (Bedmar et al., 2005, Monza et al., 2006, Sameshima-Saito et al., 2006, Kraft et al., 2011). The copper containing version, *nirK,* is the most common nitrite reductase encoding gene found in

Bradyrhizobium (Velasco et al., 2001, Delgado et al., 2003, Bedmar et al., 2005, Sameshima-Saito et al., 2006). However, some *Bradyrhizobium* isolates have been shown to contain *nirS* which contains cd_1 cytochrome (Mergel et al., 2001, Ishii et al., 2011). Both the *nirK* and *nirS* genes are found in a wide range of unrelated bacteria (Coyne et al., 1989). The *nirK* sequence from *B. japonicum* USDA 110 was found to be highly similar to a number of other *nirK* sequences from a range of taxa (Velasco et al., 2001). Ammonium oxidising bacteria have also shown to possess *nirK* for nitrifier denitrification (Shaw et al., 2006, Kraft et al., 2011)

The third step of denitrification, the reduction of nitric oxide, is the step which directly contributes to global climate change through the formation of nitrous oxide if the final stage is denitrification does not occur (Mahne and Tiedje, 1995, Meilhoc et al., 2010). Similar to the previous step, there are two major types of nitric oxide reductase (NOR); cNOR involves cytochrome to receive electrons and qNOR involves quinol (Braker and Tiedje, 2003, Bedmar et al., 2005, Meakin et al., 2006). A hybrid version of NOR has been described in *Bacillus azotoformans* but this is not commonly found in other genera (Spiro, 2012). In *Bradyrhizobium,* it is the cNOR version present with *norB* and *norC* encoding the cytochrome *b* and *c* containing subunits respectively (Mesa et al., 2002, Bedmar et al., 2005). Both *norB* and *norC* are required for growth in oxygen limited environments (Mesa et al., 2002, Spiro, 2012). The genes required for nitric oxide reduction are part of a five gene cluster in *B. japonicum* USDA 110; *norECBQD* (Sameshima-Saito et al., 2006, Fernandez et al., 2008). Leghaemoglobin produced by the plant during nodulation, has also been shown to bind to NO and it is thought that NOR is a mechanism bacteroids use to overcome high levels of toxic NO within the nodule (Meakin et al., 2006). Nitric oxide has been suggested to play a role in the initiation of symbiotic associations by aiding the exclusion of non-symbiotic bacteria (Meilhoc et al., 2010). Another source of nitrous oxide is autotrophic ammonia oxidising bacteria involved in the nitrification step of the nitrogen cycle - these have also shown to produce nitrous oxide through denitrification (Shaw et al., 2006, Kraft et al., 2011).

The final stage of denitrification is the conversion of nitrous oxide to dinitrogen gas which is catalysed by nitrous oxide reductase, a periplasmic enzyme encoded by the *nosZ* gene (Mahne and Tiedje, 1995, Bedmar et al., 2005, Monza et al., 2006, Kraft et al., 2011). This gene is in a cluster of seven genes; *nosRZDFYLX* (Delgado et al., 2003, Bedmar et al., 2005, Sameshima-Saito et al., 2006). This step of denitrification is important in mitigating nitrous oxide emissions and is agriculturally important as agriculture is the major contributor to global nitrous oxide emissions (Sameshima-Saito et al., 2006, Shaw et al., 2006, Fernandez et al., 2008, Graf et al., 2014). The *nosZ* gene is found in both bacteria and archaea and it can be present alone or as part of the

wider denitrification pathway and is the only known enzyme found to reduce nitrous oxide (Graf et al., 2014).

B. diazoefficiens USDA 110 has been shown to contain all four genes required for the full denitrification pathway, often expressing these genes while in a symbiotic state where oxygen levels are limited (Kaneko et al., 2002, Sameshima-Saito et al., 2004, Sameshima-Saito et al., 2006, Delmotte et al., 2014). Denitrification within the nodule is important for the removal of cytotoxic nitrite and NO (Fernandez et al., 2008). Both free-living isolates and bacteroids have been shown to be able to denitrify (van Berkum and Keyser, 1985). Denitrification ability is not universal in *Bradyrhizobium* and has been shown to be variable and isolates usually fall into one of three groups; full denitrifier able to complete all four steps, truncated denitrifier able to complete the first three steps but unable to reduce nitrous oxide or a non-denitrifier (van Berkum and Keyser, 1985, Sameshima-Saito et al., 2004, Monza et al., 2006, Sameshima-Saito et al., 2006).

Fernandez et al. (2008) screened 250 symbiotic *Bradyrhizobium* isolates for denitrification ability and found that 41 were considered denitrifiers, less than 20%; and Sameshima-Saito et al. (2004) showed that out of 65 screened *Bradyrhizobium* field isolates, 28 were full denitrifiers (43%) and 19 (29%) were non-denitrifiers. However, Asakawa (1993) tested the denitrification ability of 103 strains of *Bradyrhizobium* isolated from fields where soybean (*G. max*) or rice (*Oryza sativa*) was grown as the summer crop and either barley (*Hordeum vulgare* L.) or wheat (*Triticum aestivum*) was grown as the winter crop. They found that all of these strains were able to denitrify and generally formed two groups; rapid denitrifiers and slow denitrifiers where the faster denitrifiers were more likely to perform the full denitrification pathway and slower denitrifiers more likely to omit the nitrous oxide reduction step (Asakawa, 1993). This could be due to the strains being isolated from paddy soil and which is submerged in water for long periods and could act as a selection pressure for isolates which are able to respire in anaerobic conditions. Alternatively, growing soybeans will select for symbiotic strains adapted to being inside the nodules where oxygen is limited. Understanding whether the strains in the culture collection are able to denitrify will give an indication on how soil treatment affects a key denitrifier group in soil.

7.1.4 Aims

In this chapter, I aim to:

- Assess whether native *Bradyrhizobium* in soil from the Highfield experiment are able to nodulate soybeans (*Glycine max*)
- Screen the culture collection for genes associated with the nitrogen cycle to determine differences in presence or absence of these genes
- Determine the ability of the culture collection to grow on nitrate-containing medium
- Evaluate the denitrification ability of the culture collection to determine if the gene encoding nitrous oxide reductase is present and functional by using gas chromatography
- Quantify the copy number of the *nosZ* gene in Highfield and Broadbalk soils to look for patterns in the nitrous oxide emissions assay

7.2 Materials and methods

7.2.1 Soybean (*Glycine max*) inoculation

The Highfield soil from the permanent grassland (plots 10, 17 and 26) and permanent bare fallow (plots 3, 4 and 8) plots were tested for the presence of bradyrhizobia capable of nodulating the roots of soybeans (*Glycine max* L. Merr., cv. Elena). The Elena variety was chosen as it had previously been successfully grown in the glasshouses at Rothamsted Research. The seeds were rinsed in 70% ethanol followed by sdH_2O. Once washed, the seeds were soaked in 11% sodium hypochlorite for 15 minutes and then transferred to sdH_2O for 1 minute. This washing step was repeated six times with fresh sdH_2O each time. The seeds were germinated on tap water agar (tap water and 1.5% w/v (15 g L^{-1}) Becton, Dickinson & Company bacto-agar) in the dark for 3-4 days before being planted. Seed sterilisation is required to remove any bacteria and fungi which could inhibit growth of the plant or inhibit colonisation by bacteria present in the Highfield soil and remove any symbiotically effective bradyrhizobia carried over from the parent plant and soil.

The grassland soil was divided into 3 treatment pots and 2 control pots and the bare fallow soil was divided into 4 treatment pots and 2 control pots. Each pot was sown with 3 germinated seeds. The controls were inoculated with 100 µl of *B. japonicum* USDA 6 (approximately 10^8 cells). The pots were placed in a glasshouse and kept ambient day length and the temperature set to $12^oC/16^oC$. After 8 weeks, the soil was shaken and washed from the plant to expose the root system and the roots were examined for nodules. The roots were separated from the shoots and the dry weight of the shoots was calculated after drying at 80^oC overnight. Differences in dry weights were assessed using *t*-tests.

7.2.2 Nitrogen cycling genes screenings

The culture collection was screened for genes involved in nitrogen cycling. The PCR reaction mix was as listed previously in General Methods 2.4. The PCR primers used are listed in Table 7.1. The PCR conditions for each PCR are listed in Table 7.2. All PCR amplicons were held at 10^oc at the end of the PCR program until storage at -20^oc. The amplicons were examined by agarose gel electrophoresis and imaged as described in General Methods 2.5. A 100bp or 1kb ladder was used as a size comparison depending on expected product size.

Table 7.1: List of functional genes primer information

Gene	Primer name	Primer sequence (5'-3')	Size (bp)	Reference
nifH	PolF	TGC GAY CCS AAR GCB GAC TC	342	Poly et al. (2001)
	PolR	ATS GCC ATCATY TCR CCG GA		
nod	TSnodD1-1a	CAG ATC NAG DCC BTT GAA RCG CA	2000	Moulin et al. (2004)
	TSnodB1	AGG ATA YCC GTC GTG CAG GAG CA		
nirK	nirK876F	ATY GGC GGV CAY GGC GA	204	Henry et al. (2004)
	nirK1040R (Redesigned)	GGC TCG ATC AGR TTR TGV GY		
nirS	nirS85F	GAG CGC TTC CTG CAT GAC G	262	Designed
	nirS347R	GCG AAC TGC TTG TGC TTC TC		
	nirS_cd3aF	GTS AAC GTS AAG GAR ACS GG	387	Throback et al. (2004)
	nirS_R3cdR	GAS TTC GGR TGS GTC TTG A		
nosZ	nosZ2F Henry	CGC RAC GGC AAS AAG GTS MSS GT	267	Henry et al. (2006)
	nosZ2R Henry	CAK RTG CAK SGC RTG GCA GAA		

Table 7.2: PCR conditions for each primer set

Gene	1	2	3	4	5	Number of cycles Steps 2-4
nifH	95°C 10 mins	95°C 15 secs	56°C 20 secs	72°C 20 secs	72°C 10 mins	35
nodD	94°C 2 mins	94°C 1 min	55°C 2 min	72°C 2 min	72°C 7 mins	20
nirK	95°C 5 mins	95°C 15 secs	60°C 15 secs	72°C 1 min	72°C 10 mins	30
nirS (Designed)	95°C 5 mins	95°C 15 secs	60°C 15 secs	72°C 1 min	72°C 10 mins	30
nirS (Throback et al., 2004)	94°C 2 mins	94°C 30 secs	51°C 1 min	72°C 1 min	72°C 10 mins	35
nosZ	95°C hold	95°C 15 secs	60°C 15 secs	72°C 1 min	72°C 10 mins	30

7.2.3 Denitrification assays

7.2.3.1 Denitrification growth medium

To test for the presence and completeness of the denitrification pathway, isolates were cultured on modified yeast extract mannitol (YEM) medium (Mahne and Tiedje, 1995, Delgado et al., 2003, Monza et al., 2006). The YEM medium consisted of (per litre): yeast extract (Merck KGaA, Germany) 0.5 g, mannitol (Fisher Chemicals, USA) 10 g, dipotassium phosphate (K_2HPO_4) (Fisons, UK) 0.5 g, magnesium sulphate ($MgSO_4.7H_2O$) (Fisons, UK) 0.2 g, sodium chloride (BDH Lab Supplies, UK) 0.1 g. For a nitrate-containing medium, potassium nitrate (KNO_3) (Duchefa, Netherlands) 1 g per litre was also added. All chemicals were of analytical grade. The pH was adjusted to 6.8 using 5 M hydrochloric acid. For solid media, 1.5% w/v (15 g L^{-1}) bacto-agar (Becton, Dickinson & Company, USA) was added to the YEM broth. The medium was sterilised by autoclaving at 120^{o}C for 15 minutes and allowed to cool before pouring. Warm medium (<50^{o}C) (approximately 20 ml) was poured in to Petri dishes (Thermo Fisher Scientific, USA) in a sterile flow hood and stored at 4^{o}C once set.

7.2.3.2 Growth on nitrate plate test

Five isolates per plate and one negative control (no bacteria) were streaked onto YEM agar plates with and without KNO_3. All isolates in the *Bradyrhizobium* culture collection were used (See Chapter 3 for full list of isolates). A random number generator was used to determine which isolates were on any one plate and each replicate plate was randomised so each plate contained a different combination of isolates. Three replicate plates were set up resulting in 30 plates. This was then duplicated for YEM with and without KNO_3 (60 plates). The whole experiment was duplicated resulting in 120 plates in total; half were incubated aerobically and half were incubated in aerobic jars (GasPak 150, Becton, Dickinson & Co, USA). In order to make the jars anaerobic, three Oxoid™ AnaeroGen™ 2.5L sachets (Thermo Fisher Scientific Inc, USA) were added before the jars were sealed. Two anaerobic jars were used as each jar was able to hold up to 30 plates each. As a result, the plates were numbered 1-30 and randomly divided into two groups; one for each jar to avoid any bias arising from the anaerobic jar. The plates were incubated at 30^{o}C. After 14 days, growth was assessed and graded on a scale of 1-5 where 1 is no growth and 5 is high level of growth. Figure 7.1 shows an example of the grading system used to assess growth. Three negative controls were used which were known to not contain *nirK/S*. These were *Rhizobium leguminosarum* JI300 (Hirsch et al., 1980), *Rhizobium* sp. RCR3812 (Dudman and

Belbin, 1988) and *R. leguminosarum* bv. *viciae* VF39 (Prell et al., 2002). See Appendix O for blocking structure.

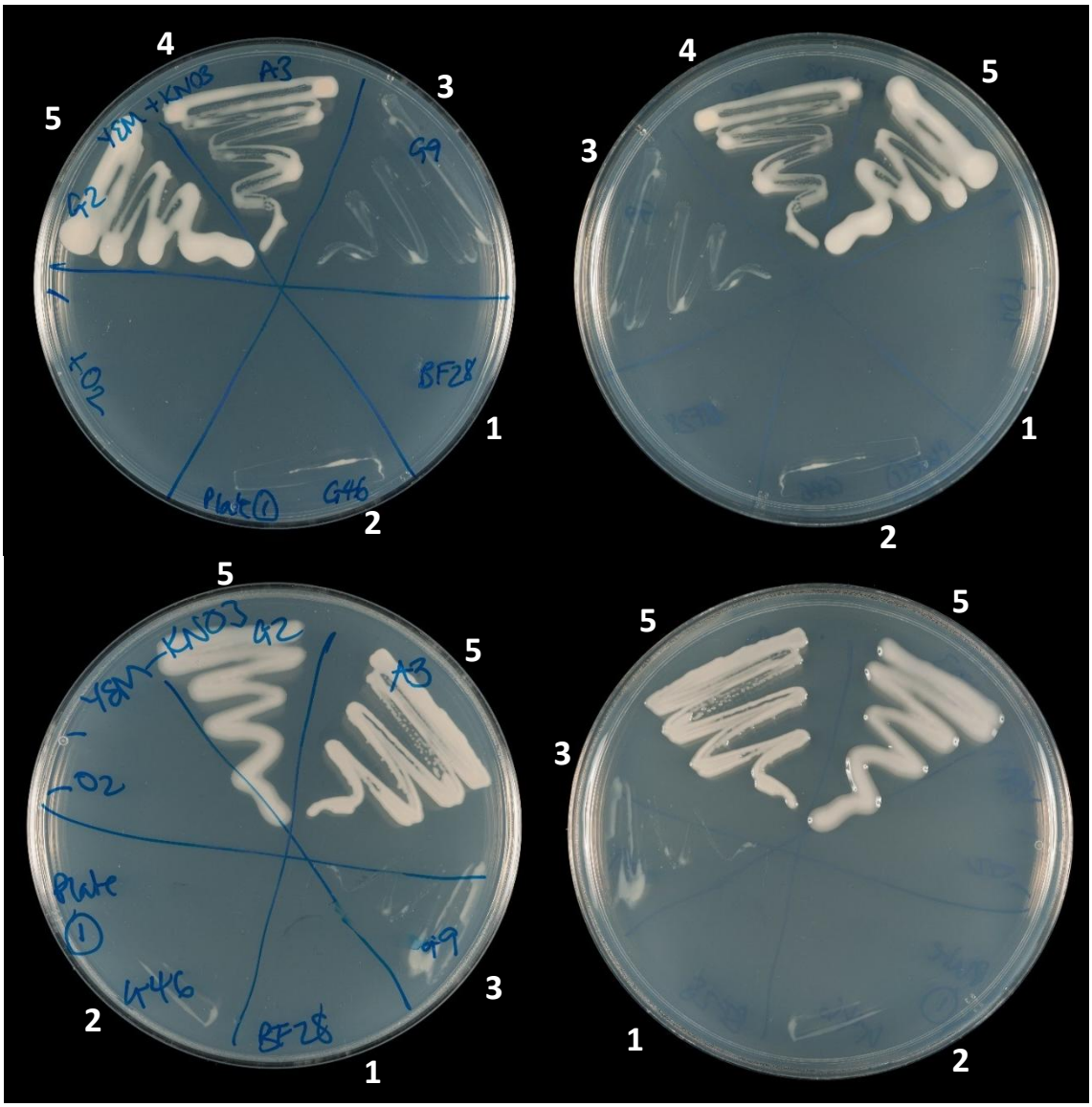


Figure 7.1: Example of the grading system used to assess growth for the denitrification plate assay.

7.2.3.3 Growth on nitrate plate test statistical analysis

In order to determine whether the presence of nitrate influences the growth score, a Wilcoxon-Mann Whitney test was used. A Wilcoxon-Mann Whitney test was also used to determine if the presence of oxygen during incubation affected the growth score. The dependent variable was growth score which is ordinal data. A Kruskal Wallis test was used to determine whether growth score was affected by the growth treatment (nitrate and aerobic, nitrate and anaerobic, no nitrate and aerobic, no nitrate and anaerobic). All statistics were performed using R (v 3.2.2).

7.2.3.4 Nitrous oxide emissions assay

Each isolate within the culture collection was grown aerobically in 15 ml of YEM (no KNO_3) incubated for two weeks at 28°C, shaking at 100 rpm until high cell density and the cultures were opaque. After two weeks, 140 µl of culture was aliquoted into a flat bottom, clear 96-well plate (Thermo Fisher Scientific, USA). The OD was read using a Varioscan SkanIt plate reader (Thermo Fisher Scientific, USA) at 660 nm. Cultures were centrifuged at 14,000 xg and the supernatant was removed. Cultures were resuspended in sdH_2O to the same OD (0.05).

Aliquots of 8 ml of YEM broth supplemented with KNO_3 were measured into 20 ml serum bottles (Wheaton Industries Inc, USA). Six replicate serum bottles per isolate or negative (no bacteria) control were set up (three replicates to receive acetylene and three replicates without acetylene) resulting in 300 serum bottles. Rubber seals and metal crimping seals were loosely placed on each bottle and bottles were autoclaved at 120°C for 15 minutes before storage at 4°C until needed. Each bottle was inoculated with 0.5 ml of bacterial culture (except for the negative control which contained only sterile culture medium) and sealed with a rubber seal and aluminium crimp seal to ensure the serum bottles were airtight. The headspace gas was replaced with dinitrogen gas passed through a 0.45 µm filter and injected into the serum bottles using 19G needles for 30 seconds (Becton, Dickinson & Co, USA). Acetylene gas was made by reacting calcium carbide with distilled water and collecting in a balloon (See Appendix P for full protocol). For three bottles per isolate, 1 ml (approximately 8% vol/vol) acetylene gas was added to the headspace. Cultures were incubated for 14 days, at 28°C, shaking at 100 rpm. After 14 days, 12 ml of the headspace was taken using a syringe and 19G needles and placed into evacuated 12 ml Exetainer® vials (LabCo, UK). Gas samples were stored at 15°C and in the dark until analysed at the University of Reading using gas chromatography.

7.2.3.5 Gas chromatography

Gas chromatography was carried out by the gas chromatography technician at the University of Reading.

Analysis was done on an Agilent 7890 B-GC system fitted with a HayeSep Q 60/80 packed column at 60°C initial temperature (Agilent Technologies, USA). N_2O was analysed using a ^{63}Ni electron capture detector (ECD) and carbon dioxide analysed using a flame ionisation detector (FID) after its conversion to methane using a methaniser. Nitrogen was the carrier gas, Ar/CH_4 the makeup gas for the ECD and air and N_2 the makeup gas for the FID. The ECD was heated to 300°C with a makeup flow rate of 2 ml/min. The FID was heated to 250°C with a H_2 flow rate of 80 ml/min and air flow rate of 450 ml/min. The samples were injected using a CombiPal autosampler (injection

volume 1 ml). The GC was calibrated using external gas mix standards: standard 1 - laboratory air (0.32 ppm N_2O); standard 2 - 5 ppm N_2O and 2500 ppm CO_2; standard 3 - 50 ppm N_2O and 5000 ppm CO_2. The balance gas for standards 2 and 3 was N_2.

7.2.3.6 Nitrous oxide emissions assay statistical analysis

To test if there were significant differences in the level of N_2O produced, measured in ppm, between cultures with and without acetylene added to the headspace for each isolate, ANOVA was used. Significant differences indicate that the presence of C_2H_2 did affect the level of N_2O produced suggesting there is a functional nitrous oxide reductase being produced. Isolates which gave a significant difference in N_2O were examined to determine whether the culture with or without C_2H_2 gave the highest level of N_2O production. Acetylene has been shown to inhibit nitrous oxide reduction (Mahne and Tiedje, 1995, Sameshima-Saito et al., 2004). If there is a functional *nosZ* gene, the culture without C_2H_2 would be expected to contain a higher level of N_2O. Isolates without *nosZ* but with other denitrification genes will produce N_2O to a similar level both with and without C_2H_2. If no or low N_2O is produced, the CO_2 measurements can be assessed to determine if the culture was viable (Sameshima-Saito et al., 2004, Monza et al., 2006). CO_2 production for isolates giving a significant difference for N_2O production with and without C_2H_2 was assessed using a one sample t-test. This was to test if the cells were viable testing if the CO_2 production was significantly greater than the CO_2 levels (measured in ppm) in the blank samples (no bacteria). The mean CO_2 production for the blank samples was 799.2245 ppm. All statistics were performed using R (v 3.2.2). Barplots were drawn using the R package *ggplot2* (Wickham, 2009).

7.2.4 Nitrous oxide reductase (*nosZ*) qPCR

Soil DNA was extracted using the MoBio PowerSoil DNA isolation kit (MoBio, USA) from 0.25 g soil taken from the Highfield plots; permanent grassland (plots 10, 17 and 26), permanent arable (plots 14, 20 and 24), and bare fallow (plots 8.1, 3 and 4). In addition, DNA from three woodland samples (0.25 g) from the Broadbalk experiment (W1, W2 and W3) was extracted (Sites described in Chapter 3, Section 3.2.1). The manufacturer's instructions were used. The beadbeating step was set at 5.5 m/s and two runs of 30 seconds each were carried out using FastPrep-24™ 5G (MP Biomedicals, USA). Extracted DNA was quantified as described in General Methods 2.3.

Copy number for the nitrous oxide reductase (*nosZ*) gene was determined using quantitative PCR using a BioRad CFX 384™ real-time system with a C1000 Touch™ thermal cycler (BioRad, USA). The reaction mix consisted of 5 µl of QuantiFast SYBR® green PCR master mix (Qiagen, Germany),

1 µM final concentration of each primer; nosZF (5'-YTGTTCGACGACAAGATCAA-3') and nosZR (5'-AGCGAGATCAGCCATTTTCC-3') (Sugawara and Sadowsky, 2013), 20 ng of DNA per reaction (2 µl) and nuclease-free water was added to a reaction mix volume of 10 µl. The qPCR conditions were as follows: 95^{o}C for 5 minutes, followed by 40 cycles of 95^{o}C for 15 seconds, 60^{o}C for 45 seconds and 76^{o}C for 30 seconds. The plate was read at both 60^{o}C and 76^{o}C. The melt curve was run from 65^{o}C to 95^{o}C with increments of 0.5^{o}C. A series of seven standards were used from 10^{1} to 10^{7} gene copies from the Park Grass Experiment at Rothamsted Research (a long-term permanent grassland equivalent to the Highfield permanent grassland plots) and three negative controls containing no DNA were used for comparison. Standard curves, melt curves and amplification curves can be found in Appendix Q.

7.2.5 qPCR statistical analysis

The starting quantity of *nosZ* in each DNA samples copy number of the *nosZ* gene was calculated by the BioRad software for the BioRad CFX384™ real-time system (BioRad, USA). The DNA quantification was multiplied by the elution dilution factor to give the quantity in the entire DNA sample before dividing by the quantity of DNA added to each qPCR reaction to give the fraction of the DNA sample in each reaction. The copy number was then multiplied by this fraction and then by 4 to give the copy number per gram of dry weight soil. Differences in gene copy number between soil treatments were analysed using a generalised linear model with quasipoisson errors. All statistics were performed using R (v 3.2.2). GLM output can be found in Appendix R.

7.3 Results

7.3.1 Soybean *(Glycine max)* inoculation

7.3.1.1 Nodulation ability

After 8 weeks of growth, irrespective of soil type, only the plants inoculated with *B. japonicum* USDA 6 culture developed nodules (Figure 7.2). This shows that the dominant *Bradyrhizobium* strains in the Highfield soil are unable to form nodules on soybean (*G. max*).

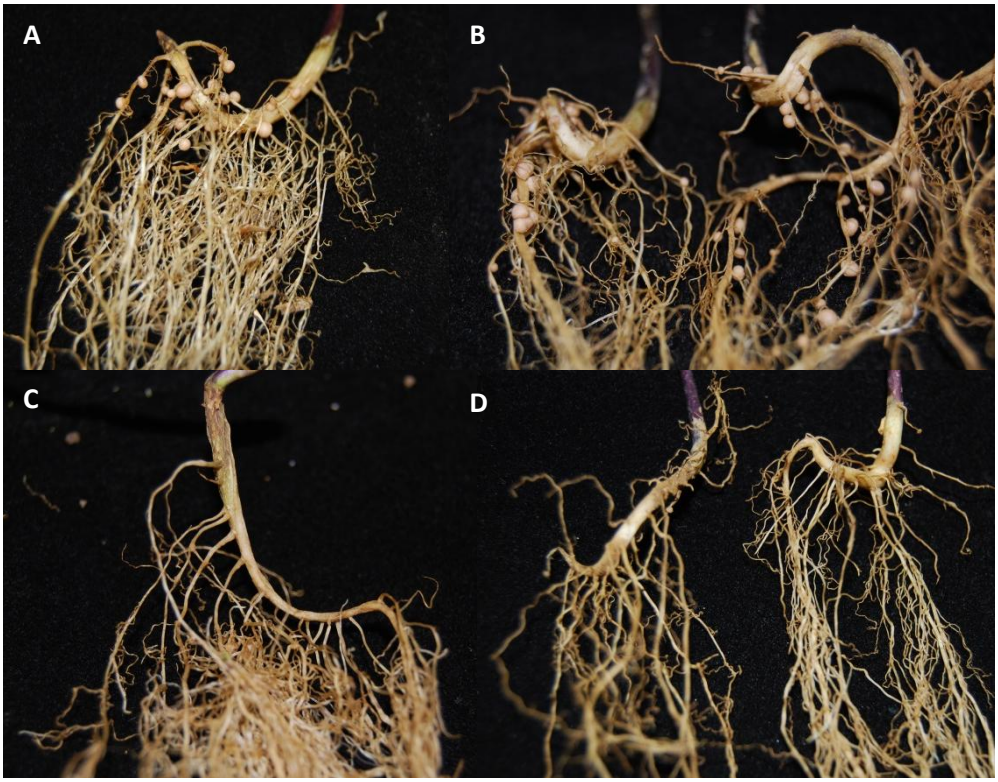


Figure 7.2: Nodules on the roots of soybean plants (*G. max*).
A and B show the presence of nodules on plants grown in grassland soil (A) and bare fallow soil (B) both inoculated with symbiotically-effective strain USDA 6. C and D show the absence of nodules on un-inoculated plants grown in grassland (C) and bare fallow (D) soil.

7.3.1.2 Comparing inoculation effects

Soybean plants which were grown in grassland soil showed a larger growth response when inoculated with *B. japonicum* USDA 6 (Figure 7.3-A). Dry weights indicate the differences in growth between grassland inoculated and non-inoculated plants to be statistically significant (t-test: t=7.0711, df=2.0000, *P*-value=0.0194). In grassland soil, the root system showed little difference in growth between those plants which were inoculated and those which were left un-inoculated suggesting inoculation affects above ground plant growth to a larger extent than below ground growth (Figure 7.3-B).

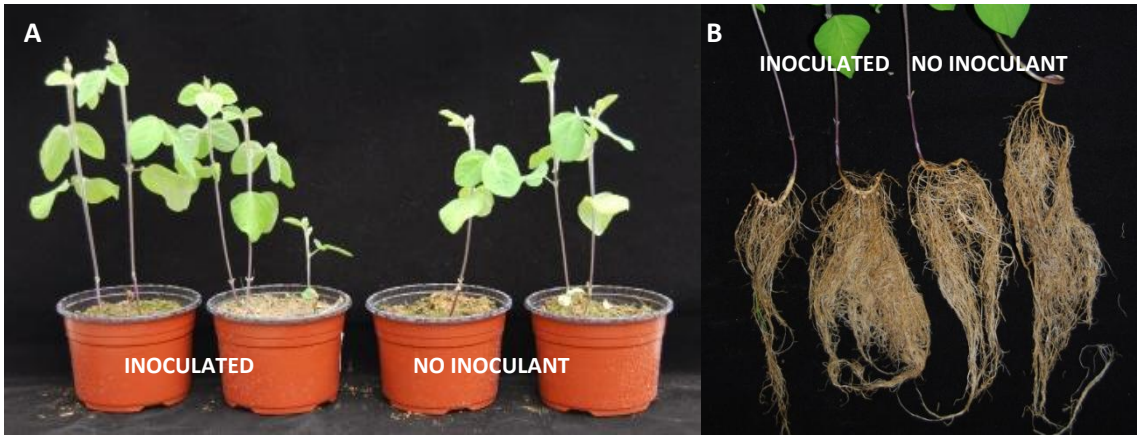


Figure 7.3: Soybean plants grown in grassland soil.
A: Differences in above ground biomass evident between inoculated and un-inoculated plants. B: Differences in below ground biomass small between inoculated and un-inoculated plants.

Plants grown in bare fallow soil showed very little difference in both above and below ground between inoculated and un-inoculated plants (Figure 7.4). These were shown to not be statistically significant (t-test: t=-0.2425, df=4.8500, *P*-value=0.8183). This shows that despite the presence of *Bradyrhizobium* to fix nitrogen and increase the nitrogen availability to the plant, the poor soil structure and lack of other nutrients particularly phosphorus which are at low levels in the bare fallow soil inhibited the growth of the plant which is not corrected by the high abundance of *Bradyrhizobium* on the roots.

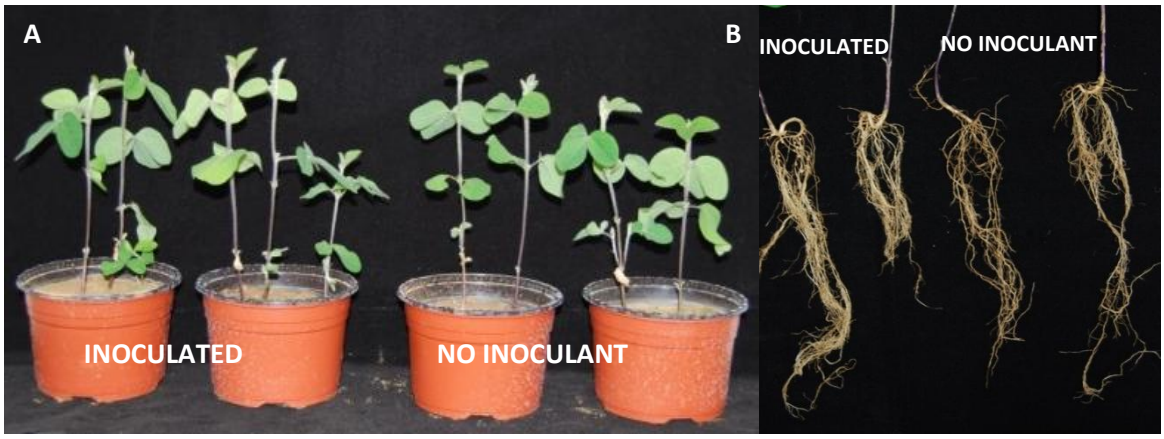


Figure 7.4: Soybean plants grown in bare fallow soil.
A: Differences in above ground biomass small between inoculated and un-inoculated plants.
B: Differences in below ground biomass small between inoculated and uninoculated plants.

Plants grown in grassland soil and not inoculated did not have significantly different above ground biomass than bare fallow, un-inoculated plants (t-test: t=0.2526, df=3.9990, *P*-value=0.8130). A significant difference in above ground biomass was seen between inoculated plants grown in grassland and bare fallow soils (t-test: t=6.5464, df=2.8820, *P*-value=0.0082).

7.3.2 Nitrogen cycling genes screenings

Table 7.3 summarises which genes were present or absent in the *Bradyrhizobium* culture collection. All agarose gels used in screening for functional genes correspond to the order listed in Table 7.3.

Table 7.3: Summary of functional genes presence or absence for each isolate

Lane Number	Isolate	Origin of isolate	*nifH*	*nodD*	*nirK*	*nosZ*
1	G2	Grassland	N	N	**Y**	N
2	G3	Grassland	N	N	**Y**	N
3	G5	Grassland	N	N	N	N
4	G6	Grassland	N	N	**Y**	N
5	G8	Grassland	N	N	**Y**	N
6	G9	Grassland	N	N	N	N
7	G15	Grassland	N	N	**Y**	N
8	G22	Grassland	N	N	**Y**	N
9	G31	Grassland	N	N	**Y**	N
10	G46	Grassland	N	N	**Y**	N
11	G54	Grassland	N	N	**Y**	N
12	BF2	Bare fallow	N	N	**Y**	N
13	BF6	Bare fallow	N	N	**Y**	**Y**
14	BF20	Bare fallow	N	N	N	N
15	BF26	Bare fallow	N	N	N	N
16	BF28	Bare fallow	N	N	**Y**	N
17	BF30	Bare fallow	N	N	N	N
18	BF41	Bare fallow	N	N	N	N
19	BF48	Bare fallow	N	N	N	N
20	BF49	Bare fallow	N	N	**Y**	**Y**
21	BF52	Bare fallow	N	N	N	**Y**
22	A1	Arable	N	N	**Y**	**Y**
23	A3	Arable	N	N	**Y**	**Y**
24	A2	Arable	N	N	**Y**	N
25	A4	Arable	N	N	**Y**	N
26	A5	Arable	N	N	**Y**	N
27	A10	Arable	N	N	N	N
28	A12	Arable	N	N	N	N
29	A16	Arable	N	N	N	**Y**
30	A19	Arable	N	N	**Y**	N
31	A20	Arable	N	N	**Y**	N
32	W4	Woodland	N	N	N	N
33	W5	Woodland	N	N	N	N
34	W9	Woodland	N	N	N	N
35	W19	Woodland	N	N	N	N
36	W22	Woodland	N	N	N	N
37	W23	Woodland	N	N	N	N
38	W24	Woodland	N	N	N	N
39	W46	Woodland	N	N	N	N
40	W49	Woodland	N	N	N	N
41	W74	Woodland	N	N	N	N
42	W91	Woodland	N	N	N	N
43	USDA6	Symbiotic	**Y**	**Y**	**Y**	N
44	GOR1	Symbiotic	**Y**	**Y**	**Y**	N

Lane Number	Isolate	Origin of isolate	*nifH*	*nodD*	*nirK*	*nosZ*
45	GOR2	Symbiotic	**Y**	**Y**	**Y**	N
46	GOR4	Symbiotic	**Y**	N	**Y**	N
47	BRM1	Symbiotic	**Y**	**Y**	N	N
48	BRM2	Symbiotic	**Y**	**Y**	N	N
49	RD5-C2	2,4-D degrader	N	N	N	N

7.3.2.1 Nitrogenase, *nifH*

The *nifH* gene is involved in nitrogen fixation; the conversion of nitrogen gas to ammonia and was absent from all isolates except the symbiotic isolates USDA 6 (43), GOR1 (44), GOR2 (45), GOR4 (46), BRM1 (47) and BRM2 (48) (Table 7.3 & Figure 7.5).

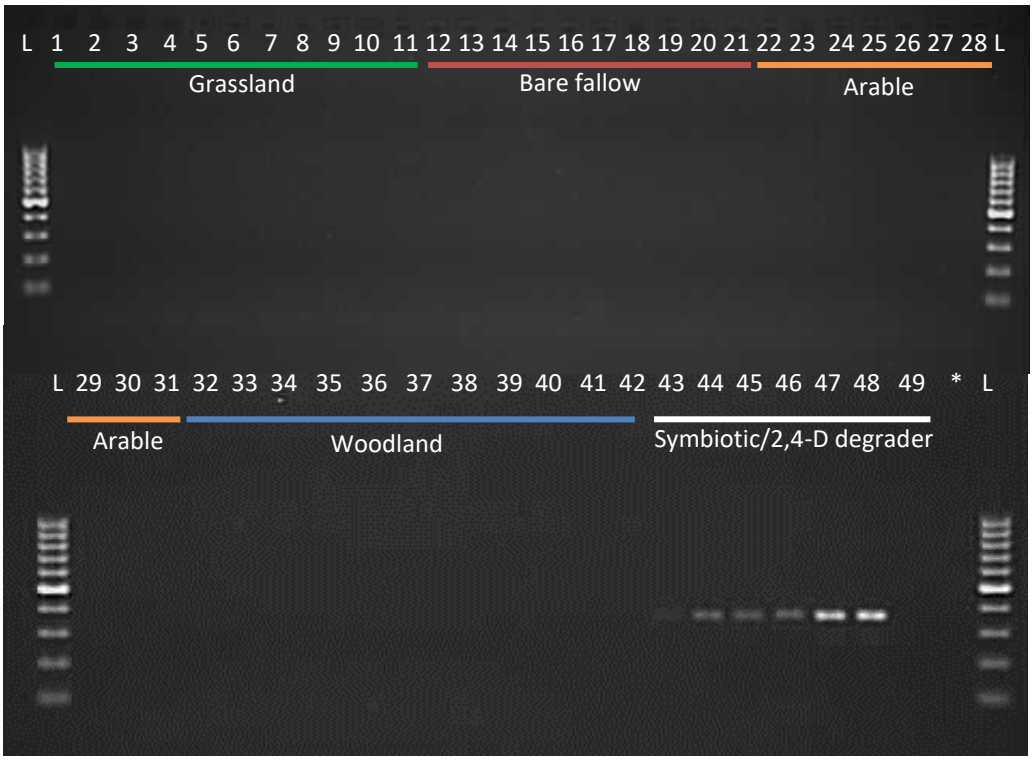


Figure 7.5: Agarose electrophoresis gel used in screening for *nifH*.
A band indicates a positive for *nifH*.
The numbers correspond with the isolates listed in Table 7.3.
L=100bp ladder *=negative control

7.3.2.2 Nodulation regulation, *nodD*

The *nodD* gene is a regulatory protein and is vital in the nodulation process. This gene was only present in USDA 6 (43), GOR1 (44), GOR2 (45), BRM1 (47) and BRM2 (48) (Table 7.3 & Figure 7.6).

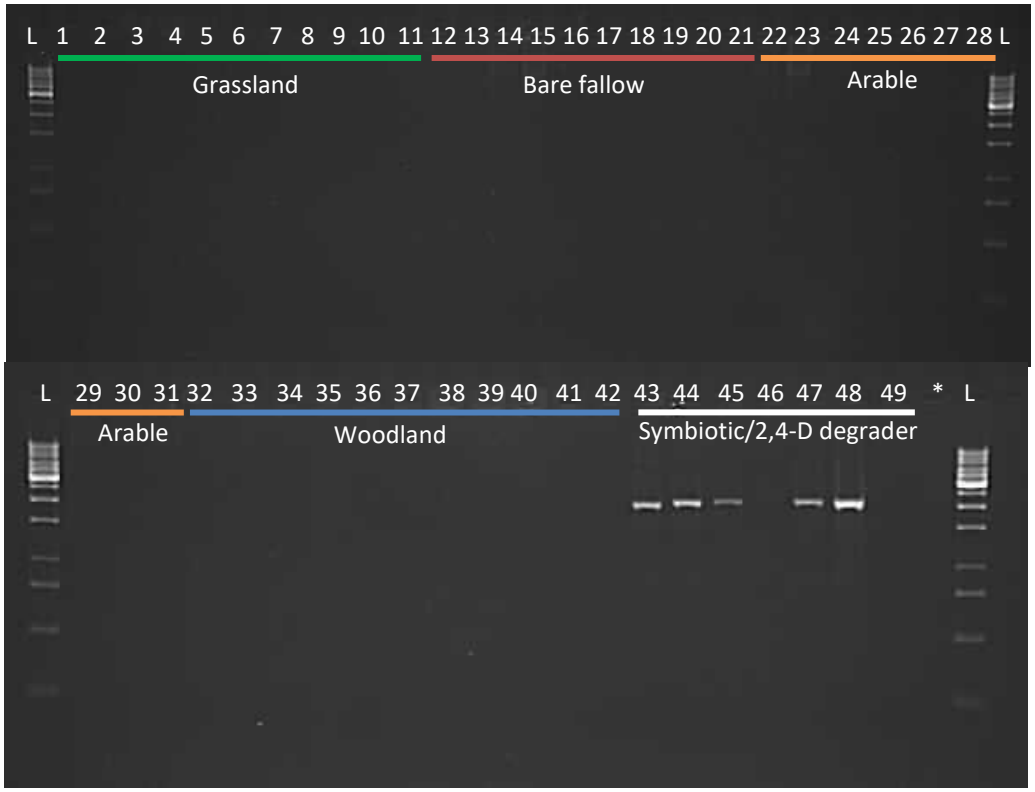


Figure 7.6: Agarose electrophoresis gel used in screening for *nodD*.
A band indicates a positive for *nodD*.
The numbers correspond with the isolates listed in Table 7.3.
L=100bp ladder *=negative control

7.3.2.3 Nitrite reductase, *nirK/S*

The *nirK* and *nirS* genes code for nitrite reductase which catalyses the same step in denitrification; nitrite reduction to nitric oxide. The difference between these genes is that *nirK* contains copper as the redox-active transition metal and *nirS* contains a *c*-type and *d*-type haem as the redox-active centre (Bedmar et al., 2005). The *nirK* gene is present in a number of the isolates in the culture collection: G2 (1), G3 (2), G6 (4), G8 (5), G15 (7), G22 (8), G31 (9), G46 (10), G54 (11), BF2 (12), BF6 (13), BF28 (16), BF49 (20), A1 (22), A3 (23), A2 (24), A4 (25), A5 (26), A19 (30), A20 (31), USDA 6 (43), GOR1 (44), GOR2 (45), GOR4 (46). There was also a faint band in lane 40 (W49) (Figure 7.7 & Table 7.3). The *nirS* PCR failed for both primer sets tested and this is probably due to a primer issue. There are very few *Bradyrhizobium nirS* sequences in the database and so to design primers for this is problematic.

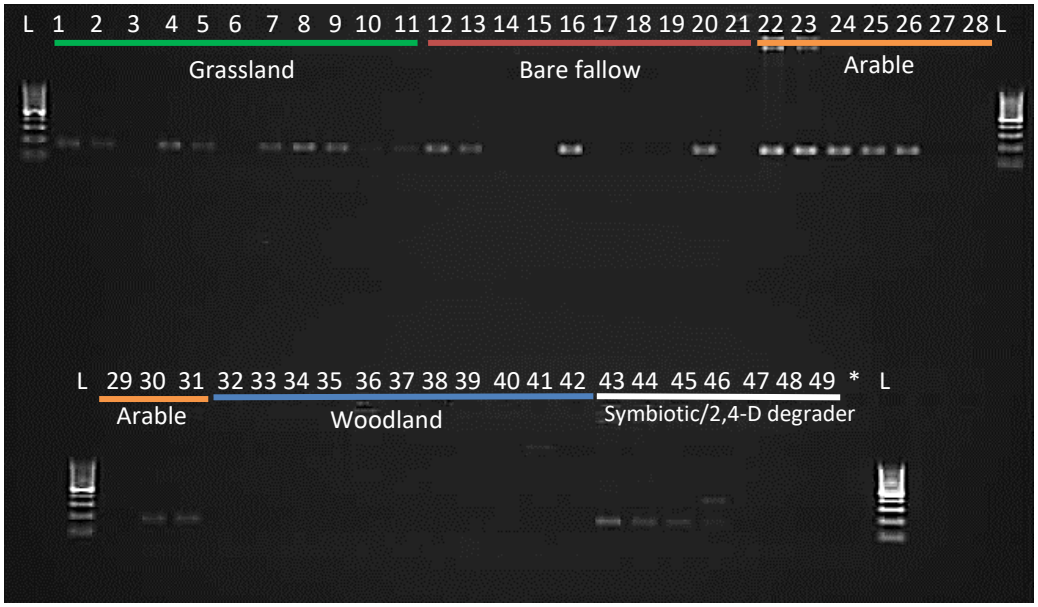


Figure 7.7: Agarose electrophoresis gel used in screening for *nirK*.
A band indicates a positive for *nirK*.
The numbers correspond with the isolates listed in Table 7.3.
L=100bp ladder *=negative control

7.3.2.4 Nitrous oxide reductase, *nosZ*

The *nosZ* gene is the final stage of denitrification which encodes an enzyme, nitrous oxide reductase, which catalyses the conversion of nitrous oxide to nitrogen gas, was present in only BF6 (13), BF49 (20), BF52 (21), A1 (22), A3 (23) and A16 (29) (Table 7.3 & Figure 7.8).

Figure 7.8: Agarose electrophoresis gel used in screening for *nosZ*.
A band indicates a positive for *nosZ*.
The numbers correspond with the isolates listed in Table 7.3.
L=100bp ladder *=negative control

These PCR results showed that there were potential differences in presence or absence of key genes in the denitrification process and therefore should be examined further. This led to looking at anaerobic growth on nitrate containing plates and analysing nitrous oxide emissions from broth cultures.

7.3.3 Denitrification plate assay

Differences in growth score according to whether nitrate was present in the medium was tested using a Wilcoxon-Mann Whitney test and it was found to significantly affect the growth score (W=49143, *P*-value=0.003) (Figure 7.9). Fewer isolates were scored 5 for growth when the medium contained nitrate compared to when the medium did not contain nitrate. Isolates grown on nitrate were graded 2, 3 and 4 more often than when grown on medium not containing nitrate.

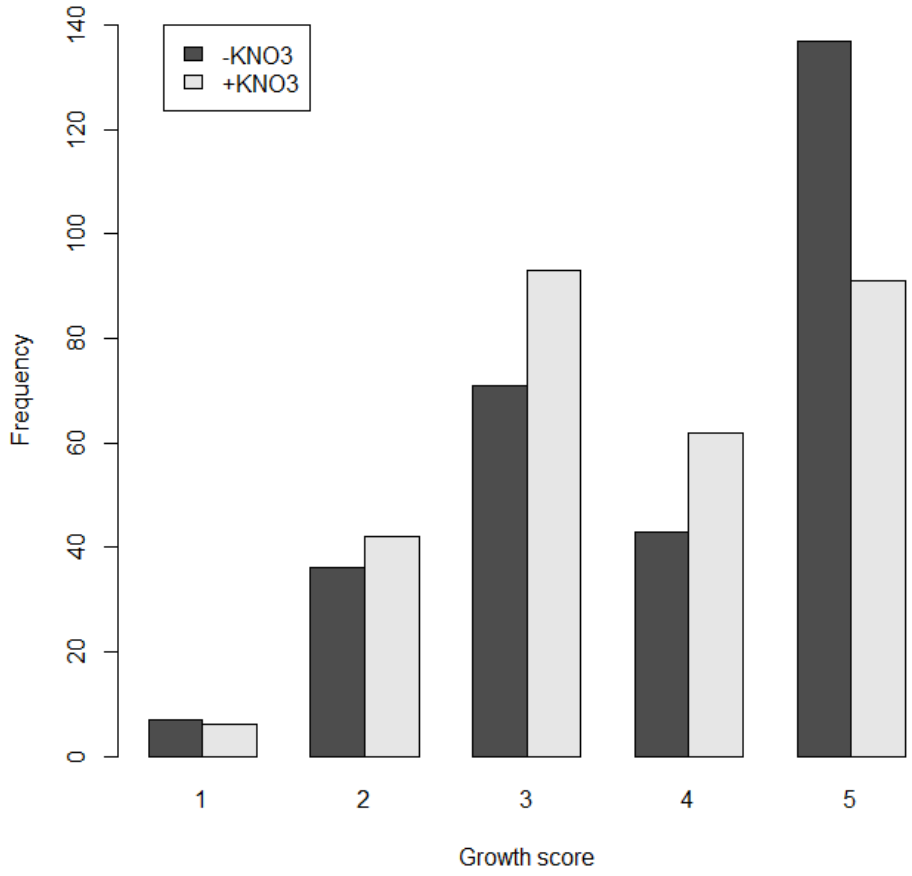


Figure 7.9: Bar plot showing the number of cultures per growth score and whether the medium contained nitrate.

The presence of oxygen was shown by a Wilcoxon-Mann Whitney test to not significantly affect the growth score (W=39456, *P*-value=0.06) (Figure 7.10). More isolates grown aerobically were scored 4 or 5 for growth than those grown anaerobically. More isolates scored 2 for growth when grown anaerobically compared to aerobically. However, these differences are not statistically different. This shows that the presence of nitrate is more important for growth than whether the growth conditions are aerobic.

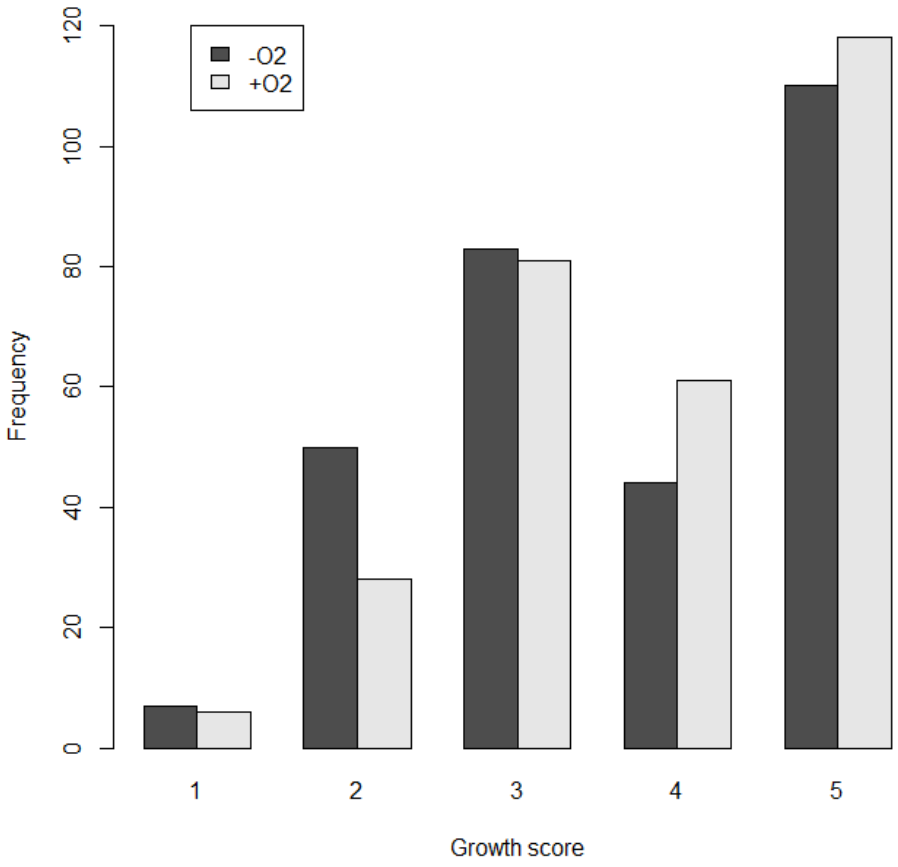


Figure 7.10: Bar plot showing the number of cultures per growth score and whether they were incubated aerobically or anaerobically.

The different growth treatments; nitrate and aerobic, nitrate and anaerobic, no nitrate and aerobic, no nitrate and anaerobic, were tested using a Kruskal-Wallis test to determine if they significantly affect the growth scores. This was shown to be statistically significant (Chi-squared=12.739, df=3, *P*-value = 0.005) (Figure 7.11). Isolates were more likely to be scored 5 if the medium did not contain nitrate. Differences in the other growth scores were not as large as those present at growth score 5.

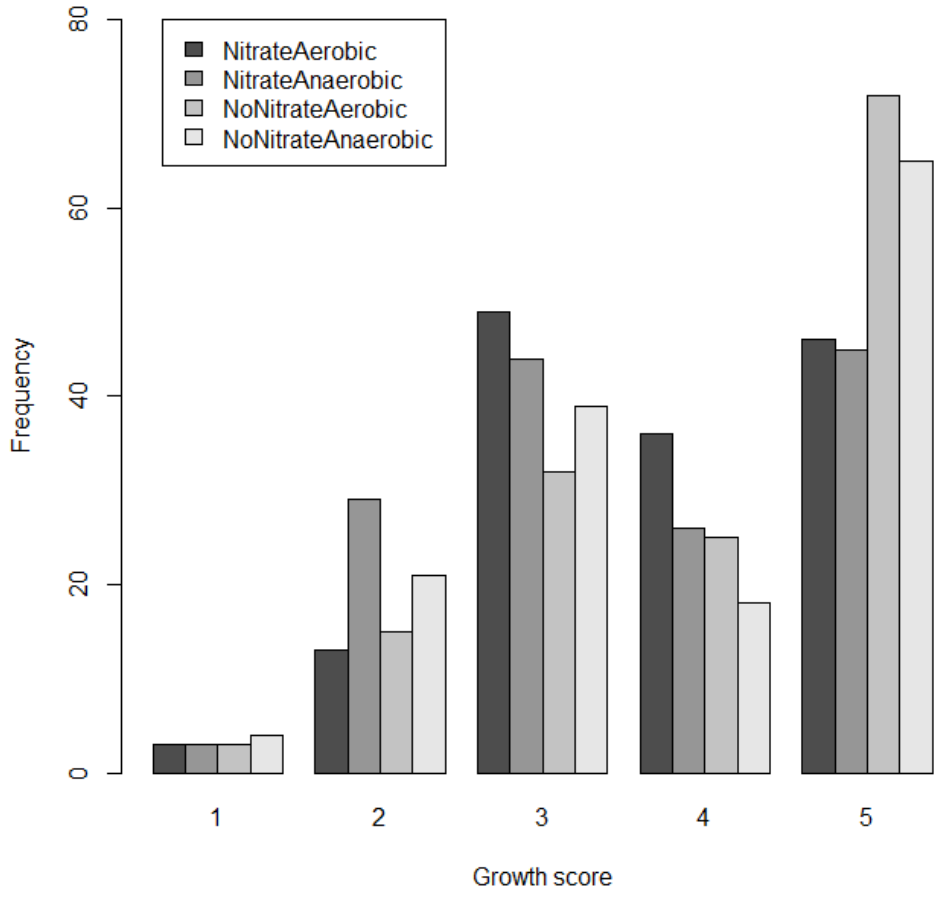


Figure 7.11: Bar plot showing the number of cultures per growth score and for each growth treatment.

The woodland isolates rarely were scored 4 or 5 for growth. This only occurred twice where there was no nitrate present in the medium (Figure 7.12-C & Figure 7.12-D). The bare fallow isolates have similar distribution of growth score across all growth treatments. There is also a similar distribution of growth scores for arable and grassland isolates over treatments with nitrate (Figure 7.12-A & Figure 7.12-B). When no nitrate is present, the arable and grassland isolates are more likely to be scored 4 or 5 for growth (Figure 7.12-C & Figure 7.12-D). The symbiotic isolates were more likely to score 4 or 5 across all growth treatments.

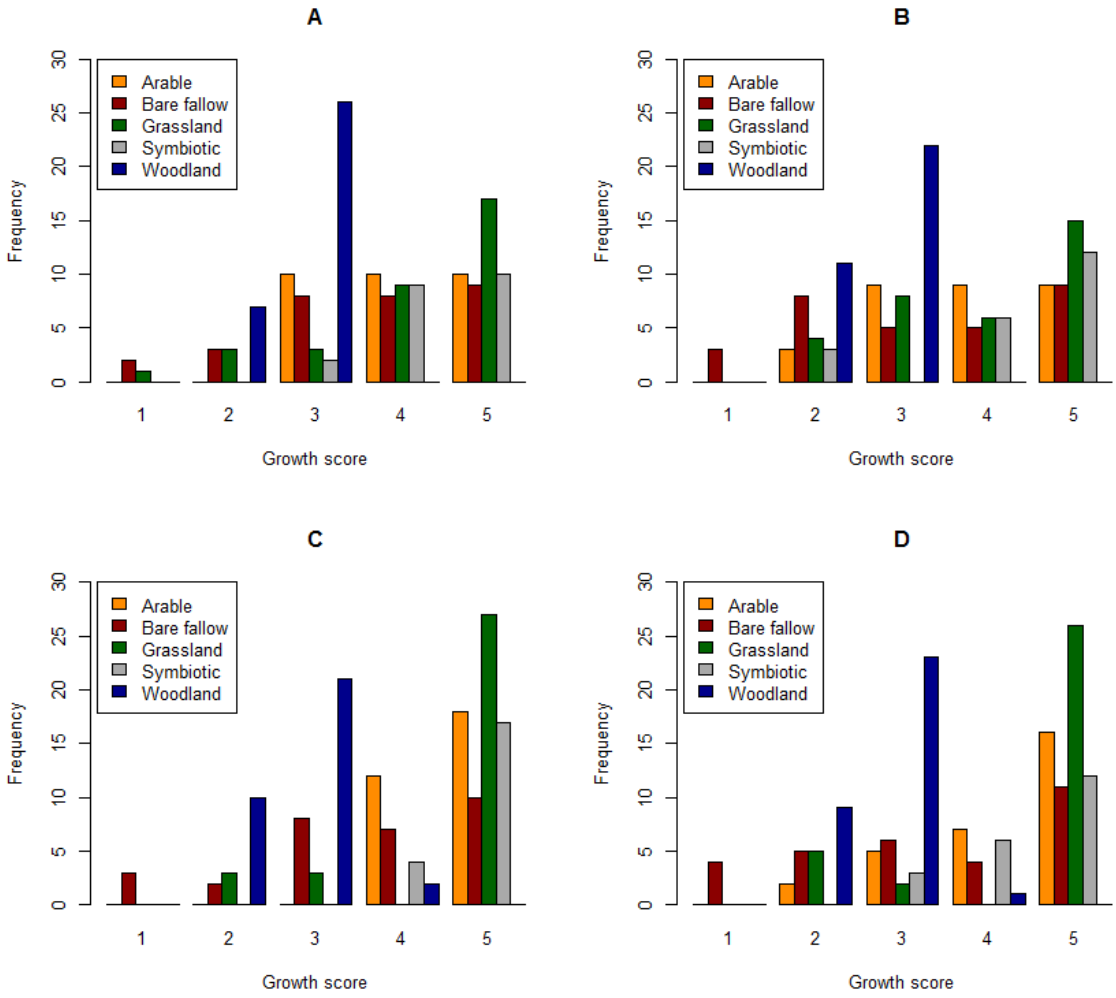


Figure 7.12: Bar plot showing the number of cultures per growth score for each soil treatment and for each growth treatment.
A: Nitrate and aerobic, B: Nitrate and anaerobic, C: No nitrate and aerobic, D: No nitrate and anaerobic.

It was also noted that some isolates had an altered colour when grown in the presence of nitrate. Grassland isolate G3, bare fallow isolate BF2 and symbiotic isolates USDA 6 and GOR1 all showed orange/brown colouration when grown on potassium nitrate both aerobically and anaerobically. Grassland isolate G2 and G15, arable isolates A4 and A19 and reference strains RD5-C2 and GOR2 only showed orange/brown colouration when grown on nitrate aerobically. Grassland isolates G22 and G54 and arable isolate A20 only showed orange/brown colouration when grown on potassium nitrate anaerobically. Figure 7.13 and Figure 7.14 shows examples of this colouration.

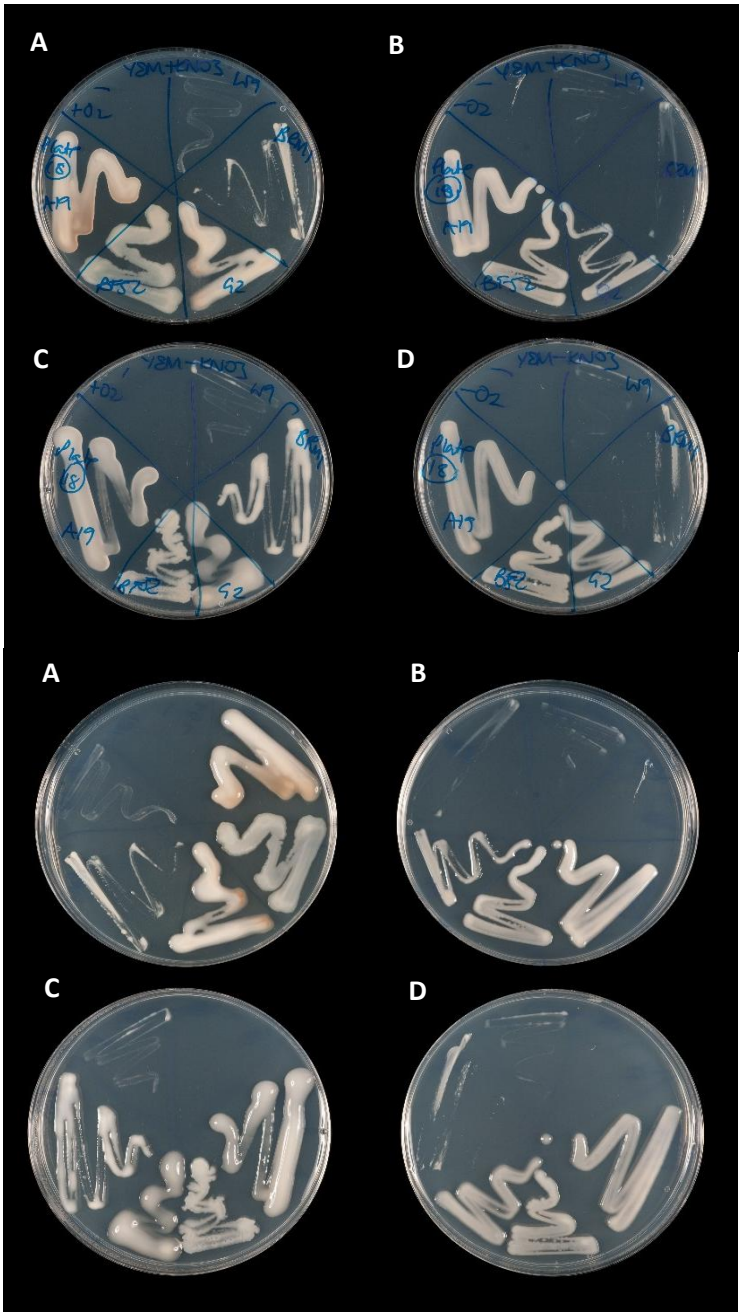


Figure 7.13: Photographs of four plates containing the same combinations of isolates incubated in all growth conditions.
Plates were photographed from underneath and without the lid. A: Nitrate, aerobic, B: Nitrate, anaerobic, C: No nitrate, aerobic, D: No nitrate, anaerobic.

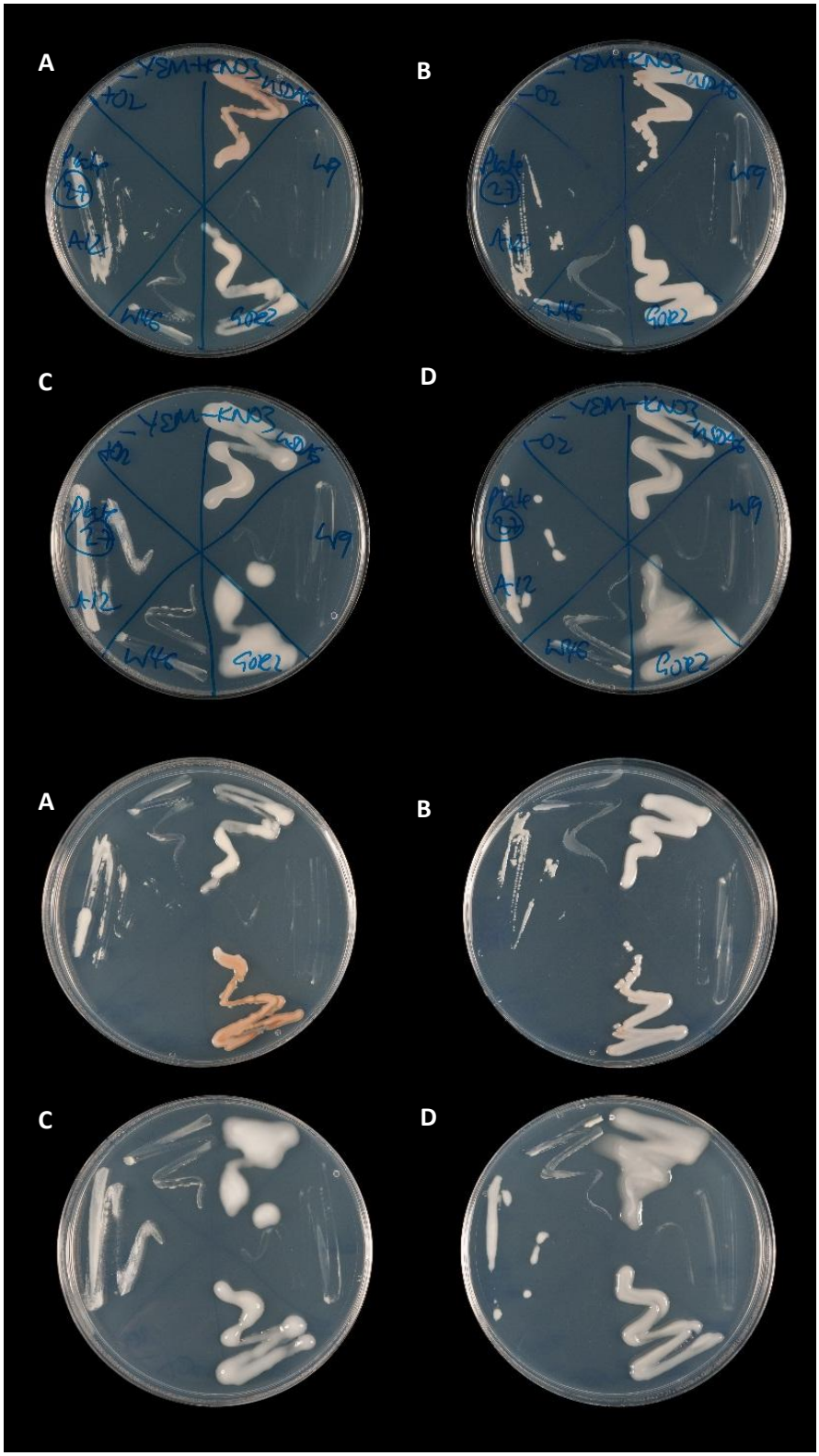


Figure 7.14: Photographs of four plates containing the same combinations of isolates incubated in all growth conditions.

Plates were photographed from underneath and without the lid. A: Nitrate, aerobic, B: Nitrate, anaerobic, C: No nitrate, aerobic, D: No nitrate, anaerobic.

7.3.4 Nitrous oxide emissions assay

ANOVA was used to test if there was a significant difference in the level of N_2O production between cultures with and without C_2H_2 added to the headspace for each isolate (Table 7.4). A total of 9 isolates showed a significant difference. These were one grassland isolate (G5), four bare fallow isolates (BF6, BF26, BF48 and BF49), two arable isolates (A1 and A3) and two woodland isolates (W22 and W74) suggesting that these isolates have a functional *nosZ* gene. Figure 7.15 shows the level of N_2O produced on a log scale with and without C_2H_2. Isolate G2, G46, BRM2 and RD5-C2 were removed for the analysis as there was an autosampler fault and the exetainer lids had become loose for replicate 2.

Table 7.4: ANOVA results testing for difference in N_2O level with and without C_2H_2

Isolate	F-value	Degrees of freedom	*P*-value	*nosZ* present (PCR result)	*nosZ* active
G3	5.732	1	0.0748	N	N
G5	**322**	**1**	**<0.001**	**N**	**Y**
G6	0.993	1	0.375	N	N
G8	0.968	1	0.381	N	N
G9	5.331	1	0.0821	N	N
G15	0.001	1	0.983	N	N
G22	0.316	1	0.604	N	N
G31	1.122	1	0.349	N	N
G54	0.248	1	0.644	N	N
BF2	0.067	1	0.809	N	N
BF6	**35.32**	**1**	**0.004**	**Y**	**Y**
BF20	0.003	1	0.956	N	N
BF26	**1146**	**1**	**<0.001**	**N**	**Y**
BF28	0.963	1	0.382	N	N
BF30	2.579	1	0.184	N	N
BF41	2.024	1	0.228	N	N
BF48	**9.227**	**1**	**0.0385**	**N**	**Y**
BF49	**35.7**	**1**	**0.00394**	**Y**	**Y**
BF52	**0.75**	**1**	**0.435**	**Y**	**N**
A1	**31.39**	**1**	**0.00499**	**Y**	**Y**
A2	0.561	1	0.496	N	N
A3	**500.8**	**1**	**<0.001**	**Y**	**Y**
A4	5.363	1	0.0815	N	N
A5	0.51	1	0.515	N	N
A10	0.593	1	0.484	N	N
A12	1.352	1	0.31	N	N

Isolate	F-value	Degrees of freedom	*P*-value	*nosZ* present (PCR result)	*nosZ* active
A16	**1**	**1**	**0.374**	**Y**	**N**
A19	0.447	1	0.541	N	N
A20	0.924	1	0.391	N	N
W4	0.835	1	0.413	N	N
W5	3.497	1	0.135	N	N
W9	0.213	1	0.668	N	N
W19	6.138	1	0.0684	N	N
W22	**36.16**	**1**	**0.00385**	**N**	**Y**
W23	0.163	1	0.707	N	N
W24	1.695	1	0.263	N	N
W46	0.964	1	0.382	N	N
W49	1.129	1	0.348	N	N
W74	**20.36**	**1**	**0.0107**	**N**	**Y**
W91	0.008	1	0.935	N	N
USDA 6	0.001	1	0.971	N	N
GOR1	0.376	1	0.573	N	N
GOR2	0.094	1	0.774	N	N
GOR4	7.366	1	0.0533	N	N
BRM1	0.113	1	0.753	N	N
BLANK	0.149	1	0.719	N	N

The bar plot shows that there was variation in the amount of nitrous oxide produced (Figure 7.15). The woodland isolates generally produced less nitrous oxide. The amount of nitrous oxide produced was variable within soil treatment and this is probably due to growth rate (Discussed in Chapter 6). The level of nitrous oxide produced was above the maximum in the calibration curve which was 50 ppm and therefore the amount of nitrous oxide recorded should be taken as a relative measure of nitrous oxide production and reduction. A number of isolates also recorded higher levels of nitrous oxide without C_2H_2; G6, A16, A19 and BF41. The difference between with and without C_2H_2 however, was not significantly different.

Some of the isolates recorded low level of nitrous oxide similar to the blank samples and so the carbon dioxide measurements for these were assessed to determine if the cells were viable. These isolates were: G9, A16, W19, W22, W23, W24 and W74. All other isolates produced nitrous oxide and so we can conclude that for these, there is a functional nitrite reductase gene (*nirK/S*) present.

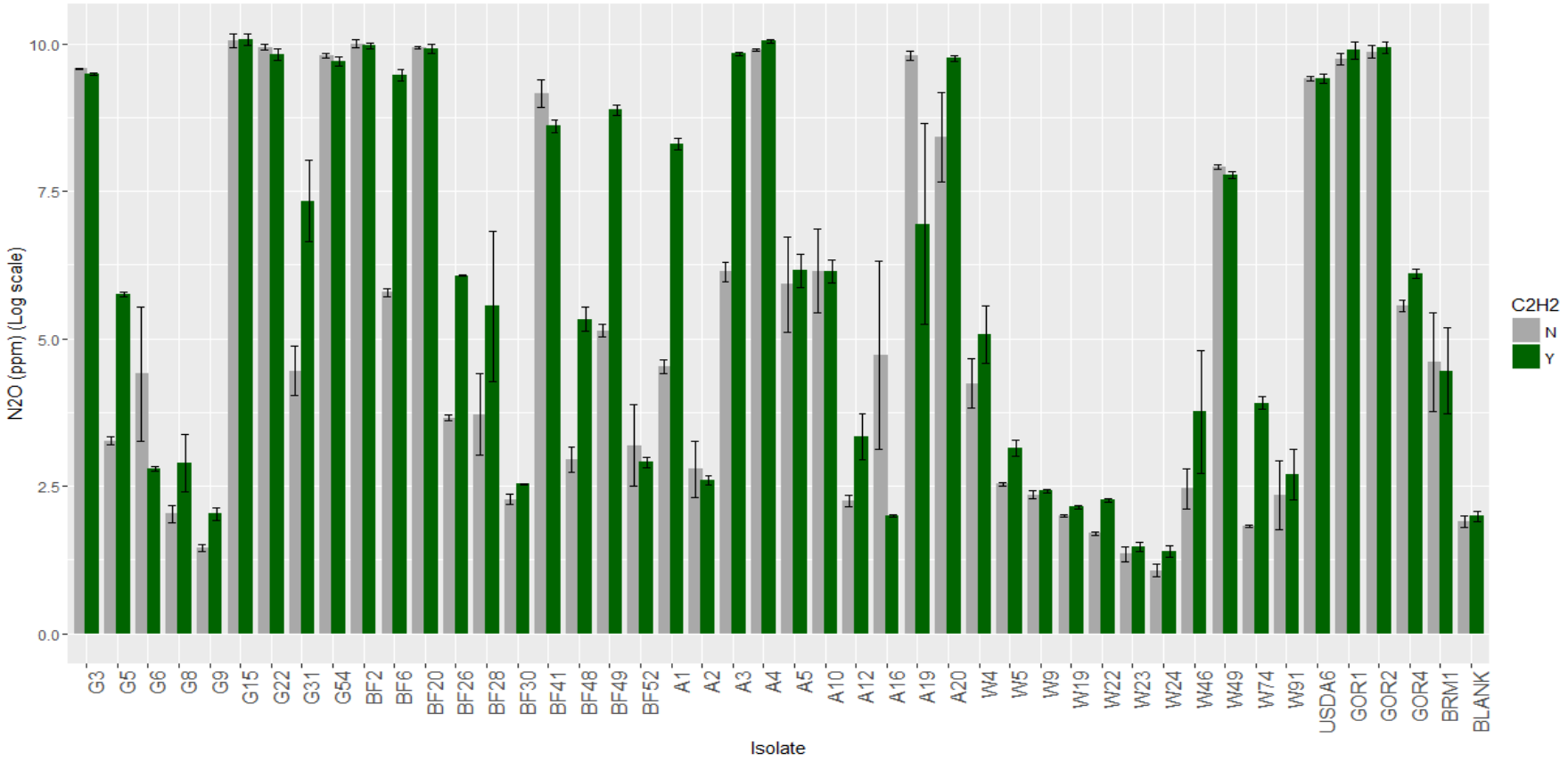


Figure 7.15: Bar plot showing the level of nitrous oxide produced (log scale) for each isolate.
Grey bars are cultures without acetylene added to the headspace and green are cultures with acetylene added. Error bars show the standard error for each mean.

Those isolates showing a significant difference in level of nitrous oxide with and without acetylene were examined more closely (Figure 7.16). For all of these isolates, levels of nitrous oxide were higher in the culture containing acetylene than those without acetylene and so we can be confident that this is due to the presence of a functional *nosZ* gene. The differences noted in W22 were small but were statistically significant but this could be due to low levels of variation in the data. At least one isolate from each soil treatment (grassland, bare fallow, arable, woodland) was found to have a functional *nosZ* gene and none of the symbiotic isolates contained a functional *nosZ* gene.

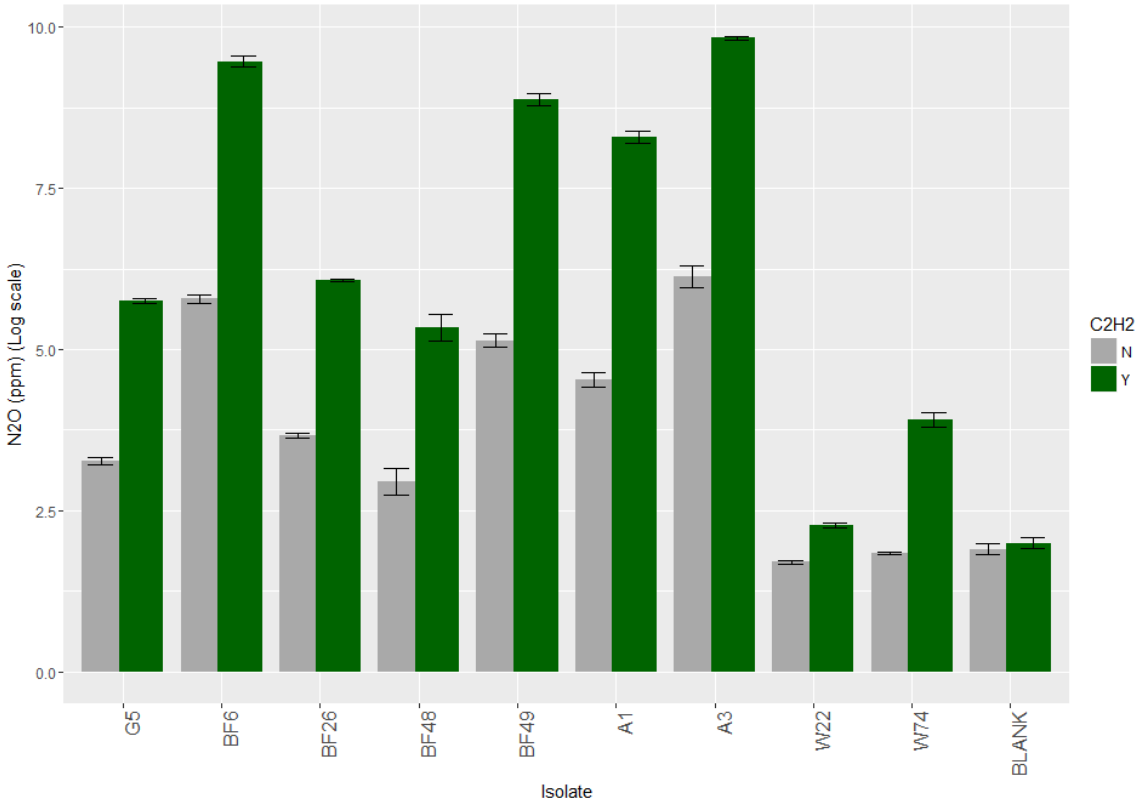


Figure 7.16: Bar chart showing the level of N_2O produced (log scale) for selected isolates.
Each isolate showing a significant difference in N_2O with and without C_2H_2 added to the headspace is shown. Grey bars are cultures without C_2H_2 added to the headspace and green are cultures with C_2H_2 added. Error bars show the standard error for each mean.

The carbon dioxide measurements were assessed to determine whether the cultures were no longer viable and this was the explanation of the low level of nitrous oxide produced in relation to the blank samples. Figure 7.17 shows the amount of CO_2 produced both with and without C_2H_2 for G9, A16, W19, W22, W23, W24 and W74. A one sample t-test was used (using with and without acetylene) to test whether these samples were significantly different from the blank samples. G9 (t=1.2473, df=5, *P*-value=0.1338), A16 (t=0.479, df=5, *P*-value=0.326), W19 (t=1.414, df=5, *P*-value=0.108), W24 (t=1.713, df=5, *P*-value=0.074) and W74 (t=-0.719, df=5, *P*-value=0.748) were all not significantly different from the blank sample. This shows that these cultures were not viable. Two woodland isolates were significantly greater than the mean CO2 production for the blank samples; W22 (t=2.788, df=5, *P*-value=0.019) and W23 (t=3.2196, df=5, *P*-value=0.0012).

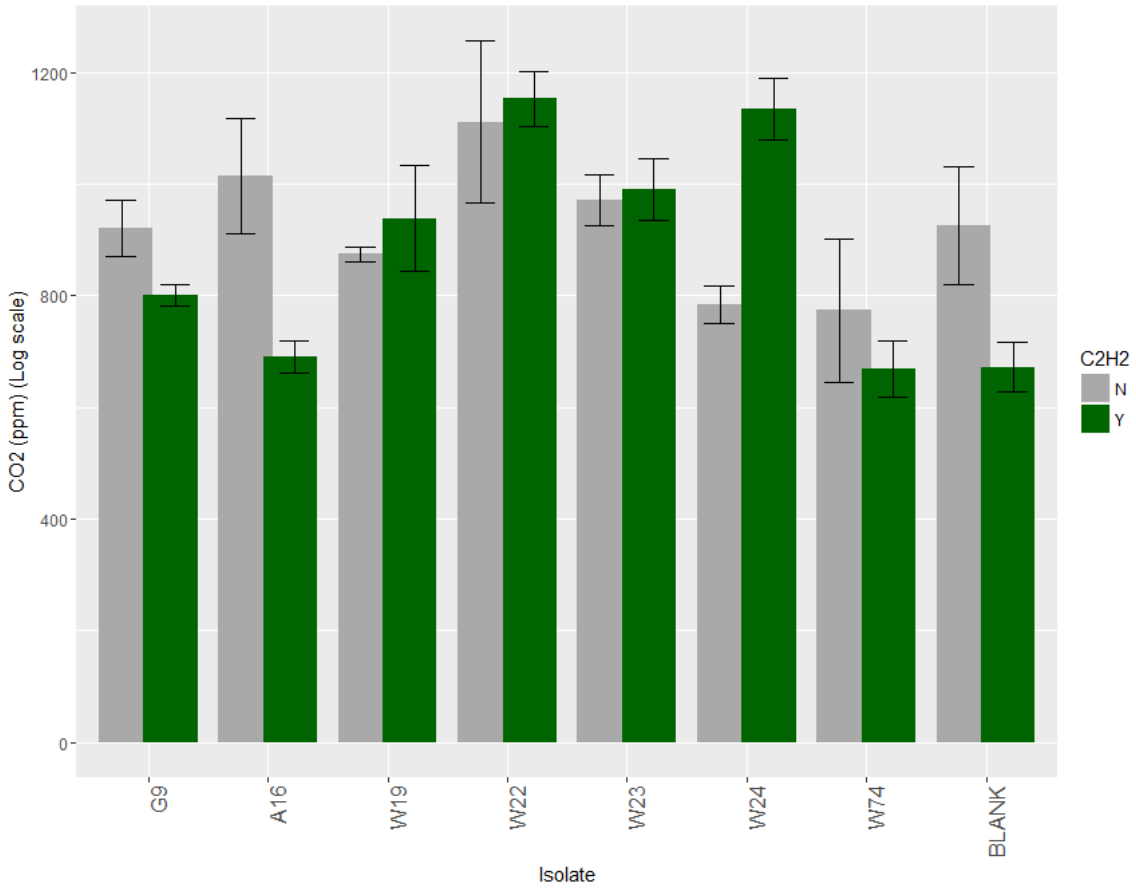


Figure 7.17: Bar chart showing the level of CO_2 produced for each isolate recording a lower N_2O reading than the blank samples.
Grey bars are cultures without C_2H_2 added to the headspace and green are cultures with C_2H_2 added. Error bars show the standard error for each mean.

The carbon dioxide measurements were assessed to determine whether the cultures were viable for isolates which produced differing levels of N_2O with and without C_2H_2. Figure 7.18 shows the amount of CO_2 produced both with and without C_2H_2 for G5, BF6, BF26, BF48, BF49, A1, A3, W22 and W74. A one sample t-test was used (using with and without acetylene) to test whether these samples were significantly different from the blank samples. G5 (t=1.2473, df=5, *P*-value=0.1338) and W74 (t=1.2473, df=5, *P*-value=0.134) were not significantly different from the blank sample. This shows that these culture were not viable. All others were significantly different from the blank samples; BF6 (t=7.118, df=5, *P*-value<0.001), BF26 (t=2.250, df=5, *P*-value=0.037), BF48 (t=2.0294, df=5, *P*-value=0.049), BF49 (t=2.1405, df=5, *P*-value=0.043), A1 (t=9.906, df=5, *P*-value<0.001), A3 (t=3.515, df=5, *P*-value=0.009) and W22 (t=2.788, df=5, *P*-value=0.019).

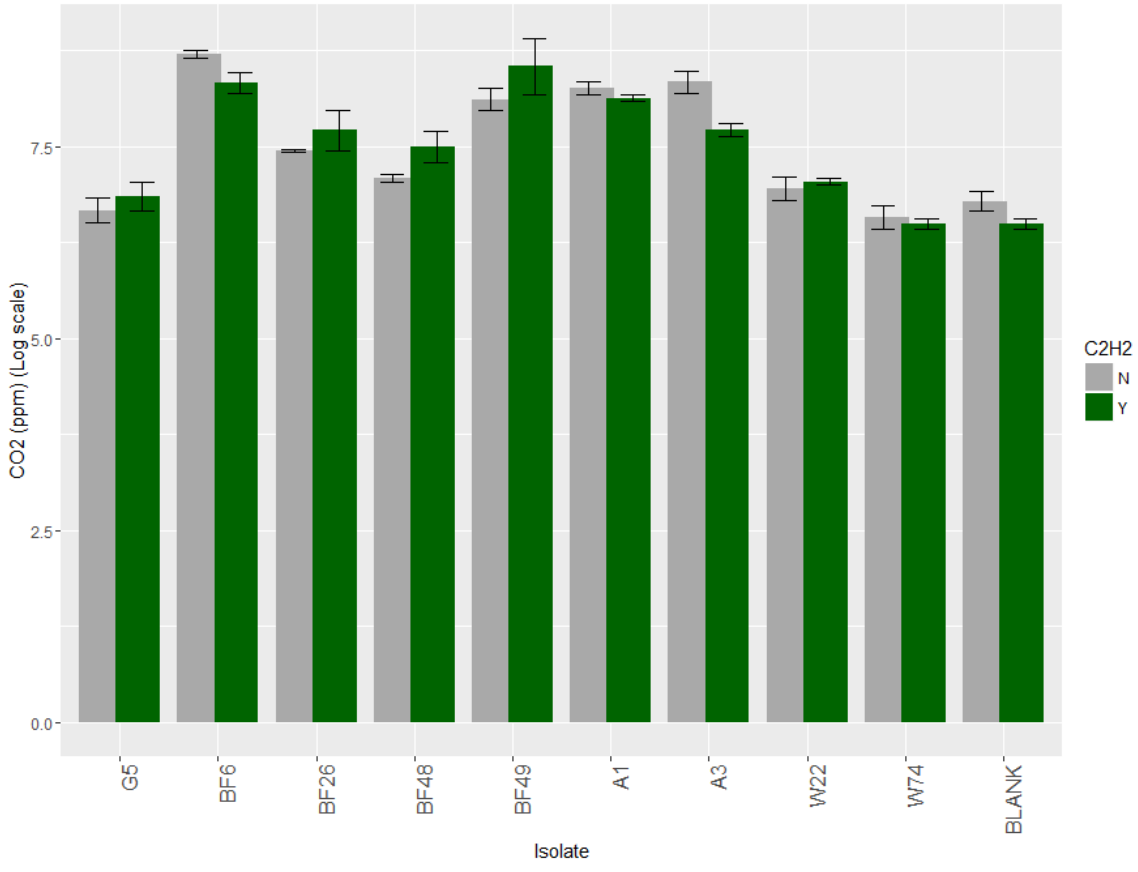


Figure 7.18: Bar chart showing the level of CO2 produced for each isolate recording a lower N2O reading than the blank samples.
Grey bars are cultures without C2H2 added to the headspace and green are cultures with C2H2 added. Error bars show the standard error for each mean.

7.3.5 Nitrous oxide reductase copy number in Highfield and Broadbalk soil

The woodland and grassland soils had a significantly higher copy number than the other soil treatments (Woodland: GLM, Estimate=1.2763, *P*<0.001, Grassland: GLM, Estimate=1.4097, *P*<0.001). The bare fallow had significantly lower *nosZ* copy number (GLM, Estimate=-1.5287, *P*=0.013) (Figure 7.19). The arable soil had 21.8×10^6, bare fallow had 4.7×10^6, grassland had 89.1×10^6 and woodland had 78×10^6 *nosZ* copies per gram of dry weight soil.

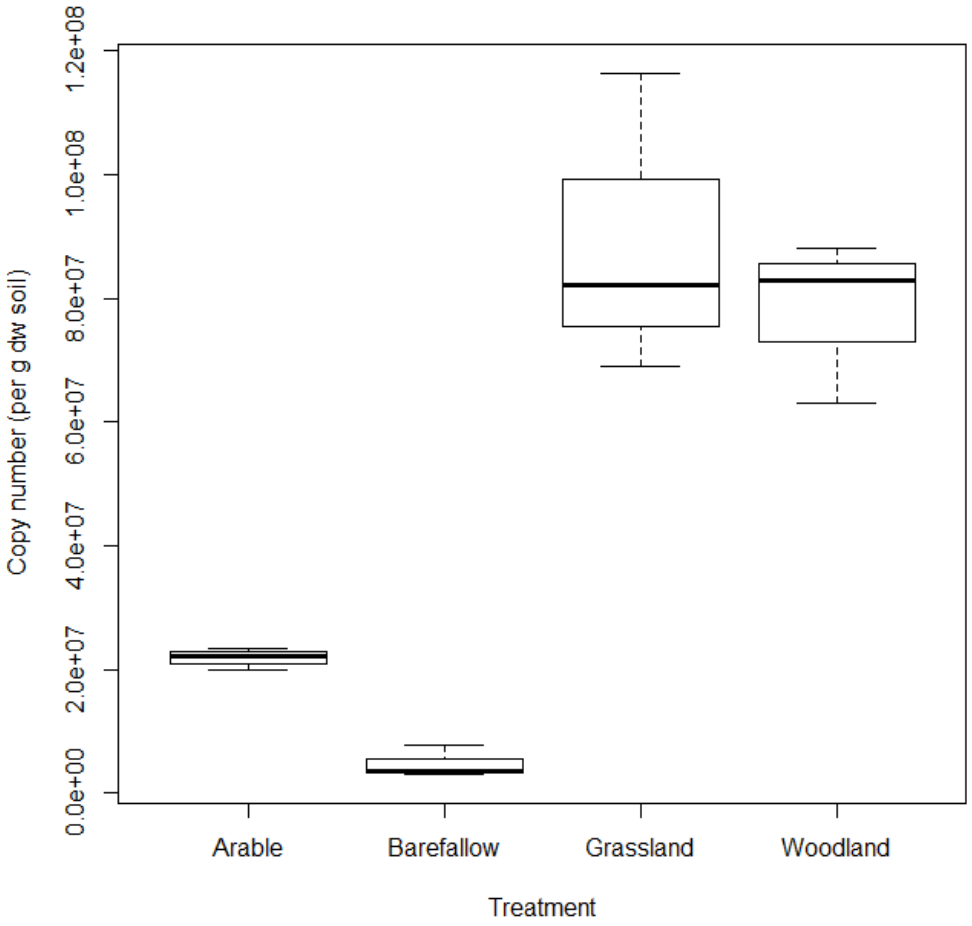


Figure 7.19: Boxplot showing the copy number of *nosZ* in each soil treatment per gram of dry weight soil.

7.4 Discussion

7.4.1 Nitrogen fixation

Nitrogen fixation (*nifH*) and nodulation (*nodD*) genes were absent from all strains which were isolated from soil; the only *Bradyrhizobium* isolates which contained one or both of these genes were isolated from gorse or broom bushes or the USDA 6 strain which was originally isolated from a soybean nodule (Kaneko et al., 2011). This confirms that these free-living *Bradyrhizobium* isolates from the Highfield and Broadbalk sites are not involved in these processes and have either lost these genes or never acquired them. The lack of nodulation genes also explains why nodules were not formed on soybean plants grown in Highfield soil rather than the lack of a suitable host.

The addition of *B. japonicum* to the grassland soils allowed the soybean plants have a higher above ground biomass than plants which were inoculated but grown in bare fallow soils. This difference was not seen in plants which were not inoculated and suggests that the plants are primarily nitrogen limited in grassland soils however other nutrients in addition to nitrogen may be limiting in the bare fallow soil and the poor soil structure might cause additional problems by impeding root growth.

7.4.2 Growth on nitrate

Overall, the aim of the plate test was to determine whether the isolates could use nitrate as the terminal electron acceptor in anaerobic conditions. Denitrification yields the highest amount of energy after oxygen reduction and therefore growth using nitrate as the terminal election acceptor is lower than growth in aerobic conditions (Koike and Hattori, 1975, Strohm et al., 2007). It was found that the presence of nitrate in the growth medium reduced growth in all cultures: whether the cultures were incubated aerobically or anaerobically had no effect. All cultures were able to grow in all growth treatments. In anaerobic conditions where nitrate was not present fermentation could be the source of energy. *Bradyrhizobium* has been shown to readily ferment and has been suggested as a way of preparing inoculants (Graham-Weiss et al., 1987). Woodland isolates had lower growth scores which correspond to the growth rate analysis in Chapter 6. Growth treatment (the combination of aerobic and anaerobic, with and without nitrate) had no effect on bare fallow isolates. Nitrate limited growth for grassland and arable isolates: when there

was no nitrate added to the medium, these isolates were scored highly (4 or 5) for growth. When nitrate was present there was an even spread of growth scores for grassland and arable isolates.

Some isolates also showed a different morphology when grown on nitrate and this was noted when incubated both aerobically and anaerobically. *B. japonicum* USDA 6 is known to have pigment genes (Kanehisa and Goto, 2000, Kanehisa et al., 2016) and the expression of one or more of these genes could be responsible for the colour change. No other records of this colour change have been found in the literature and so the ecological importance for this is unclear.

7.4.3 Nitrous oxide emissions

The nitrous oxide emissions assay showed that a functional *nosZ* gene was present in eight isolates representing the Highfield soil treatments: grassland (1), bare fallow (4), woodland (1) and arable (2). The woodland isolate W74 which also showed a significantly different N_2O production with and without C_2H_2 were found to produce similar levels of carbon dioxide to the blank samples and therefore was considered not a viable culture. This result is different from the PCR gene screening for *nosZ* where only three bare fallow and three arable isolates gave a positive result. Of the six isolates which were shown to be positive from the PCR gene screenings, two were shown to not have a functional *nosZ* gene under the assay conditions; A16 and BF52.

The denitrification pathway is controlled by a network of transcriptional regulators. These regulators are inhibited by high oxygen levels and are expressed in anoxic conditions (Vairinhos et al., 1989, Delgado et al., 2003, Bedmar et al., 2005, Bakken et al., 2012, Delmotte et al., 2014). The protein responsible for regulating *nosZ* is NosR (encoded by *nosR*), a membrane-bound iron-sulphur flavoprotein allowing for electron transfer during nitrous oxide reduction (Wunsch and Zumft, 2005, Spiro, 2012). It is critical for NosZ synthesis and the expression of *nosR* is regulated by the DnrD transcription factor in *Pseudomonas stutzeri* (Wunsch and Zumft, 2005). The expression of the *nos* genes are not controlled by levels of nitrous oxide however, the accessory proteins in the *nos* operon are not well studied (Spiro, 2012). Isolates which were shown to be negative from the PCR but positive from the nitrous oxide emissions assay were BF26, BF48 and W22. This could be due to a primer issue resulting in no amplification (Ishii et al., 2011).

The presence or absence of uptake hydrogenase (Hup) has been previously suggested to be closely linked to denitrification ability with Hup-positive phenotypes being able to complete the full denitrification pathway and Hup-negative phenotypes only being able to perform nitrate reduction (van Berkum and Keyser, 1985, Sameshima-Saito et al., 2004). This pattern is not found for the culture collection as the bare fallow isolate BF49, was shown in Chapter 5 (Section 5.3.6)

to not contain *hup* genes required for synthesising uptake hydrogenase but did contain a functional *nosZ* gene.

All viable isolates produced nitrous oxide and therefore must have functional *napA*, *nirK/S* and *norB/C* genes in order to convert nitrate into nitrous oxide. The gene required for nitrate ammonification, the conversion of nitrite to ammonia, is *nrfA* (Mohan et al., 2004). A search in the NCBI database revealed that although members of the Rhizobiales contain this gene, this does not include *Bradyrhizobium.* This corresponds to the *nirK* PCR where the majority of the isolates gave a positive result. Some *Bradyrhizobium* isolates have been shown to carry *nirS* although the majority contain *nirK* (Velasco et al., 2001, Delgado et al., 2003, Bedmar et al., 2005, Sameshima-Saito et al., 2006, Ishii et al., 2011). The *nirK* PCR showed that not all isolates were positive for this gene and so we predict that the isolates negative for *nirK* contain *nirS.* However, we were unable to amplify any *nirS* products so whether the isolates contain *nirK* or *nirS* cannot be confirmed.

7.4.4 Denitrification ability

Bradyrhizobium has a definite role in denitrification and the isolates in the culture collection are no exception. The majority of the culture collection is missing the *nosZ* gene however eight isolates from a range of soil treatments were shown to have a functional *nosZ* gene which is important for mitigating nitrous oxide emissions which contribute to global climate change. The fitness benefits of the traits associated with the change between oxic and anoxic respiration is not clear but phenotyping strains under standard conditions has been suggested to attempt to clarify these benefits (Bergaust et al., 2011, Bakken et al., 2012). One benefit could be the ability to reduce levels of cytotoxic nitric oxide (Meakin et al., 2006, Fernandez et al., 2008). Denitrification ability in *Bradyrhizobium* has been extensively studied although the majority of the literature focuses on symbiotic strains of *Bradyrhizobium* rather than free-living strains as examined in this chapter. Forest isolates which were shown to be free-living and not to contain either nitrogen fixation or nodulation genes, were also shown to lack the entire denitrification pathway (VanInsberghe et al., 2015). This is in contrast to the woodland isolates in the culture collection; all viable cultures emitted nitrous oxide which shows that these isolates are involved in the truncated denitrification pathway and one woodland isolate was able to reduce nitrous oxide into nitrogen gas.

The qPCR results showed that there were significant differences in *nosZ* copy number between the soils with the bare fallow soil containing significantly lower *nosZ* copies per gram of soil. The higher copy numbers seen in the grassland, arable and woodland soils could be *nosZ* copies which

are not functional as qPCR amplifies only DNA. It has been previously reported that woodland soils emitted larger amounts of nitrous oxide compared with arable plots with varying levels of fertiliser inputs on the Broadbalk experiment at Rothamsted Research (Clark et al., 2012). Only one woodland isolate had a functional *nosZ* gene and all produced nitrous oxide which supports the report of high levels of nitrous oxide measured in the Broadbalk woodland compared to arable soil (Clark et al., 2012). Another explanation for significantly lower copy number in bare fallow soil but high proportion of the culture collection containing *nosZ,* could be that the abundance of bacteria in bare fallow soil was shown to be lower than both arable and grassland soil and so it could be that a higher proportion of *Bradyrhizobium* have *nosZ* but the absolute abundance of bacteria is lower than is seen in both arable and grassland soils (Hirsch et al., 2009). A qPCR on *Bradyrhizobium* 16S is needed to determine whether this hypothesis is correct.

Despite *Bradyrhizobium* being able to rapidly denitrify, it is thought that the numbers found in soils are unlikely to be high enough to significantly affect nitrous oxide emissions in isolation (Breitenbeck and Bremner, 1989). The presence of *nosZ* has been previously shown to co-occur alongside *nirS* more often than with *nirK* and this is thought to be due to *nirS* presence being strongly linked to *nor* presence which results in the production of nitrous oxide (Graf et al., 2014). However, in *Bradyrhizobium* it is more common to contain *nirK* than *nirS* (Bedmar et al., 2005, Monza et al., 2006, Sameshima-Saito et al., 2006, Ishii et al., 2011, Kraft et al., 2011). Around 30% of 652 genomes assessed showed to only contain *nosZ* and neither *nir* or *nor* (Graf et al., 2014). However, this is not the case with this culture collection as all viable isolates were able to produce nitrous oxide. Overall, soil treatment does not appear to affect the denitrification ability of *Bradyrhizobium* and isolates from all soil treatments are denitrifiers within these communities.

8 General Discussion

8.1 Summary

The 16S analysis shows that all of the strains isolated from the Highfield and Broadbalk soils and the gorse and broom root nodules were *Bradyrhizobium* however, a wider range of genes should be sequenced in order to assess diversity. The 16S rRNA gene is useful as an identifier of taxonomy however, as it is a highly conserved gene which is slow to evolve, other core genes should be used to assess phylogeny (Fox et al., 1980, Lane et al., 1985, Kundig et al., 1995). The ERIC sequence DNA fingerprinting technique is useful as it gives an overview of full genome diversity between strains. These sequences are assumed to be in non-coding regions of the genome and so a high level of diversity in these regions is expected (Versalovic et al., 1991). Assessing the banding profile is subjective and different resolutions will give varying phylogeny results leading to over or under estimations of diversity within the culture collection. Also, as it is a PCR based method, DNA is needed from all strains which are going to be compared at the same time; this limits the scope of the comparisons. It is not possible to include isolates from the database which is possible for protein-encoding sequences which are submitted to public databases and can be downloaded and included in phylogenetic analyses. The ERIC sequence DNA fingerprinting gel for the *Bradyrhizobium* culture collection showed considerable variation and that each soil treatment contained isolates with different banding profiles showing that there is a high level of diversity in each soil treatment. A lower level of diversity was seen in the symbiotic isolates which were from gorse and broom root nodules; however, the nodules were taken from the same plant and so to fully assess the diversity of UK symbiotic *Bradyrhizobium* isolates, many more nodules from a range of plants should be assessed. The gorse and broom isolates in this project were included as originally we had access to only *B. japonicum* USDA 6 as a symbiotic representative. Symbiotic isolates were not the focus of the project and so including a small number of these isolates from one gorse and one broom plant was sufficient for this purpose and provided a geographically relevant comparison. The ERIC sequence is an appropriate method as an initial step for assessing diversity but in order to fully understand the heterogeneity in the group of isolates in question, protein encoding genes should be sequenced.

The multilocus sequence analysis showed that there was a high level of diversity in protein encoding genes for core functions. The five genes which were sequenced are often chosen in multilocus sequence analysis as they provide variation between strains and so can be used to distinguish between samples. None of the phylogenies for any of the individual genes or the

concatenated gene sequences grouped the isolates according to soil treatment. This is consistent with the results from the ERIC sequence dendrogram where non-consistent banding profiles were seen for each soil treatment. This shows that each soil treatment is home to a range of *Bradyrhizobium* phylotypes. The genes sequenced showed differing levels of diversity; for example, *recA* was found to have low resolution with many clades containing single isolates. A previous study (Menna et al., 2009) showed *atpD* had high levels of diversity attributed to recombination of this gene. The high level of diversity seen across the culture collection is consistent with other diversity studies of *Bradyrhizobium* and with each study taking place, the estimates of diversity increase (Giongo et al., 2008, Menna et al., 2009, Menna and Hungria, 2011). Other diversity studies have focused on culture collections of isolates from a range of hosts and to our knowledge no multilocus sequence analysis has been carried out on a culture collection dominated by free-living and non-symbiotic isolates. Rhizosphere and symbiont recruitment is carried out from the pool of isolates in the soil and understanding more about this pool is vital to understanding the interactions taking place between above and below ground communities.

In a previous study looking at the same five genes as sequenced in this project in symbiotic *Bradyrhizobium*, *dnaK* was found to be the most informative (Rivas et al., 2009). The *dnaK* gene has been previously used to assess taxonomy in addition to the 16S rRNA gene (Moulin et al., 2004): for my culture collection, the *dnaK* phylogeny showed some clustering based on 16S top BLAST hit however, it had a much lower resolution than the *gyrB, atpD* and *rpoB* phylogenies. The *rpoB* gene in this project was found to have the highest level of diversity, presumably sequence heterogeneity and primer mis-match is why it was not possible to get sequences for three of the isolates from the culture collection. Many of the primers which were from the Rivas et al. (2009) paper had to be redesigned in order to get a product to be sent for sequencing despite these primers being used in previous published work on *Bradyrhizobium* isolates (Rivas et al., 2009). Overall, this confirms that *Bradyrhizobium* is a genus with high sequence diversity: within my culture collection, redesigned primers were necessary on a subset of the isolates from all soil treatments. Housekeeping genes grouping isolates congruently with 16S rRNA gene phylotype has been previously noted in symbiotic *Bradyrhizobium* using *atpD*, *glnII*, *recA* and *rpoB* (Vinuesa et al., 2008). This corresponds to the concatenated loci phylogeny in this project. The core genome of the isolates did not appear to be affected by soil treatment. This genus is able to be involved in a wide range of functions and therefore a high level of genetic diversity is needed to adapt to the heterogeneous environment found in soil (Kowalchuk et al., 2002, Hirsch et al., 2010).

Housekeeping genes give an insight into the level of diversity however, it is the functions linked to their niche which would give clues to how the environment influences bacterial communities.

The growth rate analysis hinted at the influence of the soil treatment: the woodland isolates had significantly lower growth rates and maximum OD readings than the other soil treatments. There were no differences between the three soil treatments from Highfield (grassland, arable and bare fallow). This reflects a survival strategy for woodland isolates where many of the energy sources are from complex carbon sources like lignin and tannins (Deschaseaux and Ponge, 2001, Waldrop and Firestone, 2004)**.** A major difference in the soils from the different treatments on Highfield is the level of organic carbon where bare fallow soil contained 82% less and arable soil contained 70% less than grassland soil (Hirsch et al., 2009). The Biolog assays showed that the largest amount of variation is determined by substrates involved in central metabolism suggesting that growth rate was an important factor in the first principal component which explains the greatest amount of the variation in the data. The second principal component was determined by many amino acid substrates which were important in determining differences between symbiotic and non-symbiotic isolates. There were three substrates that only the symbiotic isolate tested was able to utilise and these were all amino acids and the symbiotic isolate used a high number of substrates. This could be due to the presence of photosynthate containing a range of carbon, nitrogen and other micronutrients present within the nodules and released by the plant during symbiosis: the bacterium maintains pathways to be able to take advantage of a rich carbon supply which is often limiting to microbial growth in soil (Alden et al., 2001, Ekblad and Nordgren, 2002, Ilstedt and Singh, 2005, Demoling et al., 2007, Delmotte et al., 2010).

The carbon metabolism assays showed that grassland isolates were more consistent in their utilisation profiles and the bare fallow and arable isolates were more variable. The woodland isolates appeared to use very few substrates. The woodland isolates had a significantly reduced growth rate and this could mean that their metabolism of these substrates were not able to be recorded within the timeframe of the experiment. Guenet et al. (2011) showed that bare fallow communities were able to metabolise straw and cellulose despite fresh organic matter not being applied to the bare fallow soil for 80 years. Both carbohydrates and carboxylic acids were important substrate guilds and were utilised more than any other guilds (amino acids, amines and amides, polymers and miscellaneous). Rhizobia are known to be able to utilise a range of substrates even if they are in low concentration due to the high number of ABC transporters present (Prell and Poole, 2006). The bare fallow soil contains low levels of organic carbon and therefore bacteria isolated from this soil will be adapted to being opportunistic having to scavenge carbon when it becomes available (Hirsch et al., 2009). The grassland is a nutrient rich

environment with a constant supply of carbon through rhizodeposits and exudates from plants (Hogberg and Read, 2006, Lakshmanan et al., 2014). The carbon metabolism results give clues to the carbon sources previously or currently present in the grassland, arable, bare fallow and woodland soils. The majority of the literature using Biolog focuses on community analysis or medical strains. To date, to our knowledge, a culture collection dominated by free-living, non-symbiotic strains has not been tested using Biolog MicroPlates.

A high level of diversity in 16S gene sequences, multilocus sequence analysis, growth rates and carbon metabolism assays was seen and so delving deeper into the genomes of this genus should highlight differences in key pathways of interest relevant to both agriculture and the environment. Initially it was predicted that G22 would have a smaller genome which would be tailored to the substrate-replete environment it lives in. It was also predicted initially that BF49 would have a larger genome because of selective pressure to retain more pathways that allowed for opportunistic carbon scavenging in a severely carbon depleted environment. Soil bacteria sequenced have shown to have large genome sizes and typical genomes are between 6.5 Mb and 9 Mb with nitrogen-fixing bradyrhizobia typically being 9Mb whereas the minimum genome size for *Escherichia coli* is 4 Mb (Perna et al., 2001, Kaneko et al., 2002, Gonzalez et al., 2006, Young et al., 2006, Kaneko et al., 2011). However, once the genomes were assembled, this was the opposite; G22 was much larger (G22: 9 Mbp, BF49: 7.5 Mbp) and similar in size to the symbiotic isolates in the database (USDA 110: 9 Mb). The bare fallow isolate appears to have gone through a genome streamlining process to retain only genes which are essential, reducing the energy burden of DNA repair and maintenance. Genome streamlining is a hypothesis that there is a strong selection pressure for minimising cellular respiration costs in some environments (Mira et al., 2001, Giovannoni et al., 2005). The oceanic bacterium *Pelagibacter ubique* is a member of the SAR11 clade with the smallest genome size of this clade. Despite, the small size, due to a median intergenic spacer size of just three base pairs, nearly all basic functions are included. The small genome size is thought to be due to severe nutrient limitation and the need for efficient resource use (Giovannoni et al., 2005). This suggests that examining BF49 in more detail could provide clues to surviving in this hostile environment.

The genome analysis also highlighted differences in carbon metabolism supported by the Biolog assays. In particular, malonic acid which is a root exudate often release by legumes (Dimroth and Hilbi, 1997, Kim, 2002, Suvorova et al., 2012). Only the grassland isolate G22 contained genes involved in the metabolism of this substrate and it was only grassland isolates which were able to utilise this substrate. The G22 genome also contained genes required for the production of marinocine, a broad spectrum antibacterial compound. These genes were missing from the bare

fallow genome suggesting that the grassland isolate could require these genes to gain a competitive advantage over other members of the grassland microbial community. By sequencing the genomes, it allowed for a deeper understanding of the processes these isolates have the potential to be involved in and by comparing genomes of free-living isolates to symbiotic strains could help identify parts of the genomes where the environment is a driver of evolution and adaptation.

Bradyrhizobium is traditionally known as a legume symbiont however, the isolates obtained from Highfield soil lacked both nodulation and nitrogen fixation genes and the inability to form nodules on soybeans (*G. max*). This is consistent with recent genomes being sequenced from American woodland soils (VanInsberghe et al., 2015) and in contrast with the majority of other sequenced *Bradyrhizobium* strains which contain either both *nif* and *nod* genes (Kaneko et al., 2002, Kaneko et al., 2011, Torres et al., 2015) or just *nif* genes (Giraud et al., 2007, Okubo et al., 2012, Okubo et al., 2013). The American woodland strains were sequenced from shotgun sequencing of the soil and were not from cultured samples, therefore functional tests cannot be carried out on these isolates. The metagenome and metatranscriptome data obtained previously from the Highfield plots (unpublished data from the Hirsch group at Rothamsted Research, Ian Clark can be contacted for more information), showed that *Bradyrhizobium* was one of the most abundant and active genera and the lack of involvement in the nitrogen fixation and nodulation processes shows that this group must be participating in other soil processes.

Denitrification is a soil process which is of particular importance globally and especially in agriculture due to the loss of nitrate from the soil, greenhouse gas emissions and the potential mitigation of these emissions. Denitrification genes were shown to be present in both genome-sequenced isolates which differentiates them from the American woodland strains which were missing the entire denitrification pathway (VanInsberghe et al., 2015). The major difference in presence or absence of denitrification genes was the final stage of denitrification; the conversion of nitrous oxide to nitrogen gas, which is catalysed by nitrous oxide reductase encoded by the *nosZ* gene. This step of denitrification is important for mitigating nitrous oxide emission from agriculture. The denitrification process begins with the conversion of nitrate into nitrite, although this step of the denitrification is not unique to denitrifiers. When oxygen is limited, nitrate can be used for growth by rhizobia and *Bradyrhizobium* is able to denitrify nitrate: the presence of oxygen inhibits denitrification as it competes with nitrate, nitrite, nitric oxide and nitrous oxide to be the terminal electron acceptor (Vairinhos et al., 1989, Delgado et al., 2003, Bedmar et al., 2005, Delmotte et al., 2014). The first unique step of denitrification is the dissimilatory reduction of nitrite to nitric oxide, and the ability to grow anaerobically on nitrate or nitrite is an indication

of this. Denitrification ability is common across *Bradyrhizobium* and so the ability to grow on nitrate anaerobically was expected. Using nitrate as an electron accepter instead of oxygen is less efficient and so lower growth scores on medium containing nitrate was an expected result. A colouration change was noted in a range of isolates from different soil treatments including *B. japonicum* USDA 6 when grown on nitrate. Pigment genes have been found in other bacteria to be involved in protection against UV radiation and photosynthetic strains of *Bradyrhizobium* have been shown to produce bacteriochlorophyll and carotenoids for photosynthesis (Mohammadi et al., 2012, Noisangiam et al., 2012). *B. japonicum* USDA 6 is known to have pigment genes (Kanehisa and Goto, 2000, Kanehisa et al., 2016) but is not photosynthetic and so the ecological significance of expressing pigment genes in response to nitrate is unknown. No reports of changes in colouration in response to nitrate can be found in the literature to date.

The qPCR results showed that there were significant differences in *nosZ* copy number between the soils with the bare fallow soil containing significantly fewer *nosZ* copies per gram of soil. The abundance of bacteria in bare fallow soil was shown to be lower than both arable and grassland soil and so it could be that a higher proportion of *Bradyrhizobium* have *nosZ* but the absolute abundance of bacteria is lower than is seen in both arable and grassland soils (Hirsch et al., 2009). A qPCR of *Bradyrhizobium* 16S is needed to determine whether this hypothesis is correct. One issue with this method is that the *nosZ* copies may not represent functional genes, as the qPCR amplified only DNA; mRNA extraction followed by reverse transcriptase and qPCR would be needed along with functional assays to determine whether *Bradyrhizobium* is a major denitrifier in these communities.

Non-symbiotic forest isolates from American forest soils were shown to lack the entire denitrification pathway in contrast to the woodland isolates in our culture collection (VanInsberghe et al., 2015). All viable woodland isolates emitted nitrous oxide which shows that these isolates are at least involved in the truncated denitrification pathway. The fitness benefits of the traits associated with the change between oxic and anoxic respiration is not clear but one benefit could be the ability to reduce levels of cytotoxic nitric oxide (Meakin et al., 2006, Fernandez et al., 2008, Bergaust et al., 2011, Bakken et al., 2012). The nitrous oxide emissions assay showed that a functional *nosZ* gene was present in eight isolates representing all of the soil treatments and the majority of isolates in my culture collection are missing the *nosZ* gene. This means that nitrous oxide is the final product of this process rather than nitrogen gas. Despite *Bradyrhizobium* being able to rapidly denitrify, it is thought that the numbers found in soils are unlikely to be high enough to significantly affect nitrous oxide emissions in isolation (Breitenbeck

and Bremner, 1989). Overall, soil treatment does not appear to affect the denitrification ability of *Bradyrhizobium* and isolates from all soil treatments are denitrifiers within these communities.

These results of the nitrous oxide emissions assay also highlighted potential issues with screening for presence or absence using PCR. Of the six isolates which were shown to be positive for *nosZ* from the PCR gene screenings, two were not shown to have a functional *nosZ* gene. Three isolates were shown to be negative for *nosZ* from the PCR but positive from the nitrous oxide emissions assay. All isolates produced nitrous oxide and so must have either *nirK* or *nirS* but the PCR to amplify the *nirS* gene failed. The inability to obtain clear, single bands for sequencing means that which of these genes is present cannot be confirmed. Although most *Bradyrhizobium* have been shown to have *nirK,* some do contain *nirS* (Velasco et al., 2001, Delgado et al., 2003, Bedmar et al., 2005, Sameshima-Saito et al., 2006, Ishii et al., 2011). There are relatively few *Bradyrhizobium nirS* sequences available in the databases and so designing primers is not straightforward. Table 8.1 shows a summary of all nitrogen cycling results. Isolates where there was a difference between the *nosZ* PCR and the denitrification assay result are in bold and shaded in grey.

Table 8.1: Summary of nitrogen cycling gene PCR and denitrification assay results.

Isolate name	Soil treatment	*nifH* PCR	*nodD* PCR	*nirK* PCR	Anaerobic growth on nitrate	N_2O produced	*nosZ* PCR	*nosZ* active
G2	Grassland	N	N	Y	4	-	N	-
G3	Grassland	N	N	Y	3	Y	N	N
G5	**Grassland**	**N**	**N**	**N**	**3**	**Y**	**N**	**Y**
G6	Grassland	N	N	Y	5	Y	N	N
G8	Grassland	N	N	Y	4	Y	N	N
G9	Grassland	N	N	N	3	-	N	-
G15	Grassland	N	N	Y	5	Y	N	N
G22	Grassland	N	N	Y	5	Y	N	N
G31	Grassland	N	N	Y	3	Y	N	N
G46	Grassland	N	N	Y	2	-	N	-
G54	Grassland	N	N	Y	5	Y	N	N
BF2	Bare fallow	N	N	Y	5	Y	N	N
BF6	Bare fallow	N	N	Y	4	Y	Y	Y
BF20	Bare fallow	N	N	N	5	Y	N	N
BF26	**Bare fallow**	**N**	**N**	**N**	**2**	**Y**	**N**	**Y**
BF28	Bare fallow	N	N	Y	1	Y	N	N
BF30	Bare fallow	N	N	N	2	Y	N	N
BF41	Bare fallow	N	N	N	3	Y	N	N
BF48	**Bare fallow**	**N**	**N**	**N**	**2**	**Y**	**N**	**Y**
BF49	Bare fallow	N	N	Y	4	Y	Y	Y
BF52	**Bare fallow**	**N**	**N**	**N**	**5**	**Y**	**Y**	**N**

Isolate name	Soil treatment	*nifH* PCR	*nod* PCR	*nirK* PCR	Anaerobic growth on nitrate	N_2O produced	*nosZ* PCR	*nosZ* active
A1	Arable	N	N	Y	4	Y	Y	Y
A3	**Arable**	**N**	**N**	**Y**	**4**	**Y**	**Y**	**N**
A2	**Arable**	**N**	**N**	**Y**	**3**	**Y**	**N**	**Y**
A4	Arable	N	N	Y	5	Y	N	N
A5	Arable	N	N	Y	3	Y	N	N
A10	Arable	N	N	N	4	Y	N	N
A12	Arable	N	N	N	3	Y	N	N
A16	**Arable**	**N**	**N**	**N**	**3**	**Y**	**Y**	**N**
A19	Arable	N	N	Y	5	Y	N	N
A20	Arable	N	N	Y	5	Y	N	N
W4	Woodland	N	N	N	3	Y	N	N
W5	Woodland	N	N	N	2	Y	N	N
W9	Woodland	N	N	N	3	Y	N	N
W19	Woodland	N	N	N	3	-	N	-
W22	Woodland	N	N	N	2	-	N	-
W23	Woodland	N	N	N	3	-	N	-
W24	Woodland	N	N	N	3	-	N	-
W46	Woodland	N	N	N	3	Y	N	N
W49	Woodland	N	N	N	3	Y	N	N
W74	Woodland	N	N	N	3	-	N	-
W91	Woodland	N	N	N	3	Y	N	N
USDA6	Symbiotic	Y	Y	Y	5	Y	N	N
GOR1	Symbiotic	Y	Y	Y	5	Y	N	N
GOR2	Symbiotic	Y	Y	Y	5	Y	N	N
GOR4	Symbiotic	Y	N	Y	5	Y	N	N
BRM1	Symbiotic	Y	Y	N	3	Y	N	N
BRM2	Symbiotic	Y	Y	N	3	-	N	-
RD5-C2	2,4-D degrader	N	N	N	5	-	N	-

This highlights the importance of functional assays and not relying purely on culture-independent approaches. Culturing a number of isolates from a range of soils with varying treatments has allowed for a deeper understanding of the biology of this widespread and active group in soil. The metagenome and metatranscriptome analyses carried out on soil from the Highfield experiment indicated that *Bradyrhizobium* was abundant and active in all soil treatments irrespective of how the land was managed (unpublished data from the Hirsch group at Rothamsted Research, Ian Clark can be contacted for more information). The main goal was to determine whether the soil treatment and land management practices impacted on the diversity and role that this genus was involved in within the soil community. Metagenomics has increased the amount of information

available on soil communities. Highlighting key groups of interest and unusual, interesting pathways will increase the impact of metagenomics research and has a number of future applications for biotechnology and product discovery. However, culture-independent techniques should always be combined with culture-based methods to get a comprehensive understanding of function. By using the metagenomes from Highfield we were able to identify that *Bradyrhizobium* is a key group in soil and deserves further investigation.

Bradyrhizobium has been noted in numerous metagenomes as a genus of interest and was shown to be in high abundance in various soil environments in a number of papers in the literature. Zhalnina et al. (2013) used 16S rRNA amplicon sequencing and found that *Bradyrhizobium* was dominant in both agricultural and non-agricultural soils making up 0.5% and 1-1.5% of the population respectively. However, a comparison between cultivated with two tillage practices and non-cultivated soil found that *Bradyrhizobium* was more dominant in the arable soils than non-cultivated soils (Souza et al., 2016). Although, the arable soil was a soybean/maize rotation where enrichment of *Bradyrhizobium* populations is expected. High *Bradyrhizobium* abundance was seen in grassland soil at Rothamsted Research (4.89% ± 0.635%); other highly represented genera included *Rhodopseudomonas*, *Pseudomonas* and *Burkholderia* (Delmont et al., 2012). Forest soils have also been shown to be dominated by *Bradyrhizobium*. Metagenomics on high Andean forest soil showed that the community was dominated by *Mycobacterium, Streptomyces, Acidobacteria, Bradyrhizobium* and *Frankia,* with *Bradyrhizobium* making up 2.5-5.8% of the community depending on the annotation software used (Montana et al., 2012). Oak forest soil, both the rhizosphere (3.42-4.56%) and bulk soil (3.34-4.22%), are dominated by *Bradyrhizobium* in addition to *Acidobacteria* and *Rhodoplanes* (Uroz et al., 2010). Northern coniferous forest soil 16S rRNA amplicon sequencing showed that *Bradyrhizobium* are assigned to 15.2% of the sequences with *Rhodoplanes* accounting for 10% (Hartmann et al., 2012). This was the same experimental site where non-symbiotic *Bradyrhizobium* were found to dominate these communities (VanInsberghe et al., 2015). The isolation method showed that it is possible to target abundant groups in soil and isolate these strains for future experiments. Through culturing *Bradyrhizobium* from a range of soil treatments we were able to observe the cultures, notice changes in colouration and assay the cultures to look into pathways of interest.

Isolation of previously uncultured or understudied microorganisms is essential for understanding the diversity of the community, notwithstanding that the majority of soil microorganisms are thought to be unculturable (Janssen et al., 2002, Zengler et al., 2002, Joseph et al., 2003, Stevenson et al., 2004, Lewis et al., 2010). Altering nutrient concentrations, incubation time and presence of humic acids have enabled culture of previously unidentified and unstudied organisms

from Acidobacteria, Verrucomicrobia, Actinobacteria and Proteobacteria (Janssen et al., 2002, Joseph et al., 2003, Stevenson et al., 2004). Zengler et al. (2002) developed a novel method for culturing novel organisms using gel microdroplets to trap the cells from a range of environments including seawater and soil. This allowed for the culturing of organisms in the absence of faster-growing potentially inhibitory competitors, with the aim of understanding the physiology and metabolism of the microbial community. Another method for isolating hard to cultivate organisms is by using a diffusion chamber; a washer placed between two membranes enables microorganisms to be cultured in semi-solid agar and for it to be incubated in the native environment (Kaeberlein et al., 2002). This allows for nutrients to diffuse into the chamber to maximise the chance of culturing novel organisms with specific nutrient requirements which are hard to replicate in the laboratory.

The isolation chip (ichip) builds on the information obtained from the diffusion chamber method as it allows for high-throughput culturing of organisms through mini diffusion chambers. It is an assembly of flat plates with multiple holes covered in a membrane to allow diffusion of nutrients but not bacterial cells. The ichip is incubated in the native environment providing a culture medium suited to the community being analysed (Nichols et al., 2010). Recently, the ichip and the culturing of novel organisms has generated a large amount of media and scientific interest for its application in the discovery of novel products, antibiotics and the associated potential impacts on biotechnology (Kaeberlein et al., 2002, Zengler et al., 2002, Nichols et al., 2010, Ling et al., 2015). A novel antibiotic, teixobactin, has been identified from the previously uncultured bacterium, *Eleftheria terrae*, using the ichip. This antibiotic shows significant potential as *Staphylococcus aureus* and *Mycobacterium tuberculosis* showed no signs of resistance (Ling et al., 2015). The ichip is a sound method for culturing hard to cultivate organisms by providing culture conditions close to natural environmental conditions and has a huge potential for culturing environmental strains and uncovering the diversity of soil microorganisms. Targeting strains via a specific pathway, for example the trap plant method used for rhizobia, requires prior knowledge of the biology of the group and therefore will not be suitable for culturing microorganisms where little is known about their biology or function in the natural environment. Metagenomics and metatranscriptomics will aid this process, however, these -omics technologies often miss rarer members of the community as their abundance is too low for the sequencing technologies. Culturing allows for numerous characteristics to be measured within the laboratory including both molecular and biochemical functions. It also allows for the comparison of isolates within the collection (Gorlach et al., 1994).

Fortunately, non-symbiotic *Bradyrhizobium* proved to be amenable to growth on common culture media and in laboratory conditions. Overall, the results of this project show that the

Bradyrhizobium culture collection has high diversity in 16S rRNA gene sequence, housekeeping gene sequences, carbon metabolism and nitrogen cycling irrespective of soil treatment. The main influence of soil treatment was on growth rate, where woodland isolates showed significantly slower growth. These results have also highlighted the importance of both culture dependent and culture independent approaches to understanding the biology of *Bradyrhizobium*.

8.2 Future experiments and application of this project

A number of other experiments could be carried out to extend this work and to understand more about the biology of this group. *Bradyrhizobium* is also known for the presence of the *tfd* gene family which is involved in the degradation of 2,4-dichlorophenoxyacetic acid (2,4-D). This is a widely-used synthetic auxin that acts as a systemic herbicide. It is often used as a model compound for studying how microorganisms have evolved the capacity to able to degrade xenobiotic organic compounds in soil using *tfd* genes (Kamagata et al., 1997, Itoh et al., 2002). The *tfd* genes encode the enzyme 2,4-D alpha-ketoglutarate dioxygenase which catalyses the first step of 2,4-D mineralisation (Itoh et al., 2000, Zaprasis et al., 2010). The *tfdA* gene is widespread and is present alpha-, beta- and gamma-Proteobacteria and can be categorised into groups (Kamagata et al., 1997, Itoh et al., 2002, Zaprasis et al., 2010). One group contains both beta- and gamma-Proteobacteria and is a large group and another group contains species closely related to *Sphingomonas* (Kamagata et al., 1997, Zaprasis et al., 2010). Group two contains species closely related to *Bradyrhizobium* and the presence of *tfdA*-like genes are widespread throughout the genus (Zaprasis et al., 2010). The bradyrhizobial strains RD5-C2, BTH, HWK12 and HW13 were isolated from a pristine environment in Japan and were found to harbour genes encoding for the same protein and therefore their existence is independent to 2,4-D exposure suggesting herbicide degradation is not the primary function of these *tfd*-like genes (Kamagata et al., 1997, Itoh et al., 2000, Itoh et al., 2002). It is possible that the primary target is auxin itself, present in plant detritus and root exudates. Xenobiotic organic compound degradation by many groups in soil is an important factor in understanding the resilience of the microbes present and to determine how soils recover after xenobiotic compound application. A 2,4-D degrader, *Bradyrhizobium* sp. RD5-C2 has been included in the experiments included in this project and so could be used as a positive control. The arable isolates would be most exposed to herbicides however *tfd* genes have been found in *Bradyrhizobium* isolates which have never been exposed to herbicides and so herbicide degradation ability is not expected to be confined to the arable isolates. Appendix S shows the PCR protocol and gel image for the *tfdA* gene screening. All of the arable isolates, the majority of grassland, bare fallow and symbiotic isolates and fewer than half of the woodland

isolates were positive for *tfdA*. 2,4-D degradation assays would need to be carried out to determine if this gene is functional or if PCR-negative isolates are able to catabolise 2,4-D.

The genomes sequenced and assembled in this thesis indicated differences in size, plasmid presence and gene presence or absence. Forty other genomes for isolates from all soil treatments and the gorse and broom symbionts have been sequenced and to date, have not been assembled and analysed. The difference in size for G22 and BF49 was substantial and analysing the other sequenced genomes would determine whether all grassland isolates in the culture collection are always larger than the bare fallow isolates. The assembled genomes could also act as reference sequences for the genomes of free-living *Bradyrhizobium* which will be sequenced from this culture collection. It would also be interesting to see how the genome sizes of isolates from arable and woodland soil treatments compare. Analysing these genomes will also help determine whether *nirK* or *nirS* is present and will help clarify primer issues with the nitrogen cycling genes PCR. If the isolates are found to contain *nirS*, adding these to the database will help future analysis of *nirS* presence in *Bradyrhizobium*. Using the genome or gene sequences to mine the metagenomes from the Highfield plots would allow for an understanding of the *Bradyrhizobium* community as a whole. The Biolog assays could also be used to identify pathways important for distinguishing between the isolates and determining whether genes involved in this process are present across the population. The metatranscriptome could also be used to determine if these genes are expressed in the field and if they are substrates the *Bradyrhizobium* community is exposed to in their natural environments.

We were able to determine that *Bradyrhizobium* is a highly diverse group and denitrification is a process that *Bradyrhizobium* is involved in irrespective of soil treatment. Also, that the main impact of soil treatment is to growth rate and the metabolism of carbon substrates involved in core metabolism was highlighted as the biggest source of variation in the carbon metabolism assays. The approaches used in this project could also be used to explore the biology of other soil bacteria involved in key processes or rarer members of the community to better understand the soil microbial community.

9 References

Acinas, S. G., Marcelino, L. A., Klepac-Ceraj, V. & Polz, M. F. 2004. Divergence and redundancy of 16S rRNA sequences in genomes with multiple *rrn* operons. *Journal of Bacteriology,* 186**,** 2629-2635.

Adler, J. 1966. Chemotaxis in bacteria. *Science,* 153**,** 708-716.

Alden, L., Demoling, F. & Baath, E. 2001. Rapid method of determining factors limiting bacterial growth in soil. *Applied and Environmental Microbiology,* 67**,** 1830-1838.

Ali, M. B. & Mcnear, D. H. 2014. Induced transcriptional profiling of phenylpropanoid pathway genes increased flavonoid and lignin content in *Arabidopsis* leaves in response to microbial products. *BMC Plant Biology,* 14**,** 84.

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics,* 12**,** 402.

Alvarez-Anorve, L. I., Calcagno, M. L. & Plumbridge, J. 2005. Why does *Escherichia coli* grow more slowly on glucosamine than on N-acetylglucosamine? Effects of enzyme levels and allosteric activation of GlcN6P deaminase (NagB) on growth rates. *Journal of Bacteriology,* 187**,** 2974-2982.

Andrade, D. S., Murphy, P. J. & Giller, K. E. 2002. Effects of liming and legume/cereal cropping on populations of indigenous rhizobia in an acid Brazilian Oxisol. *Soil Biology & Biochemistry,* 34**,** 477-485.

Apel, A. K., Sola-Landa, A., Rodriguez-Garcia, A. & Martin, J. F. 2007. Phosphate control of *phoA, phoC* and *phoD* gene expression in *Streptomyces coelicolor* reveals significant differences in binding of PhoP to their promoter regions. *Microbiology,* 153**,** 3527-3537.

Arrighi, J. F., Cartieaux, F., Chaintreuil, C., Brown, S., Boursot, M. & Giraud, E. 2013. Genotype delimitation in the *Nod*-independent model legume *Aeschynomene evenia*. *PLOS ONE,* 8**,** e63836.

Asakawa, S. 1993. Denitrifying ability of indigenous strains of *Bradyrhizobium japonicum* isolated from fields under paddy-upland rotation. *Biology and Fertility of Soils,* 15**,** 196-200.

Attard, E., Recous, S., Chabbi, A., De Berranger, C., Guillaumaud, N., Labreuche, J., Philippot, L., Schmid, B. & Le Roux, X. 2011. Soil environmental conditions rather than denitrifier abundance and diversity drive potential denitrification after changes in land uses. *Global Change Biology,* 17**,** 1975-1989.

Aziz, R. K., Bartels, D., Best, A. A., Dejongh, M., Disz, T., Edwards, R. A., Formsma, K., Gerdes, S., Glass, E. M., Kubal, M., Meyer, F., Olsen, G. J., Olson, R., Osterman, A. L., Overbeek, R. A., Mcneil, L. K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G. D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A. & Zagnitko, O. 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics,* 9**,** 75.

Baath, E. 1996. Adaptation of soil bacterial communities to prevailing pH in different soils. *FEMS Microbiology Ecology,* 19**,** 227-237.

Bachler, C., Schneider, P., Bahler, P., Lustig, A. & Erni, B. 2005. *Escherichia coli* dihydroxyacetone kinase controls gene expression by binding to transcription factor DhaR. *EMBO Journal,* 24**,** 283-293.

Baek, J. H. & Lee, S. Y. 2006. Novel gene members in the Pho regulon of *Escherichia coli*. *FEMS Microbiology Letters,* 264**,** 104-109.

Baker, B., Zambryski, P., Staskawicz, B. & Dineshkumar, S. P. 1997. Signaling in plant-microbe interactions. *Science,* 276**,** 726-733.

Bakken, L. R., Bergaust, L., Liu, B. B. & Frostegard, A. 2012. Regulation of denitrification at the cellular level: a clue to the understanding of N_2O emissions from soils. *Philosophical Transactions of the Royal Society B,* 367**,** 1226-1234.

Bala, A., Murphy, P. & Giller, K. E. 2003. Distribution and diversity of rhizobia nodulating agroforestry legumes in soils from three continents in the tropics. *Molecular Ecology,* 12**,** 917-929.

Balasooriya, W. K., Huygens, D., Denef, K., Roobroeck, D., Verhoest, N. E. C. & Boeckx, P. 2013. Temporal variation of rhizodeposit-C assimilating microbial communities in a natural wetland. *Biology and Fertility of Soils,* 49**,** 333-341.

Bardgett, R. D., Streeter, T. C. & Bol, R. 2003. Soil microbes compete effectively with plants for organic-nitrogen inputs to temperate grasslands. *Ecology,* 84**,** 1277-1287.

Bardin, S. D. & Finan, T. M. 1998. Regulation of phosphate assimilation in *Rhizobium* (*Sinorhizobium*) *meliloti*. *Genetics,* 148**,** 1689-1700.

Barre, P., Eglin, T., Christensen, B. T., Ciais, P., Houot, S., Katterer, T., Van Oort, F., Peylin, P., Poulton, P. R., Romanenkov, V. & Chenu, C. 2010. Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments. *Biogeosciences,* 7**,** 3839-3850.

Bastida, F., Hernández, T., Albaladejo, J. & García, C. 2013. Phylogenetic and functional changes in the microbial community of long-term restored soils under semiarid climate. *Soil Biology and Biochemistry,* 65**,** 12-21.

Batista, J. S. D. S. & Hungria, M. 2012. Proteomics reveals differential expression of proteins related to a variety of metabolic pathways by genistein-induced *Bradyrhizobium japonicum* strains. *Journal of Proteomics,* 75**,** 1211-1219.

Batista, L., Tomasco, I., Lorite, M. J., Sanjuán, J. & Monza, J. 2013. Diversity and phylogeny of rhizobial strains isolated from *Lotus uliginosus* grown in Uruguayan soils. *Applied Soil Ecology,* 66**,** 19-28.

Bedmar, E. J., Robles, E. F. & Delgado, M. J. 2005. The complete denitrification pathway of the symbiotic, nitrogen-fixing bacterium *Bradyrhizobium japonicum*. *Biochemical Society Transactions,* 33**,** 141-144.

Bejarano, A., Ramirez-Bahena, M.-H., Velazquez, E. & Peix, A. 2014. *Vigna unguiculata* is nodulated in Spain by endosymbionts of Genisteae legumes and by a new symbiovar (vignae) of the genus *Bradyrhizobium*. *Systematic and Applied Microbiology,* 37**,** 533-540.

Bell, C. W., Asao, S., Calderon, F., Wolk, B. & Wallenstein, M. D. 2015. Plant nitrogen uptake drives rhizosphere bacterial community assembly during plant growth. *Soil Biology & Biochemistry,* 85**,** 170-182.

Bergaust, L., Bakken, L. R. & Frostegard, A. 2011. Denitrification regulatory phenotype, a new term for the characterization of denitrifying bacteria. *Biochemical Society Transactions,* 39**,** 207-212.

Berman, T. & Magasani, B. 1966. Pathway of myo-inositol degradation in *Aerobacter aerogenes* - dehydrogenation and dehydration. *Journal of Biological Chemistry,* 241**,** 800-806.

Bird, J. A., Herman, D. J. & Firestone, M. K. 2011. Rhizosphere priming of soil organic matter by bacterial groups in a grassland soil. *Soil Biology & Biochemistry,* 43**,** 718-725.

Black, L. K., Fu, C. L. & Maier, R. J. 1994. Sequences and characterization of *hupU* and *hupV* genes of *Bradyrhizobium japonicum* encoding a possible nickel-sensing complex involved in hydrogenase expression. *Journal of Bacteriology,* 176**,** 7102-7106.

Bobik, C., Meilhoc, E. & Batut, J. 2006. FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *Journal of Bacteriology,* 188**,** 4890-4902.

Bradshaw, R. A., Cancedda, F., Ericsson, L. H., Neumann, P. A., Piccoli, S. P., Schlesinger, M. J., Shriefer, K. & Walsh, K. A. 1981. Amino acid sequence of *Escherichia coli* alkaline phosphatase. *Proceedings of the National Academy of Sciences of the United States of America,* 78**,** 3473-3477.

Braker, G. & Tiedje, J. M. 2003. Nitric oxide reductase (*norB*) genes from pure cultures and environmental samples. *Applied and Environmental Microbiology,* 69**,** 3476-3483.

Breitenbeck, G. A. & Bremner, J. M. 1989. Ability of free-living cells of *Bradyrhizobium japonicum* to denitrify in soils. *Biology and Fertility of Soils,* 7**,** 219-224.

Bren, A. & Eisenbach, M. 2000. How signals are heard during bacterial chemotaxis: Protein-protein interactions in sensory signal propagation. *Journal of Bacteriology,* 182**,** 6865-6873.

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

Brown, S. D., Utturkar, S. M., Klingeman, D. M., Johnson, C. M., Martin, S. L., Land, M. L., Lu, T.-Y. S., Schadt, C. W., Doktycz, M. J. & Pelletier, D. A. 2012. Twenty-one genome sequences from *Pseudomonas* species and 19 genome sequences from diverse bacteria isolated from the rhizosphere and endosphere of *Populus deltoides*. *Journal of Bacteriology,* 194**,** 5991-5993.

Buchanan, R. E., Stjohnbrooks, R. & Breed, R. S. 1948. International bacteriological code of nomenclature. *Journal of Bacteriology,* 55**,** 287-306.

Buckley, D. H., Huangyutitham, V., Hsu, S.-F. & Nelson, T. A. 2007. Stable isotope probing with $^{15}N_2$ reveals novel noncultivated diazotrophs in soil. *Applied and Environmental Microbiology,* 73**,** 3196-3204.

Busset, N., De Felice, A., Chaintreuil, C., Gully, D., Fardoux, J., Romdhane, S., Molinaro, A., Silipo, A. & Giraud, E. 2016. The LPS O-antigen in photosynthetic *Bradyrhizobium* strains is dispensable for the establishment of a successful symbiosis with *Aeschynomene* legumes. *PLOS ONE,* 11**,** 14.

Campbell, C. D., Grayston, S. J. & Hirst, D. J. 1997. Use of rhizosphere carbon sources in sole carbon source tests to discriminate soil microbial communities. *Journal of Microbiological Methods,* 30**,** 33-41.

Caporaso, J. G., Lauber, C. L., Walters, W. A., Berg-Lyons, D., Lozupone, C. A., Turnbaugh, P. J., Fierer, N. & Knight, R. 2011. Global patterns of 16S rRNA diversity at a depth of millions of sequences per sample. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 4516-4522.

Cariss, S. J. L., Constantinidou, C., Patel, M. D., Takebayashi, Y., Hobman, J. L., Penn, C. W. & Avison, M. B. 2010. YieJ (CbrC) mediates CreBC-dependent Colicin E2 tolerance in *Escherichia coli*. *Journal of Bacteriology,* 192**,** 3329-3336.

Castro, H., Barrico, L., Rodriguez-Echeverria, S. & Freitas, H. 2016. Trends in plant and soil microbial diversity associated with Mediterranean extensive cereal-fallow rotation agro-ecosystems. *Agriculture Ecosystems & Environment,* 217**,** 33-40.

Cesarz, S., Fender, A. C., Beyer, F., Valtanen, K., Pfeiffer, B., Gansert, D., Hertel, D., Polle, A., Daniel, R., Leuschner, C. & Scheu, S. 2013. Roots from beech (*Fagus sylvatica* L.) and ash (*Fraxinus excelsior* L.) differentially affect soil microorganisms and carbon dynamics. *Soil Biology & Biochemistry,* 61**,** 23-32.

Chaparro, J. M., Badri, D. V. & Vivanco, J. M. 2014. Rhizosphere microbiome assemblage is affected by plant development. *ISME Journal,* 8**,** 790-803.

Chauhan, N. S., Ranjan, R., Purohit, H. J., Kalia, V. C. & Sharma, R. 2009. Identification of genes conferring arsenic resistance to *Escherichia coli* from an effluent treatment plant sludge metagenomic library. *FEMS Microbiology Ecology,* 67**,** 130-139.

Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L. & Walsh, C. T. 1990. Molecular biology of carbon-phosphorus bond cleavage - cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *Journal of Biological Chemistry,* 265**,** 4461-4471.

Chen, F., Tan, M., Yang, Y. J., Ma, J., Zhang, S. L. & Li, G. 2015. The diversity changes of soil microbial communities stimulated by climate, soil type and vegetation type analyzed via a functional gene array. *World Journal of Microbiology & Biotechnology,* 31**,** 1755-1763.

Clark, I. M., Buchkina, N., Jhurreea, D., Goulding, K. W. T. & Hirsch, P. R. 2012. Impacts of nitrogen application rates on the activity and diversity of denitrifying bacteria in the Broadbalk Wheat Experiment. *Philosophical Transactions of the Royal Society B,* 367**,** 1235-1244.

Cochran, W. G. 1950. Estimation of bacterial densities by means of the most probable number. *Biometrics,* 6**,** 105-116.

Cole, J. R., Wang, Q., Fish, J. A., Chai, B., Mcgarrell, D. M., Sun, Y., Brown, C. T., Porras-Alfaro, A., Kuske, C. R. & Tiedje, J. M. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. *Nucleic Acids Research,* 42**,** D633-D642.

Cole, M. A. & Elkan, G. H. 1973. Transmissible resistance to penicillin-G, neomycin and chloramphenicol in *Rhizobium japonicum*. *Antimicrobial Agents and Chemotherapy,* 4**,** 248-253.

Compant, S., Clement, C. & Sessitsch, A. 2010. Plant growth-promoting bacteria in the rhizo- and endosphere of plants: their role, colonization, mechanisms involved and prospects for utilization. *Soil Biology & Biochemistry,* 42**,** 669-678.

Coyne, M. S., Arunakumari, A., Averill, B. A. & Tiedje, J. M. 1989. Immunological identification and distribution of dissimilatory heme cd_1 and nonheme copper nitrite reductases in denitrifying bacteria. *Applied and Environmental Microbiology,* 55**,** 2924-2931.

Cullimore, J. V., Ranjeva, R. & Bono, J. J. 2001. Perception of lipo-chitooligosaccharidic Nod factors in legumes. *Trends in Plant Science,* 6**,** 24-30.

Czechowska, K. & Van Der Meer, J. R. 2012. Reversible and irreversible pollutant-induced bacterial cellular stress effects measured by ethidium bromide uptake and efflux. *Environmental Science & Technology,* 46**,** 1201-1208.

Da Rocha, U. N., Van Overbeek, L. & Van Elsas, J. D. 2009. Exploration of hitherto-uncultured bacteria from the rhizosphere. *FEMS Microbiology Ecology,* 69**,** 313-328.

Davis, K. E. R., Sangwan, P. & Janssen, P. H. 2011. Acidobacteria, Rubrobacteridae and Chloroflexi are abundant among very slow-growing and mini-colony-forming soil bacteria. *Environmental Microbiology,* 13**,** 798-805.

De Bruijn, F. J. 1992. Use of repetitive (repetitive extragenic palindromic and enterobacterial repetitive intergeneric consensus) sequences and the polymerase chain reaction to fingerprint the genomes of *Rhizobium meliloti* isolates and other soil bacteria. *Applied and Environmental Microbiology,* 58**,** 2180-2187.

De Campos, S. B., Deakin, W. J., Broughton, W. J. & Passaglia, L. M. P. 2011. Roles of flavonoids and the transcriptional regulator Ttsl in the activation of the type III secretion system of *Bradyrhizobium elkanii* SEMIA587. *Microbiology,* 157**,** 627-635.

De Carvalho, B., Batista, J. S., Marcelino-Guimarães, F. C., Do Nascimento, L. C. & Hungria, M. 2013. Transcriptional analysis of genes involved in nodulation in soybean roots inoculated with *Bradyrhizobium japonicum* strain CPAC 15. *BMC Genomics,* 14**,** 153.

Dean-Ross, D. & Mills, A. L. 1989. Bacterial community structure and function along a heavy metal gradient. *Applied and Environmental Microbiology,* 55**,** 2002-2009.

Deangelis, K. M., Lindow, S. E. & Firestone, M. K. 2008. Bacterial quorum sensing and nitrogen cycling in rhizosphere soil. *FEMS Microbiology Ecology,* 66**,** 197-207.

Debelle, F., Moulin, L., Mangin, B., Denarie, J. & Boivin, C. 2001. *Nod* genes and Nod signals and the evolution of the rhizobium legume symbiosis. *Acta Biochimica Polonica,* 48**,** 359-365.

Del Papa, M. F., Balague, L. J., Sowinski, S. C., Wegener, C., Segundo, E., Abarca, F. M., Toro, N., Niehaus, K., Puhler, A., Aguilar, O. M., Martinez-Drets, G. & Lagares, A. 1999. Isolation and characterization of alfalfa-nodulating rhizobia present in acidic soils of Central Argentina and Uruguay. *Applied and Environmental Microbiology,* 65**,** 1420-1427.

Delgado, M. J., Bonnard, N., Tresierra-Ayala, A., Bedmar, E. J. & Muller, P. 2003. The *Bradyrhizobium japonicum napEDABC* genes encoding the periplasmic nitrate reductase are essential for nitrate respiration. *Microbiology,* 149**,** 3395-3403.

Delmont, T. O., Prestat, E., Keegan, K. P., Faubladier, M., Robe, P., Clark, I. M., Pelletier, E., Hirsch, P. R., Meyer, F., Gilbert, J. A., Le Paslier, D., Simonet, P. & Vogel, T. M. 2012. Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME Journal,* 6**,** 1677-1687.

Delmotte, N., Ahrens, C. H., Knief, C., Qeli, E., Koch, M., Fischer, H.-M., Vorholt, J. A., Hennecke, H. & Pessi, G. 2010. An integrated proteomics and transcriptomics reference data set provides new insights into the *Bradyrhizobium japonicum* bacteroid metabolism in soybean root nodules. *Proteomics,* 10**,** 1391-1400.

Delmotte, N., Mondy, S., Alunni, B., Fardoux, J., Chaintreuil, C., Vorholt, J. A., Giraud, E. & Gourion, B. 2014. A proteomic approach of *Bradyrhizobium*/*Aeschynomene* root and stem symbioses reveals the importance of the *fixA* locus for symbiosis. *International Journal of Molecular Sciences,* 15**,** 3660-3670.

Demoling, F., Figueroa, D. & Baath, E. 2007. Comparison of factors limiting bacterial growth in different soils. *Soil Biology & Biochemistry,* 39**,** 2485-2495.

Deschaseaux, A. & Ponge, J. F. 2001. Changes in the composition of humus profiles near the trunk base of an oak tree (*Quercus petraea* (Mattus.) Liebl.). *European Journal of Soil Biology,* 37**,** 9-16.

Dimroth, P. & Hilbi, H. 1997. Enzymic and genetic basis for bacterial growth on malonate. *Molecular Microbiology,* 25**,** 3-10.

Doi, Y. & Paterson, D. L. 2007. Detection of plasmid-mediated class C beta-lactamases. *International Journal of Infectious Diseases,* 11**,** 191-197.

Dudman, W. F. & Belbin, L. 1988. Numerical taxonomic analysis of some strains of *Rhizobium* spp that uses a qualitative coding of immunodiffusion reactions. *Applied and Environmental Microbiology,* 54**,** 1825-1830.

Dupuy, N., Willems, A., Pot, B., Dewettinck, D., Vandenbruaene, I., Maestrojuan, G., Dreyfus, B., Kersters, K., Collins, M. D. & Gillis, M. 1994. Phenotypic and genotypic characterization of bradyrhizobia nodulating the leguminous tree *Acacia albida*. *International Journal of Systematic Bacteriology,* 44**,** 461-473.

Dupuy, N. C. & Dreyfus, B. L. 1992. *Bradyrhizobium* populations occur in deep soil under the leguminous tree *Acacia albida*. *Applied and Environmental Microbiology,* 58**,** 2415-2419.

Edgar, R. C. 2004. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics,* 5**,** 1-19.

Edwards, U., Rogall, T., Blocker, H., Emde, M. & Bottger, E. C. 1989. Isolation and direct complete nucleotide determination of entire genes - Characterization of a gene coding for 16S ribosomal RNA. *Nucleic Acids Research,* 17**,** 7843-7853.

Ekblad, A. & Nordgren, A. 2002. Is growth of soil microorganisms in boreal forests limited by carbon or nitrogen availability? *Plant and Soil,* 242**,** 115-122.

Epidemiology, C. F. G.2011. Available: http://www.genomicepidemiology.org/ [Accessed 11th November 2015 2015].

Fabre, S., Gully, D., Poitout, A., Patrel, D., Arrighi, J. F., Giraud, E., Czernic, P. & Cartieaux, F. 2015. Nod factor-independent nodulation in *Aeschynomene evenia* required the common plant-microbe symbiotic toolkit. *Plant Physiology,* 169**,** 2654-2664.

Falk, S., Liu, B. & Braker, G. 2010. Isolation, genetic and functional characterization of novel soil *nirK*-type denitrifiers. *Systematic and Applied Microbiology,* 33**,** 337-347.

Falke, J. J., Bass, R. B., Butler, S. L., Chervitz, S. A. & Danielson, M. A. 1997. The two-component signaling pathway of bacterial chemotaxis: A molecular view of signal transduction by receptors, kinases, and adaptation enzymes. *Annual Review of Cell and Developmental Biology,* 13**,** 457-512.

Ferguson, B. J., Indrasumunar, A., Hayashi, S., Lin, M.-H., Lin, Y.-H., Reid, D. E. & Gresshoff, P. M. 2010. Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology,* 52**,** 61-76.

Ferguson, B. J. & Mathesius, U. 2014. Phytohormone regulation of legume-rhizobia interactions. *Journal of Chemical Ecology,* 40**,** 770-790.

Fernandez, L. A., Beatriz Perotti, E., Antonio Sagardoy, M. & Anahi Gomez, M. 2008. Denitrification activity of *Bradyrhizobium* sp isolated from Argentine soybean cultivated soils. *World Journal of Microbiology & Biotechnology,* 24**,** 2577-2585.

Ferreira, M. C., Andrade, D. D., Chueire, L. M. D., Takemura, S. M. & Hungria, M. 2000. Tillage method and crop rotation effects on the population sizes and diversity of bradyrhizobia nodulating soybean. *Soil Biology & Biochemistry,* 32**,** 627-637.

Ferrieres, L., Francez-Charlot, A., Gouzy, J., Rouille, S. & Kahn, D. 2004. Fix-regulated genes evolved through promoter duplication in *Sinorhizobium meliloti*. *Microbiology,* 150**,** 2335-2345.

Fierer, N., Bradford, M. A. & Jackson, R. B. 2007. Toward an ecological classification of soil bacteria. *Ecology,* 88**,** 1354-1364.

Fox, G. E., Stackebrandt, E., Hespell, R. B., Gibson, J., Maniloff, J., Dyer, T. A., Wolfe, R. S., Balch, W. E., Tanner, R. S., Magrum, L. J., Zablen, L. B., Blakemore, R., Gupta, R., Bonen, L., Lewis, B. J., Stahl, D. A., Luehrsen, K. R., Chen, K. N. & Woese, C. R. 1980. The phylogeny of prokaryotes. *Science,* 209**,** 457-463.

Franck, W. L., Chang, W.-S., Qiu, J., Sugawara, M., Sadowsky, M. J., Smith, S. A. & Stacey, G. 2008. Whole-genome transcriptional profiling of *Bradyrhizobium japonicum* during chemoautotrophic growth. *Journal of Bacteriology,* 190**,** 6697-6705.

Fraser, T., Lynch, D. H., Entz, M. H. & Dunfield, K. E. 2015. Linking alkaline phosphatase activity with bacterial *phoD* gene abundance in soil from a long-term management trial. *Geoderma,* 257**,** 115-122.

Fuhrmann, J. 1990. Symbiotic effectiveness of indigenous soybean bradyrhizobia as related to serological, morphological, rhizobitoxine and hydrogenase phenotypes. *Applied and Environmental Microbiology,* 56**,** 224-229.

Fullam, E., Pojer, F., Bergfors, T., Jones, T. A. & Cole, S. T. 2012. Structure and function of the transketolase from *Mycobacterium tuberculosis* and comparison with the human enzyme. *Open Biology,* 2**,** 110026.

Gaby, J. C. & Buckley, D. H. 2012. A comprehensive evaluation of PCR primers to amplify the *nifH* gene of nitrogenase. *PLOS ONE,* 7.

Gage, D. J. 2004. Infection and invasion of roots by symbiotic, nitrogen-fixing rhizobia during nodulation of temperate legumes. *Microbiology and Molecular Biology Reviews,* 68**,** 280-300.

Galbraith, M. P., Feng, S. F., Borneman, J., Triplett, E. W., De Bruijn, F. J. & Rossbach, S. 1998. A functional myo-inositol catabolism pathway is essential for rhizopine utilization by *Sinorhizobium meliloti*. *Microbiology,* 144**,** 2915-2924.

Garbeva, P., Van Veen, J. A. & Van Elsas, J. D. 2004. Microbial diversity in soil: selection of microbial populations by plant and soil type and implications for disease suppressiveness. *Annual Review of Phytopathology,* 42**,** 243-270.

Garland, J. L. & Mills, A. L. 1991. Classification and characterization of heterotrophic microbial communities on the basis of patterns of community-level sole-carbon-source utilization. *Applied and Environmental Microbiology,* 57**,** 2351-2359.

Garrity, G., Brenner, D. J., Krieg, N. R. & Staley, J. T. (eds.) 2005. *Bergey's Manual® of Systematic Bacteriology.* : Springer US.

Gaworzewska, E. T. & Carlile, M. J. 1982. Positive chemotaxis of *Rhizobium leguminosarum* and other bacteria towards root exudates from legumes and other plants. *Journal of General Microbiology,* 128**,** 1179-1188.

Gerard, F., Pradel, N. & Wu, L. F. 2005. Bactericidal activity of colicin V is mediated by an inner membrane protein, SdaC, of *Escherichia coli*. *Journal of Bacteriology,* 187**,** 1945-1950.

Geurts, R., Xiao, T. T. & Reinhold-Hurek, B. 2016. What does it take to evolve a nitrogen-fixing endosymbiosis? *Trends in Plant Science,* 21**,** 199-208.

Giongo, A., Ambrosini, A., Vargas, L. K., Freire, J. R. J., Bodanese-Zanettini, M. H. & Passaglia, L. M. P. 2008. Evaluation of genetic diversity of bradyrhizobia strains nodulating soybean *Glycine max* (L.) Merrill isolated from South Brazilian fields. *Applied Soil Ecology,* 38**,** 261-269.

Giovannoni, S. J., Tripp, H. J., Givan, S., Podar, M., Vergin, K. L., Baptista, D., Bibbs, L., Eads, J., Richardson, T. H., Noordewier, M., Rappe, M. S., Short, J. M., Carrington, J. C. & Mathur, E. J. 2005. Genome streamlining in a cosmopolitan oceanic bacterium. *Science,* 309**,** 1242-1245.

Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J. C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., Bena, G., Hannibal, L., Fardoux, J., Kojadinovic, M., Vuillet, L., Lajus, A., Cruveiller, S., Rouy, Z., Mangenot, S., Segurens, B., Dossat, C., Franck, W. L., Chang, W. S., Saunders, E., Bruce, D., Richardson, P., Normand, P., Dreyfus, B., Pignol, D., Stacey, G., Emerich, D., Vermeglio, A., Medigue, C. & Sadowsky, M. 2007. Legumes symbioses: Absence of *Nod* genes in photosynthetic bradyrhizobia. *Science,* 316**,** 1307-1312.

Giraud, E., Xu, L., Chaintreuil, C., Gargani, D., Gully, D. & Sadowsky, M. J. 2013. Photosynthetic *Bradyrhizobium* sp strain ORS285 is capable of forming nitrogen-fixing root nodules on soybeans *(Glycine max)*. *Applied and Environmental Microbiology,* 79**,** 2459-2462.

Golchin, A., Oades, J. M., Skjemstad, J. O. & Clarke, P. 1994. Soil structure and carbon cycling. *Australian Journal of Soil Research,* 32**,** 1043-1068.

Gonzalez, V., Santamaria, R. I., Bustos, P., Hernandez-Gonzalez, I., Medrano-Soto, A., Moreno-Hagelsieb, G., Janga, S. C., Ramirez, M. A., Jimenez-Jacinto, V., Collado-Vides, J. & Davila, G. 2006. The partitioned *Rhizobium etli* genome: Genetic and metabolic redundancy in seven interacting replicons. *Proceedings of the National Academy of Sciences of the United States of America,* 103**,** 3834-3839.

Gorlach, K., Shingaki, R., Morisaki, H. & Hattori, T. 1994. Construction of eco-collection of paddy field soil bacteria for population analysis. *Journal of General and Applied Microbiology,* 40**,** 509-517.

Gottel, N. R., Castro, H. F., Kerley, M., Yang, Z. M., Pelletier, D. A., Podar, M., Karpinets, T., Uberbacher, E., Tuskan, G. A., Vilgalys, R., Doktycz, M. J. & Schadt, C. W. 2011. Distinct microbial communities within the endosphere and rhizosphere of *Populus deltoides* roots across contrasting soil types. *Applied and Environmental Microbiology,* 77**,** 5934-5944.

Gourion, B., Delmotte, N., Bonaldi, K., Nouwen, N., Vorholt, J. A. & Giraud, E. 2011. Bacterial RuBisCO is required for efficient *Bradyrhizobium*/*Aeschynomene* symbiosis. *PLOS ONE,* 6**,** e21900.

Gourion, B., Sulser, S., Frunzke, J., Francez-Charlot, A., Stiefel, P., Pessi, G., Vorholt, J. A. & Fischer, H.-M. 2009. The PhyR-sigma(EcfG) signalling cascade is involved in stress response and symbiotic efficiency in *Bradyrhizobium japonicum*. *Molecular Microbiology,* 73**,** 291-305.

Graf, D. R. H., Jones, C. M. & Hallin, S. 2014. Intergenomic comparisons highlight modularity of the denitrification pathway and underpin the importance of community structure for N_2O emissions. *PLOS ONE,* 9**,** e114118.

Graham-Weiss, L., Bennett, M. L. & Paau, A. S. 1987. Production of bacterial inoculants by direct fermentation on nutrient-supplemented vermiculite. *Applied and Environmental Microbiology,* 53**,** 2138-2141.

Graham, P. H. 1969. Selective medium for growth of Rhizobium. *Applied Microbiology,* 17**,** 769-770.

Graneli, W., Bertilsson, S. & Philibert, A. 2004. Phosphorus limitation of bacterial growth in high Arctic lakes and ponds. *Aquatic Sciences,* 66**,** 430-439.

Gregor, J. & Klug, G. 1999. Regulation of bacterial photosynthesis genes by oxygen and light. *FEMS Microbiology Letters,* 179**,** 1-9.

Gregory, A. S., Ritz, K., Mcgrath, S. P., Quinton, J. N., Goulding, K. W. T., Jones, R. J. A., Harris, J. A., Bol, R., Wallace, P., Pilgrim, E. S. & Whitmore, A. P. 2015. A review of the impacts of degradation threats on soil properties in the UK. *Soil Use and Management,* 31**,** 1-15.

Grosse, C., Grass, G., Anton, A., Franke, S., Santos, A. N., Lawley, B., Brown, N. L. & Nies, D. H. 1999. Transcriptional organization of the *czc* heavy-metal homeostasis determinant from *Alcaligenes eutrophus*. *Journal of Bacteriology,* 181**,** 2385-2393.

Guckert, J. B., Carr, G. J., Johnson, T. D., Hamm, B. G., Davidson, D. H. & Kumagai, Y. 1996. Community analysis by Biolog: Curve integration for statistical analysis of activated sludge microbial habitats. *Journal of Microbiological Methods,* 27**,** 183-197.

Guenet, B., Juarez, S., Bardoux, G., Pouteau, V., Cheviron, N., Marrauld, C., Abbadie, L. & Chenu, C. 2011. Metabolic capacities of microorganisms from a long-term bare fallow. *Applied Soil Ecology,* 51**,** 87-93.

Guimaraes, A. A., Jaramillo, P. M. D., Nobrega, R. S. A., Florentino, L. A., Silva, K. B. & Moreira, F. M. D. 2012. Genetic and Symbiotic diversity of nitrogen-fixing bacteria isolated from agricultural soils in the Western Amazon by using cowpea as the trap plant. *Applied and Environmental Microbiology,* 78**,** 6726-6733.

Gupta, R., Bisaria, V. S. & Sharma, S. 2015. Effect of agricultural amendments on *Cajanus cajan* (Pigeon Pea) and its rhizospheric microbial communities - A comparison between chemical fertilizers and bioinoculants. *PLOS ONE,* 10**,** e0132770.

Gutknecht, R., Beutler, R., Garcia-Alles, L. F., Baumann, U. & Erni, B. 2001. The dihydroxyacetone kinase of *Escherichia coli* utilizes a phosphoprotein instead of ATP as phosphoryl donor. *EMBO Journal,* 20**,** 2480-2486.

Hall, B. G., Acar, H., Nandipati, A. & Barlow, M. 2014. Growth rates made easy. *Molecular Biology and Evolution,* 31**,** 232-238.

Halverson, L. J. & Stacey, G. 1986. Signal exchange in plant-microbe interactions. *Microbiological Reviews,* 50**,** 193-225.

Hamilton, E. W. & Frank, D. A. 2001. Can plants stimulate soil microbes and their own nutrient supply? Evidence from a grazing tolerant grass. *Ecology,* 82**,** 2397-2402.

Harris, R. S. 2007. *Improved pairwise alignment of genomic DNA.* PhD, The Pennsylvania State University.

Harrison, K. A., Bol, R. & Bardgett, R. D. 2007. Preferences for different nitrogen forms by coexisting plant species and soil microbes. *Ecology,* 88**,** 989-999.

Hartmann, M., Howes, C. G., VanInsberghe, D., Yu, H., Bachar, D., Christen, R., Nilsson, R. H., Hallam, S. J. & Mohn, W. W. 2012. Significant and persistent impact of timber harvesting on soil microbial communities in Northern coniferous forests. *ISME Journal,* 6**,** 2199-2218.

Hastings, J., De Matos, P., Dekker, A., Ennis, M., Harsha, B., Kale, N., Muthukrishnan, V., Owen, G., Turner, S., Williams, M. & Steinbeck, C. 2013. The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013. *Nucleic Acids Research,* 41**,** D456-D463.

Hattori, R. & Hattori, T. 1980. Sensitivity to salts and organic compounds of soil bacteria isolated on diluted media. *Journal of General and Applied Microbiology,* 26**,** 1-14.

Henry, S., Baudoin, E., Lopez-Gutierrez, J. C., Martin-Laurent, F., Brauman, A. & Philippot, L. 2004. Quantification of denitrifying bacteria in soils by *nirK* gene targeted real-time PCR. *Journal of Microbiological Methods,* 59**,** 327-335.

Henry, S., Bru, D., Stres, B., Hallet, S. & Philippot, L. 2006. Quantitative detection of the *nosZ* gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. *Applied and Environmental Microbiology,* 72**,** 5181-5189.

Hernandez, D., Francois, P., Farinelli, L., Osteras, M. & Schrenzel, J. 2008. *De novo* bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. *Genome Research,* 18**,** 802-809.

Hirsch, A. M., Lum, M. R. & Downie, J. A. 2001. What makes the rhizobia-legume symbiosis so special? *Plant Physiology,* 127**,** 1484-1492.
Hirsch, P. R. 1996. Population dynamics of indigenous and genetically modified rhizobia in the field. *New Phytologist,* 133**,** 159-171.
Hirsch, P. R., Gilliam, L. M., Sohi, S. P., Williams, J. K., Clark, I. M. & Murray, P. J. 2009. Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. *Soil Biology & Biochemistry,* 41**,** 2021-2024.
Hirsch, P. R., Mauchline, T. H. & Clark, I. M. 2010. Culture-independent molecular techniques for soil microbial ecology. *Soil Biology & Biochemistry,* 42**,** 878-887.
Hirsch, P. R., Vanmontagu, M., Johnston, A. W. B., Brewin, N. J. & Schell, J. 1980. Physical identification of bacteriocinogenic, nodulation and other plasmids in strains of *Rhizobium leguminosarum*. *Journal of General Microbiology,* 120**,** 403-412.
Hogberg, P. & Read, D. J. 2006. Towards a more plant physiological perspective on soil ecology. *Trends in Ecology & Evolution,* 21**,** 548-554.
Hollowell, A. C., Regus, J. U., Gano, K. A., Bantay, R., Centeno, D., Pham, J., Lyu, J., Moore, D., Bernardo, A., Lopez, G., Patil, A., Patel, S., Lii, Y. & Sachs, J. L. 2016a. Epidemic spread of symbiotic and non-symbiotic *Bradyrhizobium* genotypes across california. *Microbial Ecology,* 71**,** 700-710.
Hollowell, A. C., Regus, J. U., Turissini, D., Gano-Cohen, K. A., Bantay, R., Bernardo, A., Moore, D., Pham, J. & Sachs, J. L. 2016b. Metapopulation dominance and genomic-island acquisition of *Bradyrhizobium* with superior catabolic capabilities. *Proceedings of the Royal Society B,* 283.
Horecker, B. L., Smyrniotis, P. Z. & Hurwitz, J. 1956. Role of xylulose 5-phosphate in the transketolase reaction. *Journal of Biological Chemistry,* 223**,** 1009-1019.
Howieson, J. G., De Meyer, S. E., Vivas-Marfisi, A., Ratnayake, S., Ardley, J. K. & Yates, R. J. 2013. Novel *Burkholderia* bacteria isolated from *Lebeckia ambigua* - A perennial suffrutescent legume of the fynbos. *Soil Biology & Biochemistry,* 60**,** 55-64.
Hulton, C. S. J., Higgins, C. F. & Sharp, P. M. 1991. ERIC sequences - a novel family of repetitive elements in the genomes of *Escherichia coli, Salmonella typhimurium* and other enterobacteria. *Molecular Microbiology,* 5**,** 825-834.
Igarashi, N., Harada, J., Nagashima, S., Matsuura, K., Shimada, K. & Nagashima, K. V. P. 2001. Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *Journal of Molecular Evolution,* 52**,** 333-341.
Ilstedt, U. & Singh, S. 2005. Nitrogen and phosphorus limitations of microbial respiration in a tropical phosphorus-fixing acrisol (ultisol) compared with organic compost. *Soil Biology & Biochemistry,* 37**,** 1407-1410.
Ishii, S., Ashida, N., Otsuka, S. & Senoo, K. 2011. Isolation of oligotrophic denitrifiers carrying previously uncharacterized functional gene sequences. *Applied and Environmental Microbiology,* 77**,** 338-342.
Itoh, K., Kanda, R., Momoda, Y., Sumita, Y., Kamagata, Y., Suyama, K. & Yamamoto, H. 2000. Presence of 2,4-D-catabolising bacteria in a Japanese arable soil that belong to BANA (*Bradyrhizobium-Agromonas-Nitrobacter-Afipia*) cluster in alpha-proteobacteria. *Microbes and Environments,* 15**,** 113-117.
Itoh, K., Kanda, R., Sumita, Y., Kim, H., Kamagata, Y., Suyama, K., Yamamoto, H., Hausinger, R. P. & Tiedje, J. M. 2002. *tfdA*-like genes in 2,4-dichlorophenoxyacetic acid-degrading bacteria belonging to the *Bradyrhizobium-Agromonas-Nitrobacter-Afipia* cluster in alpha-proteobacteria. *Applied and Environmental Microbiology,* 68**,** 3449-3454.
Janssen, P. H., Yates, P. S., Grinton, B. E., Taylor, P. M. & Sait, M. 2002. Improved culturability of soil bacteria and isolation in pure culture of novel members of the divisions *Acidobacteria, Actinobacteria, Proteobacteria,* and *Verrucomicrobia*. *Applied and Environmental Microbiology,* 68**,** 2391-2396.

Jarvis, B. D. W., Pankhurst, C. E. & Patel, J. J. 1982. *Rhizobium loti*, a new species of legume root nodule bacteria. *International Journal of Systematic Bacteriology,* 32**,** 378-380.

Jhala, Y. K., Vyas, R. V., Shelat, H. N., Patel, H. K., Patel, H. K. & Patel, K. T. 2014. Isolation and characterization of methane utilizing bacteria from wetland paddy ecosystem. *World Journal of Microbiology & Biotechnology,* 30**,** 1845-1860.

Joergensen, R. G. & Scheu, S. 1999. Response of soil microorganisms to the addition of carbon, nitrogen and phosphorus in a forest Rendzina. *Soil Biology & Biochemistry,* 31**,** 859-866.

Jones, F. P., Clark, I. M., King, R., Shaw, L. J., Woodward, M. J. & Hirsch, P. R. 2016. Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes - a genome comparison. *Scientific Reports,* 6**,** 25858.

Jordan, D. C. 1982. Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen-nov, a genus of slow-growing, root nodule bacteria from leguminous plants. *International Journal of Systematic Bacteriology,* 32**,** 136-139.

Joseph, S. J., Hugenholtz, P., Sangwan, P., Osborne, C. A. & Janssen, P. H. 2003. Laboratory cultivation of widespread and previously uncultured soil bacteria. *Applied and Environmental Microbiology,* 69**,** 7210-7215.

Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. 2014. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLOS ONE,* 9**,** e104984.

Kaeberlein, T., Lewis, K. & Epstein, S. S. 2002. Isolating "uncultivable" microorganisms in pure culture in a simulated natural environment. *Science,* 296**,** 1127-1129.

Kamagata, Y., Fulthorpe, R. R., Tamura, K., Takami, H., Forney, L. J. & Tiedje, J. M. 1997. Pristine environments harbor a new group of oligotrophic 2,4-dichlorophenoxyacetic acid-degrading bacteria. *Applied and Environmental Microbiology,* 63**,** 2266-2272.

Kanehisa, M. 2000. *Post-genome Informatics* Oxford University Press.

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

Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. 2016. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Research,* 44**,** D457-D462.

Kaneko, T., Maita, H., Hirakawa, H., Uchiike, N., Minamisawa, K., Watanabe, A. & Sato, S. 2011. Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6^{T}. *Genes,* 2**,** 763-787.

Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., Kohara, M., Matsumoto, M., Shimpo, S., Tsuruoka, H., Wada, T., Yamada, M. & Tabata, S. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Research,* 9**,** 189-197.

Kardol, P., Cornips, N. J., Van Kempen, M. M. L., Bakx-Schotman, J. M. T. & Van Der Putten, W. H. 2007. Microbe-mediated plant-soil feedback causes historical contingency effects in plant community assembly. *Ecological Monographs,* 77**,** 147-162.

Kawaharada, Y., Kiyota, H., Eda, S., Minamisawa, K. & Mitsui, H. 2010. Identification of the *Mesorhizobium loti* gene responsible for glycerophosphorylation of periplasmic cyclic beta-1,2-glucans. *FEMS Microbiology Letters,* 302**,** 131-137.

Kelly, J. J., Peterson, E., Winkelman, J., Walter, T. J., Rier, S. T. & Tuchman, N. C. 2013. Elevated atmospheric CO_2 impacts abundance and diversity of nitrogen cycling functional genes in soil. *Microbial Ecology,* 65**,** 394-404.

Khalil, K., Mary, B. & Renault, P. 2004. Nitrous oxide production by nitrification and denitrification in soil aggregates as affected by O_2 concentration. *Soil Biology & Biochemistry,* 36**,** 687-699.

Khamis, A., Colson, P., Raoult, D. & La Scola, B. 2003. Usefulness of *rpoB* gene sequencing for identification of *Afipia* and *Bosea* species, including a strategy for choosing discriminative partial sequences. *Applied and Environmental Microbiology,* 69**,** 6740-6749.

Khatoon, A., Rehman, S., Salavati, A. & Komatsu, S. 2012. A comparative proteomics analysis in roots of soybean to compatible symbiotic bacteria under flooding stress. *Amino Acids,* 43**,** 2513-2525.

Kim, Y. S. 2002. Malonate metabolism: Biochemistry, molecular biology, physiology, and industrial application. *Journal of Biochemistry and Molecular Biology,* 35**,** 443-451.

Kim, Y. S., Kwon, S. J. & Kang, S. W. 1993. Malonyl-CoA synthetase from *Rhizobium trifolii:* Purification, properties, and the immunological comparison with those from *Bradyrhizobium japonicum* and *Pseudomonas fluorescens*. *Korean Biochemical Journal,* 26**,** 176-183.

Kimbrel, J. A., Thomas, W. J., Jiang, Y., Creason, A. L., Thireault, C. A., Sachs, J. L. & Chang, J. H. 2013. Mutualistic Co-evolution of Type III Effector Genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLOS pathogens,* 9**,** e1003204-e1003204.

Kinkle, B. K., Angle, J. S. & Keyser, H. H. 1987. Long-term effects of metal-rich sewage sludge application on soil populations of *Bradyrhizobium japonicum*. *Applied and Environmental Microbiology,* 53**,** 315-319.

Klappenbach, J. A., Saxman, P. R., Cole, J. R. & Schmidt, T. M. 2001. *rrn*db: the ribosomal RNA operon copy number database. *Nucleic Acids Research,* 29**,** 181-184.

Koch, A. L. 2001. Oligotrophs versus copiotrophs. *Bioessays,* 23**,** 657-661.

Kohler, P. R. A., Zheng, J. Y., Schoffers, E. & Rossbach, S. 2010. Inositol Catabolism, a Key Pathway in *Sinorhizobium meliloti* for Competitive Host Nodulation. *Applied and Environmental Microbiology,* 76**,** 7972-7980.

Koike, I. & Hattori, A. 1975. Energy yield of denitrification - estimate from growth yield in continuous cultures of *Pseudomonas denitrificans* under nitrate-limited, nitrite-limited and nitrous oxide-limited conditions. *Journal of General Microbiology,* 88**,** 11-19.

Kong, A. Y. Y., Scow, K. M., Cordova-Kreylos, A. L., Holmes, W. E. & Six, J. 2011. Microbial community composition and carbon cycling within soil microenvironments of conventional, low-input, and organic cropping systems. *Soil Biology & Biochemistry,* 43**,** 20-30.

Kowalchuk, G. A., Buma, D. S., De Boer, W., Klinkhamer, P. G. L. & Van Veen, J. A. 2002. Effects of above-ground plant species composition and diversity on the diversity of soil-borne microorganisms. *Antonie Van Leeuwenhoek,* 81**,** 509-520.

Kraft, B., Strous, M. & Tegetmeyer, H. E. 2011. Microbial nitrate respiration - Genes, enzymes and environmental distribution. *Journal of Biotechnology,* 155**,** 104-117.

Kristjansson, J. K. & Hollocher, T. C. 1980. First practical assay for soluble nitrous oxide reductase of denitrifying bacteria and a partial kinetic characterization. *Journal of Biological Chemistry,* 255**,** 704-707.

Kumar, M., Mannisto, M. K., Van Elsas, J. D. & Nissinen, R. M. 2016. Plants impact structure and function of bacterial communities in Arctic soils. *Plant and Soil,* 399**,** 319-332.

Kundig, C., Beck, C., Hennecke, H. & Goettfert, M. 1995. A single rRNA gene region in *Bradyrhizobium japonicum*. *Journal of Bacteriology,* 177**,** 5151-5154.

Kundig, W., Roseman, S. & Ghosh, S. 1964. Phosphate bound to histidine in protein as intermediate in novel phospho-transferase system. *Proceedings of the National Academy of Sciences of the United States of America,* 52**,** 1067-1074.

Kuykendall, L. D., Roy, M. A., Oneill, J. J. & Devine, T. E. 1988. Fatty-acids, antibiotic resistance and deoxyribonucleic acids homology groups of *Bradyrhizobium japonicum. International Journal of Systematic Bacteriology,* 38**,** 358-361.

Lafay, B. & Burdon, J. J. 2007. Molecular diversity of legume root-nodule bacteria in Kakadu National Park, Northern Territory, Australia. *PLOS ONE,* 2**,** e277.

Laguerre, G., Vanberkum, P., Amarger, N. & Prevost, D. 1997. Genetic diversity of rhizobial symbionts isolated from legume species within the genera *Astragalus*, *Oxytropis*, and *Onobrychis*. *Applied and Environmental Microbiology,* 63**,** 4748-4758.

Lakshmanan, V., Selvaraj, G. & Bais, H. P. 2014. Functional soil microbiome: Belowground solutions to an aboveground problem. *Plant Physiology,* 166**,** 689-700.

Lane, D. J., Pace, B., Olsen, G. J., Stahl, D. A., Sogin, M. L. & Pace, N. R. 1985. Rapid determination of 16S ribosomal-RNA sequences for phylogenetic analyses. *Proceedings of the National Academy of Sciences of the United States of America,* 82**,** 6955-6959.

Lange, R. T. 1961. Nodule bacteria associated with indigenous leguminosae of south western Australia. *Journal of General Microbiology,* 26**,** 351-359.

Lareen, A., Burton, F. & Schafer, P. 2016. Plant root-microbe communication in shaping root microbiomes. *Plant Molecular Biology,* 90**,** 575-587.

Larimer, A. L., Clay, K. & Bever, J. D. 2014. Synergism and context dependency of interactions between arbuscular mycorrhizal fungi and rhizobia with a prairie legume. *Ecology,* 95**,** 1045-1054.

Latz, E., Eisenhauer, N., Scheu, S. & Jousset, A. 2015. Plant identity drives the expression of biocontrol factors in a rhizosphere bacterium across a plant diversity gradient. *Functional Ecology,* 29**,** 1225-1234.

Lee, C. C., Ribbe, M. W. & Hu, Y. 2014. Cleaving the N,N triple bond: The transformation of dinitrogen to ammonia by nitrogenases. *In:* Kroneck, M. H. P. & Torres, S. M. E. (eds.) *The Metal-Driven Biogeochemistry of Gaseous Compounds in the Environment.* Dordrecht: Springer Netherlands.

Leekitcharoenphon P, Kaas RS, Thomsen MCF, Friis C, Rasmussen S & Aarestrup, F.M. 2012. snpTree–a web-server to identify and construct SNP trees from whole genome sequence data. *BMC Genomics,* 13, Supplement 7: S6.

Lewis, K., Epstein, S., D'onofrio, A. & Ling, L. L. 2010. Uncultured microorganisms as a source of secondary metabolites. *Journal of Antibiotics,* 63**,** 468-476.

Ley, R. E., Backhed, F., Turnbaugh, P., Lozupone, C. A., Knight, R. D. & Gordon, J. I. 2005. Obesity alters gut microbial ecology. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 11070-11075.

Li, P. L., Hwang, I., Miyagi, H., True, H. & Farrand, S. K. 1999. Essential components of the Ti plasmid trb system, a type IV macromolecular transporter. *Journal of Bacteriology,* 181**,** 5033-5041.

Libault, M., Farmer, A., Brechenmacher, L., Drnevich, J., Langley, R. J., Bilgin, D. D., Radwan, O., Neece, D. J., Clough, S. J., May, G. D. & Stacey, G. 2010. Complete transcriptome of the soybean root hair cell, a single-cell model, and its alteration in response to *Bradyrhizobium japonicum* infection. *Plant Physiology,* 152**,** 541-552.

Lima-Perim, J. E., Romagnoli, E. M., Dini-Andreote, F., Durrer, A., Dias, A. C. F. & Andreote, F. D. 2016. Linking the composition of bacterial and archaeal communities to characteristics of soil and flora composition in the Atlantic rainforest. *PLOS ONE,* 11**,** e0146566.

Lindström, K. 1989. *Rhizobium galegae*, a new species of legume root nodule bacteria. *International Journal of Systematic Bacteriology,* 39**,** 365-367.

Ling, L. L., Schneider, T., Peoples, A. J., Spoering, A. L., Engels, I., Conlon, B. P., Mueller, A., Schaberle, T. F., Hughes, D. E., Epstein, S., Jones, M., Lazarides, L., Steadman, V. A., Cohen, D. R., Felix, C. R., Fetterman, K. A., Millett, W. P., Nitti, A. G., Zullo, A. M., Chen, C. & Lewis, K. 2015. A new antibiotic kills pathogens without detectable resistance. *Nature,* 517**,** 455-459.

Liu, R. & Ochman, H. 2007. Stepwise formation of the bacterial flagellar system. *Proceedings of the National Academy of Sciences of the United States of America,* 104**,** 7116-7121.

Livermore, D. M. 1995. Beta-lactamases in laboratory and clinical resistance. *Clinical Microbiology Reviews,* 8**,** 557-584.

Lodwig, E. M., Hosie, A. H. F., Bordes, A., Findlay, K., Allaway, D., Karunakaran, R., Downie, J. A. & Poole, P. S. 2003. Amino-acid cycling drives nitrogen fixation in the legume - *Rhizobium* symbiosis. *Nature,* 422**,** 722-726.

Loman, N. J., Misra, R. V., Dallman, T. J., Constantinidou, C., Gharbia, S. E., Wain, J. & Pallen, M. J. 2012. Performance comparison of benchtop high-throughput sequencing platforms. *Nature Biotechnology,* 30**,** 434–439.

Long, A., Heitman, J., Tobias, C., Philips, R. & Song, B. 2013. Co-occurring anammox, denitrification, and codenitrification in agricultural soils. *Applied and Environmental Microbiology,* 79**,** 168-176.

Long, F., Rouquette-Loughlin, C., Shafer, W. M. & Yu, E. W. 2008. Functional cloning and characterization of the multidrug efflux pumps NorM from *Neisseria gonorrhoeae* and YdhE from *Escherichia coli*. *Antimicrobial Agents and Chemotherapy,* 52**,** 3052-3060.

Louvrier, P., Laguerre, G. & Amarger, N. 1995. Semiselective medium for isolation of *Rhizobium leguminosarum* from soils. *Soil Biology & Biochemistry,* 27**,** 919-924.

Lu, S. & Zgurskaya, H. I. 2012. Role of ATP binding and hydrolysis in assembly of MacAB-TolC macrolide transporter. *Molecular Microbiology,* 86**,** 1132-1143.

Lucas-Elio, P., Gomez, D., Solano, F. & Sanchez-Amat, A. 2006. The antimicrobial activity of marinocine, synthesized by *Marinomonas mediterranea*, is due to hydrogen peroxide generated by its lysine oxidase activity. *Journal of Bacteriology,* 188**,** 2493-2501.

Lugtenberg, B. J. J., Chin-a-Woeng, T. F. C. & Bloemberg, G. V. 2002. Microbe-plant interactions: principles and mechanisms. *Antonie Van Leeuwenhoek* 81**,** 373-383.

Luli, G. W. & Strohl, W. R. 1990. Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology,* 56**,** 1004-1011.

Luo, R., Liu, B., Xie, Y., Li, Z., Huang, W., Yuan, J., He, G., Chen, Y., Pan, Q., Liu, Y., Tang, J., Wu, G., Zhang, H., Shi, Y., Liu, Y., Yu, C., Wang, B., Lu, Y., Han, C., Cheung, D. W., Yiu, S.-M., Peng, S., Xiaoqian, Z., Liu, G., Liao, X., Li, Y., Yang, H., Wang, J., Lam, T.-W. & Wang, J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience,* 1**,** 18-18.

Lynch, M. D. J. & Neufeld, J. D. 2015. Ecology and exploration of the rare biosphere. *Nature Reviews Microbiology,* 13**,** 217-229.

Maag, M. & Vinther, F. P. 1996. Nitrous oxide emission by nitrification and denitrification in different soil types and at different soil moisture contents and temperatures. *Applied Soil Ecology,* 4**,** 5-14.

Madrzak, C. J., Golinska, B., Kroliczak, J., Pudelko, K., Lazewska, D., Lampka, B. & Sadowsky, M. J. 1995. Diversity among field populations of *Bradyrhizobium japonicum* in Poland. *Applied and Environmental Microbiology,* 61**,** 1194-1200.

Mahne, I. & Tiedje, J. M. 1995. Criteria and methodology for identifying respiratory denitrifiers. *Applied and Environmental Microbiology,* 61**,** 1110-1115.

Marcondes De Souza, J. A., Tieppo, E., Magnani, G. D. S., Carareto Alves, L. M., Cardoso, R. L., Cruz, L. M., De Oliveira, L. F., Raittz, R. T., De Souza, E. M., Pedrosa, F. D. O. & De Macedo Lemos, E. G. 2012. Draft genome sequence of the nitrogen-fixing symbiotic bacterium *Bradyrhizobium elkanii* 587. *Journal of Bacteriology,* 194**,** 3547-3548.

Marschner, P., Yang, C. H., Lieberei, R. & Crowley, D. E. 2001. Soil and plant specific effects on bacterial community composition in the rhizosphere. *Soil Biology & Biochemistry,* 33**,** 1437-1445.

Martins, L. M. V., Neves, M. C. P. & Rumjanek, N. G. 1997. Growth characteristics and symbiotic efficiency of rhizobia isolated from cowpea nodules of the north-east region of Brazil. *Soil Biology & Biochemistry,* 29**,** 1005-1010.

Mauchline, T. H., Hayat, R., Roberts, R., Powers, S. J. & Hirsch, P. R. 2014. Assessment of core and accessory genetic variation in *Rhizobium leguminosarum* symbiovar *trifolii* strains from diverse locations and host plants using PCR-based methods. *Letters in Applied Microbiology,* 59**,** 238-246.

Mccrady, M. H. 1915. The numerical interpretation of fermentation-tube results *The Journal of Infectious Diseases,* 17**,** 183-212.

Meakin, G. E., Jepson, B. J. N., Richardson, D. J., Bedmar, E. J. & Delgado, M. J. 2006. The role of *Bradyrhizobium japonicum* nitric oxide reductase in nitric oxide detoxification in soya bean root nodules. *Biochemical Society Transactions,* 34**,** 195-196.

Meilhoc, E., Cam, Y., Skapski, A. & Bruand, C. 2010. The response to nitric oxide of the nitrogen-fixing symbiont *Sinorhizobium meliloti*. *Molecular Plant-Microbe Interactions,* 23**,** 748-759.

Mellado-Vazquez, P. G., Lange, M., Bachmann, D., Gockele, A., Karlowsky, S., Milcu, A., Piel, C., Roscher, C., Roy, J. & Gleixner, G. 2016. Plant diversity generates enhanced soil microbial access to recently photosynthesized carbon in the rhizosphere. *Soil Biology & Biochemistry,* 94**,** 122-132.

Menna, P., Barcellos, F. G. & Hungria, M. 2009. Phylogeny and taxonomy of a diverse collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the 16S rRNA gene, ITS region and *glnII*, *recA*, *atpD* and *dnaK* genes. *International Journal of Systematic and Evolutionary Microbiology,* 59**,** 2934-2950.

Menna, P. & Hungria, M. 2011. Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer. *International Journal of Systematic and Evolutionary Microbiology,* 61**,** 3052-3067.

Mergel, A., Schmitz, O., Mallmann, T. & Bothe, H. 2001. Relative abundance of denitrifying and dinitrogen-fixing bacteria in layers of a forest soil. *FEMS Microbiology Ecology,* 36**,** 33-42.

Mesa, S., Velasco, L., Manzanera, M. E., Delgado, M. J. & Bedmar, E. J. 2002. Characterization of the *norCBQD* genes, encoding nitric oxide reductase, in the nitrogen fixing bacterium *Bradyrhizobium japonicum*. *Microbiology,* 148**,** 3553-3560.

Metcalf, W. W. & Wanner, B. L. 1991. Involvement of the *Escherichia coli phn* (*psiD*) gene cluster in assimilation of phosphorus in the form of phosphonates, phosphite, Pi esters, and Pi. *Journal of Bacteriology,* 173**,** 587-600.

Mikola, J., Sorensen, L. I. & Kytoviita, M. M. 2014. Plant removal disturbance and replant mitigation effects on the abundance and diversity of low-arctic soil biota. *Applied Soil Ecology,* 82**,** 82-92.

Minamisawa, K., Seki, T., Onodera, S., Kubota, M. & Asami, T. 1992. Genetic relatedness of *Bradyrhizobium japonicum* field isolates as revealed by repeated sequences and various other characteristics. *Applied and Environmental Microbiology,* 58**,** 2832-2839.

Mira, A., Ochman, H. & Moran, N. A. 2001. Deletional bias and the evolution of bacterial genomes. *Trends in Genetics,* 17**,** 589-596.

Miri, M., Janakirama, P., Held, M., Ross, L. & Szczyglowski, K. 2016. Into the root: How cytokinin controls rhizobial infection. *Trends in Plant Science,* 21**,** 178-186.

Mitsui, H., Gorlach, K., Lee, H. J., Hattori, R. & Hattori, T. 1997. Incubation time and media requirements of culturable bacteria from different phylogenetic groups. *Journal of Microbiological Methods,* 30**,** 103-110.

Mohammadi, M., Burbank, L. & Roper, M. C. 2012. Biological role of pigment production for the bacterial phytopathogen *Pantoea stewartii* subsp *stewartii*. *Applied and Environmental Microbiology,* 78**,** 6859-6865.

Mohan, S. B., Schmid, M., Jetten, M. & Cole, J. 2004. Detection and widespread distribution of the *nrfA* gene encoding nitrite reduction to ammonia, a short circuit in the biological nitrogen cycle that competes with denitrification. *FEMS Microbiology Ecology,* 49**,** 433-443.

Montana, J. S., Jimenez, D. J., Hernandez, M., Angel, T. & Baena, S. 2012. Taxonomic and functional assignment of cloned sequences from high Andean forest soil metagenome. *Antonie Van Leeuwenhoek* 101**,** 205-215.

Monza, J., Irisarri, P., Az, P. D., Delgado, M. J., Mesa, S. & Bedmar, E. J. 2006. Denitrification ability of rhizobial strains isolated from *Lotus* sp. *Antonie Van Leeuwenhoek,* 89**,** 479-484.

Morgan, J. a. W., Bending, G. D. & White, P. J. 2005. Biological costs and benefits to plant-microbe interactions in the rhizosphere. *Journal of Experimental Botany,* 56**,** 1729-1739.

Moulin, L., Bena, G., Boivin-Masson, C. & Stepkowski, T. 2004. Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus. *Molecular Phylogenetics and Evolution,* 30**,** 720-732.

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

Murrell, J. C., Gilbert, B. & Mcdonald, I. R. 2000. Molecular biology and regulation of methane monooxygenase. *Archives of Microbiology,* 173**,** 325-332.

Nacke, H., Fischer, C., Thuermer, A., Meinicke, P. & Daniel, R. 2014. Land use type significantly affects microbial gene transcription in soil. *Microbial Ecology,* 67**,** 919-930.

Narozna, D., Pudelko, K., Kroliczak, J., Golinska, B., Sugawara, M., Madrzak, C. J. & Sadowsky, M. J. 2015. Survival and competitiveness of *Bradyrhizobium japonicum* strains 20 years after introduction into field locations in Poland. *Applied and Environmental Microbiology,* 81**,** 5552-5559.

Neumann, M., Mittelstaedt, G., Iobbi-Nivol, C., Saggu, M., Lendzian, F., Hildebrandt, P. & Leimkuehler, S. 2009. A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from *Escherichia coli*. *FEBS Journal,* 276**,** 2762-2774.

Nichols, D. 2007. Cultivation gives context to the microbial ecologist. *FEMS Microbiology Ecology,* 60**,** 351-357.

Nichols, D., Cahoon, N., Trakhtenberg, E. M., Pham, L., Mehta, A., Belanger, A., Kanigan, T., Lewis, K. & Epstein, S. S. 2010. Use of ichip for high-throughput in situ cultivation of "uncultivable" microbial species. *Applied and Environmental Microbiology,* 76**,** 2445-2450.

Nies, D. H. 1992. *czcR* and *czcD*, gene products affecting regulation of resistance to cobalt, zinc, and cadmium (czc system) in *Alcaligenes eutrophus*. *Journal of Bacteriology,* 174**,** 8102-8110.

Niewiadomska, A., Sawinska, Z. & Wolna-Maruwka, A. 2012. Impact of seed dressings on microbiological activity of soil under winter triticale cultivation. *Archives of Environmental Protection,* 38**,** 89-101.

Noel, K. D. & Brill, W. J. 1980. Diversity and dynamics of indigenous *Rhizobium japonicum* populations. *Applied and Environmental Microbiology,* 40**,** 931-938.

Noh, J. G., Jeon, H. E., So, J. S. & Chang, W. S. 2015. Effects of the *Bradyrhizobium japonicum waaL* (*rfaL*) Gene on hydrophobicity, motility, stress tolerance, and symbiotic relationship with soybeans. *International Journal of Molecular Sciences,* 16**,** 16778-16791.

Noisangiam, R., Teamtisong, K., Tittabutr, P., Boonkerd, N., Toshiki, U., Minamisawa, K. & Teaumroong, N. 2012. Genetic diversity, symbiotic evolution, and proposed infection process of *Bradyrhizobium* strains isolated from root nodules of *Aeschynomene americana* L. in Thailand. *Applied and Environmental Microbiology,* 78**,** 6236-6250.

Nordin, A., Schmidt, I. K. & Shaver, G. R. 2004. Nitrogen uptake by arctic soil microbes and plants in relation to soil nitrogen supply. *Ecology,* 85**,** 955-962.

Nunan, N., Lerch, T. Z., Pouteau, V., Mora, P., Changey, E., Katterer, T., Giusti-Miller, S. & Herrmann, A. M. 2015. Metabolising old soil carbon: Simply a matter of simple organic matter? *Soil Biology & Biochemistry,* 88**,** 128-136.

Oh, E., Zhang, Q. & Jeon, B. 2014. Target optimization for peptide nucleic acid (PNA)-mediated antisense inhibition of the CmeABC multidrug efflux pump in *Campylobacter jejuni*. *Journal of Antimicrobial Chemotherapy,* 69**,** 375-380.

Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D., Arrighi, J. F., Furuta, N., Miwa, H., Yasuda, M., Nouwen, N., Teaumroong, N. & Giraud, E. 2016. *Rhizobium*-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME Journal,* 10**,** 64-74.

Okubo, T., Fukushima, S., Itakura, M., Oshima, K., Longtonglang, A., Teaumroong, N., Mitsui, H., Hattori, M., Hattori, R., Hattori, T. & Minamisawa, K. 2013. Genome analysis suggests that the soil oligotrophic bacterium *Agromonas oligotrophica* (*Bradyrhizobium oligotrophicum*) is a nitrogen-fixing symbiont of *Aeschynomene indica*. *Applied and Environmental Microbiology,* 79**,** 2542-2551.

Okubo, T., Tsukui, T., Maita, H., Okamoto, S., Oshima, K., Fujisawa, T., Saito, A., Futamata, H., Hattori, R., Shimomura, Y., Haruta, S., Morimoto, S., Wang, Y., Sakai, Y., Hattori, M., Aizawa, S.-I., Nagashima, K. V. P., Masuda, S., Hattori, T., Yamashita, A., Bao, Z., Hayatsu, M., Kajiya-Kanegae, H., Yoshinaga, I., Sakamoto, K., Toyota, K., Nakao, M., Kohara, M., Anda, M., Niwa, R., Park, J.-H., Sameshima-Saito, R., Tokuda, S.-I., Yamamoto, S., Yamamoto, S., Yokoyama, T., Akutsu, T., Nakamura, Y., Nakahira-Yanaka, Y., Hoshino, Y. T., Hirakawa, H., Mitsui, H., Terasawa, K., Itakura, M., Sato, S., Ikeda-Ohtsubo, W., Sakakura, N., Kaminuma, E. & Minamisawa, K. 2012. Complete genome sequence of *Bradyrhizobium* sp S23321: Insights into symbiosis evolution in soil oligotrophs. *Microbes and Environments,* 27**,** 306-315.

Oldroyd, G. E. D. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nature Reviews Microbiology,* 11**,** 252-263.

Op Den Camp, R. H. M., Polone, E., Fedorova, E., Roelofsen, W., Squartini, A., Op Den Camp, H. J. M., Bisseling, T. & Geurts, R. 2012. Non-legume *Parasponia andersonii* deploys a broad *Rhizobium* host range strategy resulting in largely variable symbiotic effectiveness. *Molecular Plant-Microbe Interactions,* 25**,** 954-963.

Ormeno-Orrillo, E., Servin-Garciduenas, L. E., Imperial, J., Rey, L., Ruiz-Argueso, T. & Martinez-Romero, E. 2013. Phylogenetic evidence of the transfer of *nodZ* and *nolL* genes from *Bradyrhizobium* to other rhizobia. *Molecular Phylogenetics and Evolution,* 67**,** 626-630.

Overbeek, R., Olson, R., Pusch, G. D., Olsen, G. J., Davis, J. J., Disz, T., Edwards, R. A., Gerdes, S., Parrello, B., Shukla, M., Vonstein, V., Wattam, A. R., Xia, F. F. & Stevens, R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Research,* 42**,** D206-D214.

Panke-Buisse, K., Poole, A. C., Goodrich, J. K., Ley, R. E. & Kao-Kniffin, J. 2015. Selection on soil microbiomes reveals reproducible impacts on plant function. *ISME Journal,* 9**,** 980-989.

Parker, G. F., Higgins, T. P., Hawkes, T. & Robson, R. L. 1999. *Rhizobium* (*Sinorhizobium*) *meliloti phn* genes: Characterization and identification of their protein products. *Journal of Bacteriology,* 181**,** 389-395.

Pastorelli, R., Landi, S., Trabelsi, D., Piccolo, R., Mengoni, A., Bazzicalupo, M. & Pagliai, M. 2011. Effects of soil management on structure and activity of denitrifying bacterial communities. *Applied Soil Ecology,* 49**,** 46-58.

Paynel, F., Murray, P. J. & Cliquet, J. B. 2001. Root exudates: a pathway for short-term N transfer from clover and ryegrass. *Plant and Soil,* 229**,** 235-243.

Peix, A., Helena Ramirez-Bahena, M., David Flores-Felix, J., Alonso De La Vega, P., Rivas, R., Mateos, P. F., Igual, J. M., Martinez-Molina, E., Trujillo, M. E. & Velazquez, E. 2015. Revision of the taxonomic status of the species *Rhizobium lupini* and reclassification as *Bradyrhizobium lupini* comb. nov. *International Journal of Systematic and Evolutionary Microbiology,* 65**,** 1213-1219.

Peng, C. J., Lai, S. S., Luo, X. S., Lu, J. W., Huang, Q. Y. & Chen, W. L. 2016. Effects of long term rice straw application on the microbial communities of rapeseed rhizosphere in a paddy-upland rotation system. *Science of the Total Environment,* 557**,** 231-239.

Perna, N. T., Plunkett, G., Burland, V., Mau, B., Glasner, J. D., Rose, D. J., Mayhew, G. F., Evans, P. S., Gregor, J., Kirkpatrick, H. A., Posfai, G., Hackett, J., Klink, S., Boutin, A., Shao, Y., Miller, L., Grotbeck, E. J., Davis, N. W., Limk, A., Dimalanta, E. T., Potamousis, K. D., Apodaca, J., Anantharaman, T. S., Lin, J. Y., Yen, G., Schwartz, D. C., Welch, R. A. & Blattner, F. R. 2001. Genome sequence of enterohaemorrhagic *Escherichia coli* O157 : H7. *Nature,* 409**,** 529-533.

Pieterse, C. M. J. & Dicke, M. 2007. Plant interactions with microbes and insects: from molecular mechanisms to ecology. *Trends in Plant Science,* 12**,** 564-569.

Plumbridge, J. & Vimr, E. 1999. Convergent pathways for utilization of the amino sugars N-acetylglucosamine, N-acetylmannosamine, and N-acetylneuraminic acid by *Escherichia coli*. *Journal of Bacteriology,* 181**,** 47-54.

Poly, F., Monrozier, L. J. & Bally, R. 2001. Improvement in the RFLP procedure for studying the diversity of *nifH* genes in communities of nitrogen fixers in soil. *Research in Microbiology,* 152**,** 95-103.

Poole, P. S., Blyth, A., Reid, C. J. & Walters, K. 1994. Myo-inositol catabolism and catabolite regulation in *Rhizobium leguminosarum bv viciae*. *Microbiology,* 140**,** 2787-2795.

Porter, S. S. & Rice, K. J. 2013. Trade-offs, spatial heterogeneity and the maintenance of microbial diversity. *Evolution,* 67**,** 599-608.

Prell, J., Boesten, B., Poole, P. & Priefer, U. B. 2002. The *Rhizobium leguminosarum* bv. *viciae* VF39 gamma-aminobutyrate (GABA) aminotransferase gene (*gabT*) is induced by GABA and highly expressed in bacteroids. *Microbiology,* 148**,** 615-623.

Prell, J. & Poole, P. 2006. Metabolic changes of rhizobia in legume nodules. *Trends in Microbiology,* 14**,** 161-168.

Preston-Mafham, J., Boddy, L. & Randerson, P. F. 2002. Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - a critique. *FEMS Microbiology Ecology,* 42**,** 1-14.

Puga-Freitas, R. & Blouin, M. 2015. A review of the effects of soil organisms on plant hormone signalling pathways. *Environmental and Experimental Botany,* 114**,** 104-116.

Rappe, M. S. & Giovannoni, S. J. 2003. The uncultured microbial majority. *Annual Review of Microbiology,* 57**,** 369-394.

Remigi, P., Zhu, J., Young, J. P. W. & Masson-Boivin, C. 2016. Symbiosis within symbiosis: Evolving nitrogen-fixing legume symbionts. *Trends in Microbiology,* 24**,** 63-75.

Reynolds, H. L., Packer, A., Bever, J. D. & Clay, K. 2003. Grassroots ecology: Plant-microbe-soil interactions as drivers of plant community structure and dynamics. *Ecology,* 84**,** 2281-2291.

Richardson, D., Felgate, H., Watmough, N., Thomson, A. & Baggs, E. 2009. Mitigating release of the potent greenhouse gas N_2O from the nitrogen cycle - could enzymic regulation hold the key? *Trends in Biotechnology,* 27**,** 388-397.

Riley, M. A. & Wertz, J. E. 2002. Bacteriocins: Evolution, ecology, and application. *Annual Review of Microbiology,* 56**,** 117-137.

Rivas, R., Martens, M., De Lajudie, P. & Willems, A. 2009. Multilocus sequence analysis of the genus *Bradyrhizobium*. *Systematic and Applied Microbiology,* 32**,** 101-110.

Roman-Ponce, B., Zhang, Y. J., Vasquez-Murrieta, M. S., Sui, X. H., Chen, W. F., Padilla, J. C. A., Guo, X. W., Gao, J. L., Yan, J., Wei, G. H. & Wang, E. T. 2016. *Rhizobium acidisoli* sp nov., isolated from root nodules of *Phaseolus vulgaris* in acid soils. *International Journal of Systematic and Evolutionary Microbiology,* 66**,** 398-406.

Roset, M. S., Ciochini, A. E., Ugalde, R. A. & De Iannino, N. I. 2006. The *Brucella abortus* cyclic beta-1,2-glucan virulence factor is substituted with O-ester-linked succinyl residues. *Journal of Bacteriology,* 188**,** 5003-5013.

Sachs, J. L., Ehinger, M. O. & Simms, E. L. 2010. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. *Journal of Evolutionary Biology,* 23**,** 1075-1089.

Sadowsky, M. J., Tully, R. E., Cregan, P. B. & Keyser, H. H. 1987. Genetic diversity in *Bradyrhizobium japonicum* serogroup-123 and its relation to genotype-specific nodulation of soybean. *Applied and Environmental Microbiology,* 53**,** 2624-2630.

Saeki, Y., Ozumi, S., Yamamoto, A., Umehara, Y., Hayashi, M. & Sigua, G. C. 2010. Changes in population occupancy of bradyrhizobia under different temperature regimes. *Microbes and Environments,* 25**,** 309-312.

Sagova-Mareckova, M., Cermak, L., Omelka, M., Kyselkova, M. & Kopecky, J. 2015. Bacterial diversity and abundance of a creek valley sites reflected soil pH and season. *Open Life Sciences,* 10**,** 61-70.

Sait, M., Hugenholtz, P. & Janssen, P. H. 2002. Cultivation of globally distributed soil bacteria from phylogenetic lineages previously only detected in cultivation-independent surveys. *Environmental Microbiology,* 4**,** 654-666.

Sameshima-Saito, R., Chiba, K. & Minamisawa, K. 2004. New method of denitrification analysis of *Bradyrhizobium* field isolates by gas chromatographic determination of ^{15}N-labeled N_2. *Applied and Environmental Microbiology,* 70**,** 2886-2891.

Sameshima-Saito, R., Chiba, K. & Minamisawa, K. 2006. Correlation of denitrifying capability with the existence of *nap*, *nir*, *nor* and *nos* genes in diverse strains of soybean bradyrhizobia. *Microbes and Environments,* 21**,** 174-184.

Santos-Beneit, F. 2015. The Pho regulon: a huge regulatory network in bacteria. *Frontiers in Microbiology,* 6**,** 402.

Schaller, K. & Nomura, M. 1976. Colicin E2 is a DNA endonuclease. *Proceedings of the National Academy of Sciences of the United States of America,* 73**,** 3989-3993.

Schirrmeister, J., Friedrich, L., Wenzel, M., Hoppe, M., Wolf, C., Gottfert, M. & Zehner, S. 2011. Characterization of the self-cleaving effector protein NopE1 of *Bradyrhizobium japonicum*. *Journal of Bacteriology,* 193**,** 3733-3739.

Schmid, M., Berg, M., Hilbi, H. & Dimroth, P. 1996. Malonate decarboxylase of *Klebsiella pneumoniae* catalyses the turnover of acetyl and malonyl thioester residues on a coenzyme-A-like prosthetic group. *European Journal of Biochemistry,* 237**,** 221-228.

Schultze, M. & Kondorosi, A. 1998. Regulation of symbiotic root nodule development. *Annual Review of Genetics,* 32**,** 33-57.

Sechi, L. A., Zanetti, S., Dupre, I., Delogu, G. & Fadda, G. 1998. Enterobacterial repetitive intergenic consensus sequences as molecular targets for typing of *Mycobacterium tuberculosis* strains. *Journal of Clinical Microbiology,* 36**,** 128-132.

Sekse, C., Bohlin, J., Skjerve, E. & Vegarud, G. E. 2012. Growth comparison of several *Escherichia coli* strains exposed to various concentrations of lactoferrin using linear spline regression. *Microbial informatics and experimentation,* 2**,** 5-5.

Shade, A., Hogan, C. S., Klimowicz, A. K., Linske, M., Mcmanus, P. S. & Handelsman, J. 2012. Culturing captures members of the soil rare biosphere. *Environmental Microbiology,* 14**,** 2247-2252.

Shaw, L. J., Nicol, G. W., Smith, Z., Fear, J., Prosser, J. I. & Baggs, E. M. 2006. *Nitrosospira* spp. can produce nitrous oxide via a nitrifier denitrification pathway. *Environmental Microbiology,* 8**,** 214-222.

Shiina, Y., Itakura, M., Choi, H., Saeki, Y., Hayatsu, M. & Minamisawa, K. 2014. Relationship between soil type and N_2O reductase genotype (*nosZ*) of indigenous soybean bradyrhizobia: *nosZ*-minus populations are dominant in andosols. *Microbes and Environments,* 29**,** 420-426.

Shimodaira, H. 2002. An approximately unbiased test of phylogenetic tree selection. Systematic Biology, 51, 492-508.

Silva, F. V., De Meyer, S. E., Simoes-Araujo, J. L., Barbe, T. D., Xavier, G. R., O'hara, G., Ardley, J. K., Rumjanek, N. G., Willems, A. & Zilli, J. E. 2014. *Bradyrhizobium manausense* sp nov., isolated from effective nodules of *Vigna unguiculata* grown in Brazilian Amazonian rainforest soils. *International Journal of Systematic and Evolutionary Microbiology,* 64**,** 2358-2363.

Soares, R. A. & Pereira Passaglia, L. M. 2010. Application of representational difference analysis to indentify genomic differences between *Bradyrhizobium elkanii* and *Bradyrhizobium japonicum* species. *Brazilian Journal of Microbiology,* 41**,** 1142-1151.

Souza, R. C., Mendes, I. C., Reis, F. B., Carvalho, F. M., Nogueira, M. A., Vasconcelos, A. T. R., Vicente, V. A. & Hungria, M. 2016. Shifts in taxonomic and functional microbial diversity with agriculture: How fragile is the Brazilian Cerrado? *BMC Microbiology,* 16**,** 42.

Spiro, S. 2012. Nitrous oxide production and consumption: regulation of gene expression by gas-sensitive transcription factors. *Philosophical Transactions of the Royal Society B,* 367**,** 1213-1225.

Sprent, J. I. 2007. Evolving ideas of legume evolution and diversity: a taxonomic perspective on the occurrence of nodulation. *New Phytologist,* 174**,** 11-25.

Steenkamp, E. T., Stepkowski, T., Przymusiak, A., Botha, W. J. & Law, I. J. 2008. Cowpea and peanut in southern Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes that belong to the large pantropical clade common in Africa. *Molecular Phylogenetics and Evolution,* 48**,** 1131-1144.

Steer, J. & Harris, J. A. 2000. Shifts in the microbial community in rhizosphere and non-rhizosphere soils during the growth of *Agrostis stolonifera*. *Soil Biology & Biochemistry,* 32**,** 869-878.

Stepkowski, T., Zak, M., Moulin, L., Kroliczak, J., Golinska, B., Narozna, D., Safronova, V. I. & Madrzak, C. J. 2011. *Bradyrhizobium canariense* and *Bradyrhizobium japonicum* are the two dominant rhizobium species in root nodules of lupin and serradella plants growing in Europe. *Systematic and Applied Microbiology,* 34**,** 368-375.

Stevenson, B. S., Eichorst, S. A., Wertz, J. T., Schmidt, T. M. & Breznak, J. A. 2004. New strategies for cultivation and detection of previously uncultured microbes. *Applied and Environmental Microbiology,* 70**,** 4748-4755.

Streeter, J. G. 2003. Effect of trehalose on survival of *Bradyrhizobium japonicum* during desiccation. *Journal of Applied Microbiology,* 95**,** 484-491.

Strohm, T. O., Griffin, B., Zumft, W. G. & Schink, B. 2007. Growth yields in bacterial denitrification and nitrate ammonification. *Applied and Environmental Microbiology,* 73**,** 1420-1424.

Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., Gyaneshwar, P., Denny, R., Mudge, J., Bharti, A. K., Farmer, A. D., May, G. D., Woodward, J. E., Medigue, C., Vallenet, D., Lajus, A., Rouy, Z., Martinez-Vaz, B., Tiffin, P., Young, N. D. & Sadowsky, M. J. 2013. Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biology,* 14**,** R17.

Sugawara, M. & Sadowsky, M. J. 2013. Influence of elevated atmospheric carbon dioxide on transcriptional responses of *Bradyrhizobium japonicum* in the soybean rhizoplane. *Microbes and Environments,* 28**,** 217-227.

Sun, B., Roberts, D. M., Dennis, P. G., Caul, S., Daniell, T. J., Hallett, P. D. & Hopkins, D. W. 2014. Microbial properties and nitrogen contents of arable soils under different tillage regimes. *Soil Use and Management,* 30**,** 152-159.

Suvorova, I. A., Ravcheev, D. A. & Gelfand, M. S. 2012. Regulation and Evolution of Malonate and Propionate Catabolism in Proteobacteria. *Journal of Bacteriology,* 194**,** 3234-3240.

Suzuki, R. & Shimodaira, H. 2006. Pvclust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics,* 22**,** 1540-1542.

Tan, Z. Y., Hurek, T., Vinuesa, P., Muller, P., Ladha, J. K. & Reinhold-Hurek, B. 2001. Specific detection of *Bradyrhizobium* and *Rhizobium* strains colonizing rice (*Oryza sativa*) roots by 16S-23S ribosomal DNA intergenic spacer-targeted PCR. *Applied and Environmental Microbiology,* 67**,** 3655-3664.

Tang, J., Bromfield, E. S. P., Rodrigue, N., Cloutier, S. & Tambong, J. T. 2012. Microevolution of symbiotic *Bradyrhizobium* populations associated with soybeans in east North America. *Ecology and Evolution,* 2**,** 2943-2961.

Terakado-Tonooka, J., Fujihara, S. & Ohwaki, Y. 2013. Possible contribution of *Bradyrhizobium* on nitrogen fixation in sweet potatoes. *Plant and Soil,* 367**,** 639-650.

Throback, I. N., Enwall, K., Jarvis, A. & Hallin, S. 2004. Reassessing PCR primers targeting *nirS, nirK* and *nosZ* genes for community surveys of denitrifying bacteria with DGGE. *FEMS Microbiology Ecology,* 49**,** 401-417.

Tong, Z. K. & Sadowsky, M. J. 1994. A selective medium for the isolation and quantification of *Bradyrhizobium japonicum* and *Bradyrhizobioum elkanii* strains from soils and inoculants. *Applied and Environmental Microbiology,* 60**,** 581-586.

Torres, D., Revale, S., Obando, M., Maroniche, G., Paris, G., Perticari, A., Vazquez, M., Wisniewski-Dye, F., Martinez-Abarca, F. & Cassan, F. 2015. Genome sequence of *Bradyrhizobium japonicum* E109, one of the most agronomically used nitrogen-fixing rhizobacteria in Argentina. *Genome announcements,* 3**,** e01566-14.

Tritt, A., Eisen, J. A., Facciotti, M. T. & Darling, A. E. 2012. An integrated pipeline for *de novo* assembly of microbial genomes. *PLOS ONE,* 7**,** e42304.

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.

Vairinhos, F., Wallace, W. & Nicholas, D. J. D. 1989. Simultaneous assimilation and denitrification of nitrate by *Bradyrhizobium japonicum*. *Journal of General Microbiology,* 135**,** 189-193.

Van Berkum, P. 1987. Expression of uptake hydrogenase and hydrogen oxidation during heterotrophic growth of *Bradyrhizobium japonicum*. *Journal of Bacteriology,* 169**,** 4565-4569.

Van Berkum, P. 1990. Evidence for a third uptake hydrogenase phenotype among the soybean bradyrhizobia. *Applied and Environmental Microbiology,* 56**,** 3835-3841.

Van Berkum, P. & Fuhrmann, J. J. 2000. Evolutionary relationships among the soybean bradyrhizobia reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. *International Journal of Systematic and Evolutionary Microbiology,* 50**,** 2165-2172.

Van Berkum, P. & Keyser, H. H. 1985. Anaerobic growth and denitrification among different serogroups of soybean rhizobia. *Applied and Environmental Microbiology,* 49**,** 772-777.

Van Berkum, P. & Maier, R. J. 1988. Lack of carbon substrate repression of uptake hydrogenase activity in *Bradyrhizobium japonicum* SR. *Journal of Bacteriology,* 170**,** 1962-1964.

Van Berkum, P., Terefework, Z., Paulin, L., Suomalainen, S., Lindstrom, K. & Eardly, B. D. 2003. Discordant phylogenies within the *rrn* loci of rhizobia. *Journal of Bacteriology,* 185**,** 2988-2998.

Van Der Heijden, M. G. A., Bardgett, R. D. & Van Straalen, N. M. 2008. The unseen majority: soil microbes as drivers of plant diversity and productivity in terrestrial ecosystems. *Ecology Letters,* 11**,** 296-310.

Van Diepeningen, A. D., De Vos, O. J., Korthals, G. W. & Van Bruggen, A. H. C. 2006. Effects of organic versus conventional management on chemical and biological parameters in agricultural soils. *Applied Soil Ecology,* 31**,** 120-135.

Van Elsas, J. D., Garbeva, P. & Salles, J. 2002. Effects of agronomical measures on the microbial diversity of soils as related to the suppression of soil-borne plant pathogens. *Biodegradation,* 13**,** 29-40.

Van Rhijn, P. & Vanderleyden, J. 1995. The *Rhizobium*-plant symbiosis. *Microbiological Reviews,* 59**,** 124-142.

VanInsberghe, D., Maas, K. R., Cardenas, E., Strachan, C. R., Hallam, S. J. & Mohn, W. W. 2015. Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME Journal,* 9**,** 2435–2441.

Vartoukian, S. R., Palmer, R. M. & Wade, W. G. 2010. Strategies for culture of 'unculturable' bacteria. *FEMS Microbiology Letters,* 309**,** 1-7.

Vauclare, P., Bligny, R., Gout, E. & Widmer, F. 2013. An overview of the metabolic differences between *Bradyrhizobium japonicum* 110 bacteria and differentiated bacteroids from soybean (*Glycine max*) root nodules: an in vitro13C- and 31P-nuclear magnetic resonance spectroscopy study. *FEMS Microbiology Letters***,** 49-56.

Velasco, L., Mesa, S., Delgado, M. J. & Bedmar, E. J. 2001. Characterization of the *nirK* gene encoding the respiratory, Cu-containing nitrite reductase of *Bradyrhizobium japonicum*. *Biochimica Et Biophysica Acta,* 1521**,** 130-134.

Versalovic, J., Koeuth, T. & Lupski, J. R. 1991. Distribution of repetitive DNA sequences in eubacteria and application to fingerprinting of bacterial genomes. *Nucleic Acids Research,* 19**,** 6823-6831.

Vinuesa, P., Rademaker, J. L. W., De Bruijn, F. J. & Werner, D. 1998. Genotypic characterization of *Bradyrhizobium* strains nodulating endemic woody legumes of the Canary Islands by PCR-restriction fragment length polymorphism analysis of genes encoding 16S rRNA (16S rDNA) and 16S-23S rDNA intergenic spacers, repetitive extragenic palindromic PCR genomic fingerprinting, and partial 16S rDNA sequencing. *Applied and Environmental Microbiology,* 64**,** 2096-2104.

Vinuesa, P., Rojas-Jimenez, K., Contreras-Moreira, B., Mahna, S. K., Prasad, B. N., Moe, H., Selvaraju, S. B., Thierfelder, H. & Werner, D. 2008. Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent. *Applied and Environmental Microbiology,* 74**,** 6987-6996.

Waldrop, M. P. & Firestone, M. K. 2004. Microbial community utilization of recalcitrant and simple carbon compounds: impact of oak-woodland plant communities. *Oecologia,* 138**,** 275-284.

Wang, E. T., Van Berkum, P., Beyene, D., Sui, X. H., Dorado, O., Chen, W. X. & Martinez-Romero, E. 1998. *Rhizobium huautlense* sp. nov., a symbiont *of Sesbania herbacea* that has a close phylogenetic relationship with *Rhizobium galegae*. *International Journal of Systematic Bacteriology,* 48**,** 687-699.

Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Research,* 43**,** W78-W84.

Wanner, B. L. & Chang, B. D. 1987. The *phoBR* operon in *Escherichia coli* K-12. *Journal of Bacteriology,* 169**,** 5569-5574.

Weir, B. S., Turner, S. J., Silvester, W. B., Park, D. C. & Young, J. A. 2004. Unexpectedly diverse *Mesorhizobium* strains and *Rhizobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species. *Applied and Environmental Microbiology,* 70**,** 5980-5987.

Weisburg, W. G., Barns, S. M., Pelletier, D. A. & Lane, D. J. 1991. 16S ribosomal DNA amplification for phylogenetic study. *Journal of Bacteriology,* 173**,** 697-703.

Whitman, W. B., Woyke, T., Klenk, H. P., Zhou, Y. G., Lilburn, T. G., Beck, B. J., De Vos, P., Vandamme, P., Eisen, J. A., Garrity, G., Hugenholtz, P. & Kyrpides, N. C. 2015. Genomic encyclopedia of bacterial and archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains. *Standards in Genomic Sciences,* 10.

Wickham, H. 2009. *ggplot2: Elegant Graphics for Data Analysis*, Springer-Verlag New York}.

Woese, C. R. 1987. Bacterial evolution. *Microbiological Reviews,* 51**,** 221-271.

Woomer, P. L., Singleton, P. W. & Bohlool, B. B. 1988. Reliability of the most-probable-number technique for enumerating rhizobia in tropical soils. *Applied and Environmental Microbiology,* 54**,** 1494-1497.

Wu, J. & Rosen, B. P. 1991. The ArsR protein is a trans-acting regulatory protein. *Molecular Microbiology,* 5**,** 1331-1336.

Wunsch, P. & Zumft, W. G. 2005. Functional domains of NosR, a novel transmembrane iron-sulfur flavoprotein necessary for nitrous oxide respiration. *Journal of Bacteriology,* 187**,** 1992-2001.

Xi, H. L., Schneider, B. L. & Reitzer, L. 2000. Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *Journal of Bacteriology,* 182**,** 5332-5341.

Yang, C., Rodionov, D. A., Li, X., Laikova, O. N., Gelfand, M. S., Zagnitko, O. P., Romine, M. F., Obraztsova, A. Y., Nealson, K. H. & Osterman, A. L. 2006. Comparative genomics and experimental characterization of N-acetylglucosamine utilization pathway of *Shewanella oneidensis*. *Journal of Biological Chemistry,* 281**,** 29872-29885.

Yang, Q., Wang, X. & Shen, Y. 2013. Comparison of soil microbial community catabolic diversity between rhizosphere and bulk soil induced by tillage or residue retention. *Journal of Soil Science and Plant Nutrition,* 13**,** 187-199.

Yoon, J.-H., Kang, S.-J., Yi, H.-S., Oh, T.-K. & Ryu, C.-M. 2010. *Rhizobium soli* sp nov., isolated from soil. *International Journal of Systematic and Evolutionary Microbiology,* 60**,** 1387-1393.

Yoshida, K. I., Aoyama, D., Ishio, I., Shibayama, T. & Fujita, Y. 1997. Organization and transcription of the myo-inositol operon, *iol*, of *Bacillus subtilis*. *Journal of Bacteriology,* 179**,** 4591-4598.

Young, J. P. W., Crossman, L. C., Johnston, A. W. B., Thomson, N. R., Ghazoui, Z. F., Hull, K. H., Wexler, M., Curson, A. R. J., Todd, J. D., Poole, P. S., Mauchline, T. H., East, A. K., Quail, M. A., Churcher, C., Arrowsmith, C., Cherevach, I., Chillingworth, T., Clarke, K., Cronin, A., Davis, P., Fraser, A., Hance, Z., Hauser, H., Jagels, K., Moule, S., Mungall, K., Norbertczak, H., Rabbinowitsch, E., Sanders, M., Simmonds, M., Whitehead, S. & Parkhill, J. 2006. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biology,* 7**,** R34.

Young, J. P. W., Downer, H. L. & Eardly, B. D. 1991. Phylogeny of the phototrophic Rhizobium strain BTAi1 by polymerase chain reaction-based sequencing of a 16S rRNA gene segment. *Journal of Bacteriology,* 173**,** 2271-2277.

Zabaloy, M. C. & Gomez, M. A. 2005. Diversity of rhizobia isolated from an agricultural soil in Argentina based on carbon utilization and effects of herbicides on growth. *Biology and Fertility of Soils,* 42**,** 83-88.

Zaprasis, A., Liu, Y.-J., Liu, S.-J., Drake, H. L. & Horn, M. A. 2010. Abundance of novel and diverse *tfdA*-like genes, encoding putative phenoxyalkanoic acid herbicide-degrading dioxygenases, in soil. *Applied and Environmental Microbiology,* 76**,** 119-128.

Zengler, K., Toledo, G., Rappe, M., Elkins, J., Mathur, E. J., Short, J. M. & Keller, M. 2002. Cultivating the uncultured. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 15681-15686.

Zhalnina, K., De Quadros, P. D., Gano, K. A., Davis-Richardson, A., Fagen, J. R., Brown, C. T., Giongo, A., Drew, J. C., Sayavedra-Soto, L. A., Arp, D. J., Camargo, F. a. O., Daroub, S. H., Clark, I. M., Mcgrath, S. P., Hirsch, P. R. & Triplett, E. W. 2013. Ca. *Nitrososphaera* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. *Frontiers in microbiology,* 4**,** 104-104.

Zhang, H. Y., Ishige, K. & Kornberg, A. 2002. A polyphosphate kinase (PPK2) widely conserved in bacteria. *Proceedings of the National Academy of Sciences of the United States of America,* 99**,** 16678-16683.

Zhang, N., Wang, D. D., Liu, Y. P., Li, S. Q., Shen, Q. R. & Zhang, R. F. 2014a. Effects of different plant root exudates and their organic acid components on chemotaxis, biofilm formation and colonization by beneficial rhizosphere-associated bacterial strains. *Plant and Soil,* 374**,** 689-700.

Zhang, X. X., Guo, H. J., Wang, R., Sui, X. H., Zhang, Y. M., Wang, E. T., Tian, C. F. & Chen, W. X. 2014b. Genetic divergence of *Bradyrhizobium* strains nodulating soybeans as revealed by multilocus sequence analysis of genes inside and outside the symbiosis island. *Applied and Environmental Microbiology,* 80**,** 3181-3190.

Zhao, S. C., Qiu, S. J., Cao, C. Y., Zheng, C. L., Zhou, W. & He, P. 2014. Responses of soil properties, microbial community and crop yields to various rates of nitrogen fertilization in a wheat-maize cropping system in north-central China. *Agriculture Ecosystems & Environment,* 194**,** 29-37.

Zhu, L., Ding, W., Feng, L.-J., Kong, Y., Xu, J. & Xu, X.-Y. 2012. Isolation of aerobic denitrifiers and characterization for their potential application in the bioremediation of oligotrophic ecosystem. *Bioresource Technology,* 108**,** 1-7.

Zuber, M., Harker, A. R., Sultana, M. A. & Evans, H. J. 1986. Cloning and expression of *Bradyrhizobium japonicum* uptake hydrogenase structural genes in *Escherichia coli*. *Proceedings of the National Academy of Sciences of the United States of America,* 83**,** 7668-7672.

10 Appendices

Appendix A: 16S sequence NCBI accession number

Table A-1: Table of 16S sequence accession numbers for submission SUB1933222

Identity	Strain	Soil treatment/ Host	NCBI accession number
B. japonicum	G2	Grassland	KX838339
Bradyrhizobium sp.	G3	Grassland	KX838340
Bradyrhizobium sp.	G5	Grassland	KX838341
B. japonicum	G6	Grassland	KX838342
Bradyrhizobium sp.	G8	Grassland	KX838343
Bradyrhizobium sp.	G9	Grassland	KX838344
Bradyrhizobium sp.	G15	Grassland	KX838345
Bradyrhizobium sp.	G22	Grassland	KX838346
Bradyrhizobium sp.	G31	Grassland	KX838347
Bradyrhizobium sp.	G46	Grassland	KX838348
Bradyrhizobium sp.	G54	Grassland	KX838349
Bradyrhizobium sp.	BF2	Bare fallow	KX838329
B. canariense	BF6	Bare fallow	KX838330
Bradyrhizobium sp.	BF20	Bare fallow	KX838331
Bradyrhizobium sp.	BF26	Bare fallow	KX838332
Bradyrhizobium sp.	BF28	Bare fallow	KX838333
Bradyrhizobium sp.	BF30	Bare fallow	KX838334
Bradyrhizobium sp.	BF41	Bare fallow	KX838335
B. jicamae	BF48	Bare fallow	KX838336
B. canariense	BF49	Bare fallow	KX838337
Bradyrhizobium sp.	BF52	Bare fallow	KX838338
B. canariense	A1	Arable	KX838319
B. canariense	A2	Arable	KX838320
B. canariense	A3	Arable	KX838321
Bradyrhizobium sp.	A4	Arable	KX838322
Bradyrhizobium sp.	A5	Arable	KX838323
Bradyrhizobium sp.	A10	Arable	KX838324
Bradyrhizobium sp.	A12	Arable	KX838325
Bradyrhizobium sp.	A16	Arable	KX838326
Bradyrhizobium sp.	A19	Arable	KX838327
Bradyrhizobium sp.	A20	Arable	KX838328
Bradyrhizobium sp.	W4	Woodland	KX838355
Bradyrhizobium sp.	W5	Woodland	KX838356
Bradyrhizobium sp.	W9	Woodland	KX838357
Bradyrhizobium sp.	W19	Woodland	KX838358
Bradyrhizobium sp.	W22	Woodland	KX838359
Bradyrhizobium sp.	W23	Woodland	KX838360
Bradyrhizobium sp.	W24	Woodland	KX838361

Identity	Strain	Soil treatment/Host	NCBI accession number
Bradyrhizobium sp.	W46	Woodland	KX838362
Bradyrhizobium sp.	W49	Woodland	KX838363
Bradyrhizobium sp.	W74	Woodland	KX838364
Bradyrhizobium sp.	W91	Woodland	KX838365
Bradyrhizobium sp.	GOR1	*Ulex europeaus*	KX838352
B. japonicum	GOR2	*Ulex europeaus*	KX838353
B. japonicum	GOR4	*Ulex europeaus*	KX838354
Bradyrhizobium sp.	BRM1	*Cytisus scoparius*	KX838350
Bradyrhizobium sp.	BRM2	*Cytisus scoparius*	KX838351

Appendix B: Housekeeping genes accession numbers

Table B-1: Table of NCBI submission IDs for the five housekeeping genes awaiting accessions

Housekeeping gene	Reading frame	Submission IDs
gyrB	1	1955074
	2	1955113 & 1955366
	3	1955114
atpD	1	1955115
	2	1955118 & 1955347
	3	1955120
recA	1	1955129
	2	1955130 & 1955360
	3	1955131
dnaK	1	1955135 & 1955362
	2	1955138
	3	1955139
rpoB	1	1955121
	2	1955476 & 1955364
	3	1955125

NB: Final submission IDs and accession numbers, once assigned, may be different to those listed.

Table B-2: Assigned accession numbers for the NCBI database

Isolate	Accession	Gene	Isolate	Accession	Gene	Isolate	Accession	Gene
G8	KX881084	*atpD*	BF6	KX889423	*rpoB*	W9	KX889418	*rpoB*
A12	KX881085	*atpD*	A1	KX889424	*rpoB*	W19	KX889419	*rpoB*
G9	KX881086	*atpD*	A2	KX889425	*rpoB*	W24	KX889420	*rpoB*
W19	KX881087	*atpD*	A5	KX889426	*rpoB*	W91	KX889421	*rpoB*
W24	KX881088	*atpD*	A10	KX889427	*rpoB*	A3	KX889422	*rpoB*
BRM1	KX881080	*dnaK*	A12	KX889428	*rpoB*	BF41	KX889443	*rpoB*
G8	KX881081	*dnaK*	G31	KX889429	*rpoB*	BF26	KX889444	*rpoB*
G9	KX881082	*rpoB*	GOR4	KX889430	*rpoB*	BF48	KX889445	*rpoB*
W4	KX881083	*rpoB*	G2	KX889431	*rpoB*	G46	KX889446	*rpoB*
A19	KX881089	*rpoB*	G54	KX889432	*rpoB*			
BF49	KX881090	*rpoB*	BF20	KX889433	*rpoB*			
GOR1	KX881091	*rpoB*	A4	KX889434	*rpoB*			
GOR2	KX881092	*rpoB*	G6	KX889435	*rpoB*			
BRM1	KX881093	*rpoB*	A20	KX889436	*rpoB*			
BRM2	KX881094	*rpoB*	G3	KX889437	*rpoB*			
W5	KX881095	*rpoB*	G15	KX889438	*rpoB*			
W22	KX881096	*rpoB*	G8	KX889439	*rpoB*			
W46	KX881097	*rpoB*	A16	KX889440	*rpoB*			
W49	KX881098	*rpoB*	W23	KX889441	*rpoB*			
W74	KX881099	*rpoB*	BF28	KX889442	*rpoB*			

Appendix C: Isolate numbers and identifiers

Table C-1: Table of isolate numbers and identifiers

Isolate number	Isolate name	Soil treatment
1	G2	Grassland
2	G3	Grassland
3	G5	Grassland
4	G6	Grassland
5	G8	Grassland
6	G9	Grassland
7	G15	Grassland
8	G22	Grassland
9	G31	Grassland
10	G46	Grassland
11	G54	Grassland
12	BF2	Bare fallow
13	BF6	Bare fallow
14	BF20	Bare fallow
15	BF26	Bare fallow
16	BF28	Bare fallow
17	BF30	Bare fallow
18	BF41	Bare fallow
19	BF48	Bare fallow
20	BF49	Bare fallow
21	BF52	Bare fallow
22	A1	Arable
23	A3	Arable
24	A2	Arable
25	A4	Arable
26	A5	Arable
27	A10	Arable
28	A12	Arable
29	A16	Arable
30	A19	Arable
31	A20	Arable
32	W4	Woodland
33	W5	Woodland
34	W9	Woodland
35	W19	Woodland
36	W22	Woodland
37	W23	Woodland
38	W24	Woodland
39	W46	Woodland
40	W49	Woodland

Isolate number	Isolate name	Soil treatment
41	W74	Woodland
42	W91	Woodland
43	*B. japonicum* USDA 6	Symbiotic/Reference
44	GOR1	Symbiotic
45	GOR2	Symbiotic
46	GOR4	Symbiotic
47	BRM1	Symbiotic
48	BRM2	Symbiotic
49	RD5-C2	Reference

Appendix D: Genome assembly code

Config file

```
[LIB]
max_rd_len=615
reverse_seq=0
asm_flags=3
rank=1
q=../Ion_Torrent_data/22/Galaxy24-[Trimmomatic_on_data_4_22].fastq
[LIB]
avg_ins=6000
reverse_seq=1
asm_flags=3
rank=2
pair_num_cutoff=5
#minimum aligned length to contigs for a reliable read location
(at least 35 for large insert size)
map_len=35
q1=../Illumina_6kb_mate_library/22/Galaxy58-
[FASTQ_Quality_Trimmer_on_data_10_22.1].fastq
q2=../Illumina_6kb_mate_library/22/Galaxy59-
[FASTQ_Quality_Trimmer_on_data_13_22.2].fastq
```

Commands:

```
./SOAPdenovo-127mer all -s configtrimmed.txt -p 66 -K 61 –m 83 -R
-o 22_FC 1> 22_FCass.log 2> 22_FCass.err

./GapCloser -a 22_FCtrimmed.scafSeq -o 22_FC_GCtrimmed.fasta -b
configtrimmed.txt -t 20 -l 615
```

Appendix E: Gap closing primers

Table E-1: Primers used for gap closing genomes with associated melting temperatures

Genome	Primer name	Primer sequence	T_m (°C)
G22	Gap1_F	GTGAAGCTGGACGCTCCGTGACCG	69.5
	Gap1_R	GCACGCCGTCGTTCAGTGCCG	67.6
	Gap2_F	CGGCACTGAACGACGGCGTGC	67.6
	Gap2_R	GACGCTGCCGCCCCAGTTC	65.3
	Gap3_F	GAACTGGGGCGGCAGCGTC	65.3
	Gap3_R	GTCATGCGAAGATCTCACCTCTCG	64.4
	Gap4_F	CGAGAGGTGAGATCTTCGCATGAC	64.4
	Gap4_R	CATGCCGCGCATGGTGGTGC	65.5
	Gap5_F	TCGGACTTGATGGTGGCAAC	59.4
	Gap5_R	GTTGACGAGCATGGTCCGC	61
	Gap6_F	CGGGGCGACGGCTACAAG	62.8
	Gap6_R	GGTGATTTGCGTGCAGAAAT	55.3
	Gap7_F	CAGCATCCAGATCTCGCGTA	59.4
	Gap7_R	TCAAGGGATGAGCGAATAGACC	60.3
	Gap8_F	GTTCCTCTAGGCGGGCCG	62.8
	Gap8_R	AGTCTGAACACAAACGGCGC	59.4
	Gap9_F	ACCGACTCCGCGGCTGAC	62.8
	Gap9_R	GGTGTATTCGGTGTCGACCGTCG	66
	Gap10_F	CTGGCTGAGCACGACCGCGACCAC	71.3
	Gap10_R	GCCGAAGCTTACTTGAACAG	59.3
	Gap11_F	TCACCGTCTCGAACACCTC	58.8
	Gap11_R	GATGTGCCATCGACCCAGAT	59.4
	Gap12_F	GAATTTTAGCCGCCGATGG	56.7
	Gap12_R	CACCGCGACCTTGATTCAGA	59.4
	Gap13_F	AGACCAGTCAACGGTGAACAG	59.8
	Gap13_R	CGACTTCCCCGGCTAGCATC	63.5
	Gap14_F	TCAACTCTCGGTCGAGGAGT	59.4
	Gap14_R	TTGTGCACATCCATGCTGA	54.5
	Gap15_F	ACGGGAGCGGATCATTGC	58.2
	Gap15_R	GCCAGTCAAACTCAGGCATG	59.4
	Gap16_F	GTAACACGGCTATCGCGCA	58.8
	Gap16_R	TGTTAGAATCCAGAATCCAGATGC	59.3
	Gap17_F	CGTCGGTCTGCTCGCCGAAT	63.5
	Gap17_R	ACCACGGCGCCGACCAGC	65.1
	Gap18_F	CCCGATAGATGATGAAACGGGC	62.1
	Gap18_R	CCGAACAGCGCGCCTGCC	65.1
	Gap19_F	CGCTCGCCAGAAAACACC	58.2
	Gap19_R	CGAAGGCTTGTTGATCGGCG	61.4
	Gap20_F	GCCAGGAGGATGTTGTGCG	61
	Gap20_R	GTGACAGACAGGTATGGCTTGC	62.1
	Gap21_F	CACCGCGACGCCGATTCCTG	65.5

	Primer name	Primer sequence	T_m (^oC)
	Gap21_R	CATCGCAACTGTCGCGGCAT	61.4
	Gap22_F	CAACACCGTGATCAACCGG	58.8
	Gap22_R	GCGCCATAGTTTCAATTCTCCG	60.3
	Gap23_F	TTCTGATCCTTGTTCTGGTCCTT	58.9
	Gap23_R	ATGAGGAAGCTGTTGGCCAC	59.4
	Gap24_F	TTGCGGGCGTTGGGGCAGTT	63.5
	Gap24_R	TTCTATCTATTTCGGCTGACACG	58.9
	Gap25_F	CGGCTTTTCTGCTGACGCAC	61.4
	Gap25_R	GCCCAGCACCTCCTATTTGC	61.4
	Gap26_F	CGAGGAACGGGCCGATCG	62.8
	Gap26_R	TGGACTGTAGAAAGCATTCTCCG	60.6
	Gap27_F	GTTTCGGGCAACATCGCTG	58.8
	Gap27_R	GAAGCGCCCGTCTTCTCGAA	61.4
	Gap28_F	TCTCAGCAAAAATCCCACGG	57.3
	Gap28_R	AAGCGCTAACCACGACGG	58.2
	Gap29_F	GTCCTGTGCCTATACGTGGTG	61.8
	Gap29_R	ATAGTCGATTAGGTAGGTGCGGA	60.6
	Gap30_F	TCCTCTTGCGTGATGTCGGT	59.4
	Gap30_R	CGGCTGCGCTTCCAGAATG	61
	Gap31_F	CGCATCCATGTCCTTCTGC	58.8
	Gap31_R	TTGGCTGGAGCTCGGTCCG	63.1
	Gap8F_redesign	CTTGATGTAGTATTCGATTTTGCA	55.9
	Gap8R_redesign	GGATGGAAGCTGTGACGCG	61.0
	Gap11F_redesign	CCACATGTCTTTGCGCCAG	58.8
	Gap11R_redesign	AACTGTTGGCCGATCGCG	58.2
	Gap13F_redesign	ATATCGGCCCTTCAGCGAC	58.8
	Gap13R_redesign	GGAAGTCGGCGTGAAGTGG	61.0
	Gap14F_redesign	TCTAAATCAACAGAAGTCCGCT	56.5
	Gap14R_redesign	CTGATCGACGCAACACTTTATCT	58.9
	Gap15F_redesign	GATACGGCGAGTCTCCAC	58.2
	Gap15R_redesign	TACCACGTCAGCATCAAGGT	57.3
	Gap16F_redesign	GCGCGTGTTGCTGGTCTG	60.5
	Gap16R_redesign	GCATTTCGTTGTCGTTCCCA	57.3
	Gap24F_redesign	ACCGTCCCCAAAGCCACTTA	59.4
	Gap24R_redesign	TCATCGGAACCTCCCTGACG	61.4
	Gap25F_redesign	CGTATTTGGGACGCCGGGTAACG	66.0
	Gap25R_redesign	GCGATGATTGCGGCCATC	58.2
	Gap29F_redesign	AGTGGCTCTGCACAGTGAT	56.7
	Gap29R_redesign	AGGTCAATCACCACGCGGGCGTCG	69.5
	Gap30F_redesign	TCTGCACAGTGATTTTGACAGG	58.4
	Gap30R_redesign	AGGTCTCGATCCTCTGTCCG	61.4
BF49	Gap1_F	GACTTGGAATTGCATTCGACAT	56.5
	Gap1_R	TGACGCGGGACTCTTTGAGATTTAT	61.3
	Gap2_F	GCGGCGGGAATCTCGACATT	61.4
	Gap2_R	TCTCCGAGGACGCGCAGG	62.8

Appendix F: Accession numbers of for sequences used in 16S phylogeny in Chapter 5

Sequence name	Sequence length	Accession number
Bradyrhizobium neotropicale strain BR 10247	1301 bp	NR_133987.1
Bradyrhizobium icense strain LMTR 13	1413 bp	NR_133707.1
Bradyrhizobium oligotrophicum strain S58	1484 bp	NR_118384.1
Bradyrhizobium rifense strain CTAW71	1516 bp	NR_116361.1
Bradyrhizobium pachyrhizi strain PAC48	1483 bp	NR_043037.1
Bradyrhizobium jicamae strain PAC68	1484 bp	NR_043036.1
Bradyrhizobium ottawaense strain OO99	1420 bp	NR_133988.1
Bradyrhizobium manausense strain BR 3351	1353 bp	NR_133986.1
Bradyrhizobium ingae strain BR 10250	1358 bp	NR_133985.1
Bradyrhizobium ganzhouense strain RITF806	1363 bp	NR_133706.1
Bradyrhizobium valentinum strain LmjM3	1490 bp	NR_125638.1
Bradyrhizobium elkanii strain USDA 76	1340 bp	NR_117947.1
Bradyrhizobium daqingense strain CCBAU 15774	1300 bp	NR_117944.1
Bradyrhizobium arachidis strain CCBAU 051107	1347 bp	NR_117791.1
Bradyrhizobium lablabi strain CCBAU 23086	1356 bp	NR_117513.1
Bradyrhizobium cytisi strain CTAW11	1516 bp	NR_116360.1
Bradyrhizobium yuanmingense strain B071	1426 bp	NR_028768.1
Bradyrhizobium paxllaeri strain LMTR 21	1413 bp	NR_133708.1
Bradyrhizobium huanghuaihaiense strain CCBAU 23303	1348 bp	NR_117945.1
Bradyrhizobium liaoningense strain 2281	1336 bp	NR_114611.1
Bradyrhizobium canariense strain BTA-1	1481 bp	NR_042177.1
Bradyrhizobium lupini strain USDA 3051	1480 bp	NR_134836.1
Bradyrhizobium denitrificans strain LMG 8443	1443 bp	NR_118982.1
Bradyrhizobium japonicum strain LMG 6138	1441 bp	NR_118981.1
Bradyrhizobium denitrificans strain LMG 8443	1443 bp	NR_118770.1
Bradyrhizobium daqingense strain CCBAU 15774	1372 bp	NR_118648.1
Bradyrhizobium retamae strain Ro19	1345 bp	NR_118548.1
Bradyrhizobium liaoningense strain LMG 18230	1430 bp	NR_114733.1
Bradyrhizobium japonicum strain NBRC 14783	1414 bp	NR_112930.1
Bradyrhizobium liaoningense strain NBRC 100396	1414 bp	NR_112929.1
Bradyrhizobium yuanmingense strain NBRC 100594	1414 bp	NR_112928.1
Bradyrhizobium elkanii strain NBRC 14791	1414 bp	NR_112927.1
Bradyrhizobium iriomotense strain EK05	1413 bp	NR_112671.1
Bradyrhizobium liaoningense strain 2281	1425 bp	NR_112095.1
Bradyrhizobium liaoningense strain 2281	1476 bp	NR_041785.1
Bradyrhizobium elkanii strain USDA 76	1480 bp	NR_036953.1
Bradyrhizobium japonicum strain 3I1b6	1481 bp	NR_036865.1
Bradyrhizobium betae strain PL7HG1	1483 bp	NR_029104.1
Bradyrhizobium japonicum strain DSM 30131	1482 bp	NR_119191.1
Bradyrhizobium elkanii strain ATCC 49852	1042 bp	NR_114610.1

Sequence name	Sequence length	Accession number
Bradyrhizobium canariense strain NBRC 103049	1414 bp	NR_114199.1
Bradyrhizobium betae strain NBRC 103048	1414 bp	NR_114198.1
Bradyrhizobium iriomotense strain NBRC 102520	1414 bp	NR_114138.1
Bradyrhizobium denitrificans strain IFAM 1005	1477 bp	NR_041827.1
Bradyrhizobium sp. LTSP849	1497 bp	NZ_JYMR01000018.1
Bradyrhizobium sp. LTSP857	1497 bp	NZ_JYMS01000032.1
Bradyrhizobium sp. LTSP885	1499 bp	NZ_JYMT01000017.1
Bradyrhizobium sp. LTSPM299	1499 bp	NZ_JYMU01000004.1
Bradyrhizobium sp. S23321	1490 bp	NC_017082
B. diazoefficiens USDA 110	1490 bp	NC_004463
B. japonicum USDA 6	1490 bp	NC_017249
B. japonicum E109	1498 bp	NZ_CP010313
B. oligotrophicum S58	1493 bp	NC_020453
Bradyrhizobium sp. ORS278	1489 bp	NC_009445
Bradyrhizobium sp. BTAi1	1489 bp	NC_009485

Appendix G: GrowthRates output

Table G-1: Table of maximum OD for each replicate and each isolate generated by the GrowthRates program

Isolate	Treatment	Rep1 MaxOD	Rep2 MaxOD	Rep3 MaxOD
A1	Arable	0.557	0.71	0.571
A10	Arable	0.96	0.581	0.501
A12	Arable	0.69	0.795	0.707
A16	Arable	0.583	0.817	0.805
A19	Arable	0.981	0.892	0.015
A2	Arable	0.586	0.599	0.589
A20	Arable	0.858	0.69	0.798
A3	Arable	0.626	0.553	0.751
A4	Arable	0.783	0.618	0.574
A5	Arable	0.52	0.558	0.688
BF2	Bare fallow	0.77	0.556	0.875
BF20	Bare fallow	0.855	0.566	0.667
BF30	Bare fallow	0.363	1.156	0.785
BF41	Bare fallow	0.432	0.602	0.432
BF48	Bare fallow	0.361	0.388	0.429
BF49	Bare fallow	0.836	0.51	0.678
BF52	Bare fallow	0.901	1.083	0.857
BF6	Bare fallow	0.689	0.81	0.891
BRM1	Symbiotic	0.62	0.883	0.615
BRM2	Symbiotic	0.621	0.891	0.77
G2	Grassland	0.858	0.81	0.634
G15	Grassland	0.727	0.671	0.653
G22	Grassland	0.479	0.319	0.242
G3	Grassland	0.722	0.704	0.73
G5	Grassland	0.614	0.739	0.803
G54	Grassland	0.731	0.755	0.743
G6	Grassland	0.965	0.778	0.735
G8	Grassland	0.817	0.928	0.703
G9	Grassland	0.113	0.093	0.112
GOR1	Symbiotic	0.672	0.6	0.805
GOR2	Symbiotic	0.547	0.57	0.565

Isolate	Treatment	Rep1 MaxOD	Rep2 MaxOD	Rep3 MaxOD
GOR4	Symbiotic	0.771	0.688	0.655
RD5C2	Symbiotic	0.881	0.76	0.902
USDA 6	Symbiotic	0.648	0.75	0.672
W19	Woodland	0.341	0.297	0.208
W22	Woodland	0.642	0.855	0.509
W23	Woodland	0.629	0.553	0.601
W24	Woodland	0.13	0.174	0.176
W4	Woodland	0.199	0.622	0.964
W46	Woodland	0.465	0.519	0.457
W49	Woodland	0.516	0.535	0.322
W5	Woodland	0.226	0.211	0.174
W74	Woodland	0.188	0.179	0.154
W9	Woodland	0.106	0.121	0.126
W91	Woodland	0.404	0.336	0.46

Appendix H: Growth curves

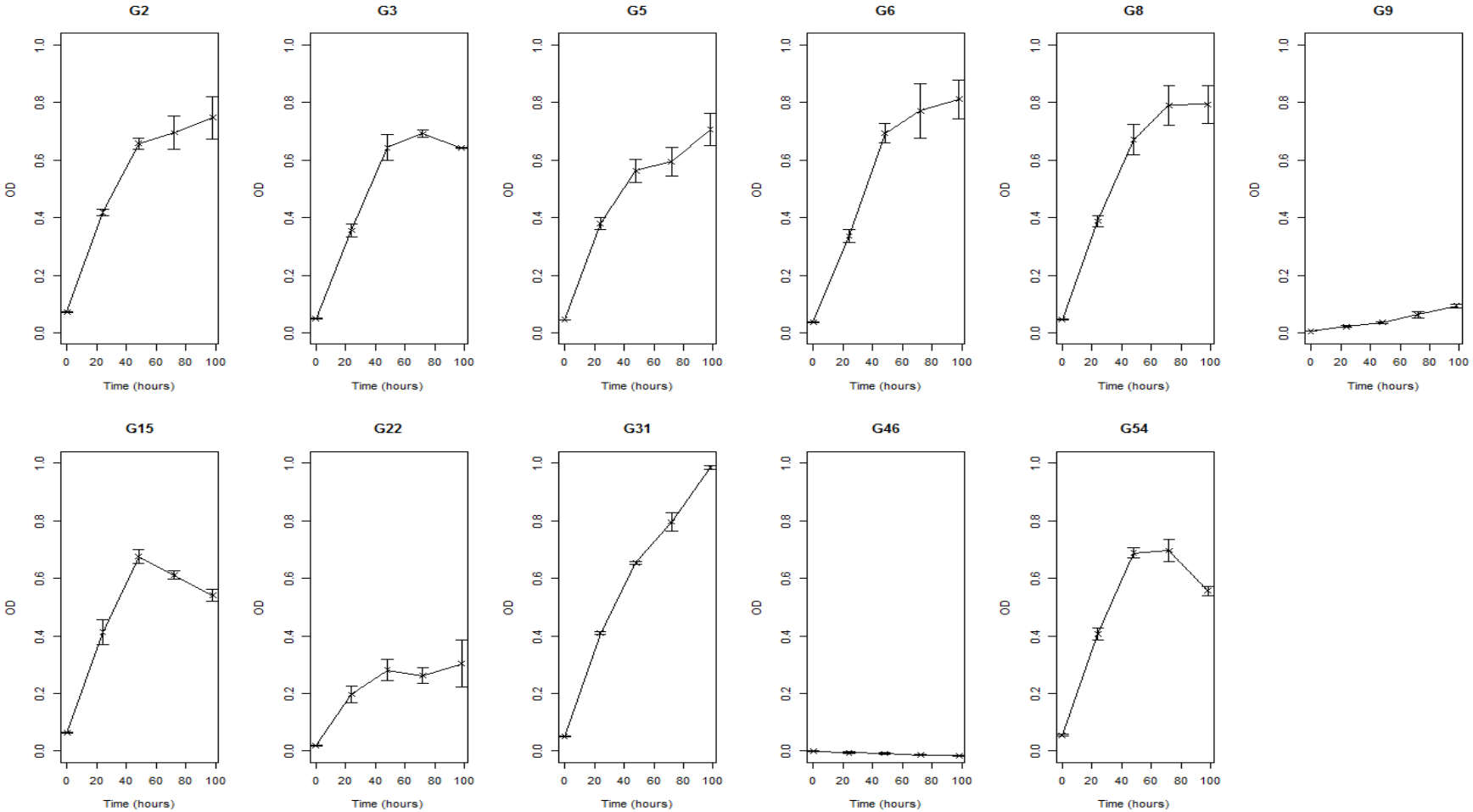


Figure H-1: Growth curves for grassland isolates

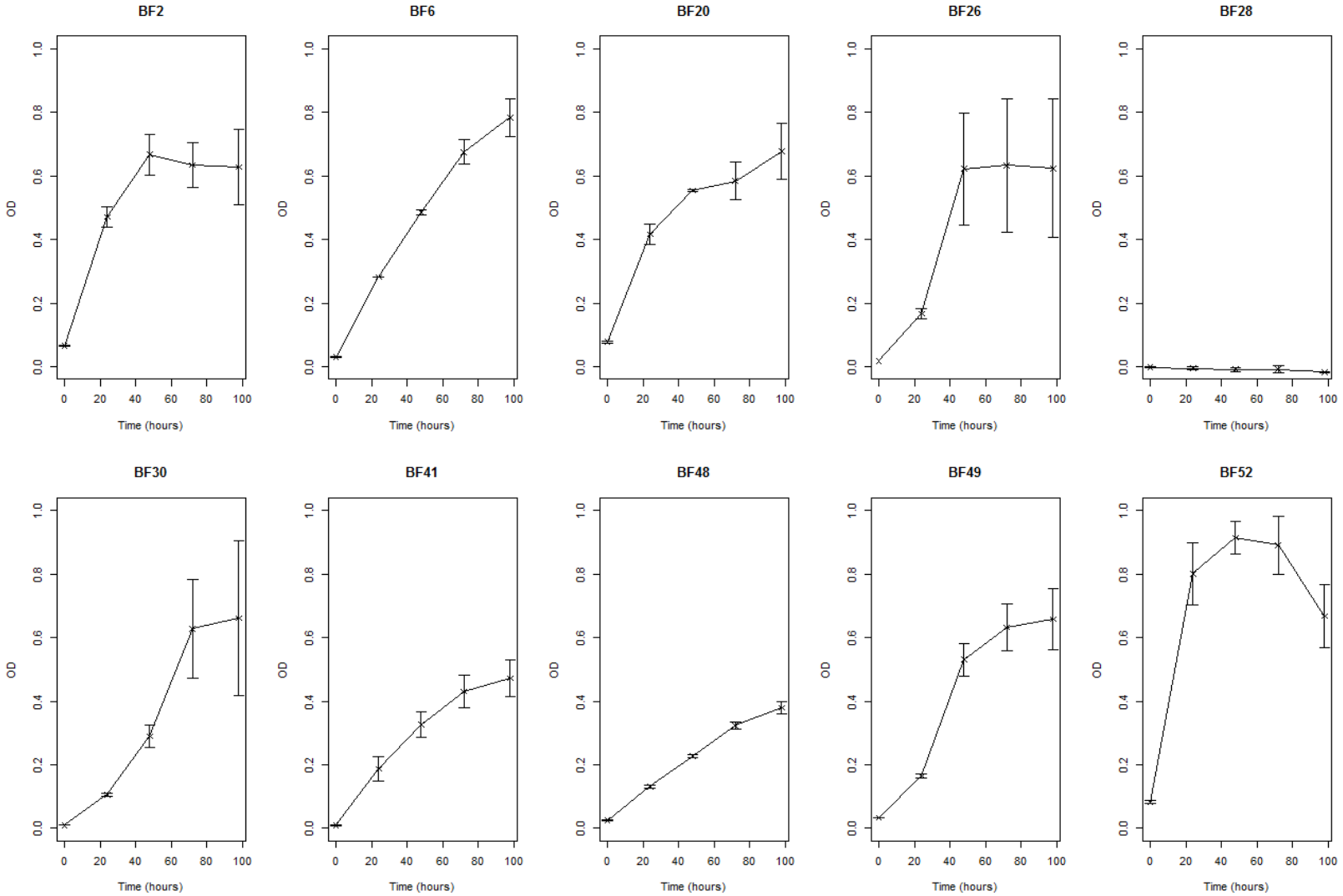


Figure H-2: Growth curves of bare fallow isolates

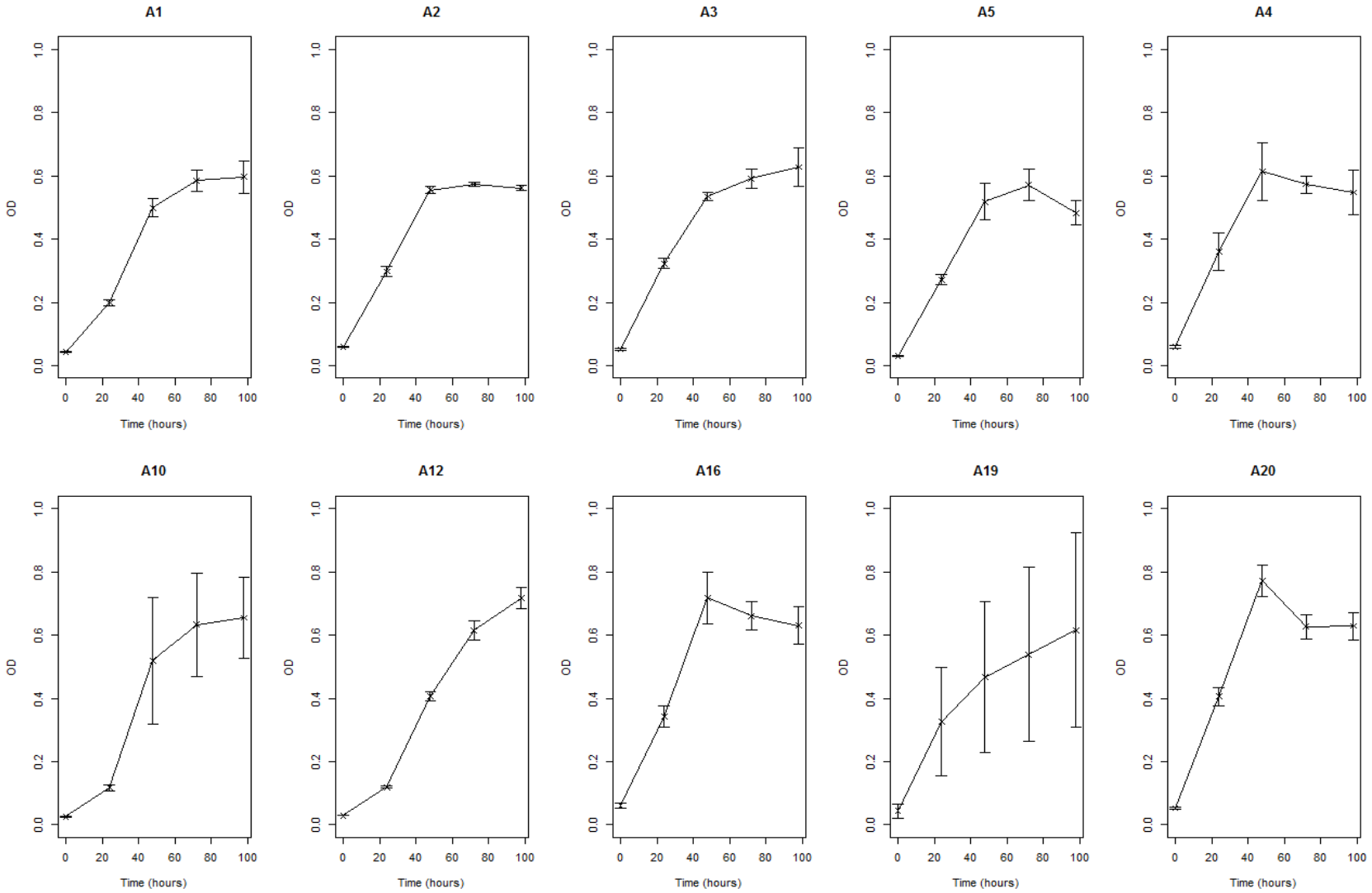


Figure H-3: Growth curves of arable isolates

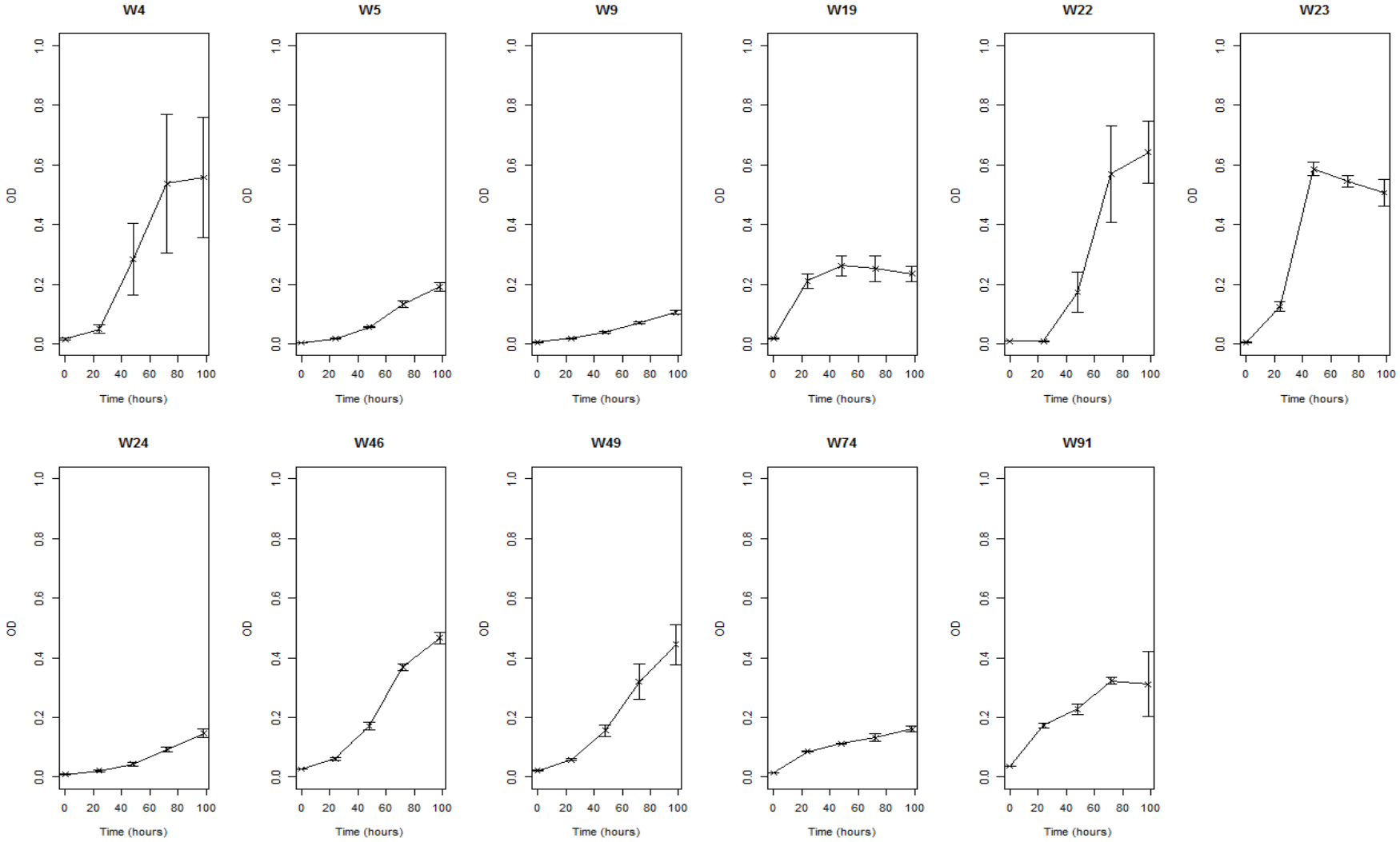


Figure H-4: Growth curves for woodland isolates

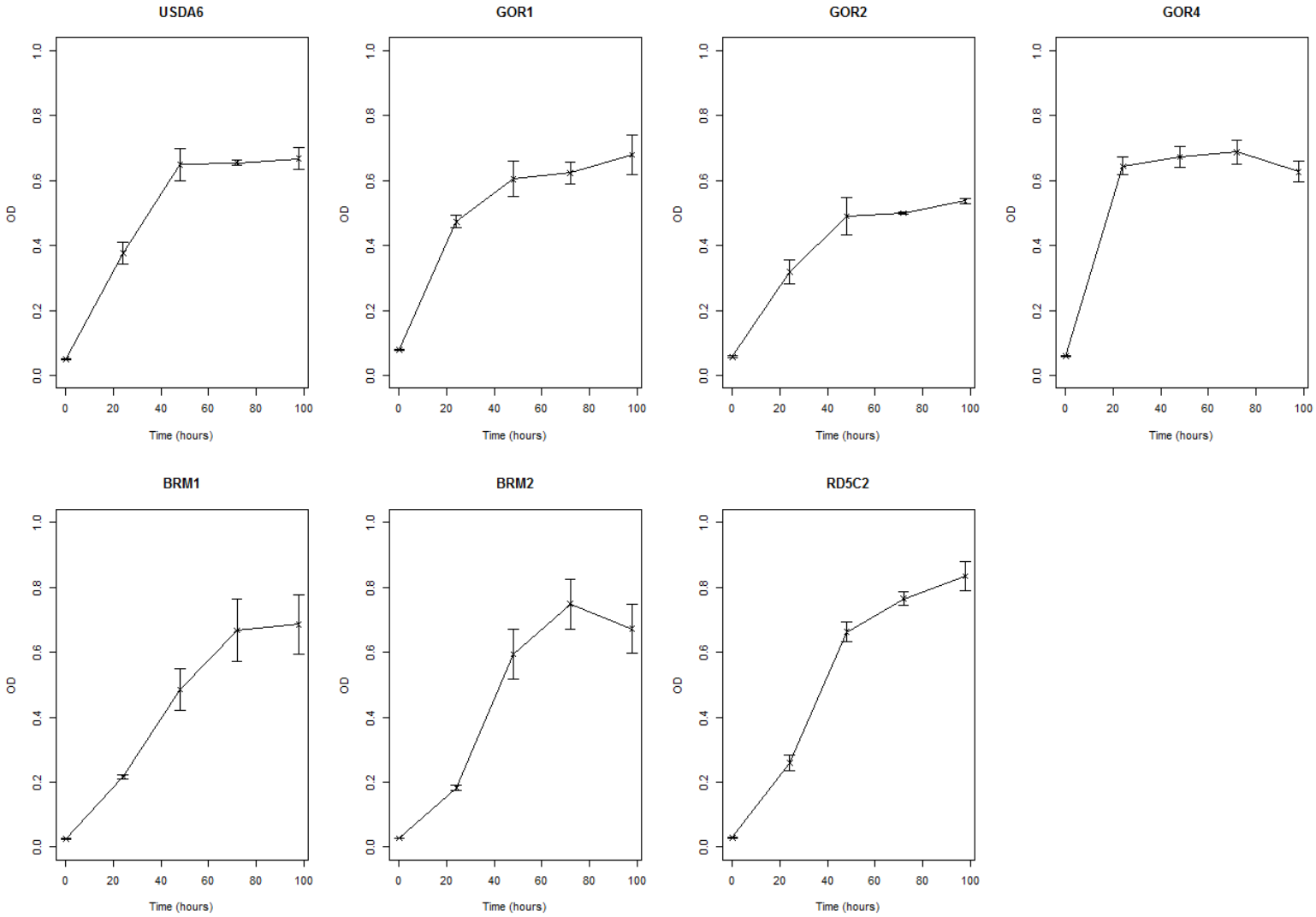


Figure H-5: Growth curves of symbiotic and reference isolates

Appendix I: Tukey's HSD results for growth rate

Table I-1: Tukey's HSD output comparing growth rate for each soil treatment

Comparison	diff	lwr	upr	p adj
Bare fallow-Arable	0.002384624	-0.06945773	0.07422698	0.9999809
Grassland-Arable	-0.006645095	-0.07623477	0.06294458	0.998733
Symbiotic-Arable	0.022048751	-0.05259006	0.09668757	0.9152069
Woodland-Arable	-0.087566672	-0.15374299	-0.02139036	0.0044285
Grassland-Bare fallow	-0.009029719	-0.08262458	0.06456514	0.9966304
Symbiotic-Bare fallow	0.019664127	-0.05872231	0.09805056	0.9514875
Woodland-Bare fallow	-0.089951296	-0.16032732	-0.01957527	0.006365
Symbiotic-Grassland	0.028693846	-0.0476333	0.10502099	0.8188006
Woodland-Grassland	-0.080921577	-0.14899643	-0.01284673	0.0127998
Woodland-Symbiotic	-0.109615423	-0.18284393	-0.03638692	0.0010366

Table I-2: Tukey's HSD output comparing maximum OD for each soil treatment

Comparison	diff	lwr	upr	p adj
Bare fallow-Arable	0.02196667	-0.2095441	0.25347746	0.9987643
Grassland-Arable	-0.02897778	-0.2532294	0.19527381	0.9958825
Symbiotic-Arable	0.04365714	-0.1968652	0.28417949	0.9850029
Woodland-Arable	-0.28947273	-0.5027248	-0.07622064	0.0033491
Grassland-Bare fallow	-0.05094444	-0.2881027	0.18621379	0.9721094
Symbiotic-Bare fallow	0.02169048	-0.2309085	0.27428946	0.9991655
Woodland-Bare fallow	-0.31143939	-0.538225	-0.08465381	0.0029387
Symbiotic-Grassland	0.07263492	-0.173328	0.31859788	0.9152999
Woodland-Grassland	-0.26049495	-0.479865	-0.04112487	0.0129223
Woodland-Symbiotic	-0.33312987	-0.5691075	-0.09715222	0.0021345

Appendix J: Loadings matrices from PCA

Table J-1: Loadings matrix for all time points

Substrate	PC1	PC2	PC3	PC4
4	-0.09444	0.059451	-0.03477	0.060646
5	-0.1261	-0.06345	0.04563	0.279871
6	-0.15345	-0.08551	-0.02725	-0.03288
10	-0.1359	0.067664	-0.17573	0.18627
12	-0.11604	-0.18242	-0.13247	-0.15822
13	-0.10654	-0.10956	0.185748	-0.21508
14	-0.1661	-0.05971	-0.01357	0.059924
18	-0.15754	-0.08459	-0.00681	-0.14093
20	-0.15788	-0.0948	-0.03096	0.041345
22	-0.15493	-0.08711	-0.01813	-0.15204
23	-0.15415	0.037206	0.034383	0.095237
26	-0.16426	-0.06044	0.034046	-0.05703
28	-0.15083	0.026012	0.113466	0.023246
29	-0.16155	-0.01808	0.033953	-0.08813
30	-0.07265	0.216757	-0.08446	0.140997
33	-0.06526	0.238843	-0.24601	-0.17778
35	-0.128	-0.17643	-0.00639	-0.22142
36	-0.15877	-0.12527	0.020967	0.095104
37	-0.14144	0.017083	-0.11835	-0.0885
38	-0.11374	0.237576	-0.08598	-0.04672
39	-0.11465	-0.12113	0.114348	0.057626
41	-0.11266	0.107233	-0.10199	-0.0646
42	-0.16657	0.002525	-0.0225	-0.02884
43	-0.04	0.298899	-0.01041	0.114884
44	-0.16367	-0.04447	0.02396	-0.05919
45	-0.15664	0.066954	0.023176	-0.03474
46	-0.06873	0.334324	-0.00756	-0.01375
47	-0.0476	0.240357	0.259718	0.058065
52	-0.15357	-0.02615	-0.00631	0.189008
53	-0.07702	-0.20962	-0.05288	0.251103
54	-0.0944	0.074064	0.22452	-0.2789
55	-0.15838	0.010048	0.142911	0.071972
60	-0.14967	-0.13896	0.004221	0.121476
61	-0.10598	0.235041	-0.15859	-0.06718
62	-0.15724	-0.00567	-0.08676	-0.03135
65	-0.08686	-0.02926	-0.2631	-0.20841
68	-0.11615	0.077005	0.240062	0.121548
69	-0.13885	-0.09046	-0.05986	0.127185
70	-0.1622	-0.00579	0.056143	0.00026
72	-0.15714	0.051584	-0.00591	-0.08964

Substrate	PC1	PC2	PC3	PC4
73	-0.16583	-0.02989	-0.03273	0.035049
74	-0.0039	0.011689	0.375682	-0.18987
76	-0.09308	0.215493	-0.15415	-0.10029
77	-0.1628	-0.08734	0.002797	0.053486
78	-0.12103	0.243286	-0.08969	0.049401
79	-0.04197	0.314439	0.078804	-0.13113
81	-0.16061	-0.00042	-0.0576	0.061404
82	-0.15986	0.015521	-0.03798	-0.11254
83	-0.16588	-0.03002	0.011376	0.073432
84	-0.16362	-0.0177	-0.02556	0.098367
85	-0.1103	0.073941	0.100489	0.324057
86	-0.06255	0.15618	0.319603	0.185778
90	-0.15136	-0.0597	-0.03834	-0.10352
91	-0.16446	0.012547	-0.05992	0.046074
92	-0.13412	-0.00796	0.130945	-0.15158
93	-0.1559	-0.10086	0.005024	-0.07931
94	-0.06002	0.043956	0.39071	-0.15895

Table J-2: Loadings matrix for 98 hours

Substrate	PC1	PC2	PC3	PC4
4	0.102915	-0.08703	-0.00588	-1.48E-05
5	0.122317	0.021575	0.158349	0.284132
6	0.159493	0.078455	-0.0028	-0.03838
10	0.128822	-0.06893	-0.04638	0.271431
12	0.114053	0.213504	-0.15413	0.011647
13	0.101493	0.161122	0.053105	-0.24005
14	0.165721	0.036593	0.030127	0.064049
18	0.15816	0.102983	-0.05418	-0.09004
20	0.161667	0.059462	0.002073	0.06456
22	0.155829	0.102456	-0.09512	-0.0725
23	0.152852	-0.07773	0.061857	0.060598
26	0.161204	0.082056	-0.00353	-0.08629
28	0.156045	-0.0332	0.035503	-0.05796
29	0.161554	0.018007	-0.02086	-0.08548
30	0.04971	-0.23662	-0.07556	0.160117
33	0.061469	-0.13807	-0.37821	-0.05083
35	0.123211	0.228887	-0.06098	-0.12533
36	0.160577	0.077687	0.106354	0.082405
37	0.141138	0.043787	-0.10793	0.004922
38	0.113164	-0.20878	-0.18752	0.006528
39	0.106831	0.083979	0.20599	0.054219
41	0.105985	-0.06933	-0.14813	-0.00826
42	0.167069	0.005195	-0.00981	-0.05649
43	0.034418	-0.29447	0.047933	0.017623
44	0.163952	0.035153	-0.01893	-0.04381
45	0.150806	-0.06772	-0.04242	-0.09432
46	0.06728	-0.31848	-0.09517	-0.05863
47	0.047442	-0.30544	0.146422	-0.05364
52	0.15031	-0.03374	0.121567	0.145826
53	0.063964	0.15895	0.165308	0.317262
54	0.136836	0.019474	-0.02495	-0.24527
55	0.157569	-0.04122	0.155765	-0.03799
60	0.150583	0.090411	0.116582	0.125757
61	0.104554	-0.16482	-0.25275	0.007386
62	0.15861	0.035874	-0.0565	0.032255
65	0.082012	0.126681	-0.32254	0.061533
68	0.114909	-0.15354	0.203921	-0.05241
69	0.133152	0.054406	-0.00245	0.190526
70	0.163524	-0.0067	0.034512	-0.04604
72	0.161962	0.008029	-0.05299	-0.09993
73	0.168036	0.009616	-0.00319	0.051873
74	0.009628	-0.01305	0.251538	-0.36526
76	0.089864	-0.15721	-0.21305	-0.04567

Substrate	PC1	PC2	PC3	PC4
77	0.164037	0.065531	0.04041	0.052835
78	0.123552	-0.23292	-0.11668	0.039366
79	0.051942	-0.2853	0.005602	-0.20256
81	0.156359	-0.00554	-0.00825	0.108241
82	0.164552	0.020023	-0.07935	-0.04725
83	0.165262	0.009308	0.064334	0.031866
84	0.161846	-0.00186	0.040009	0.086445
85	0.089303	-0.18423	0.176623	0.236115
86	0.06937	-0.23913	0.259371	0.025038
90	0.149642	0.122151	-0.02668	-0.03979
91	0.161774	-0.00614	-0.00996	0.069227
92	0.13827	0.032565	0.062284	-0.20301
93	0.160036	0.087529	0.000769	-0.10945
94	0.079411	-0.05623	0.284976	-0.302

Appendix K: Biolog utilisation curves

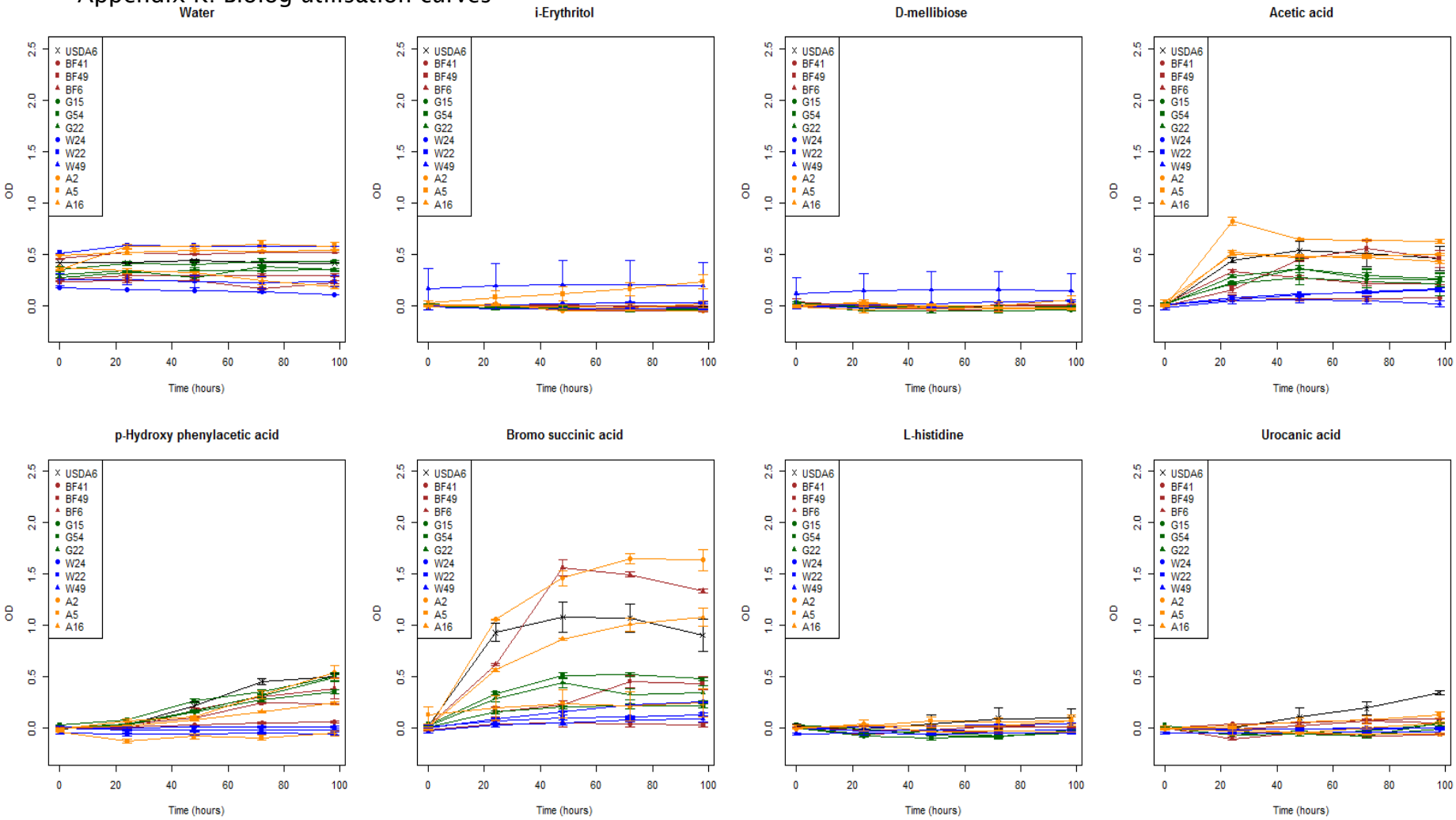


Figure K-1: Biolog utilisation curves for substrates 1-8.

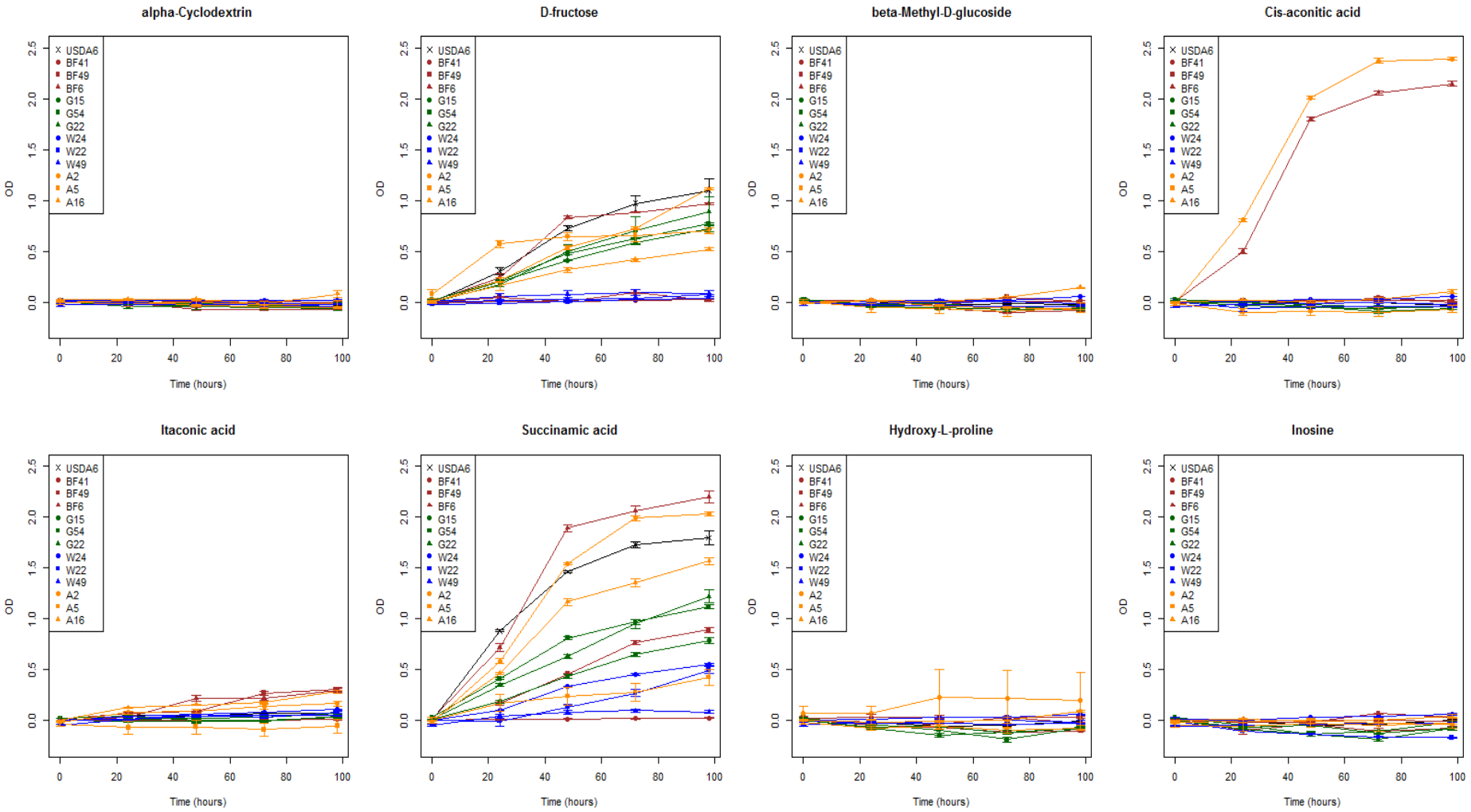


Figure K-2: Biolog utilisation curves for substrates 9-16.

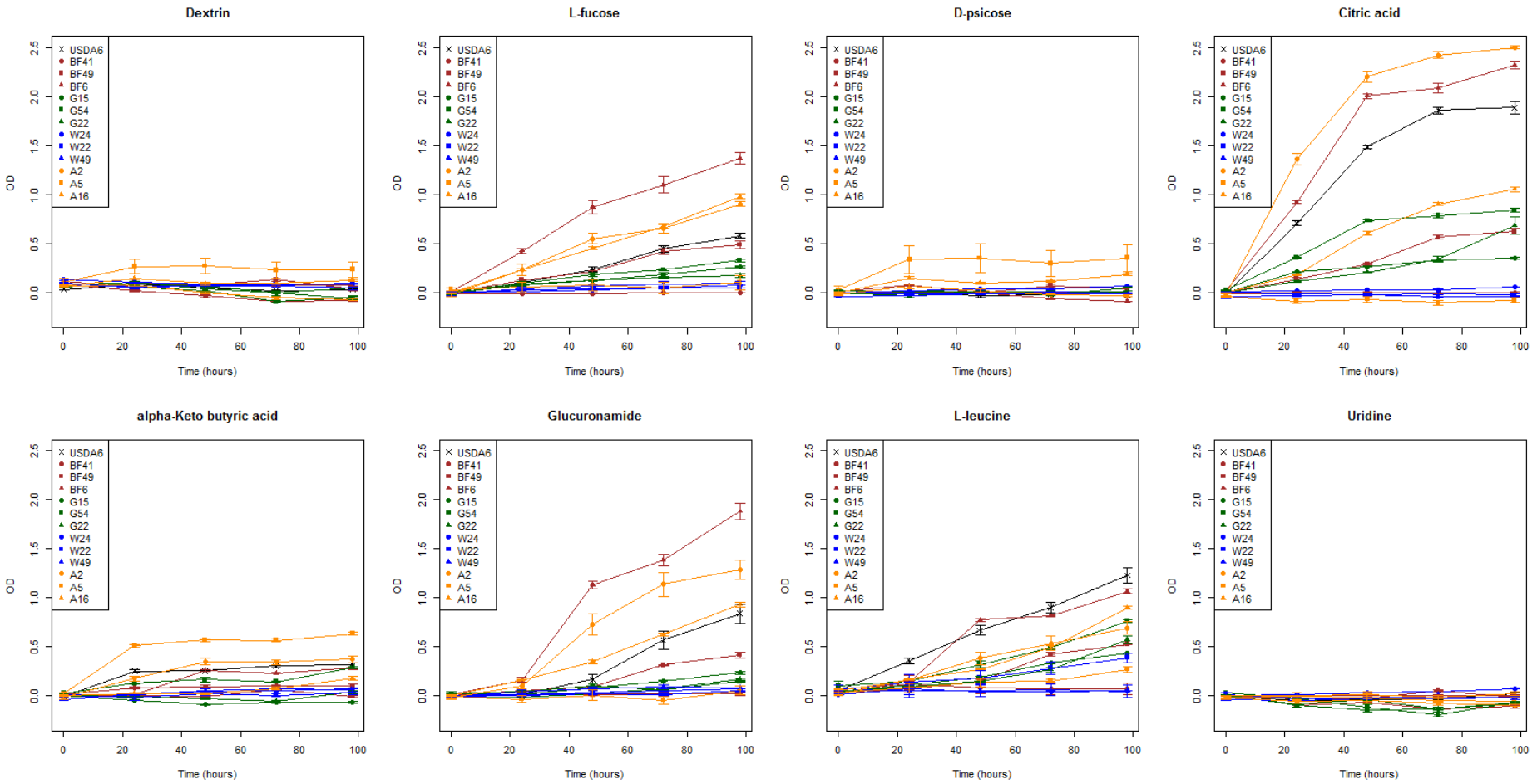


Figure K-3: Biolog utilisation curves for substrates 17-24

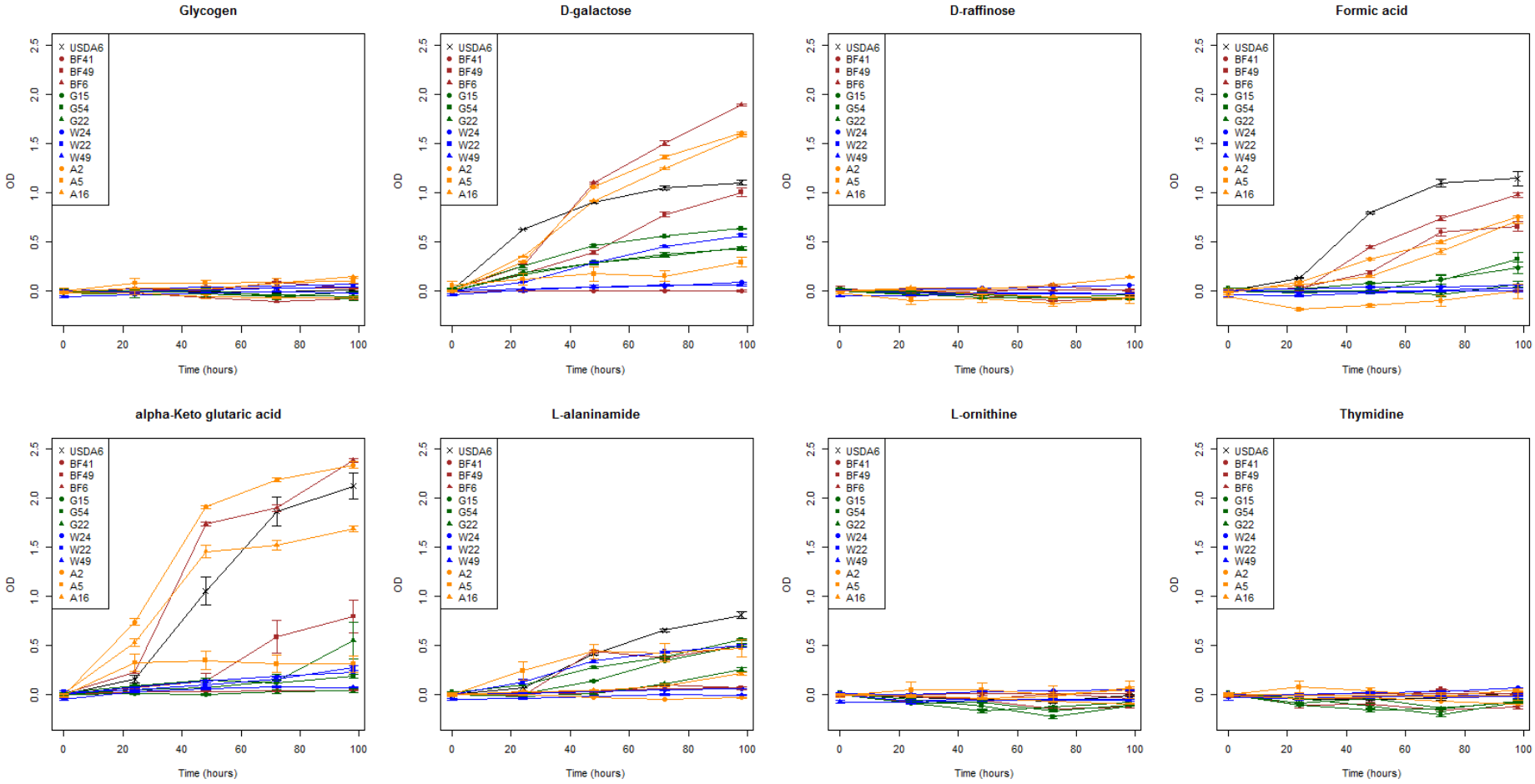


Figure K-4: Biolog utilisation curves for substrates 25-32.

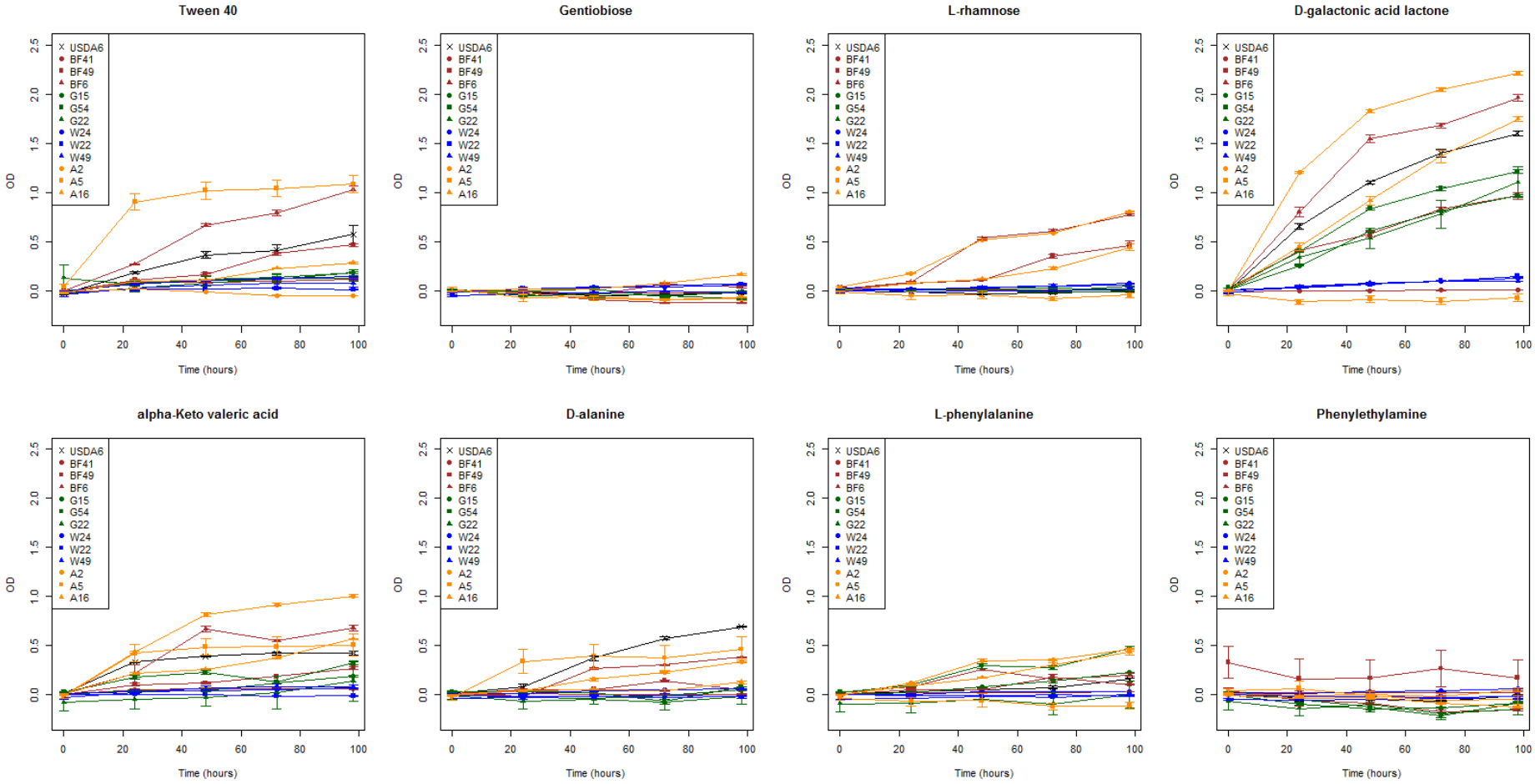


Figure K-5: Biolog utilisation curves for substrates 33-40

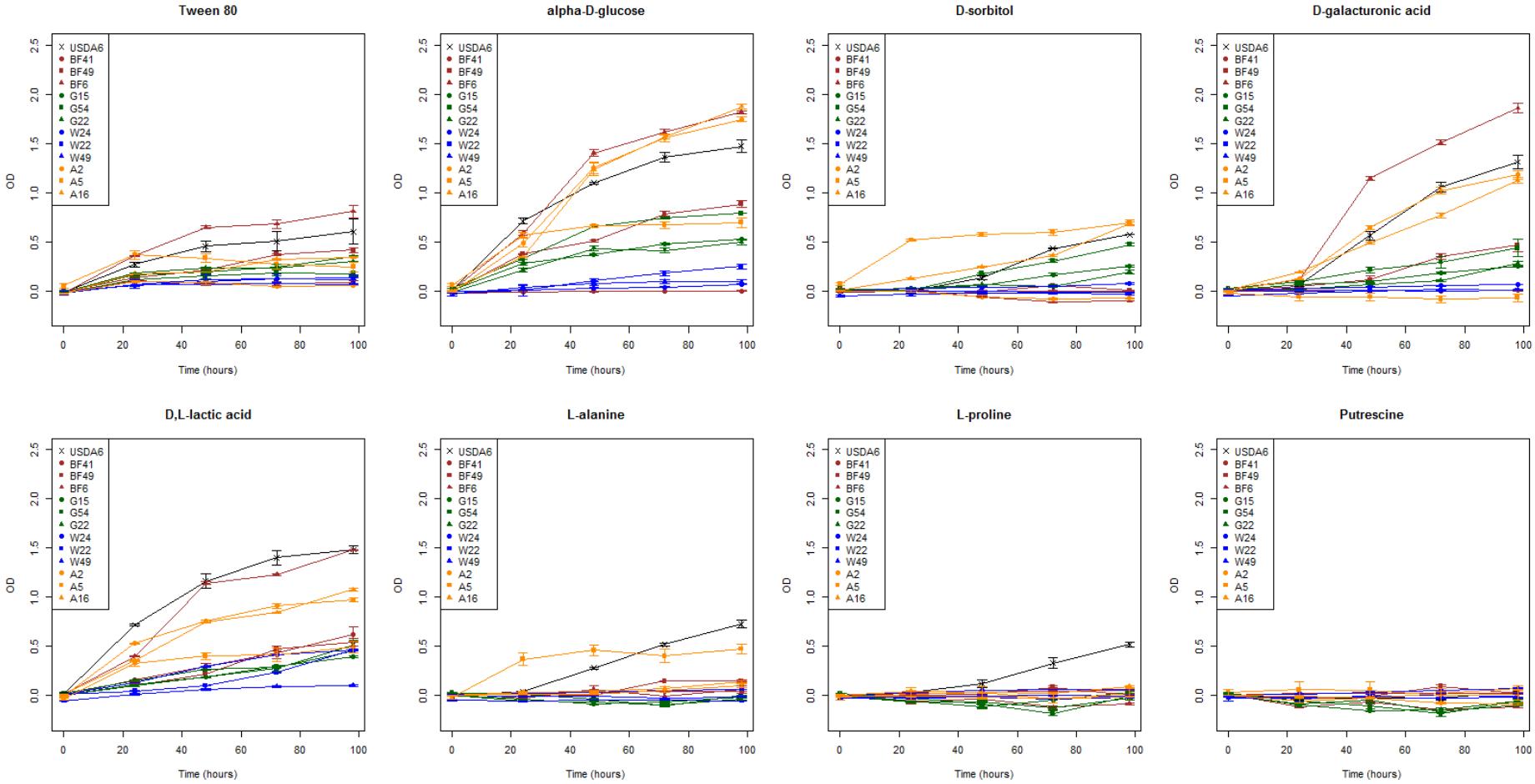


Figure K-6: Biolog utilisation curves for substrates 41-48

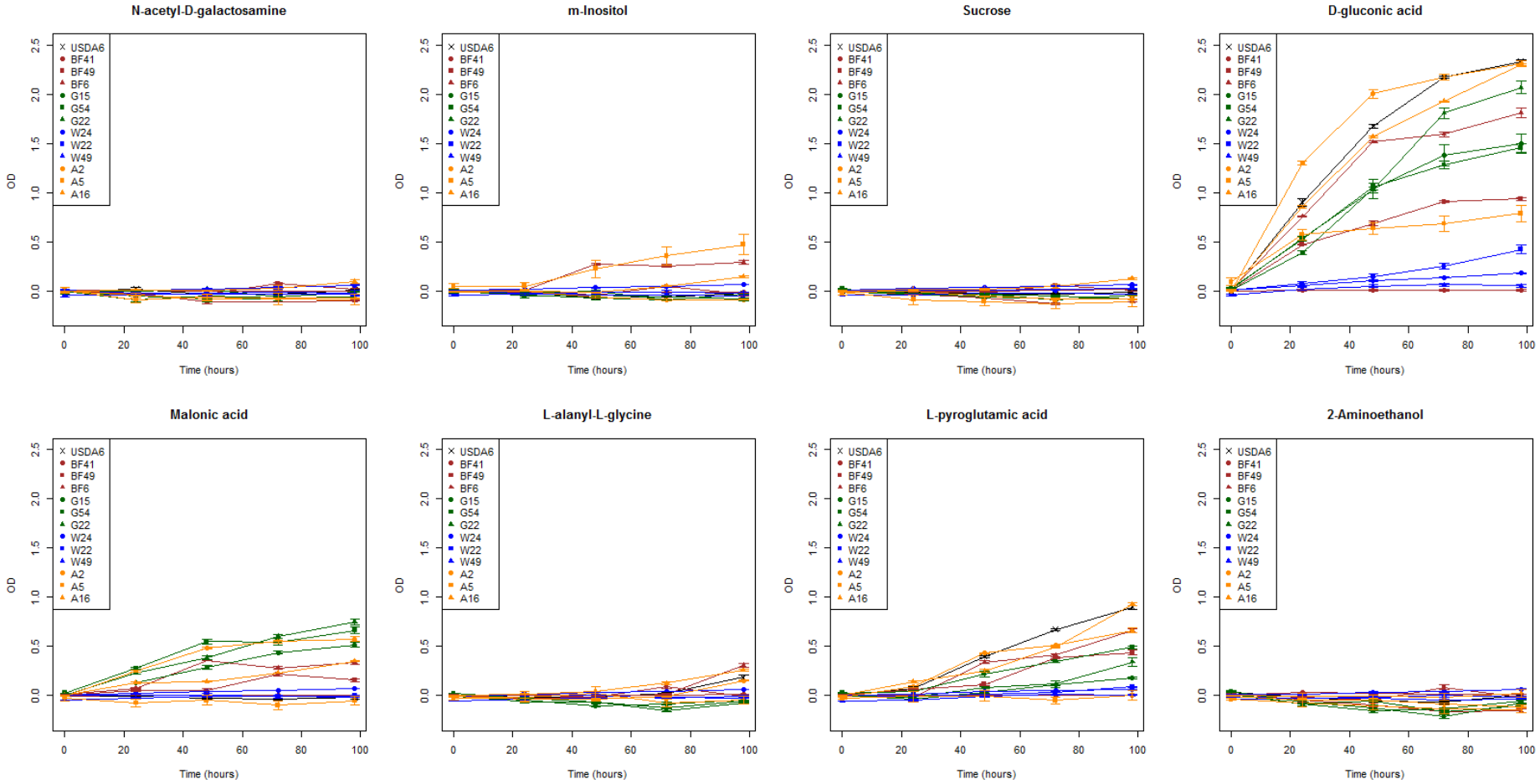


Figure K-7: Biolog utilisation curves for substrates 49-56

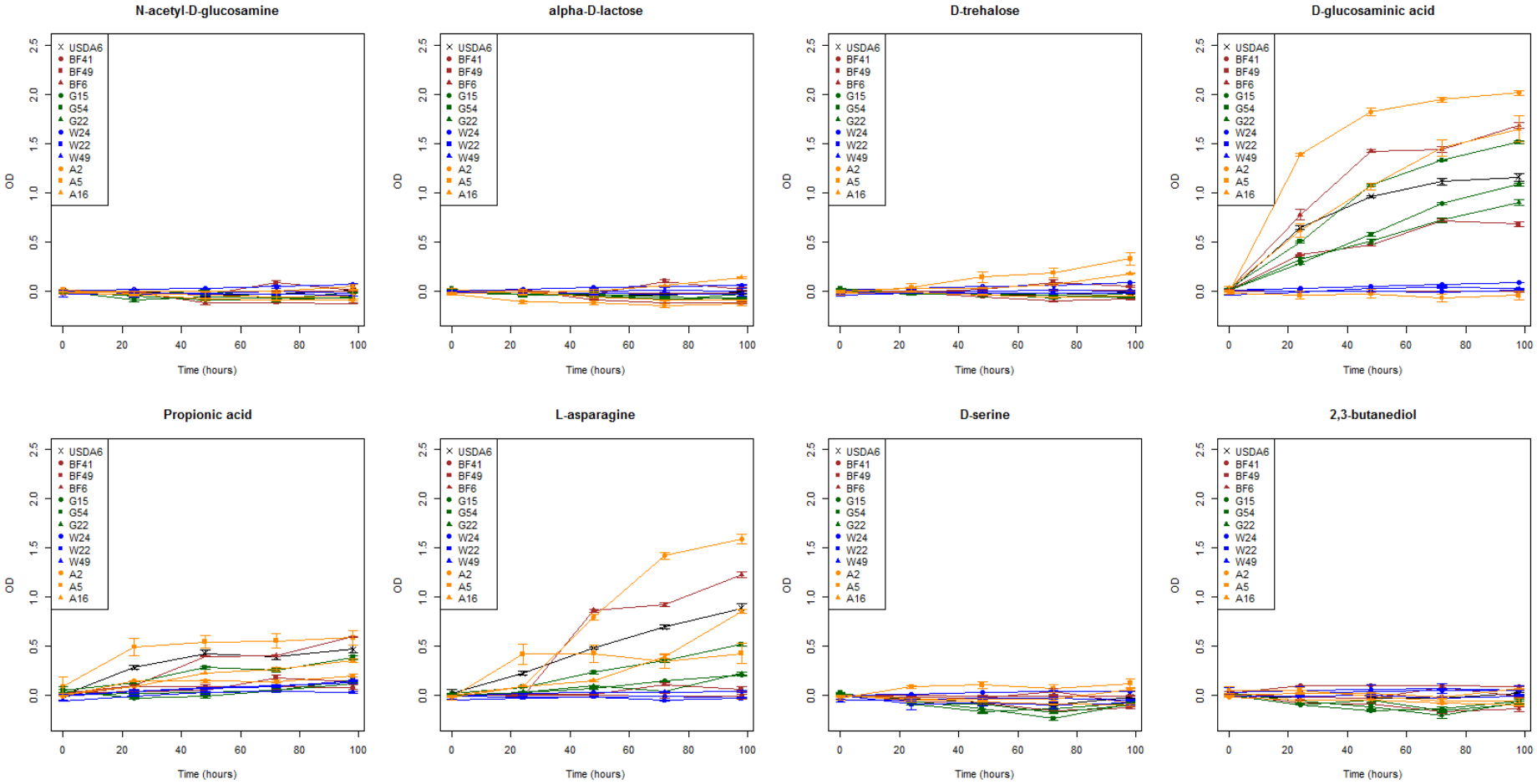


Figure K-8: Biolog utilisation curves for substrates 57-64

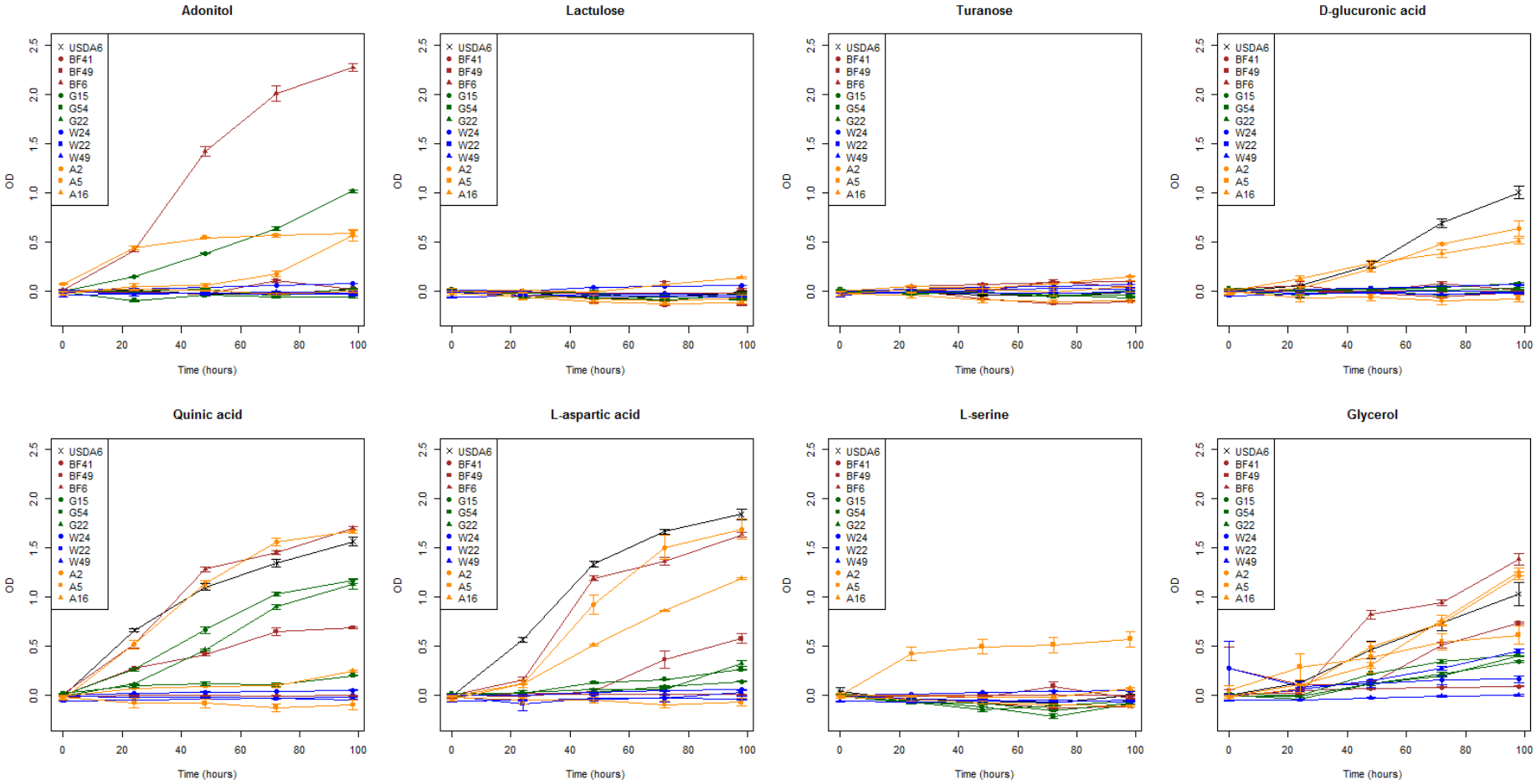


Figure K-9: Biolog utilisation curves for substrates 65-72

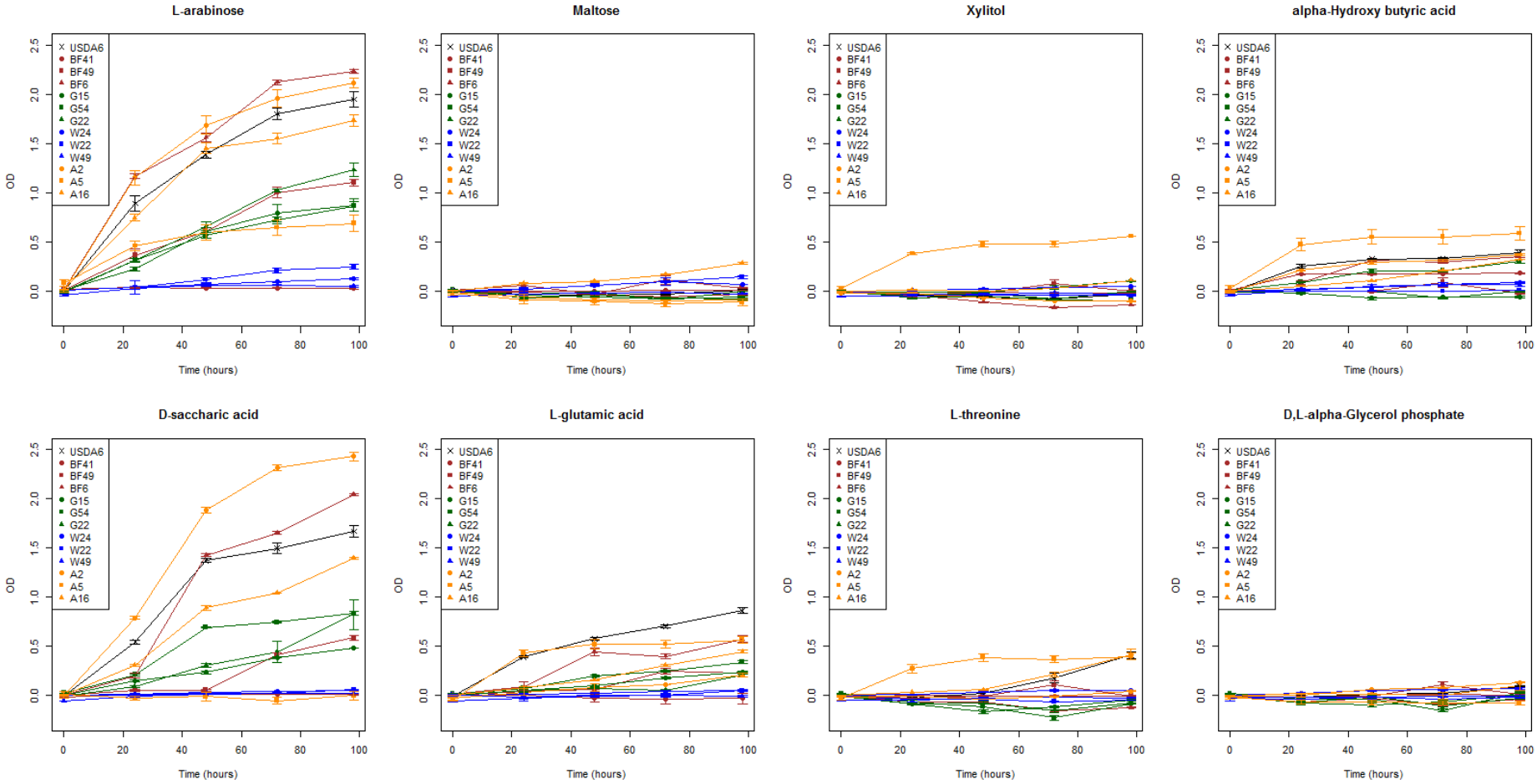


Figure K-10: Biolog utilisation curves for substrates 73-80

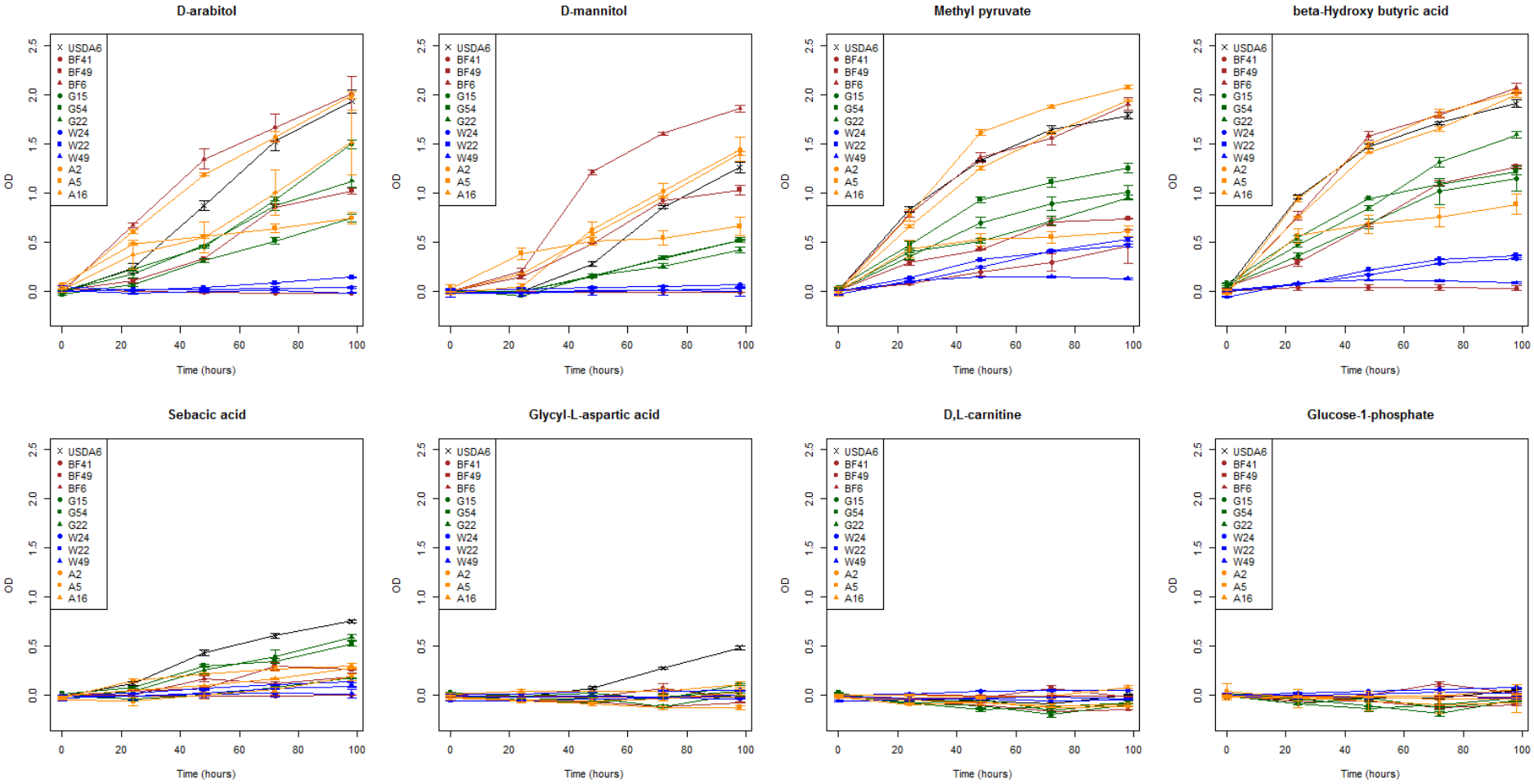


Figure K-11: Biolog utilisation curves for substrates 81-88

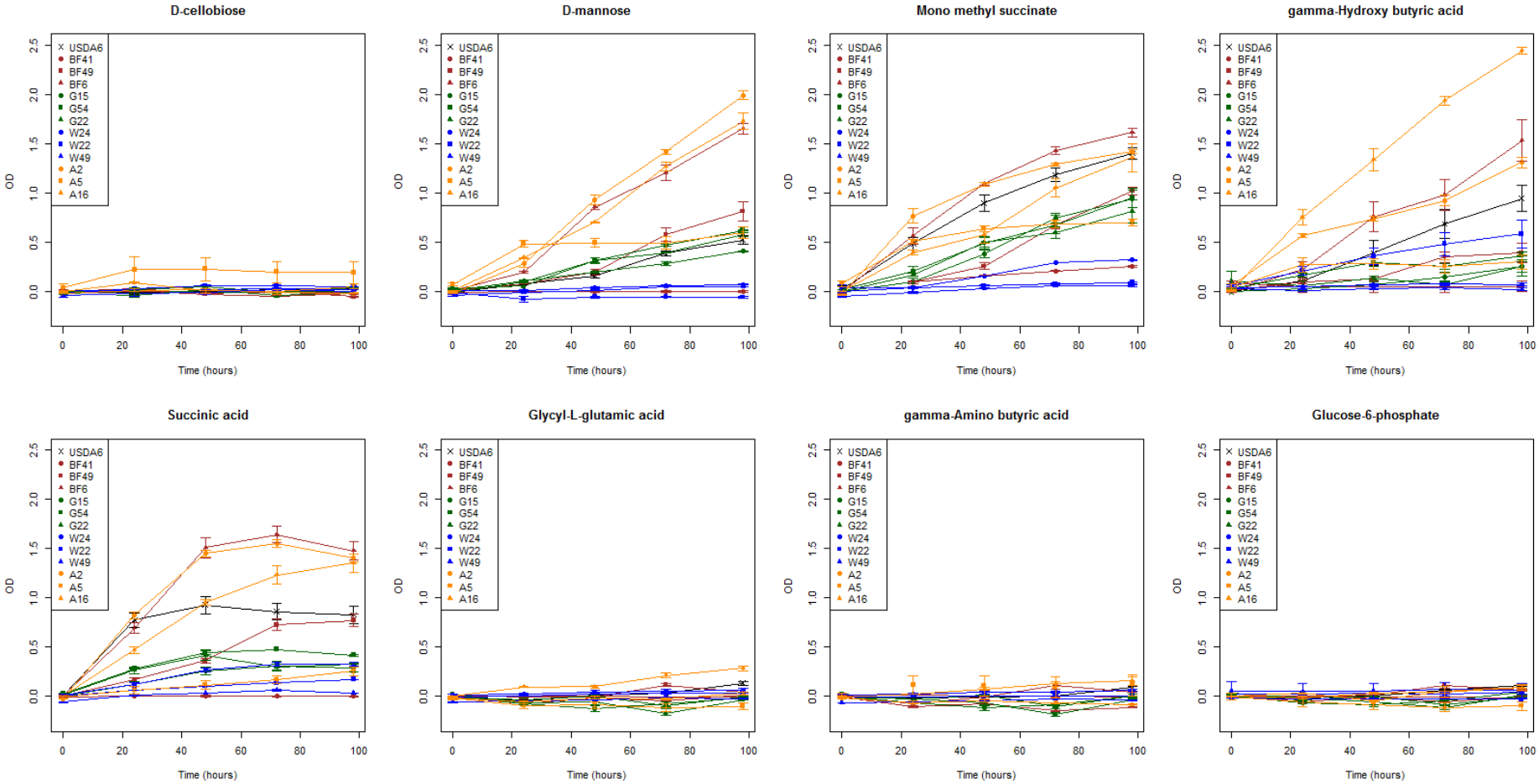


Figure K-12: Biolog utilisation curves for substrates 89-96

Appendix L: PCA biplots for 0h, 24h, 48h and 72h

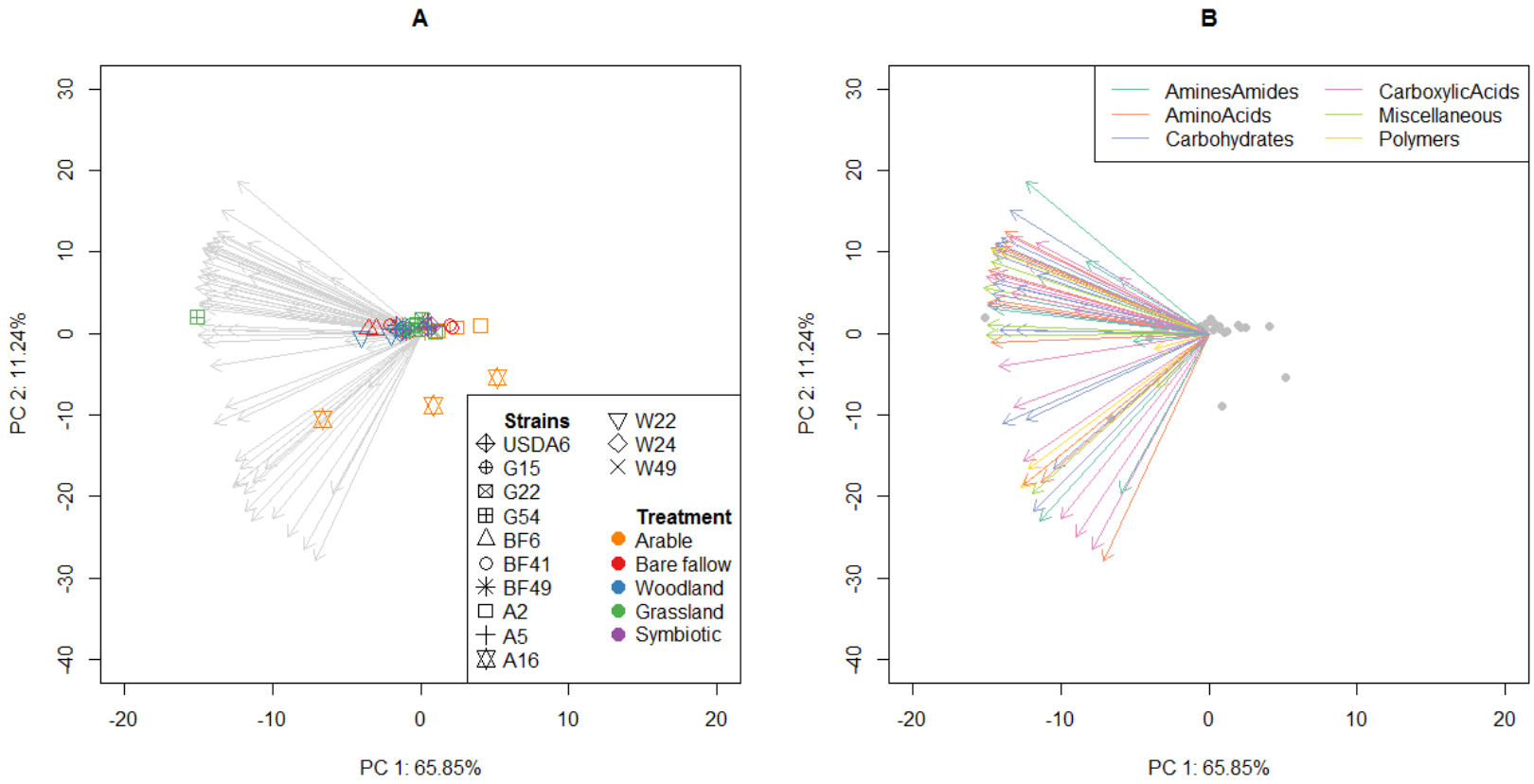


Figure L-1: Principal components analysis biplot showing PC1 and PC2 accounting for 77.09% of the variation for time 0 hours. A shows isolates colour coded according to soil treatment and B shows the 95 substrates classified into 6 guilds.

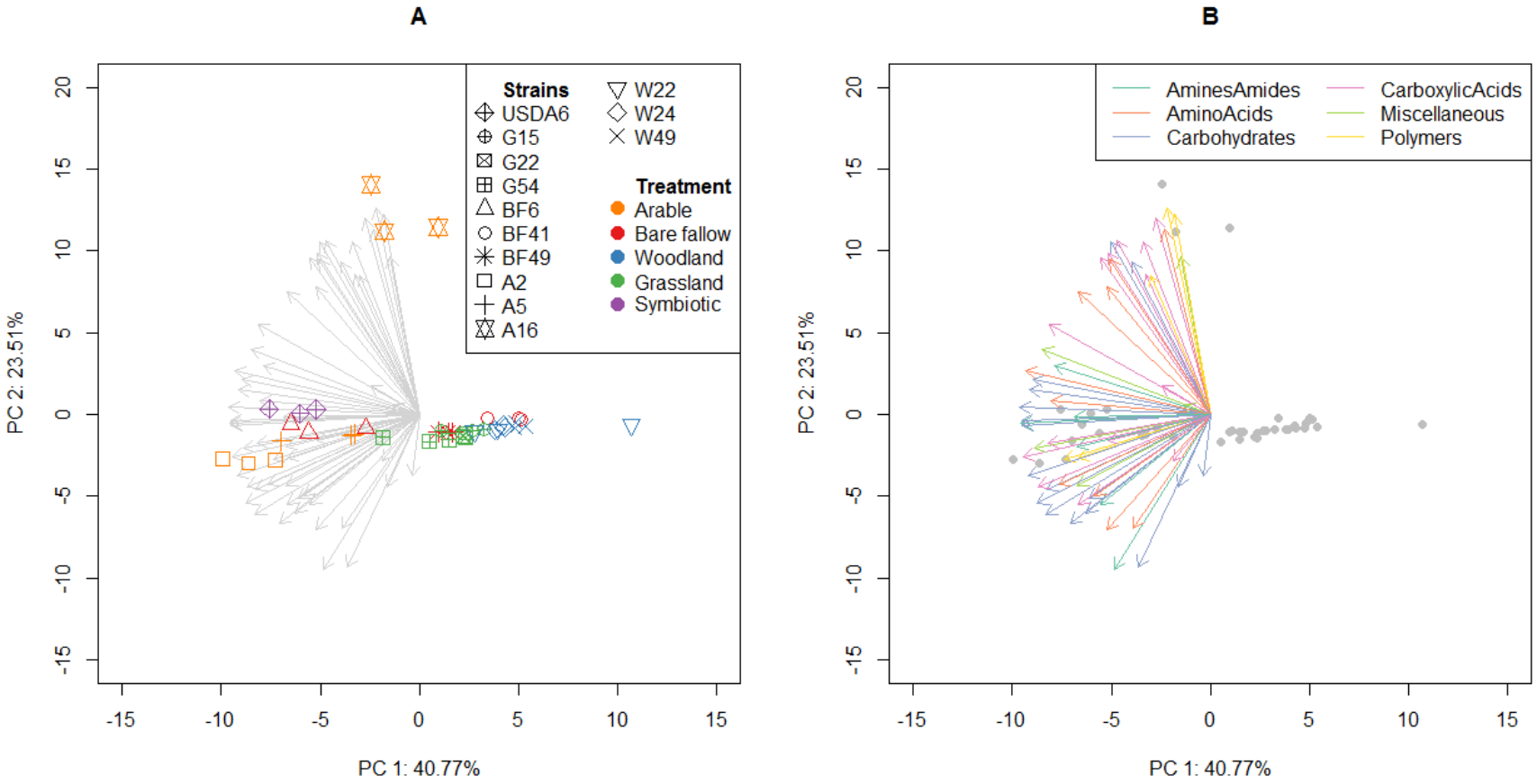


Figure L-2: Principal components analysis biplot showing PC1 and PC2 accounting for 64.28% of the variation for time 24 hours. A shows isolates colour coded according to soil treatment and B shows the 95 substrates classified into 6 guilds.

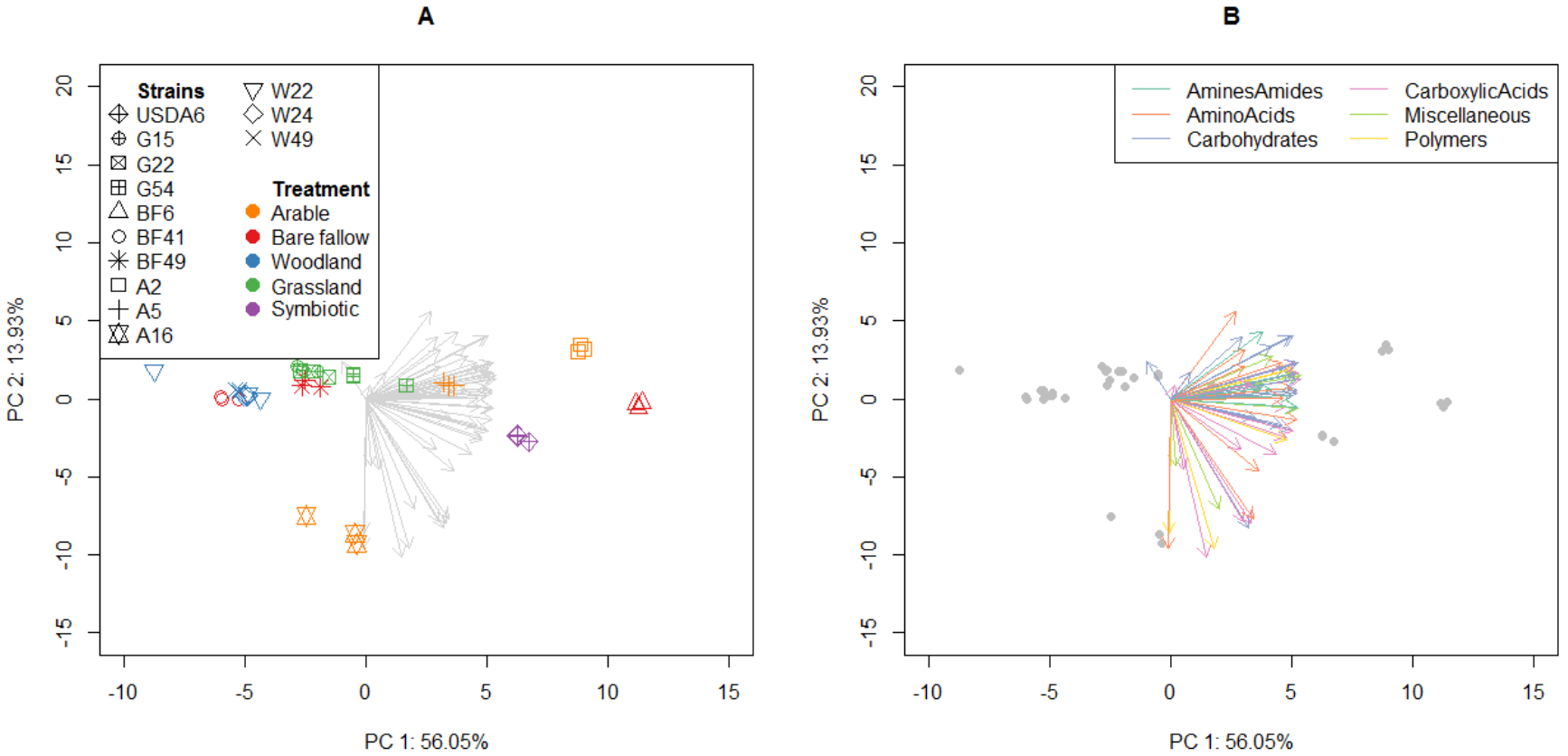


Figure L-3: Principal components analysis biplot showing PC1 and PC2 accounting for 69.98% of the variation for time 48 hours. A shows isolates colour coded according to soil treatment and B shows the 95 substrates classified into 6 guilds.

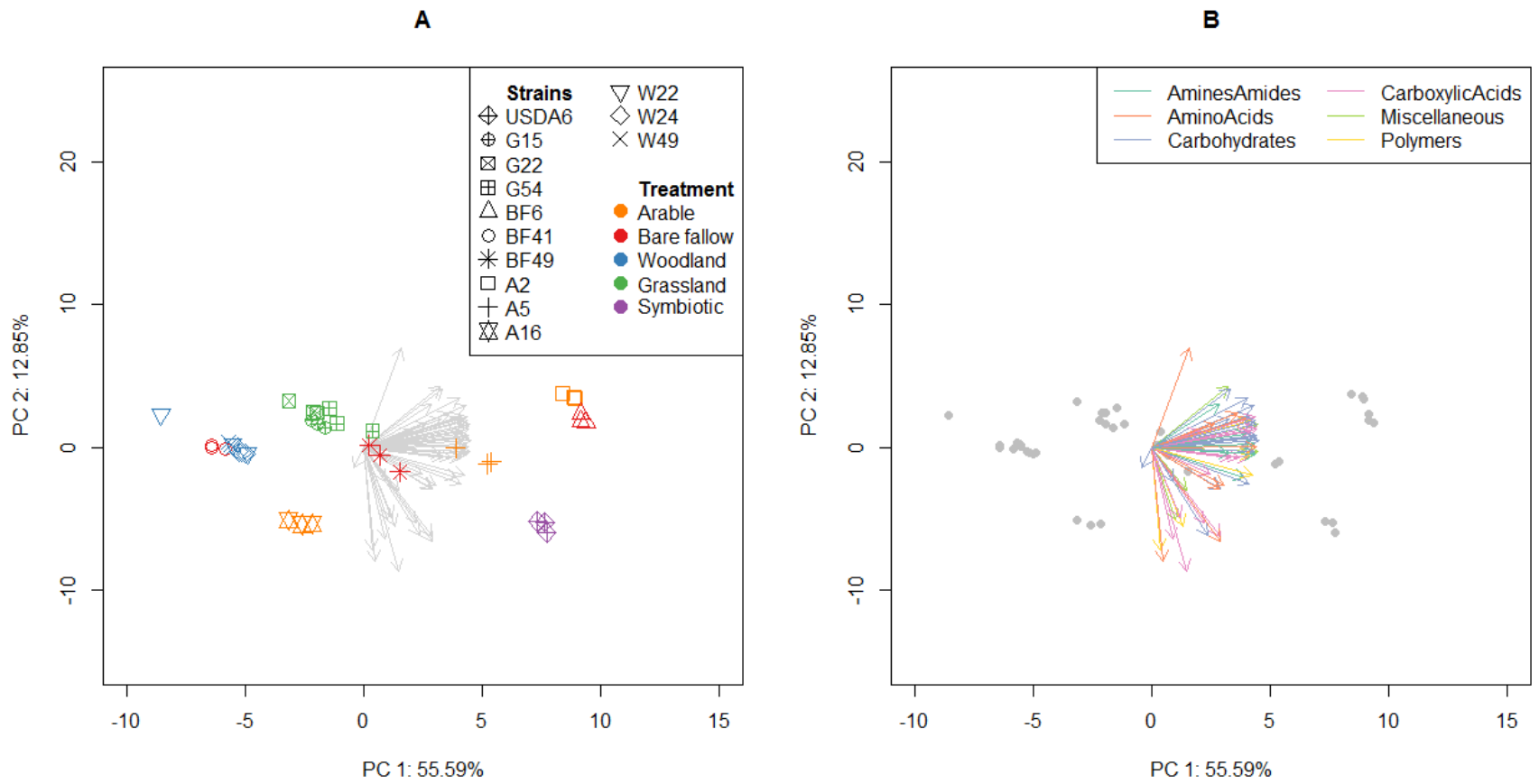


Figure L-4: Principal components analysis biplot showing PC1 and PC2 accounting for 68.44% of the variation for time 72 hours. A shows isolates colour coded according to soil treatment and B shows the 95 substrates classified into 6 guilds.

Appendix M: Binary matrix for time point 98 hours averaged across replicate plates

Table M-1: Binary matrix for time point 98 hours, 1 indicates substrates utilised to an OD of twice the water well OD

Substrate	A16	A5	A2	BF6	BF41	BF49	G15	G22	G54	W22	W24	W49	USDA 6
2	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	1	1	0	0	1	0	0	0	0	1	0	0
5	0	1	1	0	0	1	1	1	1	0	0	0	1
6	0	1	1	1	0	1	0	0	1	0	1	0	1
7	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	0	0	0
10	1	1	1	1	0	0	1	1	1	0	0	0	1
11	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	1	1	0	0	0	0	0	0	0	0	0
13	0	1	0	0	0	1	0	0	0	0	0	0	0
14	0	1	1	1	0	1	1	1	1	0	1	0	1
15	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	1	1	1	0	1	0	0	0	0	0	0	1
19	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	1	1	1	0	1	0	1	1	0	0	0	1
21	1	0	0	0	0	0	0	0	0	0	0	0	0
22	0	1	1	1	0	1	0	0	0	0	0	0	1
23	0	1	1	1	0	1	1	1	1	0	0	0	1
24	0	0	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0	0	0
26	0	1	1	1	0	1	1	1	1	0	0	0	1
27	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	1	1	1	0	1	0	0	0	0	0	0	1
29	0	1	1	1	0	1	0	1	0	0	1	0	1
30	0	1	0	0	0	0	1	0	1	0	0	0	1
31	0	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0
33	1	1	0	1	0	1	0	0	0	0	1	0	1
34	0	0	0	0	0	0	0	0	0	0	0	0	0

Substrate	A16	A5	A2	BF6	BF41	BF49	G15	G22	G54	W22	W24	W49	USDA 6
35	0	1	1	1	0	1	0	0	0	0	0	0	0
36	0	1	1	1	0	1	1	1	1	0	1	0	1
37	0	1	1	1	0	1	0	0	0	0	0	0	1
38	0	0	0	0	0	0	0	0	0	0	0	0	1
39	0	1	0	0	0	0	0	0	0	0	0	0	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	1	0	1	0	1	0	0	0	0	1	0	1
42	1	1	1	1	0	1	1	1	1	0	0	0	1
43	1	1	0	0	0	0	0	0	1	0	0	0	1
44	0	1	1	1	0	1	0	0	1	0	0	0	1
45	0	1	1	1	1	1	1	1	0	0	1	0	1
46	0	0	0	0	0	0	0	0	0	0	0	0	1
47	0	0	0	0	0	0	0	0	0	0	0	0	1
48	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0	0	0
52	1	1	1	1	0	1	1	1	1	0	1	0	1
53	0	1	1	0	0	0	1	1	1	0	0	0	0
54	0	1	0	0	0	0	0	0	0	0	0	0	0
55	0	1	1	1	0	1	0	0	1	0	0	0	1
56	0	0	0	0	0	0	0	0	0	0	0	0	0
57	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0
60	0	1	1	1	0	1	1	1	1	0	0	0	1
61	1	1	0	1	0	0	0	0	0	0	1	0	1
62	0	1	1	1	0	0	0	0	1	0	0	0	1
63	0	0	0	0	0	0	0	0	0	0	0	0	0
64	0	0	0	0	0	0	0	0	0	0	0	0	0
65	1	0	1	1	0	0	1	0	0	0	0	0	0
66	0	0	0	0	0	0	0	0	0	0	0	0	0
67	0	0	0	0	0	0	0	0	0	0	0	0	0
68	0	1	1	0	0	0	0	0	0	0	0	0	1
69	0	1	1	1	0	1	1	1	0	0	0	0	1
70	0	1	1	1	0	1	0	1	0	0	0	0	1
71	1	0	0	0	0	0	0	0	0	0	0	0	0
72	1	1	1	1	0	1	0	1	0	0	1	0	1
73	1	1	1	1	0	1	1	1	1	0	1	0	1
74	0	1	0	0	0	0	0	0	0	0	0	0	0

Substrate	A16	A5	A2	BF6	BF41	BF49	G15	G22	G54	W22	W24	W49	USDA 6
75	0	0	0	0	0	0	0	0	0	0	0	0	0
76	0	1	0	0	0	0	0	0	0	0	0	0	0
77	0	1	1	1	0	1	1	1	1	0	0	0	1
78	0	1	0	1	0	1	0	0	0	0	0	0	1
79	0	1	0	0	0	0	0	0	0	0	0	0	0
80	0	0	0	0	0	0	0	0	0	0	0	0	0
81	1	1	1	1	0	1	1	1	1	0	0	0	1
82	1	1	1	1	0	1	1	1	1	0	0	0	1
83	1	1	1	1	1	1	1	1	1	1	1	0	1
84	1	1	1	1	0	1	1	1	1	0	1	0	1
85	0	1	0	0	0	1	0	1	1	0	1	0	1
86	0	0	0	0	0	0	0	0	0	0	0	0	1
87	0	0	0	0	0	0	0	0	0	0	0	0	0
88	0	0	0	0	0	0	0	0	0	0	0	0	0
89	0	0	0	0	0	0	0	0	0	0	0	0	0
90	1	1	1	1	0	1	1	1	1	0	0	0	1
91	1	1	1	1	0	1	1	1	1	0	1	0	1
92	0	1	1	1	0	1	0	0	0	0	0	1	1
93	0	1	1	1	0	1	0	0	0	0	1	0	1
94	0	1	0	0	0	0	0	0	0	0	0	0	0
95	0	0	0	0	0	0	0	0	0	0	0	0	0
96	0	0	0	0	0	0	0	0	0	0	0	0	0

Appendix N: Generalised linear model output for substrate utilisation

Table N-1: Output for testing for interaction between treatment and guild

```
glm(formula = Count ~ Treatment * Guild, family = quasipoisson)
Deviance Residuals:
                                  Min  1Q           Median       3Q       Max
                                -4.0521      -0.7271            0   0.5371      2.5584
Coefficients:
                                       Estimate     Std Error    t value  Pr(>|t|)
(Intercept)                               5.11E-01     6.13E-01     0.833     0.40904
TreatmentBarefallow                      -2.23E-01     9.20E-01    -0.243     0.80939
TreatmentGrassland                        1.12E-15     8.67E-01         0           1
TreatmentSymbiotic                        5.88E-01     1.00E+00     0.587     0.56004
TreatmentWoodland                        -1.61E+00     1.50E+00    -1.071      0.2894
GuildAminoAcid                            1.03E+00     7.15E-01     1.441     0.15607
GuildCarbohydrate                         1.86E+00     6.60E-01     2.815  0.00706 **
GuildCarboxylicAcid                       2.20E+00     6.47E-01     3.399  0.00137 **
GuildMiscellaneous                        4.70E-01     7.82E-01     0.601     0.55057
GuildPolymers                            -5.11E-01     1.00E+00     -0.51     0.61237
TreatmentBarefallow:GuildAminoAcid       -2.19E-01     1.09E+00      -0.2     0.84194
TreatmentGrassland:GuildAminoAcid        -8.47E-01     1.10E+00    -0.773     0.44307
TreatmentSymbiotic:GuildAminoAcid         6.90E-02     1.16E+00     0.059     0.95283
TreatmentWoodland:GuildAminoAcid         -1.72E+01     2.75E+03    -0.006     0.99502
TreatmentBarefallow:GuildCarbohydrate    -1.98E-01     9.97E-01    -0.199     0.84342
TreatmentGrassland:GuildCarbohydrate     -2.08E-01     9.40E-01    -0.221      0.8261
TreatmentSymbiotic:GuildCarbohydrate     -6.52E-01     1.12E+00    -0.583     0.56232
TreatmentWoodland:GuildCarbohydrate      -7.58E-01     1.72E+00    -0.442      0.6607
```

```
TreatmentBarefallow:GuildCarboxylicAcid    -2.82E-02     9.71E-01    -0.029     0.97696
TreatmentGrassland:GuildCarboxylicAcid     -3.73E-01     9.25E-01    -0.403     0.68868
TreatmentSymbiotic:GuildCarboxylicAcid     -4.06E-01     1.07E+00    -0.378     0.70696
TreatmentWoodland:GuildCarboxylicAcid       1.05E-01     1.58E+00     0.067     0.94701
TreatmentBarefallow:GuildMiscellaneous     -6.45E-02     1.18E+00    -0.055     0.95665
TreatmentGrassland:GuildMiscellaneous      -4.70E-01     1.17E+00    -0.402     0.68911
TreatmentSymbiotic:GuildMiscellaneous      -4.70E-01     1.37E+00    -0.344     0.73224
TreatmentWoodland:GuildMiscellaneous        6.29E-01     1.77E+00     0.356     0.72345
TreatmentBarefallow:GuildPolymers           5.11E-01     1.39E+00     0.366     0.71567
TreatmentGrassland:GuildPolymers           -1.73E+01     2.75E+03    -0.006       0.995
TreatmentSymbiotic:GuildPolymers            1.05E-01     1.60E+00     0.066     0.94788
TreatmentWoodland:GuildPolymers             1.20E+00     1.96E+00     0.616     0.54104
Signif. codes:                          0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
1
(Dispersion parameter for quasipoisson family taken to be
1.88089)
Null deviance: 443.63                   on 77        degrees of freedom
Residual deviance: 114.52               on 48        degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 15
```

Table N-2: Simplified model output

```
Call:
glm(formula = Count ~ Treatment + Guild, family =
quasipoisson)

Deviance Residuals:
                      Min        1Q    Median    3Q           Max
                                   -       -
                      -3.9761  1.1493  0.0439        0.6121  2.9922

Coefficients:
                            Std     t
                     Estimate Error   value   Pr(>|t|)
(Intercept)            0.6869 0.3372   2.037      0.0455 *
TreatmentBarefallow   -0.3034 0.2012  -1.508        0.1362
TreatmentGrassland    -0.3824 0.2059  -1.857      0.0676 .
TreatmentSymbiotic     0.2324  0.241   0.965        0.3381
TreatmentWoodland     -1.7284 0.3376  -5.119   2.71e-06 ***
GuildAminoAcid         0.7472 0.3881   1.926      0.0583 .
GuildCarbohydrate      1.6314 0.3495   4.667   1.48e-05 ***
GuildCarboxylicAcid    2.0441 0.3397   6.017   7.94e-08 ***
GuildMiscellaneous     0.3285 0.4192   0.784         0.436
GuildPolymers         -0.4925  0.519  -0.949        0.3461
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’
0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for quasipoisson family taken to be
1.839269)

Null deviance:        443.63  on 77  degrees of freedom
Residual deviance: 135.50  on 68  degrees of
freedom
AIC: NA

Number of Fisher Scoring
iterations: 6
```

Appendix O: Denitrification plate assay blocking structure

Table O-1: Blocking structure of which isolate culture on each plate

Plate number	Rep1	Plate number	Rep2	Plate number	Rep3
1	G2	11	BRM2	21	W24
	A3		G54		W74
	G9		G4		W91
	BF28		W4		W5
	G46		RD5C2		GOR1
2	G22	12	W49	22	G46
	G54		A10		BF30
	W23		GOR1		G54
	A10		USDA6		W19
	G15		BF20		G3
3	BF30	13	A2	23	BF28
	W22		A1		A3
	BRM1		W5		G31
	USDA6		G31		BRM1
	W9		BF26		A4
4	A5	14	A12	24	GOR4
	BF6		BF6		RD5C2
	G6		A5		G6
	BF41		BF28		G2
	G3		BF41		A19
5	G31	15	A4	25	BF49
	GOR1		W46		G22
	GOR4		G3		W22
	A19		W23		G15
	W5		GOR2		G4
6	W74	16	W24	26	A1
	W4		BF49		BF26
	G4		GOR4		A5
	W91		W91		BF6
	RD5C2		W74		A2
7	BF2	17	G9	27	USDA6
	A4		BF2		W9
	W19		G46		GOR2
	A12		G6		W46
	A20		G22		A12

Plate number	Rep1	Plate number	Rep2	Plate number	Rep3
8	W49	18	W9	28	A20
	A16		BRM1		G8
	BF20		G2		W23
	BF52		BF52		BF2
	BRM2		A19		BF52
9	BF48	19	BF30	29	A10
	BF49		A16		W49
	A2		W19		G9
	G8		G8		BF48
	GOR2		G15		A16
10	BF26	20	W22	30	BF41
	W24		A3		BRM2
	W46		A20		BF20
	A1		BF48		W4

Table O-2: Summary of which anaerobic jar each plate number was incubated in

Chamber 1	Chamber 2
26	24
17	29
4	20
19	11
16	13
25	28
5	3
9	15
23	27
12	6
1	10
7	2
14	30
8	22
18	21

Appendix P: Acetylene gas production

Small amounts for laboratory work can be prepared by reacting calcium carbide with water. About 15 ml of water is used for each gram of calcium carbide. A simple apparatus for generating acetylene is shown in Figure P-1. Acetylene generated this way contains very minute quantities of phosphine, ethylene and methane.

1. Place a layer of cleaned sand into the bottom the gas generation flask (A).
2. Add 1-2g (several pieces) of calcium carbide into flask (A).
3. Place bug with funnel and elimination tubing attached into flask mouth (A).
4. Connect end of elimination tubing to gas collection apparatus (B). Ensure end of elimination tube is at the base of the collection vessel and collecting vessel is full of water (B)
5. Place end of collection syringe tubing at the top of the collection vessel (C).
6. Add drop wise distilled water onto calcium carbide.
7. As the reaction proceeds acetylene (ethylene) gases should be produced and collected.

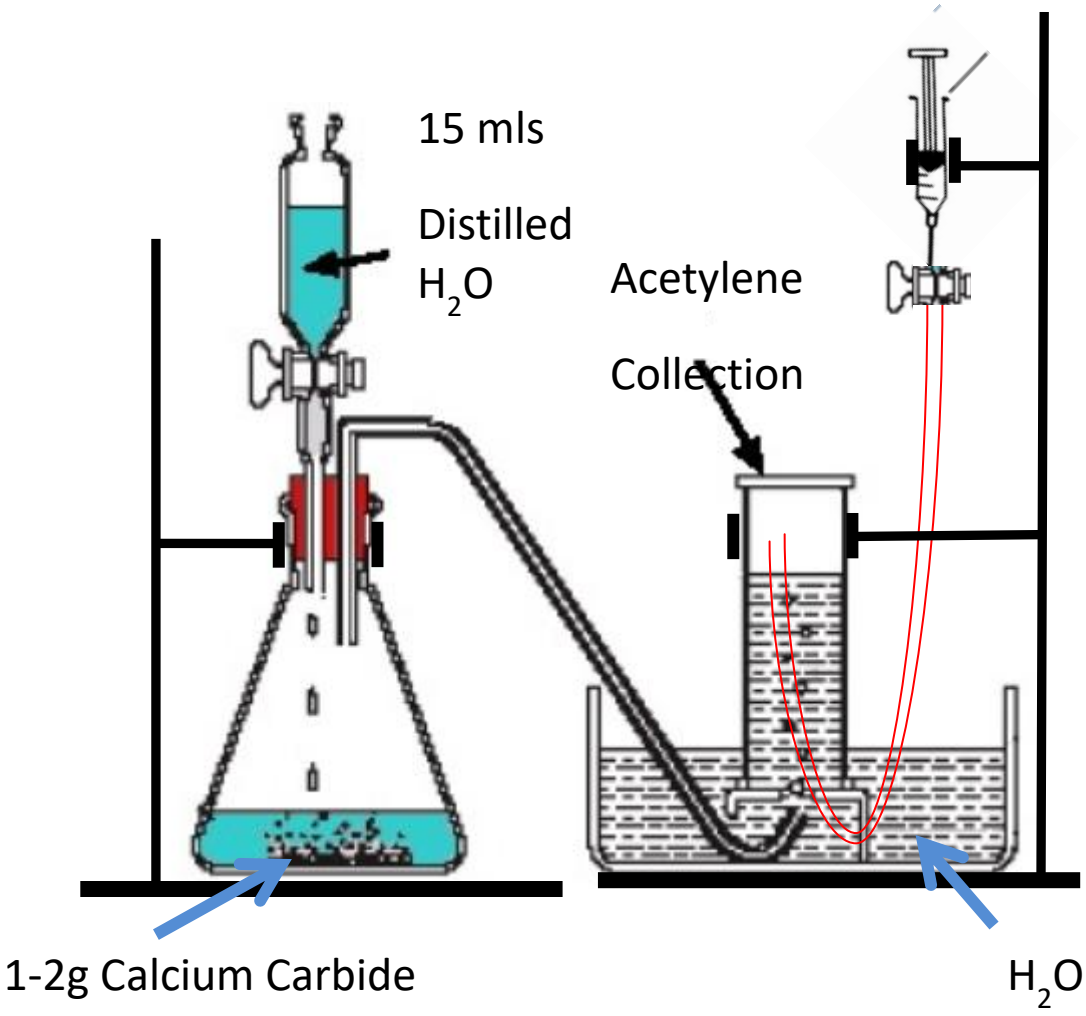


Figure P-1: Experimental set up for the production of acetylene gas.

Appendix Q: qPCR standard curves, melt curves and amplification curves

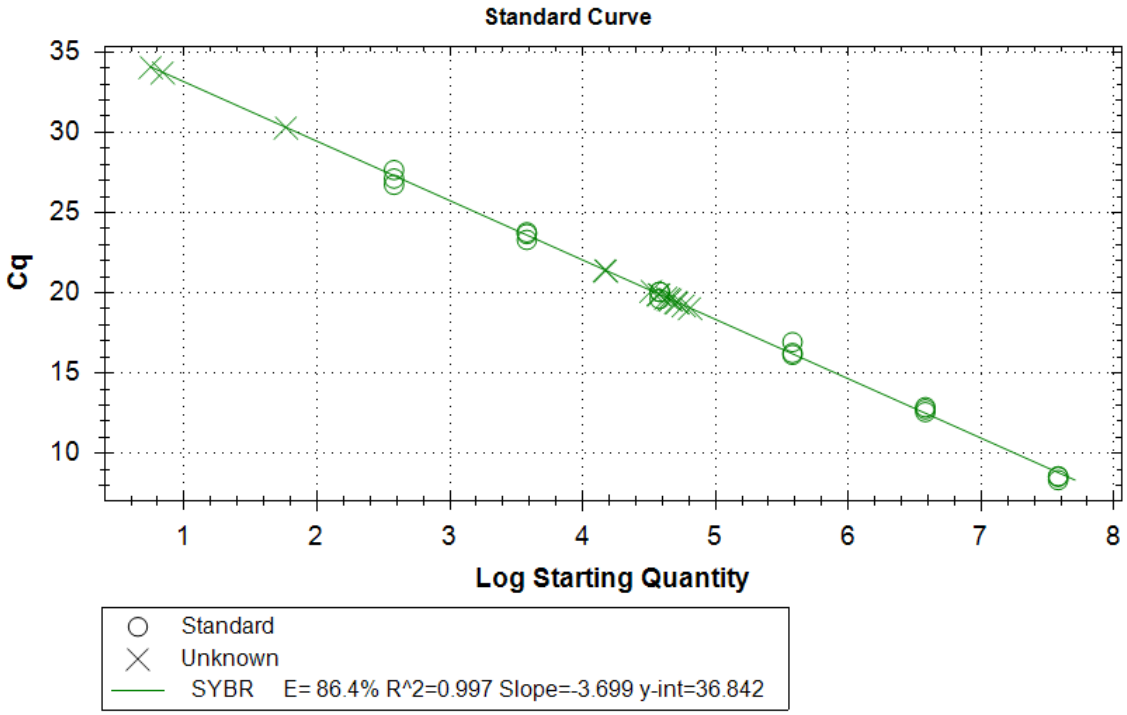


Figure Q-1: Standard curve for *nosZ* qPCR

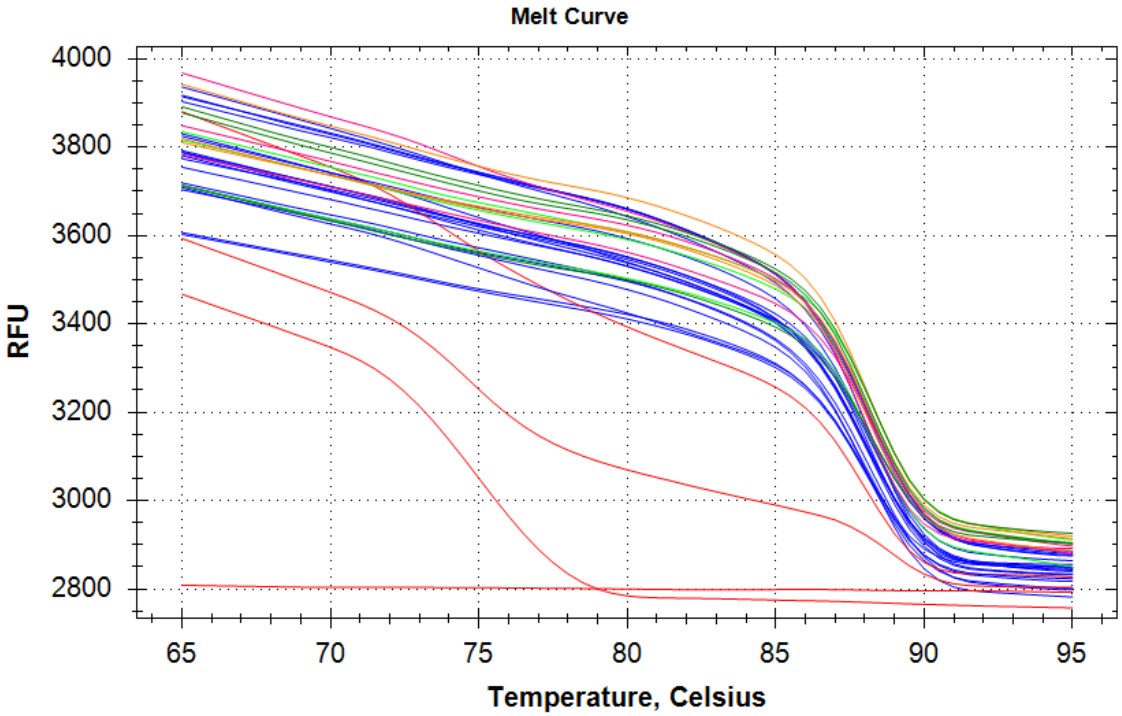


Figure Q-2: Melt curve for *nosZ* qPCR

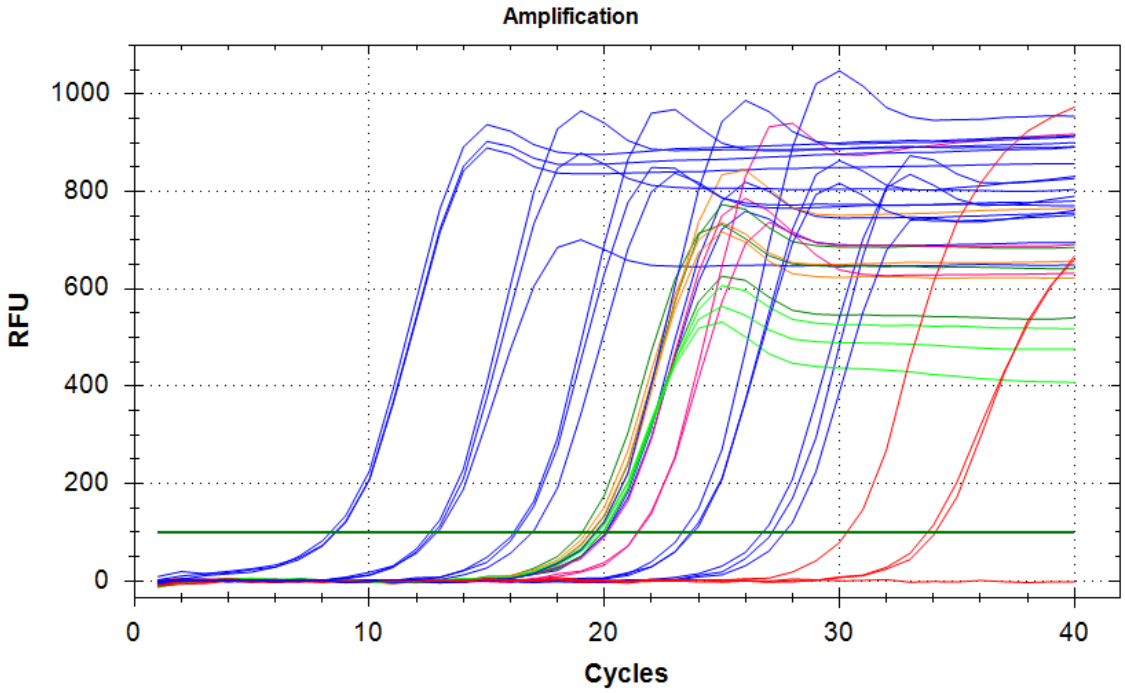


Figure Q-3: Amplification curve for *nosZ* qPCR

Appendix R: Generalised linear model output for qPCR

Table R-1: Generalised linear model output for *nosZ* qPCR

```
glm(formula = copygdw ~ Treatment, family =
quasipoisson)
Deviance Residuals:
Min        1Q           Median     3Q       Max
  -2232.4       -786.9     -181.8    698.1  2762.5
Coefficients:
                                    Std.
                     Estimate       Error     t value     Pr(>|t|)
(Intercept)                16.8959   0.2016      83.822   4.58E-13
TreatmentBarefallow        -1.5287   0.4775      -3.201   0.012586
TreatmentGrassland          1.4097   0.2248        6.27   0.000241
TreatmentWoodland           1.2763    0.228       5.599   0.000511
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
(Dispersion parameter for quasipoisson family taken to be 2653086)
Null deviance: 400336749  on 11  degrees of freedom
Residual deviance:  20865314  on  8  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4
```

Appendix S: *tfdA* PCR protocol and gel image

Each PCR reaction mixture was as stated in General Methods 2.4. The primers used to amplify a 400 bp fragment of the *tfdA* gene were tfdAalphaFItoh (5'- ACSGAGTTCKSCGACATGCG-3') and tfdAalphaRItoh (5'-GCGGTTGTCCCACATCAC-3') (Itoh et al., 2002).The PCR conditions were as follows: held at 95°C, followed by 30 cycles of 95°C for 15 seconds, 52°C for 15 seconds and 72°C for 1 minute. The product was held at 72°C for 10 minutes for full elongation of the product and held at 10°C until storage at -20°C. The PCR product was examined on a 1.5% (w/v) agarose gel as described in General Methods 2.5 using either a 100 bp marker for comparison. Negative controls were PCR reactions containing no DNA.

The *tfdA* gene is involved in the first step of 2,4-D degradation and was present in the majority of the culture collection (Figure S-1). All of the arable isolates were positive for *tfdA,* the other soils were more variable in whether this gene was present or absent.

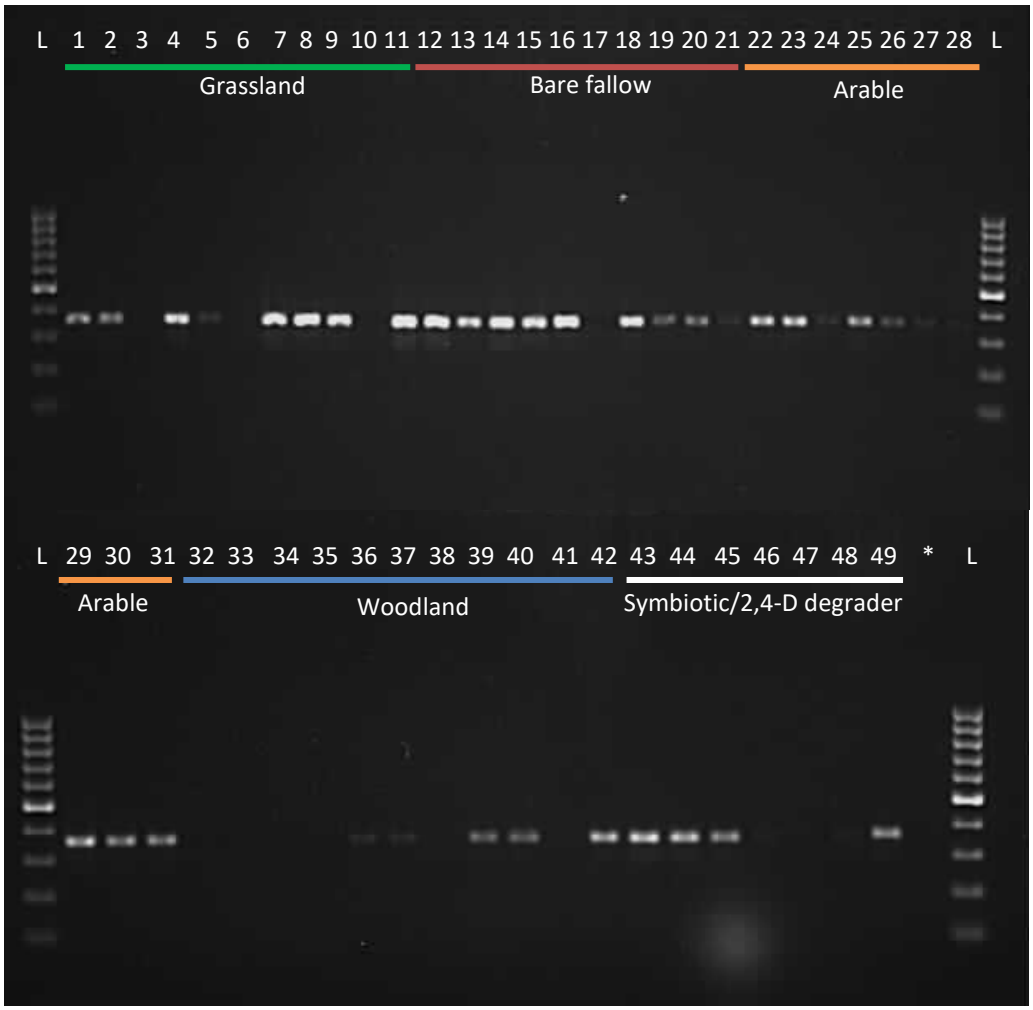
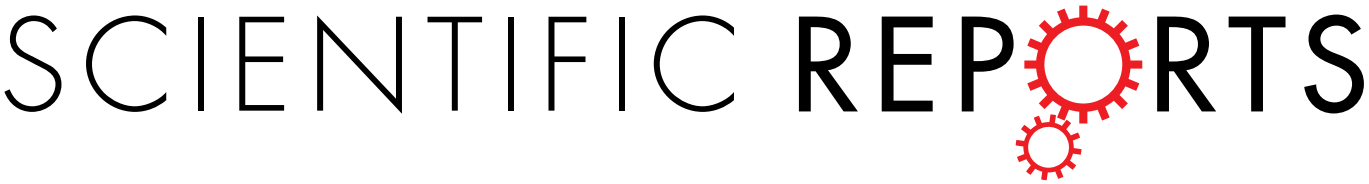


Figure S-1: Agarose electrophoresis gel used in screening for *tfdA*. A band indicates a positive for *tfdA.* The numbers correspond with the isolates listed in table B-1.
'L' is a 100bp ladder. * indicates the negative control

Appendix T: Publication

Jones, F.P., Clark, I.M., King, R., Shaw, L., Woodward, M.J. & Hirsch, P.R., 2016. Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes - a genome comparison. *Scientific Reports* **6**: 25858. DOI: 10.1038/srep25858.



OPEN

Received: 14 February 2016
Accepted: 25 April 2016
Published: 10 May 2016

Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes – a genome comparison

Frances Patricia Jones[1,2], Ian M. Clark[1], Robert King[3], Liz J. Shaw[2], Martin J. Woodward[4] & Penny R. Hirsch[1]

The slow-growing genus *Bradyrhizobium* is biologically important in soils, with different representatives found to perform a range of biochemical functions including photosynthesis, induction of root nodules and symbiotic nitrogen fixation and denitrification. Consequently, the role of the genus in soil ecology and biogeochemical transformations is of agricultural and environmental significance. Some isolates of *Bradyrhizobium* have been shown to be non-symbiotic and do not possess the ability to form nodules. Here we present the genome and gene annotations of two such free-living *Bradyrhizobium* isolates, named G22 and BF49, from soils with differing long-term management regimes (grassland and bare fallow respectively) in addition to carbon metabolism analysis. These *Bradyrhizobium* isolates are the first to be isolated and sequenced from European soil and are the first free-living *Bradyrhizobium* isolates, lacking both nodulation and nitrogen fixation genes, to have their genomes sequenced and assembled from cultured samples. The G22 and BF49 genomes are distinctly different with respect to size and number of genes; the grassland isolate also contains a plasmid. There are also a number of functional differences between these isolates and other published genomes, suggesting that this ubiquitous genus is extremely heterogeneous and has roles within the community not including symbiotic nitrogen fixation.

The slow-growing bacterial genus *Bradyrhizobium* has been shown to be one of the most abundant groups in soil[1,2] including soils sampled from long-term field experiments in the UK (Rothamsted Research, Harpenden)[3,4]. A key characteristic of the order Rhizobiales including the genus *Bradyrhizobium* is the ability to form nitrogen-fixing symbioses with legumes to increase nitrogen availability to plants[5–8]. This ability is thought to have evolved through horizontal gene transfer as the genes involved in this process are usually located on symbiosis islands on the chromosomes of bradyrhizobia[7–9] or on symbiotic plasmids in many rhizobia[5]. Some isolates of *Bradyrhizobium* have been shown to be non-symbiotic and do not possess the ability to form nodules. The absence of nodulation ability has been noted in the strain *Bradyrhizobium* sp. S23321 isolated from paddy soil in Japan[8] although nitrogen fixation (*nif*) genes were present. Recently, *Bradyrhizobium* ecotypes from forest soils have been shown to lack both nodulation and nitrogen fixation genes[2].

Bradyrhizobium is biologically important in soils, with different representatives found to perform a wide range of biochemical functions including photosynthesis, nitrogen fixation during symbioses, denitrification

[1]Department of AgroEcology, Rothamsted Research, Harpenden, AL5 2JQ, UK. [2]Department of Geography and Environmental Science, University of Reading, Reading, RG6 6AH, UK. [3]Department of Computational and Systems Biology, Rothamsted Research, Harpenden, AL5 2JQ, UK. [4]Department of Food and Nutritional Sciences, University of Reading, Reading, RG6 6AH, UK. Correspondence and requests for materials should be addressed to F.P.J. (email: frances.jones@rothamsted.ac.uk)

Taxonomy	Strain	Host	Origin	Genome size (bp)	GC content	Proteins	rRNA operon	tRNA	Gene	Accession number	Reference
Bradyrhizobium sp.	BF49	Free-living	UK	7,547,693	63.80%	7380	1	48	7431	ERS954959	This paper
Bradyrhizobium sp.	G22	Free-living	UK	9,022,917	63.70%	8653	1	49	8706	ERS955657	This paper
Bradyrhizobium sp.	G22 plasmid	Free-living	UK	364,482	60.70%	541	1	2	546	ERS955536	This paper
Bradyrhizobium sp.	G22 unplaced	Free-living	UK	68,438	62.10%	81	0	0	81	ERS955657	This paper
Bradyrhizobium sp.	S23321	Free-living	Japan	7,231,841	64.30%	6898	1	45	6951	NC_017082	8
B. diazoefficiens	USDA 110	*Glycine max*	USA	9,105,828	64.10%	8317	1	50	8373	NC_004463	10
B. japonicum	USDA 6	*Glycine max*	Japan	9,207,384	63.70%	8829	2	51	8888	NC_017249	59
B. japonicum	E109	*Glycine max*	Argentina	9,224,208	63.70%	8233	2	54	8621	NZ_CP010313	60
B. oligotrophicum	S58	*Aeschynomene*	Japan	8,264,165	65.10%	7228	2	51	7285	NC_020453	11
Bradyrhizobium sp.	ORS278	*Aeschynomene*	Africa	7,456,587	65.50%	6752	2	50	6818	NC_009445	6
Bradyrhizobium sp.	BTAi1	*Aeschynomene*	N. America	8,264,687	64.90%	7394	2	52	7553	NC_009485	6
Bradyrhizobium sp.	BTAi1 plasmid	*Aeschynomene*	N. America	228,826	60.70%	203	0	0	216	NC_009475	6

Table 1. Summary of the seven published complete *Bradyrhizobium* genomes and the two novel strains G22 and BF49.

and aromatic compound degradation[8]. Nitrogen removal through heterotrophic denitrification is an important step in the global nitrogen cycle carried out by many groups including *Bradyrhizobium*[10]. The multiple roles of *Bradyrhizobium* in the nitrogen cycle make the ecology of this group important for agriculture.

Bradyrhizobium is studied extensively due to its symbiotic relationship with soybean and consequently several genomes have been published. Currently, there are seven complete *Bradyrhizobium* genomes in the NCBI database. Six of these are symbiotic and are able to fix nitrogen and form root nodules on legumes (*B. diazoefficiens* USDA 110, *B. japonicum* USDA 6, *B. japonicum* E109, *Bradyrhizobium* sp. ORS278, *Bradyrhizobium* sp. BTAi1 and *B. oligotrophicum* S58) with ORS278, BTAi1 and S58 able to form both stem and root nodules on the aquatic legume *Aeschynomene*[6,11]. The other bradyrhizobial genome (*Bradyrhizobium* sp. S23321) is free-living because it is unable to form nodules; however, it still contains the genes required for nitrogen fixation. Four genomes sequenced from North American forest soils were also missing nodulation and nitrogen fixation genes (*Bradyrhizobium* sp. LTSP849, *Bradyrhizobium* sp. LTSP857, *Bradyrhizobium* sp. LTSP885 and *Bradyrhizobium* sp. LTSPM299). These genomes were sequenced using shotgun sequencing of the soil community and were assembled to near completion[2]. Due to the availability of a diverse array of genome reference sequences, *Bradyrhizobium* is an appropriate model to study other soil bacteria: understanding the mechanisms of *Bradyrhizobium* adaptation to independent living in agricultural soils under contrasting management may reveal the genetic potential of this globally important genus.

Here we present the genome and gene annotations and carbon metabolism profiles of two free-living *Bradyrhizobium* isolates from the Highfield experiment at Rothamsted Research that has three long-term treatment regimes: grassland, arable (wheat) and bare fallow tilled regularly to maintain a plant-free soil. Maintenance of these treatments for 60 years has led to distinct differences in soil properties and the soil microbiome[12]. *Bradyrhizobium* sp. G22 and *Bradyrhizobium* sp. BF49 were isolated from soil taken from the permanent grassland and permanent bare fallow plots of the Highfield experiment respectively. These *Bradyrhizobium* strains are the first to be isolated and genome sequenced from European soil and the first free-living and non-diazotrophic isolates, without the presence of either nodulation or nitrogen fixation genes, to have their genomes sequenced and assembled from cultured samples. The isolates were interrogated for differences to determine the level of genetic heterogeneity in carbon metabolism between these isolates.

Results and Discussion

General genome description and comparisons. The genome of the grassland isolate G22 is 9,022,917 bp in size while the bare fallow isolate BF49 genome is 7,547,693 bp, constituting a 1.5 Mbp size difference in addition to a 364,482 bp plasmid in G22. The genome size for G22 is similar to nodulating strains *B. diazoefficiens* USDA 110, *B. japonicum* USDA 6, *B. japonicum* E109, *B. oligotrophicum* S58 and *Bradyrhizobium* sp. BTAi1, whereas for BF49 it is closer in size to the free-living strain *Bradyrhizobium* sp. S23321 and the photosynthetic, nodulating strain *Bradyrhizobium* sp. ORS278. Table 1 summarises the genome information of the two novel strains G22 and BF49 along with the other seven completed *Bradyrhizobium* genomes in the database. G22 had 19 contigs which could not be placed in the chromosome or plasmid. These were annotated and used in the subsequent analysis.

G22 shows 1356 more genes (8787) than BF49 (7431) and this rises to 1902 when including the genes contained on the plasmid (541). GC content is similar between the strains at 63.7% and 63.8% for G22 and BF49 respectively, consistent with other *Bradyrhizobium* strains listed in Table 1. The G22 plasmid has a GC content of 60.7%, identical to the plasmid of BTAi1. The G22 and BF49 genomes show high pairwise sequence identity in comparison to both USDA 110 (G22: 84.4%, BF49: 83.4%) and S23321 (G22: 83.4%, BF49: 83.2%) using LASTZ (Large-Scale Genome Alignment Tool).

Orthologous gene clusters and core genome phylogeny. G22 and BF49 were compared with the free-living isolate, S23321, and the symbiotic isolate, USDA 110 (Fig. 1). This suggests that there is a core genome of 4562 genes which are present in all four genomes assessed. The 103 genes present only in the USDA 110 genome include those involved in nodulation and uptake hydrogenase. The 171 genes which are only present in

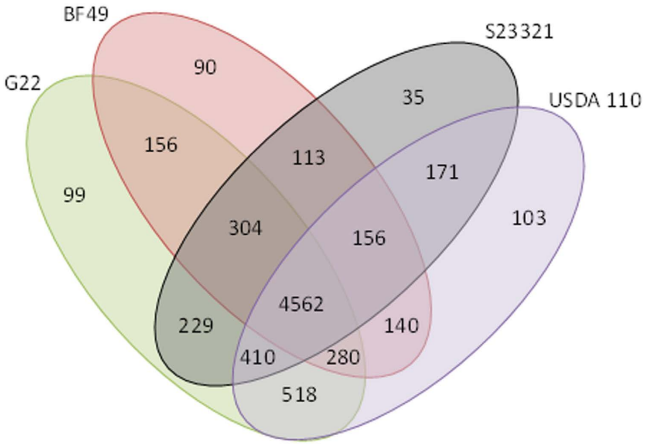


Figure 1. Venn diagram showing orthologous gene clusters for G22, BF49 and two published genomes, S23321 and USDA 110. Output generated using data from OrthoVenn.

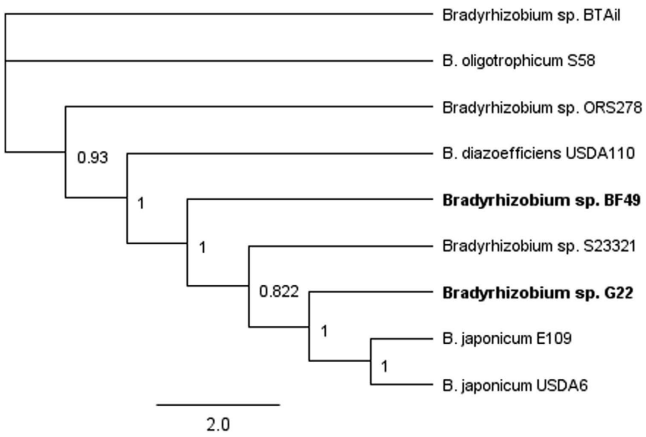


Figure 2. Phylogenetic tree for the 3442 homologous gene clusters in the core genome for the nine *Bradyrhizobium* genomes. The node labels represent the certainty of that node in the phylogenetic tree where one is maximum certainty. Reference sequence: *B. diazoefficiens* USDA 110.

the USDA 110 and S23321 genomes include nitrogen fixation genes. Only a small number of genes are unique to any one isolate (G22: 99, BF49: 90, S23321: 35, USDA 110: 103). OrthoVenn identified a core genome of 3442 homologous gene families across the nine complete genomes available in the database. The core genome SNP analysis (Fig. 2) shows that G22 clusters closest with the *B. japonicum* strains; E109 and USDA 6. BF49 is separate from G22 being more basal. The closest relatives are the free-living strain S23321 and the soybean nodulating type-strain USDA 110. Two of the photosynthetic, *Aeschynomene*-nodulating isolates (BTAi1 and S58) cluster together.

Isolate identification and 16S phylogeny. The 16S rRNA sequence from G22 shows 100% identity with *Bradyrhizobium* sp. VUPMI37 [Accession number: HG940535] and *Bradyrhizobium* sp. ICMP12674 [Accession number: AY491080][13] from the NCBI and RDP databases respectively. *Bradyrhizobium* sp. VUPMI37 was originally isolated from *Vigna unguiculata* nodules and *Bradyrhizobium* sp. ICMP12674 was originally isolated from *Ulex europaeus*[13,14]. The BF49 16S rRNA fragment shows 100% identity with *B. canariense* SEMIA928 from the NCBI database [Accession number: FJ390904][15] and *B. canariense* LG-6 from the RDP database [Accession number: GU306140]. *B. canariense* SEMIA928 was originally isolated from *Lupinus* spp. and *B. canariense* LG-6 was originally isolated from *Lupinus angustifolius* root nodules[16]. The 16S phylogeny (Fig. 3) clusters G22 with VUPMI37 and ICMP12674, which were the top hits from the NCBI and RDP databases. Similarly, BF49 clusters with the top hits from the NCBI and RDP databases (SEMIA928 and LG-6) in addition to *B. lupini* USDA 3051 which was the second hit from the NCBI database [Accession number: NR_134836]. This strain was originally isolated from *Lupinus* and was reclassified from *Rhizobium lupini* to *B. lupini* in 2015[17]. Both G22 and BF49 are in the same clade as the free-living strain, S23321, and the soybean-nodulating strains, E109, USDA 6 and USDA 110. Two of the four strains from North American forest soils, *Bradyrhizobium* sp. LTSP849 and LTSP857[2] were also in this clade. The photosynthetic strains (ORS278, BTAi1 and S58) are in a separate clade including the remaining two strains from North American forest soils Bradyrhizobium sp. LTSP885 and LTSPM299.

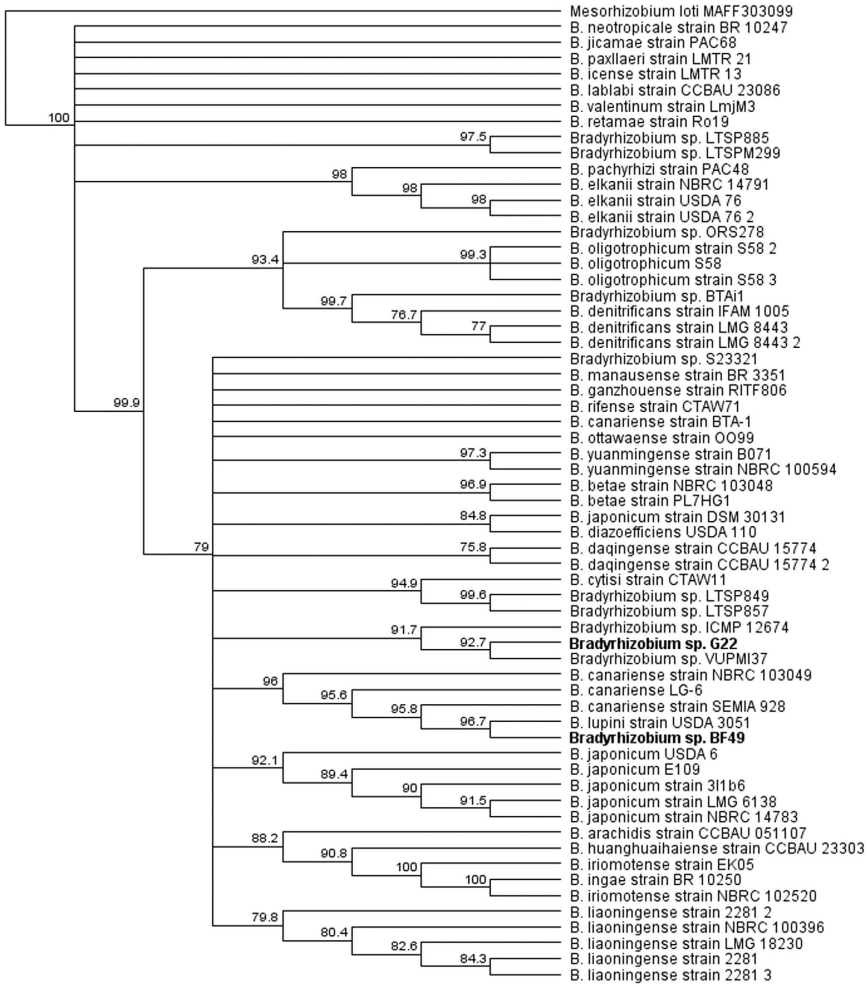


Figure 3. Phylogenetic tree for the 16S rRNA gene using Neighbour-Joining clustering method with 1000 bootstraps.

Genes involved in nitrogen fixation and nodulation. Nitrogen fixation and nodulation genes including *nifDKH, nodD* and *nodABC;* are all absent from both G22 and BF49 (Table 2) and so we suggest that these isolates are not able to either form nodules or fix atmospheric nitrogen. The absence of both *nif* and *nod* genes is in contrast to other published, complete *Bradyrhizobium* genomes and indicates similarity with forest soil bacteria (LTSP849, LTSP857, LTSP885 and LTSPM299)[2]. *B. diazoefficiens* USDA 110 (previously *B. japonicum* USDA 110) is a known symbiotic strain and contains all *nif* and *nod* genes listed. The seven complete *Bradyrhizobium* genomes which have previously been published either contain both *nif* and *nod* genes (the soybean-nodulating strains *B. diazoefficiens* USDA 110, *B. japonicum* USDA 6 and *B. japonicum* E109) or just *nif* genes (the *Aeschynomene*-nodulating strains *Bradyrhizobium* sp. BTAi1, *Bradyrhizobium* sp. S58 and *Bradyrhizobium* sp. ORS278 and the free-living strain *Bradyrhizobium* sp. S23321) (Fig. 4). These strains use a *nod*-independent route for stem and root nodulation of *Aeschynomene*[11]. The FixLJ two component system is present in both the grassland and bare fallow isolate (Table 2). FixLJ acts in response to low oxygen conditions in soil and in the nodule and controls the expression of genes for both nitrogen fixation and denitrification[18,19]. This two-component system has also been shown to regulate the response to nitric oxide in *Sinorhizobium meliloti* being shown to regulate a high proportion of genes induced by the presence of nitric oxide[20]. The presence of *fixLJ* in G22 and BF49 is consistent with other *Bradyrhizobium* isolates including all seven completed genomes.

Genes involved in denitrification. Both genomes encode a nitrate reductase, NapA/B, which catalyses the reduction of nitrate to nitrite; the first stage in denitrification[21–23]. This is common among *Bradyrhizobium* including all seven complete genomes (Table 2 and Fig. 4) and the four genomes from North American forest soils (Table 2). In addition, all fully-sequenced isolates including both G22 and BF49 contain *nirK,* encoding a respiratory copper-containing nitrite reductase which is involved in the second stage of denitrification reducing nitrite to nitric oxide. The third stage of denitrification is the conversion of nitric oxide into nitrous oxide, a potent greenhouse gas, and is catalysed by a nitric oxide reductase encoded by *norB/C*[22,23]. The presence of a nitric oxide reductase gene has been noted in all previously published and complete *Bradyrhizobium* genomes in addition to G22 and BF49. The denitrification pathway is not present in the four genomes of strains from North American forest soils.

Function	Gene(s)	G22	BF49	S23321	ORS278	BTAi1	S58	USDA 110	USDA 6	E109	LTSP849	LTSP857	LTSP885	LTSPM299
Nitrogen fixation	*nifDKH*	−	−	+	+	+	+	+	+	+	−	−	−	−
Nitrogen fixation	*fixLJ*	+	+	+	+	+	+	+	+	+	−	−	−	−
Nodulation	*nodD*	−	−	−	−	−	−	+	+	+	−	−	−	−
Nodulation	*nodABC*	−	−	−	−	−	−	+	+	+	−	−	−	−
Denitrification	Nitrate reductase *napA/B*	+	+	+	+	+	+	+	+	+	+	+	+	+
Denitrification	Nitrite reductase *nirK*	+	+	+	+	+	+	+	+	+	−	−	−	−
Denitrification	Nitric oxide reductase *norB/C*	+	+	+	+	+	+	+	+	+	−	−	−	−
Denitrification	Nitrous oxide reductase *nosZ*	−	+	−	−	+	+	+	−	−	−	−	−	−
Uptake hydrogenase	*hup*	−	−	−	+	+	+	+	+	+	−	−	−	−
Photosynthesis	Bacteriochlorophyll *bchCXYZ/FNBHLM*	−	−	+	+	+	+	−	−	−	−	−	−	−
Photosynthesis	Carotenoids *crtEF*	−	−	+	+	+	+	−	−	−	−	−	−	−
Photosynthesis	LIght harvesting complexes *pucBAC/pufBA*	−	−	+	+	+	+	−	−	−	−	−	−	−
Photosynthesis	Reaction centre subunits *puhA/pufLM*	−	−	+	+	+	+	−	−	−	−	−	−	−
Carbon fixation	RuBisCo	+	+	+	+	+	+	+	+	+	+	+	+	+

Table 2. Summary of presence or absence of genes of interest in the seven published complete *Bradyrhizobium* genomes, four isolates from North America forest soils and the two novel strains G22 and BF49 (+ = present, − = absent).

The ability to convert the greenhouse gas, nitrous oxide into environmentally benign nitrogen gas in the final stage of denitrification is an attribute with global importance[24]. Nitrous oxide reductase encoded by *nosZ* catalyses this process but it is not ubiquitous across *Bradyrhizobium*[22,23,25,26]. Of the seven published complete genomes, only *B. diazoefficiens* USDA 110, *Bradyrhizobium* sp. BTAi1 and *B. oligotrophicum* S58 contain the *nosZ* gene. It is absent from the grassland isolate, G22 but present in the bare fallow isolate, BF49. The presence of the gene shows a potential for BF49 to perform this function and increases the agricultural and environmental importance of this isolate.

Uptake hydrogenase. The uptake of hydrogen is catalysed by uptake hydrogenase which is encoded by the *hup* genes[27–29]. This process produces ATP which is used by nitrogen-fixing bacteria to mediate for energy lost through the nitrogen fixation process[27,29]. The nickel-iron hydrogenase, encoded by *hupSL*[29], is absent from both G22 and BF49 but is present in all of the symbiotic strains of *Bradyrhizobium* with complete genomes. These genes are also absent from LTSP849, LTSP857, LTSP885 and LTSPM299 (Table 2).

Photosynthesis and carbon fixation. Genes for photosynthesis are present in four of the published complete *Bradyrhizobium* genomes; S23321 (free-living), S58, BTAi1 and ORS278 (*Aeschynomene* host) (Fig. 4b). These include genes for bacteriochlorophyll (*bchCXYZ*/*FNBHLM*), carotenoids (*crtEF*), light harvesting polypeptides (*pucBAC*/*pufBA*) and reaction centre subunits (*puhA*/*pufLM*)[30,31]. They are not present in the soybean-nodulating isolates (USDA 110, USDA 6, and E109), G22 or BF49 genomes or the four isolates from forest soils (LTSP849, LTSP857, LTSP885 and LTSPM299) (Table 2). Many heterotrophic bacteria including *Bradyrhizobium* can fix carbon using the Calvin-Benson-Bassham cycle (CBB cycle). The significance of this is unclear in most cases although the photosynthetic *Aeschynomene*-nodulating strain ORS278 has been shown to require an active CBB cycle for symbiotic nitrogen fixation[32]. The first stage of the CBB cycle is catalysed by Ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCo) which is present in both G22 and BF49 genomes, consistent with other *Bradyrhizobium* isolates (Table 2). Transketolase is an important enzyme in both the CBB cycle and pentose phosphate pathway, catalysing the interconversion of sugars[33,34]. Genes for the transketolase enzyme are present in both G22 and BF49 consistent with all other complete *Bradyrhizobium* isolates which have been published.

Carbon metabolism. The principal components analysis (PCA) biplot (Fig. 5A) shows that there is a separation of the time points and USDA 6 from G22 and BF49 across PC1 which accounts for 44.86% of the variation. Across PC2, G22 and BF49 separate and PC2 accounts for 30.52% of the total variation. The first two principal components were visualised as together they accounted for 75.38% of the variation. The third PC accounted for 7.5%. Figure 5B shows the 95 substrates colour coded according to category. More carboxylic acids and amino acid substrates are associated with the separation across PC1; USDA 6 from G22 and BF49. Carbohydrates tend to have a negative direction across PC2 being more closely associated with BF49. One carboxylic acid, malonic acid, was the only substrate which has a positive direction for PC2 and negative for PC1 associating more closely with G22.

Malonic acid can be found in plant tissues, including legumes being first characterised from alfalfa leaves in 1925 and has been found to be in very high concentrations in soybeans[35–37]. This pathway was found to be closely associated with the symbiotic nitrogen fixation pathway in *Rhizobium leguminosarium* bv *trifolii*[38]. Malonic acid is activated before being broken down into acetate and carbon dioxide through decarboxylation[39]. It is often converted to malonyl-CoA by a CoA transferase, which is present in G22 (Phosphoribosyl-dephospho-CoA

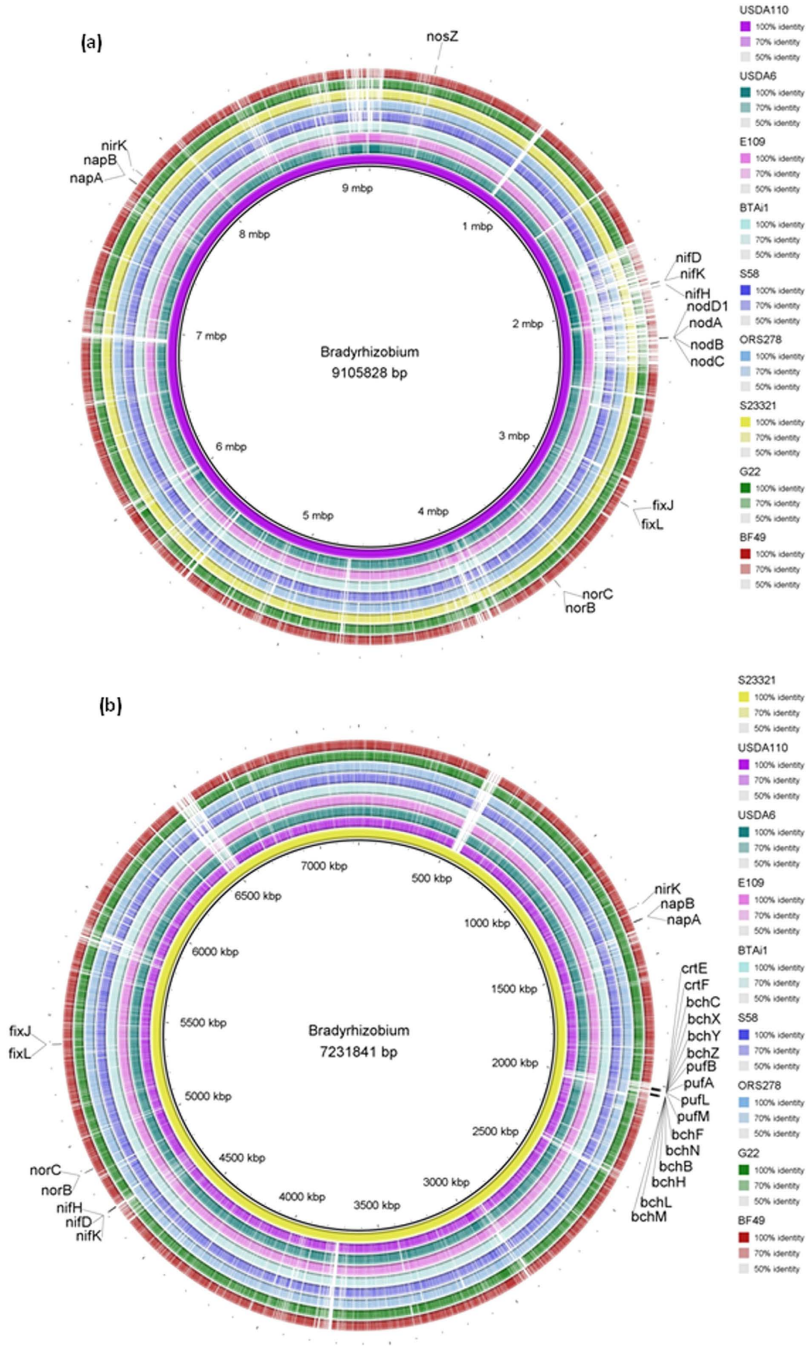


Figure 4. Whole chromosome comparisons between the seven complete *Bradyrhizobium* genomes, BF49 and G22 showing positions of genes involved with nitrogen cycling, nodulation and photosynthesis on the reference genome sequence. Reference sequence: USDA110 (**a**) and S23321 (**b**).

transferase) and the malonyl-CoA is then decarboxylated by malonate decarboxylase[35,39]. Malonate decarboxylase has four subunits; alpha, beta, gamma and delta; all of which are present in G22. The grassland isolate also contains two *mad* genes; *madL* and *madM*. These genes have previously been reported to be part of the malonate decarboxylase operon as transport proteins thought to be involved in malonate uptake[35]. The grassland isolate G22 also contains a malonyl CoA acyl carrier protein transacylase and triphosphoribosyl-dephospho-CoA synthetase. All genes involved in malonate decarboxylation are absent in the bare fallow isolate BF49. Malonate transport and utilisation genes are also present in other *Bradyrhizobium* isolates including BTAi1, ORS278, USDA 110, S23321 and all four of the forest strains (LTSP849, LTSP857, LTSP885 and LTSPM299). When grown in malonic acid, only G22 was able to metabolise it whereas BF49 and USDA 6 were not able to utilise this carbon source to the same extent.

The highest loadings and therefore the substrates which make the largest contribution to PC1 were L-pyroglutamic acid, L-leucine and D-galacturonic acid. L-pyroglutamic acid is an amino acid which is also

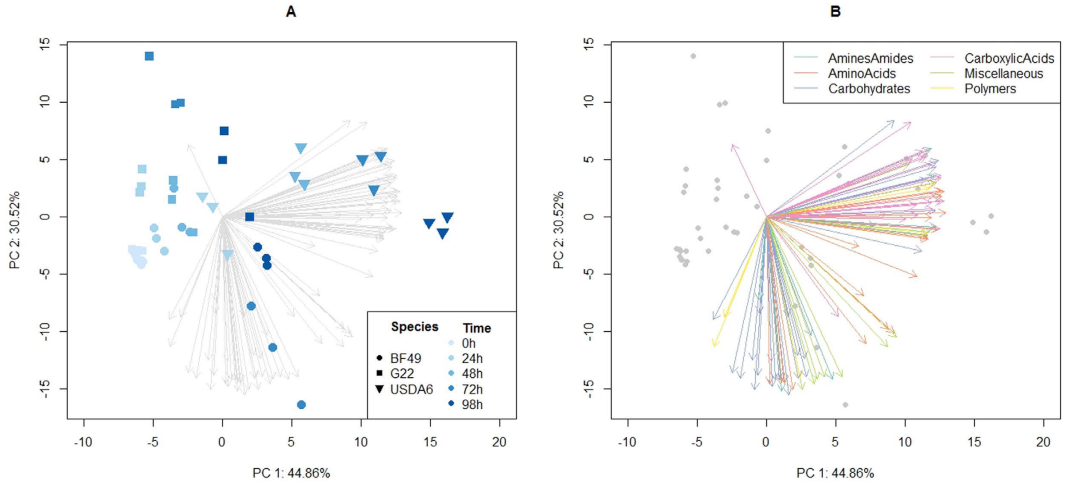


Figure 5. Principal components analysis biplot showing PC1 and PC2 accounting for 75.38% of the variation. (**A**) shows individual replicates across the 5 time points and (**B**) shows the 95 substrates classified into 6 categories.

known as 5-oxoproline. It is involved in the glutathione pathway and is converted to L-glutamic acid by the enzyme 5-oxoprolinase (EC 3.5.2.9)[40,41]. This gene is present in USDA 6 but is absent in both G22 and BF49. USDA 6 recorded a larger OD across all time points for this substrate than G22 and BF49. L-leucine is also an amino acid and is involved in numerous pathways including valine, leucine and isoleucine biosynthesis and degradation[40,41]. USDA 6 was able to utilise this substrate to a much greater extent than G22 and BF49 and the main difference in genes involved in L-leucine metabolism was the presence of leucine transaminase (EC 2.6.1.6) in USDA 6 and absence in G22 or BF49. D-galacturonic acid is a carbohydrate involved in numerous pathways including pentose and glucoronate interconversions, starch and sucrose metabolism and amino sugar and nucleotide sugar metabolism[40,41]. It is metabolised by USDA 6 more readily than G22 and BF49. This is likely due to the presence of pectinase (EC 3.2.1.15) and urinate dehydrogenase (EC 1.1.1.203) in USDA 6 but not in G22 or BF49. OD curves for the substrates discussed can be found in Supplementary Information S1.

Plasmid. Plasmid replication genes *repABC* and *parAB* are only found on the G22 plasmid and not on either the G22 or BF49 chromosomes. The *trb* operon for conjugal transfer consists of 12 genes; *traI, trbBCDEJKLFGHI*[42]. From this operon, 10 out of the 12 genes are present on the G22 chromosome; only *trbH* and *traI* are missing. Conjugal transfer genes *trbBCDEFGIJL* are absent from BF49. Conjugative transfer DNA nicking endonuclease genes *traR*/*traO* are present only in the G22 plasmid. The other genes which were unique to the plasmid were two genes involved in purine utilisation (*yagS* and a putative xanthine transporter) and one for osmotic stress (*opgC*). The purine utilisation genes, *yagS,* encodes a periplasmic aromatic aldehyde oxidoreductase. YagS is usually part of an operon *yagTSRQ* which encodes a molybdenum-containing iron sulphur flavoprotein, where YagS is the FAD-containing subunit. The role of this protein has been suggested to be detoxification of aromatic aldehydes[43,44]. The flavoprotein produced from *yagTSRQ* shows homology with xanthine dehydrogenase[43]. The osmotic stress gene, *opgC,* is involved in the synthesis of osmoregulated periplasmic glucans (OPGs). The exact role of OPGs is not understood however they have been shown to potentially play a role in the interaction between bacteria and their eukaryotic host[45,46]. The OpgC protein has been examined in *Rhodobacter sphaeroides* and was shown to encode a succinyltransferase homolog involved in the succinyl modification of OPGs[45,46]. The other genes contained on the plasmid are also on the chromosome of both G22 and BF49 including genes for DNA ligases, DNA repair, cAMP signalling and RNA processing and modification.

Conclusions

The strains described here are the first to be isolated and genome-sequenced from European soil and are unique compared to other completed *Bradyrhizobium* genomes due to the absence of previously characterised genes and gene clusters for symbiosis, nitrogen fixation and photosynthesis. They are also distinct from the North American forest isolates as G22 and BF49 contain genes for denitrification. They represent a major group, likely to play a key role in denitrification. The presence or absence of the terminal denitrification gene, *nosZ,* may determine whether the end product of denitrification is the potent greenhouse gas, nitrous oxide or the less problematic nitrogen gas. The carbon metabolism analysis shows that G22 and BF49 show different metabolic profiles over time and this are also distinct from a nodulating strain, USDA 6. The genomes and carbon metabolism analysis indicate that the free-living soil *Bradyrhizobium* have the potential to carry out many degradative and transformative functions in soil; the marked differences between two isolates from comparable soils that have undergone different management indicates that they form part of an extremely heterogeneous group.

Methods

Isolation and identification. *Bradyrhizobium* were isolated from the permanent grassland and bare fallow plots in the Highfield experiment at Rothamsted Research[12]. Serially diluted soil samples were

plated onto modified arabinose gluconate (MAG) agar plates incubated at 28 °C[47]. Colonies forming after 7 days were picked, DNA extracted using MicroLYSIS-Plus using manufacturer's instructions (Microzone, UK) and identified by PCR by the production of a 1360 bp 16S rRNA fragment using custom-designed *Bradyrhizobium* specific primers: Bradj16S70F (5′-GCGGGCGTAGCAATACGTCAGC-3′) and Bradj16S1430R (5′-GCCGGCTGCCTCCCTTGCGGGTTA-3′). Each PCR mixture (20 μl) consisted of 1x NH4 reaction buffer, 0.5 mM of each dNTP, 2.5 mM $MgCl_2$, 0.1 μM of each primer, 1 U Biotaq polymerase (Bioline, UK) and 10 ng of genomic DNA. The PCR conditions were as follows: 95 °C 2 min, 30 cycles of 95 °C 15 sec, 68 °C 15 sec and 72 °C 1 min and extension at 72 °C for 10 min. The PCR products were examined on a 1.5% agarose gel stained with ethidium bromide (0.2 μg ml^{-1}) at 100 V for 60 min. PCR products were purified using the Wizard SV gel and PCR clean up system (Promega, USA), sequenced as described in Mauchline *et al.* (2014)[48] and identification confirmed using BLAST and searching the NCBI and RDP databases. Two isolates, one from grassland soil and one from bare fallow soil were chosen at random from the generated *Bradyrhizobium* culture collection to have their genomes sequenced *de novo*.

De novo genome sequencing. DNA was extracted from isolates grown in MAG broth incubated at 28 °C and shaken at 100 rpm using the GenElute Bacterial Genomic DNA kit (Sigma Aldrich, USA). Extracted DNA was quantified using Qubit Fluorometric Quantification (Life Technologies, USA). Two sequencing platforms were used; Ion Torrent PGM™ (Life Technologies, USA) (located at Rothamsted Research, UK) and Illumina HiSeq 2000 (Illumina Inc, USA) (located at BGI, the Beijing Genomics Institute, China). For Ion Torrent sequencing, two barcoded, unamplified sequencing libraries were constructed using an Ion Plus Fragment Library Kit (Life Technologies, USA) using 1 μg template DNA. Libraries were pooled and sequence template generated using an Ion PGM™ Template OT2 400 Kit. Sequencing was performed using the Ion PGM™ Sequencing 400 Kit and an Ion 318™ Chip. The G22 sample generated 3,652,179 reads and the BF49 sample generated 3,129,252 reads. Illumina sequencing was performed by BGI and 40 ug of genomic DNA was used. A 6 kb mate library with 90 bp read length and 50X coverage was created using the Illumina HiSeq 2000 sequencing platform resulting in 2,997,699 paired reads for G22 and 3,026,005 paired reads for BF49.

De novo genome assembly and annotation. The sequence data was assembled using SOAPdenovo2 assembler and gap closer[49] using a maximum read length of 615 bp and the average insert size of 6 kb. A range of assemblies were produced using kmer values; 61–83, 81, 83, 85, 87, 89, 91. The genomes were manually curated using Geneious (Biomatters Ltd v8.1.5). Gap closing was carried out by PCR. Any remaining contigs from the 61–83 kmer reference assembly over 500 bp were investigated while those under were discarded. The G22 assembly had 19 contigs over 500 bp unplaced within the chromosome, with some containing annotated genes. No contigs over 500 bp remained after the completion of the final BF49 reference and so we can be confident that all genes are present in the assembly. The genomes were uploaded into RAST (Rapid Annotation using Subsystem Technology) for annotation[50,51]. The default parameters were used which were as follows: Classic RAST annotation scheme, RAST gene caller and Release70 FIGfam version. Automatically fix errors and backfill gaps were selected. The sequences have been deposited in the ENA database [Study ID: PRJEB10689, Sample ID G22 and unplaced contigs: ERS955657, G22 plasmid: ERS955536, Sample ID BF49: ERS954959]. The raw sequence data was also uploaded to ENA [G22: ERR1110561, BF49: ERR1110562] for Illumina and Ion Torrent [G22: ERR1110625, BF49: ERR1110597].

Genome comparisons. Genome annotations were downloaded from RAST and examined manually using KEGG[40,41]. OrthoVenn was used to assess gene orthology using an e-value of $1e^{-10}$ [52]. BRIG (BLAST Ring Image Generator) was used to compare the genomes with other *Bradyrhizobium* isolates (NCBI blast version 2.2.31)[53]. Large-Scale Genome Alignment Tool (LASTZ) was used to assess genome-wide sequence similarity in Geneious[54].

16S rRNA phylogeny analysis. The 16S rRNA gene sequences for G22 and BF49 were compared with other Bradyrhizobium sequences in the NCBI database. All Bradyrhizobium 16S sequences from the NCBI RefSeq database and the four Bradyrhizobium isolates from North American forest soils[2] were downloaded. These sequences were aligned using MUSCLE using 8 iterations[55]. The aligned region was extracted (1220 bp) and a phylogenetic tree was created using the neighbour joining clustering method with 1000 bootstraps. A 75% support threshold was used for drawing the phylogenetic tree. Accession numbers for the sequences used in this analysis can be found in Supplementary Information S2.

Core genome phylogenetic analysis. OrthoVenn was used to identify genes which were present in all 9 genomes and were considered the "core genome". The inflation value was set to 1.5 and an e-value of $1e^{-10}$. These genes were uploaded into the call SNPs and infer phylogeny (CSI) tool[56] hosted by the Center for Genomic Epidemiology[57] using the default options using *B. diazoefficiens* USDA 110 as the reference sequence. Default options were as follows: minimum of 10x depth at SNP positions, 10% relative minimum depth at SNP positions, a minimum of 10 bp distance between SNPs, minimum SNP quality score of 30, minimum read mapping quality score of 25 and minimum z-score of 1.96. Altered FastTree was selected. The core genome phylogenetic tree (Newick file) was visualised in Geneious.

Biolog carbon assays. Isolates G22, BF49 and *B. japonicum* USDA 6 were grown in MAG broth, incubated at 28 °C and shaken at 100 rpm. Cell density was estimated from 1 ml of culture stained with 0.05% methylene blue using a haemocytometer. The cultures were diluted to a cell density of 10^6 μl^{-1}. The diluted cultures were centrifuged at 14000 × g for 1 minute, the supernatant was removed and the cells re-suspended in sterile deionised

water. Each well of a Biolog GN2 MicroPlate™ was inoculated with 140 μl (10^8 cells/ 140 μl) of bacterial culture. Three replicate plates per isolate were used. The optical density (OD) was read using a Varioscan SkanIt plate reader (Thermo Fisher Scientific Inc.) at 590 nm and 25 °C every 24 hours for a total of 98 hours (0, 24, 48, 72 and 98 hours). The plates were incubated at 25 °C and shaken at 100 rpm. Full list of substrates can be seen in Supplementary Information S3.

Statistical analyses. The Biolog data was analysed using principal components analysis (PCA) using the inbuilt function, prcomp, in R (version 3.2.2). For visualisation of the PCA, biplots were drawn using PC1 and PC2 and the loadings matrix was extracted (see Supplementary Information S3). This identified specific substrates which were associated with the isolates and also substrates making the greatest contribution to the principal component. These pathways were then examined in the genomes. The substrates were grouped into carbohydrates, carboxylic acids, amines and amides, amino acids, polymers and miscellaneous according to categories identified in previously published work[58].

References

1. Uroz, S., Buee, M., Murat, C., Frey-Klett, P. & Martin, F. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ. Microbiol. Rep*. **2,** 281–288 (2010).
2. Vaninsberghe, D. *et al.* Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME J.* **9,** 2435–2441 (2015).
3. Delmont, T. O. *et al.* Structure, fluctuation and magnitude of a natural grassland soil metagenome. *ISME J.* **6,** 1677–1687 (2012).
4. Zhalnina, K. *et al.* Ca. *Nitrososphaera* and *Bradyrhizobium* are inversely correlated and related to agricultural practices in long-term field experiments. *Front. Microbiol*. **4,** 104–104 (2013).
5. Van Rhijn, P. & Vanderleyden, J. The *Rhizobium*-plant symbiosis. *Microbiol. Rev.* **59,** 124–142 (1995).
6. Giraud, E. *et al.* Legumes symbioses: Absence of *Nod* genes in photosynthetic bradyrhizobia. *Science.* **316,** 1307–1312 (2007).
7. Rivas, R., Martens, M., De Lajudie, P. & Willems, A. Multilocus sequence analysis of the genus *Bradyrhizobium*. *Syst. Appl. Microbiol.* **32,** 101–110 (2009).
8. Okubo, T. *et al.* Complete genome sequence of *Bradyrhizobium* sp S23321: Insights into symbiosis evolution in soil oligotrophs. *Microbes Environ.* **27,** 306–315 (2012).
9. Sachs, J. L., Ehinger, M. O. & Simms, E. L. Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*. *J. Evol. Biol.* **23,** 1075–1089 (2010).
10. Kaneko, T. *et al.* Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res.* **9,** 189–197 (2002).
11. Okubo, T. *et al.* Genome analysis suggests that the soil oligotrophic bacterium *Agromonas oligotrophica* (*Bradyrhizobium oligotrophicum*) is a nitrogen-fixing symbiont of *Aeschynomene indica*. *Appl. Environ. Microbiol.* **79,** 2542–2551 (2013).
12. Hirsch, P. R. *et al.* Starving the soil of plant inputs for 50 years reduces abundance but not diversity of soil bacterial communities. *Soil Biol. Biochem.* **41,** 2021–2024 (2009).
13. Weir, B. S., Turner, S. J., Silvester, W. B., Park, D. C. & Young, J. A. Unexpectedly diverse *Mesorhizobium* strains and *Rhizobium leguminosarum* nodulate native legume genera of New Zealand, while introduced legume weeds are nodulated by *Bradyrhizobium* species. *Appl. Environ. Microbiol.* **70,** 5980–5987 (2004).
14. Bejarano, A., Ramirez-Bahena, M.-H., Velazquez, E. & Peix, A. *Vigna unguiculata* is nodulated in Spain by endosymbionts of Genisteae legumes and by a new symbiovar (vignae) of the genus *Bradyrhizobium*. *Syst. Appl. Microbiol.* **37,** 533–540 (2014).
15. Menna, P., Barcellos, F. G. & Hungria, M. Phylogeny and taxonomy of a diverse collection of *Bradyrhizobium* strains based on multilocus sequence analysis of the 16S rRNA gene, ITS region and *glnII*, *recA*, *atpD* and *dnaK* genes. *Int. J. Syst. Evol. Microbiol.* **59,** 2934–2950 (2009).
16. Menna, P. & Hungria, M. Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer. *Int. J. Syst. Evol. Microbiol.* **61,** 3052–3067 (2011).
17. Peix, A. *et al.* Revision of the taxonomic status of the species *Rhizobium lupini* and reclassification as *Bradyrhizobium lupini* comb. nov. *Int. J. Syst. Evol. Microbiol.* **65,** 1213–1219 (2015).
18. Ferrieres, L., Francez-Charlot, A., Gouzy, J., Rouille, S. & Kahn, D. Fix-regulated genes evolved through promoter duplication in *Sinorhizobium meliloti*. *Microbiology.* **150,** 2335–2345 (2004).
19. Bobik, C., Meilhoc, E. & Batut, J. FixJ: a major regulator of the oxygen limitation response and late symbiotic functions of *Sinorhizobium meliloti*. *J. Bacteriol.* **188,** 4890–4902 (2006).
20. Meilhoc, E., Cam, Y., Skapski, A. & Bruand, C. The response to nitric oxide of the nitrogen-fixing symbiont *Sinorhizobium meliloti*. *Mol. Plant. Microbe Interact.* **23,** 748–759 (2010).
21. Delgado, M. J., Bonnard, N., Tresierra-Ayala, A., Bedmar, E. J. & Muller, P. The *Bradyrhizobium japonicum napEDABC* genes encoding the periplasmic nitrate reductase are essential for nitrate respiration. *Microbiology.* **149,** 3395–3403 (2003).
22. Bedmar, E. J., Robles, E. F. & Delgado, M. J. The complete denitrification pathway of the symbiotic, nitrogen-fixing bacterium *Bradyrhizobium japonicum*. *Biochem. Soc. Trans.* **33,** 141–144 (2005).
23. Fernandez, L. A., Beatriz Perotti, E., Antonio Sagardoy, M. & Anahi Gomez, M. Denitrification activity of *Bradyrhizobium* sp isolated from Argentine soybean cultivated soils. *World J. Microbiol. Biotechnol.* **24,** 2577–2585 (2008).
24. Richardson, D., Felgate, H., Watmough, N., Thomson, A. & Baggs, E. Mitigating release of the potent greenhouse gas N_2O from the nitrogen cycle - could enzymic regulation hold the key? *Trends Biotechnol.* **27,** 388–397 (2009).
25. Henry, S., Bru, D., Stres, B., Hallet, S. & Philippot, L. Quantitative detection of the *nosZ* gene, encoding nitrous oxide reductase, and comparison of the abundances of 16S rRNA, *narG*, *nirK*, and *nosZ* genes in soils. *Appl. Environ. Microbiol.* **72,** 5181–5189 (2006).
26. Shiina, Y. *et al.* Relationship between soil type and N_2O reductase genotype (*nosZ*) of indigenous soybean bradyrhizobia: *nosZ*-minus populations are dominant in andosols. *Microbes Environ.* **29,** 420–426 (2014).
27. Zuber, M., Harker, A. R., Sultana, M. A. & Evans, H. J. Cloning and expression of *Bradyrhizobium japonicum* uptake hydrogenase structural genes in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* **83,** 7668–7672 (1986).
28. Van Berkum, P. Expression of uptake hydrogenase and hydrogen oxidation during heterotrophic growth of *Bradyrhizobium japonicum*. *J. Bacteriol.* **169,** 4565–4569 (1987).
29. Black, L. K., Fu, C. L. & Maier, R. J. Sequences and characterization of *hupU* and *hupV* genes of *Bradyrhizobium japonicum* encoding a possible nickel-sensing complex involved in hydrogenase expression. *J. Bacteriol.* **176,** 7102–7106 (1994).
30. Gregor, J. & Klug, G. Regulation of bacterial photosynthesis genes by oxygen and light. *FEMS Microbiol. Lett.* **179,** 1–9 (1999).
31. Igarashi, N. *et al.* Horizontal transfer of the photosynthesis gene cluster and operon rearrangement in purple bacteria. *J. Mol. Evol.* **52,** 333–341 (2001).
32. Gourion, B. *et al.* Bacterial RuBisCO is required for efficient *Bradyrhizobium*/*Aeschynomene* symbiosis. *PLoS One.* **6,** e21900 (2011).

33. Horecker, B. L., Smyrniotis, P. Z. & Hurwitz, J. Role of xylulose 5-phosphate in the transketolase reaction. *J. Biol. Chem.* **223,** 1009–1019 (1956).
34. Fullam, E., Pojer, F., Bergfors, T., Jones, T. A. & Cole, S. T. Structure and function of the transketolase from Mycobacterium tuberculosis and comparison with the human enzyme. *Open Biol.* **2,** 110026 (2012).
35. Dimroth, P. & Hilbi, H. Enzymic and genetic basis for bacterial growth on malonate. *Mol. Microbiol.* **25,** 3–10 (1997).
36. Kim, Y. S. Malonate metabolism: Biochemistry, molecular biology, physiology, and industrial application. *J. Biochem. Mol. Biol.* **35,** 443–451 (2002).
37. Suvorova, I. A., Ravcheev, D. A. & Gelfand, M. S. Regulation and Evolution of Malonate and Propionate Catabolism in Proteobacteria. *J. Bacteriol.* **194,** 3234–3240 (2012).
38. Kim, Y. S., Kwon, S. J. & Kang, S. W. Malonyl-CoA synthetase from *Rhizobium trifolii:* Purification, properties, and the immunological comparison with those from *Bradyrhizobium japonicum* and *Pseudomonas fluorescens*. *Korean Biochem. J.* **26,** 176–183 (1993).
39. Schmid, M., Berg, M., Hilbi, H. & Dimroth, P. Malonate decarboxylase of *Klebsiella pneumoniae* catalyses the turnover of acetyl and malonyl thioester residues on a coenzyme-A-like prosthetic group. *Eur. J. Biochem.* **237,** 221–228 (1996).
40. Kanehisa, M. & Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **28,** 27–30 (2000).
41. Kanehisa, M., Sato, Y., Kawashima, M., Furumichi, M. & Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **44,** D457–D462 (2016).
42. Li, P. L., Hwang, I., Miyagi, H., True, H. & Farrand, S. K. Essential components of the Ti plasmid trb system, a type IV macromolecular transporter. *J. Bacteriol.* **181,** 5033–5041 (1999).
43. Neumann, M. *et al.* A periplasmic aldehyde oxidoreductase represents the first molybdopterin cytosine dinucleotide cofactor containing molybdo-flavoenzyme from *Escherichia coli*. *Febs Journal.* **276,** 2762–2774 (2009).
44. Xi, H. L., Schneider, B. L. & Reitzer, L. Purine catabolism in *Escherichia coli* and function of xanthine dehydrogenase in purine salvage. *J. Bacteriol.* **182,** 5332–5341 (2000).
45. Kawaharada, Y., Kiyota, H., Eda, S., Minamisawa, K. & Mitsui, H. Identification of the *Mesorhizobium loti* gene responsible for glycerophosphorylation of periplasmic cyclic beta-1,2-glucans. *FEMS Microbiol. Lett.* **302,** 131–137 (2010).
46. Roset, M. S., Ciochini, A. E., Ugalde, R. A. & De Iannino, N. I. The *Brucella abortus* cyclic beta-1,2-glucan virulence factor is substituted with O-ester-linked succinyl residues. *J. Bacteriol.* **188,** 5003–5013 (2006).
47. Van Berkum, P. Evidence for a third uptake hydrogenase phenotype among the soybean bradyrhizobia. *Appl. Environ. Microbiol.* **56,** 3835–3841 (1990).
48. Mauchline, T. H., Hayat, R., Roberts, R., Powers, S. J. & Hirsch, P. R. Assessment of core and accessory genetic variation in *Rhizobium leguminosarum* symbiovar *trifolii* strains from diverse locations and host plants using PCR-based methods. *Lett. Appl. Microbiol.* **59,** 238–246 (2014).
49. Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience.* **1,** 18–18 (2012).
50. Aziz, R. K. *et al.* The RAST Server: rapid annotations using subsystems technology. *BMC Genomics.* **9,** 75 (2008).
51. Overbeek, R. *et al.* The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). *Nucleic Acids Res.* **42,** D206–D214 (2014).
52. Wang, Y., Coleman-Derr, D., Chen, G. & Gu, Y. Q. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res.* **43,** W78–W84 (2015).
53. Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. & Beatson, S. A. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics.* **12,** 402 (2011).
54. Harris, R. S. *Improved pairwise alignment of genomic DNA*. PhD, The Pennsylvania State University. (2007).
55. Edgar, R. C. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics.* **5,** 1–19 113 (2004).
56. Kaas, R. S., Leekitcharoenphon, P., Aarestrup, F. M. & Lund, O. Solving the problem of comparing whole bacterial genomes across different sequencing platforms. *PLoS One.* **9,** e104984 (2014).
57. *Center for Genomic Epidemiology*. (2011) Available at: http://www.genomicepidemiology.org/. (Accessed: 11th November 2015).
58. Preston-Mafham, J., Boddy, L. & Randerson, P. F. Analysis of microbial community functional diversity using sole-carbon-source utilisation profiles - a critique. *FEMS Microbiol. Ecol.* **42,** 1–14 (2002).
59. Kaneko, T. *et al.* Complete genome sequence of the soybean symbiont *Bradyrhizobium japonicum* strain USDA6^{T}. *Genes.* **2,** 763–787 (2011).
60. Torres, D. *et al.* Genome sequence of *Bradyrhizobium japonicum* E109, one of the most agronomically used nitrogen-fixing rhizobacteria in Argentina. *Genome announcements*. **3** (2015).

Acknowledgements

We would like to thank Dr Steve Hanley, Rothamsted Research, for carrying out the Ion Torrent sequencing and Dr Kirsty Hassell, Rothamsted Research, for statistical advice. This study was supported by a University of Reading, Lawes Agricultural Trust and BBSRC DTP studentship and BBSRC (BBS/E/C/00005196). Rothamsted Research receives strategic funding from the BBSRC.

Author Contributions

F.P.J., I.M.C., E.J.S., M.J.W. and P.R.H. were involved in the project conception, initiation and planning. F.P.J. isolated the strains and prepared the samples for sequencing. F.P.J. and R.K. performed the genome assembly and analysis. F.P.J. wrote the main manuscript text. All authors edited and reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Jones, F. P. *et al.* Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes – a genome comparison. *Sci. Rep.* **6**, 25858; doi: 10.1038/srep25858 (2016).