Legume-rhizobia interactions in a complex microbiome

Ming-Dao Chia

Supervised by Prof Justin Borevitz

Word count (Introduction, Results, Discussion): 6016

October 2018

Australian National University

plant energy **biology**
ARC CENTRE OF EXCELLENCE

1

Table of Contents

2

Declaration

I declare this thesis to be my original work, completed for an Honours year project at the Australian National University.

Acknowledgements

List of abbreviations

APPF: Australian Plant Phenomics Facility

BNF: biological nitrogen fixation

gDNA: Genomic DNA

HGT: Horizontal gene transfer

OD600: Optical density at 600 nm

ONT: Oxford Nanopore Technologies

PCR: polymerase chain reaction

YMA: yeast-mannitol agar

YMB: yeast-mannitol broth

Abstract

Biological nitrogen fixation is important for agriculture, carbon sequestration, and ecosystem restoration. This is primarily conducted by rhizobia (nitrogen fixing bacteria) in association with legume plants.

Most research in improving rhizobial strains involves single strain experiments. However, improved metagenomics methods have demonstrated considerable differences between single strain inoculations and strain behaviour when exposed to a complex microbiome.

To identify some these differences, this experiment applies two treatment factors in a controlled environment of containers with autoclaved sand. All main experimental containers were inoculated with several strains of *Bradyrhizobia japonicum*. The first treatment factor was the planting of surface-sterilised seedlings of host plant *Acacia acuminata*; the second treatment factor was inoculation with an external soil microbiome. Several negative controls without planting or inoculation were also present.

A novel method of whole genome metagenomic sequencing to observe known strain abundance, without amplification or culturing, was developed. Using this method, abundance patterns of these *B. japonicum* strains were compared between initial inoculation and the end of a growth period of several weeks. Analysis reveals a single strain as the preferred nodulation strain within this experiment, but also shows that all strains inoculated continued to persist in the substrate at detectable levels.

The use of long reads with the MinION DNA sequencer also allowed the potential of identification of horizontal gene transfer events. None were detected in an initial screen, but a framework for further inspection of this dataset for such events is described.

1. Introduction

Nitrogen fixation is important for a considerable amount of global ecological and economic processes. Most anthropomorphic bioavailable nitrogen is produced industrially via the Haber-Bosch process, currently responsible for over a third of the global nitrogen cycle (Gruber and Galloway, 2008). The Haber-Bosch process has been recently re-estimated to be responsible for the nitrogen needs of 48% of humans (Erisman et al., 2008) and has been essential for human nutrition (Wood and Baudron, 2018). However, the Haber-Bosch process is highly energy intensive, and, as an industrial product, is not equally distributed throughout the world; a considerable proportion of its use has been directed towards intensive agricultural production of meat and dairy in wealthier countries (Erisman et al., 2008).

Although increased anthropomorphic nitrogen has contributed towards some carbon sequestration (Gruber and Galloway, 2008), increased atmospheric carbon has resulted in a corresponding increase of nitrogen limited ecosystems, with associated consequences on ecosystem services and production of certain systems, such as reduced protein from rangeland cattle (Craine et al., 2018). It has been postulated that nitrogen may be a limiting factor for carbon sequestration (Gruber and Galloway, 2008). As such, enhancing terrestrial biological nitrogen fixation (BNF) may be a vital part of mitigating climate change.

As part of efforts to mitigate anthropomorphic climate change and reduce inequality, some authors have proposed the concept of a global "safe operating space" with maximum limits on human perturbation of earth system effects, including the nitrogen cycle (Häyhä et al., 2016). Enhancing BNF is important to maintaining the safe operating space of the nitrogen cycle (Vitousek et al., 2013).

BNF is primarily carried out by the rhizobia-legume system, but not exclusively so. Some non-legume plants can also carry out BNF, such as actinorhizal plants and

Frankia (Santi et al., 2013) as well as a recently confirmed indigenous landrace of maize (Van Deynze et al., 2018). BNF is also carried out by cyanobacteria in oceanic environments (Capone et al., 2005). Some attempts at harnessing BNF as a sustainable alternative to the Haber-Bosch process include the use of electrical bioreactors (Rago et al., 2019)

In this experiment, the legume *Acacia acuminata* and strains of its rhizobial symbiont *Bradyrhizobium japonicum* are used as a model system for methodological development.

Acacias

Australian acacias are vital for ecosystem restoration, as legumes that can grow in nitrogen-poor soils common in Australia (Thrall et al., 2007). The raspberry jam wattle *Acacia acuminata* has a home range in south Western Australia, and is an important host of economically valuable sandalwood (*Santalum spicatum*), with a market price up to AUD 15,000 per tonne of wood and AUD 1,100 per kg of oil (Environment and Public Affairs Committee, 2014; Woodall et al., 2002)

Legume hosts, including *A. acuminata*, form a mutualistic relationship with rhizobia. Hosts provide sugars to feed and grow rhizobia; in exchange, rhizobia fix atmospheric nitrogen and provide it to the plant.

Acacias can also provide nitrogen to Eucalyptus trees through mycorrhizal fungi networks, further supporting a potential for carbon sequestration and rhizobia use in supporting agriculture beyond legumes (Paula et al., 2015)

Rhizobia

Rhizobia are Gram-negative bacteria that grow either free-living in soil or in symbiosis with a legume host during their life cycle. Although many rhizobial symbionts are fairly well studied, most of what we know about them in controlled environments is from

experiments involving single strains. Some work indicates results from such experiments may not be generalisable to more complex environments with multiple rhizobial strains present (Kiers et al., 2013)

Previous work in this area includes work in the *Medicago-Ensifer* system, inferring strain fitness under different treatments (Burghardt et al., 2018). Observations of wild samples indicate that legumes often have multiple strains within individual plants (Dinnage et al., 2018).

Not all rhizobia found in an environment are necessarily well adapted to that environment. Some strains of rhizobia are less capable of performing nitrogen fixation in certain conditions, but may persist through host nutrient investment even if they would normally disadvantage a host (Wendlandt et al., 2018)

Rhizobia are often studied exclusively in nodules or in pure culture form (Burghardt et al., 2018). As rhizobia can be difficult to isolate directly from soil, soil metagenomics allows us to observe rhizobia in a complex soil environment.

Microbiomes

An increasing body of research indicates that microbiomes can have considerable effects on host organisms (Zilber-Rosenberg and Rosenberg, 2008), including plants (Vandenkoornhuyse et al., 2015). For example, overall microbiome composition can enhance disease resistance in plants (Chen et al., 2018; Kwak et al., 2018). A variety of positive effects from microbiome bacteria have been proposed (Goh et al., 2013)

Microbiomes can also have effects on microbial symbionts that are part of the microbiome. One such effect is that adaptation to abiotic factors (e.g. drought, nutrients) within a strain population can reduced by other bacteria in the microbiome (Hall et al., 2018). This may have considerable effects on the application of strains optimised in lab environments; strains developed for a particular purpose via soil inoculation may not

maintain their effects if such strains are not tested in exposure to a complex environmental microbiome.

Rhizosphere microbiome diversity is often reduced in areas of increased human activity, such as intensive agriculture and signified by effects such as metal contamination (Gans et al., 2005). Rhizosphere microbiome engineering with microbial consortia has been proposed to restore soil microbiome services in line with newer agricultural practices (Wallenstein, 2017).

Despite advances in culture methods for soil bacteria (Manh Nguyen et al., 2018), the majority of microbes in soil are still not culturable with conventional culture methods. Soil metagenomics aims to resolve this through extracting DNA from microbial cells in a soil sample and sequencing it. This can either be done with a known variable sequence flanked by conserved sequence for amplification (e.g. 16S for bacteria, ITS for fungi) or through shotgun metagenomics, which sequences all DNA rather than specific amplicons. Shotgun metagenomics can offer insights into the abundance of genes and assist in inferring community function, but is usually worse at identifying the community composition (particularly with short read sequencing) (Leveau, 2007; Tringe et al., 2005).

Several sources of potential bias are common in metagenomic sequencing. This can occur with sample collection, which is often done in small sample tubes, observing only a small portion of heterogeneous soil (Mocali and Benedetti, 2010). It can also happen in the analysis stage, with inappropriate application of statistical methods often used in other complex communities (McMurdie and Holmes, 2014)

Long read sequencing

Most metagenomic sequencing is conducted with Illumina sequencing by synthesis. This is referred to as "short read" sequencing, as read lengths are up to 300 bases long (Illumina, 2018). Short reads are usually assembled into contigs, but are often not long

enough to resolve repetitive regions of genomes, a problem exacerbated in metagenomes (Forouzan et al., 2018).

New sequencing systems increase read lengths considerably. One main technique is the PacBio sequencing system, with read lengths >200kb, enough to acquire entire complete 16S sequences for metagenomics (Pacific Biosciences, 2018; Pootakham et al., 2017). An alternative long read sequencing technology is the Oxford Nanopore MinION, with even longer read lengths up to 2mb, and a variety of other benefits, being portable and having a faster turnaround time, allowing for adaptive experiments (Pomerantz et al., 2018). Both systems have higher error rates than Illumina sequencing, particularly when sequencing long homopolymer regions; as a result, both long and short reads are often combined to increase accuracy (Loman et al., 2015). This is not limited to single strains, but can be done in metagenomic samples (Bankevich and Pevzner, 2018). Long read sequencing has also been applied to identify horizontal gene transfer (HGT) events (Ashton et al., 2015).

Genome assembly

De novo assembly of genomes with short read shotgun sequencing often results in fragmented genome assemblies. These can make it hard to align other sequences, and determine gene position in chromosomes (Thomma et al., 2016). The application of long read sequencing allows for scaffolding short read contigs to form complete genomes.

Aims

The primary aims of this project are to:
1. Observe abundance patterns of *Bradyrhizobium japonicum* within nodules of host *Acacia acuminata* when exposed to a mixture of closely related strains, and compared with additional exposure to a complex microbiome.
2. Observe abundance patterns of free-living B. japonicum in soil outside a host, both in the presence or absence of a host plant.

3. Develop a pipeline to identify potential horizontal gene transfer events from metagenomic reads, using the full genomes of added strains.

As part of this, several milestones were identified at the start of the project:

1. Compare long and short read data of the same samples.
2. Compare rhizobia and other microbes in different samples.
3. Identify HGT events in one or more metagenomic reads.

Several results were hypothesised for these aims.

Figure 1. Simplified example of expected results. Numbers represent proportions within the heatmap. In this example, both strains are in equal abundance in the inoculation solution. Strain 1 is more abundant in nodules, and Strain 2 is more abundant in sand.

Sequencing the initial pure culture mix should show equal abundances of each strain, if the mixing protocol works as intended. From sequencing results after the growth period, some strains are expected to be more abundant in nodules, and other strains are expected to be more abundant in a free-living state.

HGT events are likely to occur between closely related strains (Ravenhall et al., 2015), such as the ones with hybrid assembled genomes in this experiment, and also carry features such as flanking regions or base composition which may enable them to be identified from single reads (Hendrickson et al., 2018; Tsirigos and Rigoutsos, 2005).

2. Methods

Methods listed here with considerable modifications from published protocols are also available online (see Supplementary Materials).

Rhizobia pure culture

8 rhizobial strains were transferred from frozen stock to yeast-mannitol agar (YMA) plates (strains were previously cultivated in yeast-mannitol media). These were incubated for 7 days in a 30ºC incubator, with no light.

Table 1. Strains initially provided for this experiment. Note that Illumina and long read columns here refer to sequencing conducted prior to this experiment; refer to Table 2 for final strains successfully resequenced and used in inoculation.

Strain ID	Illumina	Long read	platform
25.1	yes	yes	PacBio
38.8	yes	yes	PacBio
41.2	yes	yes	PacBio
65.7	yes	yes	PacBio
182.5	yes	yes	PacBio
36.1	yes	no	none
36.8	yes	no	none
213.6	yes	yes	ONT

Strains with visible colonies were transferred to yeast-mannitol broth (YMB) for culture amplification and left for at least 7 days. Visibly turbid broth was considered as a successful culture. One culture was later identified as a contaminated *Bacillus* culture from sequencing results, and consequently excluded from the culture mix.

To form the culture mix, each culture used was diluted to 0.2 absorbance at 600nm and mixed in equal volumes before being further diluted to form the culture mix inoculum.

Strain DNA extraction

Successful YMB cultures for strains lacking a previous complete genome assembly were pelleted and extracted using a magnetic bead based extraction. This is based on Schalamun et al. modification of the Mayjonade et al. protocol for high molecular weight gDNA extraction, but excluding the further chloroform-isoamyl alcohol cleanup as this was not needed for pure bacterial cultures (full protocol linked in Supplementary Materials) (Schalamun et al., 2018). DNA extractions were checked for quality via gel electrophoresis (assessment of fragment size and RNA contamination), Qubit fluorometer (higher accuracy DNA quantitation), and Nanodrop spectrophotometer (other contaminants). The gDNA library was then prepared and sequenced with the Oxford Nanopore Technologies (ONT) MinION system according to ONT protocols.

Some phenotypic differences between strains may have resulted in differences in DNA quality, as shown in differences in NanoDrop results (Fig 1).

Strain de novo genome assembly

MinION raw data was basecalled with Scrappie and Albacore. This was combined with Illumina reads acquired before this project in a hybrid assembly using Unicycler (Wick et al., 2017). Flye (Lin et al., 2016) was also used in assembly, yielding a genome with fewer contigs, particularly with nanopore-only assemblies. Unicycler was used for all final hybrid assemblies due to integrated circularisation and additional polishing steps. However, not all assemblies were finished, most likely due to insufficient long read coverage (Table 1).

Soil slurry preparation

Soil samples were acquired from the base of an *A. acuminata* accession in the Australian National Botanic Gardens in 15ml tubes. Water was added to the tubes and vortexed to create a soil slurry suspension. 1ml of supernatant was removed from each tube and diluted to make the soil slurry inoculum.

Growth curve observation

Several strains appeared successful (based on turbidity) in initial broth culture. These strains were subcultured in duplicate in 50ml Falcon tubes, along with a blank control. Absorbance was measured every 24 hours, starting at 48 hours post inoculation. At each time point, 500µl of culture was removed and diluted with 1500µl YMB (to accommodate an optimal loading volume) and absorbance measured at 600nm in a Varian Cary 50 UV-Vis Spectrophotometer. Data was converted from a Varian report format and loaded into R with a custom script (see Supplementary Materials).

Soil DNA extraction

Prior to DNA extraction, each pot was homogenised by taking a 25ml soil sample, vortexing with MilliQ water, and filtering by Miracloth. This was followed by an initial centrifugation step for sand samples to remove water.

To avoid DNA extraction bias, allsequenced inoculation, nodule, and soil samples (except the pure broth cultures) had DNA extracted using the Qiagen PowerSoil kit, including the concentrated culture mix and soil slurry inoculum, the autoclaved sand substrate, crushed nodules, and each original soil sample. PowerSoil kits were chosen due to their reputation for high recovery with low contamination and ease of use (Lear et al., 2018). Kits were used mostly according to manufacturer protocol, but DNA elution times were extended to 20 minutes instead of 5 minutes, as per ONT recommendations. To maximise yield, the full volume from the PowerSoil kit (~100µl/sample) was used in the library preparation. The library was later concentrated to the volume required for MinION library loading, via magnetic bead cleanup after index addition.

Experimental design

8 treatments were used overall. Host effects were applied by either the presence or absence of an *A. acuminata* plant. Two primary inoculations were used; the soil slurry and culture mix (see sections above). A third inoculation treatment applied both

inoculations together in the same pot. As a control for contamination, some pots had no inoculation treatment applied.
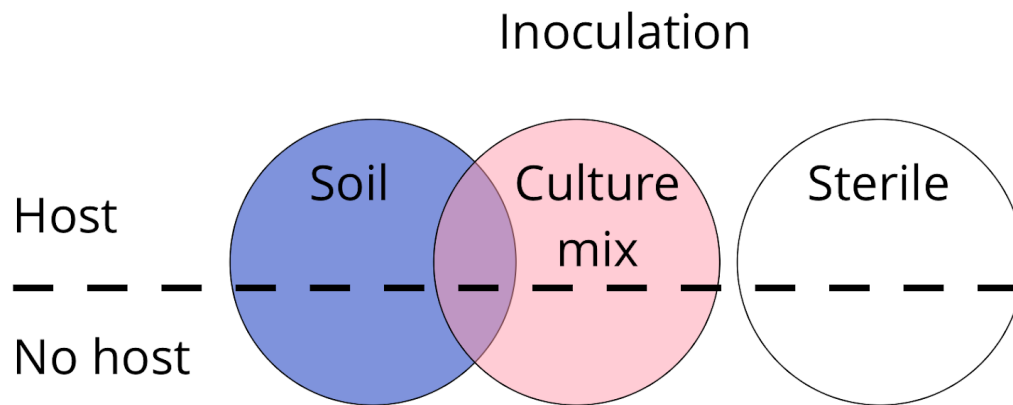
Figure 2. Schematic of treatment categories applied. Soil refers to soil slurry inoculation, and not the substrate; all plants were grown in autoclaved sand.

Table 2. Treatments used. Soil+Culture refers to adding equal quantities of soil slurry solution and rhizobial culture mix. Treatment 5 (host absent, no inoculation, control) has two less pots to accommodate for color checkers, and was deemed to be the least necessary treatment, to replace for that purpose.

Treatment designation	Inoculation	Host presence	Amount of pots
1	None (sterile)	Yes	16
2	Soil slurry	Yes	16
3	Culture mix	Yes	60
4	Soil+culture	Yes	60
5	None (sterile)	No	14
6	Soil slurry	No	16
7	Culture mix	No	60
8	Soil+culture	No	60

Acacia germination

Acacia acuminata seeds were purchased online from Nindethana Seed Company. Seeds were surface-sterilised with ethanol and bleach before soaking in MilliQ water overnight. This was followed by incubation on 1% agar at 30ºC for 2 days.

Acacia growth and maintenance

Successful germinants (visible radicles) were transferred to autoclaved sand in standard APPF 5cm pots. To avoid cross contamination from watering, bottom trays were removed; pots were directly placed on a rack over the chamber drainage system. Autoclaved McKnight's solution (a nitrogen-free nutrient solution) was prepared bi-weekly from 200x stock solution. Plants were top-watered daily over the duration of the experiment with 10ml of solution. Water was delivered via serological pipette in order to minimise contamination.

A total of 300 pots across 12 trays were used, of which 152 had plants. Treatments were randomised within trays with a custom Python script (see Supplementary Materials). Lighting, humidity, and temperature were adjusted to simulate springtime conditions within the home range of *A. acuminata* in Western Australia (home range based on data from the ALA (Atlas of Living Australia, 2018)) with conditions provided by the APPF using pysolarcalc (https://github.com/borevitzlab/pysolarcalc), specifically at GPS coordinates -37.0, 117.0 and starting from a simulated date of October 1st, 2018. (full solarcalc file linked in Supplementary Data)
As part of the APPF growth chamber system, photos were taken at regular intervals.

Sample DNA sequencing

Sequencing was done by trays, with each tray being prepared for one sequencing run. Given that plant ID ordering was already randomised, the first (by ID) suitable pot was prepared for sequencing. Suitable pots were pots that did not have visible substrate

abnormalities (particularly fungal or other visible microbial contamination, and errant plant matter that was missed in initial sand transfer and decontamination), and if it was a host-present treatment, the pot had to have a visibly growing plant. Initial experiment design called for 96 samples, but only 24 of these were completed due to technical difficulties.

After 5 weeks, plants were gently removed from each pot with forceps and visually inspected for root nodules. Nodules were cut from roots with a scalpel and pooled by pot in 1.5ml sample tubes. Pooled nodule samples were surface cleaned with 70% ethanol and sodium hypochlorite. Cleaned nodules were crushed with a glass pestle immediately prior to DNA extraction.

All instruments were decontaminated between samples by a 70% ethanol wash and 100% ethanol dip followed by exposure to a flame.

Assessment of microbial diversity

Microbial community analysis often requires a large database of microbial genomes. An online resource, the EDGE bioinformatics server, was used due to local file storage limitations. Selected samples were run through the EDGE bioinformatics server, using BWA-MEM mapping and Kraken with Nanopore default settings (Li et al., 2017)

Strain abundance measurement

MinION raw data was basecalled with Albacore. Sample demultiplexing and barcode/adapter removal was conducted with Porechop. Reads were filtered to target strain genomes through K-mer analysis using Readfilter (Ng et al., 2018) with Kraken (Wood and Salzberg, 2014). These filtered reads were passed through the DiTASic pipeline (Fischer et al., 2017), involving the usage of a generalised linear model to resolve similarities between genomes and estimate actual abundance from read counts. Abundance estimates for each sample were then further processed in R, including basic normalisation (conversion to proportion of total abundance estimates in the sample) and grouped by treatment. In treatment grouping, Edgington's method for combining

p-values (Edgington, 1972) was used, as DiTASic often assigns zero p values and no particularly large values of p are observed. The mean of abundance values was used in combining each abundance estimate. For combining variance, the Satterthwaite approximation was applied through the R package limma (Ritchie et al., 2015).
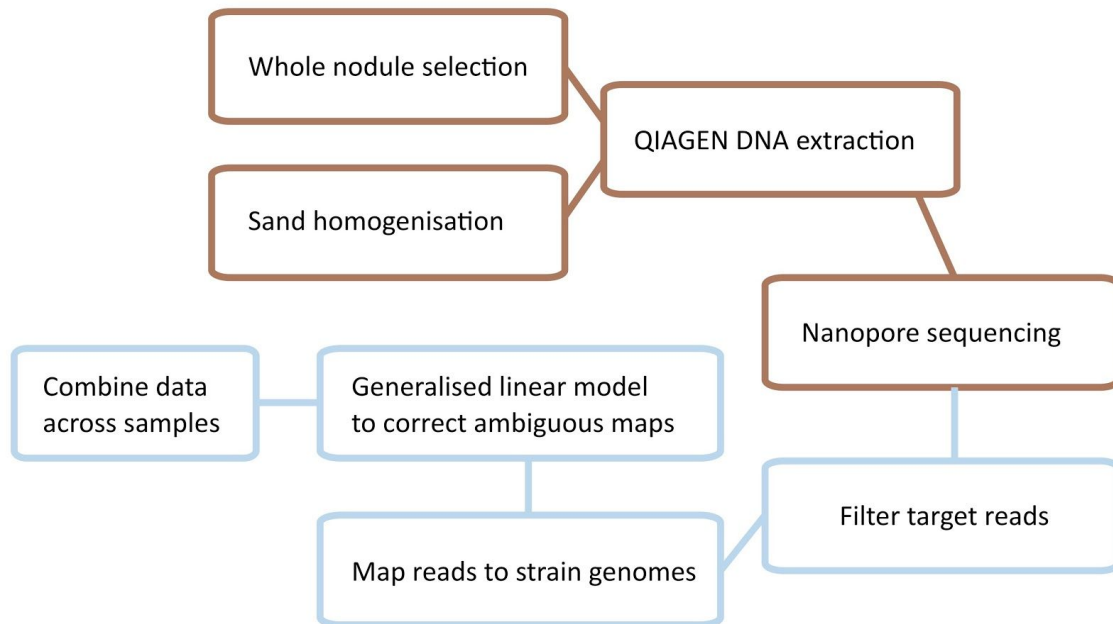
Figure 3. Overall schematic of data collection and abundance measurement, starting with either sand homogenisation or whole nodule selection.

Read simulation

As a control for bias in the bioinformatics pipeline, an equal mix of reads in the same distribution were simulated from all 4 genomes using Nanosim-H (Yang et al., 2017). Considering that there was a considerable difference in absolute number of reads between samples, simulations were conducted at 100-100,000 total reads with equal amounts of reads between samples (see Supplementary Materials for simulation parameters). To ensure that the model controlling for ambiguous read assignment worked as intended with these genomes, simulation of 10,000 reads from each genome was also conducted. These simulated readsets were passed through the same pipeline.

3. Results

Pure colony DNA extraction
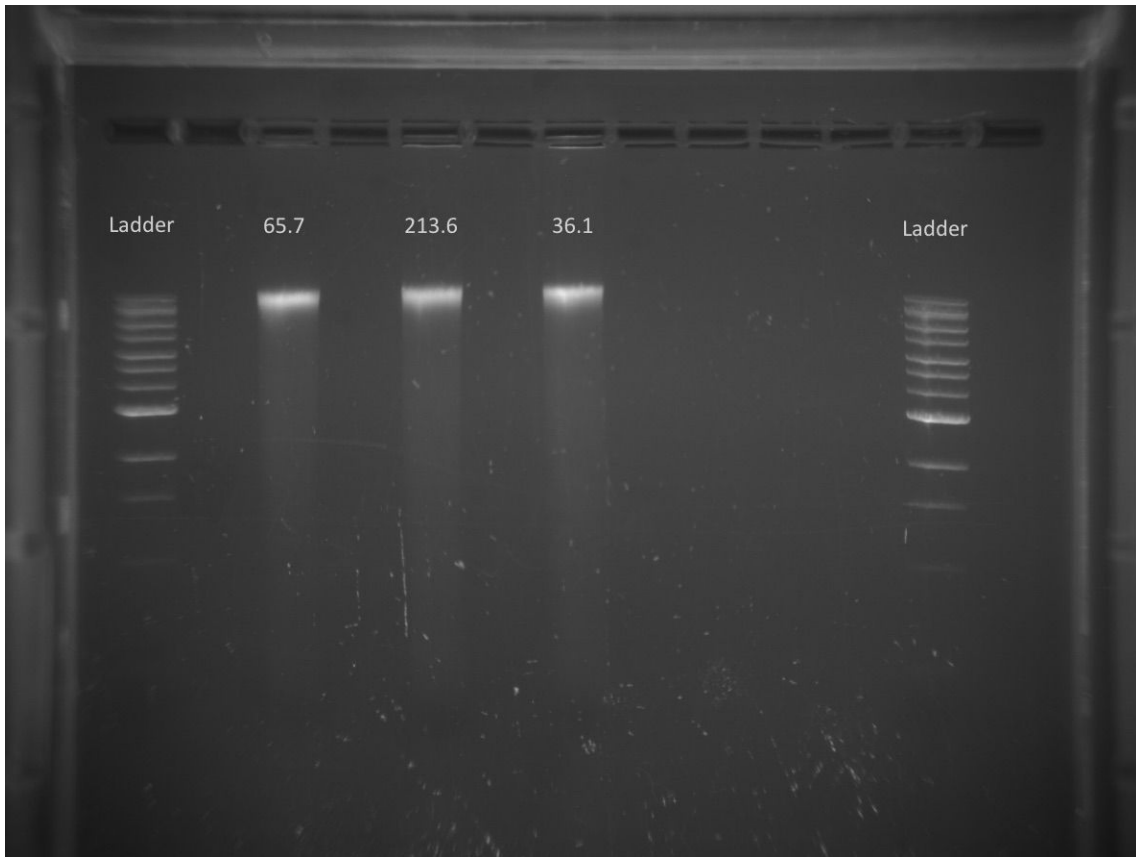
Figure 4. Gel electrophoresis in 1% agarose of initial gDNA extraction, with extraction optimised for long reads. Lane labels refer to *B. japonicum* strain IDs.

Several DNA extraction methods were evaluated, as gDNA MinION sequencing libraries generally require a large mass of relatively pure DNA compared to Illumina libraries or PCR-based systems (Schalamun et al., 2018). The modified bead-based protocol (see Methods) resulted in the usable DNA in Fig. 3. Nanodrop results also indicated low contaminants, suitable for MinION sequencing (Fig. 5). Quality control tests for other extractions were primarily conducted to confirm extraction success prior to sequencing (see links in Supplementary Materials).

Given the aim of using long reads for HGT and to maximise genome completeness through scaffolding Illumina data, a few attempts were made at using the Josh Quick protocol for ultra-long reads (Quick, 2018). This other method, however, did not consistently yield enough DNA for sequencing.
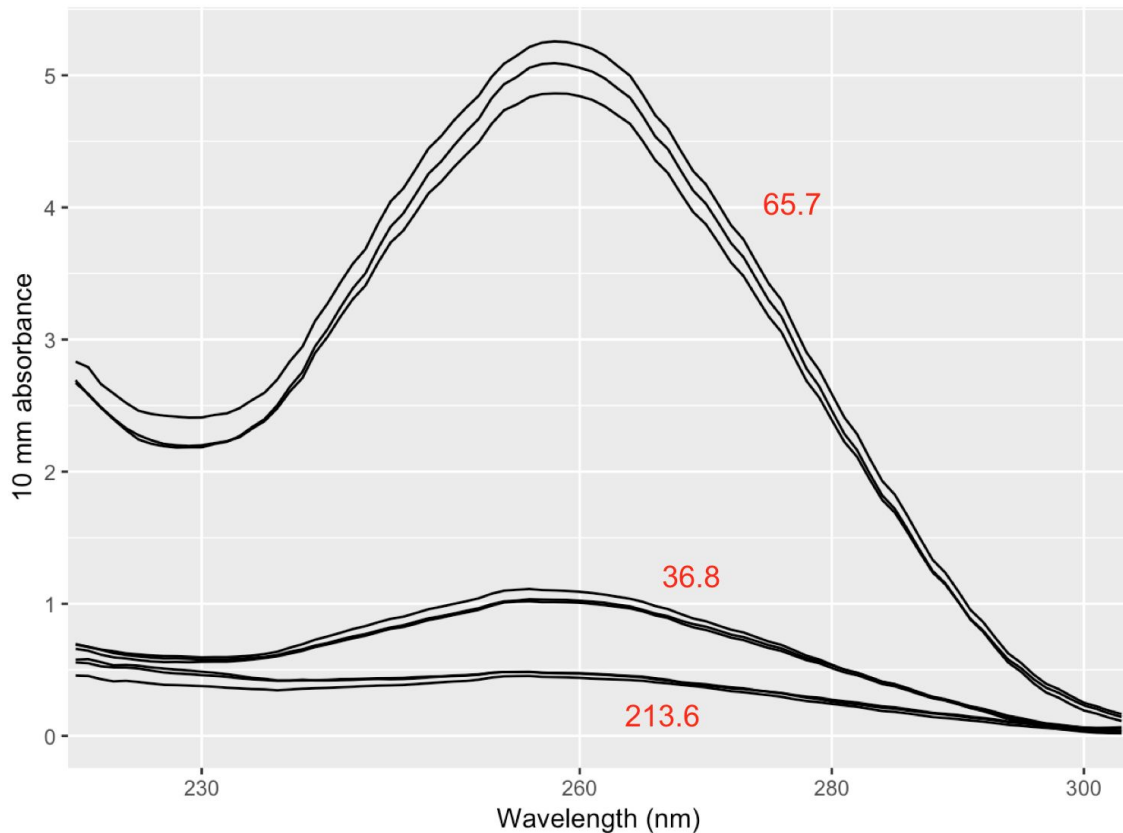
Figure 5. Nanodrop results for a set of DNA extractions conducted in triplicate for each strain. Note the groupings of three curves per strain; the top three curves are for strain 65.7, the middle three curves for strain 36.8, and bottom three curves for strain 213.6.

Genome assemblies

All 8 strains were previously sequenced with Illumina. Genome assemblies combined both Illumina and ONT reads, but low coverage from ONT resulted in incomplete assemblies with both Unicycler and Flye. Per-base accuracy should be maintained with high Illumina coverage (approximate 2Gb Illumina data for each strain).

Table 3. Genome assemblies used for strain abundance assignment. Annotation was performed by the RAST server (Aziz et al., 2008; Overbeek et al., 2014). Note the much higher number of coding sequences for the complete assembly of 36.1. More contigs indicates a fragmented assembly; increasing sequencing coverage reduced fragmentation. Similar GC content and RNAs is likely indicative of accurate, high-coverage short read data.

B. japonicum strain	Size (bases)	GC content (%)	Contigs	Subsystems	Coding sequences	RNAs	Nanopore bases	Nanopore coverage (x genome)
36.1	8112753	63.2	1	503	14513	52	1,161,226,230	143.1
36.8	7885263	63	2	372	7970	53	332,025,880	42.1
65.7	7058791	63.1	51	366	7174	51	29,206,276	4.1
213.6	7943014	63.6	6	375	8081	53	62,241,391	7.8

Growth curve

Initial estimates based on visible turbidity for strain growth projected negligible growth within the first 48 hours; no initial baseline data points were taken. Additionally, the earliest data points taken within the first 48 hours were incorrectly measured due to human error. As a result, no accurate estimation of doubling time could be calculated for comparison between strains.
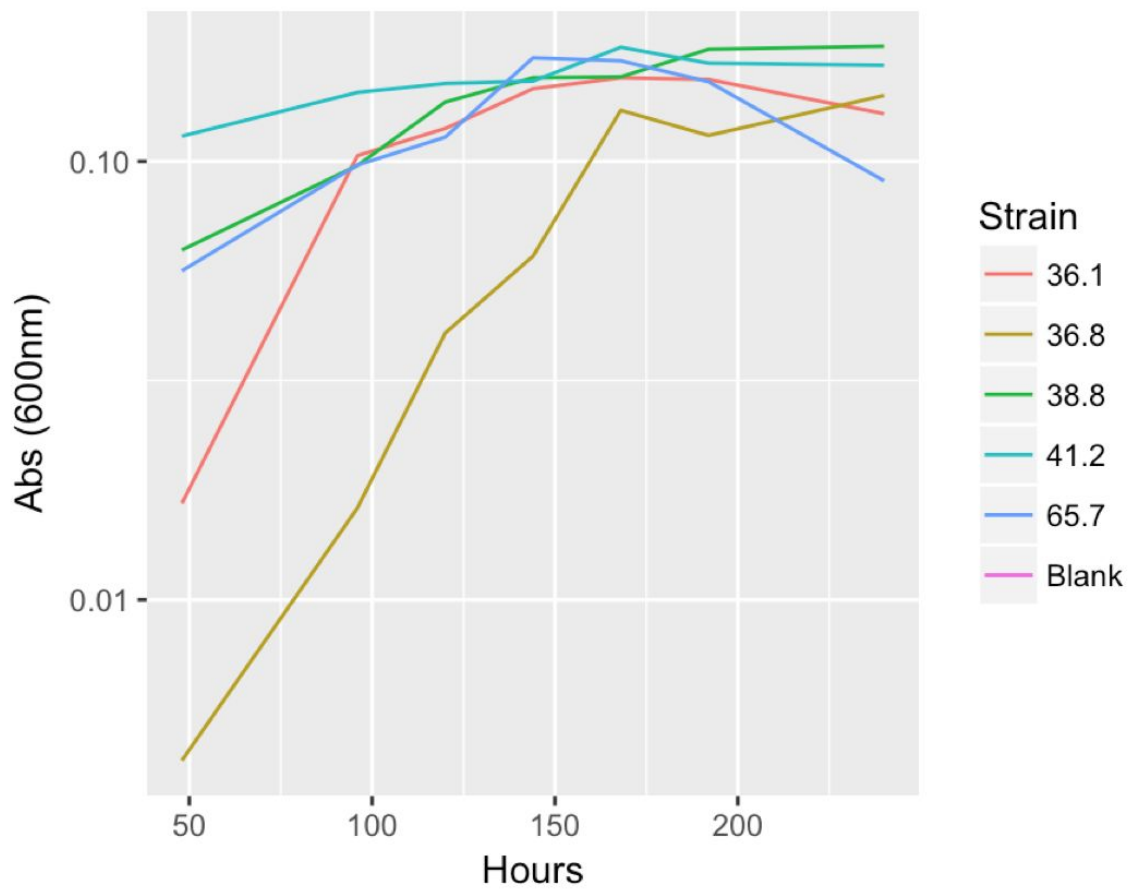
Figure 6. Absorbance over time for each strain. Note the different starting points due to invalidation of earliest results.

Colony phenotype

Rhizobia cultured on identical media and growth conditions showed considerable qualitative difference in colony morphology.
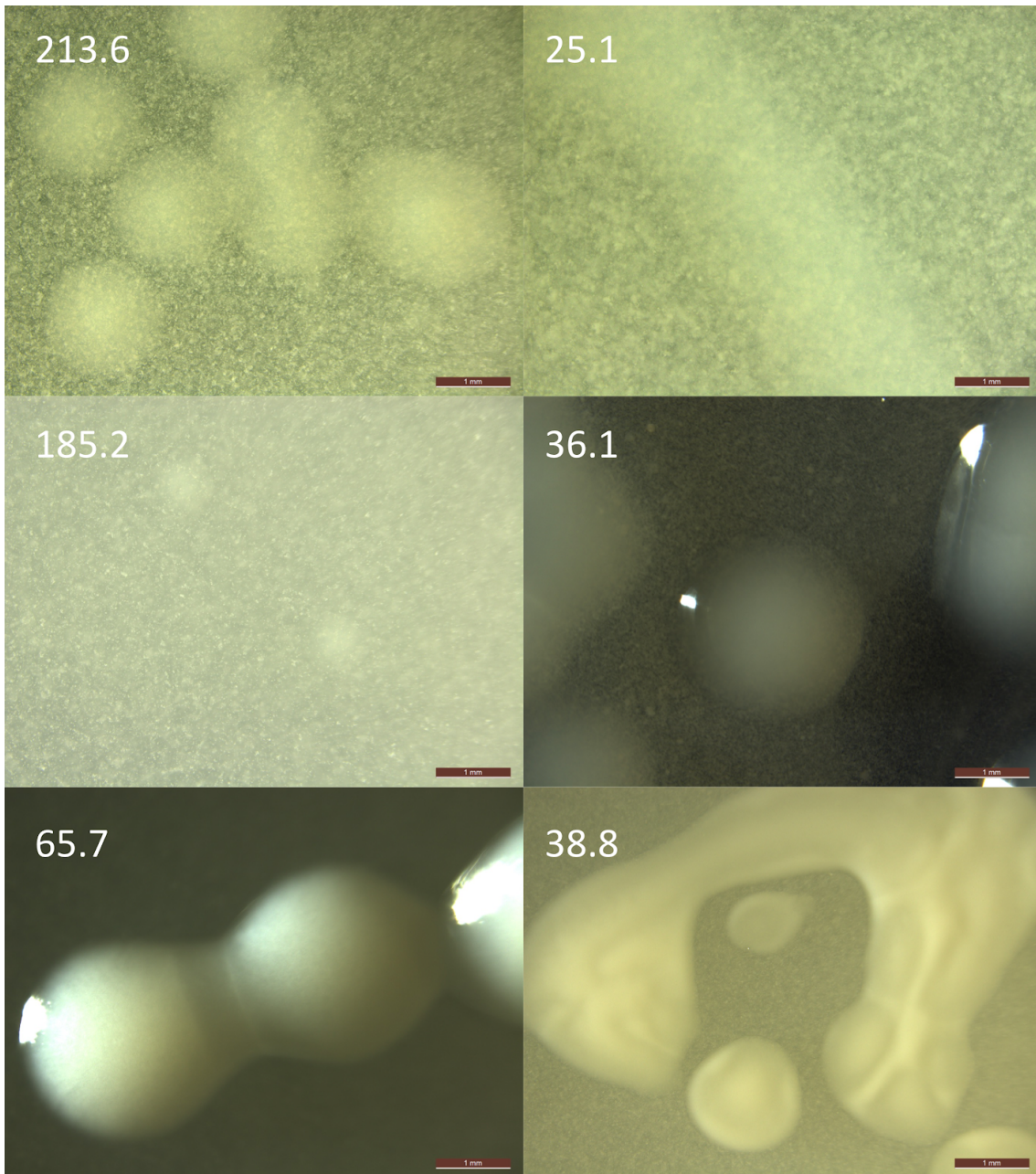
Figure 7. *B. japonicum* colony morphology, with associated strain IDs. Note that all plate agar media had visually similar opacity and were taken from the same batch of media, but increased exposure was needed on smaller colonies to allow imaging. The smallest dots visible are undissolved components of the media used.

Plant phenotype

Almost all agar plates of germinated seedlings appeared to have some contamination of colonies of an unidentified microbe. Only seedlings not on the contaminated areas of

plates were planted. Experimental planning accommodated for this, as a 90% germination failure rate was expected in the worse case and the quantity of seeds adjusted appropriately. Only half of the germination plates were needed as a result.
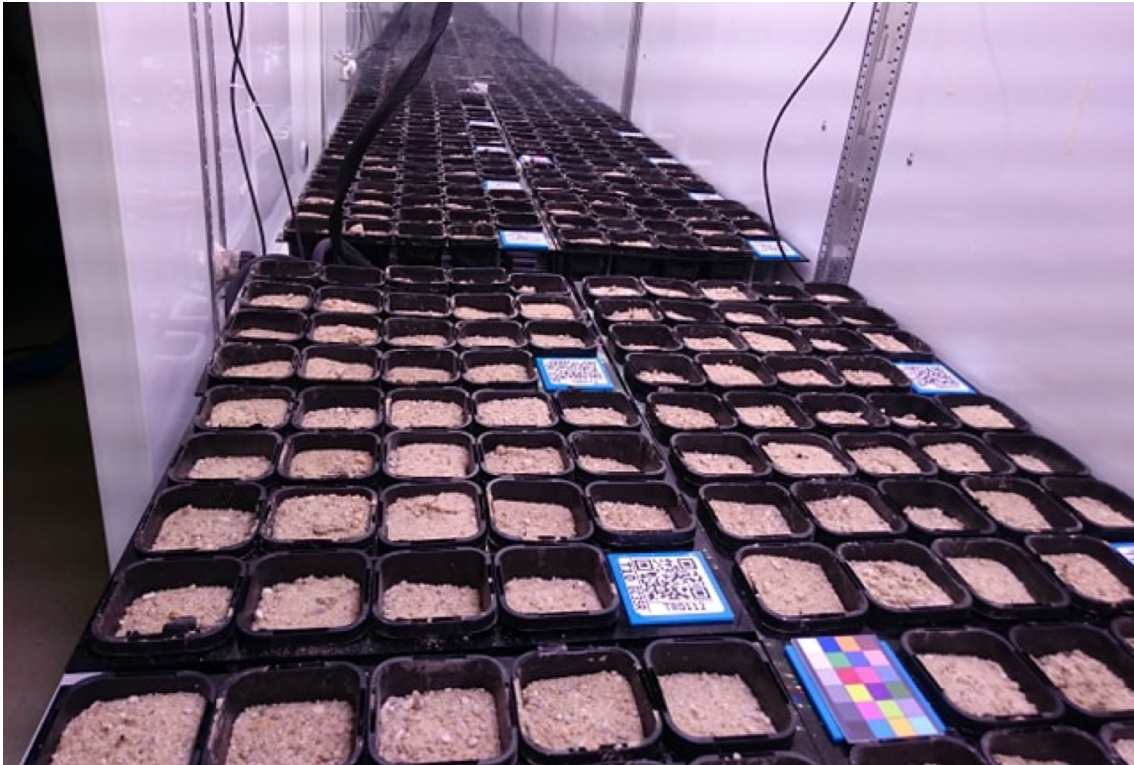
Figure 8. Chamber used for the experiment, several days after planting. At this early stage, no obvious emergence is visible.

Plants were imaged regularly in the APPF chambers. Several issues resulted in loss of early capture, preventing observation of emergence time; no further image analysis was conducted, although a leaf folding circadian rhythm was visible.
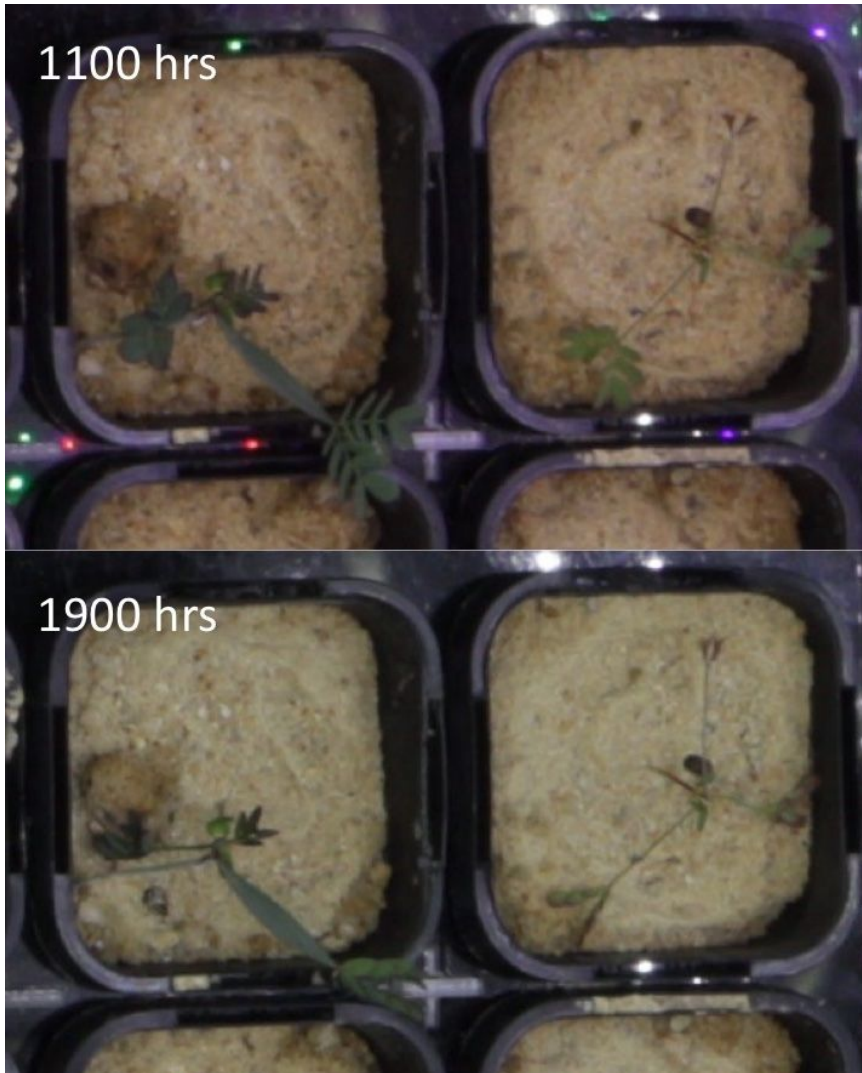
Figure 9. Two plants, imaged at 1100 hrs (top) and 1900 hrs (bottom) on the same day at about 5 weeks after planting. Note the folded leaves in the second image.
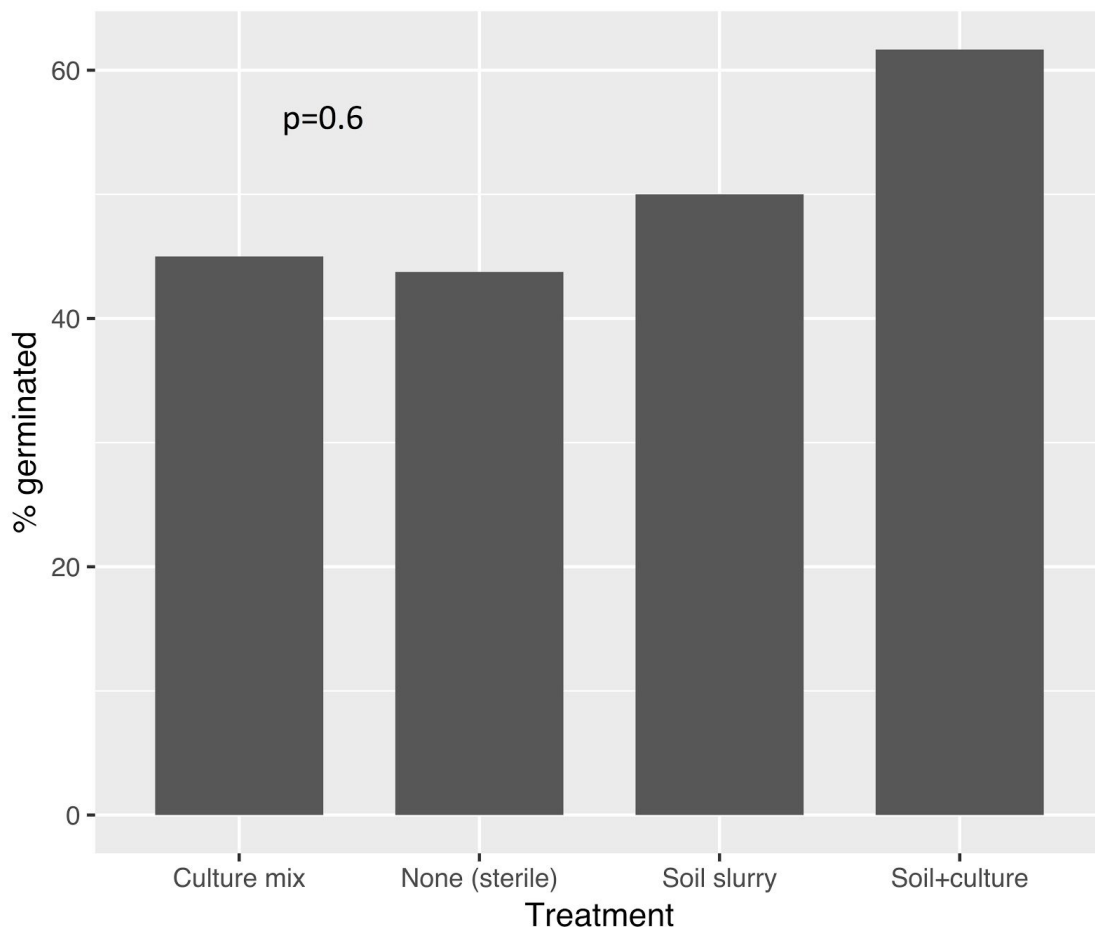
Figure 10. Percent germinated by treatment, taken at week 4. No significant difference between counts found with chi-sq test.

Emergence was manually taken several weeks into the experiment. Soil slurry provided a slight but statistically insignificant increase of emergence.

DNA sequencing

Previous attempts at barcoding with ONT standard kits resulted in around 30% reads that could not be demultiplexed (Borevitz Lab, unpublished). In the first run for this experiment, 23.7% of reads accounting for 18.5% of total bases could not be demultiplexed. In the second run, 45.0% of reads accounting for 35.5% of total bases could not be demultiplexed.

In total, 248,863 reads were successfully demultiplexed, accounting for 474 Mb of sequencing data. In comparison to BASE (Biomes of Australian Soil Environments) recommendations for soil samples, only two of 24 samples had more than 20,000 reads, but 19 of 24 exceed the 2,000 reads needed to preserve between-sample differences (Bissett et al., 2016). BASE recommendations are intended for short read amplicon data; most reads in this dataset would carry more information allowing for species or strain identification than typical short reads (Benítez-Páez and Sanz, 2017).

No filtering to remove host reads was done, as there was no available full *A. acuminata* genome at the time, and the DNA extraction was optimised for microbial DNA. Given the high proportions of reads mapping to bacteria, it is unlikely that host DNA contamination had a considerable effect on the results, particularly given the already high variability in absolute read yield between samples.

Although some work indicates that dead microbes may form a considerable proportion of microbiomes (Emerson et al., 2017), this risk is reduced with a short experiment starting with autoclaved soil (which was sequenced and yielded almost no reads), and is further mitigated by not using amplification, and with the relatively high requirements for DNA structural integrity for ONT gDNA sequencing.

Complex community composition

A considerable amount of the soil slurry inoculum contained *Bradyrhizobium* strains, which likely accounts for the low level of unique maps to strain genomes, possibly from more conserved sequences.
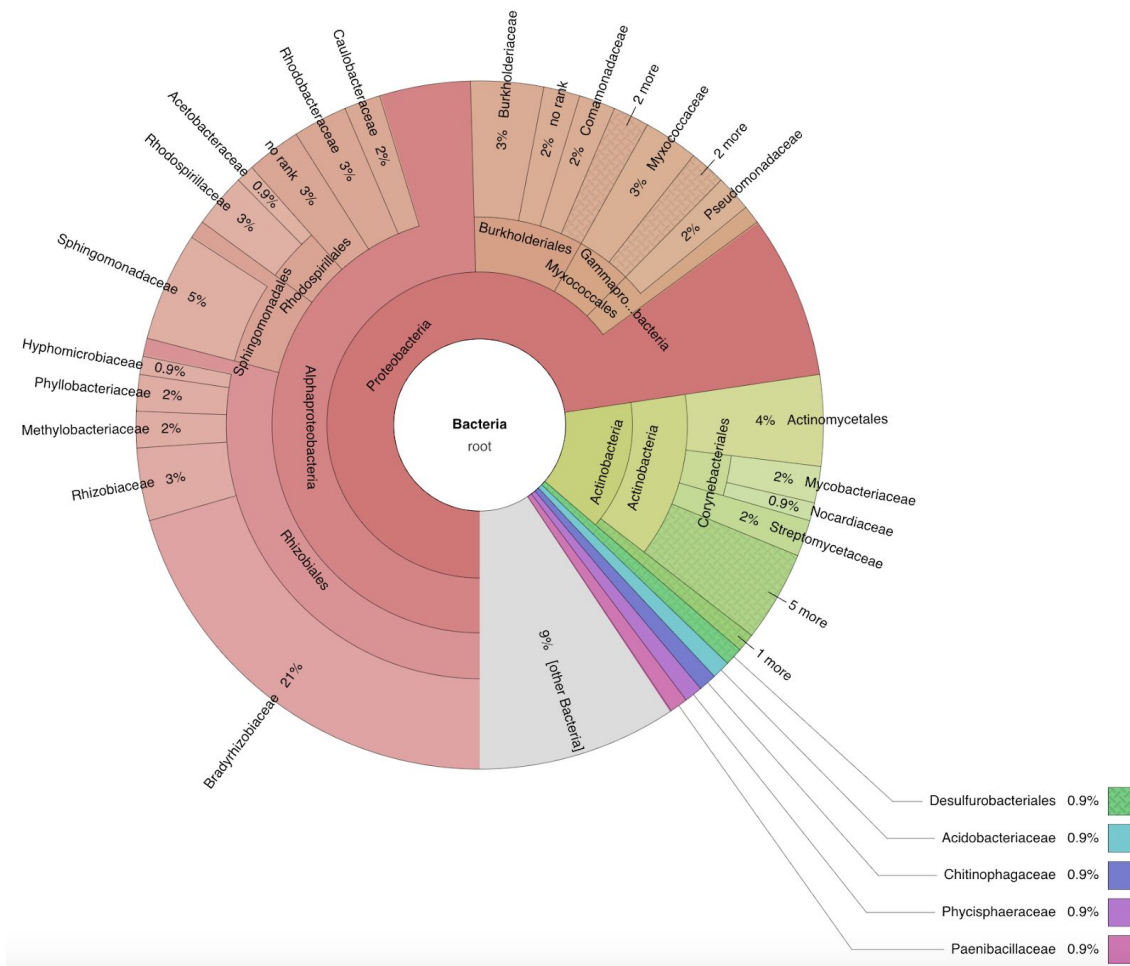
Figure 11. Kraken analysis of soil slurry inoculation. Unassigned sequence is not shown. Note the high levels of unassigned bacteria - often denoting classification of uncultivated bacteria.

Manual inspection of taxonomic assignments indicates that many community members of the soil slurry persisted in the low nitrogen, low organic content, sand environment that was used.
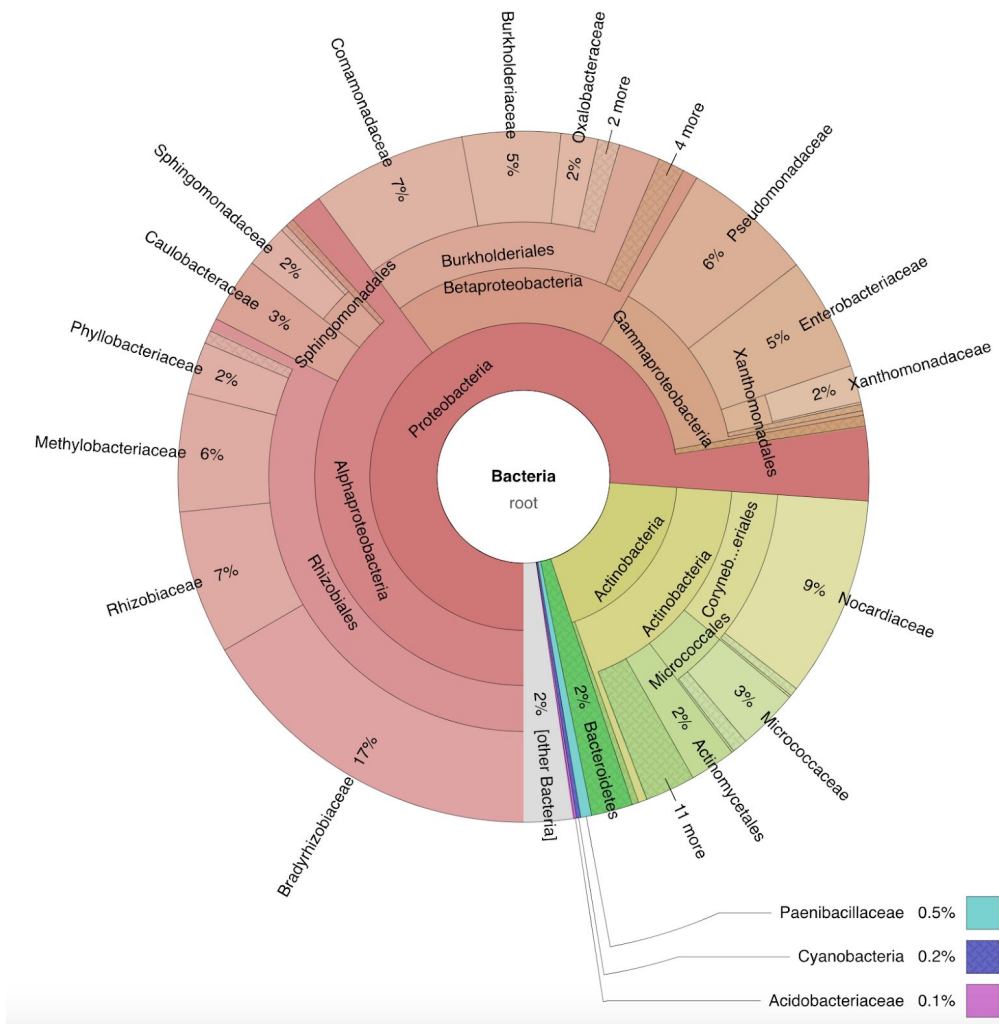
Figure 12. Kraken taxonomic assignment of pot 228 sand sample, as a fairly representative community composition of a soil slurry + culture mix treatment pot. Unassigned reads are excluded from this graph.

General taxonomic assignment of nodule sequence assigns most sequence to *Bradyrhizobium* spp. Minor presence of other bacteria may be indicative of insufficient surface cleaning of nodules. No major (>1%) presence of other bacteria were found, except for pot 207, with a considerable amount of Pseudomonas (particularly *Pseudomonas fluorescens*) assignment. This sample also had a high level of unassigned reads compared to other nodule samples. Most reads that did not map to target genomes or minor bacterial presence were unassigned, and may represent host DNA.
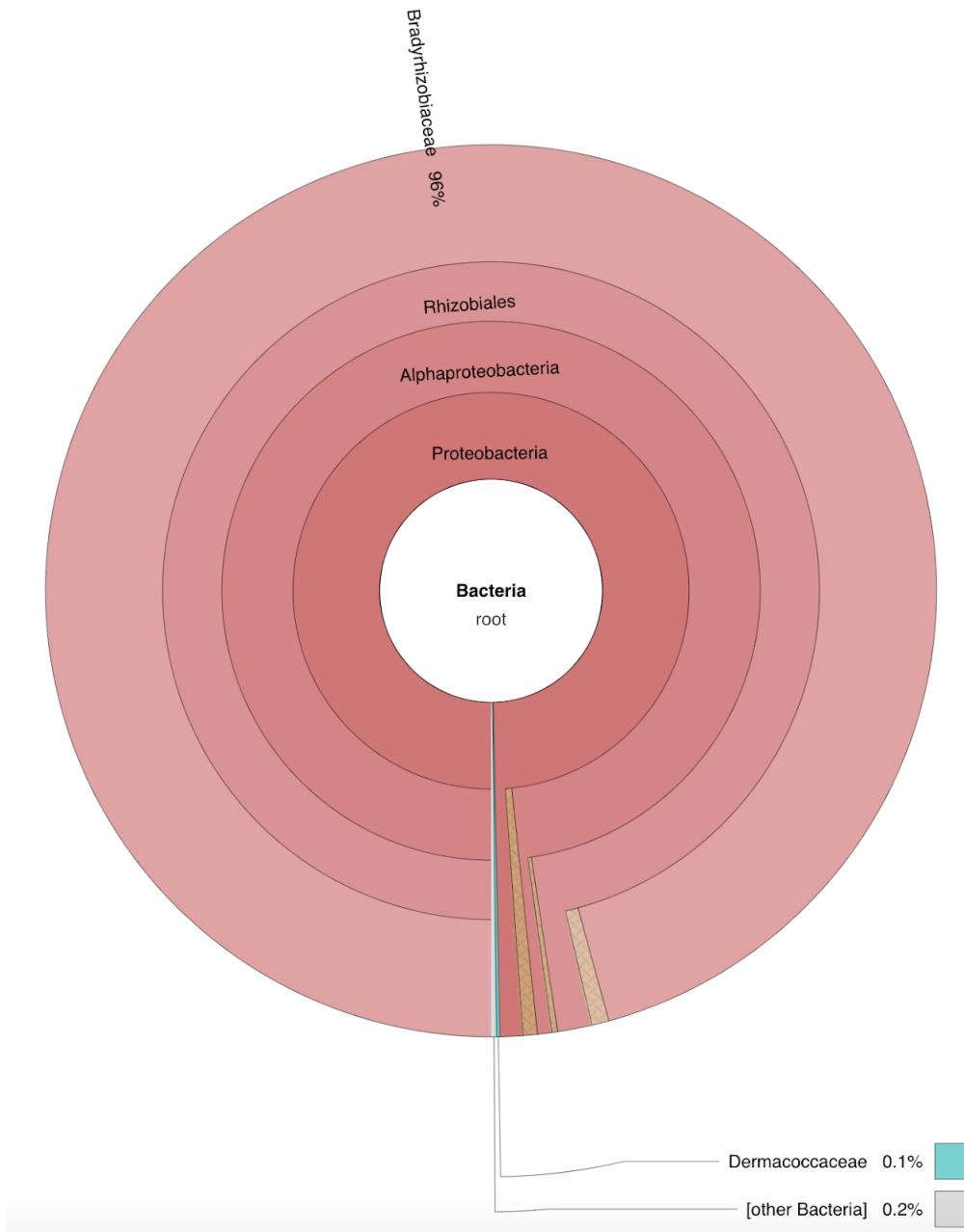
Figure 13. Pot 228 nodules, with Kraken assignments. Almost all sequence is assigned to Bradyrhizobiaceae, as expected.

Sample 207 nodules are an interesting outlier, as only 7.1% of nodule reads matched strain genomes. Further Kraken analysis indicates that 88% of these reads could not be assigned to microbial genomes.
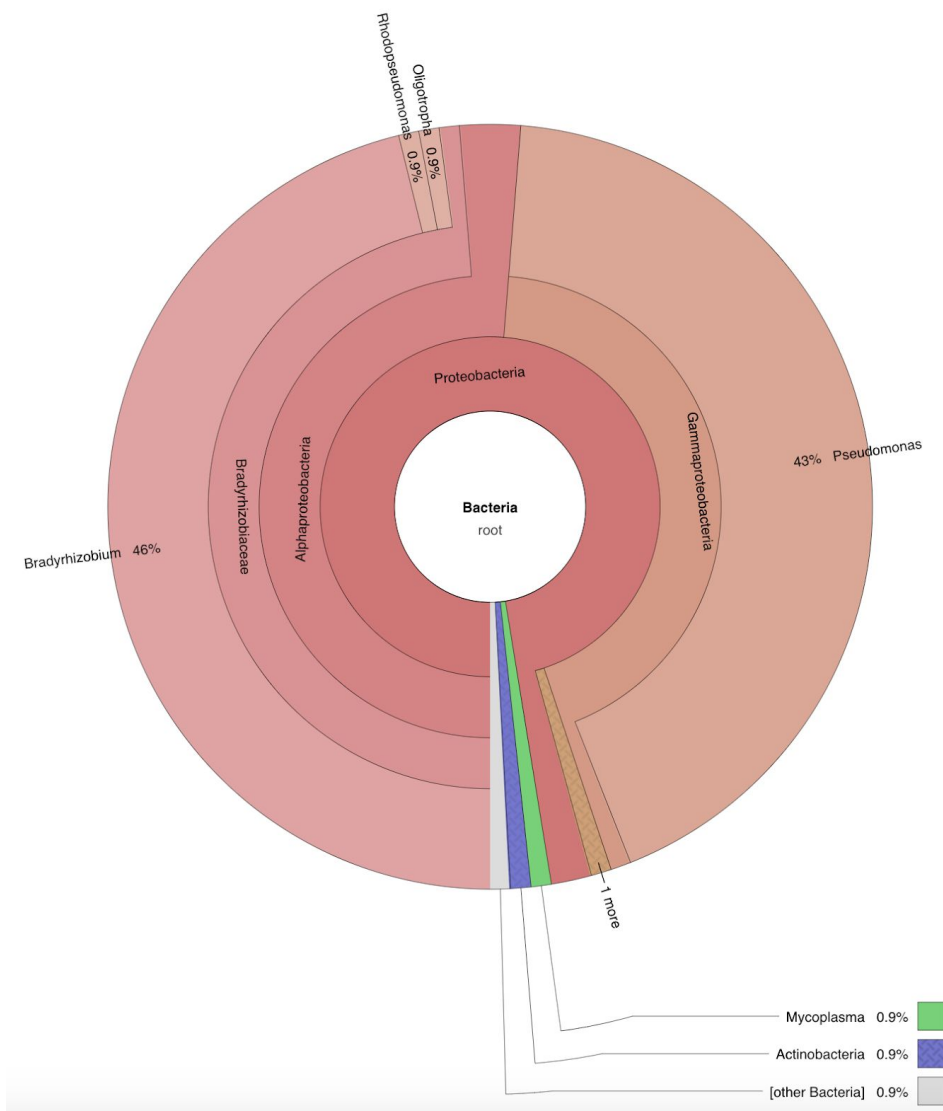
Figure 14. Kaiju results for the sample 207 (nodules). 88% of reads could not be classified and was excluded from this graph.

Abundance estimates

Abundance estimates were conducted separately across samples, across combined treatments, and across all simulated treatments.
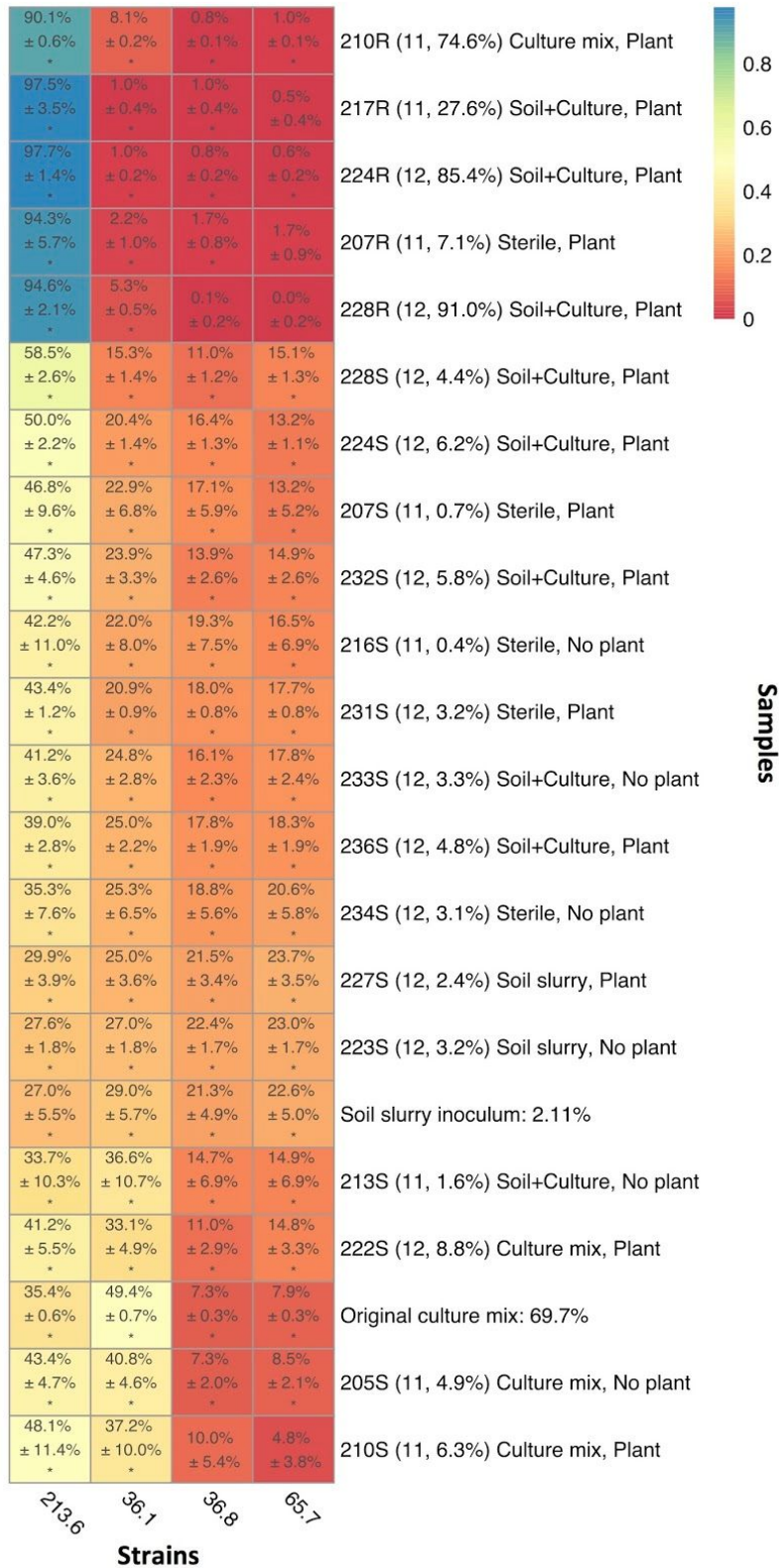
Figure 15. Proportional abundance estimates of each sample. The soil slurry inoculation and culture mix have percentages of reads mapping to strain genomes.

For all other samples, sample label denotes, in order: Pot ID, Tray ID, percentage of reads mapped to target strain genomes, inoculation treatment, and host presence/absence. Abundances marked as significant are denoted with an asterisk (*).

Summary abundances over treatment appear to show four overall groupings. Strain 213.6 has the highest abundance across all nodule treatments. Even though 36.1 appeared to have an initial advantage in being the most abundant in the inoculation treatment, both strains 213.6 and 36.1 are the most abundant in culture mix only treatments. Although 213.6 had higher abundance in these samples, this cannot be confidently assigned given the large error bars from working with a relatively low amount of reads. Strain 213.6 was the most abundant not only in nodules, but also in soil+culture samples.

Sand from containers inoculated with soil slurry only (and not culture mix) all reflected a similar pattern of relatively equal abundance between strains. This may be due to other rhizobial strains in the soil slurry matching to strain genomes.
Due to time constraints, no DNA was extracted from nodules of soil inoculation only plants (although all plants removed from that treatment had visible nodules similar to plants from other treatments of a similar size).
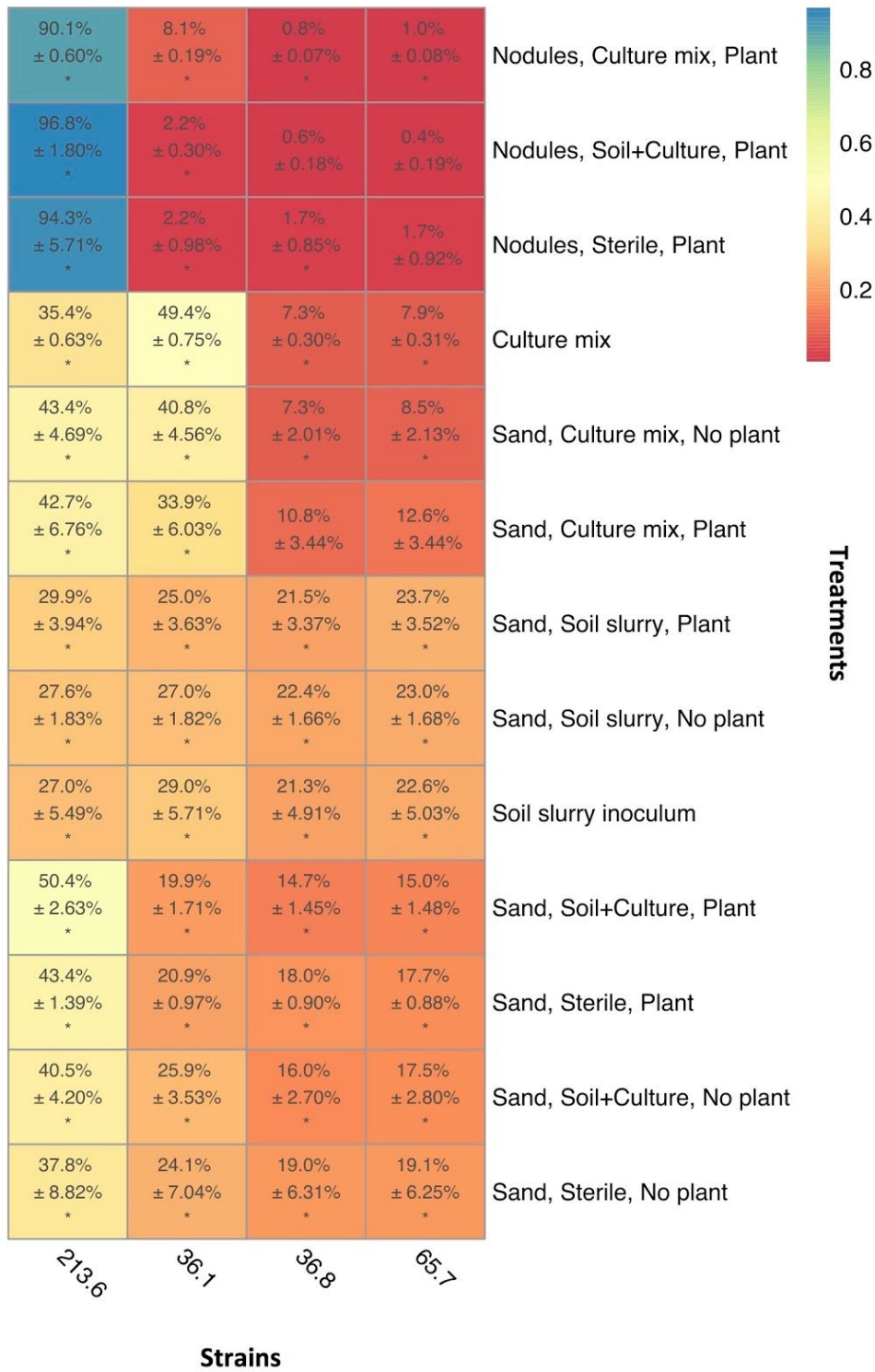
Figure 16. Overall abundance estimates by treatment. Treatment label denotes sample type, inoculation treatment, and host presence, in order (except for the culture mix and soil slurry).
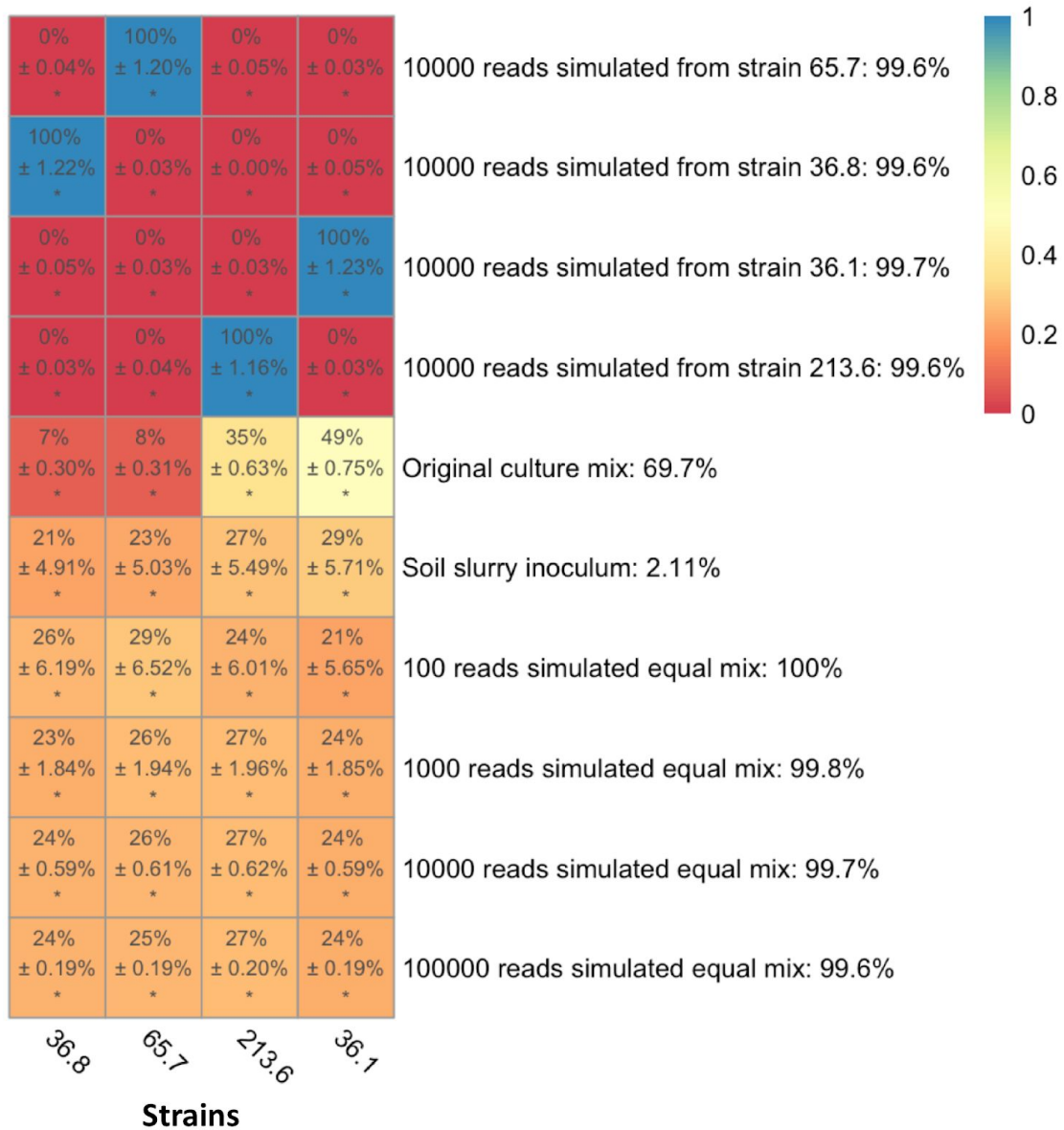
Figure 17. Comparison of inoculum abundance with simulated equal proportion reads and simulated single-strain reads. Note the slight bias towards strains 213.6 and 65.7. All abundance estimates in this test were marked as significant by DiTASiC.

Potential horizontal gene transfer

Although a pipeline for identification of potential horizontal gene transfer events from individual reads was planned, this was not completed. A considerable number of long reads suitable for this purpose were identified, with the longest reads being 46kb long (Fig. 16).
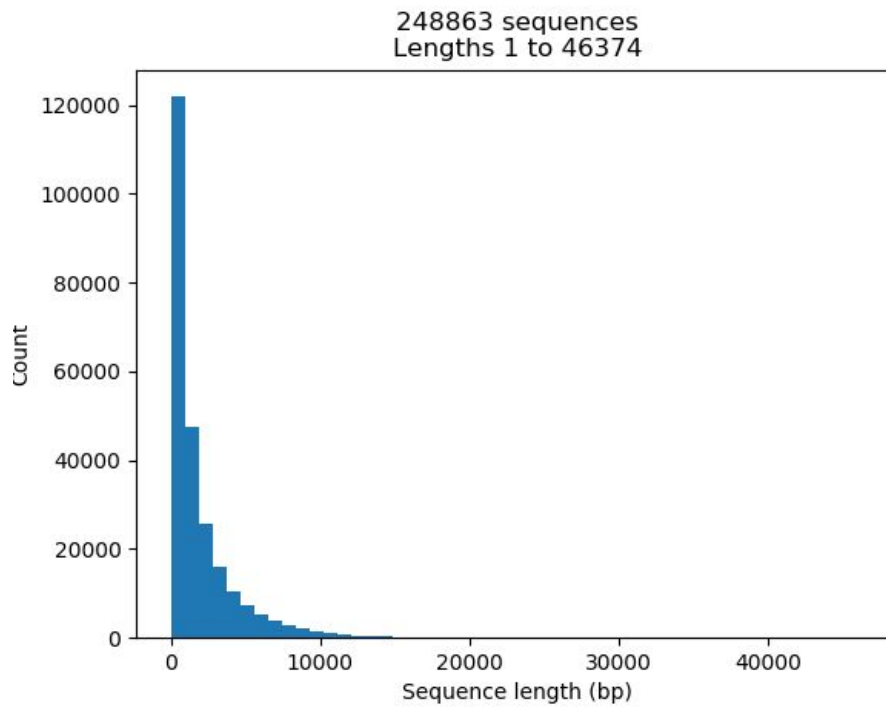
Figure 18. Length distribution of read fragments across all pot (sand and nodules) samples. No filtering of short (<1kb) fragments was conducted, as these were kept for the purpose of strain abundance assignment.

4. Discussion

Abundance patterns

The main result of this experiment largely confirms our initial hypothesis that inoculated rhizobia strain abundance patterns would change under different circumstances, with certain strains being better suited to plant nodulation, and other strains with better adaptation to a free-living soil environment. As expected, nodulation was necessary for plant growth in a very low nitrogen soil environment.

Previous work based on strain mutation rate indicates generation time in an external environment is likely considerably longer than in a lab environment, particularly in rich growth media (Gibson et al., 2018). Given the already slow growth of these strains (taking up to a week for visible colonies on agar), abundance levels may not have reached an equilibrium. Taking samples at more time points may help to resolve the extent to which abundance patterns are still changing.

Some evidence for sufficient microbial growth is that sterile controls show similar abundance patterns with each other and to soil slurry with culture mix inoculations, pointing to early cross contamination and within-community interactions. A pure founder effect bias would likely result in differing abundance patterns from order of contamination in sterile controls.

One evidence against abundance stabilisation is in having minimal abundance pattern difference between host present and host absent treatments; the effect of plant root exudates is likely to cause some effect, potentially both directly on strain abundance and on other microbes with secondary effects on the rhizobial strains (Bauer and Mathesius, 2004; Wendlandt et al., 2018). Plants may also carry their own microbes in seeds and vertically transfer microbiome community members, but no evidence of this was seen in strain abundance patterns (Aguilar et al., 2018). Further analysis of the rest of the

microbiome, both in identification of community members and in observation of gene families present, would likely help in identifying host effects.

Preliminary analysis led to a suspected bias from genomic differences between strains, particularly given the differences in genome quality. As such, read simulation was used to assess the extent of this bias (Fig. 13). The differential abundance in most samples exceeds the level of bias found through read simulation; as such, it is likely that actual biological factors are responsible for the differential abundance observed in most samples.

Soil abundance was not directly representative of nodule abundance, although the most abundant strain in sand was also the strain that appears primarily responsible for nodulation (Fig. 13). Strain 213.6 was the most successful overall; however, low sample size indicates that this result may be specific to the samples tested. Particularly given that multiple nodules were found on each plant, neutral nodulation selection (i.e. inoculated strains encountered by roots have an equal probability of forming nodules) is unlikely to produce the results seen. This may indicate that other factors such as signalling with the particular genotype of these plants were affecting nodulation. Survivorship bias may also have an impact, given high losses of germinated plants. This likely had an additional selection effect on these strains. One explanation of this would be that plants that selected other strains had insufficient nitrogen fixed, and failed to grow to a suitable size for selection. Another possible explanation is that plants which were better adapted to surviving in the experimental environment also had other phenotypic traits that led to selection of the favoured strain Given the overall success of strain 213.6, it may be worth further evaluating this strain for further use.

Variation by tray and spatial effects was planned, but not completed due to insufficient trays being sequenced. All plants with successful emergence had nodules, but insufficient phenotypic data allows for determining impact of strains on plant success.

Abundance estimate error and significance were provided by the DiTASic pipeline (Fischer et al., 2017). Modifications to statistical procedures may be necessary for other experiments, particularly in the grouping of p-values (if some samples have much higher p-values than others, or if abundance patterns follow a different distribution (Heard and Rubin-Delanchy, 2018)). No threshold has been set or statistically considered for the presence or absence of a given strain; this may be necessary in other applications (Fischer et al., 2017)

Although all strains were cultured from *A. acuminata* nodules, the plants used in this experiment have not been genotyped, and as such may favour this strain for nodulation based on signalling than others. Inoculation strains were collected along a transect within the home range of *A. acuminata*, and the controlled environment may have altered host preferences for nodulation. The simulated chamber environment was calibrated against one particular point over a given time period of the *A. acuminata* home range; altering the environmental parameters may influence nodulation. Water was added consistently but not at a saturating level due to time constraints, and likely had an additional influence on strain selection, both in affecting plant phenotype and in the abundance of free-living rhizobia (Swaine et al., 2007). Individual genotyping would have provided additional data on the extent that host effects were a factor, but was outside the original scope of this project, and given a current absence of an *A. acuminata* genome, is likely currently impractical. Usage of seed from a single maternal line can also mitigate this, but would likely be an additional source of bias from host genotype effects.

Exclusively using autoclaved sand was done to minimise substrate contamination, induce strong nodulation from lack of environmental bioavailable nitrogen, and minimise substrate chemistry changes in autoclaving. While most of the *A. acuminata* home range is in primarily sandy soil, some strains would likely be adapted to areas with more organic content.

Given that results are largely in line with available literature and initial hypotheses, it can be concluded that the method developed here allows for accurate abundance measurement of inoculated high abundance community member bacteria strains.

Method application

This method allows for observation of bacterial distribution without bias from amplification or culture-based methods. One potential source of bias is from PCR, which may under-represent abundance of DNA fragments with very high or low GC content (Aird et al., 2011)

One drawback of this method is an inability to detect rare (<1%) strains. Inefficiencies in library preparation and sequencing result in a loss of total DNA fragments at several stages, and DNA fragments from rare strains may be lost entirely through this process; amplification-based methods lose DNA before these stages, and are more likely to identify rare strains. As rare strains are often tested more for presence or absence than distribution, conventional microbiome sequencing techniques and different statistical methods (Gans et al., 2005) may be more appropriate.

Handling low biomass without amplification and the resulting need to pool samples together results in a loss of resolution in cases where individual nodules have diverse strain types, i.e. inter-nodule diversity cannot be distinguished from intra-nodule diversity. This is particularly true in the case of larger plants (Dinnage et al., 2018). For these results, container and plant size were likely too small for considerable intra-sample heterogeneity.

ONT reports on enhancing nanopore sequencing accuracy (Brown, 2018) indicate that consensus accuracy should increase considerably with the use of multiple nanopore variants, as systematic error differs between variants. However, this is unlikely to improve sequencing accuracy for this method, as abundance estimation and potential

horizontal gene transfer is conducted on individual reads, and assembly is generally not conducted at any stage.

The future use of the more recent ONT $1D^2$ sequencing method, which sequences both strands of dsDNA fragments, should increase the overall sequencing accuracy from calling internal consensus, but does not reduce systematic pore-based error. Given that the $1D^2$ method is fairly new, both usage in metagenomics and extensive independent verification of accuracy is still lacking.

Agricultural application of commercial rhizobia strains often requires frequent reapplication. Guidelines for application are often based on rough heuristics, with effectiveness judged based on crop nitrogen or yield (GRDC, 2013). These metrics are also dependent on other factors, and not necessarily indicative of the effectiveness of added inoculation (van Kessel and Hartley, 2000). As this method relies more on subsampling total DNA and as such has a reduced requirement for the clean room controls for typical amplification-based microbiome analysis, and uses portable MinION sequencing, it may be possible to use this method for rapid site-specific validation of strain effectiveness, both for soil persistence to optimise application, and nodulation of crops. This is particularly applicable for future ONT devices with reduced cost and improved field capabilities, such as integrated basecalling and single-tube DNA extraction with library preparation (Brown, 2018)

Community composition

Most root nodule samples yielded high overall reads mapping to strain genomes. An initial concern was that high levels of host DNA would be present in samples, but the DNA extraction method appears to have minimised the amount of *A. acuminata* DNA present in bacterial samples. Plant cell walls allow for selective extraction of non-host DNA (Lear et al., 2018), and it appears that the PowerSoil kit was effective in this regard, given the high rates of mapped reads in most nodule samples.

One interesting result was the high levels of *Pseudomonas* from nodule samples in pot 207 (Fig. 14). *Pseudomonas* spp. have previously been shown to be present in and potentially capable of forming *Medicago* nodules (Wigley et al., 2017) and may promote plant growth through antifungal effects (Kumar et al., 2002). Although this was from only one sample, further research into this area may prove insightful, given previous research on *Pseudomonas*-plant interactions.

Surface sterilisation of seeds may not have been sufficient to prevent vertical transfer of microbiome community members. This is supported by widespread apparent contamination of surface-sterilised seeds across multiple batches, but will need different methods to confirm (Aguilar et al., 2018). If so, these did not appear to have a noticeable effect on target strain abundance.

The accuracy of this method indicates potential future use particularly with inoculated strains, where strain genomes are available for use in mapping. However, it is likely that this method can still be used in environments where most environmental strains have previously been sequenced.

Some reports indicate that there should be observable levels of archaea in the soil microbiome (Chen et al., 2008). The almost complete absence of archaea in these samples may indicate a bias in the extraction method (Plassart et al., 2012) or poor suitability to a lab environment; further testing with positive archeal controls may assist in this regard. It may also indicate unsuitability of archaeal growth on the media used.

Horizontal gene transfer

The initial plan for HGT identification was to use BLAST on each read using a local database with the sequences of the inoculated strain genomes and filter reads that have a unique alignment to one strain genome, flanked by alignments to different strain genome. As such, this method only works on HGT that results in chromosomal integration and not plasmid transfer.

Although no actual instances of HGT were observed in individual reads through manual BLAST search on a small subset, the presence of reads mapping to other strains within nodules may indicate HGT events and the transferred sequence being picked up as most of a given read, rather than an actual separate strain present in the nodule. Obtaining strains from the nodules and sequencing pure strain colonies may allow confirming if actual HGT has occurred, if the particular strain that had an HGT event was picked up in culturing. If the resequenced strain had no HGT events, this would still not rule out that HGT events had not occured, as the particular colonies with HGT or other strains may not have been picked up in subculturing. Additionally, several strain genomes were not completed into single chromosomal contigs, further complicating identification from chromosome position (Hendrickson et al., 2018)

Several issues with HGT identification using this method were identified through its development. Firstly, as flanking regions are common to HGT events (Ashton et al., 2015), an HGT event may occur in a flanking sequence that maps to more than one sequence, but with an inserted sequence from a non-overlapping set of strains (e.g. flanking sequence present in strains A and B, with an insertion from a sequence found in strains C and D). Secondly, chimeric reads (false positives) may occur from DNA library preparation, particularly in ligation-based library preparation. DNA fragments that enter a nanopore in quick succession may also result in the software falsely reading two or more fragments as a single read (White et al., 2017). Both of these may be mitigated by requiring flanking alignments rather than simple unique alignments to different strain genomes. Thirdly, homologous recombination areas are likely to have overlap where the recombination occurs; a threshold of overlap has to be set to accommodate nanopore sequencing errors for a given BLAST threshold. Lastly, using only strain genomes means that HGT events incorporating DNA from outside the inoculated strains would not be detected. It may be possible to resolve this through further BLAST of unknown sequence insertion and matching it to other bacterial strains

found in the same sample; however, this is not necessarily indicative of which particular organism the sequence came from, as a previous HGT event may have occurred.

Other future directions

A wide variety of analyses can be conducted using this dataset alone. Correlation of abundance patterns to specific community member bacterial groups should be possible, given that general taxonomic assignment on individual samples was fairly informative. A considerable amount of phenotypic data in regular imaging and manual emergence counts was taken, but was not analysed due to time constraints. Given both plant phenotypic data and whole microbiome data was collected, this data can be analysed for correlations between overall microbiome composition and plant phenotype, although this was outside the scope of this project.

Additional controls that were considered but not completed are the use of single-strain inoculations for each strain, to confirm nodulation potential of each strain with the given batch of *A. acuminata*, and to consider founder effects in a free-living state. Culturing rhizobial strains at different OD600 levels and quantifying colony-forming units would likely result in improved equal abundance inoculation.

Only one complex microbiome inoculation was used, from the base of an established *A. acuminata* tree. Inoculation with a different microbiome, possibly from low-rhizobia soil, or a simple mix of cultured strains (Vorholt et al., 2017), may impact results considerably, particularly with the presence or absence of other bacteria that have a specific interaction with these rhizobial strains (Hall et al., 2018)

Use of different soils are also likely to change abundance patterns. Unlike the sand used in this experiment, other soil types may be difficult to sterilise and are also likely to have considerable changes in chemistry through this process, and may no longer be representative of the original soil type (Wolf et al., 1989). Different nutrients and the

availability of preexisting bioavailable nitrogen may also have an impact on abundance patterns (Chen et al., 2016; Ouyang et al., 2018).

Many amplification-based kits have common contaminants that can impact abundance or strain presence data (Salter et al., 2014). This method is sample agnostic, and should be applicable beyond soil and nodule metagenomics, and as such offers an alternative with reduced bias in high biomass samples, such as in fecal or mucosal samples.

Conclusion

In this experiment, the effectiveness of direct gDNA sequencing for abundance estimation of known closely related strains was demonstrated. This allows for avoidance of culture or amplification bias. The bioinformatics pipeline used was also validated for distinguishing closely related strains and for accuracy of abundance estimation through the use of read simulation.

Long reads from MinION sequencing also allows the possibility of HGT identification from single reads. Manual inspection of a small subset did not yield any results. Although no software was successfully developed for this purpose, a framework for identification of HGT from single reads has been described.

Abundance patterns found show that only one strain, 213.6, formed nodules with *A. acuminata* plants, even though all strains used were originally obtained from wild *A. acuminata* nodules. All strains persisted at observable levels in soil, but abundance patterns appeared to be altered in the presence of external microbiome inoculation. Inoculation with an external microbiome also increased emergence slightly, but this was not statistically significant.

These results contribute toward an enhanced understanding of rhizobia within and beyond the lab, and may also direct a variety of future experiments in clarifying the biological processes affecting strain abundance. The method developed as a part of this

process opens new avenues for research and agricultural diagnostics, particularly with increased need for solutions in a changing climate for a world with changing needs.

References

Aguilar, A., Mora, Y., Dávalos, A., Girard, L., Mora, J., Peralta, H., 2018. Analysis of genome sequence and symbiotic ability of rhizobial strains isolated from seeds of common bean (Phaseolus vulgaris). BMC Genomic- 19, 645. doi:10.1186/s12864-018-5023-0

Aird, D., Ross, M.G., Chen, W.-S., Danielsson, M., Fennell, T., Russ, C., Jaffe, D.B., Nusbaum, C., Gnirke, A., 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 12.

Ashton, P.M., Nair, S., Dallman, T., Rubino, S., Rabsch, W., Mwaigwisya, S., Wain, J., O'Grady, J., 2015. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat. Biotechnol. 33, 296–300. doi:10.1038/nbt.3103

Atlas of Living Australia, 2018. Atlas of Living Australia [WWW Document].

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 Genomic- 9, 75. doi:10.1186/1471-2164-9-75

Bankevich, A., Pevzner, P.A., 2018. Joint Analysis of Long and Short Reads Enables Accurate Estimates of Microbiome Complexity. Cell Syst. 7, 192–200.e3. doi:10.1016/j.cels.2018.06.009

Bauer, W.D., Mathesius, U., 2004. Plant responses to bacterial quorum sensing signals. Current opinion in plant biology 7, 429–433. doi:10.1016/j.pbi.2004.05.008

Benítez-Páez, A., Sanz, Y., 2017. Multi-locus and long amplicon sequencing approach to study microbial diversity at species level using the MinION$_{TM}$ portable nanopore sequencer. GigaScience 6, 1–12. doi:10.1093/gigascience/gix043

Bissett, A., Fitzgerald, A., Court, L., Meintjes, T., Mele, P.M., Reith, F., Dennis, P.G., Breed, M.F., Brown, B., Brown, M.V., Brugger, J., Byrne, M., Caddy-Retalic, S., Carmody, B., Coates, D.J., Correa, C., Ferrari, B.C., Gupta, V.V.S.R., Hamonts, K., Haslem, A., Hugenholtz, P., Karan, M., Koval, J., Lowe, A.J., Macdonald, S., McGrath, L., Martin, D., Morgan, M., North, K.I., Paungfoo-Lonhienne, C., Pendall, E., Phillips, L., Pirzl, R., Powell, J.R., Ragan, M.A., Schmidt, S., Seymour, N., Snape, I., Stephen, J.R., Stevens, M., Tinning, M., Williams, K., Yeoh, Y.K., Zammit, C.M., Young, A., 2016. Introducing BASE: the Biomes of Australian Soil Environments soil microbial diversity database. GigaScience 5, 21. doi:10.1186/s13742-016-0126-5

Brown, C., 2018. Oxford Nanopore Technical Update - Clive Brown: Plenary from London Calling 2018.

Burghardt, L.T., Epstein, B., Guhlin, J., Nelson, M.S., Taylor, M.R., Young, N.D., Sadowsky, M.J., Tiffin, P., 2018. Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. Proc. Natl. Acad. Sci. United States Am. 115, 2425–2430. doi:10.1073/pnas.1714246115

Capone, D.G., Burns, J.A., Montoya, J.P., Subramaniam, A., Mahaffey, C., Gunderson, T., Michaels, A.F., Carpenter, E.J., 2005. Nitrogen fixation by Trichodesmium spp.: An important source of new nitrogen to the tropical and subtropical North Atlantic Ocean: NITROGEN FIXATION IN THE NORTH ATLANTIC. Glob. Biogeochem. Cycles 19. doi:10.1029/2004GB002331

Chen, C., Zhang, J., Lu, M., Qin, C., Chen, Y., Yang, L., Huang, Q., Wang, J., Shen, Z., Shen, Q., 2016. Microbial communities of an arable soil treated for 8 years with organic and inorganic fertilizers. Biology and Fertility of Soils 52. doi:10.1007/s00374-016-1089-5

Chen, X.-P., Zhu, Y.-G., Xia, Y., Shen, J.-P., He, J.-Z., 2008. Ammonia-oxidizing archaea: important players in paddy rhizosphere soil? Environ. Microbiol. 10, 1978–1987. doi:10.1111/j.1462-2920.2008.01613.x

Chen, Y., Wang, J., Yang, N., Wen, Z., Sun, X., Chai, Y., Ma, Z., 2018. Wheat microbiome bacteria can reduce virulence of a plant pathogenic fungus by altering

histone acetylation. Nat. Commun. 9, 3429. doi:10.1038/s41467-018-05683-7

Craine, J.M., Elmore, A.J., Wang, L., Aranibar, J., Bauters, M., Boeckx, P., Crowley, B.E., Dawes, M.A., Delzon, S., Fajardo, A., Fang, Y., Fujiyoshi, L., Gray, A., Guerrieri, R., Gundale, M.J., Hawke, D.J., Hietz, P., Jonard, M., Kearsley, E., Kenzo, T., Makarov, M., Marañón-Jiménez, S., McGlynn, T.P., McNeil, B.E., Mosher, S.G., Nelson, D.M., Peri, P.L., Roggy, J.C., Sanders-DeMott, R., Song, M., Szpak, P., Templer, P.H., Van der Colff, D., Werner, C., Xu, X., Yang, Y., Yu, G., Zmudczyńska-Skarbek, K., 2018. Isotopic evidence for oligotrophication of terrestrial ecosystems. Nat. Ecol. & Evol. doi:10.1038/s41559-018-0694-0

Dinnage, R., Simonsen, A.K., Barrett, L.G., Cardillo, M., Raisbeck-Brown, N., Thrall, P.H., Prober, S.M., 2018. Larger plants promote a greater diversity of symbiotic nitrogen-fixing soil bacteria associated with an Australian endemic legume. J. Ecol. doi:10.1111/1365-2745.13083

Edgington, E.S., 1972. An Additive Method for Combining Probability Values from Independent Experiments. The Journal of Psychology 80. doi:10.1080/00223980.1972.9924813

Emerson, J.B., Adams, R.I., Román, C.M.B., Brooks, B., Coil, D.A., Dahlhausen, K., Ganz, H.H., Hartmann, E.M., Hsu, T., Justice, N.B., Paulino-Lima, I.G., Luongo, J.C., Lymperopoulou, D.S., Gomez-Silvan, C., Rothschild-Mancinelli, B., Balk, M., Huttenhower, C., Nocker, A., Vaishampayan, P., Rothschild, L.J., 2017. Schrödinger's microbes: Tools for distinguishing the living from the dead in microbial ecosystems. Microbiome 5, 86. doi:10.1186/s40168-017-0285-3

Environment and Public Affairs Committee, 2014. Inquiry into the Sandalwood Industry in Western Australia.

Erisman, J.W., Sutton, M.A., Galloway, J., Klimont, Z., Winiwarter, W., 2008. How a century of ammonia synthesis changed the world. Nat. Geosci. 1. doi:10.1038/ngeo325

Fischer, M., Strauch, B., Renard, B.Y., 2017. Abundance estimation and differential testing on strain level in metagenomics data. Bioinformatics 33, i124–i132.

Forouzan, E., Shariati, P., Mousavi Maleki, M.S., Karkhane, A.A., Yakhchali, B., 2018.

Practical evaluation of 11 de novo assemblers in metagenome assembly. J. Microbiol. methods 151, 99–105. doi:10.1016/j.mimet.2018.06.007

Gans, J., Wolinsky, M., Dunbar, J., 2005. Computational improvements reveal great bacterial diversity and high metal toxicity in soil. Sci. 309, 1387–1390. doi:10.1126/science.1112665

Gibson, B., Wilson, D.J., Feil, E., Eyre-Walker, A., 2018. The distribution of bacterial doubling times in the wild. Proceedings. Biological sciences 285. doi:10.1098/rspb.2018.0789

Goh, C.-H., Veliz Vallejos, D.F., Nicotra, A.B., Mathesius, U., 2013. The impact of beneficial plant-associated microbes on plant phenotypic plasticity. Journal of chemical ecology 39, 826–839. doi:10.1007/s10886-013-0326-8

GRDC, 2013. Rhizobial Inoculants. Grains Research and Development Corporation.

Gruber, N., Galloway, J.N., 2008. An Earth-system perspective of the global nitrogen cycle. Nature 451, 293–296. doi:10.1038/nature06592

Hall, J.P.J., Harrison, E., Brockhurst, M.A., 2018. Competitive species interactions constrain abiotic adaptation in a bacterial soil community. Evol. Lett. doi:10.1002/evl3.83

Häyhä, T., Lucas, P.L., van Vuuren, D.P., Cornell, S.E., Hoff, H., 2016. From Planetary Boundaries to national fair shares of the global safe operating space — How can the scales be bridged? Glob. Environ. Chang. 40. doi:10.1016/j.gloenvcha.2016.06.008

Heard, N.A., Rubin-Delanchy, P., 2018. Choosing between methods of combining p-values. Biometrika 105. doi:10.1093/biomet/asx076

Hendrickson, H.L., Barbeau, D., Ceschin, R., Lawrence, J.G., 2018. Chromosome architecture constrains horizontal gene transfer in bacteria. PLoS genetics 14, e1007421. doi:10.1371/journal.pgen.1007421

Illumina, 2018. Illumina sequencing platforms [WWW Document].

Kiers, E.T., Ratcliff, W.C., Denison, R.F., 2013. Single-strain inoculation may create spurious correlations between legume fitness and rhizobial fitness. New Phytol. 198, 4–6. doi:10.1111/nph.12015

54

Kumar, N.R., Arasu, V.T., Science, P.G.-C., 2002, 2002. Genotyping of antifungal compounds producing plant growth-promoting rhizobacteria, Pseudomonas fluorescens.

Kwak, M.-J., Kong, H.G., Choi, K., Kwon, S.-K., Song, J.Y., Lee, J., Lee, P.A., Choi, S.Y., Seo, M., Lee, H.J., Jung, E.J., Park, H., Roy, N., Kim, H., Lee, M.M., Rubin, E.M., Lee, S.-W., Kim, J.F., 2018. Rhizosphere microbiome structure alters to enable wilt resistance in tomato. Nat. Biotechnol. doi:10.1038/nbt.4232

Lear, G., Dickie, I., Banks, J., Boyer, S., Buckley, H., Buckley, T., Cruickshank, R., Dopheide, A., Handley, K., Hermans, S., Kamke, J., Lee, C., MacDiarmid, R., Morales, S., Orlovich, D., Smissen, R., Wood, J., Holdaway, R., 2018. Methods for the extraction, storage, amplification and sequencing of DNA from environmental samples. New Zeal. J. Ecol. doi:10.20417/nzjecol.42.9

Leveau, J.H.J., 2007. The magic and menace of metagenomics: prospects for the study of plant growth-promoting rhizobacteria. European Journal of Plant Pathology 119. doi:10.1007/s10658-007-9186-9

Li, P.-E., Lo, C.-C., Anderson, J.J., Davenport, K.W., Bishop-Lilly, K.A., Xu, Y., Ahmed, S., Feng, S., Mokashi, V.P., Chain, P.S.G., 2017. Enabling the democratization of the genomics revolution with a fully integrated web-based bioinformatics platform. Nucleic acids Res. 45, 67–80. doi:10.1093/nar/gkw1027

Lin, Y., Yuan, J., Kolmogorov, M., Shen, M.W., Chaisson, M., Pevzner, P.A., 2016. Assembly of long error-prone reads using de Bruijn graphs. Proc. Natl. Acad. Sci. United States Am. 113, E8396–E8405. doi:10.1073/pnas.1604560113

Loman, N.J., Quick, J., Simpson, J.T., 2015. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat. methods 12, 733–735. doi:10.1038/nmeth.3444

Manh Nguyen, T., Seo, C., Ji, M., Paik, M.-J., Myung, S.-W., Kim, J., 2018. Effective soil extraction method for cultivating previously uncultured soil bacteria. Appl. Environ. Microbiol. doi:10.1128/AEM.01145-18

McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLoS Comput. Biol. 10, e1003531.

doi:10.1371/journal.pcbi.1003531

Mocali, S., Benedetti, A., 2010. Exploring research frontiers in microbiology: the challenge of metagenomics in soil microbiology. Res. Microbiol. 161, 497–505. doi:10.1016/j.resmic.2010.04.010

Ng, K.L., Johannesen, T.B., Osterlund, M., bioRxiv, K.K.-, 2018, 2018. ReadFilter-Filtering reads of interest for quicker downstream analysis.

Ouyang, Y., Evans, S.E., Friesen, M.L., Tiemann, L.K., 2018. Effect of nitrogen fertilization on the abundance of nitrogen cycling genes in agricultural soils: A meta-analysis of field studies. Soil Biol. Biochem. 127. doi:10.1016/j.soilbio.2018.08.024

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., Stevens, R., 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic acids Res. 42, D206–D214. doi:10.1093/nar/gkt1226

Pacific Biosciences, 2018. SMRT SEQUENCING: READ LENGTHS [WWW Document].

Paula, R.R., Bouillet, J.-P., Ocheuze Trivelin, P.C., Zeller, B., Leonardo de Moraes Gonçalves, J., Nouvellon, Y., Bouvet, J.-M., Plassard, C., Laclau, J.-P., 2015. Evidence of short-term belowground transfer of nitrogen from Acacia mangium to Eucalyptus grandis trees in a tropical planted forest. Soil Biol. Biochem. 91. doi:10.1016/j.soilbio.2015.08.017

Plassart, P., Terrat, S., Thomson, B., Griffiths, R., Dequiedt, S., Lelievre, M., Regnier, T., Nowak, V., Bailey, M., Lemanceau, P., Bispo, A., Chabbi, A., Maron, P.-A., Mougel, C., Ranjard, L., 2012. Evaluation of the ISO standard 11063 DNA extraction procedure for assessing soil microbial abundance and community structure. PloS one 7, e44279. doi:10.1371/journal.pone.0044279

Pomerantz, A., Peñafiel, N., Arteaga, A., Bustamante, L., Pichardo, F., Coloma, L.A., Barrio-Amorós, C.L., Salazar-Valenzuela, D., Prost, S., 2018. Real-time DNA barcoding in a rainforest using nanopore sequencing: opportunities for rapid

biodiversity assessments and local capacity building. GigaScience 7. doi:10.1093/gigascience/giy033

Pootakham, W., Mhuantong, W., Yoocha, T., Putchim, L., Sonthirod, C., Naktang, C., Thongtham, N., Tangphatsornruang, S., 2017. High resolution profiling of coral-associated bacterial communities using full-length 16S rRNA sequence data from PacBio SMRT sequencing system. Sci. reports 7, 2774. doi:10.1038/s41598-017-03139-4

Quick, J., 2018. Ultra-long read sequencing protocol for RAD004.

Rago, L., Zecchin, S., Villa, F., Goglio, A., Corsini, A., Cavalca, L., Schievano, A., 2019. Bioelectrochemical Nitrogen fixation (e-BNF): Electro-stimulation of enriched biofilm communities drives autotrophic nitrogen and carbon fixation. Bioelectrochemistry 125. doi:10.1016/j.bioelechem.2018.10.002

Ravenhall, M., Škunca, N., Lassalle, F., Dessimoz, C., 2015. Inferring horizontal gene transfer. PLoS Comput. Biol. 11, e1004095. doi:10.1371/journal.pcbi.1004095

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., Smyth, G.K., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic acids Res. 43, e47. doi:10.1093/nar/gkv007

Salter, S.J., Cox, M.J., Turek, E.M., Calus, S.T., Cookson, W.O., Moffatt, M.F., Turner, P., Parkhill, J., Loman, N.J., Walker, A.W., 2014. Reagent and laboratory contamination can critically impact sequence-based microbiome analyses. BMC Biol. 12, 87. doi:10.1186/s12915-014-0087-z

Santi, C., Bogusz, D., Franche, C., 2013. Biological nitrogen fixation in non-legume plants. Ann. Bot. 111, 743–767. doi:10.1093/aob/mct048

Schalamun, M., Nagar, R., Kainer, D., Beavan, E., Eccles, D., Rathjen, J.P., Lanfear, R., Schwessinger, B., 2018. Harnessing the MinION: An example of how to establish long-read sequencing in a laboratory using challenging plant tissue from Eucalyptus pauciflora. Mol. Ecol. Resour. doi:10.1111/1755-0998.12938

Swaine, E.K., Swaine, M.D., Killham, K., 2007. Effects of drought on isolates of Bradyrhizobium elkanii cultured from Albizia adianthifolia seedlings of different provenances. Agroforestry Systems 69. doi:10.1007/s10457-006-9025-6

Thomma, B.P.H.J., Seidl, M.F., Shi-Kunne, X., Cook, D.E., Bolton, M.D., van Kan, J.A.L., Faino, L., 2016. Mind the gap; seven reasons to close fragmented genome assemblies. Fungal genetics and biology: FG & B 90, 24–30. doi:10.1016/j.fgb.2015.08.010

Thrall, P.H., Slattery, J.F., Broadhurst, L.M., Bickford, S., 2007. Geographic patterns of symbiont abundance and adaptation in native Australian Acacia-rhizobia interactions. J. Ecol. 95. doi:10.1111/j.1365-2745.2007.01278.x

Tringe, S.G., von Mering, C., Kobayashi, A., Salamov, A.A., Chen, K., Chang, H.W., Podar, M., Short, J.M., Mathur, E.J., Detter, J.C., Bork, P., Hugenholtz, P., Rubin, E.M., 2005. Comparative metagenomics of microbial communities. Sci. 308, 554–557. doi:10.1126/science.1107851

Tsirigos, A., Rigoutsos, I., 2005. A new computational method for the detection of horizontal gene transfer events. Nucleic acids Res. 33, 922–933. doi:10.1093/nar/gki187

Van Deynze, A., Zamora, P., Delaux, P.-M., Heitmann, C., Jayaraman, D., Rajasekar, S., Graham, D., Maeda, J., Gibson, D., Schwartz, K.D., Berry, A.M., Bhatnagar, S., Jospin, G., Darling, A., Jeannotte, R., Lopez, J., Weimer, B.C., Eisen, J.A., Shapiro, H.-Y., Ané, J.-M., Bennett, A.B., 2018. Nitrogen fixation in a landrace of maize is supported by a mucilage-associated diazotrophic microbiota. PLoS biology 16, e2006352. doi:10.1371/journal.pbio.2006352

van Kessel, C., Hartley, C., 2000. Agricultural management of grain legumes: has it led to an increase in nitrogen fixation? Field Crop. Res. 65, 165–181.

Vandenkoornhuyse, P., Quaiser, A., Duhamel, M., Le Van, A., Dufresne, A., 2015. The importance of the microbiome of the plant holobiont. New Phytol. 206, 1196–1206. doi:10.1111/nph.13312

Vitousek, P.M., Menge, D.N.L., Reed, S.C., Cleveland, C.C., 2013. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philos. Trans. R. Soc. London. Ser. B, Biol. Sci. 368, 20130119. doi:10.1098/rstb.2013.0119

Vorholt, J.A., Vogel, C., Carlström, C.I., Müller, D.B., 2017. Establishing Causality:

Opportunities of Synthetic Communities for Plant Microbiome Research. Cell host & microbe 22, 142–155. doi:10.1016/j.chom.2017.07.004

Wallenstein, M.D., 2017. Managing and manipulating the rhizosphere microbiome for plant health: A systems approach. Rhizosphere 3. doi:10.1016/j.rhisph.2017.04.004

Wendlandt, C.E., Regus, J.U., Gano-Cohen, K.A., Hollowell, A.C., Quides, K.W., Lyu, J.Y., Adinata, E.S., Sachs, J.L., 2018. Host investment into symbiosis varies among genotypes of the legume Acmispon strigosus, but host sanctions are uniform. New Phytol. doi:10.1111/nph.15378

White, R., Pellefigues, C., Ronchese, F., Lamiable, O., Eccles, D., 2017. Investigation of chimeric reads using the MinION. F1000Research 6. doi:10.12688/f1000research.11547.2

Wick, R.R., Judd, L.M., Gorrie, C.L., Holt, K.E., 2017. Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. PLoS Comput. Biol. 13, e1005595. doi:10.1371/journal.pcbi.1005595

Wigley, K., Moot, D., Wakelin, S.A., Laugraud, A., Blond, C., Seth, K., Ridgway, H., 2017. Diverse bacterial taxa inhabit root nodules of lucerne (Medicago sativa L.) in New Zealand pastoral soils. Plant and Soil 420. doi:10.1007/s11104-017-3395-6

Wolf, D.C., Dao, T.H., Scott, H.D., Lavy, T.L., 1989. Influence of Sterilization Methods on Selected Soil Microbiological, Physical, and Chemical Properties. Journal of Environment Quality 18. doi:10.2134/jeq1989.00472425001800010007x

Wood, D.E., Salzberg, S.L., 2014. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol. 15, R46. doi:10.1186/gb-2014-15-3-r46

Wood, S.A., Baudron, F., 2018. Soil organic matter underlies crop nutritional quality and productivity in smallholder agriculture. Agric. Ecosyst. & Environ. 266. doi:10.1016/j.agee.2018.07.025

Woodall, G.S., Western …, C.R.-C.S., 2002, 2002. Direct seeding Acacias of different form and function as hosts for Sandalwood (Santalum spicatum).

Yang, C., Chu, J., Warren, R.L., Birol, I., 2017. NanoSim: nanopore sequence read simulator based on statistical characterization. GigaScience 6, 1–6.

doi:10.1093/gigascience/gix010

Zilber-Rosenberg, I., Rosenberg, E., 2008. Role of microorganisms in the evolution of animals and plants: the hologenome theory of evolution. FEMS Microbiol. Rev. 32, 723–735. doi:10.1111/j.1574-6976.2008.00123.x

Supplementary Materials

Code

All code written for this project is available at:

https://github.com/borevitzlab/ming-honours-code

Further large raw data files will be uploaded and made available at:

https://mdchia.github.io/legume-rhizobia-complex-microbiome/

Full protocols

Where applicable, protocols used in this project have been documented and archived
using protocols.io. A full listing of protocols used in this project is available at:

https://www.protocols.io/private/25F4013B95EBD3B555505D8802E25573

61

Growth chamber layout

Inoculation?		Host?	Amount:
1	None (sterile)	Yes	16
2	Soil slurry	Yes	16
3	Culture mix	Yes	60
4	Soil+culture	Yes	60
5	None (sterile)	No	14
6	Soil slurry	No	16
7	Culture mix	No	60
8	Soil+culture	No	60

B	Barcode	(16)
C	Color checker	(2)

Plants needed	152
Culture (ml)	240
Soil slurry (ml)	152

1	6	11	16	21	26	31	36
2	7	12	17	22	27	32	37
3	8	13	18	23	28	33	38
4	9	14	19	24	29	34	39
5	10	15	20	25	30	35	40
81	86	91	96	101	106	111	116
82	87	92	97	102	107	112	117
83	88	93	98	103	108	113	118
84	89	94	99	104	109	114	119
85	90	95	100	105	110	115	120

41	46	51	56	61	66	71	76
42	47	52	57	62	67	72	77
43	48	53	58	63	68	73	78
44	49	54	59	64	69	74	79
45	50	55	60	65	70	75	80
121	126	131	136	141	146	151	156
122	127	132	137	142	147	152	157
123	128	133	138	143	148	153	158
124	129	134	139	144	149	154	159
125	130	135	140	145	150	155	160

161	166	171	176	181	186	191	196
162	167	172	177	182	187	192	197
163	168	173	178	183	188	193	198
164	169	174	179	184	189	194	199
165	170	175	180	185	190	195	200
241	246	251	256	261	266	271	276
242	247	252	257	262	267	272	277
243	248	253	258	263	268	273	278
244	249	254	259	264	269	274	279
245	250	255	260	265	270	275	280

201	206	211	216	221	226	231	236
202	207	212	217	222	227	232	237
203	208	213	218	223	228	233	238
204	209	214	219	224	229	234	239
205	210	215	220	225	230	235	240
281	286	291	296	301	306	311	316
282	287	292	297	302	307	312	317
283	288	293	298	303	308	313	318
284	289	294	299	304	309	314	319
285	290	295	300	305	310	315	320

Figure S1 (previous page). Physical planting layout and associated treatment codes. Total quantity of inoculum is based on application of 1ml of inoculum to each pot treated.

Sequencing layout

Table S1. Pots with completed sequencing. (See treatment code list for pot treatment.) Empty cells indicate either no plant (for host absent treatments) or that nodules were not transferred to sequencing for that run.

PotID	Treatment	Tray	Sand run	Sand barcode	Nodule run	Nodule barcode
205	7	11	1	9		
207	1	11	1	11	1	4
210	3	11	1	10	1	2
211	6	11	1	12		
212	7	11	1	7		
213	8	11	1	1		
216	5	11	1	5		
217	4	11	1	3	1	8
222	3	12	2	4		
223	6	12	2	9		
224	4	12	2	5	2	10
227	2	12	2	1		
228	4	12	2	6	2	8
231	1	12	2	11		
232	4	12	2	7		
233	8	12	2	2		
234	5	12	2	3		
236	4	12	2	12		

Sequencing results

Table S2. Summary of sequencing results. In the "Sample" field, S denotes sand samples, R denotes nodules, and C is control, used for a non-nodule root sample.

Run	Barcode	Pot	Sample	Reads	Bases	Mean read length	Reads filtered
1	1	213	S	1,901	1,708,173	899	31
1	2	210	R	36,778	58,289,150	1,585	27434
1	3	217	S	6,366	1,027,349	161	20
1	4	207	R	4,224	3,568,210	845	298
1	5	216	S	5,394	1,575,389	292	20
1	6	207	C	2,002	193,375	97	4
1	7	212	S	2,691	623,196	232	4
1	8	217	R	3,095	1,801,050	582	855
1	9	205	S	4,515	3,862,137	855	219
1	10	210	S	712	599,975	843	45
1	11	207	S	5,786	4,784,815	827	41
1	12	211	S	683	394,669	578	6
1	Unresolved			22,990	17,801,228	774	-
2	1	227	S	6,153	13,655,187	2,219	148
2	2	233	S	9,112	24,106,377	2,646	301
2	3	234	S	1,525	3,537,730	2,320	47
2	4	222	S	1,972	4,400,535	2,232	174
2	5	224	S	18,811	41,776,135	2,221	1167

2	6	228	S	21,096	20,159,207	956	934
2	7	232	S	3,715	9,917,605	2,670	216
2	8	228	R	2,955	6,015,611	2,036	2688
2	9	223	S	19,672	52,110,450	2,649	636
2	10	224	R	7,044	12,883,071	1,829	6013
2	11	231	S	72,266	181,352,297	2,510	2278
2	12	236	S	10,395	25,603,796	2,463	499
2	Unresolved			143,145	217,907,838	1,522	-