Title

How Competition Shapes the Fitness Landscape of Rhizobia

Permalink

Author

Rahman, Arafat

Publication Date

2023

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

How Competition Shapes the Fitness Landscape of Rhizobia

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Genetics, Genomics, and Bioinformatics

by

Arafat Rahman

June 2023

Dissertation Committee:
     Dr. Joel L. Sachs, Chairperson
     Dr. Howard S Judelson
     Dr. Ansel Hsiao

The Dissertation of Arafat Rahman is approved:

_____

_____

_____

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

How Competition Shapes the Fitness Landscape of Rhizobia

by

Arafat Rahman

Doctor of Philosophy, Graduate Program in Genetics, Genomics, and Bioinformatics
University of California, Riverside, June 2023
Dr. Joel L. Sachs, Chairperson

In the hundred-plus years since the invention of the Haber-Bosch process, industrial production of nitrogen (N) fertilizers has increased agronomic production and, consequently, the human population. Synthetic N fertilizers also promote the emission of reactive nitrogen forms, which can significantly damage the environment and human health. One alternative to N fertilizer is the use of beneficial nitrogen fixing soil bacteria (rhizobia) that form tumor-like structures called nodules on the roots of legumes. The most effective rhizobia for nitrogen fixation have been used to develop bioinoculants for field application, which have the potential to decrease synthetic N fertilizer use. But importantly, inoculant strains are often outcompeted by indigenous strains that fix little nitrogen.

Because of this, bioinoculants are often only marginally effective, and plant growth suffers. To resolve this rhizobial competition problem and improve the N fixing benefits to plants, we must gain a better understanding of i) competitive interactions among rhizobia, ii) the epidemiological patterns of strains that dominate natural populations and spread among them, and iii) gain insight of N fixing efficiency and developmental variation of bacteroids inside the nodule. This dissertation research uses native Californian legumes in the genus *Acmispon,* which naturally host *Bradyrhizobium* strains. The three parallel aims are to: i) characterize rhizobial competition and model plant performance based on nodule occupancy of competing strains, ii) investigate the epidemiology of rhizobial genotypes across multiple hosts in a large transect (>800 km) of California, and iii) examine the nodule-bacteroid morphology and nitrogen fixation efficiency of rhizobia that vary in N fixation. The main findings from this dissertation reveals a competitive-dominance hierarchy of rhizobial strains which leads to reduced host-benefit, the epidemic spread of certain dominant *Bradyrhizobium* strains in natural populations, and ultrastructural variation in nodule bacteroid morphology during different stages of nodule development. This dissertation research holds significant importance in improving crop yields, reducing dependency on synthetic nitrogen fertilizers, and directly benefiting agricultural research aimed at boosting bioproduct yields from legumes without the use of synthetic nitrogen fertilizers.

Table of Contents

List of Figures

List of Tables

Introduction

Nitrogen is an essential element for life which plays crucial roles in biological, environmental, and industrial processes. Nitrogen passes through various biogeochemical cycles including nitrogen fixation, nitrification, denitrification, and ammonification, all of which can contribute to the balance of nitrogen in the environment (Bernhard, 2010). The environmental balance of nitrogen is important to ensure biological forms of nitrogen are available for use by living organisms while preventing excessive accumulation that can lead to environmental problems such as eutrophication and pollution (J. Erisman et al., 2015). Nitrogen is a fundamental component of organic biomolecules such as nucleic acids and amino acids, which are building blocks of the cell. Therefore, nitrogen rich nutrients are a vital growth requirement for all cellular forms of life. Although dinitrogen gas (N_2) makes up about 78% of the atmosphere, this gas is not directly usable by most organisms due to the high energy requirement to break its triple-covalent bond (Raza, 2020). Thus, nitrogen is a limiting nutrient in most ecosystems.

In the beginning of the twentieth century, Fritz Haber and Carl Bosch invented a chemical process to synthesize ammonia from nitrogen and hydrogen gases (H. Liu, 2014). It was one of the most significant industrial processes developed in the history of humanity since it produces ammonia, the primary precursor of nitrogen based fertilizers. In the hundred-plus years since the invention of the Haber-Bosch process, industrial production of nitrogen (N) fertilizers has increased agronomic production and, consequently, the

human population since it provided a consistent and abundant source of nitrogen fertilizer (J. W. Erisman et al., 2008). But, the Haber-Bosch process is energetically demanding, and the use of synthetic nitrogen fertilizer causes eutrophication and environmental pollution. Synthetic N fertilizers also promote the emission of nitrogen, which (compounded with inefficient use) significantly damages the environment and human health (Townsend & Howarth, 2010). Managing and minimizing use of synthetic nitrogen fertilizer is a major challenge for our current agricultural practices.

Certain bacteria and archaea can convert atmospheric nitrogen into biologically available forms such as ammonia (NH_3) or nitrate (NO_3^-) (J. G. Chen et al., 2018). This process can be achieved by both free-living bacteria and in symbiotic interaction of N_2-fixing bacteria with hosts. The legume-rhizobia association is an elegant example of biological nitrogen fixation by symbiotic interaction, which is a mutually beneficial relationship between leguminous plants, such as peas, beans and clover, and nitrogen fixing bacteria called rhizobia (Catherine Masson-Boivin & Sachs, 2018). This symbiotic interaction occurs in specialized structures called root nodules and involves interchange of signaling molecules to determine specificity and compatibility between symbiotic partners. In this symbiosis rhizobia fix atmoshperic nitrogen to ammonia for plants, in exchange for photosynthates. One alternative to synthetic nitrogen fertilizer is the use of beneficial nitrogen fixing soil bacteria as bioinoculants.

The legume-rhizobia symbiosis is a thoroughly studied system of biological nitrogen fixation, which annually contributes at least 70 million tons of fixed nitrogen (Zahran, 1999). It can increase fertility in arable lands, improve tolerance to abiotic stress, help in phosporous solubilization, and reduce the use of synthetic fertilizers. However, using biological nitrogen fixation (BNF) as a replacement for synthetic fertilizer is challenging for various reasons (Soumare et al., 2020). There is huge variation in nitrogen fixation capability by different strains of rhizobia (Gordon et al., 2016). The most effective rhizobia for nitrogen fixation have been used to develop bioinoculants for field application, which have the potential to decrease synthetic N fertilizer use. But importantly, inoculant strains are often outcompeted by indigenous strains that fix little nitrogen (M. Mendoza-Suárez et al., 2021). Because of this, bioinoculants are often only marginally effective, and plant growth suffers. To resolve this rhizobial competition problem and improve the N fixing benefits to plants, we must gain a better understanding of i) competitive interactions among rhizobia, ii) the epidemiological patterns of strains that dominate populations and spread among them, and iii) the developmental and histological variation of nodule-bacteroids.

In this dissertation I use a native California legume in the genus *Acmispon,* and its natural symbionts in the genus *Bradyrhizobium*. I have three independent projects to study major objectives in my dissertation: **i)** to characterize rhizobial competition and model plant performance based on nodule occupancy, **ii)** to investigate the epidemiology of rhizobial genotypes across multiple hosts in a large transect (>800 km) of California, and **iii)** to examine the ultrastructural variation of bacteroid development in nodules.

Chapter 1 establishes a novel framework to examine competition among rhizobial symbionts to nodulate a host. The approach reduces the rhizobial interaction from the community level to simplified pairwise-combinations in a full-factorial design. This aims to disentangle relative competitive properties of strains apart from their nitrogen fixation trait variability and uncover competitive hierarchies. The main objective of this chapter is to characterize competitiveness, understand the relationship of rhizobial competitiveness with N fixation, and disentangle the mechanisms and drivers of competition.

Chapter 2 investigates the epidemiology and ecological spread of different rhizobial genotypes across a plant metapopulation in California. This chapter indicates that a handful of *Bradyrhizobium* chromosomal genotypes dominate most of the sampling sites among multiple host species and some of them have epidemic spread. This confirms a previously described scenario where epidemic genotypes spread geographically while incorporating locally adapted symbiosis specificity loci. The main objectives of this chapter are to examine the natural distribution of rhizobial strains in a metapopulation, understand the epidemic dominance patterns and study the molecular evolution of these bacteria.

Chapter 3 provides insights on the bacteroid development and morphological differences between beneficial and ineffective rhizobia. To control ineffective rhizobia, legumes have a set of 'host-control' traits which may allow them to selectively associate with beneficial rhizobia, while punishing harmful rhizobial strains. How the host control shapes the morphological difference of nodule-ultrastructure in different stages of development, as well as how bacteroid fitness varies among strains are the major goals of this chapter.

Understanding the factors of rhizobial competition, epidemiology and its evolutionary drivers, and morphological variation of the nodule is crucial to close knowledge-gaps as well as necessary to improve bioinoculants, increase crop yields, and reduce reliance on synthetic fertilizers. This will directly benefit agricultural research to increase legume-based bioproduct yield without help of synthetic nitrogen fertilizer.

Chapter 1

Competitive interference among rhizobia reduces benefits to hosts

Abstract

The capacity of beneficial microbes to compete for host infection – and the ability of hosts to discriminate among them – introduces evolutionary conflict that is predicted to destabilize mutualism. We investigated fitness outcomes in associations between legumes and their symbiotic rhizobia to characterize fitness impacts of microbial competition. Diverse *Bradyrhizobium* strains varying in their capacity to fix nitrogen symbiotically with a common host plant, *Acmispon strigosus,* were tested in full-factorial coinoculation experiments involving 28 pairwise strain combinations. We analyzed the effects of interstrain competition and host discrimination on symbiotic-interaction outcomes by relativizing fitness proxies to clonally infected and uninfected controls. More than one thousand root nodules of coinoculated plants were genotyped to quantify strain occupancy, and the *Bradyrhizobium* strain genome sequences were analyzed to uncover the genetic bases of interstrain competition outcomes. Strikingly, interstrain competition favored a fast-growing, minimally beneficial rhizobia strain. Host benefits were significantly diminished in coinoculation treatments relative to expectations from clonally inoculated controls, consistent with competitive interference among rhizobia that reduced both nodulation and plant growth. Competition traits appear polygenic, linked with inter-strain allelopathic interactions in the rhizosphere. This study confirms that competition among strains can destabilize mutualism by favoring microbes that are superior in colonizing host

tissues but provide minimal benefits to host plants. Moreover, our findings help resolve the paradox that despite efficient host control post infection, legumes nonetheless encounter rhizobia that vary in their nitrogen fixation.

Introduction

Microbial mutualists provide terrestrial plants with diverse services (Friesen et al., 2011), but benefits provided by microbial partners are unreliable, causing unpredictable fitness outcomes for hosts (Heath & Stinchcombe, 2014). For instance, interactions between plants and root-associated mycorrhizal fungi vary from highly beneficial to parasitic (Hoeksema et al., 2010; N. C. Johnson et al., 1997). Epiphytic bacteria that fix nitrogen for tropical host plants also vary broadly in the amount of benefit provided to hosts (Bentley, 1987; Fürnkranz et al., 2008). During host colonization, microbial interactions can favor highly competitive strains irrespective of the level of benefit provided to hosts (Frank, 1996; Joel L. Sachs et al., 2018). To optimize benefits from microbial associations plants employ host control traits, including partner choice and sanctions that reward beneficial microbes and discriminate against strains that provide insufficient benefits (Kiers et al., 2011, 2003). The capacity of microbial partners to compete for host infection, and the ability of hosts to discriminate among them, can introduce an evolutionary conflict between microbe and host partners. The effects of this conflict on the services exchanged in plant microbial mutualism – and the evolutionary stability of these associations – remain poorly understood.

The legume-rhizobia association is an excellent model of mutualism where microbial strains compete to infect hosts. Rhizobia encompasses polyphyletic groups of proteobacteria with the capacity to induce root nodulation and fix nitrogen in legume hosts (Sawada et al., 2003). Partner quality, as measured by relative growth benefit the host gets from rhizobial strains, vary quantitatively due to differences in nitrogen fixation capability

from those that fix substantial amounts to those that are ineffective and fail to provide any benefit for specific host plant partners (Burdon et al., 1999; L. Chen et al., 2002; Collins et al., 2002; Denton et al., 2000; Ehinger et al., 2014; Gano-Cohen et al., 2020). Legumes exhibit a suite of host-control traits to minimize costs of ineffective infections. Host legumes exhibit partner choice, the ability to discriminate against incompatible and uncooperative partners, in this case by detecting molecular signals of nod-factors and effector proteins deployed by type III secretion systems (Poole et al., 2018). Moreover, when ineffective rhizobia gain access to nodules, legumes sanction them by reducing *in planta* proliferation of rhizobia (Oono et al., 2011, 2009; John U. Regus et al., 2017; Westhoek et al., 2017). Given the plant hosts' capacity to select for strains with high net benefit *in planta*, less beneficial strains are predicted to be selected against in the population (Denison, 2000; K. R. Foster & Wenseleers, 2006; Kiers & Denison, 2008; Joel L. Sachs et al., 2004; Ellen L. Simms & Taylor, 2002; S. A. West et al., 2002; Stuart A. West et al., 2002). Nonetheless, strains ineffective at fixing nitrogen are common both in natural and agricultural landscapes, suggesting that other forces counteract host control. These predicted forces include, but are not limited to, mutation-selection balance, stable coexistence of strategies, and selection mosaics shaped by $G \times G$ and $G \times E$ interactions (Batstone et al., 2022; Heath, 2010; Heath & Stinchcombe, 2014; M. Mendoza-Suárez et al., 2021).

The legume-rhizobia mutualism generates the dominant natural input of nitrogen into terrestrial ecosystems (Cleveland et al., 1999) and has agronomic potential to reduce environmental damage caused by nitrogen fertilizer (Choudhury & Kennedy, 2005; Foyer

et al., 2019; Ohyama, 2017). Attempts to inoculate legumes with 'elite' rhizobia, strains that generate a high degree of benefits to plants in lab settings, often results in the inoculant strains being outcompeted by indigenous rhizobia that provide little or no benefits to the introduced host, a phenomenon referred to as the rhizobia competition problem (Irisarri et al., 2019; Yates et al., 2011). However, bioinoculants made from rhizobia that are native to the applied soils can achieve better success than commercial inoculants (Goyal et al., 2021; Koskey et al., 2017; Ouma et al., 2016), indicating importance to characterize rhizobia genotypes for competitiveness and symbiotic growth benefit. Because diverse rhizobia are typically present in natural and field settings, clonal inoculation experiments do not predict performance of rhizobia where multiple strains compete for plant derived nutrients (Bourion et al., 2017; Kiers et al., 2013). Yet, most experiments focus on inoculating clonal bacterial isolates on hosts as a means to evaluate their ability to form successful symbiosis(Friesen, 2012). New work addresses this limitation by incorporating multi-strain inoculations to test hypotheses of fitness outcomes in the legume-rhizobia symbiosis (Batstone et al., 2022; Burghardt et al., 2018).

Here, we characterized variation in competitive ability among rhizobia strains and related interstrain competition to the benefits that hosts derived from the symbiosis. Our objectives were to (i) compare fitness outcomes of rhizobia and hosts in single- and co-inoculation, (ii) characterize the nodulation occupancy distribution of competing strains that vary in symbiotic benefit to the host, and (iii) develop and test models that incorporate symbiotic benefit and nodulation capacity to predict mechanistic processes underlying competition. Experiments were conducted on *Acmispon strigosus* (formerly *Lotus*

strigosus), an annual legume native to the southwestern USA that is nodulated by diverse *Bradyrhizobium spp.* (Gano-Cohen et al., 2020; A. C. Hollowell et al., 2016). We used eight focal *Bradyrhizobium* strains, isolated originally from *A. strigosus*, that range from beneficial to ineffective at nitrogen fixation symbiosis. In prior work, strains were phenotypically characterized on multiple genotypes of *A. strigosus* for nodulation and nitrogen fixation capacity (Gano-Cohen et al., 2020). Strains that fixed nitrogen and improved host growth were classified as effective (i.e., Fix+), whereas those that provided no growth benefit were classified as ineffective (i.e., Fix-). The strains had their genomes sequenced and the Fix- strains contain intact *nif/fix* genes, indicating that ineffectiveness is not a result of deletion or pseudogenization of these genes (Weisberg, Rahman, Backus, Tyavanagimatt, Chang, et al., 2022). Because these *Bradyrhizobium* strains showed consistent response across multiple host genotypes (Gano-Cohen et al., 2020, 2019; Wendlandt et al., 2019), a single inbred line of *A. strigosus* was used in this study. When coinoculated with beneficial and ineffective strains, *A. strigosus* can discriminate among them and sanction strains that do not fix nitrogen for the host, reducing the capacity of nonfixing rhizobial strains to proliferate within nodule tissues (John U. Regus et al., 2017; J. L. Sachs, Ehinger, et al., 2010; Wendlandt et al., 2019). Strains were coinoculated onto *A. strigosus* in a full factorial pairwise experiment, which also included clonally inoculated and uninoculated treatments. Using Illumina amplicon sequencing, more than 1,100 nodules from coinoculated plants were genotyped to detect the occupying strains. Null models were developed from single-inoculation data to predict and then test effects of coinoculation treatments on host nodulation and growth. Understanding the role of

competition among rhizobia to colonize legumes is critical for managing and improving

sustainable agriculture and testing mutualism stability models.

Materials and Methods

Rhizobia and plant genotypes – Eight *Bradyrhizobium* strains were selected that varied quantitatively in the magnitude of growth benefits provided to hosts, including four strains that were categorized as effective because they elicited significant growth benefits in inoculated hosts relative to uninoculated control hosts (i.e., Fix+; #'s 4, 131, 156, 184) and four strains that were categorized as ineffective because they did not cause significant host growth benefits (i.e., Fix-; #'s 2, 186, 187, 200)(Gano-Cohen et al., 2020) **(Table 1.1)**. The host line *A. strigosus* AcS049 was selected for greenhouse experiments, and was initially sampled from the Bernard Field Station, Claremont, CA(J. U. Regus et al., 2017; John U. Regus et al., 2014). Plants were raised from a wild seed progenitor, and were allowed to self for at least two generations in a greenhouse before being used here (Wendlandt et al., 2019).

Inoculation experiment – Seeds were surface sterilized in 5% NaOCl for 3 minutes, rinsed in autoclaved reverse-osmosis water (RO-H$_2$O) for 7 minutes, nick scarified, and sowed into sterilized SC10R Ray Leach Conetainer pots (diameter 3.81 cm, depth 20.96 cm, volume 164 mL, Steuwe and Sons, Corvallis, Oregon, USA) filled with sterilized calcined clay (Turface® Pro League®, Turface Athletics, Buffalo Grove, Illinois, USA) which offers negligible nutrients. Once true leaves formed, seedlings were moved to a greenhouse and fertilized weekly with 1 mL nitrogen-free Jensen's fertilizer (Somasegaran & Hoben, 1994). Fertilization volume was increased weekly by 1 mL until a maximum of 5 mL was reached, which continued until harvest. After 4 days of hardening to greenhouse conditions

under 50% shade, plants were inoculated. Rhizobia inocula were prepared by streaking single colonies onto plates of Modified Arabinose Gluconate medium (MAG (J. L. Sachs et al., 2009a)), scraping grown cells, adjusting cell concentration based on turbidimetric readings, and washing cells in RO-H$_2$O. A Klett-Summerson 800-3 photoelectric colorimeter was used (American Laboratory Trading, Inc., San Diego, California, USA) to get turbidimetric reading of the culture on a KlettTH scale which is proportional to optical density.

The full factorial coinoculation experiment included plants that were treated with each of eight clonal strains, 28 pairwise strain combinations, and uninoculated controls. Plants received 5 mL cultures at concentrations of 1×10^8 cells/mL. This protocol was selected based on previous inoculation experiments showing consistent nodulation of *A. strigosus* at this concentration of *Bradyrhizobium* (Gano-Cohen et al., 2020; J. L. Sachs et al., 2009b; Wendlandt et al., 2019). For coinoculations, the concentration of cultures from each strain were adjusted before combining to reach a total concentration of 1×10^8 cells/mL. Control plants received 5 mL of autoclaved RO-H$_2$O. Each treatment group included 10 plant replicates, and locations for plants were randomized across all treatments. A total of 370 plants were used during the inoculation experiment (10 replicates × 37 treatments [8 single-inoculation, 28 coinoculation, 1 control]). To verify concentrations, each clonal inoculum was quantitatively plated (Somasegaran & Hoben, 2012). Plants were watered daily with 10 minutes of misting. Inoculation occurred in two batches on April 5 and 11, 2019. Each batch received inoculation in five replicates of all treatments.

Plant harvest and nodule genotyping – Plants were harvested starting 28 days post inoculation (dpi) from May 6 - June 19, 2019. Plants were harvested continuously, and individual plants were randomly selected for harvest. Plants were removed from the soil, shoots and roots were photographed, and nodules were dissected and counted. From photographs, nodule area was measured using ImageJ (v1.50i). Ten nodules from each coinoculated plant were randomly selected for rhizobia genotyping and preserved at -80°C. Roots, shoots, and remaining nodules were separated and dried in a convection oven at 60°C to quantify dry biomass. For plants selected for nodule genotyping with fewer than ten nodules, all nodules were genotyped. A total of 63 plants were removed from the dataset, 35 because of algal contamination or multiple plants growing in the same pot, 24 plants that died before harvest, and 4 were removed due to human error during data-collection process. No pattern was observed in plant mortality or in algal contamination across treatments.

A pooled dual-indexed amplicon sequencing approach was used to genotype nodules. Nodules were thawed, surface sterilized in bleach for 30 seconds, rinsed in autoclaved RO-H_2O, and using a sterile pestle ground to a slurry in 200 µL RO-H_2O. The nodule slurry was directly used for a PCR reaction since within nodules DNA from rhizobia exists in high concentrations. Two PCR steps were used to prepare the library (Cruaud et al., 2017). Primer pairs used in the first PCR were designed to amplify the *nifD* gene (target sequence for forward primer 5'-GAAAAGGATATCGTSTTCGGC-3' and reverse primer 5'-GTCRCCRCCGATGTTRTARTC-3') and also included sequences of the standard Illumina sequencing primers and a 0 to 3 bp "heterogeneity spacer" (Fadrosh et al., 2014).

The *nifD* gene contains SNPs that differentiate 27 of the 28 strain combinations (except strains #131 and #156 for which the sequences were identical). The primer pairs used in the second step added adapter sequences and eight-nucleotide index sequences sampled from an index-list using BARCOSEL (Kozich et al., 2013; Somervuo et al., 2018).

In the first PCR, reactions contained nodule slurry (2μl), 5x Q5 buffer (2μl), dNTPs (2mM, 1μl), Q5 Polymerase (0.1μl), forward and reverse primers (2mM, 0.5μl), and molecular grade water (4.7μl). DNA concentration between samples were not controlled since presence-absence of each strain is being measured based on minimal read cutoffs, thereby DNA concentration variation is unlikely to play an important role. PCR conditions were 98°C for 30 seconds for initial denaturation, then 98°C for 10 seconds and 74°C for 30 seconds for both extension and annealing for 35 cycles, and a final extension in 72°C cycle for 2 minutes. In the second PCR, amplicons were dual indexed with multiple identifiers for each sample (Cruaud et al., 2017). The same PCR conditions were used as in the first PCR step. Negative controls, where only DNA grade water instead of nodule slurry was added, were included in each PCR batch. After the second PCR, all amplicons were pooled. The pooled library was cleaned using the HighPrep PCR cleanup (MagBio, USA). An Agilent Bioanalyzer 2100 with the DNA High Sensitivity kit (Agilent Technologies, USA) was used to check for quality and quantity of the library along with a qPCR reaction targeting the adapter sequences (New England BioLabs library quantification kit for Illumina). A PhiX control library was combined with the amplicon library (3.9%). The library was sequenced (2×300 bp) on a MiSeq flowcell, using a 600 cycle V3 MiSeq sequencing kit.

To cross-validate MiSeq genotyping, 90 samples from *nifD* PCR products were genotyped using Sanger sequencing. The 4Peaks software package was used to examine single and dual peaks in electropherograms to identify SNPs that differ between strains (Griekspoor & Groothuis, 2006).

Genomic analyses – Genome content of the eight strains was investigated for presence-absence patterns of previously reported competition-associated genes, including 535 genes associated with nodulation competition in *Rhizobium leguminosarum* (Wheatley et al., 2020) and 128 genes associated with rhizosphere colonization, interstrain interference, or establishment of effective symbiosis and plant-growth promotion across multiple rhizobial taxa (M. Mendoza-Suárez et al., 2021). Gene sequences were downloaded from UniProt, and tBLASTn was used to search for presence of homologs in the eight genome sequences (Camacho et al., 2009; UniProt Consortium, 2021). A filter of e-value < 0.004 and BLAST coverage > 80% was used to summarize the BLAST output. Heatplot was used to visualize gene presence-absence patterns using R (version 4.1.3).

Pangenome analysis was performed to associate gene presence patterns with nodulation success. Prokka (version 1.14.6) was used to annotate the genomes and Roary (3.11.2) was used for pangenome analysis (Page et al., 2015; Seemann, 2014). PantherDB.org was used to test for statistical overrepresentation of gene sets unique to competitive strains with a *Bradyrhizobium diazoefficiens* genome as a reference (Mi et al., 2019).

***In vitro* experiments** – The Fix+ strains had higher nodule occupancy in the coinoculation experiment against Fix- strains, but some Fix+ strain combinations resulted in low host benefit compared to the single inoculation experiment. To quantify growth and cell-cell interactions, growth rate and competition were assayed for the four Fix+ strains using liquid and solid media experiments (Ratzke et al., 2020; Sexton & Tabor, 2020). For liquid experiments, two independent flask cultures of each strain were prepared in MAG media, grown cells were washed, and 10^8 cells per replicate were used to seed 30 mL liquid cultures in minimal Rhizobium Defined Medium (RDM)(J. L. Sachs et al., 2009a). To quantify growth rate and strain interaction effects, treatments were replicated five-fold and were either clonal (i.e., initiated with 10^8 cells each from two different flasks with the same strain) or mixed (i.e., initiated with 10^8 cells each from flasks with different strains) and all pairwise strain combinations were tested. To initially allow cell-cell interactions, cultures were incubated at 29°C and shaken at 60 RPM for 18 hours, increased to 100 RPM for 6 hours, then to 180 RPM for the rest of the experiment. Colorimeter readings quantified doubling time and carrying capacity over 150 hours. The same replicates and strain combinations were repeated for a solid media experiment, wherein 10^8 cells per replicate culture were spread-plated onto 23 mL solid RDM plates and incubated at 29°C for 8 days. After incubation, cells were scraped from plates and quantified for population size using a Klett-Summerson 800-3 colorimeter as described above.

Data analyses – Statistical analyses were carried out using R (version 4.1.3)(R Core Team, 2013). Host growth response to inoculation was estimated by dividing the shoot biomass

values of inoculated plants by the shoot biomass values of uninoculated control plants (Wendlandt et al., 2022). Nodule count and total nodule mass were used to estimate rhizobial fitness at the plant level (Pahua et al., 2018). Nodules from each plant which were selected for genotyping were used to measure mean nodule weight after drying and then multiplied by total nodule numbers to get the nodule biomass estimation. Host investment into symbiosis was quantified as nodule biomass value divided by the shoot biomass value of each inoculated plant (Ortiz-Barbosa et al., 2022). Data transformation was carried out to achieve normality and heteroscedasticity. Linear models were used to investigate variation in host growth response and nodulation traits. For each response variable, two linear models testing effects of treatments, with or without inoculation batch as a random-effect variable, were compared using a log-likelihood test for a significant random effect. Analysis of covariance (ANCOVA) was carried out with type III sum of square errors to test effects of clonal and coinoculation treatments on the response variables (relative growth, total nodules, mean nodule weight, host investment) with dpi as a covariate. Significant differences among treatments were assessed using Tukey's HSD tests. Marginal means were estimated which are useful to compare among treatments since they adjust for the covariate and other factors (Bartlett, 2018). Pearson's product-moment correlation coefficient was computed to assess the linear relationship of nodule count and mean nodule weight with host growth response separately. Coinoculation treatments were categorized as '+/+', '+/-', and '-/-' depending on whether Fix+ (strain 4, 131, 156, 184) and or Fix- (strain 2, 186, 187, 200) strains were paired together within or across groupings. Two sample t-tests were carried out between the coinoculation categories for relative

growth, total nodules, mean fresh nodule biomass and investment, using Holm–Bonferroni correction for multiple comparisons (Holm, 1979).

Nodule occupancy in the coinoculation treatments was inferred by analyzing MiSeq reads. Briefly, quality of the demultiplexed fastq files were assessed with FastQC (KDE Group et al., n.d.). FLASH (Magoč & Salzberg, 2011) was used to merge forward and reverse reads in the fastq files, using 10-100 bp overlap range, 0.3 mismatch ratio, and a 28 Phred score cutoff. For each read in each sample, unique SNPs were compared with each reference sequence (Amanda C. Hollowell et al., 2016). A custom R script (GitHub: https://github.com/acarafat/competition_experiment) was used to assign strain occupancy in the nodules, wherein reads were categorized to a strain if there were > 80% match to the unique SNPs and $\geq$ 10 reads matching the strain. If both coinoculated strains met this criterion, the nodule was classified as being coinfected. For each coinoculation treatment, all genotyped nodules were aggregated to calculate an average strain occupancy. In total, reads from 1125 nodules were analyzed. Nodule occupancy values for each coinoculation treatment (i.e., strains A + B) were tested using a goodness of fit $\chi2$ test against a null model of nodule occupancy wherein an equal number of nodules were expected to be randomly infected by A only, B only, and A + B (i.e., 1:1:1)(Wendlandt et al., 2019). Another χ^2 analysis was carried out to test whether the number of nodules occupied by each strain in co-inoculation treatments significantly differed from the empirically estimated inoculum ratio (i.e., quantitative culturing results; **Table S1**) wherein coinfected nodules were counted as being 50% infected by each strain. To cross-validate sequencing data, MiSeq and Sanger genotyping results were compared and categorized as a 'match'

(identical), 'partial match' (where one approach shows a single genotype but other shows coinfection), or 'mismatch' (where each approach identifies a different genotype). In Sanger sequencing trace files, presence of two chromatogram peaks in multiple known SNP sites was used to identify coinfected nodules. The distribution of coinfected nodules was compared by classifying coinoculated treatment groups in three categories: Fix+/Fix+, Fix+/Fix-, and Fix-/Fix- combinations and a Wilcoxon rank-sum test was used to compare the mean percentage of coinfected nodules in each category followed by adjustment of p-values for multiple testing using Holm-Bonferroni method (Holm, 1979).

Data from the relevant clonal inoculation treatments weighted by nodule occupancy of each participating strain were used to develop null models (Model I and Model II) to infer expected values of symbiosis traits in coinoculated plants. Nodule occupancy for a focal strain 'A' was calculated as the total fraction of nodules that were wholly and partially occupied by the strain (i.e., f_A).

$$Nodule \ \ occupancy, \ \ f_A \ = \ \frac{\#Nodules \ occupied \ by \ A}{\#Total \ Nodules} + \frac{\#Nodules \ coinfected}{2 \times \#Total \ Nodules} \qquad \ldots\ldots$$

(Equation I)

To predict the growth effects of two coinoculated strains, A & B (i.e., *Expected relative growth*$_{AB}$), we summed the growth effect of each relevant strain (in clonal inoculation) weighted by its relative nodule occupancy from the nodule genotyping results (i,e., f_A and f_B). This was the first approach used to predict expected value and hereby

referred as Model I. The same approach was also used to predict expected values for nodule counts and mean nodule biomass:

$$Expected\ Value_{AB} = Value_A \times f_A + Value_B \times f_B \quad \ldots\ldots \text{(Equation II)}$$

One-sample t-tests were used to test whether observed trait values were significantly different from the expected values. However, for the observed difference in relative growth, this test does not resolve if the deviation from expected is due to lower nodulation or lower benefit provided by rhizobia in infected nodules compared to clonal inoculation. To test whether the observed difference from expected was due to lower per nodule symbiotic benefit provided by the rhizobia or not, another expected relative growth was calculated. For this second approach (Model II), average per nodule growth benefit of each strain is calculated for single inoculation. The number of nodules formed by each strain in coinoculation, if they were infected by a single strain (i.e. no coinfected nodules), was calculated by multiplying the frequency of each strain from nodule occupancy data by the total number of nodules formed in coinoculation. Finally, per nodule expected growth benefit was calculated by multiplying the strain means for per nodule host benefit in single inoculation by the number of nodules formed with each strain in coinoculation and summing the resulting values for the two strains. A linear regression model was used to compare both expected values (i.e., Model I, nodule occupancy weighted relative growth and Model II, nodule-normalized relative growth) with the observed value, and a one-way

ANOVA was used to compare observed relative growth with expected values predicted by these two models, followed by a Tukey's HSD test.

A Pearson correlation between residuals of relative growth versus nodule counts and mass was used to evaluate association between observed and expected relative growth, nodule number, and nodule mass. A Durbin-Watson test was used to check for autocorrelation in the residuals (Savin & White, 1977). To understand the general performance of all coinoculation treatments that contain a specific strain, mean values for the number of nodules, relative growth, and nodule occupancy were calculated for each strain across all coinoculated treatment combinations using the following formula:

$$Mean\ Count\ Value,\ \text{MCV}_A = \frac{(\sum_{A \neq B} Count\ in\ Treatment_{AB})}{(n-1)} \quad \text{......} \quad \text{(Equation III)}$$

A Welch two-sample t-test was used to test for significant differences in mean count values of nodule occupancy, relative growth, and number of nodules between Fix+ and Fix- traits.

For the *in vitro* analyses, growth curve data from strains grown in a liquid medium were analyzed using the Growthcurver package (v0.3.1) in R which fits a logistic equation based on point estimates of bacterial population size(Sprouffske & Wagner, 2016). Using these data, doubling time and carrying capacity were estimated. ANOVA and post-hoc Tukey's HSD tests were used to investigate differences between treatments in each clonal and competition experiment. Doubling time and carrying capacity of each strain was used to make null predictions for the mixed strain combinations by calculating mean values for

strain pairs. One-sample t-tests were used to compare observed means in competition and

predicted values for each variable.

Results

Host benefit depends on clonal inoculation genotype – We first evaluated the effects of clonal inoculations on hosts to cross-validate prior results, establish baselines, and to model effects of coinoculations on relative growth and nodulation response of plants. Eight strains were individually inoculated onto *A. strigosus* plants grown in a greenhouse and caused host responses that closely matched previous results (Gano-Cohen et al., 2020; J. L. Sachs, Ehinger, et al., 2010) **(Fig. 1.1, Table S1.1, S1.2)**. Specifically, plants inoculated with the Fix+ strains 4, 131, 156, and 184 had more biomass than uninoculated control plants. Three of these strains elicited significant relative growth benefits, with the exception of marginal strain 4 (t = 1.84, df = 5, $P < 0.0626$; **Table S1.2, Fig. 1.1A).** Conversely, Fix- strains 2, 186, 187, and 200 were confirmed as ineffective, as host biomass values were not significantly greater than those of control plants ($P > 0.05$; **Table S1.2**). Relative host growth response varied significantly among the clonal strain treatments (**Table S1.3**, $F_{(7,56)}$ = 11.24, $P < 0.001$). No significant random effect of inoculation batch was found on single- or co-inoculation results, hence that the design can be analyzed as a single experiment **(Table S1.4)**.

Host nodulation response varied significantly among the clonal inoculation treatments, regardless of the Fix+/Fix- phenotype of strains (total nodules $F_{(7,56)}$ = 7.009, P = < 0.0 1; mean nodule biomass $F_{(7,56)}$ = 18.092, P = < 0.001; host investment (i.e., nodule proportion of total plant biomass $F_{(7,56)}$ = 8.358, P = < 0.001). Ineffective strains 2 and 200 elicited relatively low numbers of nodules (i.e., estimated marginal means of nodules < 10 in most cases, **Table S1.5**), while the remaining strains, including effective and ineffective

ones, formed 16-35 nodules per plant **(Fig. 1.1B; Table S1.5)**. Host growth response was positively correlated with nodule count (Pearson's product-moment correlation $R^2 = 0.697$, $t = 8.24$, df $= 72$, $P < 0.001$) and mean nodule weight (Pearson's product moment correlation $R^2 = 0.63$, $t = 6.746$, df $= 70$, $P < 0.001$). This pattern is consistent with host control over resource flow into nodules reflected by nodule-size, where within-nodule fitness of ineffective strains is reduced by plants (John U. Regus et al., 2017).

Strain 156 provided low relative growth benefit (mean shoot RG $= 1.94$, $P < 0.001$, **Fig. 1.1, Table S1.2**) while inducing the highest host investment among the strains tested (estimated marginal mean 1.111, **Table S1.5**). The host growth and nodulation response variables of strain 156 were consistent with previous findings that tested nodulation and growth benefit of this strain on multiple sympatric and allopatric host-lines (Gano-Cohen et al., 2019).

Host benefit varied in coinoculation treatments – To examine the effect of strain-strain competition on mutualistic symbioses, we tested all 28 possible pairs of strain combinations. Host benefit, nodulation count, mean nodule biomass, and host investment all varied significantly among the 28 coinoculation treatments (host benefit $F_{(27,200)} = 5.455$, $P < 0.001$; total nodules $F_{27,202} = 2.63$, $P < 0.001$; mean nodule biomass $F_{27,199} = 2.782$, $P < 0.001$; investment $F_{27,197} = 3.41$, $P < 0.001$; **Table 1.2, Fig. 1.2**). Like the clonal inoculation treatments, relative growth was positively correlated with nodule counts (Pearson's product-moment correlation $R^2 = 0.7619$, $t = 18.229$, df $= 240$, $P < 0.001$) and

with mean nodule weight (Pearson's product moment correlation $R^2 = 0.39$, $t = 6.7537$, df $= 241$, $P < 0.001$).

Only 8 of the 22 coinoculation combinations having at least one Fix+ strain caused significant host growth relative to uninoculated controls (one-sample t-test; $P < 0.05$; **Table S1.6, Fig. 1.2A**). Five additional coinoculation treatments including at least one Fix+ strain provided growth benefits that were marginal (one-sample t-test; $0.05 < P < 0.10$; **Table S1.6**). These marginal or no benefit treatments included those for which at least one or both inoculant strains provided a significant benefit in their corresponding clonal inoculation treatments. Surprisingly, only 3 out of 6 Fix+/Fix+ (+/+) coinoculation treatments elicited significant relative growth > 2.5 (**Table S1.6**). Other treatments that elicited significant host growth included 5 of 16 Fix+/Fix- (+/-) combinations **(Table S1.6)**. None of the Fix-/Fix- (-/-) coinoculation combinations provided a significant host growth benefit.

When the coinoculated treatments were grouped and compared by +/+, +/-, and -/- combinations, plants that received at least one Fix+ strain grew two-fold more and formed significantly more nodules than hosts that were coinoculated with two Fix- strains **(Fig S1.1)**. There were significant differences between +/+ vs -/- and +/- vs -/- categories for relative growth benefit (Welch's t-test, +/+ vs -/- $P = 0.0035$, +/- vs -/- $P = 0.0043$; **Fig. S1.1A**). The number of nodules also significantly differed between +/- vs. -/-, likely because plants that received Fix+ were larger (Welch's t-test, +/- vs -/- $P = 0.001$; **Fig. S1.1B**). There were no significant differences in fresh nodule biomass and host investment values among any of the +/+, +/-, and -/- categories (**Fig. S1.1C and S1.1D**). However, for

coinoculated treatments 156+200, 186+156, 187+156, 2+131 and 4+156, investment in nodulation was high (> 1) **(Table S1.7)**. These treatments were comprised by one Fix+ and one Fix- strain, and the plants received no significant growth benefit from inoculation (**Fig. 1.2D-E, Table S1.7**).

Fix+ strains dominate over Fix- strains in nodule occupancy – We used amplicon sequencing of the *nifD* locus and a dual-indices barcoding method to genotype rhizobia in nodules from each plant in coinoculation treatments. Genotyping nodule-occupying bacteria of coinoculated plants revealed that in +/- combinations, effective strains dominated nodules relative to ineffective strains (**Fig. 1.3A**). Non-random nodule occupancy was found for all coinoculated treatments (i.e., χ^2 test, rejecting the null of equal nodule-occupancy among strains). Non-random nodulation was also found for most coinoculated treatments when the inoculum ratio was used as a null (i.e., χ^2 test, rejecting the null for all treatments, except 4+131 and 4+156 where most of the nodules were coinfected; **Table S1.8**). Proportion of coinfected nodules were significantly higher in +/+ combinations (mean ± se = 67.25 ± 10.08) compared to +/- (mean ± se = 18.20 ± 3.94) or -/- (mean ± se = 23.52 ± 8.58) combinations (**Figure 1.3B**). The dominant strain in each pair was determined by a majority occupancy of nodules (**Figure 1.3C, Table S1.8**). Mean nodule occupancy values were consistent with a linear dominance hierarchy, where 131 & 156 > 4 > 184 > 186 > 187 > 2 > 200 (the top four strains are Fix+) (**Fig. 1.3D**). The presence of a Fix+ strain increased mean nodule occupancy in coinoculation (Welch two sample t-test, t = 3.7941, df = 4.6801, P = 0.01437) as well as mean relative growth benefit

(Welch two sample t-test, t = 4.0876, df = 3.7289, *P* = 0.01728), however no significant effect was observed for mean nodule biomass (Welch two sample t-test, t = 1.766, df = 4.5128, *P* = 0.144) (**Fig. 1.3E, 1.3F**). Strain 156 provided the lowest benefit among the four beneficial strains, but it dominated nodules relative to all ineffective strains, and formed a high proportion of coinfected nodules when coinoculated with effective strains, except 131 from which it could not be genetically differentiated when assessed based on *nifD* alleles (**Fig. 1.3D-F**). As in the clonal inoculation results, strain 156 induced a pattern of host growth effects and nodulation that were consistent with a strategy of evading host sanctions despite providing marginal benefits (Jones et al., 2015).

Nodule genotyping was validated by comparing Sanger and MiSeq sequence data for 90 samples. Both technologies yielded high-quality reads from 51 nodules, and both yielded complete genotype matches in 47% of them. On the other hand, in 41% of nodules sequenced, one technology detected both strains but the other detected only one of them, likely reflecting the differences in sensitivity between Sanger and MiSeq, while the remaining samples had no match (**Table S1.9**).

Coinoculated plants receive less host growth benefit than predicted from clonal inoculation – Based on data from the single strain inoculation results, models were developed to infer expected values of symbiosis traits in coinoculated plants. Two models were developed and compared. Model I weighs symbiosis traits based on nodule-occupancy and allowed us to test whether the observed overall plant growth benefit of each strain in co-inoculation was significantly different from that expected based on the values

from single-inoculation experiments. Model II normalizes data based on the number of nodules formed and allowed us to tease apart some of the underlying mechanisms driving the overall reduction in plant growth in co-inoculation experiments, including whether strains provided a lower degree of benefits, when controlling for the number of nodules formed. Twenty-seven coinoculated combinations were tested while the 156+131 combination was excluded. Under Model I, weighing for nodule occupancy, fourteen strain combinations induced significantly less host growth than expected based on clonal inoculation data (equation II). None of the coinoculated plants grew significantly more than expected **(Table S1.10)**. Similar patterns were observed for total number and area of nodules, in which observed trait values were almost always lower than the expected trait values **(Table S1.10)**. Among the 14 coinoculation treatments that produced a lower host growth response than the expected values, nine also had significantly lower observed nodulation relative to the values expected from the null model **(Fig. 1.4A-B)**. Only four of these coinoculation treatments also had significantly lower nodule areas than expected **(Fig. 1.4C)**.

Under Model II, normalizing based on nodule counts, the predicted relative growth had a better fit with the observed data (slope 0.93, $P < 0.001$; adjusted R^2 0.6824, $F_{1,232} = 501.7$, $P < 0.001$) compared to Model I (slope 0.334, $P < 0.001$; adjusted R^2 0.063, $F_{1,236} = 16.92$, $P < 0.001$). ANOVA revealed a significant difference between the expected shoot RG values in both models and the observed values. Based on post-hoc Tukey HSD the expected shoot RG values based on model I (i.e., nodule-occupancy weighted method) were significantly greater than those observed (0.639 difference in mean, $P < 0.001$). Under

model II, there was no significant difference between the expected and observed shoot RG values (0.139 difference in mean, P = 0.117). This suggests that the reduction in plant growth in coinoculation relative to single inoculation was not driven by a reduction in the per nodule benefit each strain conferred to its host.

The deviation in growth of coinoculated plants from expected values was tightly correlated with nodule number differences (Pearson's product moment correlation R^2 = 0.86, P < 0.001; **Fig. S1.2)**. No autocorrelation was found in a Durbin-Watson test (D-W = 2.544, P = 0.122). In combination with the above data, we interpret this to mean that lower than expected benefits in coinoculation is associated with a reduction in nodule counts, consistent with inter-strain interference reducing both nodulation and net growth benefits.

Gene content differences suggest that competition traits are polygenic and linked with inter-strain interactions – To characterize why some inoculant strains were more competitive relative to others, the gene content of each strain was compared. Genome sequences of the eight strains were analyzed for presence/absence variation (PAV) in 663 genes previously reported to be associated with competition (M. Mendoza-Suárez et al., 2021; Wheatley et al., 2020). Among the eight genomes, 130 genes exhibit PAV (**Fig. S1.3-4)**. PAV patterns of strain 156 and 131, the two most competitive strains, were compared to the other strains. Only one gene predicted to encode a hypothetical protein was found to be uniquely present in strain 131. No genes were uniquely present in strain 156 when strain 131 was excluded from the comparison. Compared to other Fix+ strains, strains 156 and 4

lacked genes associated with flagella-based motility (i.e., flagellin, probable flagellum biosynthesis repressor protein, putative flagellar synthesis related protein, putative chemotaxis *MotC* protein, flagellar hook-associated protein 1, flagellar motor stator protein *MotA* etc.; **Figs. S1.3-4**). Strain 156 also contains several *virB* gene homologs associated with the type IV secretion system (T4SS) which are absent in other Fix+ strains (**Fig. S1.3**). These T4SS genes are likely associated with conjugation of mobile genetic elements, such as integrative and conjugative elements and plasmids(Costa et al., 2021; Grohmann et al., 2018) . Finally, strain 156 contains several metabolic pathway genes that are not present in strain 131, including enzyme families of glucuronosyltransferase, glutamine synthetase, galactosyl transferase, phophoribosylformylglycinamidine cyclo-ligase, phosphoribosylglycinamide formyltransferase, and exopolyphosphatase (**Fig. S1.4**).

Pangenome analysis of the eight focal genomes uncovered 39,676 gene families, the vast majority of which were defined as 'cloud' genes (i.e., present in < 15% of strains, **Fig. S1.5**)(Sitto & Battistuzzi, 2020). Strain 131 and strain 156 have 984 and 1173 unique genes, respectively, in comparison to the six other strains. However, of those, only 70 had a single-copy homolog present in both genomes compared to the remaining six strains. For strain 131, 85 unique genes with functional annotations were found but 61 are predicted to be associated with functions common among insertion sequence (IS) elements. Among the remaining unique genes of strain 131, putative encoded functions included efflux pumps (nickel and cobalt resistance protein *CnrB*, cation efflux system *CusA*), and antibiotic resistance protein (*AbaF*). For strain 156, 5 out of 43 annotated unique genes are IS elements, but no gene-ontology based functional enrichment was found for the rest of the

genes. Strains 131 and 156 contain genes that encode for several light-activated proteins, including blue-light activated and photosystem I assembly proteins, a category of genes that regulate root attachment during nodulation(Bonomi et al., 2012). Sixteen unique genes encoding non-hypothetical proteins were found in both strain 131 and strain 156, including prophage integrase gene *IntA*, which is also a site-specific recombinase required for conjugative transfer of symbiotic and non-symbiotic ICEs (Hernández-Tamayo et al., 2013).

Competitive strains grow faster than less competitive strains – Four Fix+ strains were cultured in solid and liquid media in clonal and pairwise coinoculation to assess competition traits *in vitro*. During clonal growth in liquid minimal media, strain 156 had by far the fastest doubling time (8.82 hours) and strain 4 was slowest (14.468 hours), whereas strain 156 had the lowest carrying capacity (8.96×10^{10} cells), and strain 4 was the highest (9.04×10^{10} cells; **Table S1.11**), differences that were significant among strain treatments (**Tables S1.11**).

In the mixed strain experiments, carrying capacity had a significant treatment effect in the ANOVA followed by Tukey HSD post-hoc test (**Table S1.12, S1.13**). Based on null predictions from clonal results, the mixed strain treatment 4+184 was found to have significantly reduced carrying capacity (t = -2.84, df = 4, P = 0.046) and 156+184 (t = 4.897, df = 4, P = 0.008) was found to be significantly slower in doubling time (**Table S1.14**).

In the solid media experiments that estimated population size, only strain 4 was significantly different from other treatments with a lower total population (2.25×10^8 cells/mL; **Table S1.15**).

Discussion

Our results suggest four main conclusions about interactions among rhizobia strains during the nodulation process. Firstly, the linear dominance hierarchy that we uncovered indicates that competitive ability for nodulation is genetically determined, and not altered by emergent effects of specific strain interactions, which could generate a nonlinear hierarchy, or no hierarchy at all. It is likely that the hierarchy we uncovered does not perfectly reflect natural populations, given the high-density of inocula we used in pairwise combinations and the otherwise sterile conditions, whereas in nature strains are likely to compete with a multitude of other rhizobia strains, and with other microbes as well. Nonetheless, dominance hierarchies were also uncovered in rhizobia communities that were inoculated onto *Acacia* hosts, although these varied depending on host species (Vuong et al., 2017). Population data are also consistent with dominance hierarchies; a genotypic meta-analysis of rhizobial populations reported that a handful of strains dominate nodules in host populations, with individual strains occupying more than 30% of those nodules (Mcinnes, 2004). Secondly, all four effective strains dominated nodule occupancy against ineffective rhizobia in terms of number of nodules inhabited, which provides a partial explanation of the observed dominance hierarchy. This dominance pattern of Fix+ strains over Fix- strains in coinoculation is observed irrespective of the strain they were coinoculated with, consistent with previous evidence of host sanctions that are robust to diverse strain identities (J. L. Sachs, Russell, et al., 2010). Previous work also suggested that sanctions are robust to other sources of variation, showing that ineffective rhizobia are sanctioned independent of the level of extrinsic fertilization in the soil (Kiers et al., 2006; John U.

Regus et al., 2014), or the host genotype (Gano-Cohen et al., 2019; Ortiz-Barbosa et al., 2022). This type of host-control can be conditional, as data from peas suggests that host sanctions depend on the magnitude of nitrogen fixation in competing strains, where intermediate fixers are tolerated only if a better strain is not available (Westhoek et al., 2021). Our data suggests some degree of conditionality. For instance, when plants were coinoculated with two Fix+ strains, prevalence of coinfected nodules is significantly high compared to treatments with one or two Fix- strains in coinoculation. The prevalence of coinfected nodules is also high compared to other published work in similar or different host-rhizobia systems (i.e. *Bradyrhiozbium-Acmispon*(Wendlandt et al., 2019) or *Rhizobium leguminosarum/Trifolium* (M. A. Mendoza-Suárez et al., 2020)). Thirdly, while evidence suggests that hosts were able to selectively favor beneficial versus ineffective strains for nodule occupancy, the host-control among Fix+ strains – which varied quantitatively in their nitrogen fixation benefits – does not suggest discrimination against marginally beneficial strains. In particular, strain 156, the fastest growing strain *in vitro*, occupied a higher number of single-infected nodules when competed against strain 184, and formed >90% coinfected nodules against strain 4. Although we cannot differentiate strain 156 from 131, and we cannot determine the relative occupancy of each strain in coinfected nodules, taking account of its very low host-benefit in clonal inoculation and coinoculation indicates plant hosts can be less effective at sanctioning this strain in the presence of other effective strains (Gano-Cohen et al., 2019). In the broader context, such phenotypes can be viewed as an extreme end in a continuum of benefits to mutualist partners, where a strain can evade host-sanctions despite providing minimal host-benefit

36

(K. R. Foster & Wenseleers, 2006; N. C. Johnson et al., 1997; Nancy Collins Johnson & Graham, 2013; Jones et al., 2015). These results suggest that hosts and symbionts may be in conflict over the magnitude of resources exchanged, whereby some symbionts that provide minimal host benefits nonetheless receive greater host investment compared to more beneficial symbionts, thereby having a fitness-advantage over other strains (Porter & Simms, 2014; Price et al., 2015). Fourth, and perhaps most urgent for the application of rhizobia in agriculture, we found that in pairwise coinoculation, hosts received significantly less benefit from rhizobia than expectations based on the clonal inoculations. A significant reduction was found in ~50% of strain combinations (**Fig. 1.4A**), and in no cases did coinoculated hosts receive significantly greater than expected benefits. However, this study only tested pairwise combinations whereas in agricultural or natural settings, more strains are naturally present and participate in the symbiosis, and it remains to be seen how competition in a community of rhizobia impacts symbiotic benefit potential compared to clonal inoculation. These data, combined with the parallel reduction in the number of nodules formed in coinoculated plants, suggest that strain interactions reduce both the number of nodules formed and the net benefit received by hosts.

Strain interactions throughout the host infection process can favor competitiveness and erode the net benefit of symbiosis (Frank, 1996; Joel L. Sachs et al., 2018). To predict effects of rhizobia in mixed strain populations, we developed null models of host benefit parameterized with empirical data from clonal infections as well as the genotypic data on nodule occupancy in coinoculation. In a majority of coinoculated treatments, the observed relative growth was significantly lower than expected **(Fig. 1.4B)**. Strikingly, the

coinoculation treatment of 131+184 generated no significant host growth benefits (**Fig. 1.2A**), although these were the two highest benefiting strains in the clonal setting (**Fig. 1.1A**). A similar trend was also found for nodulation by these two strains, where coinoculated plants formed among the fewest nodules in any coinoculation treatment **(Fig. 1.2B)**. More generally, the lower nodulation in coinoculation compared to clonal treatments, irrespective of nitrogen-fixation effectiveness, suggests that interference among strains is reducing nodulation. Rhizobia strain interactions exhibit antagonism *in vitro* (Maan & Garcha, 2018; Schwinghamer, 1971). Native *Bradyrhizobium* and *Rhizobium* inhibit growth of other strains in culture and in coinoculation on hosts, where bacteriocin producing strains are found occupying more nodules relative to non-producing strains (Goel et al., 1999). The potential for strains to be in conflict should be taken in consideration when preparing high performing bioinocula to improve agricultural yield. A variety of interstrain competitiveness traits have been identified, including diverse bacteriocins, altered mobility, and metabolic capabilities of strains to utilize complex hydro-carbon chains (Wielbo, 2012). Our work suggests that competitiveness is determined by the rhizobia genotype, and is highly polygenic, shaped by functions such as conjugation and integration (**Figs. S1.3-4; Table S1.12**). The ability to acquire novel genomic elements could allow strains to acquire loci that affect root attachment (Bonomi et al., 2012), as well as antibiotic and resistance functions that can modulate inter-strain allelopathy (i.e., *cnrB*, *cusA*, and *abaF*). These data are also consistent with traits of competitiveness for nodulation and efficiency of nitrogen fixation being independent (Bourion et al., 2017). The *in vitro* experiments revealed that in some strain combinations,

both doubling time and carrying capacity of rhizobia can be reduced in mixed populations relative to clonal ones **(Table S1.13)**. Also, the lack of difference of observed per nodule benefit in coinoculation from expected values based on model II indicates that the per nodule growth benefit from coinoculation does not vary significantly from expected performance based on single inoculation **(Table S1.11)**. Overall, these data suggest that reduced host growth and reduced nodulation in coinoculated plants is largely driven by competitive interference among strains that occurs during the initial colonization of host roots – likely before nodule formation.

A longstanding goal in symbioses research is to resolve the degree to which fitness is aligned between partners (Burghardt et al., 2018; Friesen, 2012; Gano-Cohen et al., 2019; Heath, 2010; Kimbrel et al., 2013; Porter & Simms, 2014; Quides et al., 2021; Joel L. Sachs et al., 2018). A meta-analysis observed that the fitness interests of rhizobia and plant hosts are aligned (Friesen, 2012). However, analyses were based largely on sets of single inoculation experiments, which cannot reliably predict performance of rhizobia in natural settings where multiple strains simultaneously compete for plant derived nutrients (Kiers et al., 2013). In a handful of experiments which compared both clonal and a community inoculation, inoculation of single bacterial strains to a host plant was useful to evaluate the genotype's ability to form successful symbiosis but could not predict competitiveness with other strains (Bourion et al., 2017). Results from our coinoculation experiments indicate that beneficial strains are consistently more competitive than ineffective ones, but among beneficial strains, the dominance cannot always be confirmed due to a high number of coinfected nodules and technical limitations to determine relative nodule occupancy in

coinfected nodules. However, the higher number of coinfected nodules among Fix+/Fix+ treatments reinforces the idea that host-control has a threshold of benefit above which plants cannot effectively defend against strains that provide marginal benefits. Only in 4 of the 27 treatment combinations were nodules found to be singly infected by each strain alongside the presence of coinfected nodules, whereas in the remaining treatments, nodules were either singly infected by the dominant strain or coinfected by both strains **(Figure 1.3A)**. This indicates that although all strains have the ability to form nodules on their own during single inoculation, only the dominant strain was able to form single-infected nodules in most of the cases, whereas the competitor strain was only able to coinfect nodules. Formation of a high number of mixed nodules with no negative effect on plant growth has been reported where low- or non-beneficial strains of *Sinorhizobium meliloti* can evade sanctions by *Medicago sativa* in the presence of a highly beneficial strain(Checcucci et al., 2016). When focusing only on the strains that were effective on *A. strigosus,* all of them in coinoculation showed higher dominance over ineffective ones **(Fig. 1.3B)** and elicited high mean nodule numbers **(Fig. 1.3D)**. Although single-strain inoculation is less ecologically relevant since it excludes competitive interactions among strains, our results show that it is still predictive of the per nodule growth benefit that a strain provides to its host in coinoculation.

Host control by legumes engages rhizobia at two stages of the infection process. The first stage, partner choice, involves flavonoid signals that hosts release, promoting responsive signals in the rhizobia, including nod factors and effectors (D. Wang et al., 2012). This signal exchange winnows the pool of microbes that gain access to the root

surface, and selects against nodulation by incompatible strains (Walker et al., 2020; Q. Wang et al., 2018). Signal exchange may also involve plant immunity and rhizobial effectors secreted by the type III secretion system (T3SS)(Nelson & Sadowsky, 2015; Yasuda et al., 2016). However, this initial stage is limited in its efficacy, as many ineffective strains can and do gain access to host nodules, even strains that cannot nodulate themselves can coinfect nodules alongside nodulating strains (Gano-Cohen et al., 2020). When legumes encounter rhizobia mutants that vary markedly in their capacity to fix nitrogen, but do not otherwise differ genetically from the parental strain, the plant host cannot differentiate among them prior to nodulation(Amarger, 1981; Hahn, 1986; Westhoek et al., 2017). The second stage of host control, sanctioning that occurs within nodules, is efficient at punishing non-fixing strains (Denison, 2000; Kiers et al., 2003; Oono et al., 2009; John U. Regus et al., 2017; Wendlandt et al., 2019; Westhoek et al., 2017). In our data, strains 2 and 200 elicited the lowest nodule counts in single inoculation, as well as lowest mean nodule occupancy in coinoculation, whereas strains 186 and 187 have the opposite pattern **(Figure 1.1B and 1.3C)**, which reflects a difference in the competitiveness level of Fix- strains as well as a variation in host-control. Hosts can sanction nonperforming symbionts, but the threshold that triggers this mechanism is unknown, and there may be a cost to this action. Therefore, it may not be beneficial for the host to sanction every non- or low-performing rhizobia (Goyal et al., 2021). Our findings help resolve the paradox that despite efficient host control post infection, legumes nonetheless encounter strains that generate only moderate host benefit compared to what is possible from single infection in symbiosis (Heath & Stinchcombe, 2014). Despite

efficiency of host control after nodulation, the host appears to have limited ability to overcome the reduction in growth benefits associated with competitive strain interference in the rhizosphere.
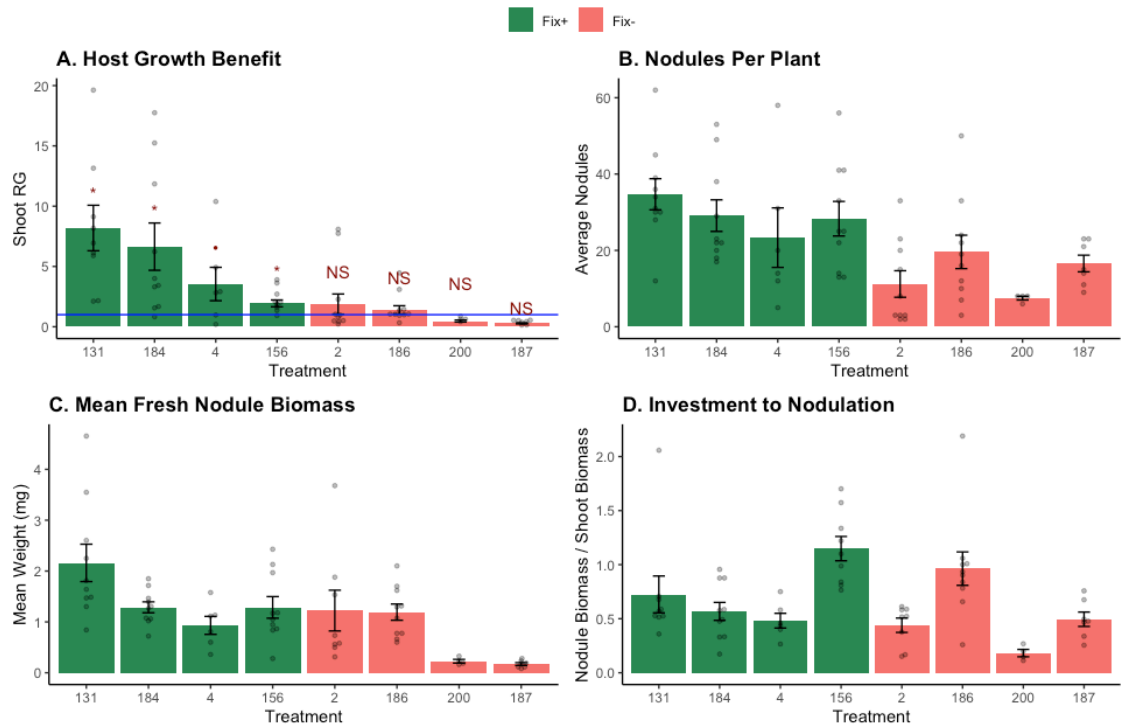
Fig. 1.1 Plant response in clonal inoculation treatments. **(A)** Shoot relative growth (RG) indicates host growth benefit from inoculation relative to uninoculated controls. *P*-values between 0.001-0.05 are indicated with single asterisk (*), between 0.05-0.1 with a period (.), and above 0.1 is NS (non-significant). **(B)** Average number of nodules per plant, **(C)** mean fresh nodule biomass, and **(D)** investment to nodulation are shown. Turquoise color represents Fix+ strains and red color represents Fix- strains. For panels b-d, the strains are ordered left to right based on being most to least beneficial based as determined by growth data in **(A)**.
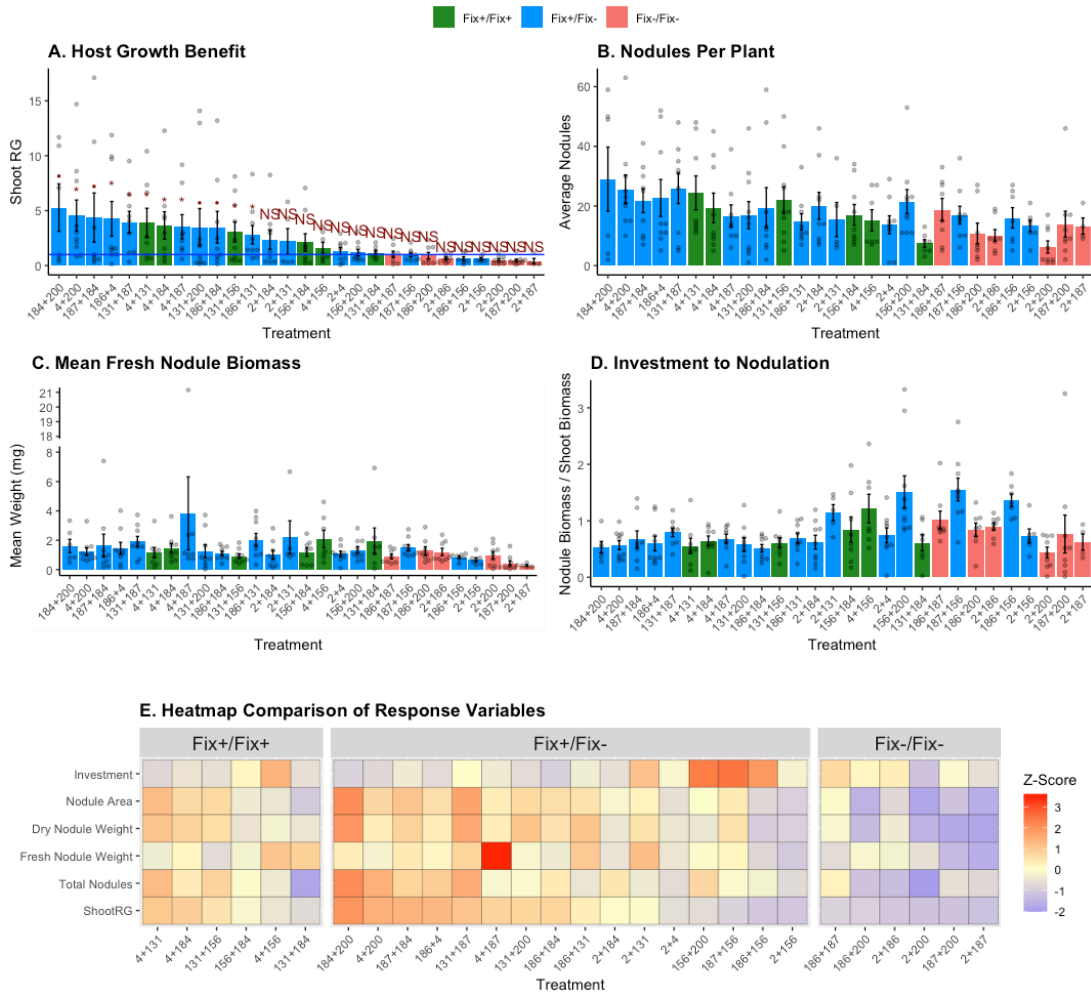
Fig. 1.2 Plant growth and nodulation responses to coinoculation treatments. **(A)** Mean host relative growth is shown for each coinoculation treatment with the horizontal blue line indicating mean value of the uninoculated controls. *P*-values between 0.001-0.05 are indicated with a single asterisk (*), between 0.05-0.1 with a period (.), and above 0.1 is NS (non-significant). **(B)** Mean nodule counts, **(C)** mean nodule biomass, and **(D)** host investment into symbiosis are also indicated. Coinoculation treatments are organized from the most to the least beneficial in panels a-d. Bars indicate means and error bars indicate the standard error of the mean. **(E)** A heatmap compares all response variables by standardizing the values into Z-scores while breaking them in Fix+/Fix+, Fix+/Fix-, and Fix-/Fix- treatment groups. Bars are colored red when the coinoculation treatment was composed of two ineffective strains (Fix-/Fix-), blue for one effective and one ineffective strain (Fix+/Fix-), and green for two effective strains (Fix+/Fix+).
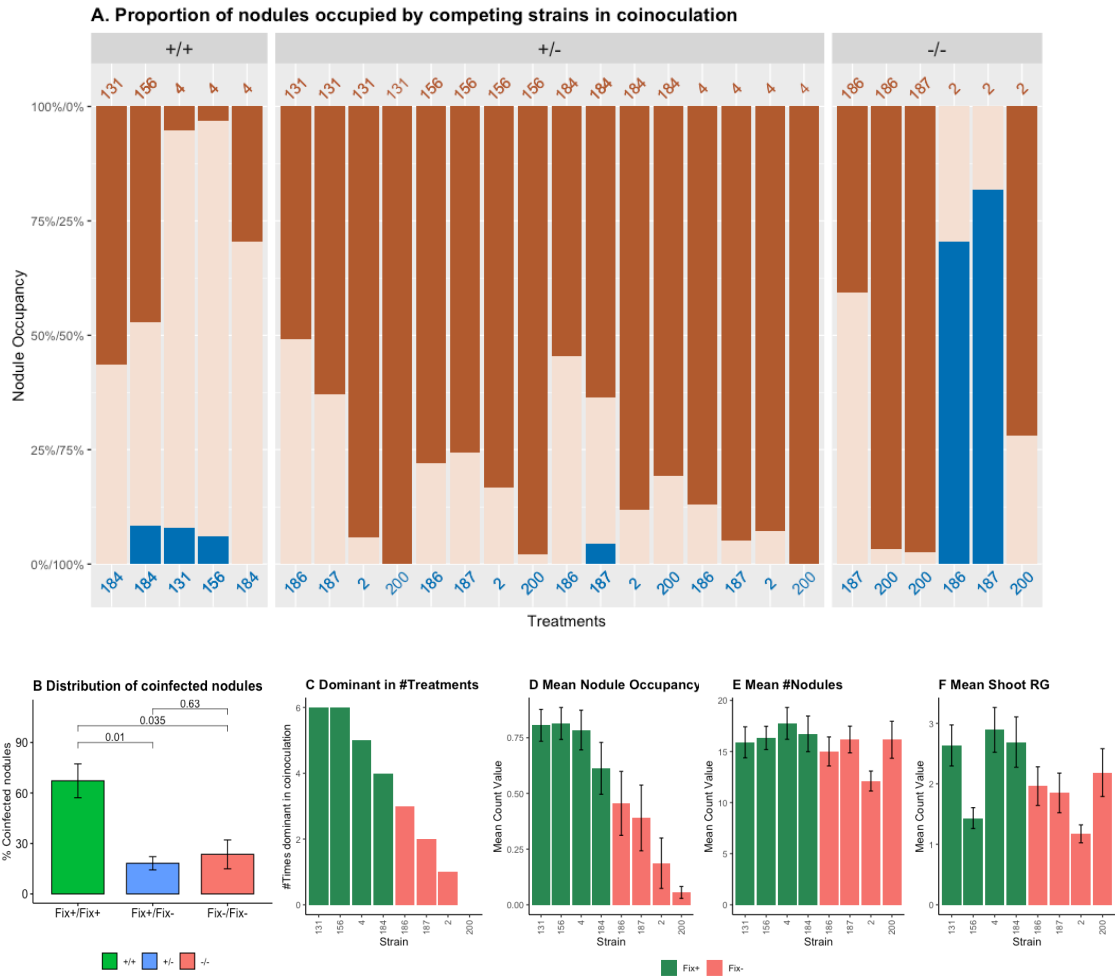
Fig. 1.3 Nodule genotyping results. (A) Nodule occupancy is illustrated with bars indicating the proportion of each competitor strain within nodules that were coinoculated. Brown color indicates occupancy by the strain labeled at the top, blue indicates occupancy of the other strain labeled at the bottom, and tan indicates nodules infected by both strains. The coinoculated treatments are divided in +/+, +/-, and -/- groups based on the nitrogen-fixing capacity of each strain. **(B)** Percent coinfected nodules in +/+, +/-, and -/- treatment groups. **(C)** Strain dominance is quantified as the number of treatments where a strain had higher non-random nodule occupancy compared to the competing strain. **(D-F)** Mean count value for number of nodules, relative growth, and nodule occupancy are indicated for each strain among all coinoculated treatments, respectively.
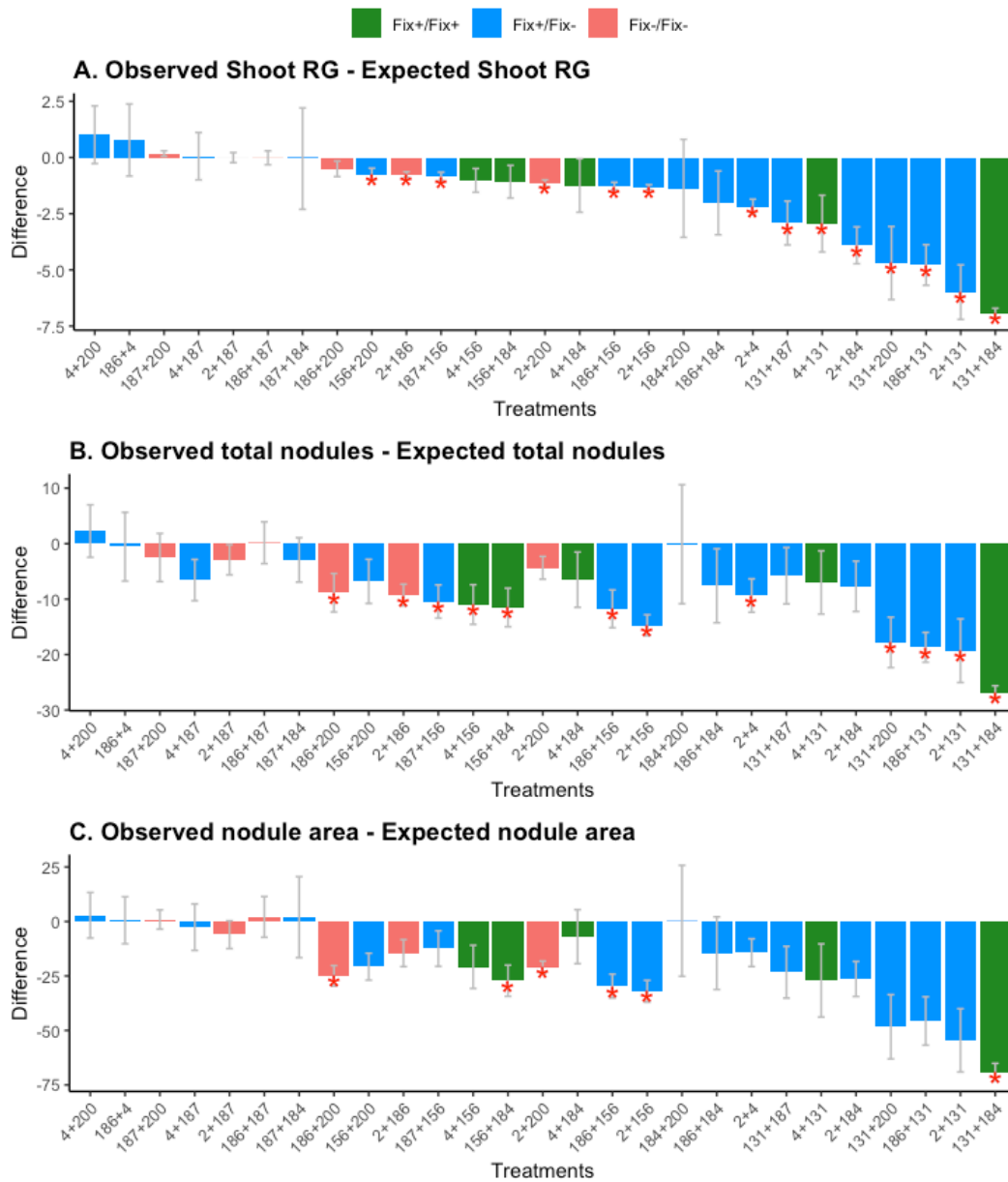
Fig. 1.4. Observed data in coinoculated treatments relative to expectations based on the clonal inoculations. Relative growth **(A)**, total nodules **(B)**, and nodule area **(C)** are each displayed with the relevant expected values subtracted from them. Negative values indicate that the observed values were less than the expected ones. Asterisks indicates significant difference between observed and expected values (i.e., $P < 0.05$).

Table 1. *Bradyrhizobium* strains and their key properties.

Strain numbers	Strain Code	Nod	Fix	T3SS	Taxon	Sampling Site
2	05LoS24R3_28	+	-	-	Novel XIII	Bodega Marin Reserve
186	11LoS6_2	+	-	+	NA	San Dimas Reservoir
4	05LoS21R5_36	+	+	+	*B. canariense*	Bodega Marin Reserve
131	13LoS28_1	+	+	+	*B. canariense*	UC Riverside
187	11LoS7_1	+	-	-	*B. canariense*	San Dimas Reservoir
156	11LoS34_4	+	-	+	*B. canariense*	Burns Piñon Ridge Reserve
184	11LoS34_10	+	+	+	NA	Burns Piñon Ridge Reserve
200	13LoS78_1	+	-	+	Novel IV	Pismo Dunes Natural Preserve

Table 1.2. Linear model testing effect of treatment and DPI on host benefit and nodulation during coinoculation. ANCOVA statistics for testing whether response variables significantly vary among single inoculation treatments using DPI as covariate since plants were harvested continuously from 28 DPI. Degrees of freedom (df), F-statistics, and associated P-values are reported for each model.

	Log_{10}(Shoot RG+0.5)[1]			$\sqrt{Total\ Nodules}$[1]			Log(Mean Nodule Biomass + 0.1) (mg)[1]			Log_{10}(Investment + 0.2)[1]		
Sample Size[2] = 231	Adj. R^2 = 0.54			Adj. R^2 = 0.40			Adj. R^2 = 0.32			Adj. R^2 = 0.29		
Shapiro-Wilk test[3]	W = 0.99 P = 0.53			W = 0.988 P = 0.063			W=0.98 P = 0.001			W=0.97 P = 0.00128		
ANCOVA	df	F	P	df	F	P	df	F	P	df	F	P
Intercept	1	48.502	$4.647×10^{-11}$	1	1.5763	0.2107	1	28.5467	$2.490×10^{-07}$	1	1.6847	0.195823
Elapsed DPI	1	164.206	$< 2.2×10^{-16}$	1	110.29	$< 2.2×10^{-16}$	1	62.34	$1.906×10^{-13}$	1	10.51	0.001389
Treatment	27	5.455	$3.756×10^{-13}$	27	2.63	$6.513×10^{-05}$	27	2.7821	$2.466×10^{-05}$	27	3.41	$3.502×10^{-07}$
Residuals	200			202			199			197		

1. Response variable transformations used in linear models to fit coinoculation data for different response variables (Shoot RG, Total nodules, Mean nodule biomass, Investment).
2. Sample size indicates number of datapoints (N) used to fit the models. Adjusted R-squared for each model are also reported.
3. Shapiro-Wilk statistics testing normality of the residuals in the linear models. Test statistics W and P-value are reported

Supplementary Figures and Tables
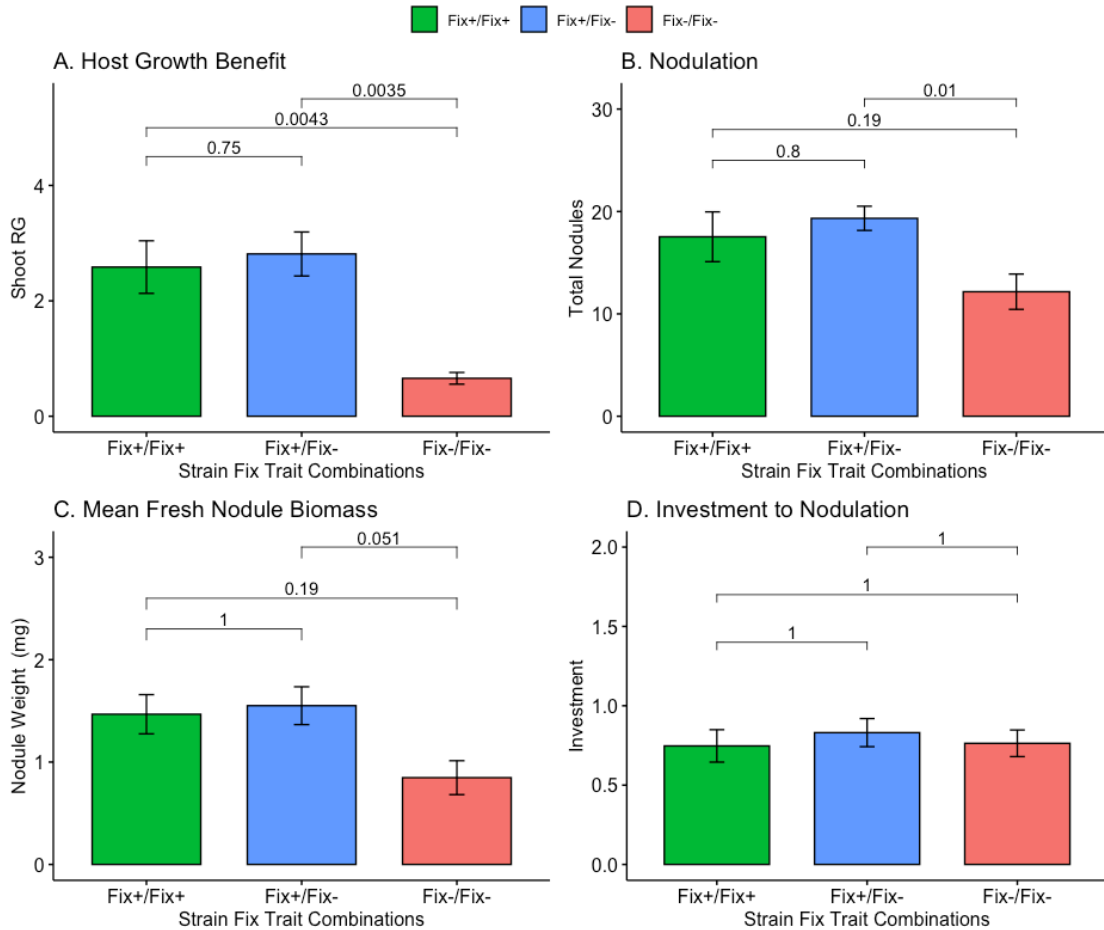
Supplementary Fig. 1.1: Coinoculation experiment results with strain combinations grouped by effectiveness in clonal inoculation results. The coinoculation treatment combinations are divided into Fix+/Fix+ (green), Fix+/Fix- (blue), and Fix-/Fix- (red) based on the participating strains' relative growth effects in the single inoculation experiments (**Fig. 1**). Pairwise t-test are carried out between all three combinations and *P*-values are shown on top of the box-plots, indicating statistical differences between groups. The panels indicate **(A)** shoot relative growth, **(B)** nodule counts, **(C)** mean nodule biomass, and **(D)** host investment for all coinoculation treatment combination categories.

$y = 1.91 + 2.59 \, x \; R^2 = 0.86$

$y = 0.386 + 0.246 \, x \; R^2 = 0.67$
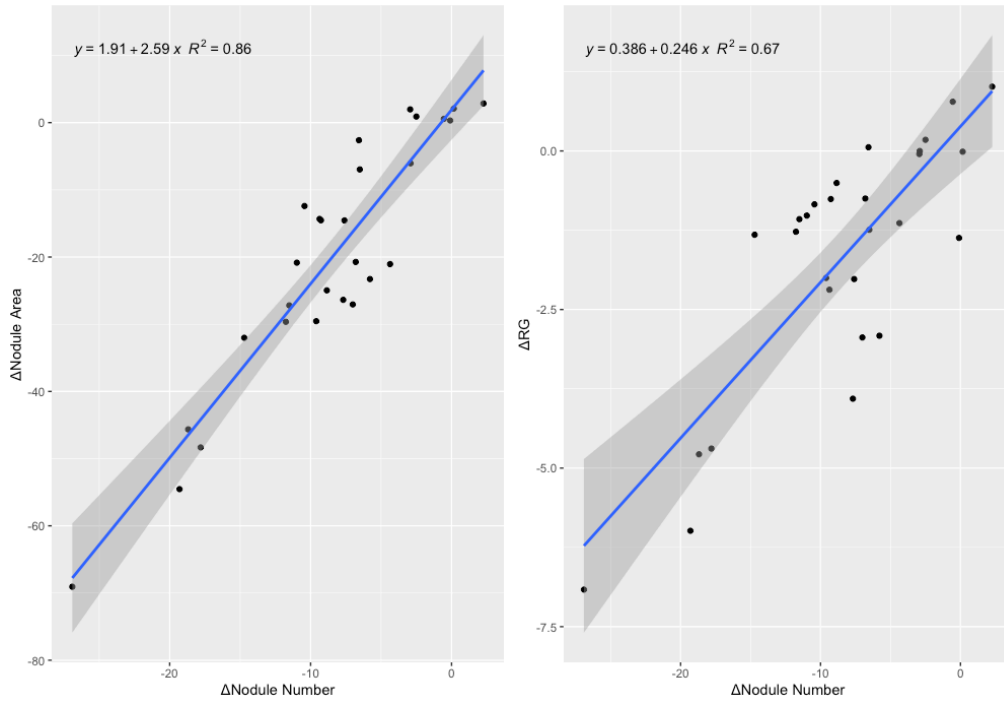
Supplementary Fig. 1.2 Correlation of residuals for expected values of nodule area and nodule number (left), and relative growth with nodule number (right).
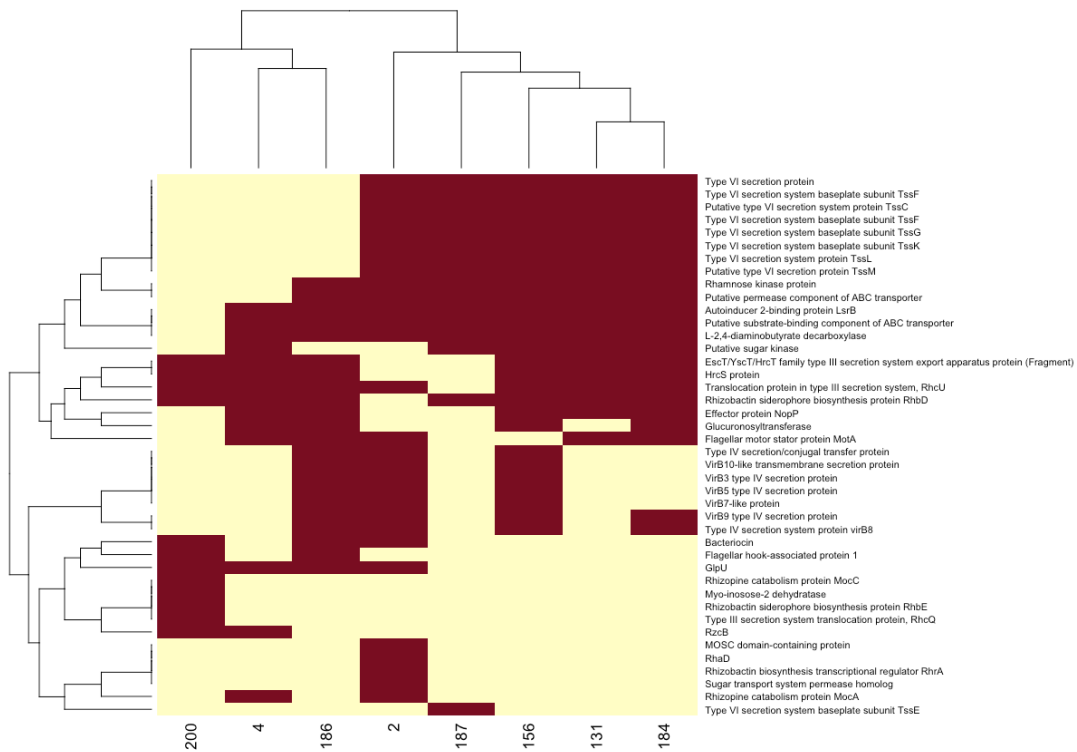
Supplementary Fig. 1.3: A heatmap shows hierarchical clustering of competition gene presence-absence among *Bradyrhizobium* strain genomes based on data from Mendoza-Suarez et al. 2021. Brown squares indicate presence of genes (by rows) in the genome of these isolates (by columns). Strain 4, 156, 131, 184 are categorized as nitrogen fixation effective (Fix+) and the rest of the strains are ineffective (Fix-). Nodule occupancy based competitive hierarchy is 131 & 156 > 4 > 184 > 186 > 187 > 2 > 200.
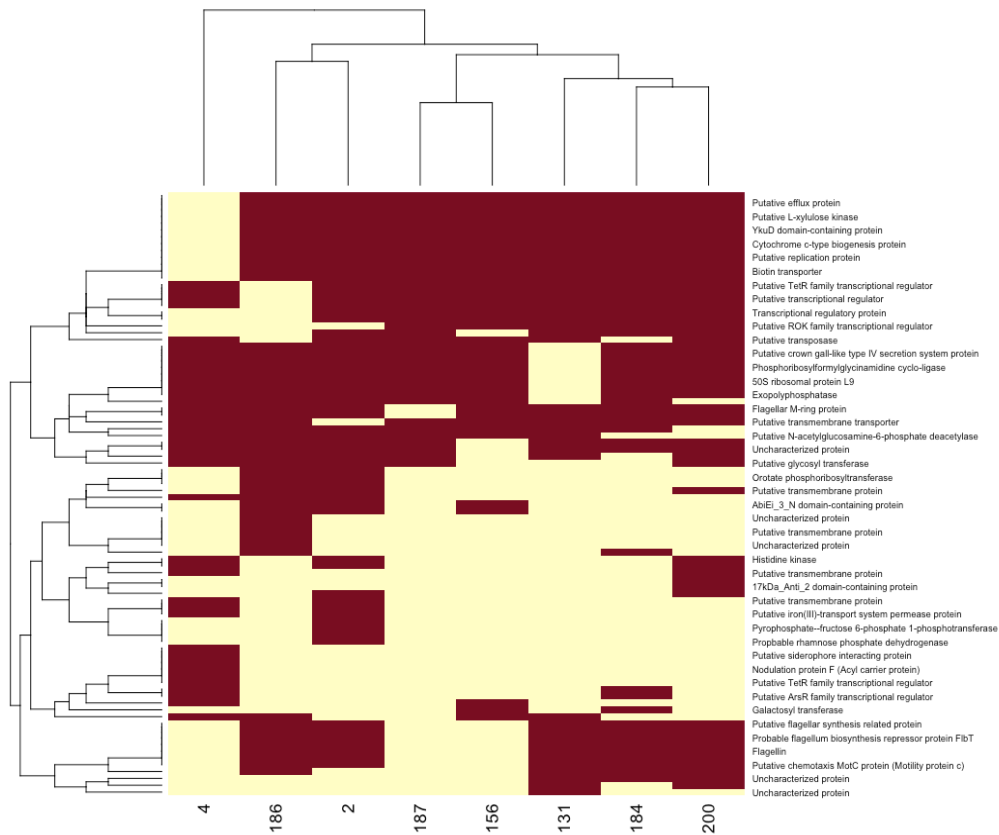
Supplementary Fig. 1.4: A heatmap shows hierarchical clustering of competition gene presence-absence among *Bradyrhizobium* strain genomes based on data from Wheatley et al. 2020. Brown squares indicate presence of genes (by rows) in the genome of these isolates (by columns). Strain 4, 156, 131, 184 are categorized as nitrogen fixation effective (Fix+) and the rest of the strains are ineffective (Fix-). Nodule occupancy based competitive hierarchy is 131 & 156 > 4 > 184 > 186 > 187 > 2 > 200.
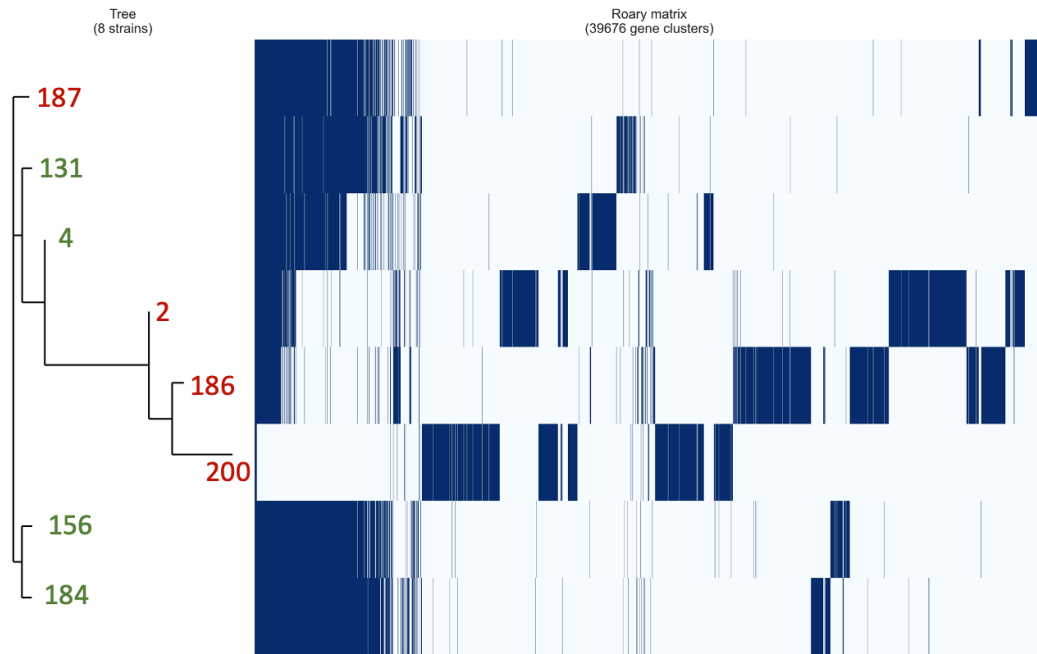
Supplementary Fig. 1.5: Gene presence-absence data are compared among the eight focal strains. (Left) Core-gene alignment tree from Roary output. The tip labels indicate isolates, green indicates nitrogen-fixation effective isolates and red indicates ineffective isolates. (Right) The heatmap Roary matrix summarizes gene-presence absence matrix of gene-clusters across the eight genomes.

Supplementary Table 1.1: Quanitative culture of inocula. CFU count carried out by spread plate technique.

Inocula Culture	CFU		Log(CFU)		
	1st Batch	2nd Batch	1st Batch	2nd Batch	Mean
2	250000000	300000000	8.39794001	8.47712125	8.43753063
186	100000000	150000000	8	8.17609126	8.08804563
4	250000000	400000000	8.39794001	8.60205999	8.5
131	NA	350000000	NA	8.54406804	8.54406804
187	150000000	250000000	8.17609126	8.39794001	8.28701563
156	200000000	100000000	8.30103	8	8.150515
184	200000000	300000000	8.30103	8.47712125	8.38907563
200	50000000	300000000	7.69897	8.47712125	8.08804563

Supplementary Table 1.2. Testing effectiveness of single-inoculation treatments on shoot relative growth and classifying nitrogen-fixing trait.

Treatment	Mean	t-value	df	*P*-value	Nitrogen-fixation Classification
131	8.19	3.81	8	0.00258	Effective
184	6.64	2.87	9	0.00909	Effective
4	3.54	1.84	5	0.06262	Marginally Effective
156	1.94	3.3	8	0.00539	Effective
186	1.4	1.17	9	0.1351	Ineffective
2	1.88	1.05	9	0.1599	Ineffective
200	0.47	-6.59	3	0.9964	Ineffective
187	0.28	-15.14	6	1	Ineffective

 Supplementary Table 1.3. Linear models testing the effects of treatment and days post inoculation on host benefit and nodulation during single inoculation. ANCOVA statistics test whether response variables significantly vary among single inoculation treatments using DPI as a covariate. Degrees of freedom (df), F-statistics, and associated P-values are reported for each model.

	Log_{10}(Shoot RG + 0.5)[1]			$\sqrt{Total\ Nodules}$			Log(Mean Nodule Biomass + 0.1)[1]			Log_{10}(Investment + 0.2)		
Sample Size[2] = 67	Adj. R^2 = 0.53			Adj. R^2 = 0.406			Adj. R^2 =0.66			Adj. R^2 = 0.448		
Shapiro-Wilk test[3]	W = 0.97 P = 0.165			W = 0.98 P = 0.487			W = 0.98 P = 0.553			W = 0.975 P = 0.22		
ANCOVA	df	F	*P*	df	F	*P*	df	F	*P*	df	F	*P*
Intercept	1	2.103	0.15259	1	1.474	0.2295	1	**3.1884**	0.0795	1	0.022	0.8827
Treatment	7	*11.24*	**8.576×10⁻⁰⁹**	7	*7.009*	**4.713×10⁻⁰⁶**	7	*18.092*	**2.542×10⁻¹²**	7	*8.358*	***6.975×10⁻⁰⁷***
DPI	1	*6.641*	**0.01262**	1	*10.413*	0.002	1	*6.16*	0.016	1	0.0164	0.8985
Residuals	56			58			56			54		

Supplementary Table 1.4: Testing two linear models (with or without inoculation batch as random variable) for each response variable.

Experiment	Response Variable	Model	#Parameter	AIC	χ2	df	Pr(>Chisq)
Single inoculation	RG	Log(RG) ~ Treatments	10	203.79	0	1	1
		Log(RG) ~ Treatments + (1|batch)	11	205.79			
	Total Nodules	Square Root(Total Nodules) ~ Treatments	10	305.75	0	1	1
		Square Root(Total Nodules) ~ Treatments + (1 | batch)	11	305.75			
	Nodule Biomass	Log(Mean Nodule Biomass + 0.1) ~ Treatments	10	156.74	0	1	1
		Log(Mean Nodule Biomass + 0.1) ~ Treatments + (1 | batch)	11	158.74			
	Investment	Log10(Investment + 0.2) ~ Treatments	10	-71.036	0	1	1
		Log10(Investment + 0.2) ~ Treatments + (1 | batch)	11	-69.036			
Coinoculation	RG	Log(RG) ~ Treatments	30	345.49	0	1	1
		Log(RG) ~ Treatments + (1|batch)	31	347.49			
	Total Nodules	Square Root(Total Nodules) ~ Treatments	30	926.02	0	1	1
		Square Root(Total Nodules) ~ Treatments + (1 | batch)	31	928.02			
	Nodule Biomass	Log(Mean Nodule Biomass + 0.1) ~ Treatments	30	726.56	0	1	1
		Log(Mean Nodule Biomass + 0.1) ~ Treatments + (1 | batch)	31	728.56			
	Investment	Log10(Investment + 0.2) ~ Treatments	30	-58.22	0	1	1
		Log10(Investment + 0.2) ~ Treatments + (1 | batch)	31	-56.22			

Supplementary Table 1.5. Estimated marginal means of response variables in single inoculation.

Treatment	Shoot RG	Total Nodules	Mean Nodule Biomass (mg)	Investment
131	7.31	35.06	2.016	0.651
184	4.543	28.4	1.271	0.526
4	2.488	20.89	0.868	0.466
156	1.753	27.46	1.159	1.111
186	1.122	16.49	1.07	0.881
2	0.981	8.07	0.842	0.408
200	0.42	6.84	0.212	0.177
187	0.29	16.35	0.166	0.476

Supplementary Table 1.6. Testing effectiveness of coinoculation treatments on shoot relative growth and classifying nitrogen-fixing trait for the coinoculated community.

Treatment	Mean	t	df	*P*-value	Nitrogen-fixation Classification
131+156	3.06	2.26	9	0.024	Effective
131+184	1.17	0.79	6	0.228	Ineffective
131+187	3.95	3.02	8	0.0082	Effective
131+200	3.49	1.47	9	0.0872	Marginally Effective
156+184	2.1	1.39	8	0.1001	Ineffective
156+200	1.186	0.688	9	0.2541	Ineffective
184+200	5.271	1.976	5	0.0525	Marginally Effective
186+131	2.8	2.148	8	0.0319	Effective
186+156	0.6295	-2.17	6	0.9635	Ineffective
186+184	3.48	1.716	8	0.0622	Marginally Effective
186+187	1.016	0.057	7	0.477	Ineffective
186+200	0.8961	-0.346	7	0.6303	Ineffective
186+4	4.2532	2.05	8	0.036	Effective
187+156	0.9569	-0.2401	8	0.5919	Ineffective
187+184	4.3683	1.503	7	0.0881	Marginally Effective
187+200	0.4626	-4.7682	8	0.999	Ineffective
2+131	2.2	1.027	4	0.1811	Ineffective
2+156	0.614	-3.528	5	0.9916	Ineffective
2+184	2.3607	1.5749	8	0.0769	Marginally Effective
2+186	0.6547	3.0589	7	0.9908	Ineffective
2+187	0.405	-2.4692	3	0.9549	Ineffective
2+200	0.5	-4.133	8	0.9984	Ineffective
2+4	1.315	0.9677	8	0.1808	Ineffective
4+131	3.894	2.192	7	0.0322	Effective
4+156	1.6225	1.244	6	0.1298	Ineffective
4+184	3.6506	2.1351	8	0.0326	Effective
4+187	3.537	2.3137	7	0.0269	Effective
4+200	4.456	2.529	9	0.0161	Effective

Supplementary Table 1.7: Estimated marginal means of response variables in coinoculated treatments after post-hoc Tukey HSD test. Nitrogen-fixation classification is based on significant difference of Shoot RG compared to uninoculated controls by one-sample t-test. Effective treatments are significantly different in Shoot RG compared to uninoculated controls whereas marginally effective treatments have borderline significance score (0.05 < P < 0.10).

Treatment	Shoot RG	Total nodules	Fresh mean nodule biomass	Investment	Nitrogen-fixation Classification
131+156	2.958408	24.76964	1.123503	0.629893	**Effective**
131+184	1.27569	8.146198	1.236413	0.527993	Ineffective
131+187	3.325353	24.83377	1.761651	0.785829	**Effective**
131+200	2.15225	17.54353	0.914434	0.518553	Marginally Effective
156+184	2.146792	20.64398	1.16225	0.746078	Ineffective
156+200	1.317434	22.94821	1.374562	1.401899	Ineffective
184+200	3.036117	22.99955	1.354337	0.502832	Marginally Effective
186+131	2.569641	15.8536	1.849267	0.679915	**Effective**
186+156	0.758019	17.21742	0.939985	1.389096	Ineffective
186+184	2.630851	18.8464	1.210992	0.513937	Marginally Effective
186+187	0.727392	16.15295	0.757798	0.94467	Ineffective
186+200	0.907689	10.76231	1.26079	0.807239	Ineffective

186+4	2.312971	18.96262	1.038796	0.519393	**Effective**
187+156	0.856231	15.70671	1.399999	1.454789	Ineffective
187+184	1.613071	18.19762	0.835473	0.55849	Marginally Effective
187+200	0.691776	15.14395	0.385114	0.511422	Ineffective
2+131	1.413616	14.57202	1.668117	1.131789	Ineffective
2+156	0.760564	13.02971	0.63102	0.717266	Ineffective
2+184	1.953993	20.90347	0.849678	0.570459	Marginally Effective
2+186	0.639928	9.769809	1.084486	0.889069	Ineffective
2+187	0.258848	11.82332	0.25012	0.572291	Ineffective
2+200	0.730364	6.998708	0.86501	0.404535	Ineffective
2+4	1.328425	13.70841	1.032358	0.670541	Ineffective
4+131	3.245427	25.51858	1.006466	0.507559	**Effective**
4+156	1.770244	17.22492	1.929563	1.144596	Ineffective
4+184	3.346087	20.89053	1.53936	0.628801	**Effective**
4+187	3.006206	16.75781	1.904734	0.656323	**Effective**
4+200	3.901178	26.95479	1.24448	0.554162	**Effective**

Supplementary Table 1.8: Goodness of fit test for nodule occupancy. Strain A and strain B represents participants in pairwise coinoculated treatments. Dominance is determined by higher non-random nodule occupancy of one strain over another in a coinoculated treatment.

Strain A	Strain B	Test of random infection					Test of nodulation by inoculum ratio				
		#Nodules infected by A	#Nodules infected by A+B	#Nodules infected by B	Chi² p-value	Dominance	#Nodules infected by A (including 50% of co-infection)	#Nodules infected by B (including 50% of co-infection)	Inocula A quantitative culture log(CFU)	Inocula B quantitative culture log(CFU)	Chi² p-value
4	131	2	33	3	2.29E-11	131	18.5	19.5	8.544	8.5	8.59E-01
4	186	20	3	0	2.57E-07	4	21.5	1.5	8.088	8.5	1.78E-05
4	156	2	60	4	3.99E-22	156	32	34	8.151	8.5	9.39E-01
4	2	38	3	0	6.56E-15	4	39.5	1.5	8.438	8.5	2.55E-09
4	184	13	31	0	6.67E-08	4	28.5	15.5	8.389	8.5	4.51E-02
4	187	37	2	0	3.42E-15	4	38	1	8.287	8.5	1.92E-09
4	200	40	0	0	4.25E-18	4	40	0	8.088	8.5	8.95E-11
2	156	0	6	30	7.58E-10	156	3	33	8.151	8.438	9.73E-07
2	186	0	8	19	4.06E-05	186	4	23	8.088	8.438	3.89E-04
2	131	0	1	16	6.97E-07	131	0.5	16.5	8.544	8.438	9.37E-05
2	187	0	4	18	5.12E-06	187	2	20	8.287	8.438	1.47E-04

2	200	23	9	0	3.39E-06	2	27.5	4.5	8.088	8.438	2.83E-05
2	184	0	5	37	3.15E-13	184	2.5	39.5	8.389	8.438	1.27E-08
187	200	36	1	0	1.58E-15	187	36.5	0.5	8.088	8.287	2.06E-09
186	187	11	16	0	5.85E-04	186	19	8	8.287	8.088	4.00E-02
186	200	30	1	0	6.28E-13	186	30.5	0.5	8.088	8.088	7.12E-08
184	200	25	6	0	6.94E-08	184	28	3	8.088	8.389	4.38E-06
184	187	14	7	1	3.11E-03	184	17.5	4.5	8.287	8.389	5.10E-03
184	186	24	20	0	1.27E-05	184	34	10	8.088	8.389	1.84E-04
156	184	17	16	3	6.20E-03	156	25	11	8.389	8.151	2.46E-02
156	186	32	9	0	2.22E-09	156	36.5	4.5	8.088	8.151	5.10E-07
156	200	45	1	0	1.98E-19	156	45.5	0.5	8.088	8.151	2.72E-11
156	187	28	9	0	6.38E-08	156	32.5	4.5	8.287	8.151	5.29E-06
131	184	22	17	0	3.60E-05	131	30.5	8.5	8.389	8.544	3.43E-04
131	187	22	13	0	2.79E-05	131	28.5	6.5	8.287	8.544	1.39E-04
131	200	29	0	0	2.54E-13	131	29	0	8.088	8.544	3.11E-08
131	186	27	26	0	1.74E-06	131	40	13	8.088	8.544	9.24E-05

Supplementary Table 1.9: MiSeq genotyping cross-validation using Sanger sequencing. Category indicates Sange and MiSeq comparison category for different match levels.

Category	Count	%Match of original Genotype
NA in both Sanger and MiSeq	1	-
NA in Sanger	29	-
NA in MiSeq	9	-
Full Match	24	47.0588235
Partial Match	21	41.1764706
No Match	6	11.7647059
Total	90	

Supplementary Table 1.10: Difference in expected vs. observed shoot relative growth, number of nodules, nodule area and mean nodule biomass. Expected value was estimated by calculating mean of each participant isolate's relative growth from single inoculation experiment weighted by nodulation occupancy in the coinoculation. The p-value represent alternative hypothesis of observed value either higher or lower than the expected value.

Treatment	Relatiive Growth		#Nodules		Nodule Area		Nodule Biomass (mg)	
	Difference	*P*-value	Difference	*P*-value	Difference	*P*-value	Difference	*P*-value
131+184	-6.91563	2.55E-08	-26.9062	7.85E-07	-69.0656	7.91E-06	0.242451	2.98E-07
131+187	-2.91313	9.86E-03	-5.78968	2.88E-01	-23.2806	9.07E-01	-0.11687	5.87E-12
131+200	-4.69218	8.36E-03	-17.8	3.53E-03	-48.33	2.60E-01	-0.57082	2.27E-11
156+184	-1.07509	8.74E-02	-11.5118	1.07E-02	-27.1981	3.83E-03	-0.08517	1.70E-11
156+200	-0.75019	2.27E-02	-6.8	1.21E-01	-20.7333	2.03E-01	0.343713	5.75E-15
184+200	-1.37083	2.93E-01	-0.1	9.93E-01	0.32	5.64E-01	0.32211	7.93E-08
186+131	-4.78227	4.61E-04	-18.6851	1.17E-04	-45.6444	6.33E-02	0.850692	2.92E-11
186+156	-1.27329	1.64E-04	-11.7563	1.35E-02	-29.6443	6.16E-03	-0.42633	1.95E-13
186+184	-2.01947	9.41E-02	-7.59722	2.88E-01	-14.5522	6.37E-01	-0.17293	4.76E-15
186+187	-0.01158	6.52E-01	0.159524	9.67E-01	2.098333	5.65E-01	0.496253	2.73E-13
186+200	-0.50493	1.81E-01	-8.85	3.80E-02	-24.9567	7.08E-03	0.410677	1.38E-11
186+4	0.776303	7.10E-01	-0.55	9.31E-01	0.561076	8.83E-01	0.265606	1.57E-11
187+156	-0.84159	2.04E-03	-10.4337	8.37E-03	-12.4065	7.18E-01	0.252332	7.37E-14
187+184	-0.04844	5.07E-01	-2.93722	4.85E-01	1.972286	6.62E-01	0.456381	2.61E-09
187+200	0.175833	9.88E-01	-2.50111	5.80E-01	0.904111	2.66E-01	0.231655	7.92E-15
2+131	-5.98829	4.25E-03	-19.3	2.83E-02	-54.54	1.64E-01	0.969178	4.63E-05
2+156	-1.32114	4.98E-05	-14.7089	2.34E-04	-32.0153	8.45E-03	-0.60299	2.92E-12

2+184	-3.90591	7.52E-04	-7.68684	1.29E-01	-26.37	1.20E-01	-0.26593	1.01E-12
2+186	-0.75826	6.25E-04	-9.265	2.08E-03	-14.5295	1.20E-01	0.014325	5.80E-12
2+187	-0.00085	6.29E-01	-2.90824	3.63E-01	-6.05385	3.09E-01	0.087211	4.92E-08
2+200	-1.13734	6.03E-05	-4.36111	6.64E-02	-21.05	1.77E-04	0.681169	1.93E-13
2+4	-2.18716	1.06E-04	-9.36333	1.43E-02	-14.3342	1.58E-01	-0.09953	2.49E-14
4+131	-2.9404	2.50E-02	-7.00972	2.58E-01	-27.0567	3.70E-01	-0.9755	6.30E-11
4+156	-1.01886	6.52E-02	-10.9788	2.18E-02	-20.8394	9.91E-02	0.905389	3.12E-08
4+184	-1.24264	1.48E-01	-6.51129	2.27E-01	-6.96719	5.24E-01	0.340034	1.98E-12
4+187	0.058572	5.23E-01	-6.5731	1.21E-01	-2.6145	4.39E-01	3.230087	1.04E-04
4+200	1.013501	7.35E-01	2.266667	6.43E-01	2.85	4.50E-01	0.32323	1.61E-14

Supplementary Table 1.11: Clonal and pairwise growth statistics of four Fix+ isolates from this study in the liquid RDM media. Parametrs of growth curve was estimated from five replicates.

Experiment	Treatment	Carrying Capacity (log10 CFU)		Doubling Time (hour)	
		Mean	SE	Mean	SE
Clonal	131	8.995415	0.009874667	11.26364	0.5366348
Clonal	156	8.958522	0.021580236	8.821987	1.0925229
Clonal	184	8.981008	0.009238915	10.86454	0.2379909
Clonal	4	9.037723	0.010204495	14.46802	0.8272705
Competition	131+156	8.967835	0.016351296	12.27995	0.9207333
Competition	131+184	8.988119	0.009456798	11.90403	0.503558
Competition	156+184	8.943167	0.011775629	11.3226	0.3021004
Competition	4+131	8.99331	0.007208692	12.70796	0.4289135
Competition	4+156	8.988606	0.006520765	12.86764	0.6633758
Competition	4+156	8.980869	0.008916358	10.46651	1.16585

Supplementary Table 1.12: ANOVA comparison of carrying capacity, growth rate, and doubling time in both clonal and competition experiemnts on liquid RDM media.

Experiment	Response Variable	Model	Independent Variable	Df	Sum Sq	Mean Sq	F	Pr(>F)
Clonal	Carrying Capacity (K)	log10(K) ~ Strain	Strain	3	0.0146	0.004867	5.134	**0.0122**
			Residuals	15	0.01422	0.000948		
Clonal	Doubling Time (DT)	DT ~ Strain	Strain	3	71.58	23.859	9.182	**0.00108**
			Residuals	15	38.98	2.598		
Competition	Carrying Capacity (K)	log10(K) ~ Strain	Strain	5	0.008564	0.0017128	3.037	**0.0301**
			Residuals	23	0.012973	0.0005641		
Competition	Doubling Time (DT)	DT ~ Strain	Strain	5	19.96	3.991	1.48	0.235
			Residuals	23	62.04	2.698		

Supplementary Table 1.13: Tukey HSD posthoc test based on significant ANOVA results.

Experiment	Response Variable	Comparison	Difference	Lower	Upper	Adj p
Clonal	Carrying Capacity (K)	156 vs. 131	-0.03689	-0.09302	0.019233	0.271406
		184 vs. 131	-0.01441	-0.07053	0.041719	0.879504
		4 vs. 131	0.042307	-0.01722	0.101838	0.214593
		184 vs. 156	0.022486	-0.03364	0.078612	0.662944
		4 vs. 156	**0.079201**	**0.01967**	**0.138732**	**0.007882**
		4 vs. 184	0.056715	-0.00282	0.116246	0.064444
Clonal	Doubling Time (DT)	156 vs. 131	-2.44165	-5.38001	0.496708	0.120864
		184 vs. 131	-0.3991	-3.33746	2.539262	0.978886
		4 vs. 131	**3.204378**	**0.087778**	**6.320978**	**0.042904**
		184 vs. 156	2.042554	-0.8958	4.980912	0.230115
		4 vs. 156	**5.646028**	**2.529428**	**8.762628**	**0.000536**
		4 vs. 184	**3.603474**	**0.486874**	**6.720074**	**0.021103**
Competition	Carrying Capacity (K)	131+184 vs. 131+156	0.020285	-0.02632	0.066894	0.754627
		156+184 vs. 131+156	-0.02467	-0.07128	0.021942	0.580734
		4+131 vs. 131+156	0.025475	-0.02396	0.074912	0.607283
		4+156 vs. 131+156	0.020771	-0.02584	0.06738	0.73644
		4+184 vs. 131+156	0.013034	-0.03358	0.059643	0.95046
		156+184 vs. 131+184	-0.04495	-0.09156	0.001657	0.063076
		4+131 vs. 131+184	0.005191	-0.04425	0.054627	0.999441
		4+156 vs. 131+184	0.000486	-0.04612	0.047096	1
		4+184 vs. 131+184	-0.00725	-0.05386	0.039359	0.996333

		4+131 vs. 156+184	**0.050143**	**0.000706**	**0.09958**	**0.04548**
		4+156 vs. 156+184	0.045439	-0.00117	0.092048	0.058949
		4+184 vs. 156+184	0.037702	-0.00891	0.084311	0.162342
		4+156 vs. 4+131	-0.0047	-0.05414	0.044732	0.999654
		4+184 vs. 4+131	-0.01244	-0.06188	0.036995	0.968059
		4+184 vs. 4+156	-0.00774	-0.05435	0.038872	0.995033

Supplementary Table 1.14: Based on carrying capacity, growth rate, and doubling time from clonal treatments in liquid RDM growth, expected value was predicted by pairwise mean and compared with competition treatments.

Variable	Comparison	Competition Mean	Clonal Prediction	t	df	p
Carrying Capacity	4+131	8.99331	9.014218	-2.9004	3	0.06248
	4+156	8.988606	8.993722	-0.78463	4	0.4765
	4+184	**8.980869**	**9.006214**	**-2.8426**	**4**	**0.04675**
	131+156	8.967835	8.976968	-0.5586	4	0.6062
	131+184	8.988119	8.988211	-0.0097304	4	0.9927
	156+184	8.943167	8.969765	-2.2587	4	0.08681
Doubling Time	4+131	12.70796	12.6878	0.046982	3	0.9655
	4+156	12.86764	11.33133	2.3159	4	0.0815
	4+184	10.46651	12.46608	-1.7151	4	0.1615
	131+156	12.27995	10.04281	2.4297	4	0.072
	131+184	11.90403	11.06409	1.668	4	0.1706
	156+184	11.3226	9.843263	4.8968	4	**0.008062**

Supplementary Table 1.15: Total population from five solid RDM plate replicates spreaded by 2x10^8 cells in clonal and pairwise competition.

Experiment	Treatment	Total Population
Clonal	4	2.25E+08
Clonal	131	2.83E+09
Clonal	156	1.77E+09
Clonal	184	3.22E+09
Competition	156+184	2.88E+09
Competition	4+156	2.95E+09
Competition	4+184	3.42E+09
Competition	131+184	3.05E+09
Competition	4+131	3.36E+09
Competition	131+156	2.63E+09

Chapter 2

Epidemic *Bradyrhizobium* haplotypes adapt to host metapopulations via acquisition of diverse, host-specific symbiosis ICEs

Abstract

Legume hosts acquire compatible rhizobia symbionts from soil communities where the environment and multiple hosts are potential selective forces. How these factors shape rhizobial genomes is poorly understood, especially in communities where multiple host and rhizobia taxa interact and vary in symbiotic compatibility. Here, a metapopulation of *Bradyrhizobium spp.* were investigated in nine native *Acmispon* host species, consisting of 507 isolates from 17 sampling-sites across an ~800 km transect in California. Rhizobia haplotypes were defined from four core chromosomal loci and four symbiosis loci encoded on integrative conjugative elements (symICEs) that can be horizontally transmitted. Six chromosomal haplotypes dominated the metapopulation, three of which had epidemic characteristics, being detected in multiple distant sampling sites. Dominant chromosomal haplotypes were massively expanded clonal lineages – with near identity based on nucleotide-similarity and protein-composition – that have acquired diverse symICE haplotypes, allowing for independent adaptation to multiple host taxa. Core and symICE regions of *Bradyrhizobium* genomes can evolve independently, with the chromosome shaped by selective sweeps and clonal expansion, that acquire symICEs that bear features of host and local adaptation. The multipartite genome architecture allows *Bradyrhizobium*

to respond to selective forces exerted by hosts and environmental factors that are experienced during tramsmission phases.

Introduction

Most bacterial mutualists have facultative associations with hosts, with a bipartite lifestyle that includes symbiotic interaction with the host partner, punctuated with transmission phases in the environment (Bright & Bulgheresi, 2010; J. L. Sachs et al., 2011). During symbiotic interaction, hosts can impose intense selection on bacterial partners, favoring strains that are compatible with the host and that provide fitness benefits (Burghardt et al., 2018; Weis et al., 2001; S. Yang et al., 2010). Host selection is thought to be the primary factor that maintains beneficial bacterial services, and that minimizes impacts of ineffective or exploitative strains (Kevin R. Foster et al., 2017; Joel L. Sachs et al., 2004). Conversely, selective forces shaping bacteria during environmental growth can be independent from adaptation to the host and need not be aligned to host fitness (Denison & Kiers, 2004; J. L. Sachs et al., 2009a, 2011). Bacterial mutualists can spend extensive periods of time in the environment (Bright & Bulgheresi, 2010; J. L. Sachs et al., 2011), and thus it is critical to understand the relative roles of selection during association with the host, versus selection during free-living phases. Characterizing these independent forces is vital to understanding the stability of beneficial host-microbial associations.

The legume-rhizobia symbiosis is critical to ecosystems as it generates the predominant natural source of nitrogen in terrestrial habitats (Cleveland et al., 1999). Rhizobia encompass polyphyletic lineages of nitrogen fixing proteobacteria that can enhance legume growth through nodulation within root (or rarely shoot) tissues (Sawada et al., 2003). Host legumes acquire rhizobia from the soil environment, where rhizobia can spend extensive periods living independent of host association (Denison & Kiers, 2004;

Amanda C. Hollowell et al., 2016b; J. L. Sachs et al., 2009a; VanInsberghe et al., 2015). To initiate the symbiosis, the legume host releases species-specific flavonoids from its roots that attract rhizobia and influence which strains nodulate the host (C.-W. Liu & Murray, 2016). In response, rhizobia produce signaling molecules, including nod-factors and effector proteins deployed by type III secretion systems, that can modulate host compatibility (C. Masson-Boivin & Sachs, 2018). Compatible rhizobia enter root cells, where they differentiate into endosymbiotic bacteroids that can fix nitrogen in exchange for photosynthates. The legume-rhizobia symbiosis exhibits features that are common to diverse microbial mutualisms, including host acquisition of microbial partners from diverse environmental pools, reciprocal exchange of fitness-enhancing services, and the expression of specificity traits in both partners (Oldroyd et al., 2011).

Bradyrhizobium spp. are found in diverse soils and habitats, and nodulate thousands of legume species (Ormeño-Orrillo & Martínez-Romero, 2019; Parker, 2015; VanInsberghe et al., 2015). *Bradyrhizobium* genomes typically include a core set of chromosomal genes that are vertically inherited, in addition to integrative and conjugative elements, a subset of which express symbiosis functions and can be horizontally transmitted and reshuffled among genomes (i.e., symICEs (Weisberg, Rahman, Backus, Tyavanagimatt, & Chang, 2022)). The symICE encodes the canonical set of loci that encode the capacity to fix nitrogen and that mediate host specificity, including Nod-factors, exo- and lipo-polysaccharides, type III or IV secretion system. A striking but poorly understood aspect of *Bradyrhizobium* populations is that they commonly include 'epidemic genotypes', i.e., strains sharing sets of alleles that achieve high local or regional frequencies (A. C.

Hollowell et al., 2016; La Pierre et al., 2017; J. L. Sachs et al., 2009a; Vinuesa et al., 2005). The most abundant epidemic genotypes share hundreds of chromosomal alleles, have expansive biogeographic ranges (i.e., > 500 km), and exhibit capability to resist multiple antibiotics and catabolize diverse carbon sources, suggesting potential mechanisms for superior fitness in the soil environment or during host infection (A. C. Hollowell et al., 2016; Amanda C. Hollowell et al., 2015, 2016b).

Here we investigated the biogeography, host range, and genomic variation of *Bradyrhizobium* symbionts that associate with an *Acmispon spp.* host metacommunity. Root-nodule isolates were cultured from nine *Acmispon* species across an ~800 km transect in California, USA that included seventeen natural sites **(Table 2.1)**. Previous work uncovered epidemic *Bradyrhizobium* genotypes from a set of 358 *A. strigosus* nodule isolates, where strains were genotyped at four chromosomal genes and four genes on the symICE, previously described as the symbiosis-island (Amanda C. Hollowell et al., 2016b). Subsequently, 86 of those *Bradyrhizobium* isolates from *A. strigosus* had their full genomes sequenced (Weisberg, Rahman, Backus, Tyavanagimatt, & Chang, 2022). In the present study we sequenced genomes of an additional 177 *Bradyrhizobium* isolates from eight additional *Acmispon* host species using a hybrid Illumina and Nanopore approach. Combining these datasets, we analyzed the eight original loci to quantify abundance and biogeographic patterns of chromosomal and symICE genotypes. Our primary goals were to i) infer the roles of host species versus local habitat in structuring *Bradyrhizobium* populations, ii) investigate epidemic genotypes and their capacity to expand across habitats and host species, and iii) examine evidence of horizontal gene transfer. In parallel, iv)

whole genome sequence datasets were used to examine the degree to which epidemic

genotypes are clonal, and v) to investigate genomic drivers that shape host and

environmental adaptation in *Bradyrhizobium*.

Materials and Methods

Field collections – *Bradyrhizobium* isolates were cultured from the root nodules of nine *Acmispon* species across an ~800 km transect in California, United States, including *A. americanus*, *A. argophyllus* var niveus, *A. dendroideus* var. dendroideus, *A. glaber*, *A. grandiflorus* var. grandiflorus, *A. heermanni, A. niveus, A. parviflorus,* and *A. strigosus* (**Table S2.1**). All of these host species were predicted to be nodulated by *Bradyrhizobium spp.*, based on the growth rate and colony morphology of nodule cultures (Torres-Martínez et al., 2021). Host plants were sampled at 17 field sites across an ~800 km transect ranging from McLaughlin Reserve in Northern California to Anza Borrego State Park in Southern California near the Mexican border (**Table 2.1**). Plants were excavated, transported to the laboratory and their roots were washed with tap water to remove all soil particles. Whole nodules were dissected from host roots, surface sterilized in bleach (5% sodium hypochlorite), rinsed in sterile water, and cultured by crushing nodules and plating contents on modified arabinose gluconate at 29°C (MAG, 1.8% w/v agar) (J. L. Sachs et al., 2009a). Nodule cultures were plated at low density to generate individual colonies, and one clone per nodule was archived at -70°C in 50% glycerol, based on the assumption that most nodules harbor a single rhizobial clone (E. L. Simms et al., 2006).

Genotyping and whole genome sequencing – Previous work genotyped 358 *A. strigosus* nodule isolates using four core chromosomal genes (i.e., *dna*K, *gln*II, ITS, *rec*A) and four genes on the symICE (i.e., *nif*D, *nod*D-A, *nod*Z, *nol*L) (Amanda C. Hollowell et al., 2016b). Among these strains 86 also had their whole genome sequenced (Weisberg,

Rahman, Backus, Tyavanagimatt, & Chang, 2022). In the present study we sequenced genomes of an additional 177 *Bradyrhizobium* isolates that were collected either in 2005 or 2019 using a hybrid approach that combined Illumina and Nanopore platforms (**Table 2.1**). BLASTn and tBLASTx (Camacho et al., 2009) were used to extract the above eight genes from the genome datasets using sequences from Hollowell and colleagues as queries (Amanda C. Hollowell et al., 2016b). If all eight loci were not present in any isolate, it was left out from the final dataset. Sequences were aligned using MAFFT (v7.471) and the filter.seqs command in Mothur (v.1.44.3) was used to remove columns containing insertions/deletions (Katoh & Standley, 2013; Schloss, 2020). Horizontal gene transfer can cause chromosomal and sym-ICE genes to have different evolutionary histories, so sequences were concatenated for the chromosomal and symICE genes separately using the EvobiR (v1.1) package in R (Blackmon & Adams, 2015; Weisberg, Rahman, Backus, Tyavanagimatt, Chang, et al., 2022). The unique.seqs command in Mothur (v.1.44.3) was used to find unique sequences for each locus and a Jupyter notebook (Kluyver et al., 2016) was used to computationally assign unique haplotype labels to the concatenated sequences (**Table 2.2**).

Abundance, range, and clonality of haplotypes – Haplotype frequencies were estimated using randomized sub-sampling with 1000 bootstrap replicates, to minimize biases caused by uneven sampling among host plants or sites. In each sub-sample, one nodule isolate per plant was randomly selected, and this process was repeated with replacement to calculate mean frequencies of each chromosomal and symICE haplotype per field sampling site, in

addition to haplotypes with all 8-loci concatenated together. Following from previous work, haplotypes were defined as 'dominant' at a field site if they were above a minimal cutoff (i.e., occurrence ≥ 4) and their frequency was at least 10% at that field site, and among haplotypes categorized as dominant, they were further defined as 'epidemic' if they were sampled in multiple locations separated by at least 10 km distance (A. C. Hollowell et al., 2016; Amanda C. Hollowell et al., 2016b). Host range variation of dominant haplotypes was analyzed from seven focal field sites where two or more host species were sampled (**Table 2.1**). An unpaired t-test was used to test whether dominant haplotypes nodulate more host species on average compared to non-dominant ones. A Pearson correlation was used to test for a relationship between haplotype abundance and host range. Nodule isolates from *A. strigosus* hosts were overrepresented (371/507 isolates), so analyses that excluded these isolates were also conducted. In this smaller dataset, a minimal cutoff of haplotype occurrence ≥ 3 was used to determine if the same haplotypes were consistently dominant in the other hosts excluding *A. strigosus*.

Average nucleotide identity (ANI) was used to quantify clonality among isolate groups, separately for the chromosome and symICE, using FastANI (Jain et al., 2018). For the chromosome, 3049 core ortholog-gene groups were extracted from 224 *Bradyrhizobium* genomes using GET_HOMOLOGUES in the default setting (Contreras-Moreira & Vinuesa, 2013). An all-against-all search was conducted to make clusters of homologous genes using the BDBH algorithm (Contreras-Moreira & Vinuesa, 2013). A nested bootstrapping method was used and repeated 500 times wherein five genomes were randomly sampled, and 40 aligned genes were randomly sampled and concatenated. For

the symICE, only 21 orthologous-gene groups could be extracted from all *Bradyrhizobium* genomes, which were concatenated, and ANI was estimated without bootstrapping. ANI was calculated for epidemic haplotypes and compared to species groups and the *Bradyrhizobium* metapopulation as a whole.

Phylogenetic and population genetic analyses – Maximum likelihood based phylogenetic trees were reconstructed. The IQTree package (v2.1.1) was used with 1000 ultrafast bootstraps, with automatic selection of substitution models with ModelFinder (Chernomor et al., 2016; Hoang et al., 2018; Kalyaanamoorthy et al., 2017; Minh et al., 2020). Three types of phylogenies were reconstructed for the study, trees that included all the isolates along with publicly available *Bradyrhizobium* genomes as outgroup taxa, trees including all the isolates used in this study but no outgroup taxa, and trees that only included unique haplotypes among our isolates.

Genetic variation was examined to infer host and or local adaptation using separate analyses that focused only on the four chromosomal and the four symICE loci. Firstly, analysis of molecular variance (AMOVA) was used to partition phylogenetic relatedness among samples into variation explained by host plant species, the field site of collection, and the interaction among these factors using the Poppr (v2.9.3) statistical package in R (Kamvar et al., 2014). Secondly, phylogenetic clustering of isolates from specific host species or field collection sites was investigated using the abundance-weighted net relatedness index (NRI) and the nearest taxon index (NTI), that measure the degree to which a group of selected samples exhibits significant phylogenetic clustering relative to a

randomized model of community assembly (J. L. Sachs et al., 2009a; Webb et al., 2002).

For both indices, positive values indicate taxon samples that are more closely related on average than expected from random samples of the population, and negative values indicate samples that are less related on average (i.e., phylogenetic evenness). NRI detects tree-wide patterns of clustering and evenness, while NTI is more sensitive to clustering near the tips of the phylogeny (Kooyman et al., 2011). Both indices were implemented in R using Picante (v1.8.2) (Kembel et al., 2010).

Cophylogenies were reconstructed to infer the differential evolutionary histories of the four chromosomal and four symICE loci. A custom-script written in R (v4.1) was used to map and visualize symICE haplotypes associated with dominant chromosomal haplotypes. Faith's phylogenetic diversity was calculated to estimate the diversity of symICE haplotypes associated with individual chromosomal haplotypes (Faith, 1992). A prediction of HGT is that recently acquired symbiosis genes will have low nucleotide diversity compared to the genomic background at both synonymous and nonsynonymous sites (Bamba et al., 2019). To test this prediction, nucleotide diversity was calculated for both chromosomal and symICE sequences along with all overall nucleotide diversity, for both chromosomal and symICE trees. Nucleotide diversity and Tajima's D were estimated for concatenated chromosomal and symICE loci and all eight component loci using the Pegas (v 1.0) package in R (Paradis, 2010). Tajima's D is a test which can infer evidence of non-random evolutionary processes.

To compare evidence of selection pressure on chromosomal and symICE genes, datasets were analyzed for each genome region. For the chromosomal dataset, 107 core

genes were selected (i.e., present in >99% isolates) from strains in which genome sequences were available, with *B. japonicum* USDA6 as an outgroup (Kaneko et al., 2011). Gene sequences were concatenated using the catfasta2phyml.pl script and a maximum likelihood-based phylogeny was constructed using IQ-Tree (Nylander, 2018). MOTHUR (v.1.47.0) was used to categorize the concatenated sequences into operational taxonomic units (OTUs) based on a 0.005 cutoff on the distance matrix generated by using command dist.seqs (Schloss et al., 2009). The phylogeny and alignment file were pruned to keep one isolate from each OTU and subsequently used for selection analysis. CodeML was used in the ETE3 toolkit for testing hypotheses of evolutionary selection on sites of each gene (negative selection M0, relaxed selection M1, positive selection M2) (Huerta-Cepas et al., 2016; Z. Yang, 2007). Finally, SLR was used to compare the models and positively selected sites were reported (Massingham & Goldman, 2005). For the symICE dataset, the same approach was used, but only 17 core symbiosis genes were found based on the above criteria. For the symICE phylogeny, *Bradyrhizobium* WSM471 (Reeve et al., 2013) was used as an outgroup since *B. japonicum* USDA6 does not share the same conserved genes.

A pan-genome association analysis was conducted on accessory genes (i.e., shared among some genomes but not found in all) with Scoary (v1.6.16) using the no_pairwise option to uncover genes associated with traits and particular groups of isolates (Brynildsrud et al., 2016; Guillier et al., 2020). Multiple traits were assigned to isolates to test association of genes with each trait, i.e., dominant, epidemic, as well as specific abundant taxa. Scoary output was filtered by Benjamini FDR-adjusted p-values < 0.05 and genes positively associated with each trait or taxon were selected (Benjamini & Hochberg, 1995). Genes

positively associated with each trait were tested for GO enrichment in PANTHER (v16.0) using a statistical overrepresentation test for biological processes (Mi et al., 2019). The default whole-genome of *B. diazoefficiens* was used as a reference in the analysis and Bonferroni correction was used for multiple testing correction in the statistical overrepresentation test.

Sequence analyses were carried out using Python (v3.9) and R (v4.1). Datasets and codes are available in a GitHub repository: https://github.com/acarafat/epidemic_haplotypes.

Results

We uncovered haplotypes of the *Bradyrhizobium* chromosome that were highly abundant, biogeographically widespread, and host generalists. A total of 241 chromosomal haplotypes were recovered in the metapopulation, and among these, six haplotypes were categorized as dominant (i.e., occurrence $\geq$ 4, local frequency $\geq$ 10%), and three were further categorized as epidemic (i.e., sampled at multiples sites with $\geq$ 10km distance; **Figure 2.1**). The most striking example was chromosomal haplotype K1_R1_I1_G1, isolated 94 times, comprising 18% of all isolates, distributed across > 700 km, uncovered in ten sampled host populations, and found to infect seven host species. Two other epidemic chromosomal haplotypes (K1_R1_I3_G1, K1_R3_I1_G1) were also uncovered in multiple host populations ranging over > 450 km and found to infect multiple host species (**Figure 2.1**). Almost 7.5% of chromosomal haplotypes were found to infect $\geq$ 2 host species (i.e., 18/241), and these haplotypes were often highly abundant (frequency range 2-94; median = 5.5). Six diverse monophyletic clades of *Bradyrhizobium* were uncovered, however all dominant and epidemic chromosomal haplotypes were categorized as *B. canariense* **(Figure S2.1)**.

When analyzing the *Bradyrhizobium* symICEs, a much smaller subset of the diversity was made up of dominant or epidemic haplotypes, and in approximately half of their occurrences, symICE haplotypes achieved local dominance only when in association with a chromosomal haplotype (i.e., hitchhiking; Hollowell et al. 2016b). A total of 338 symICE haplotypes were uncovered, five were found to be dominant, and two of these were epidemic (D12_A1_L1_Z1, D16_A14_L1_Z6; **Figure S2.2**). Both epidemic symICE

haplotypes were restricted to *A. strigosus* hosts and were categorized as *B. canariense*. The symICE haplotype D16_A14_L1_Z6 was isolated eleven times from two sampling sites (~250km distance) and D12_A1_L1_Z1 was isolated six times from two sampling sites (~40km distance). Fewer than 1.8% of symICE haplotypes were found in more than one host (i.e., 6/338; maximum two hosts) and none of them were dominant. No dominant symICE haplotype was found when *A. strigosus* was removed from the dataset. SymICE haplotypes often achieved high abundance via hitchhiking with dominant chromosomal haplotypes **(Table S2.2)**. Three whole genome haplotypes, constituting all 8 loci from the chromosome and symICE, were found to achieve local dominance based on the bootstrap analysis, all being a combination of one dominant chromosomal and one dominant symICE haplotype (**Figure S2.3**).

Host range and frequency of *Bradyrhizobium* haplotypes were correlated for the chromosomal loci, but no such correlation was found for the symICE loci. Host range of chromosomal haplotypes was positively correlated with their frequency, when analyzing field sites where multiple host taxa were collected (R = 0.42, $P = 1.5 \times 10^{-11}$; **Figure S2.4**). Moreover, dominant chromosomal haplotypes had an average host range of > 2 that was significantly higher than non-dominant haplotypes (2.667 vs. 1.115; t = -6.1184, df = 152, $P = 7.67 \times 10^{-9}$; **Figure 2.2**). With *A. strigosus* isolates excluded from the dataset, two chromosomal haplotypes, K1_R1_I1_G1 and K1_R1_I3_G1, were found to be dominant on the other hosts as well, suggesting that dominance is not restricted to *A. strigosus* symbionts **(Figure 2.3)**. Correlations could not be investigated for symICE haplotypes because of limited host range; only two symICE haplotypes were found to occur at different

field sites (listed above) and two other symICE haplotypes were found to infect two hosts (D14_A58_L36_Z5, D53_A9_L7_Z3).

The genomes of epidemic haplotypes were highly clonal across a large portion of their chromosome, but with little evidence of clonality within the symICE **(Table S2.3)**. Epidemic chromosomal haplotypes had ANI values of > 99.33% across 3049 conserved chromosomal genes (**Figure 2.4A, Table S2.3**), substantially higher than the ANI value for the same loci across *B. canariense* (97.92%), the best sampled *Bradyrhizobium* species, or the full sampled metapopulation (i.e., > 7 species; 94.37%) (**Figure 2.4A, Table S2.3**). For the symICE, only 21 conserved genes could be queried. Among this small dataset, the ANI of epidemic haplotypes was lower than in the chromosomal dataset. The epidemic symICE haplotypes had an ANI value of > 97.94%, similar to that of *B. canariense* (97.56%) or the metapopulation (96.99%) (**Figure 2.4B**, **Table S2.3**).

Phylogenies of the chromosomal and symICE loci were incongruent and each suggested independent evolutionary histories and selective pressures. Clades in the chromosomal tree were clustered into six species of *Bradyrhizobium*, and conversely, the *Bradyrhizobium* species were interspersed across multiple clades in the symICE tree (**Figure 2.5**). The symICE phylogeny exhibited striking patterns of clustering by host species and sampling sites, consistent with host and or local adaptation. Significant phylogenetic clustering was uncovered for the host species including *A. americanus, A. heermannii, A. niveus, A. parviflorus,* and *A. strigosus* (i.e., NRI p-value < 0.05; **Table 2.3**) and for the field sampling locations including Anza Borrego Desert State Park, Bodega Marine Reserve, Bernard Biological Field Station, San Gorgonio Mountains, McLaughlin

Reserve, Motte Rimrock Reserve, and University of California Riverside (**Table 2.4**). The chromosomal phylogeny exhibited negligible phylogenetic clustering by host or sampling site, and only one sampling site had significant phylogenetic clustering (Motte Rimrock Reserve, **Table 2.4**). These data are consistent with the hypothesis that the symICE region evolves in response to both specific hosts and locales. However, these factors are non-independent given that *Acmispon* host communities likely vary among these sites.

Distinct population-structure was indicated by the chromosomal versus the symICE loci. In the AMOVA, most of the genetic variance occurred among the host species for both chromosomal and symICE haplotypes (80.41%, 64.78% respectively). Focusing on variance within sampling sites (and thus controlling for among site variation), 50% of variance in symICE haplotypes occurred among hosts, whereas only ~5% of variance in chromosomal haplotypes did (**Table 2.5**). On the other hand, 15% of variance in chromosomal haplotypes occurred among the field sampling sites, but for the symICE loci, the result was non-significant. These data further support the hypothesis that the symICE region evolves in response to specific hosts but does not support a role for adaptation to local habitat for the symICE.

The cophylogenetic analyses illustrated an evolutionary history where epidemic chromosomal haplotypes acquired a diversity of symICEs (**Figure 2.6**). The most widespread and abundant chromosomal haplotype, K1_R1_I1_G1, acquired 70 different symICE haplotypes, while chromosomal haplotypes K1_R1_I3_G1 and K1_R3_I1_G1 acquired 11 and 5 symICE haplotypes, respectively. Phylogenetic diversity of the acquired symICE haplotypes varied widely, with Faith's PD values of 0.195, 0.107, and 0.07 for

epidemic haplotypes K1_R1_I1_G1, K1_R1_I3_G1, and K1_R3_I1_G1, respectively (**Table 2.6**). There was a positive correlation between the number of symICE haplotypes and their phylogenetic diversity (F = 14.36, df = 4, P = 0.0192, adjusted R-squared = 0.727) (**Figure S2.5**). Similarly, within the clades labeled as 'SYM1' through 'SYM5' in the symICE phylogeny, the clade-members often had lower nucleotide diversity in the concatenated symICE loci sequences compared to corresponding isolate's concatenated chromosomal loci sequences (except SYM-5 clade; **Figure S2.6**). This reflects the predicted outcome of acquisition of evolutionary similar symICE haplotypes in very different chromosomal haplotype background by horizontal gene transfer process.

The genomes bearing epidemic chromosomal haplotypes appeared to be shaped by two distinct evolutionary processes, population-expansion of the chromosomal region, and positive selection on the symICE. Significantly negative Tajima's D values were found for the chromosomal loci, indicating evidence for population expansion, but no such evidence was found for the symICE loci (**Table S2.4**). Tajima's D value for chromosomal haplotypes was also low compared to the symICE haplotypes, suggesting a recent population expansion restricted to the chromosomal region of the genome **(Table S2.4)**. Conversely, the symICE loci appeared to be shaped by positive selection. Almost half (47%) of the conserved symICE loci that were analyzed exhibited evidence of positive selection pressure (8/17 SI genes), whereas < 3% of chromosomal loci exhibited such evidence (3/107 chromosomal loci, comparing M2 vs M1, P < 0.05) **(Table S2.5)**. These chromosomal genes include *nagK*, *ctrA*, and *resA* which has only 1-2 sites under selection pressure. Within the symICE region, the gene *nifQ* was found to be under strong positive

selection pressure on seven sites **(Table S2.5)**. The *nifQ* gene is molybdenum doner to FeMo-co and necessary to incorporate molybdenum into nitrogenase (Hernandez et al., 2008). Other symICE genes includes *resA*, blr1754, *hisC*, *fixU*, blr1755, *nodU*, and acyltransferase which each have a handful of sites under positive selection **(Table S2.5)**.

A statistical overrepresentation test for gene ontology categories (GO terms) recovered significant enrichment for gamma-aminobutyric acid metabolism (GABA) up to 100-fold in two epidemic chromosomal haplotypes (K1_R1_I1_G1, K1_R3_I1_G1) **(Figure S2.7)**. GABA is a plant metabolite that has a role in stress adaptation and plant-microbe interactions, and has been suggested to provide additional energy generation bypassing the decarboxylation step in the TCA cycle during N_2-fixation, thereby increasing efficiency of nitrogen-fixation (Hijaz & Killiny, 2019; Sulieman, 2011). Other GO terms with >10-fold enrichment included 'fatty acid elongation' and the broad term of 'biosynthetic processes' for dominant haplotypes, whereas epidemic haplotypes additionally included the terms 'monocarboxylic acid' and 'lipid biosynthetic processes'. Lipids and lipid-like polymers such as poly-β-hydroxybutyrate are known as important storage molecule for differentiated rhizobia within nodules (Terpolilli et al., 2016). Finally, the chromosomal haplotypes K1_R1_I1_G1 and K1_R3_I1_G1, were associated with >10-fold enrichment of the GO terms 'cellular amino acid' and 'alpha amino-acid metabolic processes'. Amino acids are important for rhizobia, both in free-living and symbiotic states, which can be used not only in protein synthesis, but also carbon and nitrogen metabolism, signaling, and other processes, and amino-acid metabolism is tightly regulated

and plays a critical role in the nitrogen-fixing symbiosis between rhizobia and plants (Dunn, 2015).

Discussion

A major goal of molecular epidemiology is to understand how mutation, gene acquisition, drift, and selection shape bacterial adaptation. Addressing these questions is important for rhizobia because we have little understanding of the genomic drivers of fitness, either in natural or managed landscapes, and because of their critical role in generating nitrogen for plant hosts. Our results here suggest three important patterns about genome evolution in *Bradyrhizobium* populations. Firstly, our data demonstrate that a substantial portion of the *Bradyrhizobium* genome – herein defined as the chromosome – is highly clonal and shows evidence of a massive expansion across sites we sampled in California. A small handful of epidemic chromosomal haplotypes dominate the diverse *Bradyrhizobium* metapopulation, having spread over hundreds of kilometers of space, and forming symbiotic root nodules on up to seven different *Acmispon* host species. Variation in spatial abundance of chromosomal haplotypes is consistent with fitness variation, suggesting that the epidemic haplotypes exhibit a fitness advantage on hosts, in the soil, or in both environments. This type of epidemic clonality was also observed in other studies including multilocus analysis of different rhizobium species (Mcinnes, 2004). Secondly, the chromosome and symICE are largely divergent in their evolutionary patterns, with phylogenetic incongruence between them. Chromosomal haplotypes, especially those that are epidemic, can expand across diverse host species and distant field sites, but signals of adaptation to hosts and sites appears localized to the symICEs, that these chromosomes acquire. These data reveal that the chromosome and symICEs of *Bradyrhizobium* are responding to different selective pressures and are evolving independently. The different replicons of 'divided' bacterial

93

genomes can commonly have distinct evolutionary histories that are functionally associated with different phases of the bacterial life cycle (diCenzo & Finan, 2017). Given that horizontal transfer of symbiosis genes appears to be common in rhizobia, and is likely critical in their adaptation to novel hosts, divergent selection is expected among their chromosomal and mobile regions (Epstein & Tiffin, 2021). Thirdly, the symICE shows patterns of molecular evolution consistent with adaptation to specific host species and to local field sites, and lacks evidence of clonal expansion that was observed for the chromosome. These data suggest that legume hosts stringently select for the symbiotic haplotypes (i.e., the symICE or plasmid region), but not for the remainder of the genome (Li et al., 2016; Parker, 2012), leading to contrasting population structure in the symbiosis and chromosomal loci (Pérez Carrascal et al., 2016; Riley et al., 2022).

Our dataset demonstrated that in a large and diverse community of rhizobia, a relatively small subset of chromosomal haplotypes was extremely successful in natural host populations **(Figure 2.1)**. It showed that the chromosomal haplotypes that were most successful were restricted to *B. canariense,* which is a commonly sampled rhizobia species found in root nodules of diverse legumes including the *Lupinus*, *Ornithopus,* and *Vigna spp.* in multiple continents (Stepkowski et al., 2005; Stępkowski et al., 2011; Wade et al., 2014). A positive skew in haplotype abundance has been observed in many rhizobial populations and is consistent with a fitness hierarchy. For instance, a meta-analysis of rhizobial populations reported that a handful of rhizobia genotypes dominate nodules in host populations, with individual genotypes occupying more than 30% of those nodules (Mcinnes, 2004). On a study of *Sesbania cannabina* nodulating rhizobia from China,

including multiple genera of rhizobia, genotype variance was observed where five genospecies accounted for a third of 874 isolates (Zhang et al., 2020). In a much broader study of 100 native legume species in 8 sampling sites from North America researchers found that *Bradyrhizobium* from specific regions were significantly more similar than random, based on permutation tests on both housekeeping genes and for *nif*D, but some bacterial groups were dispersed across multiple regions and associated with diverse legume host lineages as well, reflecting epidemic patterns of spread (Koppell & Parker, 2012). Similar to our dataset, the chromosomal genotypes showed no evidence of host adaptation and very little evidence of local adaptation (Koppell & Parker, 2012). Dominance of a single lineage of *Bradyrhizobium* has also been observed in managed settings. In two soybean sites in Canada where only one sampling site had recent inoculation history, only 2-6 unique genotypes of native rhizobia accounted for >60% abundance, showing that clonal expansion appeared to be driven by host mediated selection, whereas only ~20% of isolates could be attributed to inoculation source (in the site with recent inoculation history) (Tang et al., 2012). Our results show a lack of phylogenetic structure by sampling sites and host species in the chromosomal haplotype phylogeny **(Figure 2.5, Table 2.3-4)**, which is not surprising since similar findings were previously reported (Riley et al., 2022). More specifically, in *Bradyrhizobium* it has been found that multilocus-based phylogeny clades often have divergent host species (Koppell & Parker, 2012; Parker, 2015; Riley et al., 2022).

Our genomic analyses suggested that epidemic haplotypes are equipped with diverse metabolic processes **(Figure S2.7)**, several of which might be linked with fitness

superiority. Geographic spread among diverse soils has been attributed to diverse catabolic capacity of rhizobia since epidemic strains were associated with broader carbon source utilization compared to other strains in a metapopulation (Amanda C. Hollowell et al., 2016). It has been shown that rhizobial populations can persist in the soil without legume hosts after releasing back from nodule for years (Denison & Kiers, 2011), therefore, stress adaptation in free-living stages in the soil can play an important role in the superior fitness of the epidemic strains (Kajić et al., 2016). Rhizobial symbiosis genes (and related loci on MGEs) appear to be expressed mainly when the bacteria are within host nodules, wherein the remainder of the genome (i.e., the chromosome) appears to be largely expressed under low nutrient growth conditions, similar to what rhizobia might experience in soil (Pessi et al., 2007; Uchiumi et al., 2004). However, while our data suggest that fitness differences are driving haplotype abundance patterns in these populations, we cannot rule out the role random processes like drift, moreover, we still have only scant understanding of the traits that drive fitness variation in rhizobia populations.

The symICE showed strong signatures of host adaptation, with most haplotypes being restricted to individual host species, consistent with functions encoding root nodulation specificity (Fauvart & Michiels, 2008; Rogel et al., 2011; Walker et al., 2020; Zhao et al., 2018). We uncovered significant phylogenetic structure on the symICE phylogenetic tree among strains isolated from different host species and sampling sites **(Figure 2.5, Table 2.3-4)**. These data suggest that symICEs genotypes are adapted to a restricted set of legume hosts, which can drive such phylogenetic structure. For example, in *Rhizobium* species sampled from common bean, the symbiotic plasmids shows extreme

similarity without showing any such relationship with the chromosome (Pérez Carrascal et al., 2016). However, in our dataset, not all sampling sites had equivalent sample-representation from different hosts, so we cannot dissect local adaptation from adaptation to the subsets of hosts that are present in each locale. A high proportion of symICE loci were found under strong positive selection pressure in contrast to chromosomal loci, which potentially indicates selective controls exerted by host. However, this not a universal pattern, as other studies reported absence of positive selection and predominance of purifying selection on symbiosis genes (Bailly et al., 2006; Epstein & Tiffin, 2021).

Our data also suggest that different evolutionary processes shape rhizobia in natural communities, such as we studied here, compared to agronomic settings. In our study, the symICE did not show extensive local dominance or epidemic patterns, unlike the chromosomal region. In cases where the symICE haplotypes achieved local dominance, it was often associated with hitchhiking, meaning that the symICE haplotypes were linked to dominant chromosomal haplotypes. The symICE was also highly varied in terms of gene content, with few conserved genes, and little evidence of clonality in the conserved genes, unlike what we found for the chromosome. Conversely, in studies in managed soils, single symICE haplotypes have been found to experience massive selective sweeps, where host compatible symICEs rapidly spread through a diverse population of chromosomal haplotypes (Sullivan et al., 1995). Host-specific symICE acquisition in diverse chromosomal background was also observed in the domestication of chickpea (*Cicer arietinum*), where cultivation of chickpea beyond its native range resulted to nodulate by evolutionarily divergent and geographically isolated *Mesorhizobium* taxa which share a

common set of symICE (J. Liu et al., 2020; von Wettberg et al., 2018) (Hill et al., 2021). Since managed settings are more homogeneous, both for soil characteristics and for host taxa, this might favor a single symICE haplotype, unlike what we found in our diverse plant communities.

The molecular determinants of rhizobial abundance, biogeography, and host range appear to involve independent processes in the symICE and the chromosome. Within the chromosome, population expansion has greatly reduced nucleotide diversity, whereas the symICE exhibits little gene conservation and exhibits evidence of positive selection **(Table S2.4-5)**. The low observed genetic diversity in chromosomal haplotypes reflects a recent history of population expansion. On the other hand, the incongruent phylogenetic topology among the chromosomal and symICE indicates frequent horizontal gene transmission leading to recurrent acquisition of symICE by dominant/epidemic haplotypes **(Figure 2.5, 2.6)**. The process of chromosomal halplotypes acquiring diverse symICEs allows for dominant chromosomal haplotypes to infect multiple host and spread geographically **(Figure 2.6, S2.5)**. Among rhizobial taxa, host range can vary from highly specialized to exceptionally broad (Ehinger et al., 2014; Parker, 2012; Pueppke & Broughton, 1999) and acquisition of symbiosis loci is suggested to play a prominent role in host range evolution (Remigi et al., 2016). Phylogenetic incongruency between core chromosomal and symbiotic genes provides substantial evidence of HGT shaping rhizobial evolution and population structure, driving mobilization of plant specific symbiotic genes in locally adapted rhizobial haplotypes. A study that sampled nodule associated *Mesorhizobium* from *Lotus japonicus* across a wide range of Japan observed low sequence diversity in multiple

core symbiotic genes along with phylogenetic-incongruance of chromosomal and symbiotic phylogeny, indicating signature of HGT of symICEs (Bamba et al., 2019). In a multilocus analysis of chromosomal and symbiosis genes from alpha- and beta-proteobacteria from Cape Fynbo legumes, such signature of genus- and species-level HGT resulting phylogenetic-incongruance between chromosomal and symbiosis genes was observed between distinct rhizobial genus of *Rhizobium* and *Mesorhizobium* (Lemaire et al., 2015). Geographically widespread evidences of HGT of symICEs within and between diverse genera of Rhizobia has been compiled and reviewed (Andrews et al., 2018; Wardell et al., 2022). Among rhizobial taxa, host range can vary from highly specialized to exceptionally broad (Ehinger et al., 2014; Parker, 2012; Pueppke & Broughton, 1999) and horizontal gene transfer (HGT) of mobile genetic elements is suggested to play a prominent role in host range evolution (Remigi et al., 2016). For instance, a study on 196 genomes of *Rhizobium leguminosarum* sv. *trifolii* identified 171 genes that are mobile across species-boundary in four blocks, which involves mostly symbiosis genes, but smaller proportion of chromosomal genes too (Cavassim et al., 2020). However, for the majority of the genes in the rhizobial chromosome, introgression events across species boundaries are rare (Cavassim et al., 2020; Epstein et al., 2012; Pérez Carrascal et al., 2016). The phylogenetic patterns detected for the symICE, structured both by host taxon and sampling site, along with the expansion of dominant chromosomal haplotypes in distant host-populations, taken together, can be explained by HGT driven host-specific symICEs acquisition.

Symbiotic nitrogen fixation is a joint phenotype resulting in traits expressed both by the host and the symbionts (Joel L. Sachs et al., 2018). Legumes exert selective pressure

on symbionts via host control, selecting for compatible and beneficial symbionts (Oono et al., 2009; John U. Regus et al., 2017; Ellen L. Simms & Taylor, 2002; Westhoek et al., 2021, 2017). Some researchers have argued that selection by the host is stronger than the selection in the soil. For instance, a recent study on 202 accessions of *Medicago truncatula* hosts and a community of 88 *Sinorhizobium meliloti* strains found extensive host by symbiont effects (G × G) on rhizobial fitness, suggesting that host genotype is an important driver (Epstein et al., 2022). Another study, using a cross-inoculation experiment with 9 *Mesorhizobium* in 15 *Lotus japonicus* accessions, reported variation in partner quality associated with core genome variation, i.e., chromosomal variation, suggesting that the chromosome can also interact with the host genotype (Bamba et al., 2020). In a select and resequence experiment, Burghardt and collegues found that altering the host haplotypes reshuffled rhizobia strain dominance in an experimental system, but found little effect in altering the non-host, *in vitro* environment (Burghardt et al., 2018). Nodule forming rhizobia also have host-independent saprophytic free-living stages in diverse soil-types where local adaptation might be crucial for fitness (Poole et al., 2018). Interestingly, a strong coevolutionary G × G × E effect leading to selection mosaic was reported on different genotypes of *Vicia cracca* and *Rhizobium leguminosarum* when different soild nitrogen levels were introduced (Van Cauwenberghe et al., 2016). *Bradyrhizobium* strains that were evolved in vitro for 500 generations, with no host interaction, evolved reduced host benefits, suggesting a potential tradeoff between these host associated and free-living parts of their lifecycle (Joel L. Sachs et al., 2011). These studies reinforce that selection by the environment can play important role in shaping rhizobia, beyond the effect of the host.

The role of the soil environment in selecting rhizobia is still poorly understood, but is likely critical in shaping traits for rhizobial survival during transmission and might be critical in adapting to diverse soils.
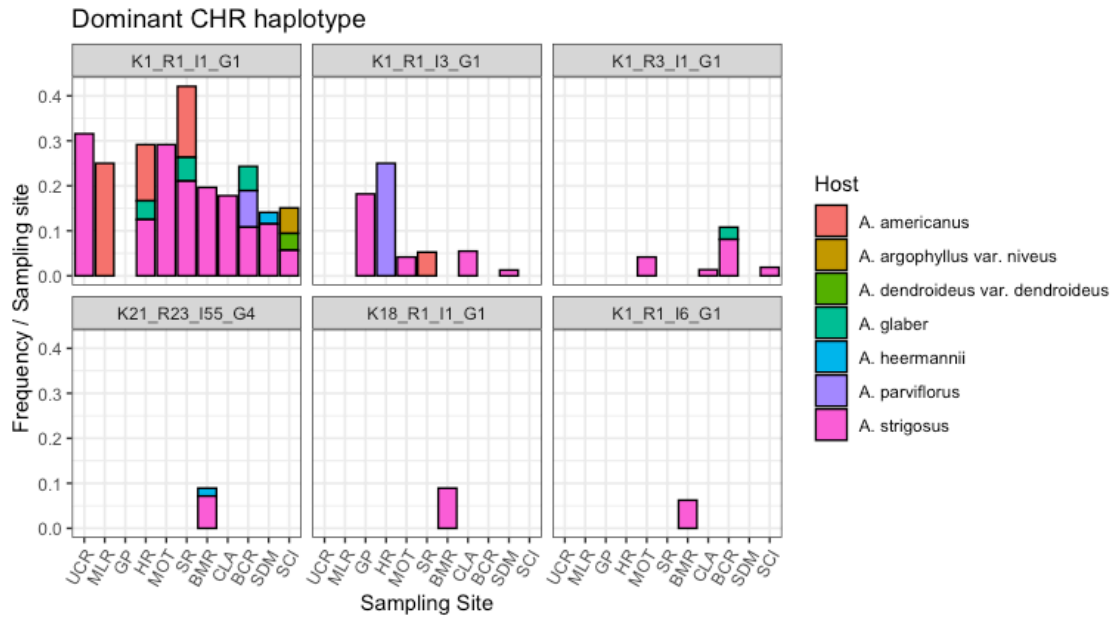
Figure 2.1: Six dominant chromosomal haplotypes were found in the meta-population; three were also epidemic (top three panels). Dominant is a haplotype having high occurrence (≥ 4) and comprise high relative frequency (≥ 10%) in a sampling site. An epidemic is a dominant haplotype found in multiple sites (> 10 km distance). K1_R1_I1_G1 is the most prevalent chromosomal haplotype, which is dominant in multiple sampling sites.
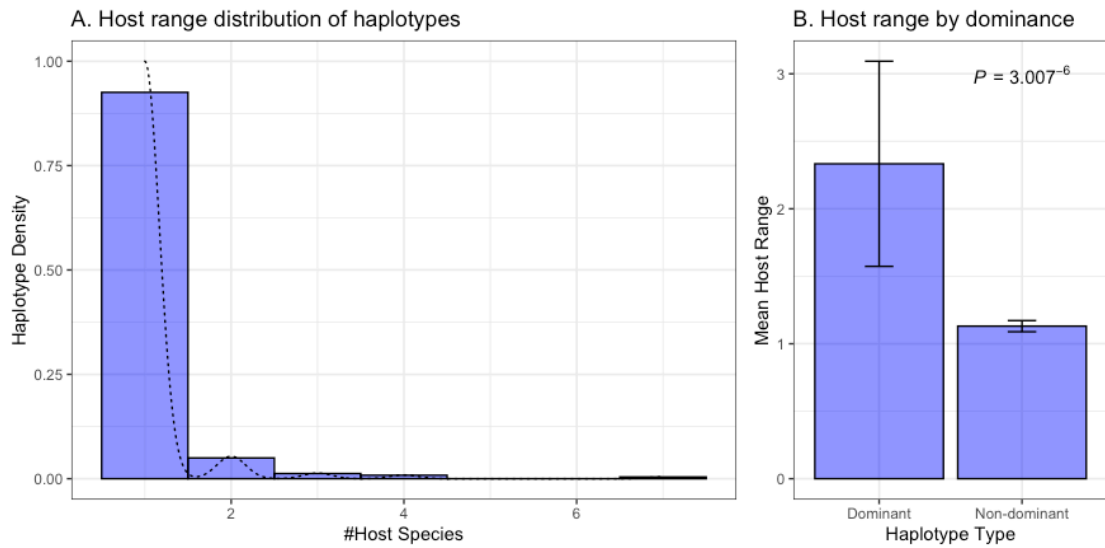
A. Host range distribution of haplotypes

B. Host range by dominance

$P = 3.007^{-6}$

Figure 2.2: Dominant haplotype has a higher host range. (a) Although most chromosomal haplotypes infects single host species, (b) but dominant haplotypes infect two or more host hosts compared to non-dominant haplotypes (Unpaired t-test: t = -6.1184, df = 152, **p-value = 7.669e-09**).
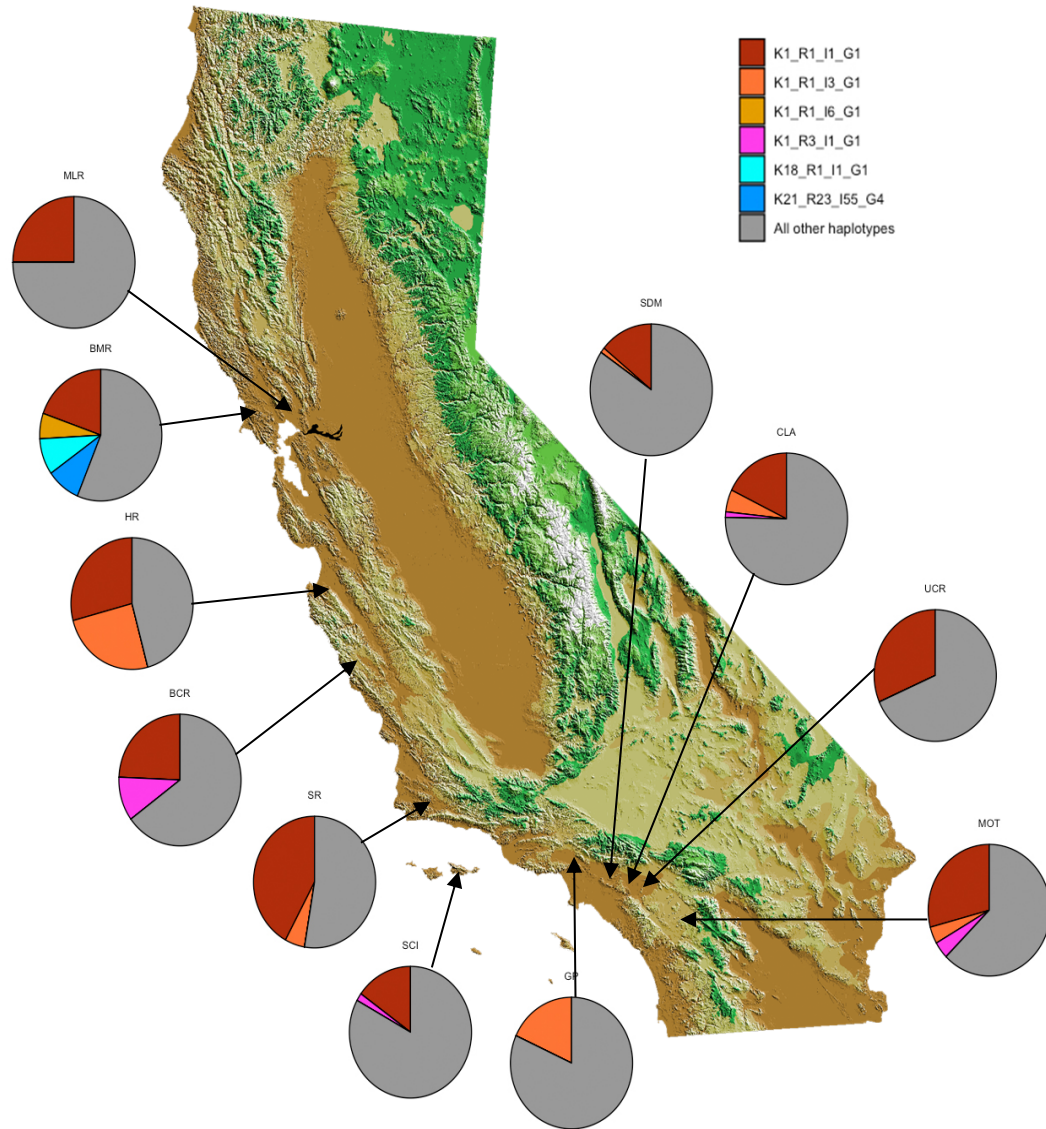
Figure 2.3: Geographic spread of chromosomal haplotypes in California metapopulation. Map credit: U.S. Geological Survey.
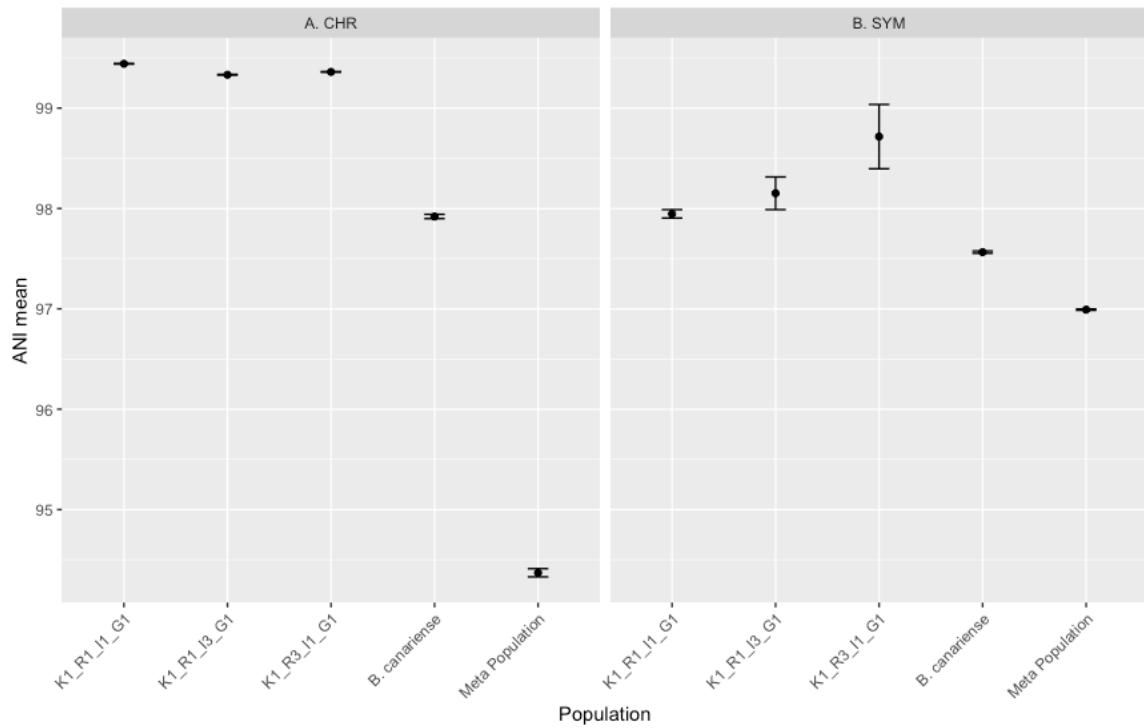
Figure 2.4: chromosomal haplotypes are conserved in genome level. **(A)** Average Nucleotide Identity (ANI) estimated for 3049 core chromosomal ortholog-group genes by bootstrapping genomes in five categories: within three epidemic haplotype groups, the *B. canariense* species group where these haplotypes belong, and the metapopulation. **(B)** ANI estimated for 21 concatenated symbiotic genes for all genomes in above mentioned categories.
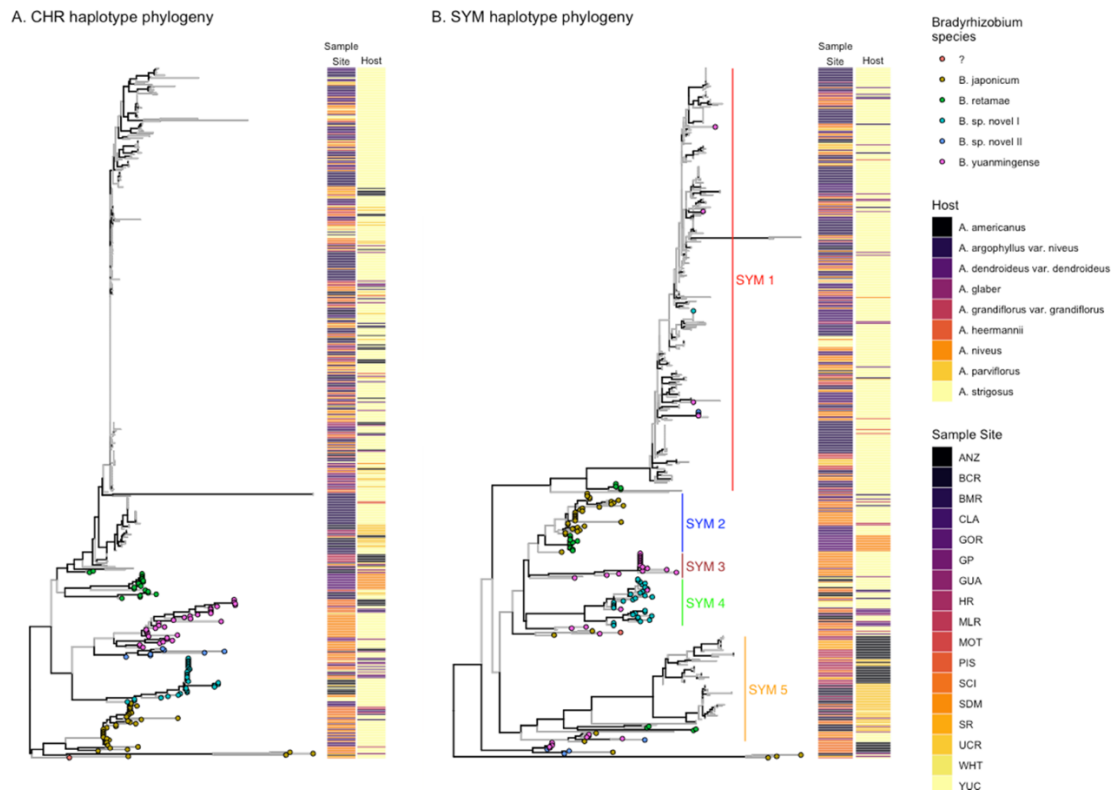
Figure 2.5: Maximum-likelihood based phylogenetic tree for concatenated chromosomal genes (A) and symbiosis genes (B) for 507 *Bradyrhizobium* isolates used in this study. Tip-points indicate *Bradyrhizobium* species. For simplicity, *B. canariense* taxa does not have any tip-points. Black branches indicates bootstrap support (SH-aLRT and UFBoot) >70. Major symICE phylogeny clades are assigned as SYM1 to SYM5. Heat-plot on the side of each phylogeny indicates sampling site and host for each isolate.

Figure 2.6: Co-Phylogeny of chromosomal (left) and symICE (right) haplotypes for three epidemic (A-C) and three dominant (D-F) haplotypes. Each line between two nodes indicates that chromosomal and symICE haplotype pair has been associated in the same genomic background.
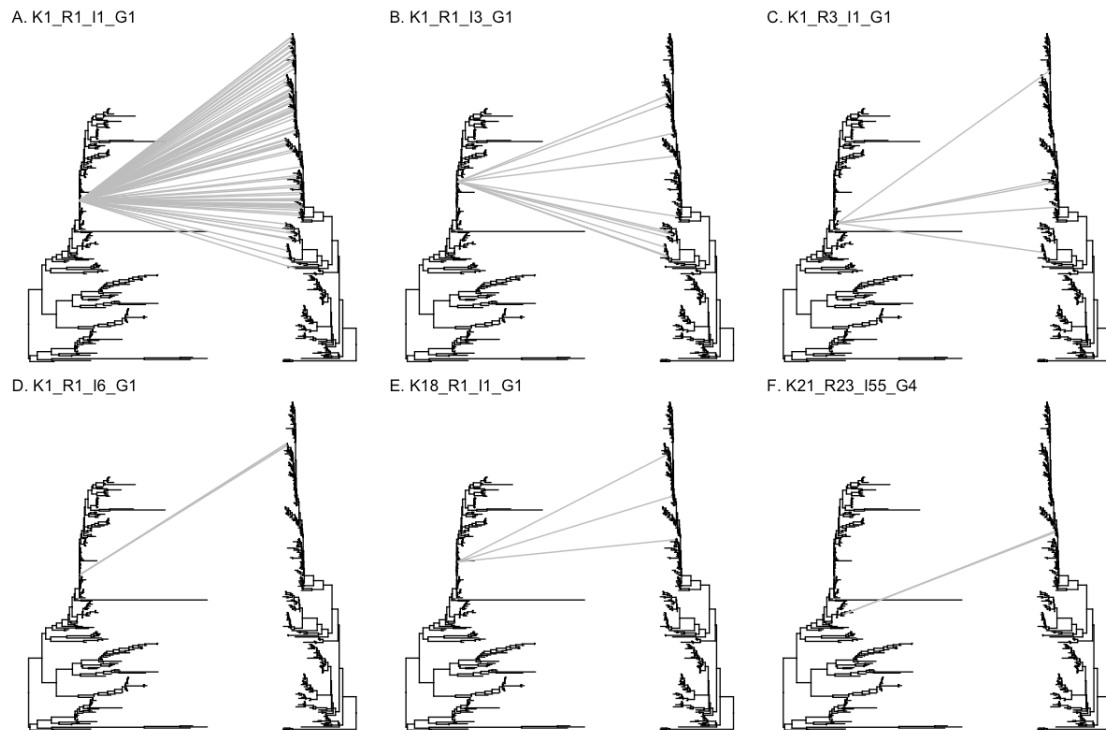
Table 2.1. Summary of the number of species collected at different sites, and isolates collected per species. In bold are highlighted sites for which we could perform host vs soil type comparisons. Numbers are excluding isolates for which we obtained pseudomonas and not rhizobia

Sampling site	Sampling Year	Host Species	Number of Isolates
Anza Borrego Desert State Park - Palm Canyon	2011	A. strigosus	9
Anza Borrego Desert State Park - Roadside	2011	A. strigosus	1
Big Creek Reserve	2019	A. glaber	10
		A. parviflorus	19
		A. strigosus	8
Bodega Marine Reserve	2005	A. strigosus	108
		A. heermannii	2
		A. parviflorus	2
Burns Piñon Ridge Reserve	2011	A. strigosus	16
Griffith Park	2013, 2019	A. glaber	2
		A. strigosus	9
Guadalupe-Nipomo Dunes National Wildlife Refuge	2013	A.strigosus	1
Hastings Reserve	2019	A. americanus	7
		A. glaber	5
		A. parviflorus	9
		A. strigosus	5
McLaughlin Reserve	2019	A. americanus	
Motte Rimrock Reserve	2009	A. strigosus	24
Pismo Dunes Natural Preserve	2013	A. strigosus	3
Robert J. Bernard Biological Field Station	2011, 2012, 2014	A. strigosus	73
San Dimas Canyon	2011, 2012, 2019	A. heermannii	2
		A. strigosus	64
San Dimas Reservoir	2011	A. strigosus	12
San Gorgonio Mountains - Santa Ana River South Fork Fishing Area	2019		
San Gorgonio Mountains - Barton flats campground		A. niveus	13

Santa Cruz Island	2019	A. americanus	9
		A. argophyllus var. niveus	10
		A. dendroideus var. dendroideus	13
		A. grandiflorus var. grandiflorus	4
		A. parviflorus	6
		A. strigosus	11
Sedgewick	2019		
		A. americanus	10
		A. glaber	3
		A. strigosus	6
University of California - Riverside Hills	2009, 2013, 2014	A.strigosus	19
Whitewater Preserve	2013	A. strigosus	2
		Total	507

Table 2.2: Dataset formation

Locus	Hollowell 2016b	Draft Genome	Combined (with overlapping from both sources)	Final Dataset
dnaK	356	272	565	507
glnII	357	333	566	
ITS	320	238	527	
recA	357	340	565	
nifD	349	311	555	507
nod-A	351	251	557	
nodZ	350	251	556	
nolL	354	251	560	

Table 2.3: Phylogenetic structure by host species (Abundance Weighted)

Host	CHR				SYM			
	NRI	NRI p-value	NTI	NTI p-value	NRI	NRI p-value	NTI	NTI p-value
A. americanus	-0.5953493	0.774	1.49884563	**0.06**	**3.979253**	**0.003**	1.0893142	0.135
A. argophyllus var. niveus	-5.2968415	0.999	0.71895412	0.357	-0.2389036	0.548	0.1851932	0.457
A. dendroideus var. dendroideus	-9.7330134	1	0.97686781	0.217	0.5114537	0.259	-0.3958305	0.698
A. glaber	-5.6082505	0.999	-3.1725846	0.997	-0.4088543	0.626	-3.2501499	0.994
A. grandiflorus var. grandiflorus	0.1941924	0.444	-0.0945152	0.572	1.3002148	0.154	0.6620587	0.364
A. heermannii	-0.9510647	0.865	NaN	0.5005	**1.5320736**	**0.034**	1.2299866	0.123
A. niveus	1.0000009	0.136	0.28822724	0.398	**2.6744731**	**0.003**	0.4276384	0.363
A. parviflorus	0.8632918	0.185	0.95127395	0.17	**3.9989458**	**0.001**	0.8019039	0.22
A. strigosus	-2.8168417	0.986	-3.6087906	1	**4.4509774**	**0.001**	-2.4681633	0.986

Table 2.4: Phylogenetic structure by sampling site (Abundance Weighted)

Sampling Site	CHR				SYM			
	NRI	NRI p-value	NTI	NTI p-value	NRI	NRI p-value	NTI	NTI p-value
ANZ	0.02853216	0.509	-0.3523819	0.642	**1.5089752**	**0.082**	-0.7227459	0.78
BCR	-1.2766995	0.9	-0.4033566	0.667	-0.3974194	0.645	0.18322322	0.48
BMR	-0.0485783	0.583	**1.75287324**	**0.028**	**3.5415746**	**0.001**	**2.12306194**	**0.006**
CLA	-1.7062134	0.939	1.11386323	0.133	**4.0798237**	**0.001**	**1.74654966**	**0.013**
GOR	0.97414031	0.149	0.35680267	0.36	**2.7226594**	**0.003**	0.49044694	0.327
GP	-3.9764682	0.996	-0.919458	0.836	0.5277182	0.245	-1.3628154	0.911
GUA	NaN	0.5005	NaN	0.5005	NaN	*NA*	NaN	*NA*
HR	0.94629349	0.085	1.26991227	0.052	-0.6288127	0.713	**1.448079**	**0.04**
MLR	0.58613975	0.309	0.77318665	0.33	**3.0712615**	**0.01**	1.04439872	0.121
MOT	**1.76143127**	**0.009**	1.50942989	0.057	**5.7302851**	**0.001**	1.05690864	0.138
PIS	-3.5133805	0.998	-3.6809988	0.998	-1.9485236	0.991	-2.3428573	0.978
SCI	-8.6796288	1	0.49426583	0.337	-0.6849315	0.76	-0.1421693	0.596
SDM	-6.4203888	0.999	0.02302634	0.516	-0.1194372	0.521	-0.0010485	0.518
SR	0.57099356	0.378	-1.5849726	0.932	-0.4532744	0.663	-2.6189443	0.986
UCR	0.28476818	0.461	0.58592815	0.311	**4.1832281**	**0.003**	**1.49070676**	**0.039**
WHT	-1.7078942	0.921	-1.7266315	0.931	-0.4170975	0.541	-0.3982885	0.538
YUC	-3.0846759	0.989	1.08133529	0.146	0.1022081	0.442	**1.8522668**	**0.001**

Table 2.5: Analysis of molecular variance

CHR					
Comparison	**Df**	**σ**	**Variance %**	**Φ**	**_P_ value**
Between Sampling Site	16	9.008	14.16655	0.14166554	0.002
Between Host within Site	16	3.448	5.423398	0.06318513	0.043
Within Hosts	258	51.130	80.410049	0.14166554	0.001
Total	290	63.587	100		
SYM					
Comparison	**Df**	**σ**	**Variance %**	**Φ**	**_P_ value**
Between Sampling Site	16	-9.203	-14.49452	-0.145	0.715
Between Host within Site	16	31.564	49.71315	0.434	0.001
Within Hosts	279	41.131	64.78136	0.352	0.001
Total	311	3.493	100.00000		

**Table 2.6:** chromosomal-symICE haplotype association summary.

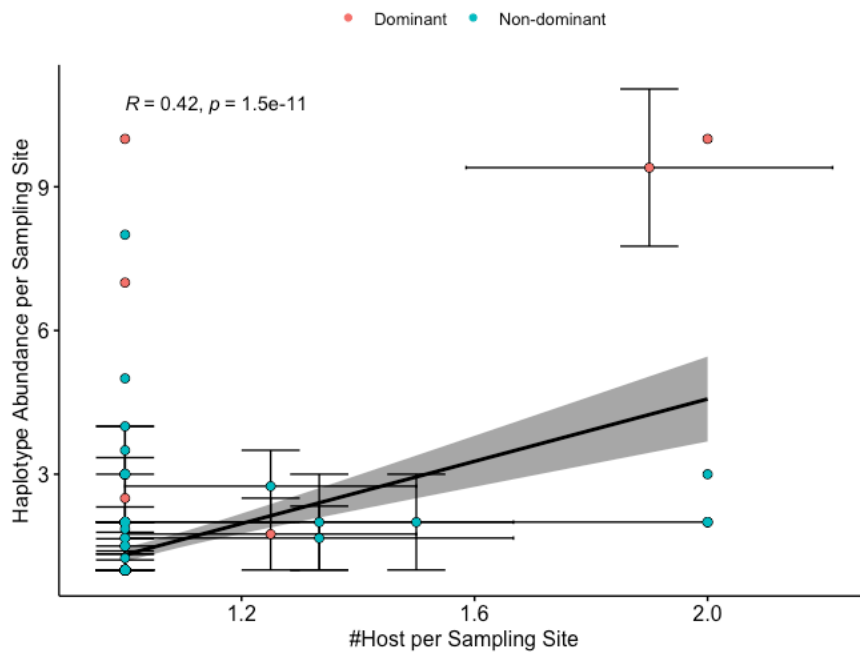| Chromosomal Haplotype	# symICE Haplotype associated	Phylogenetic Diversity of Connected symICE haplotype	# Unique Host infected	# Site
K1_R1_I1_G1	70	0.1953965	7	10
K1_R1_I3_G1	11	0.1079797	2	6
K1_R3_I1_G1	5	0.07129972	1	4
K1_R1_I6_G1	3	0.001160326	1	1
K18_R1_I1_G1	3	0.01119363	1	1
K21_R23_I55_G4	2	0.0003878465	2	1
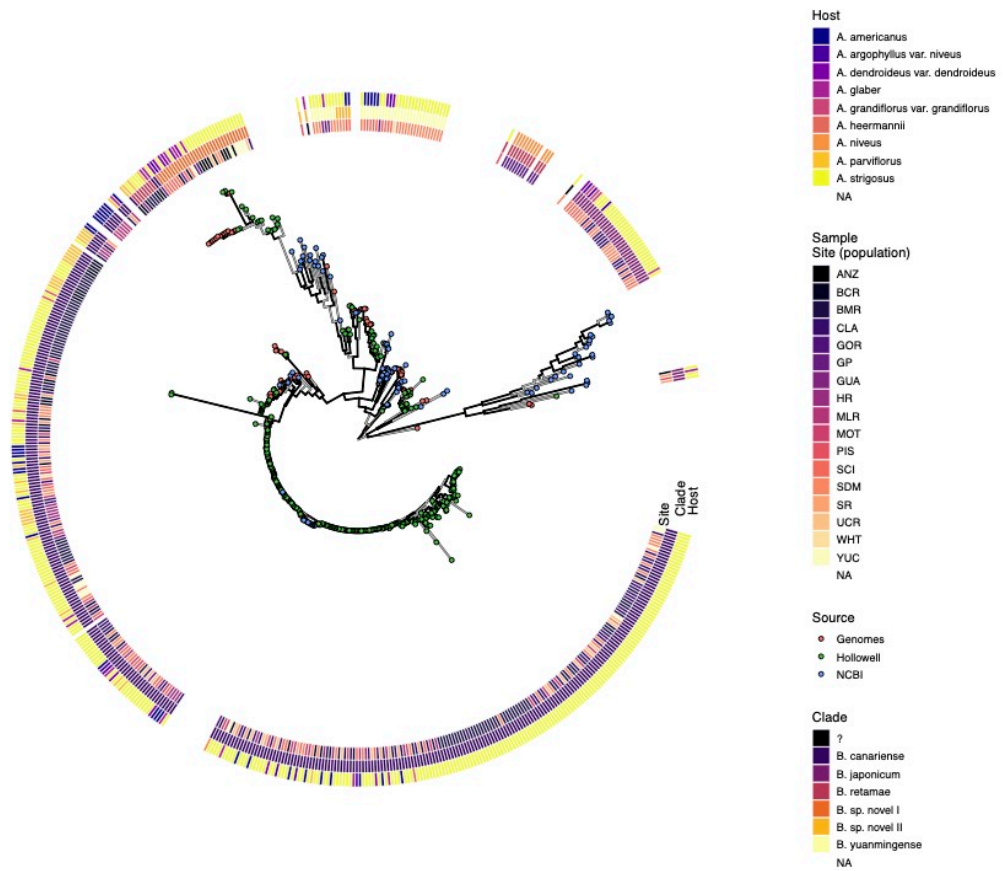
Supplementary Figure 2.1: Phylogenetic tree of all *Bradyrhizobium* species used in this study along with reference *Bradyrhizobium* species genomes downloaded from NCBI. Host, sample sites, major species groups is shown in the histogram.
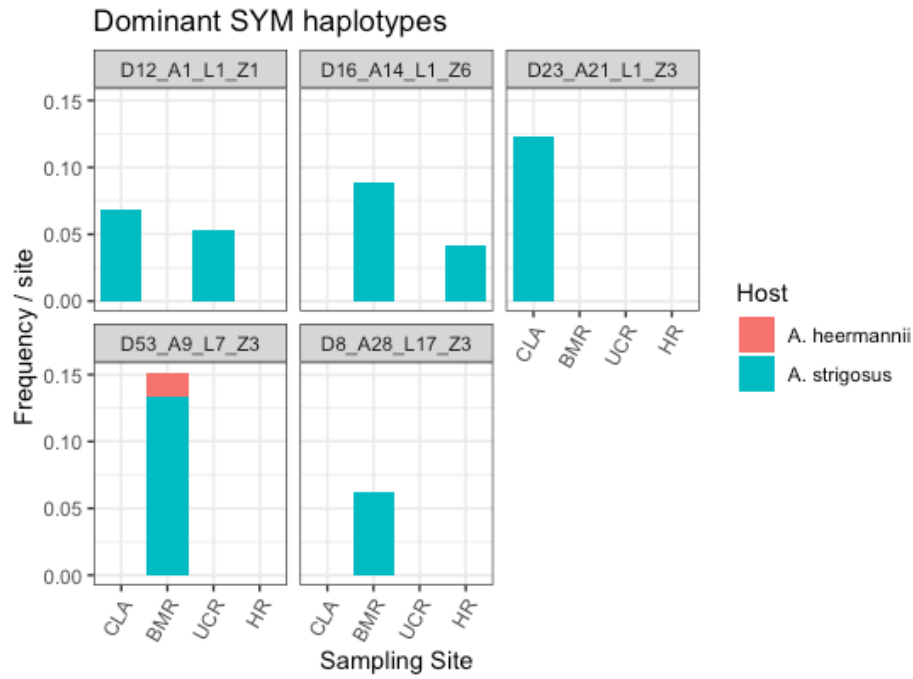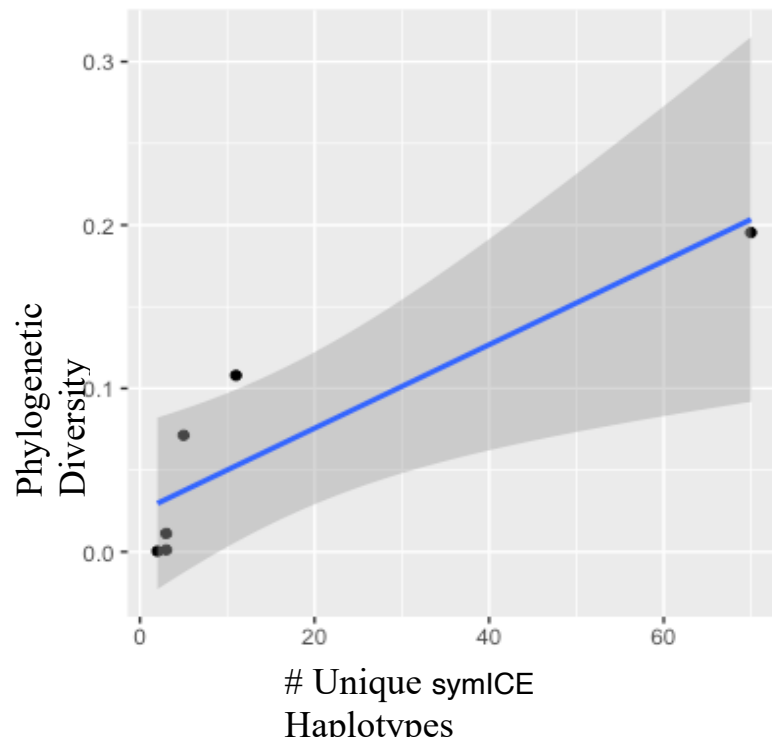
Supplementary figure 2.2: Five dominant symICE haplotype was found out of which two of them were epidemic (D12_A1_L1_Z1 and D16_A14_L1_Z6).
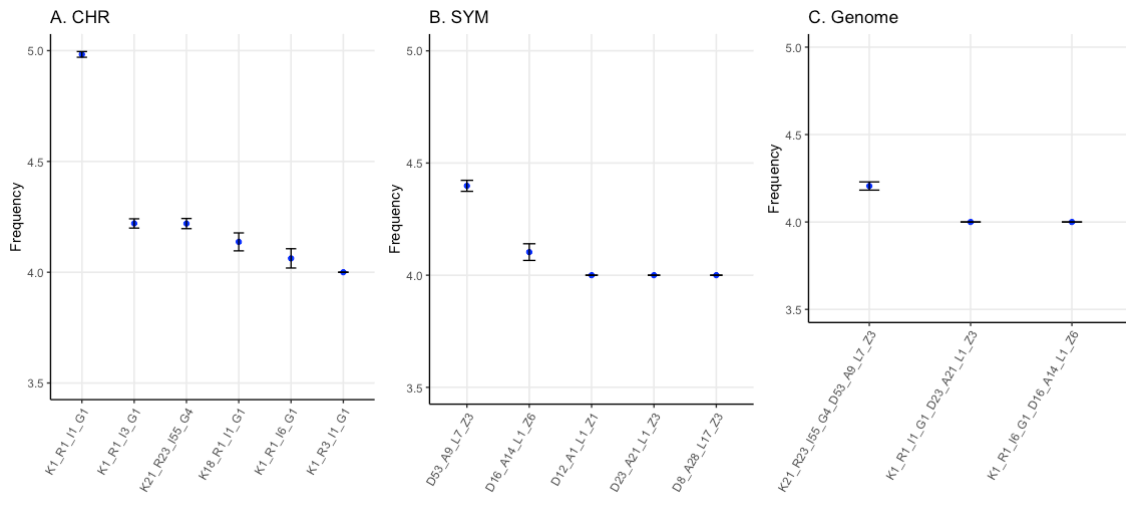
Supplementary Figure 2.3: Bootstrapping for occurrences in sampling sites for each haplotype. The error bar represents variation in haplotype count among all replicates.

Supplementary Figure 2.4: Correlation of the average number of hosts infected by each chromosomal haplotype in each sampling site with mean haplotype abundance by those haplotypes. The range indicates standard errors of mean if the chromosomal haplotype sampled from multiple population.

Supplementary Figure 2.5: Phylogenetic diversity of associated symICE haplotype increase with the higher association of symICE haplotype for each dominant haplotype. Multiple R-squared: 0.7822, Adjusted R-squared: 0.7277, F-statistic: 14.36 n 1 and 4 DF, p-value: 0.01927.
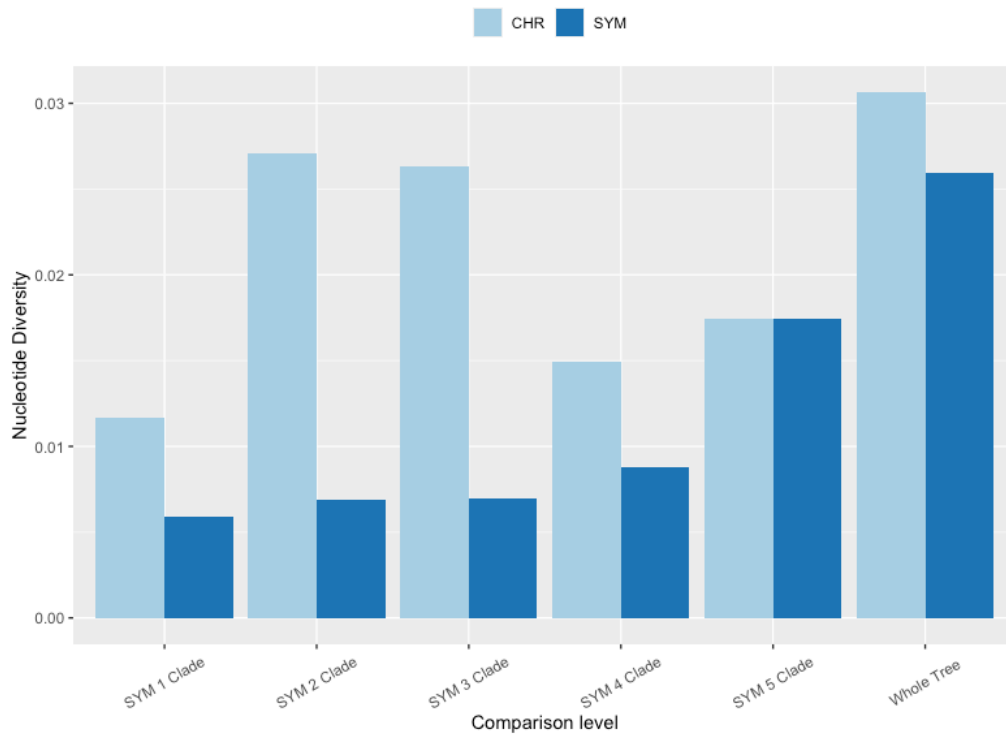
Supplementary Figure 2.6: symICE phylogeny clade members have lower nucleotide diversity in symbiosis loci compared to chromosomal loci except SYM5 clade.
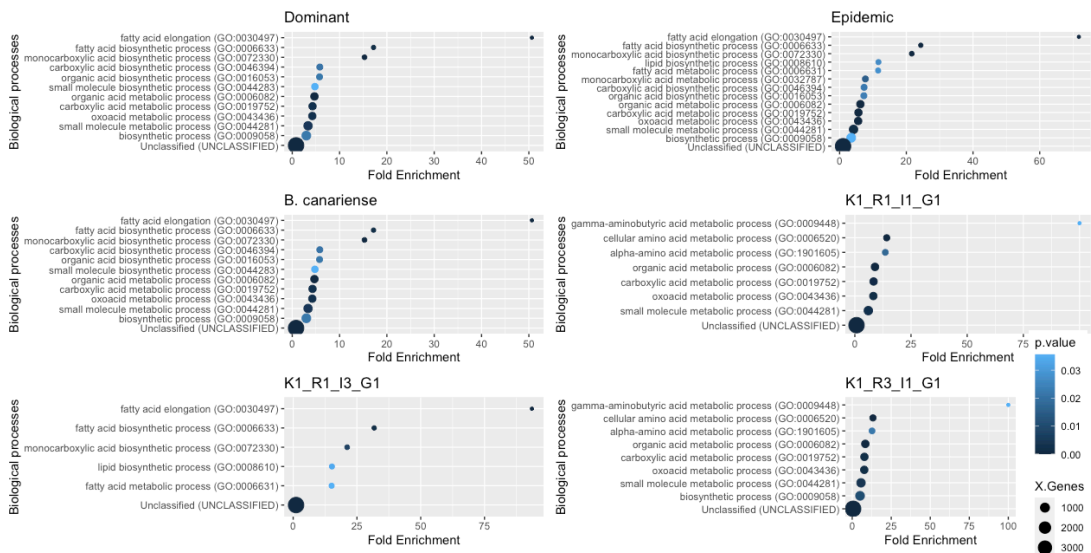
Supplementary figure 2.7: Statistical overrepresentation test based on GO enrichment.

Supplementary Table 2.1. Summary of the *Acmispon* species chromosome numbers, reproductive habit, habitat and the rhizobia lineage isolated from root nodules. The rhizobia lineage assignment was based on the most similar species reported from BLAST searches from whole-genome sequences of the isolates. Data was gathered from the Jepson Manual and chromosome numbers from Grant (1995).

Species	Variety	Habit	Rhizobia Lineage	Life Cycle
A. strigosus		Annual	Bradyrhizobium	Coastal scrub, chaparral, foothills, deserts, roadsides, other disturbed areas
A. niveus		Perennial	Bradyrhizobium	Sandy area
A. grandiflorus	grandiflorus	Perennial	Bradyrhizobium	Dry, open, disturbed sites, chaparral to yellow-pine forest
A. americanus	americanus	Annual	Bradyrhizobium	Coast, chaparral, mountain forest, water courses, roadsides, other disturbed areas
A. parviflorus		Annual	Bradyrhizobium	Coastal bluffs to oak/pine or fir woodland, open or disturbed areas;
A. argophyllus	niveus	Perennial	Bradyrhizobium	Channel Islands; Rocky slopes, dry riverbeds;
A. dendroideus	dendroideus	Perennial	Bradyrhizobium	Channel Islands
A. glaber		Subshrub	Bradyrhizobium	Chaparral, Coastal Sage Scrub, Coastal Strand, Joshua Tree Woodland
A. hermannii	hermannii	Perennial /Annual	Bradyrhizobium	Washes, riverbanks, chaparral

Supplementary Table 2.2: Dominant symICE haplotype's association with other chromosomal haplotypes.

symICE		Associated Chromosome					%Association with Dominant Chromosome
Haplotype	Dominance Type	Haplotype	Occurance	Site	Dominance Type		
D12_A1_L1_Z1	Epidemic	K1_R1_I3_G1	3	CLA	**Epidemic**		50%
		K1_R1_I3_G2	1	CLA	Non-Dominant		
		K1_R12_I1_G1	1	UCR	Non-Dominant		
		K1_R2_I3_G1	1	CLA	Non-Dominant		
D16_A14_L1_Z6	Epidemic	K1_R1_I6_G1	5	BMR	**Dominant**		45%
		K1_R23_I6_G1	2	BMR	Non-Dominant		
		K1_R24_I6_G1	1	BMR	Non-Dominant		
		K1_R25_I6_G1	1	BMR	Non-Dominant		
		K1_R5_I13_G1	1	HR	Non-Dominant		
		K1_R5_I6_G1	1	BMR	Non-Dominant		
D53_A9_L7_Z3	Dominant	K21_R23_I55_G4	9	BMR	**Dominant**		52.94%
		K21_R1_I55_G4	5	BMR	Non-Dominant		
		K21_R70_I55_G4	1	BMR	Non-Dominant		
		K21_R71_I55_G4	1	BMR	Non-Dominant		
		K21_R72_I55_G4	1	BMR	Non-Dominant		
D23_A21_L1_Z3	Dominant	K1_R1_I1_G1	4	CLA	**Epidemic**		44.44%
		K1_R1_I7_G1	1	CLA	Non-Dominant		
		K1_R2_I1_G1	1	CLA	Non-Dominant		
		K1_R6_I1_G1	2	CLA	Non-Dominant		
		K1_R9_I1_G1	1	CLA	Non-Dominant		
D8_A28_L17_Z3	Dominant	K1_R1_I1_G1	2	BMR	**Epidemic**		28.57%
		K1_R1_I1_G5	1	BMR	Non-Dominant		
		K1_R2_I1_G1	1	BMR	Non-Dominant		
		K1_R23_I1_G1	1	BMR	Non-Dominant		
		K1_R35_I1_G1	1	BMR	Non-Dominant		
		K11_R1_I1_G1	1	BMR	Non-Dominant		

Supplementary Table 2.3. ANI and POCP bootstrap statistics. ANI estimated for only the core orthologous CHR bootstrap genes across.

Population	Cromosomal ANI Mean	SE	symICE ANI Mean	SE	Available Genomes	Mean Genome Size (#nucleotide)	Chromosomal Core Gene Number	Chromosomal Core Gene Size (#nucleotide)	Symbiosis Conserved Gene Number	Symbiosis Conserved Gene Size (#nucleotide)	Coverage Core chromosomal	Conserved Symbiosis
Meta Population	94.37021851	0.04069026	96.99197	6.43E-03	224	8,662,254	1851.872	1,810,410	21	20458	20.89999	0.236174
B. canariense	97.92005028	0.02125591	97.56488	1.18E-02	129	8,662,062	1866.5	1,823,858	21	20458	21.0557	0.236179
K1_R1_I1_G1	99.44172237	0.0035523	97.94649	4.18E-02	40	8,576,512	1872.396	1,830,391	21	20458	21.3419	0.238535
K1_R1_I3_G1	99.3321361	0.00372715	98.15144	1.63E-01	9	8,550,880	1873.222	1,831,363	21	20458	21.41725	0.23925
K1_R3_I1_G1	99.36067097	0.00351164	98.71615	3.20E-01	5	8,879,986	1872.5	1,831,305	21	20458	20.62284	0.230383

Supplementary Table 2.4: Tajima's D and nucleotide diversity estimates values.

	Tajima's D	P value (normal)	P value (beta distribution)	Nucleotide Diversity
CHR Concatenated	-2.12966	**0.03319973**	**0.008003677**	0.03066633
SYM Concatenated	-1.764252	0.07768957	**0.04460523**	0.02594106
dnaK	-0.7402718	0.4591351	0.4907761	0.04273532
ITS	-1.8583	0.06312643	**0.03060637**	0.01818584
recA	0.04152628	0.1178178	0.08683944	0.04152628
glnII	-1.475946	0.1399583	0.1121744	0.04983117
nodZ	-1.605686	0.1083429	**0.07638057**	0.01672493
nodL	-0.896202	0.3701449	0.3875837	0.03410141
nodA	-1.406195	0.1596661	0.1353973	0.03023091
nifD	-2.77084	**0.005591183**	**2.54E-06**	0.01354856

Supplementary Table 2.5: Selection analysis

Locus	Gene Under Strong Positive Selection (*P* value of *LRT* $\frac{M2}{M1}$ < 0.05)	Sites under postitive selection (Probability > 95)
CHR	*nag*K	Y137, S228
	*ctr*A	V227
	*trp*S	P2
SYM	*res*A	K123, R299
	*nif*Q	A22, R114, G166, I237, R238, S239, Q241
	blr1754	R173
	*his*C	R29, T32, I69
	*fix*U	G68
	blr1755	A70
	Acyltransferase	L125, H280
	*nod*U	D48, E243

Chapter 3

Effect of Rhizobial variation in nitrogen-fixation on nodule development and senescence

Abstract

Legumes establish a symbiotic relationship with nitrogen-fixing bacteria called rhizobia, resulting in the formation of specialized root nodules that enable the conversion of atmospheric nitrogen into a usable form for the plant. This process involves the secretion of plant compounds to attract compatible rhizobia and the subsequent development of nodules that provide an optimal environment for rhizobial colonization and nitrogen fixation. Within nodules, rhizobia differentiate into bacteroids, specialized cells responsible for nitrogen fixation. However, not all rhizobial genotypes effectively fix nitrogen, leading to reduced benefits for the legume plant. Legumes have evolved mechanisms to limit their support of ineffective rhizobia, but the mechanisms underlying nodule development, maturation, and senescence are not fully understood. This study aims to investigate the impact of nitrogen-fixation variation on nodule development, particularly bacteroid differentiation and the fate of undifferentiated rhizobia during nodule senescence. In this study, five different strains of *Bradyrhizobium* with natural variation in nitrogen fixation was used to study ultrastructural variation of bacteroids in nodules during two different developmental stage. The analysis using transmission electron microscopy

revealed notable variations in the proliferation of rhizobia within fully developed nodules, with beneficial strains outcompeting nonfixing ones. Different strategies were observed among nonfixing strains, including an early increase in the number of symbiosomes and bacteroid density, as well as enhanced nodulation capacity exhibited by CW1. The symbiotic benefit derived from Fix+ bacteroids varied throughout the stages of nodule development, with the highest benefit observed during early development. However, as the nodules aged, senescence led to a decrease in the nitrogen-fixation efficiency of nodule-bacteroids. By studying the distribution and ultrastructure of bacteroids and examining host-control mechanisms at different stages of nodule development, valuable insights can be gained to enhance biological nitrogen fixation and improve agricultural yield.

Introduction

Legume plants can form specialized root nodules as a result of their symbiotic relationship with nitrogen-fixing bacteria known as rhizobia (Suzaki et al., 2015). These nodules provide a microenvironment for the rhizobia to convert atmospheric nitrogen gas into a form that can be used by the plant, in exchange for photosynthates produced by the plant (Garg & Geetanjali, 2009). In the legume-rhizobium symbiosis, nodule formation is initiated when flavonoid compounds secreted by the plant root system attract compatible rhizobia to the root surface. The rhizobia then secrete nodulation factors, which trigger the plant to develop specialized structures called nodules that provide an optimal environment for rhizobial colonization and nitrogen fixation (Walker et al., 2020). Nodule, the symbiotic organ formed by both legume and rhizobia, is the working unit of biological nitrogen fixation (BNF) process that generate predominant natural source of nitrogen in terrestrial habitats (Cleveland et al., 1999).

The legume-rhizobium symbiosis involves a series of molecular and cellular events that result in the formation of nodules, where rhizobia differentiate into bacteroids and fix nitrogen (Brewin, 1991; Q. Wang et al., 2018). During the infection stage, the rhizobia penetrate the root hairs and move towards the root cortex, where they enter the plant cell via an infection thread. The infection thread grows towards the root meristem, and the cortical cells divide to form a nodule primordium. Nodule organogenesis is the process by which the plant forms the nodule structure from the primordium that grows and develops into a mature nodule, which is composed of several regions, including the apical zone,

interzone, and basal zone (Madsen et al., 2010; Verma, 1992). Bacteroid differentiation occurs in the apical zone of the nodule, where the rhizobia differentiate into bacteroids, specialized cells that are capable of fixing atmospheric nitrogen. The bacteroids are surrounded by a peribacteroid membrane that separates them from the plant cytoplasm and provides a specialized environment for nitrogen fixation, also known as symbiosome. Finally, during nodule senescence, the plant degrades the nodule tissue and releases the fixed nitrogen to support plant growth. Ineffective rhizobia that are not able to fix nitrogen may also be expelled from the nodule during senescence, allowing the plant to conserve energy and resources (Serova et al., 2018).

Once inside the plant cell, the rhizobia are enclosed in a membrane-bound compartment called a symbiosome, which provides a protected environment for the rhizobia to differentiate into bacteroids and fix nitrogen. However, not all rhizobial genotypes are effective in fixing nitrogen (Catherine Masson-Boivin & Sachs, 2018; Joel L. Sachs & Simms, 2008). Some may not possess the necessary genetic machinery or may be inhibited by environmental factors, leading to suboptimal levels of nitrogen fixation (Liang et al., 2018; Nandasena et al., 2007). In these cases, the ineffective rhizobia may continue to reside within the nodule and access plant resources without expending energy on nitrogen fixation. This can lead to reduced benefits for the legume plant and a potential waste of resources. To avoid this situation, legumes must have evolved mechanisms to limit their support of ineffective rhizobia, which may involve selection of specific rhizobial strains or other regulatory mechanisms within the nodule (Joel L. Sachs et al., 2018; Schumpp & Deakin, 2010; Ellen L. Simms & Taylor, 2002; Westhoek et al., 2017). By

ensuring effective nitrogen fixation within nodules, legumes can optimize their resource use and reap the maximum benefits from their symbiotic relationship with rhizobia. However, even with the presence of these elegant control mechanisms, ineffective rhizobia is common in nature that drastically reduces legume host benefit (Burghardt & diCenzo, 2023; Gano-Cohen et al., 2020, 2019; Saranraj et al., 2023; Zaw et al., 2021).

While the general process of nodulation is well understood, the mechanisms underlying nodule development, maturation, and senescence remain incompletely characterized (Kazmierczak et al., 2020). Specifically, the literature lacks a comprehensive understanding of how variation in nitrogen fixation affects bacteroid development within nodules (Coba de la Peña et al., 2017). Although we understand the role of sanctioning against the non-fixing strains, but it has not been studied in great details across a variety of non-fixing strains. In this study, we aim to understand how nitrogen-fixation variation of different rhizobia may affect nodule developmental processes during nodule development and maturation, particularly regarding the differentiation of bacteroids and the fate of undifferentiated rhizobia during nodule senescence. The main objectives of this study are to understand: (i) how within-nodule bacteroid distribution and ultrastructure varies in early- and mature-stage of nodulation between set of strains with natural variation in nitrogen fixation, (ii) what are the within-nodule fitness variation of bacteroids and its effect on the legume host, and (iii) how the host-control by sanctioning varies in different stages of nodule development. To answer these questions, we use five *Bradyrhizobium* strains with nitrogen fixation variation to investigate nodule development, ultrastructure-histology, and senescence outcome in early nodule formation and late post-maturation

stage using *Acmispon strigosus* as host. Understanding how nitrogen-fixation variation among rhizobia affects bacteroid distribution during different stages of nodule development is critical to develop strategies to improve BNF and agricultural yield.

Materials and Methods

Rhizobia and plant genotypes – Five *Bradyrhizobium* strains from *B. canariense* species-clad were selected, including two effective strains (#'s 131 & 172) and three ineffective ones (#'s 187, 155, CW1) **(Table 3.1)**. The Fix- strains represents independent origin from a beneficial ancestor and share similar symICE structure with the Fix+ strains used in this study (Weisberg, Rahman, Backus, Tyavanagimatt, Chang, et al., 2022). The host line *A. strigosus* AcS049 was selected for greenhouse experiments, collected from the Bernard Field Station, Claremont, CA (J. U. Regus et al., 2017; John U. Regus et al., 2014).

Inoculation experiment – Seeds were surface sterilized in 5% NaOCl for 3 minutes, rinsed in autoclaved reverse-osmosis water (RO-H_2O) for 7 minutes, nick scarified, and sowed into sterilized SC10R Ray Leach Conetainer pots (diameter 3.81 cm, depth 20.96 cm, volume 164 mL, Steuwe and Sons, Corvallis, Oregon, USA) filled with sterilized calcined clay (Turface® Pro League®, Turface Athletics, Buffalo Grove, Illinois, USA) which offers negligible nutrients. Once true leaves formed, seedlings were moved to a greenhouse and fertilized weekly with 1 mL nitrogen-free Jensen's fertilizer. Fertilization volume was increased weekly by 1 mL until a maximum of 5 mL was reached, which continued until harvest. After 4 days of hardening to greenhouse conditions under 50% shade, plants were inoculated. Rhizobia inocula were prepared by streaking single colonies onto plates of Modified Arabinose Gluconate medium (MAG(J. L. Sachs et al., 2009a)), scraping grown

cells, adjusting cell concentration based on colorimeter readings, and washing cells in RO-H$_2$O. A Klett-Summerson 800-3 photoelectric colorimeter was used (American Laboratory Trading, Inc., San Diego, California, USA) to get turbidimetric readings of cultures which are directly proportional to optical density (OD).

Plants were inoculated with one of five clonal strains or were used as uninoculated controls. Plants received 5 mL cultures at concentrations of 1×10^8 cells/mL, whereas control plants received 5 mL of autoclaved RO-H$_2$O. Each treatment group included 20 plant replicates, and locations of plants in the greenhouse were randomized across all treatments. A total of 120 plants were used during the inoculation experiment (20 replicates × 5 single inoculation treatments, 1 control), Plants were watered daily with 10 minutes of misting. Inoculation occurred on March 29, 2020.

Plant harvest and nodule preservation – To examine the difference between early- and late-stage of nodulation, plants were harvested in two time-points, early harvest (4 weeks post inoculation, i.e., wpi, beginning on May 5, 2020) and late harvest (6-7 wpi, beginning on May 17, 2020). The early harvest is designed to measure nodule traits before sanctions has usually occurred where no significant difference in host benefit is usually not expected. Half of the plant replicates in each treatment were harvested at each timepoint; ~10 plants from each treatment were used to quantify plant biomass, root nodule counts, and nodule biomass. Plants were removed from the soil, shoots and roots were photographed, and nodules were dissected and counted. From photographs, nodule area was measured using ImageJ (v1.50i). Two plants from each treatment group in each harvest were selected for

nodule preservation for RNA extractions. For these plants, nodules were immediately dissected and transferred in liquid nitrogen for quick chilling, followed by further preservation in -80°C freezer. Two more plants from each treatment group in each harvest were selected for nodule preservation for microscopy. Plant root and shoots were dried on 55°C oven and biomass were measured by using a scale. Individual plants were randomly selected for each harvest timepoint and analysis.

Transmission Electron Microscopy – Nodules were fixed overnight in 2.2% v/v glutaraldehyde, 1.8% v/v paraformaldehyde, 50 mM phosphate buffer, pH 7.2 at 4 °C, rinsed three times in 50 mM phosphate buffer (15 min each), postfixed in 2.5 w/v the OsO_4 rinsed twice with dH_2O, and the samples were dehydrated in a graded acetone series to 100% acetone. Nodules were infiltrated in a graded series of Spurr's resin (Spurr's standard mix; EMS, Hatfi eld, Pennsylvania, USA) on a rotator to 100% Spurr's, embedded in fresh Spurr's resin and polymerized at 70°C overnight. Individual nodules were cut out of blocks of Spurr's resin and re-embedded in molds for sectioning. Sections of 50 nm thickness were prepared parallel to the long axis of the parent root with an ultramicrotome (Leica RMC XT-X), collected on copper grids and counterstained with lead citrate (0.4%, 10 min) and uranyl acetate (1% w/v, 20 min) at room temperature in a CO_2 -free environment. Images were acquired with a Tecnai12 TEM (ThermoFisher) and Gatan US1000 HR CCD Camera (Pleasanton, California, USA).

Data analysis – Statistical analyses were carried out using R (version 4.1.3)(R Core Team, 2013). Host growth response to inoculation was estimated by dividing the shoot biomass values of inoculated plants by the shoot biomass values of uninoculated control plants (Wendlandt et al., 2022). Nodule count and total nodule mass were used to estimate rhizobial fitness at the plant level (Pahua et al., 2018). Nodules from plants that were selected for genotyping were used to measure mean nodule weight after drying and then multiplied by total nodule numbers to get nodule biomass estimation. Host investment into symbiosis was quantified as nodule biomass value divided by the shoot biomass value of each inoculated plant (Ortiz-Barbosa et al., 2022). Data transformation was carried out to achieve normality and heteroscedasticity. Linear models were used to investigate variation in host growth response and nodulation traits. Analysis of variance (ANOVA) was carried out with type III sum of square errors to test effects of clonal treatments on the response variables (relative growth, total nodules, mean nodule weight, host investment). Heteroscedasticity of best fit model was tested and for significant differences was followed by Tukey's HSD test.

Histological and ultrastructural data from TEM were analyzed using ImageJ software (Schneider et al., 2012). For each nodule, bacterial population density per area, bacteroid area per nodule area, number of symbiosome were measured for both early and late harvested nodules. Also, variation in histological and ultrastructural features was analyzed among strain treatments and among times using ANOVA and t-test.

Results

We first evaluated the effects of clonal inoculations on hosts to cross-validate prior results (**Table 3.1**). The five tested strains caused host responses that closely matched previous results (Gano-Cohen et al., 2020; J. L. Sachs, Ehinger, et al., 2010) **(Fig. 3.1, Table 3.2)**. The Fix+ strains, 131 and 172, elicited significant mean relative host growth responses, indicating that inoculated plants were larger than uninoculated control plants in the second harvest (**Table 3.2**). The Fix- strains 155 and 187 were ineffective in both harvests, as relative host growth was not significantly higher than 1($P > 0.05$; **Figure 3.1**, **Table 3.2**). Strain CW1, which was previously characterized as Fix-, provided a low but significant level of growth benefit in the late harvest (**Figure 3.1**, **Table 3.2**). Relative host growth response varied significantly among the clonal strain treatments (**Table 3.3,** $F_{(4,74)} = 12.92$, $P = 5 \times 10^{-08}$). For nodulation, we have found that significant difference in nodulation among treatments (**Figure 3.2, Table 3.3**, $F_{(4,84)} = 4.046$, $P = 4.76 \times 10^{-03}$).

Nodulation increased for all treatments in late harvest, compared to early (**Figure 3.2**). Strain CW1 shows highest nodulation propensity in early harvest to the late harvest. whereas strain 172, although characterized as Fix+ strain, formed lowest number of nodules on the host in the early harvest. For nodule area, significant difference was observed between treatments t ($F_{(4,41)} = 9.71$, $P = 1.2805 \times 10^{-05}$ **Figure 3.3, Table 3.2**). Fix+ strains 131 and 172 has 2-fold more nodule area compared to the rest (**Figure 3.3**).

Although strain CW1 provided significant effect to host according to our data, but its mean nodule area is comparable to that of other Fix- strains (**Figure 3.3**).

We also analyzed ultramicroscopic features inside nodules from TEM data. Symbiosome was only detected during the early harvest, which indicate that during the late harvest nodule bacteroid started senescence **(Figure 3.4-5)**. Number of bacteroids per 100 μm^2 area has observed a significant reduction in the late harvest ($F_{(4,217)}$ = 2.801, P = 2.69×10^{-02}, **Figure 3.6, Table 3.4**). Fix+ strains tend to have relatively higher number of bacteroids compared to Fix- strains based on HSD post-hoc test. Similarly, mean bacteroid size has also observed a significant differences between harvest batch with an interaction effect between batch and treatment, but effect of treatment on bacteroid size was not significant ($F_{(4,211)}$ = 1.236, P = 0.297, **Table 3.4**). When density was calculated for bacteroid size normalizing nodule slice area, no significant density of bacteroid area was observed ($F_{(4,216)}$ = 1.405, P = 0.233**, Table 3.4**). There was no significant difference between number of symbiosomes, symbiosome size in symbiosimes between treatment during early harvest, however Fix- strains has more bacteroids apart from CW1 (**Figure 3.7**). The normalized density decreased in the late harvest for Fix+ treatments to comparable level of Fix- strains.

To estimate how much benefit the host getting from bacteroid during early or late harvest, the host benefit per nodule bacteroid was estimated and was normalized by number of nodules. Effective strains have a higher number of benefit from bacteroid level when

compared to ineffective strains, and the benefit is highest during the early harvest **(Figure 3.8)**.

Discussion

The findings of this study provide insights into the effects of different clonal inoculations on host plants. It shows that the nodulation increases with time, however the host benefit depends on the Fix trait. The results validate previous research and confirm that Fix+ strains 131 and 172 are effective in eliciting growth responses in hosts, while Fix- strains 155 and 187 are ineffective. This study also corroborates previous findings that Fix+ strains produces larger nodules. These results confirms that the presence of nitrogen-fixing bacteria in the root nodules of host plants is crucial for enhancing plant growth and development.

Furthermore, the study found that nodulation increased for all treatments in the late harvest, indicating that the host plants were able to sustain a longer-term symbiotic relationship with the bacteria, regardless Fix trait. However, one Fix- strain CW1 formed significantly more nodules compared to other Fix+ strains without providing high benefit to the host. This finding is important, as it suggests that although the host plants were able to maintain a stable and effective symbiotic relationship with the Fix+ strains over an extended period which is essential for the long-term health and growth of the host plants, however some Fix- strains can induce more nodules without providing any benefit to the host with increase investment to nodulation, thus indicate potential gap of host's partner choice and sanction mechanism.

Different strategies of Fix- strains was observed to optimize their *in planta* fitness. An early increase of number of symbiosome and bacteroid density has been observed in the Fix- strains compared to the Fix+ strains, which indicates an increase in their population

size. However, in the later stage, the bacteroid density reduced which implies host control. On the other hand, CW1 shows similar bacteroid density and symbiosome number, however it has a hypernodulator phenotype (i.e. forming an increased number of nodules), which could be another strategies to increase *in planta* fitness.

The study also found that there were significant differences in nodule area between treatments, with Fix+ strains having a 2-fold higher nodule area compared to the rest. This finding is consistent with previous research that has shown that plant can sanction nodule tissues formed by ineffective strains. The study also found that strain CW1, which is characterized as Fix-, provided a low but significant level of growth benefit in the late harvest, suggesting that even bacteria that are not classified as nitrogen-fixing can have a marginally positive effect on host plant growth.

The ultramicroscopic analysis of the root nodules provided further insights into the symbiotic relationship between the host plants and the bacteria. The finding that symbiosome was only detected during the early harvest and the significant reduction in the number of bacteroids in the late harvest indicates that the bacteria were starting to senesce.

Importantly, this study shows that plants are getting most benefit from Fix+ bacteroid during early nodule development stage. With time, nodulation increase, however the level of benefit from Fix+ bacteroid level decreases. This reflects senescence and lack of symbiosome-environment decreases nitrogen fixation efficiency of nodule bacteroids.

Overall, the findings of this study have important implications for understanding the symbiotic relationship between nitrogen-fixing bacteria and host plants. The study highlights the importance of maintaining a stable and effective symbiotic relationship over

an extended period for the long-term health and growth of the host plants. Further research is needed to explore the mechanisms underlying the symbiotic relationship between the bacteria and host plants and to identify new strains of nitrogen-fixing bacteria that are more effective in promoting plant growth and development.
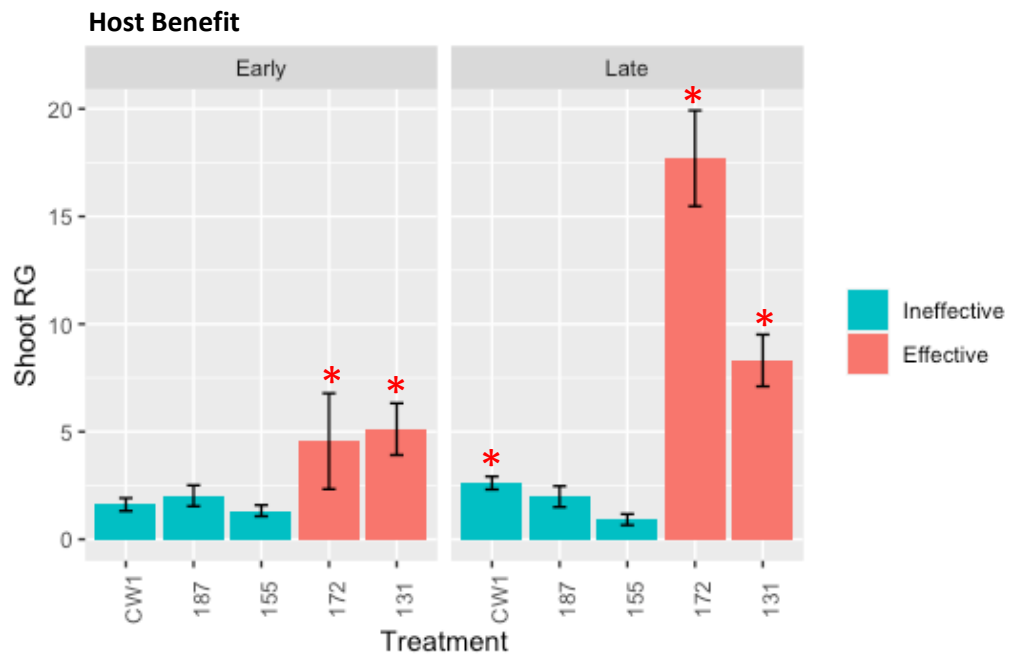
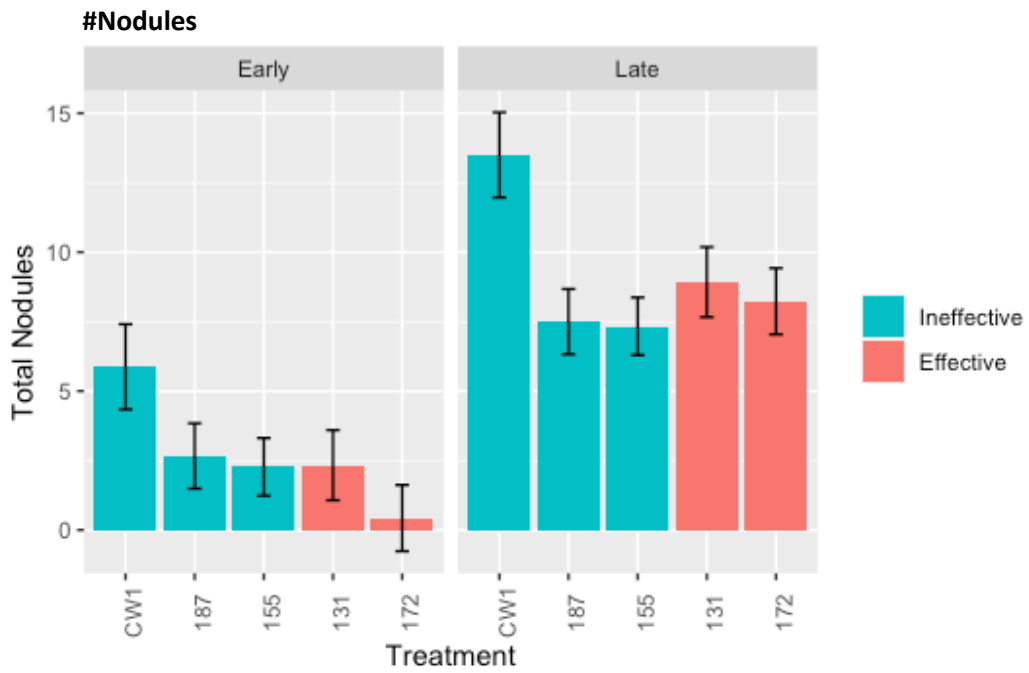**Figure 3.1:** Shoot RG in two harvest batches.
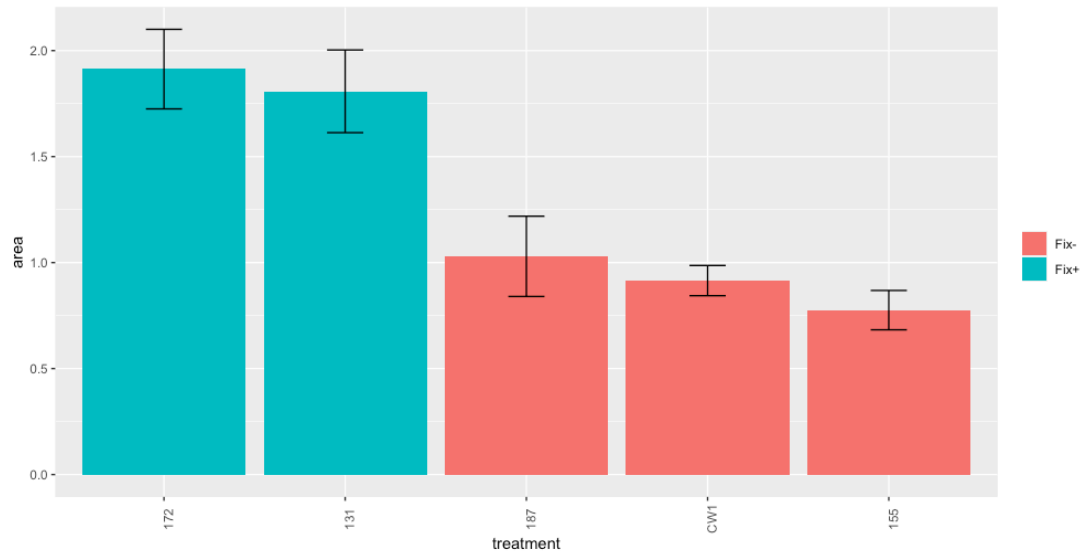
Figure 3.2: Total nodules for two harvest batches.

Figure 3.3: Nodule area.
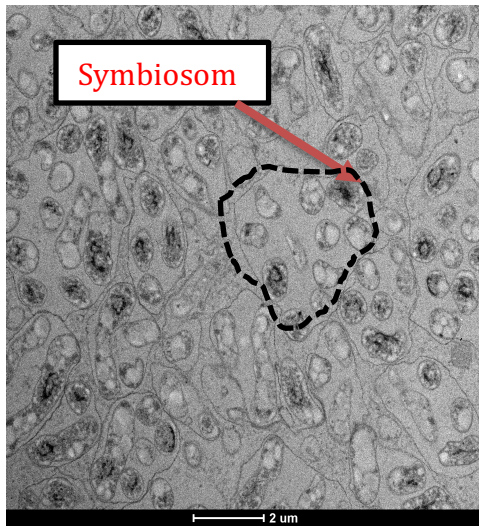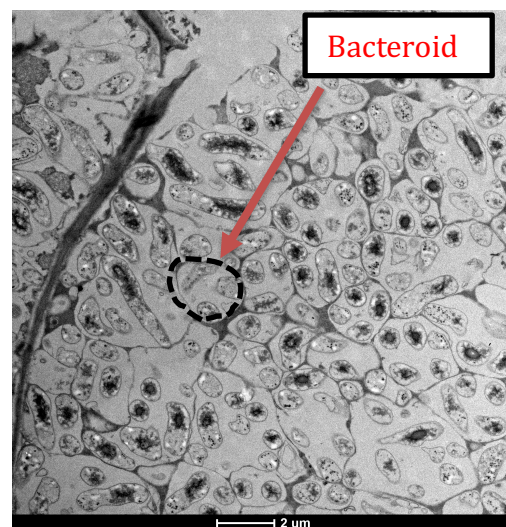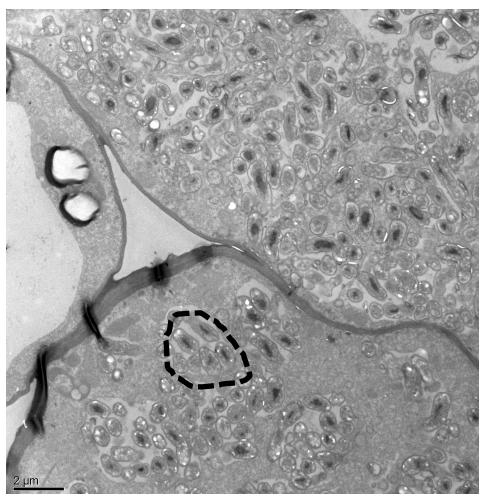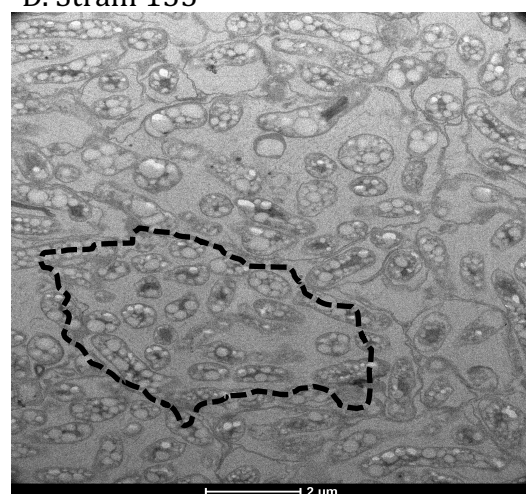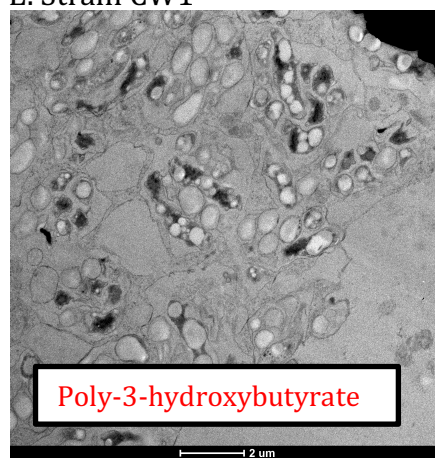
Figure 3.4: TEM of nodule bacteroid (2 μm) during early harvest.
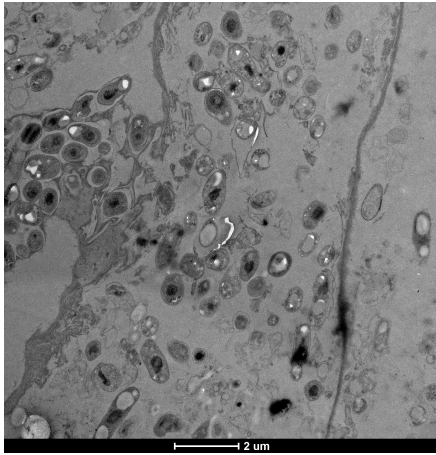
A. Strain 187 (Fix-)
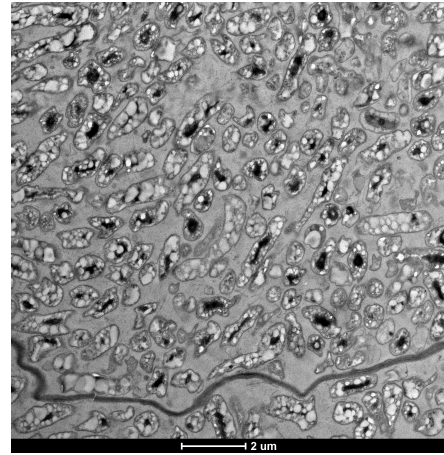
B. Strain 155 (Fix-)
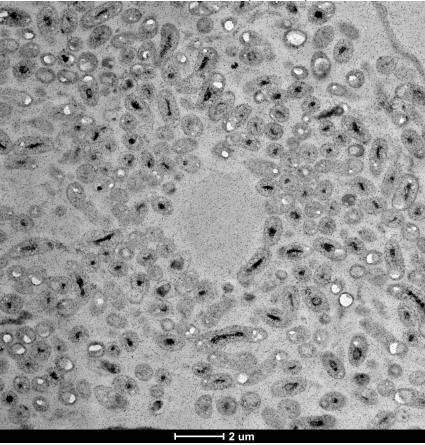
C. Strain CW1 (Fix-)
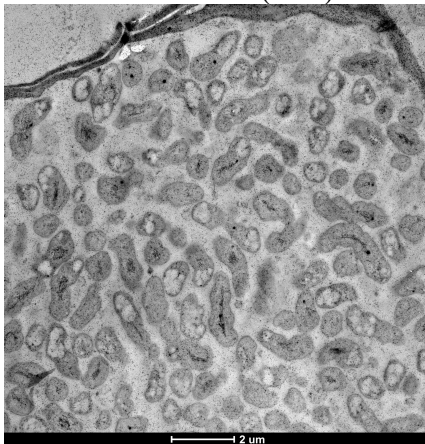
D. Strain 172 (Fix +)
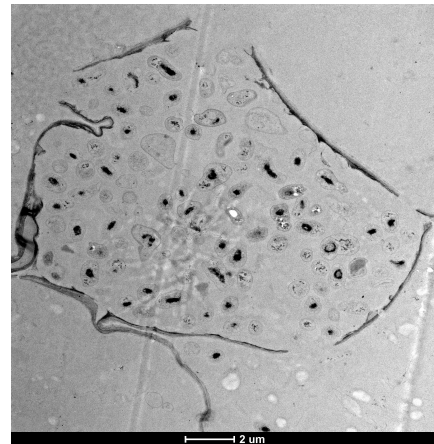
E. Strain 172 (Fix +)

Figure 3.5: TEM of nodule bacteroid (2 μm) during late harvest.
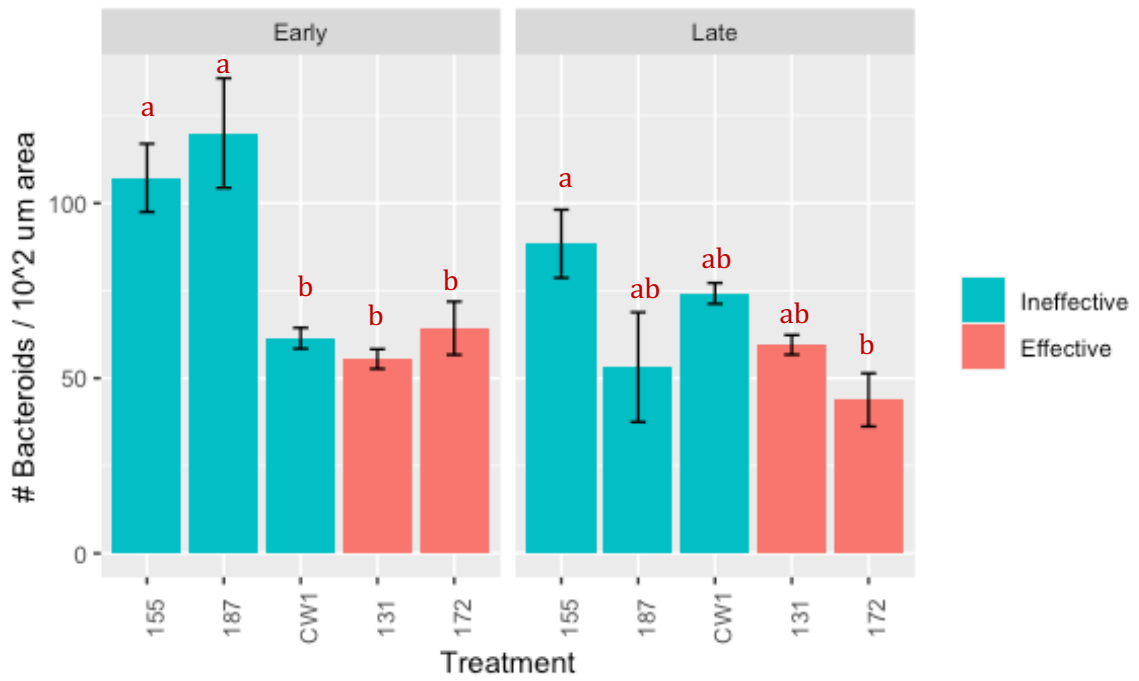
Figure 3.6: Number of bacteroid per 100 um area.

A. Number of symbiosomes

B. Symbiosome size

C. # Bacteroids in symbiosome
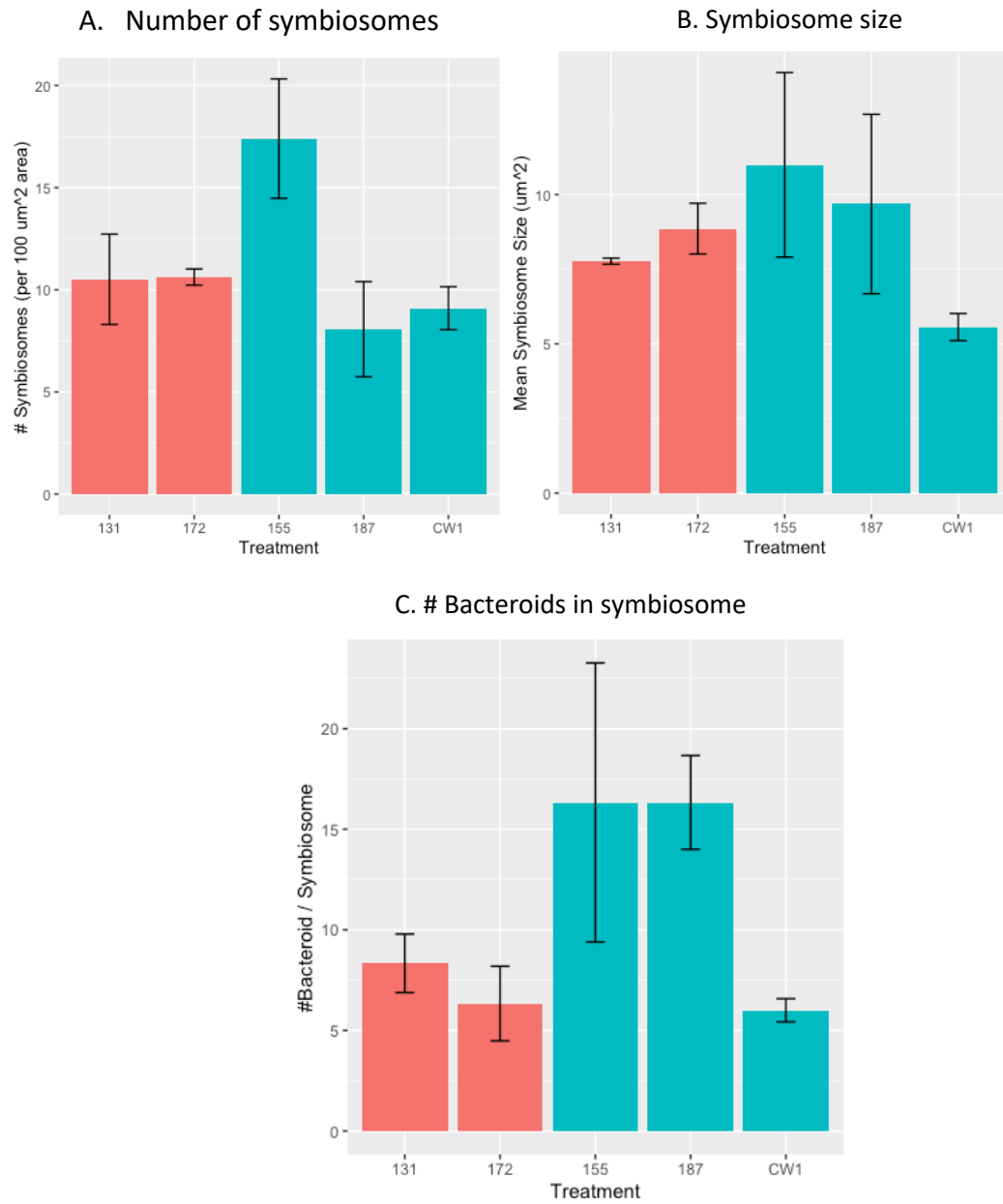
Figure 3.7: Fix- strains, except CW1, has more bacteroids within symbiosome.
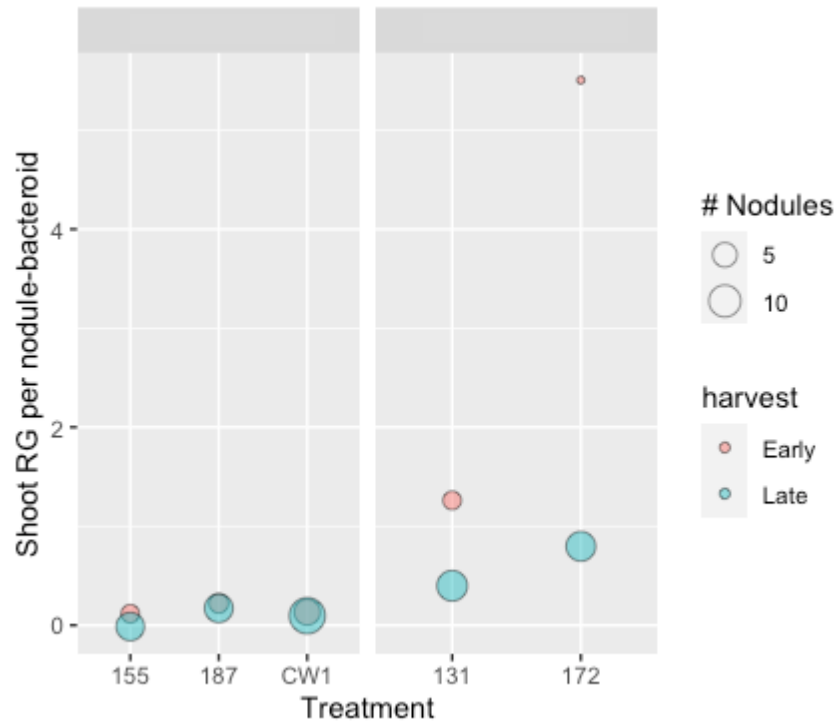
Figure 3.8: Plants getting most benefit from early nodule-bacteroids.

Table 3.1: Isolates used in this study. All from *B. canariense* clade.

Strain Code	Strain ID	Fix Trait	Type III Secretion System	Presence of Symbiosis ICE
131	13LoS28_1	Fix+	T3SS+	Present
172	11LoS18_3	Fix+	T3SS-	Present
187	11LoS7_1	Fix-	T3SS+	Lost SI on chromosome but contain on plasmid
CW1	14LoS3.1	Fix-	T3SS+	Similar ICE as the beneficial
155	11LoS34_2	Fix-	T3SS+	Similar ICE as the beneficial

Table 3.2: Testing benefit to the host by each strain in early or late harvest batch

Harvest Batch	Treatment	t	df	p-value	Category
Early	131	4.4137	5	0.003466	**Effective**
	172	1.0472	6	0.1677	Ineffective
	155	-0.34242	9	0.6301	Ineffective
	187	1.9237	5	0.0562	Ineffective
	CW1	1.1411	5	0.1528	Ineffective
Late	131	10.606	11	2.04×10^{-7}	**Effective**
	172	12.853	12	1.12×10^{-8}	**Effective**
	155	-0.71015	2	0.7244	Ineffective
	187	1.0371	9	0.1634	Ineffective
	CW1	4.7559	5	0.002539	**Effective**

Table 3.3: Estimation of different linear-models testing effect of treatment and batch on harvest variables including shootRG, number of nodules, and area of nodules.

Response Variables	Model Variables	ANOVA Variable ~ Treatment		
		df	F value	P value
Shoot RG	Treatment	4	12.92	5.00E-08
	Residuals	74		
Nodule	Treatment	4	4.046	4.76E-03
	Residuals	84		
Nodule	Treatment	4	9.71	1.28E-05
	Residuals	41		

Table 3.4: Estimation of different linear-models testing effect of treatment and batch on nodule bacteroid ultrastructural variables including bacteroid number, bacteroid size, and bacteroid area density.

Response Variables	Model Variables	ANOVA		
		Variable ~ Treatment		
		df	F value	P value
Bacteroid	Treatment	4	2.801	***2.69E-02***
	Residuals	217		
Bacteroid	Treatment	4	1.236	2.97E-01
	Residuals	216		
Bacteroid	Treatment	4	1.405	2.33E-01
	Residuals	216		

Conclusion

An unresolved question is to understand the role of competition among beneficial microbes when they colonize their host. The goal of chapter 2 was to characterize variation in competitive ability among rhizobial strains and to examine how competition among rhizobia genotypes shapes host investment and mutualistic benefit. For this an experiment was designed using a focal set of eight *Bradyrhizobium* strains that range from beneficial to ineffective in symbiotic nitrogen fixation in a full-factorial co-inoculation experiment on *Acmispon strigosus*. Using Illumina amplicon sequencing, more than 1,100 nodules were genotyped from co-inoculated plants to determine the ratios of occupying strains. Null models were developed to use single-inoculation treatment data to predict and test co-inoculation treatments' effects on host nodulation and growth. The results from chapter 2 reveal that there is a linear dominance hierarchy of rhizobia which is consistent by strain. It also shows co-inoculation treatments significantly diminished host benefits relative to expectations based on clonal controls. This reduction in host benefit is consistent with competitive interference of strains which may reduce nodulation and plant growth. Genotyping data from nodules demonstrates that legumes discriminate against nonfixing rhizobia within nodules (i.e., ineffective strains). However, interactions among rhizobia that vary quantitatively in benefits can favor strains that fix low nitrogen levels. A minimally beneficial rhizobia strain exhibited the highest level of competitive ability.

We also have a limited understanding of how rhizobia strains vary in fitness in natural landscapes, which includes a community of potential hosts. Nine *Acmispon* species were used to study a rhizobial metapopulation consisting of 507 *Bradyrhizobium* sp. strains sampled from 17 populations across an ~800 km transect in California. Chromosomal and symbiosis haplotypes were defined by using four chromosomal genes and four genes encoded on an Integrative Conjugative Element that encodes symbiosis functions (i.e., symICE). Analyses reveal the presence of six highly abundant chromosomal haplotypes in the metapopulation, three of which have a large epidemiological spread. Whole genome analyses show that dominant haplotypes exhibit near clonality genome-wide. Cophylogenetic analyses reveal that dominant chromosomal haplotypes acquire a diverse set of symbiotic haplotypes in the metapopulation. Phylogenetic structure by host and sampling sites is observed in symbiosis loci. Correlation between host range and strain abundance was also found. The results suggest that acquisition of local and host-adapted symICEs enables epidemic haplotypes to spread through the meta-population.

Rhizobia populations include beneficial and nonfixing strains, despite the latter having the necessary genes to provide fixed nitrogen to hosts. A combination of harvest and histological data were used to understand the developmental processes that differ among nitrogen fixing and nonfixing rhizobia. A greenhouse experiment examined five *Bradyrhizobium* strains that vary in their capacity to achieve nitrogen fixation on their host plant, *Acmispon strigosus*. The transmission electron microscopic analysis shows significant differences in the proliferation of rhizobia, favoring the beneficial strains over

the nonfixing ones in the fully developed nodules. Multiple strategies of ineffective strains were observed, in one case early increase of symbiosome population and bacteroid density was seen, and on the other hand one ineffective strain showed increased nodulation capacity. Symbiotic benefit from nitrogen fixing bacteroids varies during nodule development stage, where bacteria in the early nodule development have the highest benefit. Overtime, nodule senescence decreases nitrogen fixation efficiency of nodule-bacteroids.

Overall, the dissertation provides insights into rhizobial competition, the relationship between competitiveness and nitrogen fixation, the distribution of symbiotic haplotypes in metapopulations, and the developmental processes in nitrogen fixing and nonfixing rhizobia. These findings contribute to a better understanding of the complex interactions between rhizobia and their host plants and have implications for optimizing symbiotic nitrogen fixation and agricultural practices.

Reference:

Amarger, N. (1981). Competition for nodule formation between effective and ineffective strains of Rhizobium meliloti. *Soil Biology & Biochemistry*, *13*(6), 475–480.

Andrews, M., De Meyer, S., James, E. K., Stępkowski, T., Hodge, S., Simon, M. F., & Young, J. P. W. (2018). Horizontal transfer of symbiosis genes within and between rhizobial genera: Occurrence and importance. *Genes*, *9*(7). https://doi.org/10.3390/genes9070321

Bailly, X., Olivieri, I., de Mita, S., Cleyet-Marel, J.-C., & Béna, G. (2006). Recombination and selection shape the molecular diversity pattern of nitrogen-fixing Sinorhizobium sp. associated to Medicago. *Molecular Ecology*, *15*(10), 2719–2734.

Bamba, M., Aoki, S., Kajita, T., Setoguchi, H., Watano, Y., Sato, S., & Tsuchimatsu, T. (2019). Exploring Genetic Diversity and Signatures of Horizontal Gene Transfer in Nodule Bacteria Associated with Lotus japonicus in Natural Environments. *Molecular Plant-Microbe Interactions: MPMI*, *32*(9), 1110–1120.

Bamba, M., Aoki, S., Kajita, T., Setoguchi, H., Watano, Y., Sato, S., & Tsuchimatsu, T. (2020). Massive rhizobial genomic variation associated with partner quality in Lotus-Mesorhizobium symbiosis. *FEMS Microbiology Ecology*, *96*(12). https://doi.org/10.1093/femsec/fiaa202

Bartlett, J. W. (2018). Covariate adjustment and estimation of mean response in randomised trials. *Pharmaceutical Statistics*. https://doi.org/10.1002/pst.1880

Batstone, R. T., Burghardt, L. T., & Heath, K. D. (2022). Phenotypic and genomic signatures of interspecies cooperation and conflict in naturally occurring isolates of a model plant symbiont. *Proceedings. Biological Sciences*, *289*(1978), 20220477.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society*, *57*(1), 289–300.

Bentley, B. L. (1987). Nitrogen Fixation by Epiphylls in a Tropical Rainforest. *Annals of the Missouri Botanical Garden. Missouri Botanical Garden*, *74*(2), 234.

Bernhard, A. (2010). The nitrogen cycle: processes. *Players, and Human*, *3*, 10–25.

Blackmon, H., & Adams, R. H. (2015). *EvobiR: Tools for comparative analyses and teaching evolutionary biology*. Zenodo. https://doi.org/10.5281/ZENODO.30938

Bonomi, H. R., Posadas, D. M., Paris, G., Carrica, M. del C., Frederickson, M., Pietrasanta, L. I., Bogomolni, R. A., Zorreguieta, A., & Goldbaum, F. A. (2012). Light regulates attachment, exopolysaccharide production, and nodulation in Rhizobium leguminosarum through a LOV-histidine kinase photoreceptor. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(30), 12135–12140.

Bourion, V., Heulin-Gotty, K., Aubert, V., Tisseyre, P., Chabert-Martinello, M., Pervent, M., Delaitre, C., Vile, D., Siol, M., Duc, G., Brunel, B., Burstin, J., & Lepetit, M. (2017). Co-inoculation of a Pea Core-Collection with Diverse Rhizobial Strains Shows Competitiveness for Nodulation and Efficiency of Nitrogen Fixation Are Distinct traits in the Interaction. *Frontiers in Plant Science*, *8*, 2249.

Brewin, N. J. (1991). Development of the legume root nodule. *Annual Review of Cell Biology*, *7*(1), 191–226.

Bright, M., & Bulgheresi, S. (2010). A complex journey: transmission of microbial symbionts. *Nature Reviews Microbiology*, *8*(3), 218–230.

Brynildsrud, O., Bohlin, J., Scheffer, L., & Eldholm, V. (2016). Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biology*, *17*(1), 238.

Burdon, J. J., Gibson, A. H., Searle, S. D., Woods, M. J., & Brockwell, J. (1999). Variation in the effectiveness of symbiotic associations between native rhizobia and temperate AustralianAcacia: within-species interactions. *The Journal of Applied Ecology*, *36*(3), 398–408.

Burghardt, L. T., & diCenzo, G. C. (2023). The evolutionary ecology of rhizobia: multiple facets of competition before, during, and after symbiosis with legumes. *Current Opinion in Microbiology*, *72*(102281), 102281.

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. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(10), 2425–2430.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*(1), 421.

Cavassim, M. I. A., Moeskjær, S., Moslemi, C., Fields, B., Bachmann, A., Vilhjálmsson, B. J., Schierup, M. H., W Young, J. P., & Andersen, S. U. (2020). Symbiosis genes show a unique pattern of introgression and selection within a Rhizobium

leguminosarum species complex. *Microbial Genomics*, *6*(4). https://doi.org/10.1099/mgen.0.000351

Checcucci, A., Azzarello, E., Bazzicalupo, M., Galardini, M., Lagomarsino, A., Mancuso, S., Marti, L., Marzano, M. C., Mocali, S., Squartini, A., Zanardo, M., & Mengoni, A. (2016). Mixed Nodule Infection in Sinorhizobium meliloti–Medicago sativa Symbiosis Suggest the Presence of Cheating Behavior. *Frontiers in Plant Science*, *7*, 835.

Chen, J. G., Crooks, R. M., Seefeldt, L. C., Bren, K. L., Bullock, R. M., Darensbourg, M. Y., Holland, P. L., Hoffman, B., Janik, M. J., Jones, A. K., Kanatzidis, M. G., King, P., Lancaster, K. M., Lymar, S. V., Pfromm, P., Schneider, W. F., & Schrock, R. R. (2018). Beyond fossil fuel-driven nitrogen transformations. *Science (New York, N.Y.)*, *360*(6391). https://doi.org/10.1126/science.aar6611

Chen, L., Figueredo, A., Villani, H., Michajluk, J., & Hungria, M. (2002). Diversity and symbiotic effectiveness of rhizobia isolated from field-grown soybean nodules in Paraguay. *Biology and Fertility of Soils*, *35*(6), 448–457.

Chernomor, O., von Haeseler, A., & Minh, B. Q. (2016). Terrace aware data structure for phylogenomic inference from supermatrices. *Systematic Biology*, *65*(6), 997–1008.

Choudhury, A. T. M. A., & Kennedy, I. R. (2005). Nitrogen Fertilizer Losses from Rice Soils and Control of Environmental Pollution Problems. *Communications in Soil Science and Plant Analysis*, *36*(11–12), 1625–1639.

Cleveland, C. C., Townsend, A. R., Schimel, D. S., Fisher, H., Howarth, R. W., Hedin, L. O., Perakis, S. S., Latty, E. F., Von Fischer, J. C., Elseroad, A., & Wasson, M. F. (1999). Global patterns of terrestrial biological nitrogen (N2) fixation in natural ecosystems. In *Global Biogeochemical Cycles* (Vol. 13, Issue 2, pp. 623–645). https://doi.org/10.1029/1999gb900014

Coba de la Peña, T., Fedorova, E., Pueyo, J. J., & Lucas, M. M. (2017). The symbiosome: Legume and rhizobia co-evolution toward a nitrogen-fixing organelle? *Frontiers in Plant Science*, *8*, 2229.

Collins, M. T., Thies, J. E., & Abbott, L. K. (2002). Diversity and symbiotic effectiveness of Rhizobium leguminosarum bv. trifolii isolates from pasture soils in south-western Australia. *Soil Research*, *40*(8), 1319–1329.

Contreras-Moreira, B., & Vinuesa, P. (2013). GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Applied and Environmental Microbiology*, *79*(24), 7696–7701.

Costa, T. R. D., Harb, L., Khara, P., Zeng, L., Hu, B., & Christie, P. J. (2021). Type IV secretion systems: Advances in structure, function, and activation. *Molecular Microbiology*, *115*(3), 436–452.

Cruaud, P., Rasplus, J.-Y., Rodriguez, L. J., & Cruaud, A. (2017). High-throughput sequencing of multiple amplicons for barcoding and integrative taxonomy. In *Scientific Reports* (Vol. 7, Issue 1). https://doi.org/10.1038/srep41948

Denison, R. F. (2000). Legume Sanctions and the Evolution of Symbiotic Cooperation by Rhizobia. *The American Naturalist*, *156*(6), 567–576.

Denison, R. F., & Kiers, E. T. (2004). Lifestyle alternatives for rhizobia: mutualism, parasitism, and forgoing symbiosis. *FEMS Microbiology Letters*, *237*(2), 187–193.

Denison, R. F., & Kiers, E. T. (2011). Life histories of symbiotic rhizobia and mycorrhizal fungi. *Current Biology: CB*, *21*(18), R775-85.

Denton, M. D., Coventry, D. R., Bellotti, W. D., & Howieson, J. G. (2000). Distribution, abundance and symbiotic effectiveness of Rhizobium leguminosarum bv. trifolii from alkaline pasture soils in South Australia. *Australian Journal of Experimental Agriculture*, *40*(1), 25–35.

diCenzo, G. C., & Finan, T. M. (2017). The divided bacterial genome: Structure, function, and evolution. *Microbiology and Molecular Biology Reviews: MMBR*, *81*(3). https://doi.org/10.1128/mmbr.00019-17

Dunn, M. F. (2015). Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Critical Reviews in Microbiology*, *41*(4), 411–451.

Ehinger, M., Mohr, T. J., Starcevich, J. B., Sachs, J. L., Porter, S. S., & Simms, E. L. (2014). Specialization-generalization trade-off in a Bradyrhizobium symbiosis with wild legume hosts. *BMC Ecology*, *14*, 8.

Epstein, B., Branca, A., Mudge, J., Bharti, A. K., Briskine, R., Farmer, A. D., Sugawara, M., Young, N. D., Sadowsky, M. J., & Tiffin, P. (2012). Population genomics of the facultatively mutualistic bacteria Sinorhizobium meliloti and S. medicae. *PLoS Genetics*, *8*(8), e1002868.

Epstein, B., Burghardt, L. T., Heath, K. D., Grillo, M. A., Kostanecki, A., Hämälä, T., Young, N. D., & Tiffin, P. (2022). Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis. *Molecular Ecology*. https://doi.org/10.1111/mec.16602

Epstein, B., & Tiffin, P. (2021). Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. *Proceedings. Biological Sciences*, *288*(1942), 20201804.

Erisman, J. W., Sutton, M. A., Galloway, J., Klimont, Z., & Winiwarter, W. (2008). How a century of ammonia synthesis changed the world. *Nature Geoscience*, *1*, 636.

Erisman, J., Willem, J. N., Galloway, N. B., Dise, M. A., Sutton, A., Bleeker, B., Grizzetti, A. M., & Leach, W. D. (2015). Nitrogen: too much of a vital resource: Science Brief. *Science Brief. WWF Netherlands*.

Fadrosh, D. W., Ma, B., Gajer, P., Sengamalay, N., Ott, S., Brotman, R. M., & Ravel, J. (2014). An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. *Microbiome*, *2*(1), 6.

Faith, D. P. (1992). Conservation evaluation and phylogenetic diversity. *Biological Conservation*, *61*(1), 1–10.

Fauvart, M., & Michiels, J. (2008). Rhizobial secreted proteins as determinants of host specificity in the rhizobiumâ legume symbiosis. *FEMS Microbiology Letters*, *285*(1), 1–9.

Foster, K. R., & Wenseleers, T. (2006). A general model for the evolution of mutualisms. *Journal of Evolutionary Biology*, *19*(4), 1283–1293.

Foster, Kevin R., Schluter, J., Coyte, K. Z., & Rakoff-Nahoum, S. (2017). The evolution of the host microbiome as an ecosystem on a leash. *Nature*, *548*(7665), 43–51.

Foyer, C. H., Nguyen, H., & Lam, H.-M. (2019). Legumes-The art and science of environmentally sustainable agriculture. *Plant, Cell & Environment*, *42*(1), 1–5.

Frank, S. A. (1996). Host-symbiont conflict over the mixing of symbiotic lineages. *Proceedings. Biological Sciences*, *263*(1368), 339–344.

Friesen, M. L. (2012). Widespread fitness alignment in the legume–rhizobium symbiosis. *The New Phytologist*, *194*, 1096–1111.

Friesen, M. L., Porter, S. S., Stark, S. C., von Wettberg, E. J., Sachs, J. L., & Martinez-Romero, E. (2011). Microbially Mediated Plant Functional Traits. *Annual Review of Ecology, Evolution, and Systematics*, *42*(1), 23–46.

Fürnkranz, M., Wanek, W., Richter, A., Abell, G., Rasche, F., & Sessitsch, A. (2008). Nitrogen fixation by phyllosphere bacteria associated with higher plants and their

colonizing epiphytes of a tropical lowland rainforest of Costa Rica. *The ISME Journal*, *2*(5), 561–570.

Gano-Cohen, K. A., Wendlandt, C. E., Al Moussawi, K., Stokes, P. J., Quides, K. W., Weisberg, A. J., Chang, J. H., & Sachs, J. L. (2020). Recurrent mutualism breakdown events in a legume rhizobia metapopulation. In *Proceedings of the Royal Society B: Biological Sciences* (Vol. 287, Issue 1919, p. 20192549). https://doi.org/10.1098/rspb.2019.2549

Gano-Cohen, K. A., Wendlandt, C. E., Stokes, P. J., Blanton, M. A., Quides, K. W., Zomorrodian, A., Adinata, E. S., & Sachs, J. L. (2019). Interspecific conflict and the evolution of ineffective rhizobia. *Ecology Letters*, *22*(6), 914–924.

Garg, N., & Geetanjali. (2009). Symbiotic nitrogen fixation in legume nodules: Process and signaling: A review. In *Sustainable Agriculture* (pp. 519–531). Springer Netherlands.

Goel, A. K., Sindhu, S. S., & Dadarwal, K. R. (1999). Bacteriocin-producing native rhizobia of green gram (Vigna radiata) having competitive advantage in nodule occupancy. *Microbiological Research*, *154*(1), 43–48.

Gordon, B. R., Klinger, C. R., Weese, D. J., Lau, J. A., Burke, P. V., Dentinger, B. T. M., & Heath, K. D. (2016). Decoupled genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. *Ecology and Evolution*, *6*(5), 1317–1327.

Goyal, R. K., Mattoo, A. K., & Schmidt, M. A. (2021). Rhizobial-host interactions and symbiotic nitrogen fixation in legume crops toward agriculture sustainability. *Frontiers in Microbiology*, *12*, 669404.

Griekspoor, A., & Groothuis, T. (2006). 4peaks. *Available: Mekentosj. Com*.

Grohmann, E., Christie, P. J., Waksman, G., & Backert, S. (2018). Type IV secretion in Gram-negative and Gram-positive bacteria. *Molecular Microbiology*, *107*(4), 455–471.

Guillier, L., Gourmelon, M., Lozach, S., Cadel-Six, S., Vignaud, M.-L., Munck, N., Hald, T., & Palma, F. (2020). AB_SA: Accessory genes-Based Source Attribution - tracing the source of Salmonella enterica Typhimurium environmental strains. *Microbial Genomics*, *6*(7). https://doi.org/10.1099/mgen.0.000366

Hahn, M. (1986). Competitiveness of a nif− Bradyrhizobium japonicum mutant against the wild-type strain. *FEMS Microbiology Letters*, *33*(1), 143–148.

Heath, K. D. (2010). Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution; International Journal of Organic Evolution*, *64*(5), 1446–1458.

Heath, K. D., & Stinchcombe, J. R. (2014). Explaining mutualism variation: a new evolutionary paradox? *Evolution; International Journal of Organic Evolution*, *68*(2), 309–317.

Hernandez, J. A., Curatti, L., Aznar, C. P., Perova, Z., Britt, R. D., & Rubio, L. M. (2008). Metal trafficking for nitrogen fixation: NifQ donates molybdenum to NifEN/NifH for the biosynthesis of the nitrogenase FeMo-cofactor. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(33), 11679–11684.

Hernández-Tamayo, R., Sohlenkamp, C., Puente, J. L., Brom, S., & Romero, D. (2013). Characterization of IntA, a bidirectional site-specific recombinase required for conjugative transfer of the symbiotic plasmid of Rhizobium etli CFN42. *Journal of Bacteriology*, *195*(20), 4668–4677.

Hijaz, F., & Killiny, N. (2019). Exogenous GABA is quickly metabolized to succinic acid and fed into the plant TCA cycle. *Plant Signaling & Behavior*, *14*(3), e1573096.

Hill, Y., Colombi, E., Bonello, E., Haskett, T., Ramsay, J., O'Hara, G., & Terpolilli, J. (2021). Evolution of diverse effective N2-fixing microsymbionts of Cicer arietinum following horizontal transfer of the Mesorhizobium ciceri CC1192 symbiosis integrative and conjugative element. *Applied and Environmental Microbiology*, *87*(5), AEM.02558-20.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Hoeksema, J. D., Chaudhary, V. B., Gehring, C. A., Johnson, N. C., Karst, J., Koide, R. T., Pringle, A., Zabinski, C., Bever, J. D., Moore, J. C., Wilson, G. W. T., Klironomos, J. N., & Umbanhowar, J. (2010). A meta-analysis of context-dependency in plant response to inoculation with mycorrhizal fungi. *Ecology Letters*, *13*(3), 394–407.

Hollowell, A. C., Regus, J. U., Gano, K. A., Bantay, R., Centeno, D., Pham, J., Lyu, J. Y., Moore, D., Bernardo, A., Lopez, G., Patil, A., Patel, S., Lii, Y., & Sachs, J. L. (2016). Epidemic Spread of Symbiotic and Non-Symbiotic Bradyrhizobium Genotypes Across California. *Microbial Ecology*, *71*(3), 700–710.

Hollowell, Amanda C., Gano, K. A., Lopez, G., Shahin, K., Regus, J. U., Gleason, N., Graeter, S., Pahua, V., & Sachs, J. L. (2015). Native California soils are selective

reservoirs for multidrug-resistant bacteria. *Environmental Microbiology Reports*, *7*(3), 442–449.

Hollowell, Amanda C., Regus, J. U., Turissini, D., Gano-Cohen, K. A., Bantay, R., Bernardo, A., Moore, D., Pham, J., & Sachs, J. L. (2016). Metapopulation dominance and genomic-island acquisition of Bradyrhizobium with superior catabolic capabilities. *Proceedings. Biological Sciences / The Royal Society*, *283*(1829). https://doi.org/10.1098/rspb.2016.0496

Hollowell, Amanda 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: Biological Sciences*, *283*(1829), 20160496.

Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics, Theory and Applications*, *6*(2), 65–70.

Huerta-Cepas, J., Serra, F., & Bork, P. (2016). ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Molecular Biology and Evolution*, *33*(6), 1635–1638.

Irisarri, P., Cardozo, G., Tartaglia, C., Reyno, R., Gutiérrez, P., Lattanzi, F. A., Rebuffo, M., & Monza, J. (2019). *Selection of Competitive and Efficient Rhizobia Strains for White Clover* [Data set]. https://doi.org/10.3389/fmicb.2019.00768

Jain, C., Rodriguez-R, L. M., Phillippy, A. M., Konstantinidis, K. T., & Aluru, S. (2018). High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nature Communications*, *9*(1), 5114.

Johnson, N. C., Graham, J. H., & Smith, F. A. (1997). Functioning of mycorrhizal associations along the mutualism-parasitism continuum. *The New Phytologist*, *135*(4), 575–585.

Johnson, Nancy Collins, & Graham, J. H. (2013). The continuum concept remains a useful framework for studying mycorrhizal functioning. *Plant and Soil*, *363*(1), 411–419.

Jones, E. I., Afkhami, M. E., Akçay, E., Bronstein, J. L., Bshary, R., Frederickson, M. E., Heath, K. D., Hoeksema, J. D., Ness, J. H., Pankey, M. S., Porter, S. S., Sachs, J. L., Scharnagl, K., & Friesen, M. L. (2015). Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters*, *18*(11), 1270–1284.

Kajić, S., Hulak, N., & Sikora, S. (2016). Environmental stress response and adaptation mechanisms in rhizobia. *Agriculturae Conspectus Scientificus*, *81*(1), 15–19.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2*, e281.

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 USDA6T. *Genes*, *2*(4), 763–787.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780.

Kazmierczak, T., Yang, L., Boncompagni, E., Meilhoc, E., Frugier, F., Frendo, P., Bruand, C., Gruber, V., & Brouquisse, R. (2020). Legume nodule senescence: a coordinated death mechanism between bacteria and plant cells. In P. Frendo, F. Frugier, & C. Masson-Boivin (Eds.), *Advances in Botanical Research* (Vol. 94, pp. 181–212). Elsevier.

KDE Group, University of Kassel, DMIR Group, University of Würzburg, L3S Research Center, & (Germany), H. (n.d.). *FASTQC. A quality control tool for high throughput sequence data*. Retrieved July 12, 2022, from https://www.bibsonomy.org/bibtex/f230a919c34360709aa298734d63dca3

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics (Oxford, England)*, *26*(11), 1463–1464.

Kiers, E. T., & Denison, R. F. (2008). Sanctions, cooperation, and the stability of plant-rhizosphere mutualisms. *Annual Review of Ecology, Evolution, and Systematics*, *39*(1), 215–236.

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., Fellbaum, C. R., Kowalchuk, G. A., Hart, M. M., Bago, A., Palmer, T. M., West, S. A., Vandenkoornhuyse, P., Jansa, J., & Bücking, H. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science*, *333*(6044), 880–882.

Kiers, E. T., Ratcliff, W. C., & Denison, R. F. (2013). Single-strain inoculation may create spurious correlations between legume fitness and rhizobial fitness. *The New Phytologist*, *198*(1), 4–6.

Kiers, E. T., Rousseau, R. A., & Denison, R. F. (2006). Measured sanctions: legume hosts detect quantitative variation in rhizobium cooperation and punish accordingly. *Evolutionary Ecology Research*, *8*(6), 1077–1086.

Kiers, E. T., Rousseau, R. A., West, S. A., & Denison, R. F. (2003). Host sanctions and the legume-rhizobium mutualism. *Nature*, *425*(6953), 78–81.

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*(2), e1003204.

Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J., Grout, J., Corlay, S., Ivanov, P., Avila, D., Abdalla, S., Willing, C., & Jupyter Development Team. (2016). Jupyter Notebooks – a publishing format for reproducible computational workflows. In *Positioning and Power in Academic Publishing: Players, Agents and Agendas* (pp. 87–90). IOS Press.

Kooyman, R., Rossetto, M., Cornwell, W., & Westoby, M. (2011). Phylogenetic tests of community assembly across regional to continental scales in tropical and subtropical rain forests. *Global Ecology and Biogeography: A Journal of Macroecology*, *20*(5), 707–716.

Koppell, J. H., & Parker, M. A. (2012). Phylogenetic clustering of Bradyrhizobium symbionts on legumes indigenous to North America. *Microbiology*, *158*(Pt 8), 2050–2059.

Koskey, G., Mburu, S. W., Njeru, E. M., Kimiti, J. M., Ombori, O., & Maingi, J. M. (2017). Potential of native rhizobia in enhancing nitrogen fixation and yields of climbing beans (Phaseolus vulgaris L.) in contrasting environments of eastern Kenya. *Frontiers in Plant Science*, *8*, 443.

Kozich, J. J., Westcott, S. L., Baxter, N. T., Highlander, S. K., & Schloss, P. D. (2013). Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. *Applied and Environmental Microbiology*, *79*(17), 5112–5120.

La Pierre, K. J., Simms, E. L., Tariq, M., Zafar, M., & Porter, S. S. (2017). Invasive legumes can associate with many mutualists of native legumes, but usually do not. *Ecology and Evolution*, *7*(20), 8599–8611.

Lemaire, B., Van Cauwenberghe, J., Chimphango, S., Stirton, C., Honnay, O., Smets, E., & Muasya, A. M. (2015). Recombination and horizontal transfer of nodulation and ACC deaminase (acdS) genes within Alpha- and Betaproteobacteria nodulating legumes of the Cape Fynbos biome. *FEMS Microbiology Ecology*, *91*(11), fiv118.

Li, Y., Li, X., Liu, Y., Wang, E. T., Ren, C., Liu, W., Xu, H., Wu, H., Jiang, N., Li, Y., Zhang, X., & Xie, Z. (2016). Genetic diversity and community structure of rhizobia nodulating Sesbania cannabina in saline-alkaline soils. *Systematic and Applied Microbiology*, *39*(3), 195–202.

Liang, J., Hoffrichter, A., Brachmann, A., & Marín, M. (2018). Complete genome of Rhizobium leguminosarum Norway, an ineffective Lotus micro-symbiont. *Standards in Genomic Sciences*, *13*(1), 36.

Liu, C.-W., & Murray, J. D. (2016). The role of flavonoids in nodulation host-range specificity: An update. *Plants*, *5*(3), 33.

Liu, H. (2014). Ammonia synthesis catalyst 100 years: Practice, enlightenment and challenge. *Cuihua Xuebao/Chinese Journal of Catalysis*, *35*(10), 1619–1640.

Liu, J., Yu, X., Qin, Q., Dinkins, R. D., & Zhu, H. (2020). The impacts of domestication and breeding on nitrogen fixation symbiosis in legumes. *Frontiers in Genetics*, *11*, 00973.

Maan, P. K., & Garcha, S. (2018). Bacteriocins from Gram-negative Rhizobium spp. *Advances in Bioresearch*, *9*(1), 36–43.

Madsen, L. H., Tirichine, L., Jurkiewicz, A., Sullivan, J. T., Heckmann, A. B., Bek, A. S., Ronson, C. W., James, E. K., & Stougaard, J. (2010). The molecular network governing nodule organogenesis and infection in the model legume Lotus japonicus. *Nature Communications*, *1*(1), 10.

Magoč, T., & Salzberg, S. L. (2011). FLASH: fast length adjustment of short reads to improve genome assemblies. *Bioinformatics (Oxford, England)*, *27*(21), 2957–2963.

Massingham, T., & Goldman, N. (2005). Detecting amino acid sites under positive selection and purifying selection. *Genetics*, *169*(3), 1753–1762.

Masson-Boivin, C., & Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia-the roots of a success story. *Curr Opin Plant Biol*, *44*, 7–15.

Masson-Boivin, Catherine, & Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia — the roots of a success story. In *Current Opinion in Plant Biology* (Vol. 44, pp. 7–15). https://doi.org/10.1016/j.pbi.2017.12.001

Mcinnes, A. (2004). Structure and diversity among rhizobial strains, populations and communities?a review. *Soil Biology & Biochemistry*, *36*(8), 1295–1308.

Mendoza-Suárez, M. A., Geddes, B. A., Sánchez-Cañizares, C., Ramírez-González, R. H., Kirchhelle, C., Jorrin, B., & Poole, P. S. (2020). Optimizing Rhizobium-legume symbioses by simultaneous measurement of rhizobial competitiveness and N2 fixation in nodules. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(18), 9822–9831.

Mendoza-Suárez, M., Andersen, S. U., Poole, P. S., & Sánchez-Cañizares, C. (2021). Competition, Nodule Occupancy, and Persistence of Inoculant Strains: Key Factors in the Rhizobium-Legume Symbioses. *Frontiers in Plant Science*, *12*, 690567.

Mi, H., Muruganujan, A., Huang, X., Ebert, D., Mills, C., Guo, X., & Thomas, P. D. (2019). Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). *Nature Protocols*, *14*(3), 703–721.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.

Nandasena, K. G., O'Hara, G. W., Tiwari, R. P., Sezmiş, E., & Howieson, J. G. (2007). In situ lateral transfer of symbiosis islands results in rapid evolution of diverse competitive strains of mesorhizobia suboptimal in symbiotic nitrogen fixation on the pasture legume Biserrula pelecinus L. *Environmental Microbiology*, *9*(10), 2496–2511.

Nelson, M. S., & Sadowsky, M. J. (2015). Secretion systems and signal exchange between nitrogen-fixing rhizobia and legumes. *Frontiers in Plant Science*, *6*, 491.

Nylander, J. (2018). *catfasta2phyml: Concatenates FASTA formatted files to one "phyml" (PHYLIP) formatted file* (v1.1.0) [Computer software]. Github. https://github.com/nylander/catfasta2phyml

Ohyama, T. (2017). The Role of Legume-Rhizobium Symbiosis in Sustainable Agriculture. In S. Sulieman & L.-S. P. Tran (Eds.), *Legume Nitrogen Fixation in Soils with Low Phosphorus Availability: Adaptation and Regulatory Implication* (pp. 1–20). Springer International Publishing.

Oldroyd, G. E. D., Murray, J. D., Poole, P. S., & Downie, J. A. (2011). The Rules of Engagement in the Legume-Rhizobial Symbiosis. *Annual Review of Genetics*, *45*(1), 119–144.

Oono, R., Anderson, C. G., & Denison, R. F. (2011). Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings. Biological Sciences / The Royal Society*, *278*(1718), 2698–2703.

Oono, R., Denison, R. F., & Kiers, E. T. (2009). Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *The New Phytologist*, *183*(4), 967–979.

Ormeño-Orrillo, E., & Martínez-Romero, E. (2019). A Genomotaxonomy View of the Bradyrhizobium Genus. *Frontiers in Microbiology*, *10*(1334). https://doi.org/10.3389/fmicb.2019.01334

Ortiz-Barbosa, G. S., Torres-Martínez, L., Manci, A., Neal, S., Soubra, T., Khairi, F., Trinh, J., Cardenas, P., & Sachs, J. L. (2022). No disruption of rhizobial symbiosis during early stages of cowpea domestication. *Evolution; International Journal of Organic Evolution*, *76*(3), 496–511.

Ouma, E. W., Asango, A. M., Maingi, J., & Njeru, E. M. (2016). Elucidating the potential of native rhizobial isolates to improve biological nitrogen fixation and growth of common bean and soybean in smallholder farming systems of Kenya. *International Journal of Agronomy*, *2016*, 1–7.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., Fookes, M., Falush, D., Keane, J. A., & Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics (Oxford, England)*, *31*(22), 3691–3693.

Pahua, V. J., Stokes, P. J. N., Hollowell, A. C., Regus, J. U., Gano-Cohen, K. A., Wendlandt, C. E., Quides, K. W., Lyu, J. Y., & Sachs, J. L. (2018). Fitness variation among host species and the paradox of ineffective rhizobia. *Journal of Evolutionary Biology*, *31*(4), 599–610.

Paradis, E. (2010). pegas: an R package for population genetics with an integrated-modular approach. *Bioinformatics (Oxford, England)*, *26*(3), 419–420.

Parker, M. A. (2012). Legumes select symbiosis island sequence variants in Bradyrhizobium. *Mol Ecol*, *21*(7), 1769–1778.

Parker, M. A. (2015). The Spread of Bradyrhizobium Lineages Across Host Legume Clades: from Abarema to Zygia. *Microbial Ecology*, *69*(3), 630–640.

Pérez Carrascal, O. M., VanInsberghe, D., Juárez, S., Polz, M. F., Vinuesa, P., & González, V. (2016). Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing Rhizobium species associated with Phaseolus vulgaris. *Environmental Microbiology*, *18*(8), 2660–2676.

Pessi, G., Ahrens, C. H., Rehrauer, H., Lindemann, A., Hauser, F., Fischer, H. M., & Hennecke, H. (2007). Genome-wide transcript analysis of Bradyrhizobium japonicum bacteroids in soybean root nodules. *Molecular Plant-Microbe Interactions*, *20*(11), 1353–1363.

Poole, P., Ramachandran, V., & Terpolilli, J. (2018). Rhizobia: from saprophytes to endosymbionts. *Nature Reviews. Microbiology*, *16*(5), 291–303.

Porter, S. S., & Simms, E. L. (2014). Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecology Letters*, *17*(9), 1121–1129.

Price, P. A., Tanner, H. R., Dillon, B. A., Shabab, M., Walker, G. C., & Griffitts, J. S. (2015). Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and mediates symbiotic compatibility. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(49), 15244–15249.

Pueppke, S. G., & Broughton, W. J. (1999). Rhizobium sp. strain NGR234 and R. fredii USDA257 share exceptionally broad, nested host ranges. *Mol Plant Microbe Interact*, *12*(4), 293–318.

Quides, K. W., Salaheldine, F., Jariwala, R., & Sachs, J. L. (2021). Dysregulation of host-control causes interspecific conflict over host investment into symbiotic organs. *Evolution; International Journal of Organic Evolution*, *75*(5), 1189–1200.

R Core Team. (2013). R: A language and environment for statistical computing. *Http://Www. R-Project. Org/*.

Ratzke, C., Barrere, J., & Gore, J. (2020). Strength of species interactions determines biodiversity and stability in microbial communities. *Nature Ecology & Evolution*, *4*(3), 376–383.

Raza, A. (2020). *The Plant Family Fabaceae: Biology and Physiological Responses to Environmental Stresses*. 43–74.

Reeve, W., De Meyer, S., Terpolilli, J., Melino, V., Ardley, J., Tian, R., Tiwari, R., Howieson, J., Yates, R., O'Hara, G., Ninawi, M., Lu, M., Bruce, D., Detter, C.,

Tapia, R., Han, C., Wei, C.-L., Huntemann, M., Han, J., … Kyrpides, N. (2013). Genome sequence of the Ornithopus/Lupinus-nodulating Bradyrhizobium sp. strain WSM471. *Standards in Genomic Sciences*, *9*(2), 254–263.

Regus, J. U., Wendlandt, C. E., Bantay, R. M., Gano-Cohen, K. A., Gleason, N. J., Hollowell, A. C., O'Neill, M. R., Shahin, K. K., & Sachs, J. L. (2017). Nitrogen deposition decreases the benefits of symbiosis in a native legume. *Plant and Soil*, *414*(1), 159–170.

Regus, John U., Gano, K. A., Hollowell, A. C., & Sachs, J. L. (2014). Efficiency of partner choice and sanctions in Lotus is not altered by nitrogen fertilization. *Proceedings. Biological Sciences / The Royal Society*, *281*(1781), 20132587.

Regus, John U., Quides, K. W., O'Neill, M. R., Suzuki, R., Savory, E. A., Chang, J. H., & Sachs, J. L. (2017). Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *American Journal of Botany*, *104*(9), 1299–1312.

Remigi, P., Zhu, J., Young, J. P. W., & Masson-Boivin, C. (2016). Symbiosis within Symbiosis: Evolving Nitrogen-Fixing Legume Symbionts. *Trends in Microbiology*, *24*(1), 63–75.

Riley, A. B., Grillo, M. A., Epstein, B., Tiffin, P., & Heath, K. D. (2022). Discordant population structure among rhizobium divided genomes and their legume hosts. *Molecular Ecology*. https://doi.org/10.1111/mec.16704

Rogel, M. A., Ormeño-Orrillo, E., & Martinez Romero, E. (2011). Symbiovars in rhizobia reflect bacterial adaptation to legumes. *Systematic and Applied Microbiology*, *34*(2), 96–104.

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*(5), 1075–1089.

Sachs, J. L., Kembel, S. W., Lau, A. H., & Simms, E. L. (2009a). In situ phylogenetic structure and diversity of wild Bradyrhizobium communities. *Applied and Environmental Microbiology*, *75*(14), 4727–4735.

Sachs, J. L., Kembel, S. W., Lau, A. H., & Simms, E. L. (2009b). In situ phylogenetic structure and diversity of wild *Bradyrhizobium* communities. *Applied and Environmental Microbiology*, *75*(14), 4727–4735.

Sachs, J. L., Russell, J. E., Lii, Y. E., Black, K. C., Lopez, G., & Patil, A. S. (2010). Host control over infection and proliferation of a cheater symbiont. *Journal of Evolutionary Biology*, *23*(9), 1919–1927.

Sachs, J. L., Skophammer, R. G., & Regus, J. U. (2011). Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences*, *108*, 10800–10807.

Sachs, Joel L., Mueller, U. G., Wilcox, T. P., & Bull, J. J. (2004). The evolution of cooperation. *The Quarterly Review of Biology*, *79*(2), 135–160.

Sachs, Joel L., Quides, K. W., & Wendlandt, C. E. (2018). Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *The New Phytologist*, *219*(4), 1199–1206.

Sachs, Joel L., Russell, J. E., & Hollowell, A. C. (2011). Evolutionary Instability of Symbiotic Function in Bradyrhizobium japonicum. *PloS One*, *6*(11), e26370.

Sachs, Joel L., & Simms, E. L. (2008). The origins of uncooperative rhizobia. *Oikos*, *117*(7), 961–966.

Saranraj, P., Sayyed, R. Z., Sivasakthivelan, P., Kokila, M., Al-Tawaha, A. R. M., Amala, K., & Yasmin, H. (2023). Symbiotic effectiveness of Rhizobium strains in agriculture. In *Plant Growth Promoting Microorganisms of Arid Region* (pp. 389–421). Springer Nature Singapore.

Savin, N. E., & White, K. J. (1977). The Durbin-Watson test for serial correlation with extreme sample sizes or many regressors. *Econometrica: Journal of the Econometric Society*, *45*(8), 1989.

Sawada, H., Kuykendall, L. D., & Young, J. M. (2003). Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts. *The Journal of General and Applied Microbiology*, *49*(3), 155–179.

Schloss, P. D. (2020). Reintroducing mothur: 10 years later. *Applied and Environmental Microbiology*, *86*(2). https://doi.org/10.1128/AEM.02343-19

Schloss, P. D., Westcott, S. L., Ryabin, T., Hall, J. R., Hartmann, M., Hollister, E. B., Lesniewski, R. A., Oakley, B. B., Parks, D. H., Robinson, C. J., Sahl, J. W., Stres, B., Thallinger, G. G., Van Horn, D. J., & Weber, C. F. (2009). Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Applied and Environmental Microbiology*, *75*(23), 7537–7541.

Schumpp, O., & Deakin, W. J. (2010). How inefficient rhizobia prolong their existence within nodules. *Trends in Plant Science*, *15*(4), 189–195.

Schwinghamer, E. A. (1971). Antagonism between strains of Rhizobium trifolii in culture. *Soil Biology & Biochemistry*, *3*(4), 355–363.

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*, *30*(14), 2068–2069.

Serova, T. A., Tsyganova, A. V., & Tsyganov, V. E. (2018). Early nodule senescence is activated in symbiotic mutants of pea (Pisum sativum L.) forming ineffective nodules blocked at different nodule developmental stages. *Protoplasma*, *255*(5), 1443–1459.

Sexton, J. T., & Tabor, J. J. (2020). Multiplexing cell-cell communication. *Molecular Systems Biology*, *16*(7), e9618.

Simms, E. L., Taylor, D. L., Povich, J., Shefferson, R. P., Sachs, J. L., Urbina, M., & Tausczik, Y. (2006). An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proceedings of the Royal Society B-Biological Sciences*, *273*(1582), 77–81.

Simms, Ellen L., & Taylor, D. L. (2002). Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. *Integrative and Comparative Biology*, *42*(2), 369–380.

Sitto, F., & Battistuzzi, F. U. (2020). Estimating pangenomes with Roary. *Molecular Biology and Evolution*, *37*(3), 933–939.

Somasegaran, P., & Hoben, H. J. (1994). Verifying the nitrogen-fixing potential of glasshouse-selected soybean rhizobia in the field environment. In *Handbook for Rhizobia* (pp. 189–197). Springer New York.

Somasegaran, P., & Hoben, H. J. (2012). *Handbook for rhizobia: methods in legume-Rhizobium technology*. Springer Science & Business Media.

Somervuo, P., Koskinen, P., Mei, P., Holm, L., Auvinen, P., & Paulin, L. (2018). BARCOSEL: a tool for selecting an optimal barcode set for high-throughput sequencing. *BMC Bioinformatics*, *19*(1). https://doi.org/10.1186/s12859-018-2262-7

Soumare, A., Diedhiou, A. G., Thuita, M., Hafidi, M., Ouhdouch, Y., Gopalakrishnan, S., & Kouisni, L. (2020). Exploiting biological nitrogen fixation: A route towards a sustainable agriculture. *Plants*, *9*(8), 1011.

Sprouffske, K., & Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, *17*(1), 172.

Stepkowski, T., Moulin, L., Krzyzańska, A., McInnes, A., Law, I. J., & Howieson, J. (2005). European origin of Bradyrhizobium populations infecting lupins and serradella in soils of Western Australia and South Africa. *Applied and Environmental Microbiology*, *71*(11), 7041–7052.

Stępkowski, T., Zak, M., Moulin, L., Króliczak, J., Golińska, B., Narożna, D., Safronova, V. I., & Mądrzak, 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*(5), 368–375.

Sulieman, S. (2011). Does GABA increase the efficiency of symbiotic N2 fixation in legumes? *Plant Signaling & Behavior*, *6*(1), 32–36.

Sullivan, J. T., Patrick, H. N., Lowther, W. L., Scott, D. B., & Ronson, C. W. (1995). Nodulating strains of Rhizobium loti arise through chromosomal symbiotic gene transfer in the environment. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(19), 8985–8989.

Suzaki, T., Yoro, E., & Kawaguchi, M. (2015). Leguminous plants: inventors of root nodules to accommodate symbiotic bacteria. *International Review of Cell and Molecular Biology*, *316*, 111–158.

Tang, J., Bromfield, E. S. P., Rodrigue, N., Cloutier, S., & Tambong, J. T. (2012). Microevolution of symbioticBradyrhizobiumpopulations associated with soybeans in east North America. In *Ecology and Evolution* (Vol. 2, Issue 12, pp. 2943–2961). https://doi.org/10.1002/ece3.404

Terpolilli, J. J., Masakapalli, S. K., Karunakaran, R., Webb, I. U. C., Green, R., Watmough, N. J., Kruger, N. J., Ratcliffe, R. G., & Poole, P. S. (2016). Lipogenesis and redox balance in nitrogen-fixing pea bacteroids. *Journal of Bacteriology*, *198*(20), 2864–2875.

Torres-Martínez, L., Porter, S. S., Wendlandt, C., Purcell, J., Ortiz-Barbosa, G., Rothschild, J., Lampe, M., Warisha, F., Le, T., Weisberg, A. J., Chang, J. H., & Sachs, J. L. (2021). Evolution of specialization in a plant-microbial mutualism is explained by the oscillation theory of speciation. *Evolution*, *75*(5), 1070–1086.

Townsend, A. R., & Howarth, R. W. (2010). Fixing the global nitrogen problem. *Scientific American*, *302*(2), 64–71.

Uchiumi, T., Ohwada, T., Itakura, M., Mitsui, H., Nukui, N., Dawadi, P., Kaneko, T., Tabata, S., Yokoyama, T., Tejima, K., Saeki, K., Omori, H., Hayashi, M., Maekawa, T., Sriprang, R., Murooka, Y., Tajima, S., Simomura, K., Nomura, M.,

… Minamisawa, K. (2004). Expression islands clustered on the symbiosis island of the Mesorhizobium loti genome. *Journal of Bacteriology*, *186*(8), 2439–2448.

UniProt Consortium. (2021). UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489.

Van Cauwenberghe, J., Visch, W., Michiels, J., & Honnay, O. (2016). Selection mosaics differentiateRhizobium-host plant interactions across different nitrogen environments. *Oikos (Copenhagen, Denmark)*, *125*(12), 1755–1761.

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 j*, *9*(11), 2435–2441.

Verma, D. (1992). Signals in root nodule organogenesis and endocytosis of Rhizobium. *The Plant Cell*, *4*(4), 373–382.

Vinuesa, P., Silva, C., Werner, D., & Martinez-Romero, E. (2005). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in Bradyrhizobium species cohesion and delineation. *Molecular Phylogenetics and Evolution*, *34*(1), 29–54.

von Wettberg, E. J. B., Chang, P. L., Başdemir, F., Carrasquila-Garcia, N., Korbu, L. B., Moenga, S. M., Bedada, G., Greenlon, A., Moriuchi, K. S., Singh, V., Cordeiro, M. A., Noujdina, N. V., Dinegde, K. N., Shah Sani, S. G. A., Getahun, T., Vance, L., Bergmann, E., Lindsay, D., Mamo, B. E., … Cook, D. R. (2018). Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-02867-z

Vuong, H. B., Thrall, P. H., & Barrett, L. G. (2017). Host species and environmental variation can influence rhizobial community composition. *The Journal of Ecology*, *105*(2), 540–548.

Wade, T. K., Le Quéré, A., Laguerre, G., N'zoué, A., Ndione, J.-A., Dorego, F., Sadio, O., Ndoye, I., & Neyra, M. (2014). Eco-geographical diversity of cowpea bradyrhizobia in Senegal is marked by dominance of two genetic types. *Systematic and Applied Microbiology*, *37*(2), 129–139.

Walker, L., Lagunas, B., & Gifford, M. L. (2020). Determinants of host range specificity in legume-rhizobia symbiosis. *Frontiers in Microbiology*, *11*, 585749.

Wang, D., Yang, S., Tang, F., & Zhu, H. (2012). Symbiosis specificity in the legume: rhizobial mutualism. *Cellular Microbiology*, *14*(3), 334–342.

Wang, Q., Liu, J., & Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-Rhizobium interactions. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00313

Wardell, G. E., Hynes, M. F., Young, P. J., & Harrison, E. (2022). Why are rhizobial symbiosis genes mobile? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1842), 20200471.

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505.

Weis, V. M., Reynolds, W. S., deBoer, M. D., & Krupp, D. A. (2001). Host-symbiont specificity during onset of symbiosis between the dinoflagellates Symbiodinium spp. and planula larvae of the scleractinian coral Fungia scutaria. *Coral Reefs*, *20*(3), 301–308.

Weisberg, A. J., Rahman, A., Backus, D., Tyavanagimatt, P., & Chang, J. H. (2022). Pangenome evolution reconciles robustness and instability of symbiosis. *In Review*.

Weisberg, A. J., Rahman, A., Backus, D., Tyavanagimatt, P., Chang, J. H., & Sachs, J. L. (2022). Pangenome evolution reconciles robustness and instability of rhizobial symbiosis. *MBio*, e0007422.

Wendlandt, C. E., Gano-Cohen, K. A., Stokes, P. J. N., Jonnala, B. N. R., Zomorrodian, A. J., Al-Moussawi, K., & Sachs, J. L. (2022). Wild legumes maintain beneficial soil rhizobia populations despite decades of nitrogen deposition. *Oecologia*, *198*(2), 419–430.

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. (2019). Host investment into symbiosis varies among genotypes of the legume Acmispon strigosus, but host sanctions are uniform. *The New Phytologist*, *221*(1), 446–458.

West, S. A., Kiers, E. T., Pen, I., & Denison, R. F. (2002). Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology*, *15*(5), 830–837.

West, Stuart A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings. Biological Sciences / The Royal Society*, *269*(1492), 685–694.

Westhoek, A., Clark, L. J., Culbert, M., Dalchau, N., Griffiths, M., Jorrin, B., Karunakaran, R., Ledermann, R., Tkacz, A., Webb, I., James, E. K., Poole, P. S., & Turnbull, L. A. (2021). Conditional sanctioning in a legume-Rhizobium mutualism.

Wang, Q., Liu, J., & Zhu, H. (2018). Genetic and molecular mechanisms underlying symbiotic specificity in legume-Rhizobium interactions. *Frontiers in Plant Science*, *9*. https://doi.org/10.3389/fpls.2018.00313

Wardell, G. E., Hynes, M. F., Young, P. J., & Harrison, E. (2022). Why are rhizobial symbiosis genes mobile? *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *377*(1842), 20200471.

Webb, C. O., Ackerly, D. D., McPeek, M. A., & Donoghue, M. J. (2002). Phylogenies and community ecology. *Annual Review of Ecology and Systematics*, *33*(1), 475–505.

Weis, V. M., Reynolds, W. S., deBoer, M. D., & Krupp, D. A. (2001). Host-symbiont specificity during onset of symbiosis between the dinoflagellates Symbiodinium spp. and planula larvae of the scleractinian coral Fungia scutaria. *Coral Reefs*, *20*(3), 301–308.

Weisberg, A. J., Rahman, A., Backus, D., Tyavanagimatt, P., & Chang, J. H. (2022). Pangenome evolution reconciles robustness and instability of symbiosis. *In Review*.

Weisberg, A. J., Rahman, A., Backus, D., Tyavanagimatt, P., Chang, J. H., & Sachs, J. L. (2022). Pangenome evolution reconciles robustness and instability of rhizobial symbiosis. *MBio*, e0007422.

Wendlandt, C. E., Gano-Cohen, K. A., Stokes, P. J. N., Jonnala, B. N. R., Zomorrodian, A. J., Al-Moussawi, K., & Sachs, J. L. (2022). Wild legumes maintain beneficial soil rhizobia populations despite decades of nitrogen deposition. *Oecologia*, *198*(2), 419–430.

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. (2019). Host investment into symbiosis varies among genotypes of the legume Acmispon strigosus, but host sanctions are uniform. *The New Phytologist*, *221*(1), 446–458.

West, S. A., Kiers, E. T., Pen, I., & Denison, R. F. (2002). Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *Journal of Evolutionary Biology*, *15*(5), 830–837.

West, Stuart A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proceedings. Biological Sciences / The Royal Society*, *269*(1492), 685–694.

Westhoek, A., Clark, L. J., Culbert, M., Dalchau, N., Griffiths, M., Jorrin, B., Karunakaran, R., Ledermann, R., Tkacz, A., Webb, I., James, E. K., Poole, P. S., & Turnbull, L. A. (2021). Conditional sanctioning in a legume-Rhizobium mutualism.

Proceedings of the National Academy of Sciences of the United States of America, *118*(19), e2025760118.

Westhoek, A., Field, E., Rehling, F., Mulley, G., Webb, I., Poole, P. S., & Turnbull, L. A. (2017). Policing the legume-Rhizobium symbiosis: a critical test of partner choice. *Scientific Reports*, *7*(1), 1419.

Wheatley, R. M., Ford, B. L., Li, L., Aroney, S. T. N., Knights, H. E., Ledermann, R., East, A. K., Ramachandran, V. K., & Poole, P. S. (2020). Lifestyle adaptations of Rhizobium from rhizosphere to symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(38), 23823–23834.

Wielbo, J. (2012). Rhizobial communities in symbiosis with legumes: genetic diversity, competition and interactions with host plants. *Open Life Sciences*, *7*(3), 125.

Yang, S., Tang, F., Gao, M., Krishnan, H. B., & Zhu, H. (2010). R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(43), 18735–18740.

Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. *Molecular Biology and Evolution*, *24*(8), 1586–1591.

Yasuda, M., Miwa, H., Masuda, S., Takebayashi, Y., Sakakibara, H., & Okazaki, S. (2016). Effector-triggered immunity determines host genotype-specific incompatibility in legume-Rhizobium symbiosis. *Plant & Cell Physiology*, *57*(8), 1791–1800.

Yates, R. J., Howieson, J. G., Reeve, W. G., & O'Hara, G. W. (2011). A re-appraisal of the biology and terminology describing rhizobial strain success in nodule occupancy of legumes in agriculture. *Plant and Soil*, *348*(1–2), 255–267.

Zahran, H. H. (1999). Rhizobium-legume symbiosis and nitrogen fixation under severe conditions and in an arid climate. *Microbiology and Molecular Biology Reviews: MMBR*, *63*(4), 968–989, table of contents.

Zaw, M., Rathjen, J. R., Zhou, Y., Ryder, M. H., & Denton, M. D. (2021). Symbiotic effectiveness, ecological adaptation and phylogenetic diversity of chickpea rhizobia isolated from a large-scale Australian soil collection. *Plant and Soil*, *469*(1–2), 49–71.

Zhang, Z., Liu, W., Shao, S., Wang, E.-T., & Li, Y. (2020). Diverse Genomic Backgrounds Vs. Highly Conserved Symbiotic Genes in Sesbania-Nodulating Bacteria: Shaping of the Rhizobial Community by Host and Soil Properties. *Microbial Ecology*, *80*(1), 158–168.

Zhao, R., Liu, L. X., Zhang, Y. Z., Jiao, J., Cui, W. J., Zhang, B., Wang, X. L., Li, M. L., Chen, Y., Xiong, Z. Q., Chen, W. X., & Tian, C. F. (2018). Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. *The ISME Journal*, *12*(1), 101–111.