Genetic variation in host selectivity and adaptive strain enrichment within legume-rhizobia symbiosis: processes are host-dependent, far from perfect, and correlate with nodule morphology

Liana Burghardt[1 *], Patrick Sydow[1], Jeremy Sutherland[1], Brendan Epstein[2], and Peter Tiffin[2]

1. Department of Plant Science, Pennsylvania State University, University Park, PA, 16802
2. Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN 55108

* corresponding author: liana.burghardt@gmail.com

ORCID
L. Burghardt 0000-0002-0239-1071
P. Sydow 0009-0001-4761-5446
J. Sutherland: 0000-0003-3535-0302
B. Epstein 0000-0001-7083-1588
P. Tiffin 0000-0003-1975-610X

KEYWORDS: *Medicago truncatula*, sanctions, mutualism, cheaters, rewards, bacterial fitness

Author contributions: LTB and PT conceived the study, analyzed data, created figures, and wrote and revised the paper. LTB implemented the experiment. BE/PS/JS performed bioinformatic and statistical analyses, drafted text and figures, and edited the manuscript.

ABSTRACT: Mutualism breakdown can be prevented if partner species preferentially select and reward partners that provide greater benefit. We examined these two components using the legume *Medicago truncatula* and its nitrogen-fixing symbiont *Sinorhizobium meliloti*. First, we reanalyzed data from 202 accessions to show significant genetic variation in the capacity of *Medicago* to restrict strain diversity, finding that hosts with shorter nodules were more selective. A genome-wide association study on host selectivity identified genes including the hormone leginsulin, pectin degradation, multidrug and toxic compound efflux, Zinc transport, and DNA methylation. Second, we used two well-studied *Medicago* genotypes with contrasting nodule morphologies to assess the effectiveness of adaptive enrichment mechanisms by sampling the relative frequencies of rhizobial strains in pools of small nodules (indicating a lack of host investment) compared to large nodules (indicating increased host investment) and pairing these results with previous single-strain assessments of strain benefits to hosts. While both hosts enriched beneficial strains in large nodules, the host that formed larger and more variably sized nodules and thus had greater 'potential' to increase rhizobial populations was less effective. Our findings reveal that host genetic variation affects strain selectivity and suggest that nodule traits warrant attention when exploring mutualism evolution.

INTRODUCTION:

Mutualisms occur when two species derive fitness benefits from interacting with each other (Bronstein, 2015). Mutualism's stability depends on mechanisms limiting the spread of genotypes that do not benefit their partners (Sachs et al., 2004). These partner choice mechanisms may operate before resources are exchanged via, for instance, signaling and recognition (Bull and Rice, 1991), or after partners start to exchange benefits (e.g., (Denison, 2000; Kiers et al., 2003; Bever et al., 2009)). Post-exchange mechanisms, such as sanctions and rewards, both involve the host altering the fitness of symbionts but differ in that sanctions block the reproduction of less beneficial symbionts, whereas rewards direct resources to beneficial symbionts (Bever et al., 2009; Bever, 2015). Despite the centrality of partner choice mechanisms to mutualism evolution, little is known about natural genetic variation in these traits within host species (Porter et al., 2024). When and which of these mechanisms hosts use is important because they have different effects on maintaining variation in the symbiont quality (Bever, 2015). Here, we characterize natural genetic variation in host selectivity and adaptive enrichment in the model mutualism between *Medicago truncatula* and *Sinorhizobium meliloti*.

Legumes and rhizobia form a resource-based mutualism that is a model for investigating the genetics and evolution of plant-bacteria symbiosis (Sachs et al., 2004, 2018; Oldroyd et al., 2011; Friesen, 2012; Burghardt et al., 2018). This mutualism is formed when soil-inhabiting rhizobia invade a legume root, triggering the plant to develop a nodule. Inside the nodule, rhizobia convert atmospheric N_2 to a plant-usable form. In exchange, the plant provides the bacteria with carbon compounds, which, along with the environment inside the nodule, support a high rate of bacterial reproduction (Sprent et al., 1987); hundreds of thousands of bacteria can emerge from a nodule formed by a single bacterium (e.g. (Sachs et al., 2010)). Although a single bacterium invades most nodules, individual plants, which can form hundreds of nodules, can form nodules with many different rhizobial strains. It is vital to study mutualism maintenance in horizontally transmitted mutualisms like this because they are considered particularly susceptible to mutualism breakdown as rhizobia spend many generations living in soil and rhizosphere, between inhabiting nodules (Wilkinson and Sherratt, 2001).

Pre-nodulation selection mechanisms may be less effective at filtering for beneficial partners than post-nodulation mechanisms because hosts form associations before receiving feedback on whether the rhizobia they associate with are beneficial (Wang et al., 2012). However, once nodules have formed, host species may allocate resources based on the resources being provided by individual nodules, thereby differentially allocating resources to protect themselves from investing in less-beneficial rhizobia (Porter et al., 2024). For example, by reducing N_2 availability to nodules, researchers have shown that hosts can withhold resources from nodules (i.e., sanctions) when no N-fixation occurs (Denison, 2000; Kiers et al., 2003). However, it is less clear if sanctions are effective when N-fixation is reduced rather than eliminated (Kiers et al., 2006). Nevertheless, several studies suggest post-nodule formation enrichment can occur if there is a substantial difference in N-fixing capacity between strains (Quides et al., 2017, 2021; Regus et al., 2017; Burghardt et al., 2018; Wendlandt et al., 2019; Westhoek et al., 2021; Montoya et al., 2023; Underwood et al., 2024). One theme emerging from these studies is that hosts may be able to enrich strains by funneling resources to increase the size of nodules with helpful strains and limiting resources to nodules inhabited by less helpful strains (Porter et al., 2024). These findings suggest that nodule size traits and morphological variation that have received little empirical attention may provide new avenues for studying partner choice mechanisms.

Nodule development differs among legume species, with some forming nodules with determinate meristems and others forming nodules with indeterminate meristems (Brewin, 1991; Łotocka et al., 2012). This difference can have consequences for rhizobial populations. In species with determinate nodule development, nodules have a meristem that undergoes a fixed number of divisions, creating spherical nodules with little morphological variation beyond size. In contrast, in species that form indeterminate nodules, the nodule meristem keeps dividing, causing ongoing nodule growth and

elongation, increasing the zone of N-fixation, and increasing the population size of the rhizobia in the nodules. The meristem of indeterminate nodules can also branch, increasing the habitat space for rhizobial populations. This is likely to be especially salient in species where rhizobia near the meristem remain undifferentiated, but rhizobia in the nitrogen-fixing bacterioids are terminally differentiated (Brewin, 1991). Here, we examine the relationship between nodule morphology and host selectivity. Our initial prediction was that host genotypes with higher variance in nodule size have a greater potential to enrich helpful strains by boosting populations in large nodules.

We took a two-pronged approach to evaluate i) genetic variation among Medicago hosts in their ability to preferentially select or reward more beneficial rhizobia strains and ii) the extent to which nodule morphology is associated with differential rewards. First, to examine if there is natural genetic variation for host selectivity, we extended the analysis of a previously published dataset including 202 *M. truncatula* accessions (Epstein et al. 2023) by accounting for differences in nodule number across samples and expanding the nodule traits examined. Second, we collected data on two well-studied *Medicago* accessions with intermediate host selectivity and contrasting nodule morphology patterns to assess the efficacy of pre- and post-nodulation selection mechanisms. We grew each host with a mixture of 68 naturally occurring rhizobial isolates. To assess rewards, we sampled rhizobial strains in pools of large nodules (representing increased host investment) vs. small nodules (lack of investment after nodule formation). We show that host selection of beneficial rhizobia occurred both before and after nodules formed, and host selectivity is associated with shorter and more numerous nodules, and enrichment was more effective in the host with shorter, more numerous nodules.

MATERIALS AND METHODS:

Genetic variation in host selectivity and nodule morphology: To determine the breadth of genetic variation in nodule morphology and host selectivity, we expanded the analysis of a study that grew 202 accessions of *Medicago truncatula* (the HapMap panel) with a mixture of 89 strains of *Sinorhizobium meliloti* (Epstein et al., 2023) . Epstein et al. (2023) estimated strain diversity among the nodules on each plant based on the frequency of each strain in the nodules. However, nodule number was positively correlated with strain diversity ($R=0.26$, $p < 0.001$). The among-accession differences in nodule number may, therefore, obscure mechanisms that more directly affect host selectivity of differential rewards. To account for the effect of nodule number, we calculated a metric of host selectivity that controlled for nodule sampling via simulations of the nodule community. In our simulations, the number of nodules sampled from each plant determined the number of random draws from the initial 89-strain community (Figure S1). Simulations were completed in R, and the probability of drawing a strain was equal to the relative frequency of the strain in the initial inoculum. To calculate host selectivity, we subtracted each sample's observed Shannon Diversity Index (H calculated with vegan package version 2.6-4; Oksanen, 2022) from the mean H value from 10 replicate simulations to calculate host selectivity. Host selectivity values above zero indicate a host with a less diverse strain community than expected by chance, while values below zero indicate a more diverse community. One potential mechanism for observing negative host selectivity values could be a tendency for multiple strains to inhabit a single nodule. However, empirical estimates suggest co-infection is rare: in fact, entire papers have been published on artificially increasing coinfection to 3-20% in the model *Medicago* (Daubech et al., 2017). When we explored the possibility of co-infection in our assembly model, we found that host selectivity values increased slightly. Still, host selectivity values did not change relative to one another and therefore should not impact inferences below.

Associating host selectivity with nodule morphology characteristics: We used Least Absolute Shrinkage and Selection Operator (LASSO) regression in R (glmnet package version 4.1-7; Friedman, 2010) with 30

traits describing the mean and variance of nodule morphologies (Table S1) as predictors. LASSO regression is a machine learning algorithm that selects a sparse linear regression model with the smallest error. The number of parameters is determined by the size of their coefficients, with the tuning parameter (λ) determining the magnitude of the penalty. Genotypic means of nodule traits were centered and scaled prior to feature selection. The value for λ was selected following cross-validation, resulting in the lowest mean squared error with a training dataset of 150 out of 203 genotypes.

Associating Medicago genetic variation with host selectivity: To determine host genomic regions associated with the host selectivity, we conducted a Genome-wide Association Study (GWAS) using a linear mixed model (LMM) implemented with the -lmm option in GEMMA (v0.98.3) (Zhou and Stephens, 2012). We obtained the SNP data for *Medicago truncatula* A17 from the Legume Information System (legumeinfo.org, (Dash et al., 2016). The SNP data were initially filtered using bcftools v1.17 (Li, 2011) to include only biallelic SNPs with a minor allele frequency greater than 0.05. We then filtered the remaining SNPs in plink v1.90b6.21 (Purcell et al., 2007) to a missingness threshold of 0.2. We included a standardized K-matrix (r^2 >= 0.3) for the GWAS to reduce the effects of unequal relatedness among accessions. *P*-values for each SNP were calculated with a likelihood ratio test, and we used a *p*-value cut-off of $1x10^{-7.}$ Lastly, we used bedtools to find annotated genes closest to each candidate SNP in the Mt5.0 reference genome (Table S2).

Select and Resequence experiment to examine strain enrichment in small vs. large nodules: We chose two divergent host genotypes that represented the extremes of the variation in nodule morphology, and have intermediate host selectivity values, A17 and R108, which are also genetic models in this system. To estimate the fitness enrichment that individual rhizobia strains receive from these hosts, we inoculated both accessions with a 68-strain community of *S meliloti*. Most strains were from 101 strains obtained from the US Department of Agriculture (USDA) strain collection and used to investigate host x strain interactions in nodulation (Burghardt et al., 2018, 2020). Notably, this strain community represents an independent set of strains from those used in the previous experiment.

We grew six replicates of each of the two host accessions. Each replicate comprised 10-12 plants grown in a 1L pot filled with Sunshine mix #5. Before planting, we disrupted seed coats with a razor blade, incubated the seeds on wet filter paper for two days at 4°C, and then left the seeds at room temperature overnight to germinate before planting. The rhizobia inoculum was formed by mixing equal volumes of 68 single-strain cultures grown in TY media for 72 hours at 30°C with shaking at 200 rpm. When seedlings were three days old, we inoculated each pot with approximately 10^8 rhizobia from the mixed inoculum diluted in 10 ml of 0.85% NaCl solution. Thereafter, we watered pots as needed and fertilized them once a week with 150 mL of N-free fertilizer (Bucciarelli et al., 2006). Ten weeks after planting, we harvested the plants. At harvest we removed nodules from each pot and divided them into three size classes (small, medium, and large) based on visual size estimates (Figure 1); for A17, small and large pools from each replicate contained the same number of nodules. However, for R108 plants, the small nodule pools contained 1.5-2x as many nodules for three of the six pots as the paired large nodule pools. This was because R108 formed far fewer nodules with many massive nodules, and we encountered limitations in our processing volumes (Figure 2a).

We characterized the small and large nodule pools in three ways: 1) To estimate the number of rhizobia released from a single representative nodule in each size class, we counted the colony-forming units (CFUs) from serial dilutions of a randomly chosen nodule from each size class from each replicate pot. We sterilized the surface of each nodule in 20% bleach for 90 seconds, added 1 ml of sterile 0.85% by volume NaCl solution, and crushed individual nodules with a pestle. We then used the nodule slurry to create three dilution series and plated 100 µL of each series on two replicate TY plates. After 48-72 hours at 30°C, we counted the colonies on each plate. 2) To assess nodule characteristics in each group,

we counted lobes, measured nodule length (mm) of 10 representative nodules from each size class, and counted and weighed each pool of nodules. 3) To determine the frequency of each rhizobial isolate in each nodule pool, we followed Burghardt et al.'s Select and Resequence method (2018). Briefly, we surface sterilized each nodule pool in 20% bleach for 90 seconds and crushed the nodules with a pestle in a 2 mL tube in 1 mL sterile 0.85% NaCl solution. We centrifuged the slurry at 200xg for 10 min to pellet heavier plant tissue and bacteriods and create a supernatant enriched for undifferentiated rhizobial bacteria. We then pelleted undifferentiated rhizobia via fast centrifugation (20,000xg for 4 min), removed the supernatant, and froze the pellets at -20°C. Following the manufacturer's instructions, we extracted DNA using Qiagen Plant DNEasy kits.

Sequencing and analysis of large and small nodule pools: The University of Minnesota Genomics Core sequenced DNA from all samples with an Illumina HiSeq 2500, generating paired-end 125 bp reads. We removed adapters and low-quality sequences by running TrimGalore! (v0.4.1; github.com/FelixKrueger/TrimGalore) with the following settings: –quality 30, –stringency 3, –length 100, –length_1 101, –length_2 101 -e 0.1. We mapped the trimmed reads to the *S. meliloti* USDA1106 reference genome with bwa mem v0.7.17 (Li and Durbin, 2009) with default options and then used HARP (Kessner et al., 2013) to estimate strain frequencies. SNPs segregating among the 68 strains were previously identified by sequencing each strain using Illumina (Burghardt et al., 2022). The Illumina reads were trimmed with TrimGalore! (v0.4.1, settings -e 0.1, –quality 30, –stringency 3, –length 126 or 70, –length_1 and –length_2 127 or 71; shorter lengths were used for Illumina GAIIx data, longer lengths were used for Illumina HiSeq 2500 data), aligned to USDA1106 with bwa mem (v0.7.17), and the SNPs identified with freebayes (v1.2.0-2), and filtering out sites with QUAL scores < 20. Accession numbers in File S4.

Statistical analysis of nodule morphologies, strain communities, strain enrichment, and host benefits: To assess nodule morphological (length, branching, weight) and population size differences across hosts and nodule size pools, we ran linear models (ANOVA) in R on log-transformed values for nodule length, lobes, and population sizes: lm (Trait ~ host*pool size). We measured strain frequencies in the initial community (median=0.0092, 5%-95% quantile: 0.0054-0.0154).) and large and small nodule pools for each replicate pot. We calculated strain relative fitness = $\log_2$(selected frequency/initial frequency) to control for slight differences in strain frequencies in the inoculum and normalize the data. To quantify and visualize multivariate shifts in strain fitness across nodule pools, we conducted a Redundancy Analysis using the 'rda' function in the 'vegan' R package (Oksanen et al., 2017), which fits a multivariate linear regression to centered and scaled data and uses PCA to decompose the principal axes of variation in the fitted parameters. The adjusted R^2 of the RDA model provides an estimate of the proportion of variance in relative fitness explained by the model predictor(s) and, when paired with a PERMANOVA ('anova' function, 999 permutations), provides an estimate of the proportion of variance, in relative fitness explained by each predictor. To assess shifts in host benefits of strains in large and small pools, we calculated the proportion of the potential benefit each host captured in each replicate of small vs large nodule pools inferred from the benefit each strain provides to the host. The benefit each strain provides to hosts came from published data on the size of plants when inoculated with single strains in nitrogen-free conditions (Burghardt et al., 2018). Using ANOVA, we tested for 1) enrichment above random assembly based on the initial inoculum and 2) differential enrichment across nodule sizes within a host. Lastly, we sought to determine the relationship between host benefit and strain rewards by using strain representation in small vs large nodule pools as a signal of host nodule investment. To measure differences in *relative strain enrichment in large nodules*, we summed the strain frequency in small and large nodules for each replicate pot. Then we divided the relative frequency in large nodules by the total value and took the median value across pots. A value of 0.50 means that the strain is equally

represented in small and large nodules. 0.99 means 99% of the total strain frequency was found in large nodules, and 0.01 means only 1% was in large nodules. To determine whether the strains enriched in large nodules are more beneficial to hosts, we ran a linear model to see if *host benefits provided by each strain* predicted strain enrichment in large nodules (strain enrichment ~ host biomass).

RESULTS:

Host selectivity was strongly heritable and associated with maintaining shorter nodules: After removing the confounding effect of nodule number on strain diversity, our estimates of host selectivity across the Medicago GWAS panel of 202 accessions ranged from -0.22 to 0.95, with a median of 0.14 (Figure 2a). Positive host selectivity values mean that host nodule pools were less diverse than expected if the host randomly formed associations with the 89 potential rhizobial strains. Most (94%) hosts were more selective than expected by chance, with a tail of more highly selective accessions.

We performed a GWAS to identify genomic regions associated with host selectivity, which had a strong genetic basis (broad sense $H^2 = 0.465$). Based on a likelihood ratio test, the top ten associations had *P* values less than 10^{-7}, with effect sizes ranging from 0.09 to 0.16. In addition to putative proteins and transporters, they tag interesting host genes, including leginsulin 37, MtZIP11- Zinc-Iron Permease 11, two SNPs nearby a pectate lyase gene, MtMATE38 -multidrug & toxic compound extrusion family, and a putative transcription initiation Spt4, spt4/RpoE2 zinc finger (Table S2).

Lastly, we tested if any nodule morphology traits were associated with host selectivity (Figure 2b, Table S1). Using a LASSO regression, we found shorter nodules were associated with greater host selectivity, explaining 7.5% of the variation (Figure 2c). Post hoc regressions indicated an intermediate optimum for this trait, suggesting that hosts that form very large or small nodules tend to be less selective. None of the other 29 nodule morphology traits, including nodule branching or variance in nodule size, were strongly associated with host selectivity. This finding suggests that host control of nodule elongation may be more important than overall differences in nodule size for host selectivity.

Focal host genotypes differed in their ability to enrich populations of beneficial strains: With the broad approach above, we cannot determine whether selective hosts are better or worse at increasing the fitness of beneficial strains of rhizobia via increasing populations in large nodules (Figure 1). To examine differential rewards to more beneficial rhizobia, we focused on two *Medicago* hosts (A17 and R108) with intermediate selectivity but substantial differences in nodule size. First, we characterized nodule morphologies and rhizobial populations to measure the magnitude of absolute fitness enrichment in small and large nodules (Figure 3a). Compared to small nodules, the large nodules from A17 were 7.4 times heavier ($P < 0.01$), had 1.2 times more lobes ($P < 0.01$), were 2.8 times longer ($P < 0.001$), and released approximately 6.4 times more rhizobia ($P < 0.05$). By contrast, for R108, large nodules were 13 times heavier, had 8.2 more lobes, were 2.5 times longer, and released approximately 14.4 times more viable rhizobia (all $P < 0.001$). The differences between A17 and R108 were particularly evident for large nodules; large nodules from R108 were heavier, had more lobes, and released far more rhizobia than large nodules from A17 (Figure 3b-e). The larger differences in rhizobia released from large versus small in R108 than A17 suggest that while both hosts could enrich beneficial strains, R108 has more *potential* than A17 based on the larger asymmetry in nodule rhizobial population sizes.

Second, for enrichment to occur, specific strains must be consistently enriched in large nodules compared to small nodules. We estimated the relative fitness of strains in the small and large nodule pools to measure the strength and consistency of host selection and the potential for host rewards. We expect 1) strains in small nodule pools to represent young and unrewarded nodules and 2) strains that derive more rewards from hosts to be overrepresented in larger nodules. A redundancy analysis (RDA) of strain relative fitness in large and small nodules in A17 and R108 revealed that host, size, and their

interaction explained ~42% of the variation in the composition of strain communities. Moreover, the interaction between Host and Size (P <.001, df=1) suggests that hosts enriched different strains in large nodules. Subsetting the data by host reveals that nodule size explained 41% (P =0.003, df=1) of the variation in strain communities in A17 and 17% in R108 (P =0.012, df=1). Host identity was a strong determinant of representation in the nodule communities, and small and large nodules consistently supported different strain communities across replicates (Figure 4a). Thus, we established the *potential* for adaptive host enrichment in both hosts.

Third, we examined strain communities in small nodules to assess the potential for pre-nodulation partner choice mechanisms to explain rhizobial fitness. Strains in small nodule pools are a combination of both young and unrewarded nodules. We found that small nodules support rhizobial communities that are predicted to be more beneficial than expected by chance associations for A17 but not R108; strains in small A17 nodules extract 55% of the potential benefit in the initial inoculum as compared to 35% expected by chance, while strains in small R108 nodules are not more beneficial to R108 than expected by chance 30% vs 33% (Figure 4b). These results suggest that pre-nodulation partner choice mechanisms can contribute to rhizobial strain success, at least for one host genotype.

Lastly, we focused on post-nodule formation selection mechanisms by comparing the host benefits of strains in large and small pools. We took two complementary approaches to test whether post-nodulation strain enrichment was adaptively applied (Figure 1). Consistent with the preferential enrichment of beneficial strains, particularly by A17, the estimated benefit of the nodule community increased by ~ 20% in large nodules than in small nodules in A17 but < 10% in R108 (Figure 4b, both P < 0.01, $P_{host*size}$ = 0.018). We then examined the relationship between strain enrichment in large nodules and the benefit plants obtained from each strain. Consistent with host rewarding more beneficial strains, this relationship was significantly positive for both hosts (Figure 4c, Table SX, r = 0.61, P < 0.0001), although it was weaker for the less selective host R108 (Figure 3c, r = 0.3, P = 0.07). Overall, our results suggest A17 was better than R108 at enriching the fitness of more beneficial strains.

DISCUSSION:

Mutualisms are expected to break down unless mutualistic partners limit the evolution of genotypes that reap benefits but do not reward mutualistic partners (Sachs and Simms, 2006; Heath and Tiffin, 2009; Jones et al., 2015). Here, we show ample genetic variation in host selectivity in a model legume species (Henry and Nevo, 2014) and identify candidate genes contributing to this variation. The annotated functions of these genes point towards the importance of symbiont access to Zinc, cell wall modification, DNA methylation, and under-characterized plant hormones. Further, we showed an association between nodule morphology and selectivity: On average, host accessions with shorter nodules tend to be more selective. To ask whether hosts differ in their ability to enrich beneficial rhizobia adaptively is much more challenging, so we focused on two host accessions with contrasting nodule morphology, A17 and R108. While both *Medicago* accessions were able to preferentially enrich the fitness of more beneficial symbionts by increasing the size of the nodules they inhabited, the A17 host with shorter nodules was more effective at preferentially rewarding beneficial strains via both pre- and post-nodule formation mechanisms. Our results suggest the importance and feasibility of leveraging natural genetic variation in host and symbionts to decipher the mechanisms underlying mutualism maintenance and potentially breeding for that ability (Henry and Bergelson, 2023).

Host selectivity is sometimes inferred directly from bacterial diversity in nodules (Epstein et al., 2023). This whole-plant diversity may be appropriate when understanding ecological and evolutionary shifts in rhizobia communities. However, when identifying mechanisms that affect the rhizobial diversity in nodules, host sampling processes can be confounding (e.g., hosts that make more nodules may have increased diversity due to increased sampling, not host control). After accounting for among-

host variation in nodule number, we uncovered heritable variation in host selectivity within *Medicago* and identified allelic variation associated with host selectivity.

These candidates differ from those highlighted in the diversity-based GWAS in Epstein et al. 2023, suggesting that we are focusing on a different trait. The top candidates point to interesting host genes, including MtLegin27, a leginsulin-related MtN11/16/17 family gene(de Bang et al., 2017; Boschiero et al., 2020). Not much is known about these secreted hormone-like cysteine-rich peptides, but three similar genes (Legin47, 31, 42) were recently shown to be upregulated 72 hours post inoculation with rhizobia (Gao et al., 2022). Another promising candidate, MtZIP11- Zinc-Iron Permease 11, is highly expressed in nodules and is a close homolog of MtZIP6 that specifically transports Zinc into rhizobia-infected host cells (Abreu et al., 2017). We also identified two SNPs segregating at high frequency near a putative pectate lyase gene (Xie et al., 2012; Liu et al., 2019; Su, 2023; Zhang et al., 2024) that increases in expression in the first three days after exposure to rhizobia (Pereira et al., 2024). MtMATE38 is a member of a mostly uncharacterized family of 70 multidrug & toxic compound extrusion transporters that are likely biologically important (e.g., transporting citric acid, flavonoids, and Aluminum) (Wang et al., 2017). Lastly, our results point towards a putative transcription initiation Spt4, spt4/RpoE2 zinc finger involved in RNA-directed methylation in *Arabidopsis* (Köllen et al., 2015). Together, these candidates point to potential mechanisms of host selectivity in *Medicago* beyond the frequently studied NCR peptides (de Bang et al., 2017; Montiel et al., 2017; Horváth et al., 2023). Figuring out how plants enrich beneficial bacteria is an area of active exploration (Henry and Bergelson, 2023) and may lead to exciting paths forward for stabilizing mutualisms in agricultural systems (Denison et al., 2003; Kiers and Denison, 2014; Busby et al., 2017).

Associating nodule morphological characteristics with host selectivity could point us towards developmental mechanisms underlying that selectivity. Only one of the 30 morphology traits we measured was strongly associated with host selectivity. Surprisingly, we found no evidence for our original hypothesis that hosts with more variance in nodule size or branched nodules would be more selective. Therefore, intra-plant variation in nodule morphology appears to be a poor predictor of the host's ability to reward beneficial strains, at least for *Medicago*. However, we did find that hosts with shorter nodules, on average, tended to be more selective, explaining about 7.5% of the variation in selectivity. While this isn't a huge amount of variation, it suggests that mechanisms by which hosts can control the size of symbiotic organs may be important, especially as recent research suggests that rhizobial bacteria can control nodule size via expression of plant hormones (Nett et al., 2022). More broadly, this research is a starting point for connecting organ form to function in symbiosis research and suggests that expanding research in this direction would be worthwhile.

Increased host selectivity can be adaptive or not, and figuring out how plants can or cannot preferentially enrich beneficial bacteria while repelling pathogens or cheaters is an area of active exploration (Henry and Bergelson, 2023). Although both host genotypes we tested were able to enrich beneficial rhizobia strains, our results make clear that the preferential rewarding of more beneficial symbionts is far from perfect, and both pre- and post-nodulation processes affect strain frequencies in nodules. This variation in strain-specific rewards is important as it may allow less beneficial strains to reap the reproductive benefits of symbiosis, thereby contributing to maintaining mediocre rhizobia strains in populations (Bever et al., 2009; Bever, 2015). More generally, our results suggest pairing assessments of single-strain benefits with assessments of nodule strain competitive fitness can identify host genotypes that best maintain mutualistic rhizobia populations in the face of diverse rhizobial populations in natural and agricultural soils (Mendoza-Suárez et al., 2020; Batstone et al., 2022; Burghardt and diCenzo, 2023).

ACKNOWLEDGEMENTS:

This work was supported by the National Science Foundation grants IOS-1856744 and 2243819 to Burghardt and Tiffin and completed through computing resources provided by the Minnesota Supercomputing Institute (MSI) at the University of Minnesota. LTB was partly supported by USDA NIFA Grant 2022-67013-36860 and This work is/was supported by the USDA National Institute of Food and Agriculture and Hatch Appropriations under Project #PEN04760 and Accession #1025611. We thank Adam Kostanecki for taking the nodule pictures, members of the Tiffin and Heath Labs for helping with the harvesting, and Burghardt Lab members for feedback that improved the manuscript.

DATA AVAILABILITY:

All code, data sets, and results are available on Burghardt Lab Github Repository:

REFERENCES:

Abreu, I., Á. Saéz, R. Castro-Rodríguez, V. Escudero, B. Rodríguez-Haas, M. Senovilla, C. Larue, et al. 2017. Zinc-Iron Permease6 provides zinc to rhizobia-infected nodule cells. *Plant, Cell & Environment* 40: 2706–2719.

de Bang, T. C., P. K. Lundquist, X. Dai, C. Boschiero, Z. Zhuang, P. Pant, I. Torres-Jerez, et al. 2017. Genome-Wide Identification of Medicago Peptides Involved in Macronutrient Responses and Nodulation1[OPEN]. *Plant Physiology* 175: 1669–1689.

Batstone, R. T., L. T. Burghardt, and K. D. Heath. 2022. Phenotypic and genomic signatures of interspecies cooperation and conflict in naturally occurring isolates of a model plant symbiont. *Proceedings of the Royal Society B: Biological Sciences* 289: 20220477.

Bever, J. D. 2015. Preferential allocation, physio-evolutionary feedbacks, and the stability and environmental patterns of mutualism between plants and their root symbionts. *NEW PHYTOLOGIST* 205: 1503–1514.

Bever, J. D., S. C. Richardson, B. M. Lawrence, J. Holmes, and M. Watson. 2009. Preferential allocation to beneficial symbiont with spatial structure maintains mycorrhizal mutualism. *ECOLOGY LETTERS* 12: 13–21.

Boschiero, C., X. Dai, P. K. Lundquist, S. Roy, T. Christian de Bang, S. Zhang, Z. Zhuang, et al. 2020. MtSSPdb: The Medicago truncatula Small Secreted Peptide Database1 [OPEN]. *Plant Physiology* 183: 399–413.

Brewin, N. J. 1991. Development of the Legume Root Nodule. *Annual Review of Cell Biology* 7: 191–226.

Bronstein, J. L. ed. . 2015. Mutualism. First edition. Oxford University Press, Oxford, United Kingdom.

Bucciarelli, B., J. Hanan, D. Palmquist, and C. P. Vance. 2006. A Standardized Method for Analysis of Medicago truncatula Phenotypic Development. *Plant Physiology* 142: 207.

Bull, J. J., and W. Rice. 1991. Distinguishing mechanisms for the evolution of cooperation. *Journal of Theoretical Biology* 149: 63–74.

411 Burghardt, L., and G. C. diCenzo. 2023. The evolutionary ecology of rhizobia: multiple facets of
412 competition before, during, and after symbiosis with legumes. *CURRENT OPINION IN*
413 *MICROBIOLOGY* 72.

414 Burghardt, L. T., B. Epstein, J. Guhlin, M. S. Nelson, M. R. Taylor, N. D. Young, M. J. Sadowsky, and P.
415 Tiffin. 2018. Select and resequence reveals relative fitness of bacteria in symbiotic and free-
416 living environments. *Proceedings of the National Academy of Sciences* 115: 2425.

417 Burghardt, L. T., B. Epstein, M. Hoge, D. I. Trujillo, and P. Tiffin. 2022. Host-associated rhizobial fitness:
418 dependence on nitrogen, density, community complexity, and legume genotype J. B. Glass
419 [ed.],. *Applied and Environmental Microbiology* 88: e00526-22.

420 Burghardt, L. T., D. I. Trujillo, B. Epstein, P. Tiffin, and N. D. Young. 2020. A select and resequence
421 approach reveals strain-specific effects of Medicago nodule-specific PLAT-domain genes. *Plant*
422 *physiology* 182: 463–471.

423 Busby, P. E., C. Soman, M. R. Wagner, M. L. Friesen, J. Kremer, A. Bennett, M. Morsy, et al. 2017.
424 Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLOS Biology* 15:
425 e2001793.

426 Dash, S., J. D. Campbell, E. K. S. Cannon, A. M. Cleary, W. Huang, S. R. Kalberer, V. Karingula, et al. 2016.
427 Legume information system (LegumeInfo.org): a key component of a set of federated data
428 resources for the legume family. *Nucleic Acids Research* 44: D1181–D1188.

429 Daubech, B., P. Remigi, G. Doin De Moura, M. Marchetti, C. Pouzet, M.-C. Auriac, C. S. Gokhale, et al.
430 2017. Spatio-temporal control of mutualism in legumes helps spread symbiotic nitrogen fixation.
431 *eLife* 6: e28683.

432 Denison, R. F. 2000. Legume sanctions and the evolution of symbiotic cooperation by
433 rhizobia. *The American Naturalist* 156: 567–576.

434 Denison, R. F., E. T. Kiers, and S. A. West. 2003. Darwinian Agriculture: When Can Humans Find Solutions
435 Beyond the Reach of Natural Selection? *The Quarterly Review of Biology* 78: 145–168.

436 Epstein, B., L. T. Burghardt, K. D. Heath, M. A. Grillo, A. Kostanecki, T. Hämälä, N. D. Young, and P. Tiffin.
437 2023. Combining GWAS and population genomic analyses to characterize coevolution in a
438 legume-rhizobia symbiosis. *Molecular Ecology* 32: 3798–3811.

439 Friesen, M. L. 2012. Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytologist*
440 194: 1096–1111.

441 Gao, Y., B. Selee, E. L. Schnabel, W. L. Poehlman, S. A. Chavan, J. A. Frugoli, and F. A. Feltus. 2022. Time
442 Series Transcriptome Analysis in Medicago truncatula Shoot and Root Tissue During Early
443 Nodulation. *Frontiers in Plant Science* 13: 861639.

444 Heath, K. D., and P. Tiffin. 2009. STABILIZING MECHANISMS IN A LEGUME-RHIZOBIUM MUTUALISM.
445 *Evolution* 63: 652–662.

Henry, L. P., and J. Bergelson. 2023. Evolutionary implications of host genetic control for engineering beneficial microbiomes. *Current Opinion in Systems Biology* 34: 100455.

Henry, R. J., and E. Nevo. 2014. Exploring natural selection to guide breeding for agriculture. *Plant Biotechnology Journal* 12: 655–662.

Horváth, B., B. Güngör, M. Tóth, Á. Domonkos, F. Ayaydin, F. Saifi, Y. Chen, et al. 2023. The *Medicago truncatula* nodule-specific cysteine-rich peptides, NCR343 and NCR-new35 are required for the maintenance of rhizobia in nitrogen-fixing nodules. *New Phytologist* 239: 1974–1988.

Jones, E. I., M. E. Afkhami, E. Akçay, J. L. Bronstein, R. Bshary, M. E. Frederickson, K. D. Heath, et al. 2015. Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters* 18: 1270–1284.

Kessner, D., T. L. Turner, and J. Novembre. 2013. Maximum Likelihood Estimation of Frequencies of Known Haplotypes from Pooled Sequence Data. *Molecular Biology and Evolution* 30: 1145–1158.

Kiers, E. T., and R. F. Denison. 2014. Inclusive fitness in agriculture. *Philosophical Transactions of the Royal Society B: Biological Sciences* 369: 20130367.

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

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

Köllen, K., L. Dietz, N. Bies-Etheve, T. Lagrange, M. Grasser, and K. D. Grasser. 2015. The zinc-finger protein SPT4 interacts with SPT5L/KTF1 and modulates transcriptional silencing inArabidopsis. *FEBS Letters* 589: 3254–3257.

Li, H. 2011. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* 27: 2987–2993.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25: 1754–1760.

Liu, C.-W., A. Breakspear, N. Stacey, K. Findlay, J. Nakashima, K. Ramakrishnan, M. Liu, et al. 2019. A protein complex required for polar growth of rhizobial infection threads. *Nature Communications* 10: 2848.

Łotocka, B., J. Kopcińska, and M. Skalniak. 2012. Review article: The meristem in indeterminate root nodules of Faboideae. *Symbiosis* 58: 63–72.

Mendoza-Suárez, M. A., B. A. Geddes, C. Sánchez-Cañizares, R. H. Ramírez-González, C. Kirchhelle, B. Jorrin, and P. S. Poole. 2020. Optimizing Rhizobium-legume symbioses by simultaneous

481 measurement of rhizobial competitiveness and N2 fixation in nodules. *Proceedings of the*
482 *National Academy of Sciences* 117: 9822.

483 Montiel, J., J. A. Downie, A. Farkas, P. Bihari, R. Herczeg, B. Bálint, P. Mergaert, et al. 2017. Morphotype
484 of bacteroids in different legumes correlates with the number and type of symbiotic NCR
485 peptides. *Proceedings of the National Academy of Sciences of the United States of America* 114:
486 5041–5046.

487 Montoya, A. P., C. E. Wendlandt, A. B. Benedict, M. Roberts, J. Piovia-Scott, J. S. Griffitts, and S. S. Porter.
488 2023. Hosts winnow symbionts with multiple layers of absolute and conditional discrimination
489 mechanisms. *Proceedings of the Royal Society B: Biological Sciences* 290: 20222153.

490 Nett, R. S., K. S. Bender, and R. J. Peters. 2022. Production of the plant hormone gibberellin by rhizobia
491 increases host legume nodule size. *The ISME Journal* 16: 1809–1817.

492 Oksanen, J., F. Blancet, R. Kindt, P. Legendre, P. Minchin, R. O'hara, G. Simpson, et al. 2017. Package
493 'vegan': Community ecology package. *R package version*.

494 Oldroyd, G. E. D., J. D. Murray, P. S. Poole, and J. A. Downie. 2011. The rules of engagement in the
495 legume-rhizobial symbiosis. *Annual Review of Genetics* 45: 119–144.

496 Pereira, W. J., J. Boyd, D. Conde, P. M. Triozzi, K. M. Balmant, C. Dervinis, H. W. Schmidt, et al. 2024. The
497 single-cell transcriptome program of nodule development cellular lineages in Medicago
498 truncatula. *Cell Reports* 43: 113747.

499 Porter, S. S., S. E. Dupin, R. F. Denison, E. T. Kiers, and J. L. Sachs. 2024. Host-imposed control
500 mechanisms in legume–rhizobia symbiosis. *Nature Microbiology* 9: 1929–1939.

501 Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, et al. 2007. PLINK:
502 A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American*
503 *Journal of Human Genetics* 81: 559–575.

504 Quides, K. W., F. Salaheldine, R. Jariwala, and J. L. Sachs. 2021. Dysregulation of host-control causes
505 interspecific conflict over host investment into symbiotic organs. *Evolution* 75: 1189–1200.

506 Quides, K. W., G. M. Stomackin, H.-H. Lee, J. H. Chang, and J. L. Sachs. 2017. Lotus japonicus alters in
507 planta fitness of Mesorhizobium loti dependent on symbiotic nitrogen fixation U. Melcher [ed.],.
508 *PLOS ONE* 12: e0185568.

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

512 Sachs, J. L., M. O. Ehinger, and E. L. Simms. 2010. Origins of cheating and loss of symbiosis in wild
513 Bradyrhizobium. *JOURNAL OF EVOLUTIONARY BIOLOGY* 23: 1075–1089.

514 Sachs, J., L., U. Mueller G., T. P. Wilcox, and J. J. Bull. 2004. The evolution of cooperation. *Quarterly*
515 *Review of Biology* 79: 135–160.

516 Sachs, J. L., K. W. Quides, and C. E. Wendlandt. 2018. Legumes versus rhizobia: a model for ongoing
517 conflict in symbiosis. *New Phytologist* 219: 1199–1206.

518 Sachs, J. L., and E. L. Simms. 2006. Pathways to mutualism breakdown. *Trends in Ecology & Evolution* 21:
519 585–592.

520 Sprent, J. I., J. Sutherland, and S. Defaria. 1987. Some aspects of the biology of nitrogen-fixing
521 organisms. *Philosophical Transations of the Royal Society B-Biological Sciences* 317: 111–129.

522 Su, C. 2023. Pectin modifications at the symbiotic interface. *New Phytologist* 238: 25–32.

523 Underwood, T. J., B. Jorrin, L. A. Turnbull, and P. S. Poole. 2024. Pisum sativum applies sanctions to less
524 effective strains at the nodule level, but not within mixed nodules. 2024.04.25.582971.

525 Wang, J., Q. Hou, P. Li, L. Yang, X. Sun, V. A. Benedito, J. Wen, et al. 2017. Diverse functions of multidrug
526 and toxin extrusion (MATE) transporters in citric acid efflux and metal homeostasis in Medicago
527 truncatula. *The Plant Journal: For Cell and Molecular Biology* 90: 79–95.

528 Wendlandt, C. E., J. U. Regus, K. A. Gano-Cohen, A. C. Hollowell, K. W. Quides, J. Y. Lyu, E. S. Adinata, and
529 J. L. Sachs. 2019. Host investment into symbiosis varies among genotypes of the legume
530 Acmispon strigosus, but host sanctions are uniform. *New Phytologist* 221: 446–458.

531 Westhoek, A., L. J. Clark, M. Culbert, N. Dalchau, M. Griffiths, B. Jorrin, R. Karunakaran, et al. 2021.
532 Conditional sanctioning in a legume– *Rhizobium* mutualism. *Proceedings of the National*
533 *Academy of Sciences* 118: e2025760118.

534 Wilkinson, D., and T. Sherratt. 2001. Horizontally acquired mutualisms, an unsolved problem in ecology?
535 *OIKOS* 92: 377–384.

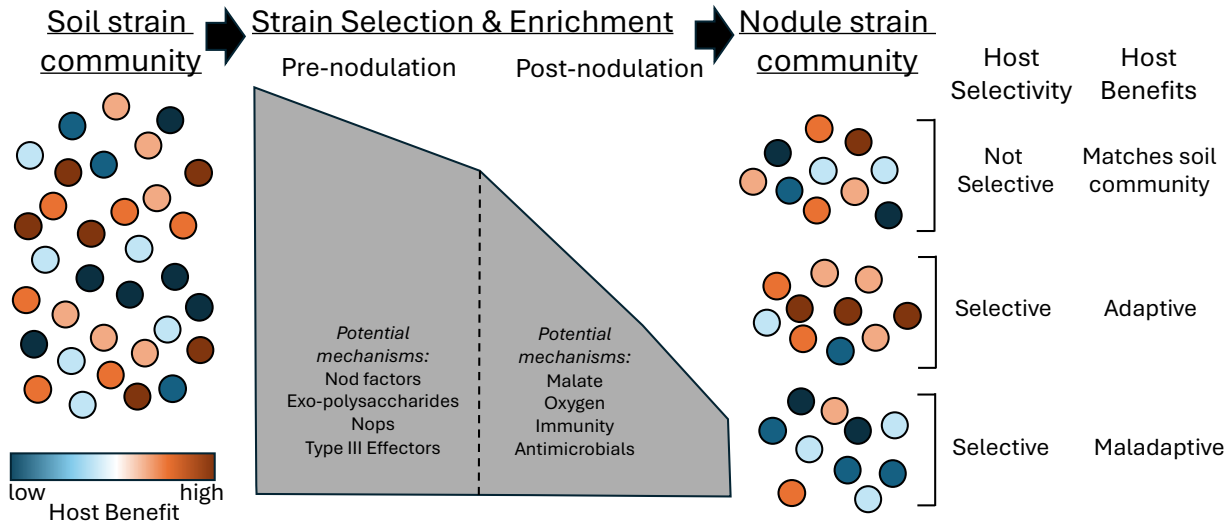
536 Xie, F., J. D. Murray, J. Kim, A. B. Heckmann, A. Edwards, G. E. D. Oldroyd, and J. A. Downie. 2012.
537 Legume pectate lyase required for root infection by rhizobia. *Proceedings of the National*
538 *Academy of Sciences of the United States of America* 109: 633–638.

539 Zhang, X., J.-X. Chen, W.-T. Lian, H.-W. Zhou, Y. He, X.-X. Li, and H. Liao. 2024. Molecular module
540 GmPTF1a/b-GmNPLa regulates rhizobia infection and nodule formation in soybean. *New*
541 *Phytologist* 241: 1813–1828.

542 Zhou, X., and M. Stephens. 2012. Genome-wide efficient mixed-model analysis for association studies.
543 *Nature Genetics* 44: 821–824.

544

Figure 1: Rhizobial Selection and Enrichment in Legume Nodules. Determining how and if hosts can choose helpful partners remains an essential but elusive question in mutualism research (Porter et. al, 2024). Here, we examine host genetic variation in selectivity in 202 accessions of *Medicago truncatula* and then focus on assessing the two hosts' ability to adaptively enrich the fitness of beneficial strains by increasing the size of nodules and, thus, the population size of those strains.



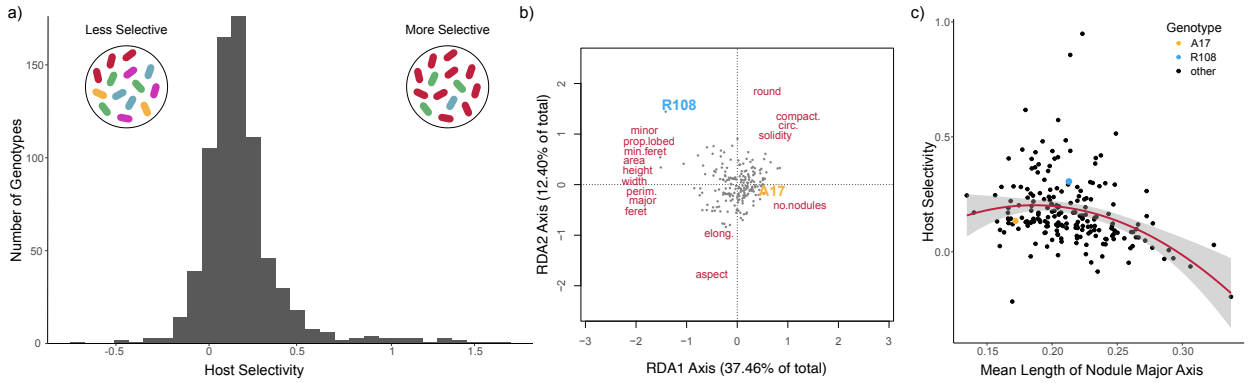


Figure 2: Host Selectivity and Nodule Morphology. For 202 accessions of *Medicago truncatula,* we calculated a novel host selectivity by controlling for nodule number differences among plants (a) and quantified nodule morphology variation in nodule pools (16 trait means are visualized here with an RDA). We used LASSO regression with $\lambda = 0.0104$ and determined the mean length of the major axis of nodules to be most strongly associated ($p = 4.2 \times 10^{-6}$, adj. $R^2 = 0.11$) with host selectivity (c). Lastly, we used GWAS to identify host candidate genes associated with host selectivity (Table S2).

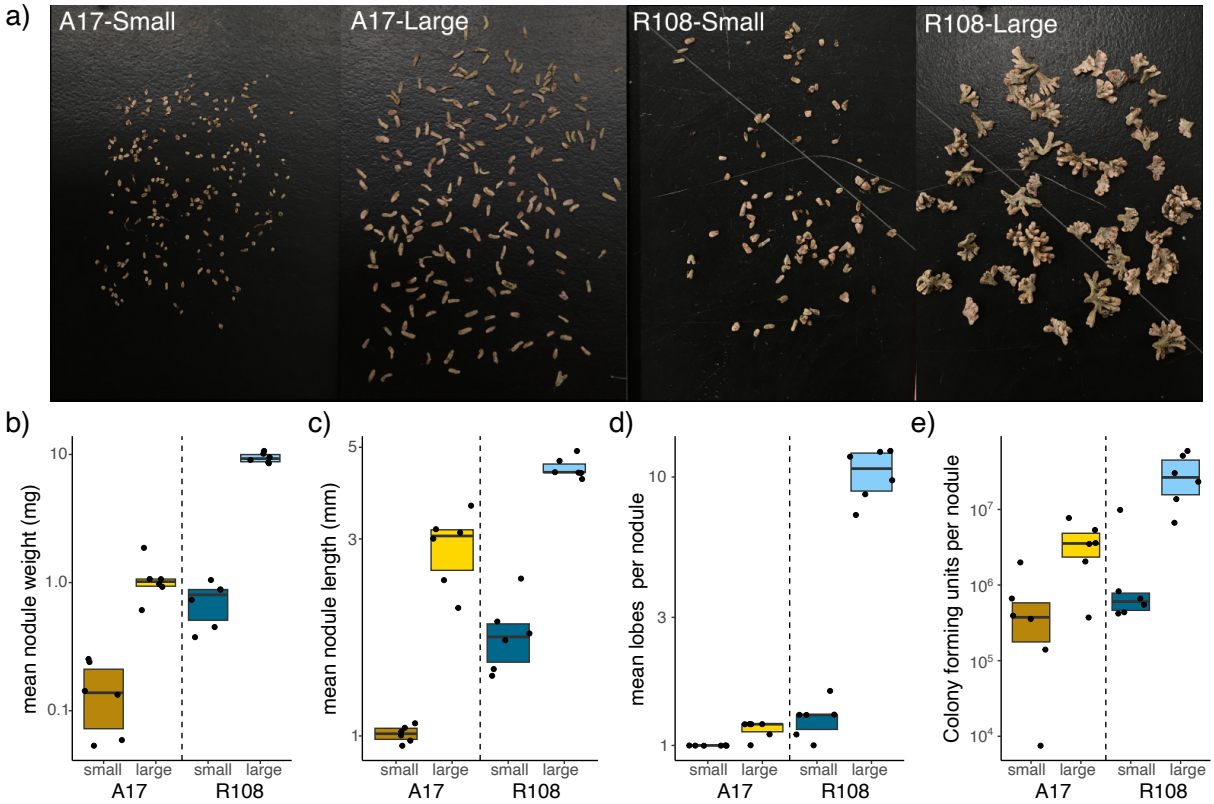


Figure 3 Nodule characteristics in two focal hosts, A17 and R108: a) Examples of single replicates of large and small nodule pools from our two host genotypes. (b-e) Nodule Characteristics (weight, length, lobes, colony forming units) of small and large nodules from A17 and R108 hosts. Each data point represents a replicate. The average weight per nodule is based on the fresh weight of the entire pool of nodules, and colony data is the average of replicate dilutions from haphazardly selected nodules from each class. All traits are on a log10 scale. All traits differed ($p<0.05$) between hosts and nodule size classes based on linear regressions and ANOVA (Table S3).

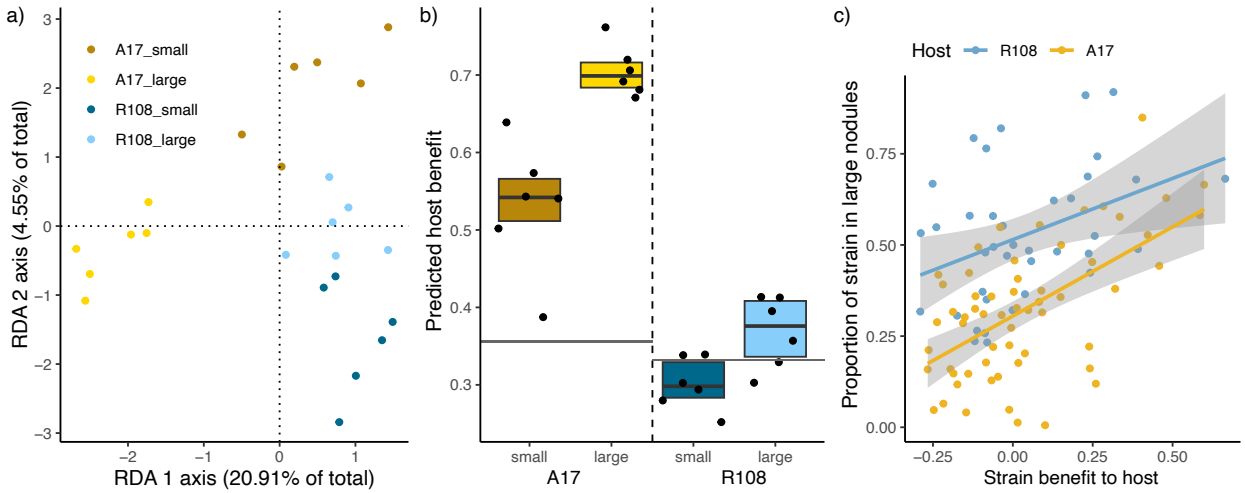


Figure 4: **Adaptive strain enrichment in two model hosts**. a) Redundancy analysis (RDA) of strain fitness in large and small nodule pools in A17 and R108 revealed that host, size, and their interaction account for approximately 41% of the variation in strain communities. The interaction between host and size (p<.001, df=1) indicated that hosts enriched different strains in larger nodules. Note that constrained axis RDA 1, which differentiates large and small A17 nodules, explains 4-5 times more variation than constrained axis 2. b) The predicted benefit of the rhizobia found in nodules differs between hosts and nodule size (p=0.024, df=1). The predicted benefit is estimated by multiplying each strain's benefit to hosts in a single-strain experiment by its frequency in the nodule pool and summing this across all strains. Since hosts vary in overall size, this sum is scaled such that one represents a nodule community composed entirely of the most beneficial strain, while zero signifies a community composed of the least beneficial strain. c) The relationship between the rewards rhizobia receive and the benefit those strains provide to plant hosts A17 and R108 during single-strain inoculations. The rewards received are measured by their proportional representation in large nodule pools versus the total representation in large and small pools. Each dot represents the median strain proportion across six replicate pots. Single strain plant benefit predicted the enrichment of a strain in larger nodules for both hosts but explained more variation in A17 (34%) than in R108 (15%). Benefits for b and c were estimated by measuring plant biomass in a single-strain inoculation experiment under nitrogen-free conditions (See Tables S4-S6 for complete statistical results).

588 **Table S1:** Nodule traits measured from 4-5 replicate pots of 202 Medicago accessions inoculated with 89
589 strains of *Sinorhizobium meliloti* (Epstein et al. 2024). Traits with abbreviations are visualized in Fig. 2b.

Nodule Trait	Abr.	Description	CoV
Mean Perimeter	perim.	Mean number of pixels in the boundary of a nodule	0.0357
Mean AR	aspect	Mean Aspect Ratio of the bounding box constraining nodules	0.0248
Mean Feret	feret	Mean 'caliper diameter' / distance between two parallel lines constraining nodules	0.0084
Mean Major	major	Mean length of the longest line that can be drawn through nodules	0.0050
Mean Width	width	Mean width of bounding boxes constraining nodules	0.0066
Mean Height	height	Mean height of bounding boxes constraining nodules	0.0064
Mean Compactness	compact.	Mean ratio of the area of nodules and the area of a circle with the same perimeter	0.0125
Mean Circularity	circ.	$4\pi * \frac{Area}{Perimeter^2}$	0.0124
Mean Roundness	round	$4 * \frac{Area}{\pi * Major^2}$	0.0011
Mean MinFeret	min.feret	Mean of the smallest possible 'caliper diameter' / distance between two parallel lines constraining nodules	0.0042
Mean Area	area	Mean number of pixels in a nodule	0.0033
Mean Minor	minor	Mean length of the shortest line that can be drawn through a nodule	0.0017
Mean Elongation	elong.	Mean ratio of the width and length of the bounding box constraining nodules	0.0024
Mean Solidity	solidity	Mean ratio of the area and convex area of nodules	0.0014
Perimeter Variance	-	Variance of Perimeter within nodule images	0.1307
AR Variance	-	Variance of AR within nodule images	0.0894
Elongation Variance	-	Variance of Elongation within nodule images	0.0213
Feret Variance	-	Variance of Feret length within nodule images	0.0058
Variance Height	-	Variance of Height within nodule images	0.0042
Variance Major	-	Variance of Major within nodule pool images	0.0040
Variance Width	-	Variance of Width within nodule images	0.0041
Variance MinFeret	-	Variance of MinFeret within nodule images	0.0042
Variance Minor	-	Variance of Minor within nodule pool images	0.0017
Variance Compactness	-	Variance of Compactness within nodule pool images	0.0018
Variance Area	-	Variance of Area of nodules in nodule pool images	0.0014
Variance Circularity	-	Variance of circularity within nodule images	0.0016
Variance Roundness	-	Variance of Roundness within nodule images	0.0011
Variance Solidity	-	Variance of Solidity within nodule images	0.0010
Predicted Proportion Lobed	prop.lobed	Modelled	0.0387
Nodule Number	no.nodules	Number of nodules within nodule images	37.0

590

591

592 **Table S2**: Top 10 candidate SNP variants for host selectivity selected based on *p*-values from likelihood
593 ratio tests. MAF is the Minor allele frequency. Beta is the effect size from the genome-wide association.

ChromosomeID-Position-BaseChange	rank	Chr	-log10(*p*)	MAF	beta	Closest gene in the Medicago V5 genome
CM010654.1-26952852-C	1	7	9.365	0.05	0.148	MtrunA17_Chr7g0236791- MtLegin27- Leginsulin
CM010650.1-51530113-A	2	3	8.299	0.06	0.124	MtrunA17_Chr3g0135691- MtZIP11- Zinc-Iron Permease 11
CM010651.1-51084261-T	3	4	7.973	0.31	0.111	MtrunA17_Chr4g0058051- putative pectate lyase
CM010653.1-25128406-A	4	6	7.660	0.05	0.156	MtrunA17_Chr6g0472811- snoR71-ncRNA
CM010651.1-51080330-T	5	4	7.586	0.26	0.112	MtrunA17_Chr4g0058051- putative pectate lyase
CM010654.1-23572058-T	6	7	7.563	0.06	0.114	MtrunA17_Chr7g0233111- putative protein
CM010653.1-32706149-T	7	6	7.556	0.08	0.105	MtrunA17_Chr6g0477961- MtMATE38 -multidrug & toxic compound extrusion family
CM010650.1-45833295-C	8	3	7.405	0.06	0.129	MtrunA17_Chr3g0127911- Putative transcription initiation Spt4, spt4/RpoE2 zinc finger
CM010649.1-16046163-T	9	2	7.329	0.12	0.093	MtrunA17_Chr2g0298091- putative transferase
CM010654.1-23572805-T	10	7	7.046	0.07	0.108	MtrunA17_Chr7g0233111- putative protein

594

Table S3: ANOVA results for linear models of individual nodule traits from sequenced pools (Figure 3).

Trait	Predictor	df	Prop. Variance	F value	*P* value	Model R^2 adj
Nodule Weight	**host**	1	0.379	11d	*0****	0.92
	size	1	0.553	168	*0****	
	replicate	5	0.013	0.76	0.59	
	host*size	1	0.006	1.96	0.18	
	residual	15	0.049	NA	NA	
Nodule length	**host**	1	0.192	75	*0****	0.94
	size	1	0.75	291	*0****	
	replicate	5	0.018	1.42	0.27	
	host*size	1	0.002	0.64	0.44	
	residual	15	0.038	NA	NA	
Nodule lobes	**host**	1	0.389	460	*0****	0.98
	size	1	0.335	394	*0****	
	replicate	5	0.006	1.4	0.28	
	host*size	1	0.257	304	*0****	
	residual	15	0.013	NA	NA	
Colony forming units	**host**	1	0.179	8.66	*0.01**	0.52
	size	1	0.474	22.9	*0****	
	replicate	5	0.025	0.24	0.94	
	host*size	1	0.011	0.52	0.48	
	residual	15	0.311	NA	NA	

597 **Table S4:** PERMANOVA results from RDA analysis of strain communities in nodules pools (Figures 4a and
598 S6).

Models	predictor	df	Prop Variance	F value	*P* value	R^2 adj
Both Hosts	**host**	1	0.18	6.10	*0.001***	0.34
	size	1	0.13	4.56	*0.001***	
	Host*size	1	0.12	4.03	*0.002***	
	residual	20	0.58	NA	NA	
A17 only	**size**	1	0.41	6.98	*0.004***	0.35
	residual	10	0.59	NA	NA	
R108 only	**size**	1	0.17	2.02	*0.016**	0.085
	residual	10	**0.83**	NA	NA	

599

600 **Table S5:** Statistical results for predicted host benefit of each strain community inferred from previous
601 single-strain experiments (Figure 4b)

Model	predictor	df	Prop Variance	F value	*P* value	R^2 adj
Host Benefit	**host**	1	0.75	169	*0****	0.90
	size	1	0.14	30.8	*0****	
	Host*Size	1	0.027	6	*0.024**	
	residual	20	0.089	NA	NA	

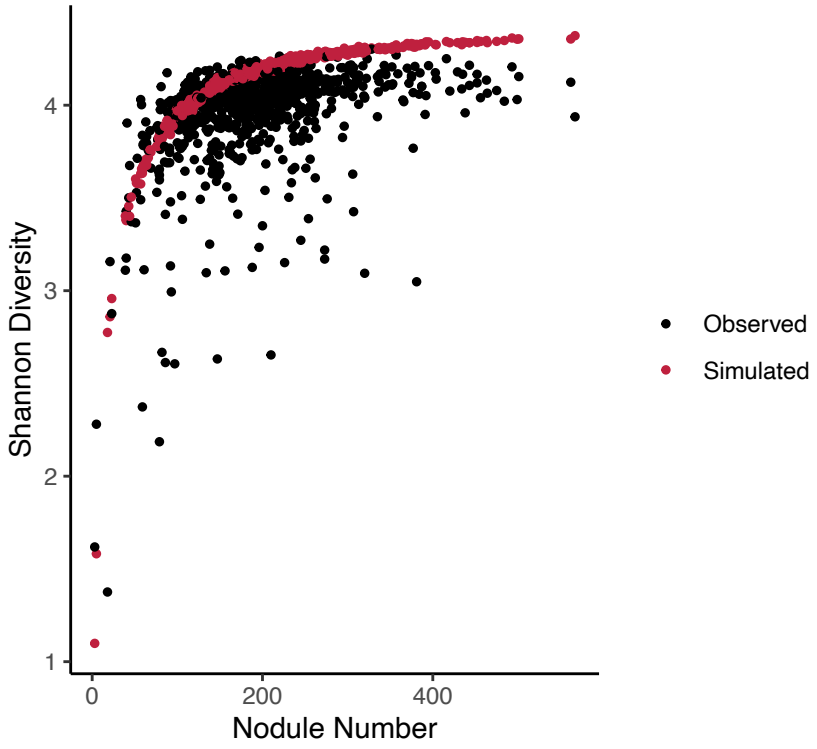
602

603 **Table S6:** ANOVA results from linear models of strain relative representation in large vs. small nodules
604 (Figure 4c).

Model	predictor	df	Prop Variance	F value	*P* value	R^2 adj
Both Hosts	**host**	1	0.20	33.5	*0****	0.42
	benefit	1	0.23	38.2	*0****	
	Host*benefit	1	0.006	0.097	0.32	
	residual	93	0.56	NA	NA	
A17 only	**benefit**	1	0.34	29.7	*0****	0.33
	residual	57	0.66	NA	NA	
R108 only	**benefit**	1	0.15	6.4	*.016**	0.13
	residual	36	0.85	NA	NA	

605

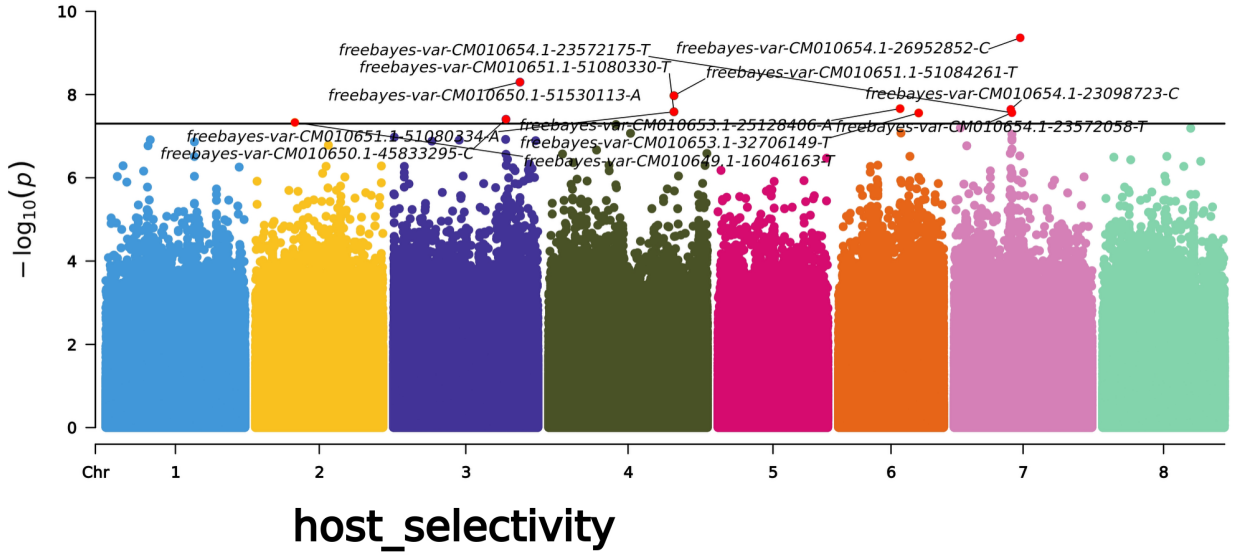
606 **Figure S1:** Relationship between nodule number and Shannon diversity for simulations and empirical
607 data. Red dots are the average of 10 simulations of neutral nodule assembly at each empirically
608 observed nodule number (red dot). Black dots represent empirical data on nodule Shannon diversity
609 estimated for each of the four replicate pots of 202 Medicago accessions. 'Host selectivity' is estimated
610 by subtracting the observed value from the simulated value for a given nodule number.

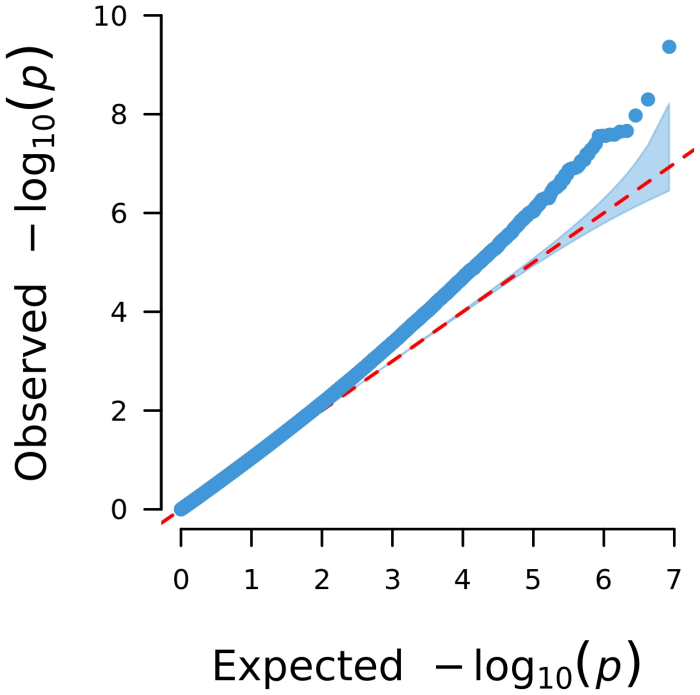


611

612

Figure S2: Visualization of results from the genome-wide association analysis of the host selectivity traits Manhattan plot (top) showing the 10 candidate variants with *p* values less than $1x10^{-7}$. QQ plot (bottom) of expected vs. observed p-value distribution. Candidate information is in Table S2.





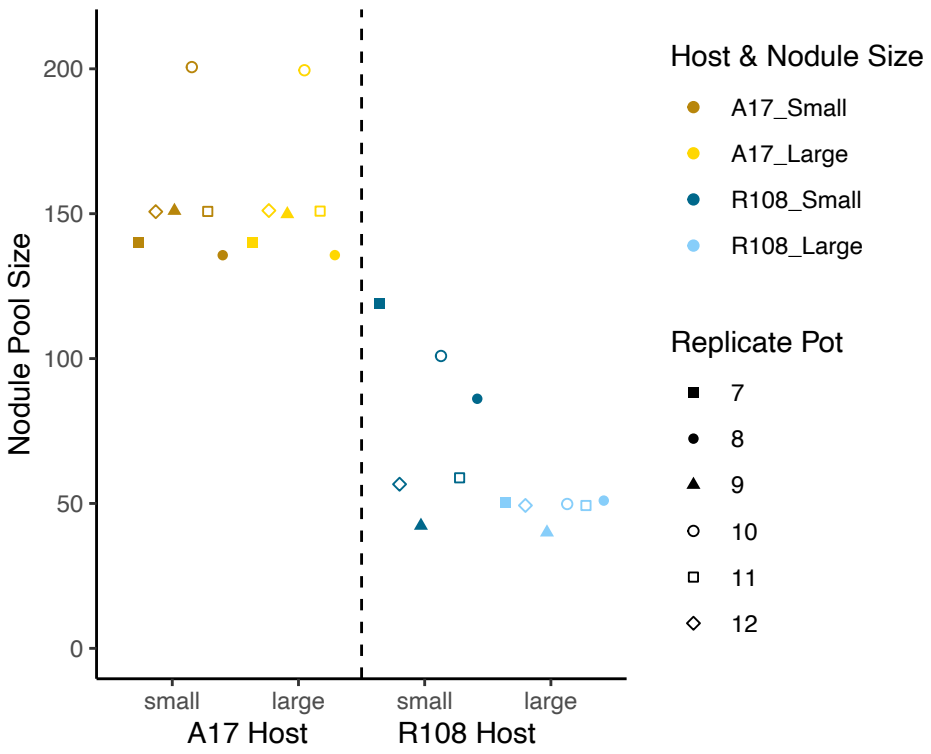


Figure S3: Size of Sequenced Nodule Pools. A17 nodules have identical nodule pool sizes, whereas 3 of the R108 pools had smaller nodules in the large pool due to processing volume and nodule size limitations. However, this did not affect Shannon diversity (Figure S4), suggesting nodule pools were saturated.

Figure S4: Strain diversity in Sequenced Nodule Pools. Note that differences in nodule pool size in Figure S3 for pots 7, 8, and 10 do not alter strain diversity patterns. This suggests that the practical size limitation in half of the large nodule pools for R108 is not affecting our inferences.

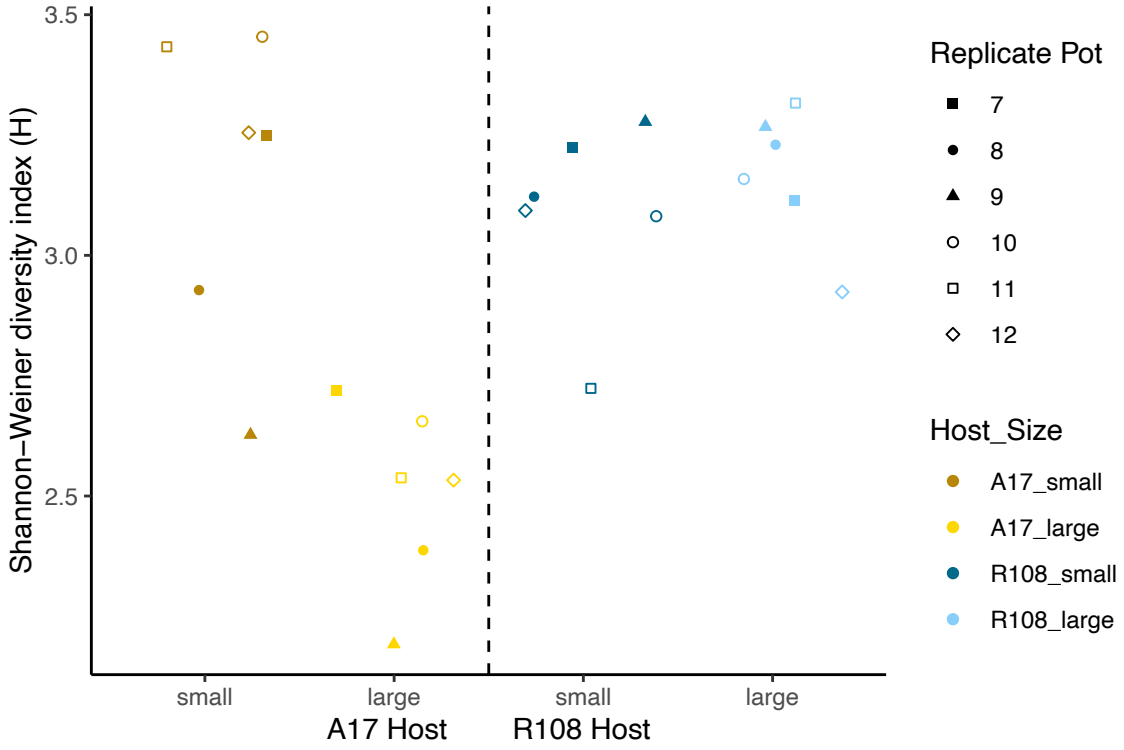
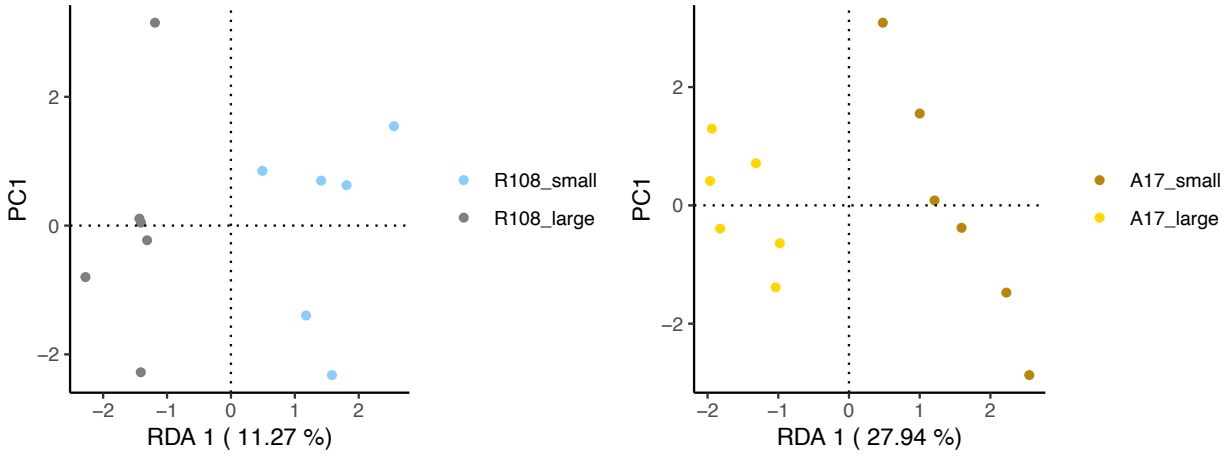



Figure S5: RDA analysis conducted independently for large and small nodule strain communities in A17 (left) and R108 (right). See Table S6 for statistical results.