Title: Mutations in legume symbiosis genes create a complex selective landscape for rhizobial symbionts

Sohini Guha[1], Regina B. Bledsoe[1], Jeremy Sutherland[1], Brendan Epstein[2], Gwendolyn M. Fry[1], Nevin D. Young[3], Peter Tiffin[2], and Liana T. Burghardt[1*]

[1]The Pennsylvania State University, Department of Plant Science, University Park, Pennsylvania
[2]University of Minnesota, Department of Plant and Microbial Biology, St. Paul, Minnesota
[3]University of Minnesota, Department of Plant Pathology, St. Paul, Minnesota
[*]Corresponding author (liana.burghardt@gmail.com)

Author for correspondence: Liana T. Burghardt, Department of Plant Science, 109 Tyson Building, University Park, PA 16802, liana.burghardt@gmail.com,1-(814)-863-6168

Running Title: Rhizobial fitness across Medicago mutants

Abstract

In the mutualism between leguminous plants and rhizobia bacteria, rhizobia live inside root nodules, creating the potential for host genes to shape the rhizobial selective environment. Many host genes that affect symbiosis have been identified; however, the extent to which these genes affect selection acting on rhizobia is unknown. In this study, we inoculated 18 *Medicago truncatula* symbiotic mutants (including mutants that alter NCR peptide production, plant defence, and nodule number regulation) with a mixture of 86 *Sinorhizobium meliloti* strains. Most mutations led to reduced host benefits, but effects on rhizobial benefit (i.e., relative strain fitness) varied widely, revealing widespread host gene by strain fitness interactions. Genome-wide association analyses identified genetic variants on rhizobial replicons pSymA and pSymB as important in mediating strain fitness responses to host mutations. While most top variants only affected rhizobial fitness with one host mutant (limited effect variants), nine affected fitness across six or more host mutations. These pervasive variants occurred primarily on pSymA, the symbiotic replicon, and include *fixL* and a few metabolic genes. In contrast to the limited effect variants, variants with pervasive positive effects on strain fitness when host genes were mutated tended to adversely affect fitness in wild-type hosts. Our results reveal how host symbiosis genes affect rhizobial fitness and set the stage for improving rhizobial inoculants and breeding legume hosts better adapted for multi-strain environments.

Introduction

Many bacteria live in close association with eukaryotic hosts, and a subset of these bacterial taxa significantly influence host fitness[1]. If host genetic variation alters selective pressures on bacteria, that variation can shape bacterial evolution. These genome x genome interactions are an essential prerequisite for co-evolution, and a potential mechanism for maintaining genetic variation in bacterial and host populations[2]. Most examples of genome x genome interaction involve pathogens, including a host immune receptor and microbial targets (e.g., human HLA and HIV epitopes [3]) or plant R genes [4]. However, examinations of quantitative pathogen resistance leveraging techniques such as co-GWAS [5,6] suggest complex genetic architectures underlie these host-pathogen interactions [7,8]. More recently, studies of genome-by-genome interactions have expanded to mutualisms [9–12]. These studies suggest that genome-by-genome interactions in mutualisms may be equally, if not more, genetically complex than in parasitism [13–18]. To what extent and in what ways host genes alter bacterial selection has important implications for bacterial evolution and deploying beneficial microbes in translational contexts such as agriculture [19,20]. Here, we take a novel approach to measuring genome x genome interactions by measuring how over a dozen host mutations alter the selective landscape experienced by scores of strains of mutualistic bacteria.

Legume hosts and rhizobial bacteria are a tractable system for studying how individual host genes alter the selection of their microbial mutualists. Rhizobia inhabit diverse environments, including soil, the rhizosphere, and root nodules, where they experience selection pressures that shape their fitness [21]. In nodules, rhizobia convert atmospheric nitrogen into plant assimilable forms and receive photosynthates for growth [22,23]. Host genetic variation can influence rhizobial fitness within root nodules [16,24–26] in ways likely shaped by nodule development and subsequent nitrogen fixation [27]. Decades of work have resulted in extensive collections of mutant lines in model legumes, including *Lotus* and *Medicago*, which have been exhaustively used to dissect the functional genetics of nodule formation and function. Most research on these mutants has focused on their effect on host benefits (e.g., shoot biomass) and investment (e.g., nodule number/size). Far less attention has been paid to the effect of host mutations on rhizobia [28]. However, doing so is necessary to clarify the potential for host genes to shape the rhizobial fitness.

Measuring the effect of host genes on rhizobia requires considering three aspects of rhizobial fitness—absolute fitness, relative fitness, and nodule strain diversity. Since most nodules are initiated by a single rhizobium, rhizobial population sizes in nodules indicate the rewards an individual bacterium receives from forming a nodule [28,29]. However, measuring population size alone does not enable an estimate of fitness difference among strains. Assessing strain-level relative fitness requires comparing strain frequencies before and after exposure to a selective environment, such as host nodules. A Select and Resequence (S&R) approach leverages whole-genome shotgun (WGS) sequencing to enable high-throughput tracking of strain frequencies [25,30]. This method allows for the direct quantification of selection pressures on rhizobia, revealing how selective an environment is (nodule strain diversity) and which strains are enriched or depleted compared to the initial community (relative strain fitness) [25,30]. A key advantage of the S&R method is that, since all strains are already sequenced, genome-wide association studies (GWAS) can be used to identify genetic variants associated with rhizobial relative fitness [16,25,30,31].

Work in the legume rhizobium mutualism suggests that host genetic variation can alter these metrics of rhizobial fitness [16,24,25,30]. For instance, host mutations that increase plant defence responses (*nad1*) reduce the number of rhizobia released from each nodule, while disrupting genes that cause hosts to form more nodules (*sunn1, sunn4, rdn1*) can increase rhizobia released at the plant level [32]. Other host genes like *rj2* and *rj4* in soybean restrict the diversity of strains that form nodules [33,34]. Disruption of the *Sinorhizobium* gene encoding HRRP peptidase, potentially involved in processing host NCR peptides, can restrict the rhizobial strains in nodules by modulating host discrimination or defence thresholds [35]. Lastly, disruptions within a host gene can dramatically alter *relative strain fitness.* For instance, members of the NPD gene family in *Medicago* differentially influenced which strain frequencies increased in nodules relative to the inoculated strain community [30]. Taken together, these studies highlight the potential for single host genes to reshape the selective environment and alter the evolution of rhizobia.

Here, we characterize how genetically disrupting host symbiosis genes alter selection on their nitrogen-fixing rhizobial symbionts *in planta*. To accomplish this goal, we chose eighteen existing *Medicago truncatula* mutant lines disrupted in symbiotic genes. These genes affect a wide range of host phenotypes and encompass the developmental trajectory of nodulation. We inoculated these mutants with a mixture of 86 *Sinorhizobium meliloti* strains (Fig. 1). We studied standard host investment and benefit traits, as well as multiple aspects of rhizobial fitness (Table 1). We then used a Select and Resequence approach to measure the frequencies of rhizobial strains (relative strain fitness) in nodules [25] and followed up with GWAS, to identify variants in rhizobial genes that altered strain fitness when host genes were perturbed. We examined the following questions: 1) Which host symbiosis genes alter rhizobial fitness? 2) Do host symbiosis genes have strain-specific effects on rhizobial fitness (i.e., is there evidence of genome x genome interactions)? 3) If yes, which rhizobial genes harbour genetic variation underlying the observed shifts in strain fitness, and are any of those variants selected in multiple host mutations? And 4) Are any host investment or benefit traits correlated with rhizobial fitness traits? While few host mutations altered rhizobial population sizes or nodule strain diversity, many host mutations did shift relative strain fitness. Most rhizobial genes associated with strain fitness shifts were found to be important within specific mutants, although eighteen variants, primarily on pSymA were important with six or more mutations, suggesting candidate genes that may play fundamental roles in in-plant rhizobial fitness.

Materials and methods

Host symbiotic mutants: The development of symbiosis in nodules is complex, with each step regulated by multiple hosts and rhizobial genes [27,36,37]. To encompass these steps, we chose 18 host symbiosis mutants (Table S1), which we grouped into three categories: early-acting, late-acting, and regulation of nodule number (Autoregulation of nodule number; AON) [38]. The early-acting genes (*hcl, ipd3*, *latd, dmi1-2*, *dmi2-3,* and *dmi3-1*) affect the initial processes, such as Nod Factor recognition and nodule organogenesis [39–43]. The late-acting genes (*nad1, dnf1, dnf2, dnf4, dnf2/dnf1, dnf5/dnf2, dnf6, dnf7*) influence nitrogen fixation [32,44], while the regulation of nodule number genes are in the AON (Autoregulation Of Nodulation) pathway (*sunn1, sunn4, rdn1*) [45,46]. While these categorizations help set expectations for the impact of the host mutations, each gene may have additional pleiotropic effects. Together, these mutants provide a broad view of the effects of host symbiosis genes on their rhizobial partners in multi-strain inoculations.

Seeds for all host mutant lines were obtained from Dr. Dong Wang, Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst.

Initial inoculum preparation: We created a mixed inoculum using 86 sequenced *Sinorhizobium meliloti* (henceforth *Sinorhizobium*) strains, which were part of a collection isolated from soils from 21 locations in Spain and France, using the nodules of *Medicago* plants as a "trap" for capturing *Sinorhizobium* isolates [47]. We grew each strain separately for 72 hours in Tryptone-Yeast extract (TY) media at 28°C, shaking at 120 rpm, before mixing equal volumes of each culture to create an initial inoculum with roughly equal proportions of each strain. To inoculate plants, we added 60 ml of the multi-strain mixture to 6 litres of sterile 0.85% NaCl solution and applied 10^7 cells per pot. This density was chosen to reflect *Sinorhizobium* populations in field soil [48] and to avoid triggering quorum sensing mechanisms [49], which typically occur at higher cell densities. We measured rhizobial density in the inoculum by plating a dilution series and counting the number of colonies. We measured the strain frequencies in the inoculum by centrifuging eight 1 ml aliquots at 18,000 x g for 2 minutes, removing the supernatant, and then freezing the pellet at -20°C before extracting DNA. Initial communities were sequenced, and frequencies inferred (Fig. S1) as outlined in the Rhizobial Fitness section below.

Greenhouse experiment to measure multi-strain symbiotic phenotypes: We grew five replicate pots of each of the 18 *Medicago* mutants plus wildtype A17 (Fig. 1). Before planting, seeds were scarified by gently abrading with sandpaper, surface sterilized by soaking in a 10% bleach solution for 90 seconds, and rinsed in de-ionized water. Seeds were stratified on wet filter paper at 4°C in the dark for 2 days, then planted in pots containing steam-sterilized Sunshine Mix #5. Germinated seedlings were thinned to three plants per pot. All pots were inoculated 48 hours after seed planting with the initial 86-strain inoculum. Following inoculation, pots were randomized into five spatial blocks with one representative of each genotype per block. Plants were watered as needed and fertilized weekly using increasing concentrations of N-free Fårhaeus medium [50]. Uninoculated plants were included to monitor cross-contamination. Pots were harvested and traits measured 8 weeks post-inoculation. We measured nodule number per plant, average area per nodule and total nodulated area per plant as measures of host investment, total biomass per plant and root-to-shoot ratio as measures of host benefit, and population size of rhizobia released per plant and per nodule, nodule strain diversity and relative strain fitness in nodule pools as measures of rhizobial fitness. After 8 weeks, we harvested inoculated plants and pooled nodules from ~ 3 plants per pot to ensure enough nodules (>100 /pot) to limit stochastic variation and saturate diversity. None of the uninoculated plants formed nodules. We photographed nodule pools per replicate to assess nodule number and size using ImageJ. Plant biomass was measured by weighing plant roots and shoots after oven drying at 75 °C for 3 days. Because not all plants survived in every pot, the number of plants occasionally varied. To control for this variation, all host traits were normalized per plant for analysis.

Rhizobial fitness: We surface sterilized the nodules from each pot in 10% bleach for 1 minute, rinsed the nodules in sterile water, and homogenized them for 90 seconds in 4 ml sterile 0.85% sodium chloride solution using an Omni TH tissue homogenizer. We estimated the number of rhizobia released from the nodule pools by dilution plating a 500-1000 ul aliquot of the homogenate and counting the number of colonies formed. We then estimated ***absolute fitness*** of the rhizobia on both a per plant and per nodule level by dividing the total number of colonies by the number of plants or nodules harvested from each pot and plotting the log10 transformation of it. The remaining nodule homogenate was spun at 400xg for 10 min at 10°C. to pellet the

bacteroids and the plant debris. We collected the supernatant enriched for undifferentiated rhizobial cells, pelleted them at 18000xg for 5 min, pipetted off the supernatant before freezing the pellets at -20 °C until DNA extraction. DNA was extracted using a Qiagen DNeasy® Plant extraction kit following the manufacturer's protocol, except we used 3 µL of RNase in the first step and 50 µL of nuclease water for the final elution.

Samples were sequenced using Illumina short-read technology at the Huck Genomic Core Facility at The Pennsylvania State University. Indexed libraries were generated using the Illumina DNA Prep Kit. 12X libraries were pooled and sequenced on a 150 x 150 paired-end NextSeq High Output run, resulting in an average read depth of ~75. We trimmed the reads using TrimGalore! (v0.4.1, www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default settings, except with minimum read length: 100 bp; quality threshold: 30 (−-quality); and minimum adapter match: 3 (−stringency), used bwa mem (v0.7.17) with default settings to align reads to the *Sinorhizobium (Ensifer) meliloti* USDA1106 genome (NCBI BioProject: PRJNA388336) and estimated strain frequencies using HARP [51]. Strain frequencies were used to 1) estimate ***nodule strain diversity*** using Shannon's H index (via the diversity function from the vegan package in R) and 2) calculate ***relative strain fitness*** of each strain in each replicate of a particular host mutant by taking the $\log_2$ of the ratio of the final strain frequency in the nodules to the initial frequency in the inoculum (median initial frequency = 0.0092, 5%-95% quantile: 0.006-0.018). The log transformation puts increases and decreases in proportional representation on the same scale (e.g., a fourfold decrease and increase will have a value of −2 and 2, respectively). Strains with frequency estimates of 0 in the selected communities were assigned a fitness value of −8, less than the lowest measurable reductions in other strains [25].

Statistical analyses. To assess the effects of the host mutants and the symbiotic stage on plant traits and rhizobial fitness, we used linear models and analysis of variance (ANOVA) for the univariate traits and redundancy analysis (RDA) for the multivariate trait of relative strain fitness. All analyses were run in the R statistical environment (v4.3.2). For each univariate trait, we fit two linear models (`lm()` function): one using host genotype and block, and a second using the symbiotic disruption stage and block as additive predictors. We compared variation explained by the two models using the adjusted R^2 (`anova()` function) to infer whether models that included mutant identity explained more variance than the symbiotic stage alone. Because the genotypes are nested within the stage of symbiosis, we ran two separate linear models to avoid introducing collinearity. We then compared the variance explained by each model. To ensure this comparison was not biased by differences in the number of levels (degrees of freedom) in each predictor, we used the adjusted R^2, which corrects for model complexity and sample size. This approach allows us to assess the relative explanatory power of broad functional categories (stage) versus specific genetic backgrounds (genotypes) on rhizobial community structure. We log10-transformed average nodule area/plant and total nodule area traits to reduce heteroscedasticity of variance. To assess pairwise differences between A17 and each mutant for each trait, we used Tukey's Honest Significant Difference (HSD) tests. Comparisons with an adjusted p-value < 0.05 were considered statistically significant. To investigate how host genotype or symbiotic stage altered the relative fitness of rhizobial strains (a multivariate trait), we performed an RDA (`rda()` function of the vegan package, v2.6-4) using relative strain fitness as the response matrix and host genotype or symbiotic stage as the explanatory variable. The significance of each model was assessed by extracting the adjusted R^2 and via permutation-based ANOVA (1,000 permutations) using the `anova()` function with `by = "terms"`. To test whether the relative strain fitness within each mutant differed from A17, we ran the model for each

mutant-A17 pairwise contrast. Predictors with $p < 0.05$ indicated significant changes in strain composition.

Genome-wide association study on relative strain fitness shifts in mutants vs. wildtype nodules

The relative strain fitness shift of each strain for each host mutant was calculated by subtracting the relative strain fitness in A17 from the relative strain fitness in the host mutants. We conducted 15 GWAS (three mutants formed no nodules) to identify rhizobial variants underlying host mutation-dependent ***strain fitness shifts*** for each nodulating mutant. We also conducted a GWAS to identify genes associated with strain fitness in A17 nodules. For each of these 16 GWAS, we ran a linear mixed model (LMM) in GEMMA (v0.98.3). Analyses were run separately for the three replicons (chromosome, pSymA, pSymB) to account for the different population structures. We used a published database of 30,076 SNPs representing previously defined linkage-disequilibrium (LD) groups within *S. meliloti*[16]. SNP filtering was performed with bcftools-1.18 [52] and PLINK 1.9 [53] based on minor allele frequency (0.05) and missingness (0.2). We selected the top five variant groups for each mutation-driven fitness shift tested based on their *p*-value ranks, cross-compared in the other 15 GWAS. To control for multiple tests while enabling comparisons across replicons that differed in variant group size from ~ 3,000 to 12,000, we adopted a Bonferroni corrected *p*-value cutoff of $p < 10^5$ to approximate a nominal $p < 0.05$ across replicons. This choice of a consistent intermediate threshold aligns with our goal of examining selection landscapes rather than identifying individual causal variants. This allowed us to categorise a rhizobial candidate as 'pervasive' if it ranked in the top ten within six or more host mutants or 'limited effect' if it appeared in five or fewer. We focused on ranks rather than *p*-values for this comparative analysis to focus on variants with the most evidence, regardless of differences in power or genetic architecture across host genetic perturbations.

Results

Effect of symbiotic mutations on rhizobial population sizes

Community inoculation resulted in nodulation across all host mutants except for *dmi1-2*, *dmi2-3, and dmi3-1,* consistent with the expectation that these genes are required for nodule formation (Fig. 2A). For the 15 mutants that formed nodules, we measured the number of rhizobia released from the nodules and tested if each mutant differed significantly from A17 plants. For instance, *ipd3* releases ~ 2.1×10^5 rhizobial cells per nodule and ~2.435×10^7 per plant, a 10-fold decrease compared to A17 (Fig. 2B, C). For both, rhizobia released per nodule and plant, symbiosis stage did not explain nearly as much variation in rhizobia population sizes as individual host mutations (Table S2). Only two mutants significantly reduced the rhizobial population size at both per plant and nodule levels: *ipd3 (*per plant p=0.0001; per nodule p=0.04) and *nad1 (*per plant p = 0.02; per nodule p=0.03), and no mutations increased population size (Table S3). Because *ipd3* and *nad1* do not produce significantly different numbers of nodules compared to A17 ($p > 0.05$), the smaller per-plant rhizobial population size is primarily due to reduced reproduction within nodules. This makes sense given that *nad1* evokes a defense-like response in nodules that kills reproductively active bacteria, and *ipd3* causes aborted infection threads, the part of the nodule in which reproductively competent rhizobia are found [43,54].

Host symbiotic genes shape relative strain success

The total rhizobial population released from individual nodules and plants does not tell us whether the effects of the mutations are strain specific. We examined strain-specific effects of the mutants in two ways. First, we evaluated the *nodule strain diversity,* as measured by

Shannon's H, in each of the 15 nodulating mutants compared to A17. This analysis revealed that *dnf7* was the only mutant significantly reducing *nodule strain diversity* ($p = 0.009$, Fig. 2C, Table S2). Second, we compared the relative fitness of each strain in the nodules of each host mutant to A17 (Fig. 2E). Hierarchical clustering revealed that strain relative fitness did not cluster solely by symbiotic stage and that the *latd* nodule strain fitness was quite different from all other hosts, including wildtype. Note that the *latd* host mutant results should be interpreted carefully—*latd* uniquely had a single estimate for rhizobial fitness traits, because plants formed so few nodules that we had to combine nodules across all pots to create a large enough sample for sequencing. A redundancy analysis with PERMANOVAs reinforced this finding. While the symbiosis stage explained some variation in relative strain fitness, symbiotic genes had strong individual effects (Fig. S2, Table S4, S5). Both early-acting (*ipd3* and *nad1)* and late-acting (all *dnf* mutants except *dnf4,* which was marginal at $p = 0.06$, and *dnf1*) significantly altered rhizobial relative fitness in nodules (Fig. 2E). The *dnf* mutants did not cluster together in the RDA space, reinforcing that defects in nitrogen fixation did not produce uniform effects on rhizobial enrichment. Some strains consistently exhibited high or low fitness across all host genotypes (e.g., MAG322 vs. MAG27), but the fitness of most strains depended on host symbiotic mutations and often changed the rank order of relative strain fitness of the strains compared to wildtype (Fig. S3). Importantly, strain shifts were often specific to individual host mutants; for example, MAG726B, MAG701B, and MAG201 had high fitness in *dnf4* but low fitness in *latd*. The host-dependent changes in strain fitness allow us to use GWAS to identify rhizobial allelic variants associated with the observed strain fitness shifts caused by each mutation relative to wildtype.

Rhizobial genes underlying strain fitness shifts in response to host mutations were not on the main chromosome

To determine genetic variants associated with strain fitness when host genes were disrupted, we performed 16 GWAS, one for the shifts in strain fitness caused by each of the fifteen mutants (relative fitness in mutant-relative fitness in A17) and one for relative fitness in A17 (Fig.3, Fig. S4). We found 9 linkage groups on pSymA and 17 linkage groups on pSymB that were strongly associated with strain fitness shifts ($p < 10^{-5}$) (Table 2, Table S6). In contrast, linkage groups on the chromosome had weaker effects (none with $p < 10^{-5}$), except for LD group 838 ($p = 6.38 \times 10^{-6}$), which was associated with increased strain fitness with *dnf6* and *latd*. The *p*-values of variants across different replicons may differ in part due to replicon-specific patterns of relatedness among strains. However, the impact of population structure on GWAS power is not always straightforward to predict. For instance, the chromosome may exhibit long-range linkage disequilibrium, resulting in large linkage blocks that tag multiple genes, which can both help or hinder association resolution.

GWAS identifies rhizobial genes with pervasive and specific effects on strain fitness shifts.

To study the genomic landscape underlying rhizobial fitness shifts in response to host mutations, we identified the top five variants for each host mutant (in total 114 variant groups) based on *p*-ranks (Table S6). Then we compared the *p*-value ranks and effect sizes of these groups across all the host mutants (Fig. 4, Fig. S5, Table S7). Most of the candidates on the chromosome were classified as 'limited effect' variant groups since they ranked among the top ten candidates, in five or fewer host mutations (Table S7). Exceptions included a massive chromosomal variant group (901) with 144 variants in 65 genes, among the top ten candidates among nine host mutants. Six variants with pervasive effects were found on pSymB. On pSymB, four small variant groups were top-ranked in at least three host mutants (ABC transporter and RKpR, a

dihydrodipicolinate synthase family protein, *pepT*, peptidase T, and a hypothetical protein (Fig. S5). However, the seven pervasive variants on pSymA were most interesting to us because they spanned similar hosts and had strong statistical support. These candidates included malonyl-CoA synthase, SGNH/GDSL hydrolase family protein, *fixL,* an oxygen sensor histidine kinase, a short-chain oxidoreductase, and a transport protein belonging to the MFS (Major Facilitator Superfamily). Interestingly, the non-reference variant often increased strain fitness within the host mutants. In addition, we noted a trend where the pervasive variants that conferred higher relative fitness in the mutant hosts, tended to have lower fitness in wildtype A17. For example, the malonyl COA synthase had an alternate allele associated with increased strain fitness (positive beta) in several host mutant backgrounds (*ipd3*, *dnf1*, *dnf2*, *dnf6*, *dnf7*, *nad1*, *rdn1*, and *sunn1*). In contrast, the same allele was associated with decreased fitness (negative beta) in the wild-type A17 (Fig. 4, Table S6). This pattern suggests a host-dependent allelic effect with the allele conferring a fitness advantage in mutant hosts becoming disadvantageous in the wild-type host.

Only 82 of the 94 limited-effect variant groups were small (having fewer than three genes). The majority of these were on pSymB (Table S7). Those with strong statistical support ($p<10^{-5}$) included variants in or nearby genes coding two ABC transporters (*sunn1, ipd3, dnf1/dnf2, rdn1)*, glutamine synthetase (*latd)* methyltransferases (*nad1*), transcriptional regulators (*rdn1*), a couple of hypothetical proteins (*ipd3, dnf1/dnf2,dnf5/dnf2,dnf2,sunn1)*and genes involved in purine metabolism (*dnf3*). Interestingly, a rhizobial glutamine synthetase gene, which converts glutamate and ammonia into glutamine to assimilate nitrogen, was the strongest candidate for variation in strain fitness shift in hosts with a mutated *latd,* which encodes a nitrate transporter [55]. This was unexpected since rhizobial glutamine synthetases are generally thought to be strongly downregulated in nodules where host-driven nitrogen assimilation processes dominate [56].

Correlations of host investment and benefit traits with rhizobial traits

Because mutualistic partners affect each other's fitness, we were curious if any host investment or benefit traits were correlated with rhizobial fitness. At the level of individuals pots, nodule area, plant biomass, and root-to-shoot ratio covaried with nodule strain diversity and position on the first RDA axis describing shifts in strain relative fitness (Fig. 5A). Total nodulated area per plant covaried with the second RDA axis describing shifts in strain relative fitness. Perhaps unsurprisingly, the host trait that covaried with CFUs released per plant was the number of nodules per plant. Across the first PC of trait space, we found the wild-type plants grouped with the AON mutants, *hcl*, and three *dnfs* (3,4,6). In contrast, the second PC tended to separate the early-acting, AON, and wildtype plants from the late-acting mutants. Genotypic trait correlations confirmed these patterns (Fig. 5B). Host mutants that were larger and had lower root-to-shoot ratios, i.e., those that benefited more from rhizobial N-fixation, tended to release more rhizobia (CFUs) from individual nodules, more rhizobia on a whole-plant level, and had nodule strain communities that resembled those in A17 nodules (Fig. 5B). Interestingly, most host traits correlated strongly with strain diversity and position along the first rhizobial relative fitness RDA axis; The exception was average nodule area/plant, which was most strongly correlated with host position on the second RDA axis. Details on mutation-specific effects on host traits can be found in Fig. S6 and Table S3. Our results suggest that symbiotic gene disruptions strongly shape both partners' fitness in directions reinforcing mutualism—mutations that reduce rhizobial fitness also reduce host fitness.

Discussion

By combining a Select-and-Resequence (S&R) experiment with GWAS, we demonstrate that mutations in host symbiotic genes create distinct environments that shape rhizobial strain success and identify rhizobial candidate genes underlying strain fitness shifts in response to each host mutation. We found rhizobial genetic variants with both pervasive and limited effects across host mutants, and that symbiotic stage categories provide little clarity about how rhizobial genes influence strain fitness. We also found that mutations that increased rhizobial fitness tend to increase host investment and benefit traits. Our findings underscore the complexity of mutualism, where both host and rhizobial genomes jointly determine the outcomes of symbiosis at the community and individual strain levels, adding to knowledge of how *Medicago* hosts determine rhizobial fitness [16,30].

Host genetic mutations impact rhizobial absolute fitness

Legumes releasing rhizobia from their senescing nodules directly contribute to the soil rhizobial population [28,48,57]. We found that most mutations did not affect the number of rhizobia released from each nodule, and those that did, *ipd3* and *nad1*, released fewer rhizobia per nodule. The reduced number of rhizobia per nodule with these two mutants is not unexpected, given that disrupting *nad1* leads to overactive plant defence responses in nodules and disrupting *ipd3* traps rhizobia within infection threads, likely reducing survival [32,43]. While our results suggest that the effects of host genetic mutations on rhizobial fitness may sometimes be gene-specific, we did observe that rhizobia released per nodule and average nodule area were positively correlated across mutations. Increasing the number of rhizobia the host supports represents one pathway to nurture robust populations of rhizobia, and identifying the host allelic variation that increases this amplification could set the stage for targeted interventions aimed at bolstering soil rhizobial populations in agricultural fields [58].

Rhizobial fitness shifts depend on host genetics

A large group of NCR peptides is one of the major known host factors for altering strain survival and growth in nodules [59–62]. Although late-acting NCRs can be important determinants of host strain specificity in *M. truncatula*, our investigation reveals a more complex landscape [27,63]. Our analyses of two NCR mutations revealed differing effects on strain fitness within nodules. The *dnf7* mutant, which lacks functional NCR169 [64], was the only mutant that reduced nodule strain diversity, and it clustered with *ipd3,* which reduced rhizobial population size. *dnf4*, which does not produce NCR211 and influences rhizobial differentiation during bacteroid formation [65], caused similar shifts in strain relative fitness as *dnf3* and *dnf6,* suggesting these uncharacterized mutants may also relate to NCRs. A few rhizobia genes (*bacA, hrrP, yejABEF*) are known to affect the efficacy of NCRs [35,59,66]; however, none of those genes were top candidates in our GWAS, suggesting additional mechanisms of host restriction of rhizobia [67–69].

Rhizobial variants with pervasive and limited effects across host mutants

Among the ~120 top rhizobial variants, ~18% affected rhizobial fitness in six or more host mutations (pervasive genes), and ~82% affected rhizobial fitness in 5 or fewer host mutants (limited effect genes). Interestingly, pervasive effect alleles tended to exhibit fitness trade-offs; alleles that conferred higher fitness in host mutants tended to confer lower fitness with wild-type hosts [16]. Many of the GWAS candidates were identified in only one or a few host mutant backgrounds, with the strongest associations on pSymB. Unlike the pervasive variants, most limited-effect variants tended to have little or no effect on strain fitness in the wild-type background. Most candidate genes with pervasive effects were highly expressed in the nitrogen

fixation zone (Zone III) [56]. For example, the *fixL* gene, which regulates the expression of nitrogen fixation genes in response to low oxygen levels, Malonyl-CoA synthase that is induced in the rhizosphere of alfalfa, responsible for malonate catabolism within root nodules [73], and a NAD(P)H-binding protein from the conserved NmrA domain family that shows high similarity to a transcriptional regulator, involved in nitrogen metabolite repression in fungi [74]. These genes represent low-hanging fruit for uncovering novel mechanisms that alter the host range of symbionts [75].

Beyond these candidates, most genes only altered strain fitness in specific host mutants, indicating that each mutation creates a unique challenge for rhizobia reproduction. Notable candidates include: A Class I SAM-dependent methyltransferase associated with strain fitness in *nad1* mutants, suggesting the importance of epigenetic modifications for rhizobial survival under host defences [32] , and *stc4*, encoding stachydrine N-demethylase reductase, which enhanced strain fitness in *dnf4* mutants, possibly by protecting differentiating rhizobia [65]. While these genes are candidates for altering strain fitness in some hosts and not others, our experiments do not pinpoint at what stage each host gene exerts the strongest selective filter. One way to follow-up on these results is with assays comparing rhizobia mutated in the candidate genes identified in our GWAS on wildtype and mutant hosts, which would allow one to identify the selective stage—rhizosphere, infection thread, or nodule—where each gene exerts its influence on strain fitness.

Considerations for interpreting mixed-strain experiments

Final strain fitness in nodules is a composite trait that reflects host genetics and microbial traits such as rhizosphere colonization, competition for infection, and proliferation within nodules. As such, shifts in relative strain fitness across host genotypes could also be influenced by differential competitive abilities among strains before infection of each host genotype. Thus, while our design allows us to detect which host symbiotic genes alter strain fitness, it does not allow us disentangle the relative contributions of microbial competition versus host filtering. This limitation is especially relevant in complex communities, where strain-strain interactions influence host effects [21]. While data suggests the effects of microbe × microbe interactions on strain relative fitness in our system are weak [26], caution that it is important to interpret our host-driven fitness shifts as emergent outcomes of host × microbe × microbe interactions. An additional consideration is that our experiment is designed to measure how growing plants of a single genotype will influence rhizobial strain fitness. Linking these fine-scale host effects to long-term dynamics of rhizobial adaptation is an important overarching goal in plant-microbe interactions that requires incorporating additional ecological and evolutionary contexts including the importance of soil selection [76] or the cooccurrence of host genotypes across time and space [13,77].

Conclusion

Despite legumes being used for soil fertility improvement and actively managed for centuries, biological nitrogen fixation in agriculture remains inconsistent spatially and temporally [79–82]. This challenge has driven interest and investment in developing microbial inoculants to enhance crop yield, stress tolerance, disease, and pest management [20]. Rhizobia have been included in microbe-based bioinoculant formulations for centuries [75]. However, most of those strain selection decisions are dominated by strain performances observed under single-strain assays (inoculating a single plant with a single strain) in controlled environments. Thus, strains which

show promise as potential bioinoculants under controlled conditions fail to reach densities high enough to influence crop yield in realistic agricultural systems [83]. Formulating ways to sustain rhizobial fitness in the agricultural fields is a critical avenue for plants to continue acquiring nitrogen from their associated rhizobia [84].

Advances in host-mediated microbiome engineering have demonstrated that plant genes can be leveraged to assemble trait-optimized microbial communities, with the selection of beneficial microbes being dependent on the host genotype[81,85]. The findings in this study suggest a route towards expanding the search for host genes that alter rhizobial fitness. If natural allelic variants of these genes also favor colonization by high-performing symbionts, it will enable the development of "designer plants" with enriched mutualistic microbiomes [86]. Here, we identify host genes that drive strain-specific selection and show that even functionally similar mutations can have divergent effects on strain fitness in multi-strain contexts [21]. While our study focused on loss-of-function mutations, studies on segregating natural genetic variation in *Medicago* suggest that specific host genes can differentially enrich rhizobia but find limited evidence for widespread balancing or positive selection [16]. Though we do not expect to find these loss-of-function mutations segregating in natural populations, this work lays a critical foundation: identifying the host genes that influence rhizobial fitness. A next step could be to systematically test the effect of naturally occurring alleles within these genes on rhizobial fitness, to pinpoint the variants that promote or reduce fitness and the mechanisms by which they do it.

Acknowledgments

We thank Prof. Dong Wang of University of Massachusetts, Amherst for generously providing the *M. truncatula* mutant seeds, Lillian Cherry, Maria Alejandra Gil-Polo, and Jennifer E. Harris for help with the tedious task of picking and processing nodules. Prof. Katy Heath of University of Illinois Urbana-Champaign and Dr. Mike Grillo (currently Assistant Professor in Loyola University, Chicago) for providing strains and sequencing data. This work was supported by the National Science Foundation grant IOS-1856744 to PT, NY, and LTB and IOS-2243819 to PT and LTB and was completed using computing resources provided by Penn State Institute for Computational and Data Science and sequencing by the Huck Genomics Core Facility. LTB's work was also supported by HATCH grant no. 1025611. The authors acknowledge the Huck Institutes' Genomics Core Facility (RRID:SCR_023645) for providing library prep and sequencing services.

Authorship Credits

LTB, NY, PT idea conception and funding, LTB experimental design, LTB, and RGB plant care; LTB, SG, RB drafted the manuscript; RB, GF, LTB data collection, SG, LTB, RGB, BE, JS analysis, SG, LTB, RB, JS, BE, PT, NY, GF manuscript revisions.

Competing Interests:

The authors declare no competing interests.

Data Accessibility

All the raw reads from sequencing the rhizobial strains in the initial community and in the rhizobial genomic DNA extracted from the nodule pools of the individual host mutants have been deposited in NCBI under the Bioproject PRJNA401437 (Submission ID: SUB14622715). All raw data and analyses code have been uploaded to the Github (https://github.com/BurghardtLab/HostMutants2021_C86meliloti)

517

References

1. Friesen, M. L. *et al.* Microbially Mediated Plant Functional Traits. *Annu. Rev. Ecol. Evol. Syst.* **42**, 23–46 (2011).
2. Percy, D. M. John N. Thompson, The Geographic Mosaic of Coevolution. *Syst. Biol.* **56**, 149–151 (2007).
3. Bartha, I. *et al.* A genome-to-genome analysis of associations between human genetic variation, HIV-1 sequence diversity, and viral control. *eLife* **2**, e01123 (2013).
4. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
5. Amandine, C. *et al.* Unraveling coevolutionary dynamics using ecological genomics. *Trends Genet.* **38**, 1003–1012 (2022).
6. Demirjian, C., Vailleau, F., Berthomé, R. & Roux, F. Genome-wide association studies in plant pathosystems: success or failure? *Trends Plant Sci.* **28**, 471–485 (2023).
7. Corwin, J. A. *et al.* The Quantitative Basis of the Arabidopsis Innate Immune System to Endemic Pathogens Depends on Pathogen Genetics. *PLOS Genet.* **12**, e1005789 (2016).
8. Newberry, E. A., Minsavage, G. V., Holland, A., Jones, J. B. & Potnis, N. Genome-Wide Association to Study the Host-Specificity Determinants of Xanthomonas perforans. *Phytopathology®* **113**, 400–412 (2023).
9. Medina, M. & Sachs, J. L. Symbiont genomics, our new tangled bank. *Genomics* **95**, 129–137 (2010).
10. Clare, E. L., Schiestl, F. P., Leitch, A. R. & Chittka, L. The promise of genomics in the study of plant-pollinator interactions. *Genome Biol.* **14**, 207 (2013).
11. *Mutualism*. (Oxford University Press, Oxford, United Kingdom, 2015).
12. Batstone, R. T. Genomes within genomes: nested symbiosis and its implications for plant evolution. *New Phytol.* **234**, 28–34 (2022).

13. Heath, K. D. & Stinchcombe, J. R. EXPLAINING MUTUALISM VARIATION: A NEW EVOLUTIONARY PARADOX?: PERSPECTIVE. *Evolution* **68**, 309–317 (2014).

14. Turkarslan, S. *et al.* Synergistic epistasis enhances the co-operativity of mutualistic interspecies interactions. *ISME J.* **15**, 2233–2247 (2021).

15. Batstone, R. T. *et al.* Genome-wide association studies across environmental and genetic contexts reveal complex genetic architecture of symbiotic extended phenotypes. *mBio* **13**, e01823-22 (2022).

16. Epstein, B. *et al.* Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis. *Mol. Ecol.* mec.16602 (2022) doi:10.1111/mec.16602.

17. Carlson, C. & Frederickson, M. E. An emerging view of coevolution in the legume–rhizobium mutualism. *Mol. Ecol.* **32**, 3793–3797 (2023).

18. Gohar, D. *et al.* Genomic evidence of symbiotic adaptations in fungus-associated bacteria. *iScience* **28**, 112253 (2025).

19. Denison, R. F., Kiers, E. T. & West, S. A. Darwinian Agriculture: When Can Humans Find Solutions Beyond the Reach of Natural Selection? *Q. Rev. Biol.* **78**, 145–168 (2003).

20. Busby, P. E. *et al.* Research priorities for harnessing plant microbiomes in sustainable agriculture. *PLOS Biol.* **15**, e2001793 (2017).

21. Burghardt, L. T. & diCenzo, G. C. The evolutionary ecology of rhizobia: multiple facets of competition before, during, and after symbiosis with legumes. *Curr. Opin. Microbiol.* **72**, 102281 (2023).

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

565 23. Poole, P., Ramachandran, V. & Terpolilli, J. Rhizobia: from saprophytes to endosymbionts.
566 *Nat. Rev. Microbiol.* **180**, 1691 (2018).
567 24. Heath, K. D. & Tiffin, P. Context dependence in the coevolution of plant and rhizobial
568 mutualists. *Proc. R. Soc. B Biol. Sci.* **274**, 1905–1912 (2007).
569 25. Burghardt, L. T. *et al.* Select and resequence reveals relative fitness of bacteria in symbiotic
570 and free-living environments. *Proc. Natl. Acad. Sci.* **115**, 2425–2430 (2018).
571 26. Burghardt, L. T., Epstein, B., Hoge, M., Trujillo, D. & Tiffin, P. *Host-Associated Rhizobia*
572 *Fitness: Dependence on Nitrogen, Density, Community Complexity, and Legume Genotype*.
573 http://biorxiv.org/lookup/doi/10.1101/2020.11.20.392183 (2020)
574 doi:10.1101/2020.11.20.392183.
575 27. Roy, S. *et al.* Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and
576 Symbiotic Nitrogen Fixation. *Plant Cell* **32**, 15–41 (2020).
577 28. Burghardt, L. T. Evolving together, evolving apart: measuring the fitness of rhizobial
578 bacteria in and out of symbiosis with leguminous plants. *New Phytol.* **228**, 28–34 (2020).
579 29. Ratcliff, W. C., Underbakke, K. & Denison, R. F. Measuring the fitness of symbiotic
580 rhizobia. *Symbiosis* **55**, 85–90 (2011).
581 30. Burghardt, L. T., Trujillo, D. I., Epstein, B., Tiffin, P. & Young, N. D. A Select and
582 Resequence Approach Reveals Strain-Specific Effects of *Medicago* Nodule-Specific PLAT-
583 Domain Genes. *Plant Physiol.* **182**, 463–471 (2020).
584 31. Epstein, B. *et al.* Genome-Wide Association Analyses in the Model Rhizobium *Ensifer*
585 *meliloti*. *mSphere* **3**, e00386-18 (2018).
586 32. Wang, C. *et al. NODULES WITH ACTIVATED DEFENSE 1* is required for maintenance of
587 rhizobial endosymbiosis in *Medicago truncatula*. *New Phytol.* **212**, 176–191 (2016).

588 33. Yang, S., Tang, F., Gao, M., Krishnan, H. B. & Zhu, H. *R* gene-controlled host specificity in
589 the legume–rhizobia symbiosis. *Proc. Natl. Acad. Sci.* **107**, 18735–18740 (2010).
590 34. Tang, F., Yang, S., Liu, J. & Zhu, H. *Rj4* , a Gene Controlling Nodulation Specificity in
591 Soybeans, Encodes a Thaumatin-Like Protein But Not the One Previously Reported. *Plant*
592 *Physiol.* **170**, 26–32 (2016).
593 35. Price, P. A. *et al.* Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and
594 mediates symbiotic compatibility. *Proc. Natl. Acad. Sci.* **112**, 15244–15249 (2015).
595 36. Oldroyd, G. E. D. Speak, friend, and enter: signalling systems that promote beneficial
596 symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**, 252–263 (2013).
597 37. Wheatley, R. M. *et al.* Lifestyle adaptations of Rhizobium from rhizosphere to symbiosis.
598 *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23823–23834 (2020).
599 38. Ferguson, B. J. *et al.* Legume nodulation: The host controls the party. *Plant Cell Environ.* **42**,
600 41–51 (2019).
601 39. Bright, L. J., Liang, Y., Mitchell, D. M. & Harris, J. M. The *LATD* Gene of *Medicago*
602 *truncatula* Is Required for Both Nodule and Root Development. *Mol. Plant-Microbe*
603 *Interactions®* **18**, 521–532 (2005).
604 40. Hogg, B. V., Cullimore, J. V., Ranjeva, R. & Bono, J.-J. The *DMI1* and *DMI2* Early
605 Symbiotic Genes of *Medicago truncatula* Are Required for a High-Affinity Nodulation
606 Factor-Binding Site Associated to a Particulate Fraction of Roots. *Plant Physiol.* **140**, 365–
607 373 (2006).
608 41. Messinese, E. *et al.* A Novel Nuclear Protein Interacts With the Symbiotic DMI3 Calcium-
609 and Calmodulin-Dependent Protein Kinase of *Medicago truncatula*. *Mol. Plant-Microbe*
610 *Interactions®* **20**, 912–921 (2007).

611 42. Smit, P. *et al.* Medicago LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling.
612 *Plant Physiol.* **145**, 183–191 (2007).
613 43. Ovchinnikova, E. *et al.* IPD3 Controls the Formation of Nitrogen-Fixing Symbiosomes in
614 Pea and *Medicago* Spp. *Mol. Plant-Microbe Interactions®* **24**, 1333–1344 (2011).
615 44. Starker, C. G., Parra-Colmenares, A. L., Smith, L., Mitra, R. M. & Long, S. R. Nitrogen
616 Fixation Mutants of *Medicago truncatula* Fail to Support Plant and Bacterial Symbiotic
617 Gene Expression. *Plant Physiol.* **140**, 671–680 (2006).
618 45. Schnabel, E., Journet, E.-P., de Carvalho-Niebel, F., Duc, G. & Frugoli, J. The Medicago
619 truncatula SUNN Gene Encodes a CLV1-like Leucine-rich Repeat Receptor Kinase that
620 Regulates Nodule Number and Root Length. *Plant Mol. Biol.* **58**, 809–822 (2005).
621 46. Schnabel, E. L. *et al.* The *ROOT DETERMINED NODULATION1* Gene Regulates Nodule
622 Number in Roots of *Medicago truncatula* and Defines a Highly Conserved, Uncharacterized
623 Plant Gene Family. *Plant Physiol.* **157**, 328–340 (2011).
624 47. Riley, A. B., Grillo, M. A., Epstein, B., Tiffin, P. & Heath, K. D. Discordant population
625 structure among rhizobium divided genomes and their legume hosts. *Mol. Ecol.* **32**, 2646–
626 2659 (2023).
627 48. Denison, R. F. & Kiers, E. T. Life histories of symbiotic rhizobia and mycorrhizal fungi.
628 *Curr. Biol. CB* **21**, R775-85 (2011).
629 49. Darch, S. E., West, S. A., Winzer, K. & Diggle, S. P. Density-dependent fitness benefits in
630 quorum-sensing bacterial populations. *Proc. Natl. Acad. Sci.* **109**, 8259–8263 (2012).
631 50. Fåhraeus, G. The Infection of Clover Root Hairs by Nodule Bacteria Studied by a Simple
632 Glass Slide Technique. *Microbiology* **16**, 374–381 (1957).
633 51. Kessner, D., Turner, T. L. & Novembre, J. Maximum Likelihood Estimation of Frequencies
634 of Known Haplotypes from Pooled Sequence Data. *Mol. Biol. Evol.* **30**, 1145–1158 (2013).

635 52. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
636 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–
637 2993 (2011).
638 53. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
639 Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
640 54. Domonkos, Á. *et al.* NAD1 Controls Defense-Like Responses in Medicago truncatula
641 Symbiotic Nitrogen Fixing Nodules Following Rhizobial Colonization in a BacA-
642 Independent Manner. *Genes* **8**, 387 (2017).
643 55. Harris, J. M. & Dickstein, R. Control of root architecture and nodulation by the *LATD/NIP*
644 transporter. *Plant Signal. Behav.* **5**, 1365–1369 (2010).
645 56. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic
646 root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**,
647 817–837 (2014).
648 57. Brewin, N. J. Development of the Legume Root Nodule. *Annu. Rev. Cell Biol.* **7**, 191–226
649 (1991).
650 58. Gopalakrishnan, S. *et al.* Plant growth promoting rhizobia: challenges and opportunities. *3*
651 *Biotech* **5**, 355–377 (2015).
652 59. Haag, A. F. *et al.* Protection of Sinorhizobium against Host Cysteine-Rich Antimicrobial
653 Peptides Is Critical for Symbiosis. *PLoS Biol.* **9**, e1001169 (2011).
654 60. Montiel, J. *et al.* Morphotype of bacteroids in different legumes correlates with the number
655 and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5041–5046 (2017).
656 61. Czernic, P. *et al.* Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and
657 Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich
658 Peptides. *Plant Physiol.* **169**, 1254–1265 (2015).

659 62. Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the
660 *Rhizobium* –legume symbiosis. *Proc. Natl. Acad. Sci.* **103**, 5230–5235 (2006).
661 63. Walker, L., Lagunas, B. & Gifford, M. L. Determinants of host range specificity in legume-
662 rhizobia symbiosis. *Front. Microbiol.* **11**, (2020).
663 64. Horváth, B. *et al.* Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes
664 symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant. *Proc. Natl. Acad. Sci.*
665 *U. S. A.* **112**, 15232–15237 (2015).
666 65. Kim, M. *et al.* An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing
667 symbiosis. *Proc. Natl. Acad. Sci.* **112**, 15238–15243 (2015).
668 66. Nicoud, Q. *et al.* Sinorhizobium meliloti Functions Required for Resistance to Antimicrobial
669 NCR Peptides and Bacteroid Differentiation. *mBio* **12**, e00895-21 (2021).
670 67. Guefrachi, I. *et al.* Extreme specificity of NCR gene expression in Medicago truncatula.
671 *BMC Genomics* **15**, 712 (2014).
672 68. Burghardt, L. T. *et al.* Transcriptomic basis of genome by genome variation in a legume-
673 rhizobia mutualism. *Mol. Ecol.* **26**, 6122–6135 (2017).
674 69. Lima, R. M., Kylarová, S., Mergaert, P. & Kondorosi, É. Unexplored Arsenals of Legume
675 Peptides With Potential for Their Applications in Medicine and Agriculture. *Front.*
676 *Microbiol.* **11**, 1307 (2020).
677 70. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat. Rev.*
678 *Genet.* **14**, 807–820 (2013).
679 71. Hoban, S. *et al.* Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions,
680 and Future Directions. *Am. Nat.* **188**, 379–397 (2016).
681 72. Elena, S. F. Local adaptation of plant viruses: lessons from experimental evolution. *Mol.*
682 *Ecol.* **26**, 1711–1719 (2017).

683 73. Karunakaran, R., East, A. K. & Poole, P. S. Malonate Catabolism Does Not Drive N_2
684 Fixation in Legume Nodules. *Appl. Environ. Microbiol.* **79**, 4496–4498 (2013).
685 74. Stammers, D. K. The structure of the negative transcriptional regulator NmrA reveals a
686 structural superfamily which includes the short-chain dehydrogenase/reductases. *EMBO J.*
687 **20**, 6619–6626 (2001).
688 75. Mus, F. *et al.* Symbiotic Nitrogen Fixation and the Challenges to Its Extension to
689 Nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).
690 76. Burghardt, L. T., Epstein, B. & Tiffin, P. Legacy of prior host and soil selection on rhizobial
691 fitness in planta. *Evolution* **73**, 2013–2023 (2019).
692 77. Wippel, K. *et al.* Host preference and invasiveness of commensal bacteria in the Lotus and
693 Arabidopsis root microbiota. *Nat. Microbiol.* **6**, 1150–1162 (2021).
694 78. Sugawara, M. *et al.* Variation in bradyrhizobial NopP effector determines symbiotic
695 incompatibility with Rj2-soybeans via effector-triggered immunity. *Nat. Commun.* **9**, 3139
696 (2018).
697 79. Kaminsky, L. M., Trexler, R. V., Malik, R. J., Hockett, K. L. & Bell, T. H. The Inherent
698 Conflicts in Developing Soil Microbial Inoculants. *Trends Biotechnol.* **37**, 140–151 (2019).
699 80. Soumare, A. *et al.* Exploiting Biological Nitrogen Fixation: A Route Towards a Sustainable
700 Agriculture. *Plants* **9**, 1011 (2020).
701 81. Trivedi, P., Batista, B. D., Bazany, K. E. & Singh, B. K. Plant–microbiome interactions
702 under a changing world: responses, consequences and perspectives. *New Phytol.* **234**, 1951–
703 1959 (2022).
704 82. Batstone, R. T., Ibrahim, A. & MacLean, L. T. Microbiomes: Getting to the root of the
705 rhizobial competition problem in agriculture. *Curr. Biol.* **33**, R777–R780 (2023).

83. Parnell, J. J. *et al.* From the Lab to the Farm: An Industrial Perspective of Plant Beneficial Microorganisms. *Front. Plant Sci.* **7**, (2016).

84. Porter, S. S. & Sachs, J. L. Agriculture and the Disruption of Plant–Microbial Symbiosis. *Trends Ecol. Evol.* **35**, 426–439 (2020).

85. Batstone, R. T., O'Brien, A. M., Harrison, T. L. & Frederickson, M. E. Experimental evolution makes microbes more cooperative with their local host genotype. *Science* **370**, 476–478 (2020).

86. Batstone, R. T., Burghardt, L. T. & Heath, K. D. Phenotypic and genomic signatures of interspecies cooperation and conflict in naturally occurring isolates of a model plant symbiont. *Proc. R. Soc. B Biol. Sci.* **289**, 20220477 (2022).

717 Figure Legends

718 **Fig. 1:** Schematic overview of the experiment used to evaluate rhizobial fitness and host traits
719 across different host mutant backgrounds. A set of 86 *Sinorhizobium* strains was cultured and
720 combined into a mixed inoculum. This inoculum was applied to *Medicago truncatula* wild-type
721 (A17) and host mutants with gene disruptions that act at different stages of symbiosis: Early,
722 Late, and Autoregulation of Nodulation (AON). The number of host mutants per stage is
723 indicated in parentheses. Host and rhizobial traits were measured at the nodule and whole-plant
724 scales. The experimental conditions are clarified under 'Experimental Details'. Each treatment
725 included five pots per host mutant with three plants per pot, grown in randomized blocks under
726 nitrogen(N)-free conditions for 8 weeks. Traits measured include indicators of rhizobial benefit,
727 host benefit, and host investment (Refer Table 1 for more details). Different coloured nodules
728 indicate nodules formed by different rhizobial strains.
729

730 **Fig. 2: Host symbiotic mutations influence rhizobial fitness.** (A) Host mutations altered
731 nodule numbers in consistent and predictable ways. (B–C) Host genotype also influenced
732 rhizobial fitness, as measured by $\log_{10}$-transformed colony-forming units per nodule and per
733 plant (CFU/nodule) (B) and CFU /plant (C). (D) Nodule strain diversity was largely unaffected
734 by host mutations. Closed triangles indicate significant increase or decrease relative to wild-type
735 A17 (based on Tukey-adjusted *p* values). Boxplots are color-coded by the symbiotic stage
736 disrupted by each host mutation (see Table S1). (E) Heatmap showing the median relative fitness
737 of each of 86 *S. meliloti* strains across the different host genotypes. Strains enriched relative to
738 their initial inoculum appear in shades of red; depleted strains appear in shades of blue. Host
739 genotypes are hierarchically clustered based on similarity in strain fitness profiles. Strains are
740 ranked by their relative fitness in A17. Mutants in bold differ significantly from A17 (see Table
741 S5 for RDA paired with the PERMANOVA results).
742

743 **Fig. 3: GWAS of strain fitness shifts in host mutants. (**A) The rhizobial strain fitness shifts
744 for each of the 86 strains, calculated by subtracting individual relative strain fitness measured in
745 A17 from the relative strain fitness in the host mutants. The heatmap is hierarchically based on
746 the similarity of median 'shift' values across hosts. Shades of orange denote positive fitness shift
747 compared to A17, while shades of purple denote negative fitness shift compared to A17. (B)
748 Manhattan plot indicating the SNP (single-nucleotide polymorphism) positions, based on S.
749 *meliloti* USDA1106 association with the relative strain fitness shifts across the 15 host mutants
750 and strain fitness in A17. The dotted line indicates the *p* value threshold used for identifying
751 candidate genes ($p < 10^{-5}$), and significant associations are color-coded according to host
752 mutation and A17. The GWAS was run separately on each replicon to account for different
753 population structures. The 'Δ' indicates 'strain fitness shift' in that host mutant. For comparison,
754 we also ran the GWAS on A17 relative strain fitness.
755

756 **Fig. 4: Rhizobial candidate genes underlying strain fitness shifts predominantly occur on**
757 **pSymA**. Heatmaps based on the inverse *p*-value ranks (red-yellow) and β values which denote
758 the effect sizes (red-blue) for the top five variant groups on the pSymA across the 15 host
759 mutants and A17. Rows represent candidate variant groups, and columns represent host mutants.
760 The host mutants are colour coded based on the stage of disruption they affect. Progressive dark
761 shades (low *p*-value ranks) indicate stronger association between the variant group and strain
762 fitness shift in that host mutant. Heatmap based on the effect sizes (β values) for the same
763 candidate variant groups was ordered to match the *p*-value ranking. Shades of red indicate host
764 mutants where the corresponding variant has a positive effect on strain fitness shift, shades of

blue indicate negative effects as compared to A17. Colour intensity corresponds to effect magnitude. Alternate allele frequencies (i.e., not present in reference genome USDA1106) for each candidate are also shown in parentheses next to the gene names. Pervasiveness of a variant group is defined has number of times a variant group is among the top 10 candidates (based on the *p* value ranks) within a host mutant. A variant group is classified as "pervasive" if it is in the top10 across 6 or more host genotypes while a variant is classified as "limited effect" if it is in the top10 candidates across 5 or less host genotypes. We have mentioned the number of times a variant appears in the top 10 under the "Pervasiveness" column in the figure and have indicated the pervasive genes in bold. (See Fig. S5 for the other replicons.)

Fig 5: Covariation and correlation structure among host and rhizobial traits.
(A) Principal component analysis (PCA) of host investment traits, benefit traits, and rhizobial fitness outcomes, conducted at the level of individual pots. Arrows represent the loadings of each trait on the first two principal components, which together explain 51.2% and 16.7% of the total variance, respectively. To avoid loss of representation due to missing data, values for two genotypes (*latd* and *rdn1*) were averaged across replicates prior to PCA, as these genes had incomplete trait data that would otherwise be excluded by listwise deletion.
(B) All pairwise combinations of genotypic trait correlations (Pearsons). Each ellipse represents a pairwise Pearson correlation between traits, with ellipse elongation and color saturation indicating the strength and direction of the correlation (blue = positive, red = negative). Strong correlations may reflect shared genetic regulation or pleiotropy.

Tables

Table 1: Definitions of the symbiotic traits examined

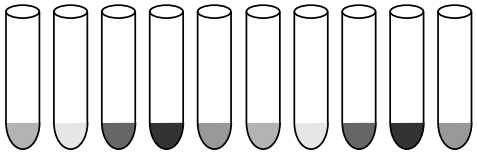
Symbiotic Traits	Category	Interpretation
CFU* per nodule	Rhizobial Benefit	Rhizobia released per nodule indicates the average individual strain fitness increase from host association
CFU* per plant	Rhizobial Benefit	Rhizobia released per plant indicates the contribution of each host plant to overall rhizobial population size
Nodule strain diversity	Rhizobial Benefit	Shannon diversity index (*H*) calculated for strains isolate from nodule indicates host selectivity
Strain relative fitness	Rhizobial Benefit	Enrichment/depletion of specific strains relative to frequency in inoculum: $\log_2$(selected freq./initial freq.)
Nodules per plant	Host Investment	Number of successful host attempts to form associations with rhizobia
Average area per nodule	Host Investment	Host investment in each nodule
Total nodule area per plant	Host Investment	Total plant investment in nodules (i.e. nitrogen fixation)
Total biomass per plant	Host Benefit	Proxy for total nitrogen fixed when plants are grown in nitrogen limited conditions
Root-to-shoot ratio	Host Benefit	Lower values indicate lower investment in roots. In a N limited environment we infer that nitrogen needs were met.

*CFU= Colony Forming Units

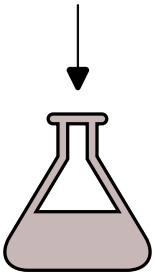
Table 2: Candidate rhizobial genes in small variant groups (three genes or fewer) associated with ($p < 10^{-5}$) variation in rhizobial relative strain fitness. See Table S6 for information on all variant groups.

LD group[1]	Host mutants	Location	Alt. allele freq[2]	Eff. alt. vars in mutants	Eff alt. vars in WT	Variants	Genes ID's[3]
9034	*dnf1,dnf5/dnf2, dnf2, dnf4, dnf7, ipd3, nad1,rdn1 sunn1, sunn4*	pSymA	0.07	+	-	1	CDO30_RS23065:SGNH/GDSL hydrolase family protein
6978.alt[4]	*dnf1, dnf1/dnf2, dnf2, dnf4, dnf7, dnf5/dnf2, ipd3, rdn1, sunn1*	pSymA	0.07	+	none	1	CDO30_RS23160:NAD(P)H-binding protein
5279	*dnf1, dnf2, dnf7, ipd3, nad1,rdn1, sunn1*	pSymA	0.07	+	-	1	CDO30_RS21460:malonyl-CoA synthase
7969	*dnf1, dnf2, dnf6, nad1, sunn4*	pSymA	0.06	+	none	1	CDO30_RS22325:hypothetical protein
6968	*dnf4, dnf7,ipd3, sunn1, sunn4*	pSymA	0.10	+	none	1	CDO30_RS23125:SDR family oxidoreductase (pseudogene)
10122	*dnf1, dnf1/dnf2, dnf2, sunn1*	pSymB	0.07	+	-	1	CDO30_RS27255:hypothetical protein
12516	*dnf1/dnf2, rdn1, sunn1*	pSymB	0.06	+	none	2	CDO30_RS28530:ABC transporter ATP-binding protein; CDO30_RS28525:*RkpR*
10601	*dnf7, ipd3, ltd*	pSymB	0.05	+	none	1	CDO30_RS28740:*pepT*, peptidase T
3770	*rdn1, sunn4*	pSymA	0.14	+	-	1	CDO30_RS18080:*fixL*, oxygen sensor histidine kinase *fixL*
12013	*dnf1/dnf2, ipd3*	pSymB	0.05	+	none	1	CDO30_RS28675:hypothetical protein
10882	*dnf4*	pSymB	0.05	+	none	1	CDO30_RS30150:glycosyltransferase
9671	*dnf3*	pSymB	0.07	+	none	3	CDO30_RS27435:four-carbon acid sugar kinase family protein
10140	*dnf3*	pSymB	0.08	+	none	1	CDO30_RS27435:four-carbon acid sugar kinase family protein
10913	*dnf3*	pSymB	0.05	+	none	2	CDO30_RS27440:xanthine dehydrogenase family protein molybdopterin-binding subunit; CDO30_RS27445:xanthine dehydrogenase family protein subunit M
10915	*dnf3*	pSymB	0.05	+	none	42	CDO30_RS27440:xanthine dehydrogenase family protein molybdopterin-binding subunit; CDO30_RS27445:xanthine dehydrogenase family protein subunit M
8294	*dnf4*	pSymA	0.07	+	none	1	CDO30_RS19605:*stc4*, stachydrine N-demethylase reductase subunit Stc4
8754	*dnf5/dnf2*	pSymB	0.1	+	none	2	CDO30_RS32445:hypothetical protein
9239	*dnf5/dnf2*	pSymB	0.09	+	none	1	CDO30_RS32445:hypothetical protein
10903	*ipd3*	pSymB	0.05	+	none	1	CDO30_RS27310:ABC transporter ATP-binding protein
838	*latd*	chr	0.09	+	none	1	CDO30_RS12270:glutamine synthetase family protein
4107	*nad1*	pSymA	0.11	-	none	1	CDO30_RS22005:amidase
4386	*nad1*	pSymA	0.09	-	none	2	CDO30_RS22060:Atu4866 domain-containing protein;CDO30_RS22055:class I SAM-dependent methyltransferase;
9865	*rdn1*	pSymB	0.07	+	none	7	CDO30_RS32685:four-carbon acid sugar kinase family protein; CDO30_RS32675:NAD(P)-dependent oxidoreductase; CDO30_RS32705:LacI family transcriptional regulator
9926	*rdn1*	pSymB	0.07	+	none	4	CDO30_RS31555:MASE1 domain-containing protein; CDO30_RS32090:TRAP transporter large permease; CDO30_RS32070:class I SAM-dependent methyltransferase
9941	*rdn1*	pSymB	0.07	+	none	1	CDO30_RS25485:pyridoxamine 5'-phosphate oxidase family protein
10728	*rdn1*	pSymB	0.5	+	none	1	CDO30_RS32910:LysR family transcriptional regulator
12841	*sunn1*	pSymB	0.07	+	none	1	CDO30_RS28530:ABC transporter ATP-binding protein

Notes: [1]LD group numbers were obtained from [16,16] , [2]AF= non-reference (alternate) allele frequency, [3]Locus tags are from the *Sinorhizobium meliloti* USDA1106 reference annotation available on MAGE (http://www.genoscope.cns.fr/agc/microscope/home/index.php), [4]A tri allelic SNP



1. Grow 86 *Sinorhizobium* isolates and mix for inoculum

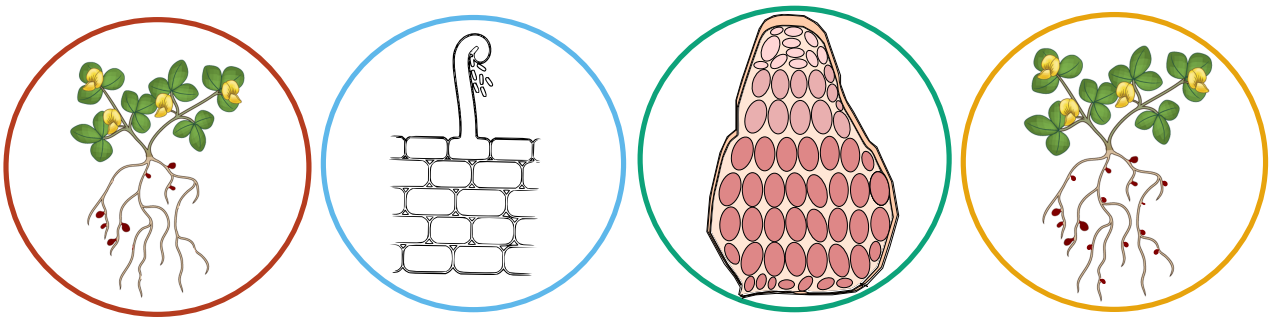


Experiment Details

- 5 pots/host mutant with 3 plants/pot
- Randomized block design
- N-free fertilizer
- Harvested at 8 weeks

2. Inoculate *Medicago* hosts mutants

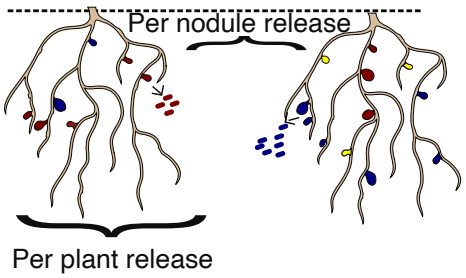
Wildtype | **Host mutants spanning stage of symbiosis**

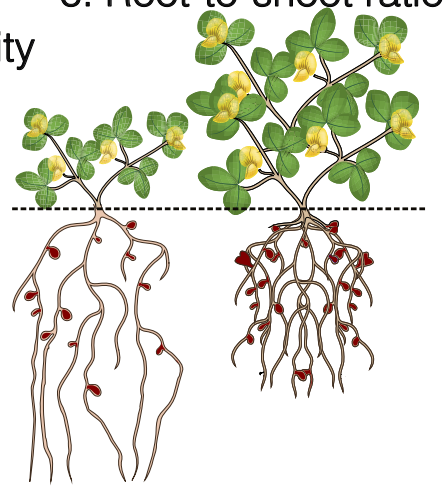


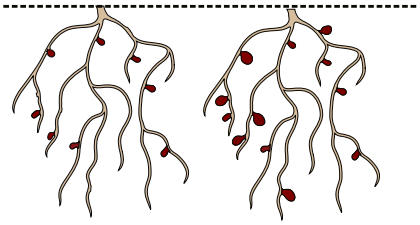
A17 | Early (6) | Late (9) | AON (3)

3. Measure host and rhizobial traits

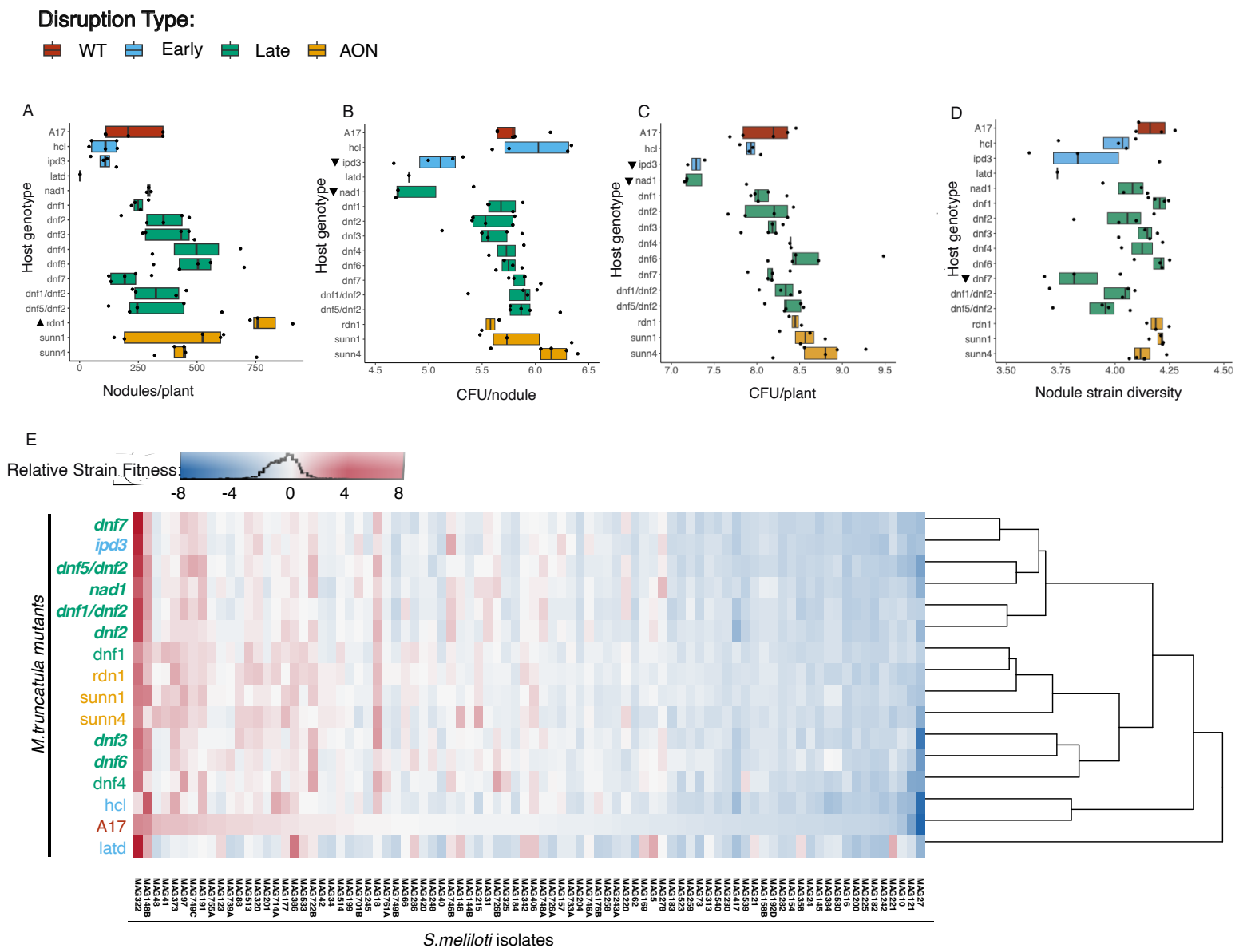
Rhizobial Benefits	Host Benefits	Host Investments
1. CFU/nodule	5. Total biomass/plant	7. Nodules/plant
2. CFU/plant	6. Root-to-shoot ratio	8. Average area/nodule
3. Nodule isolate diversity		9. Total nodulated area
4. Relative strain fitness		

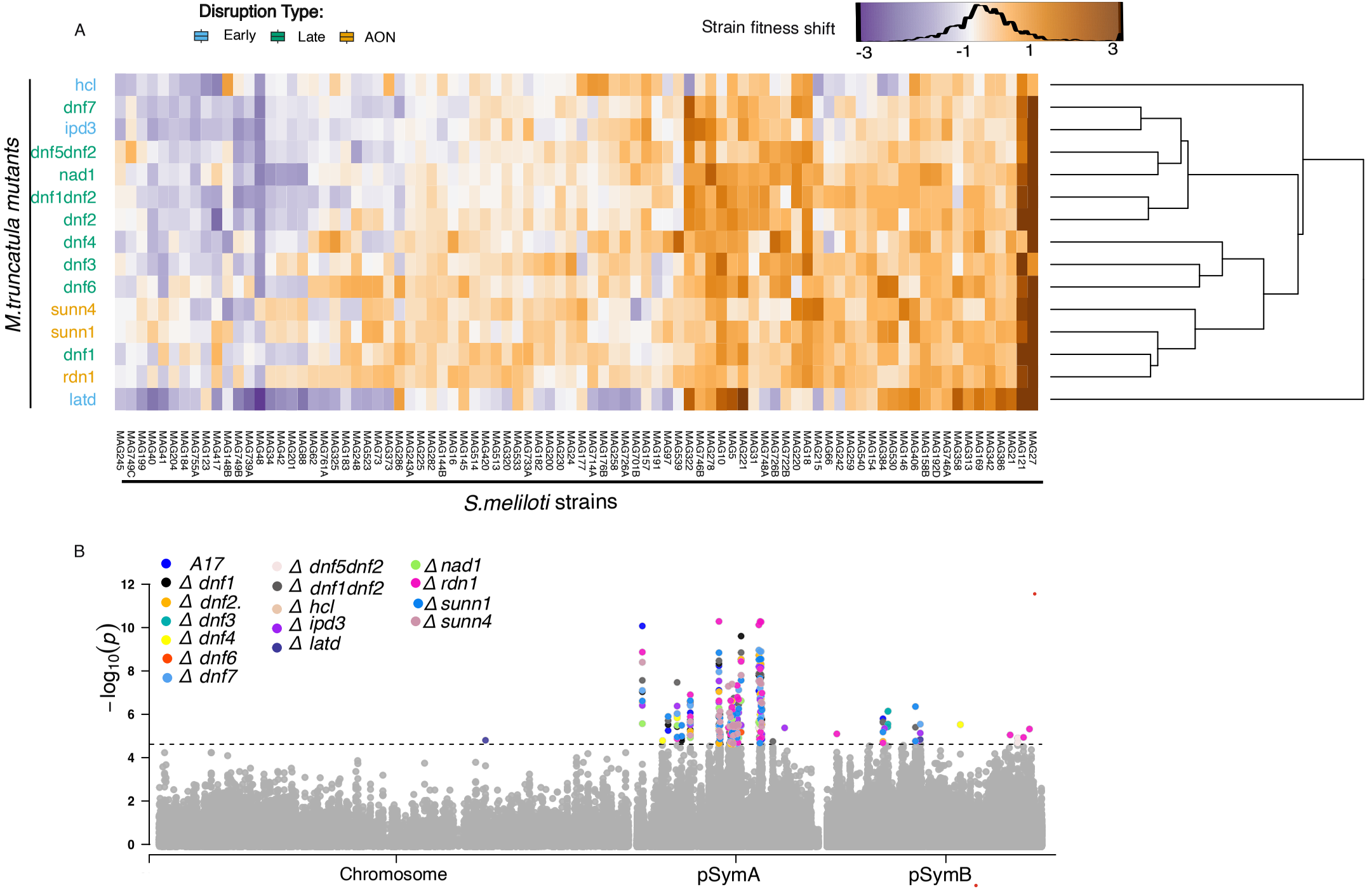


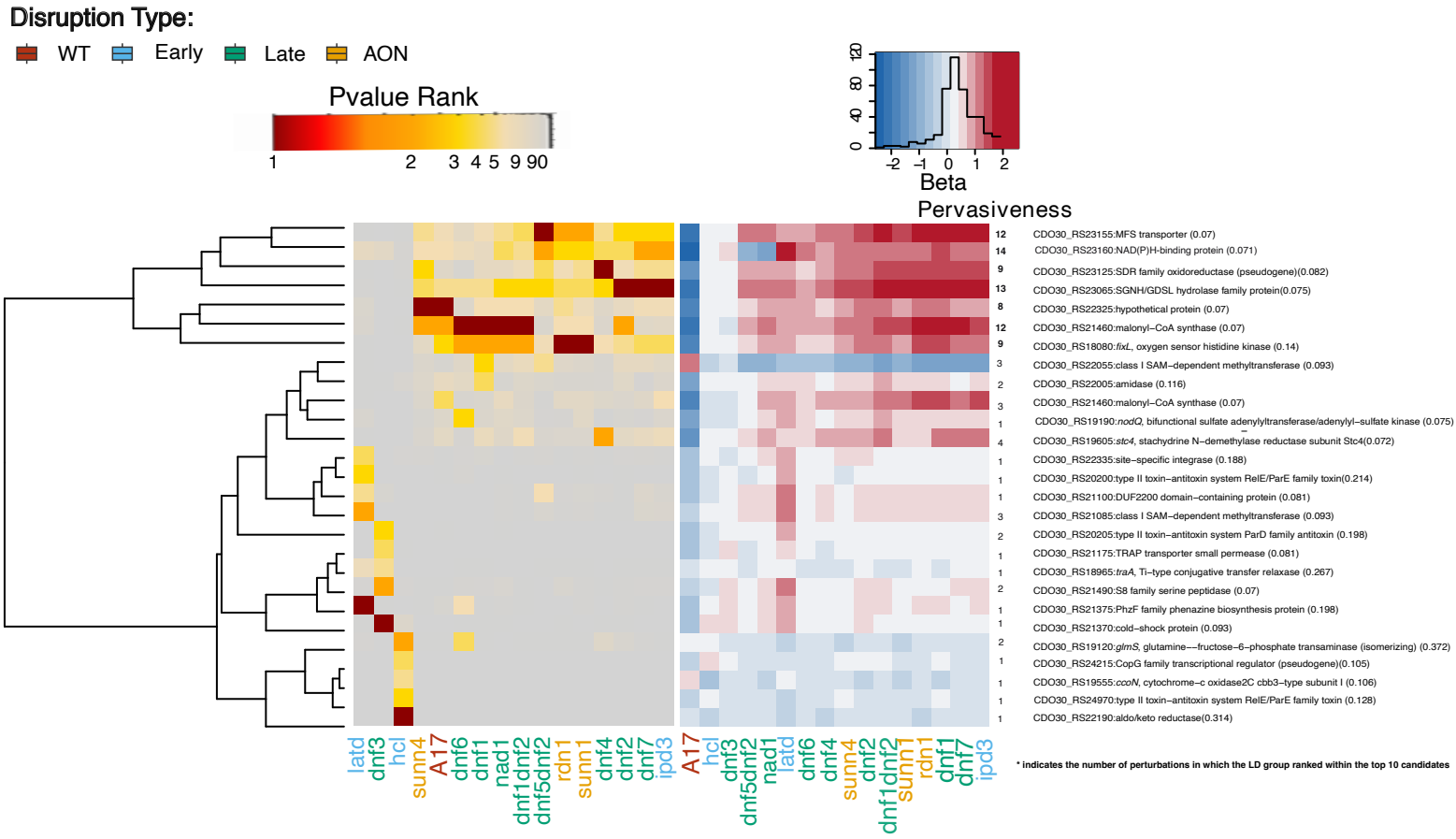


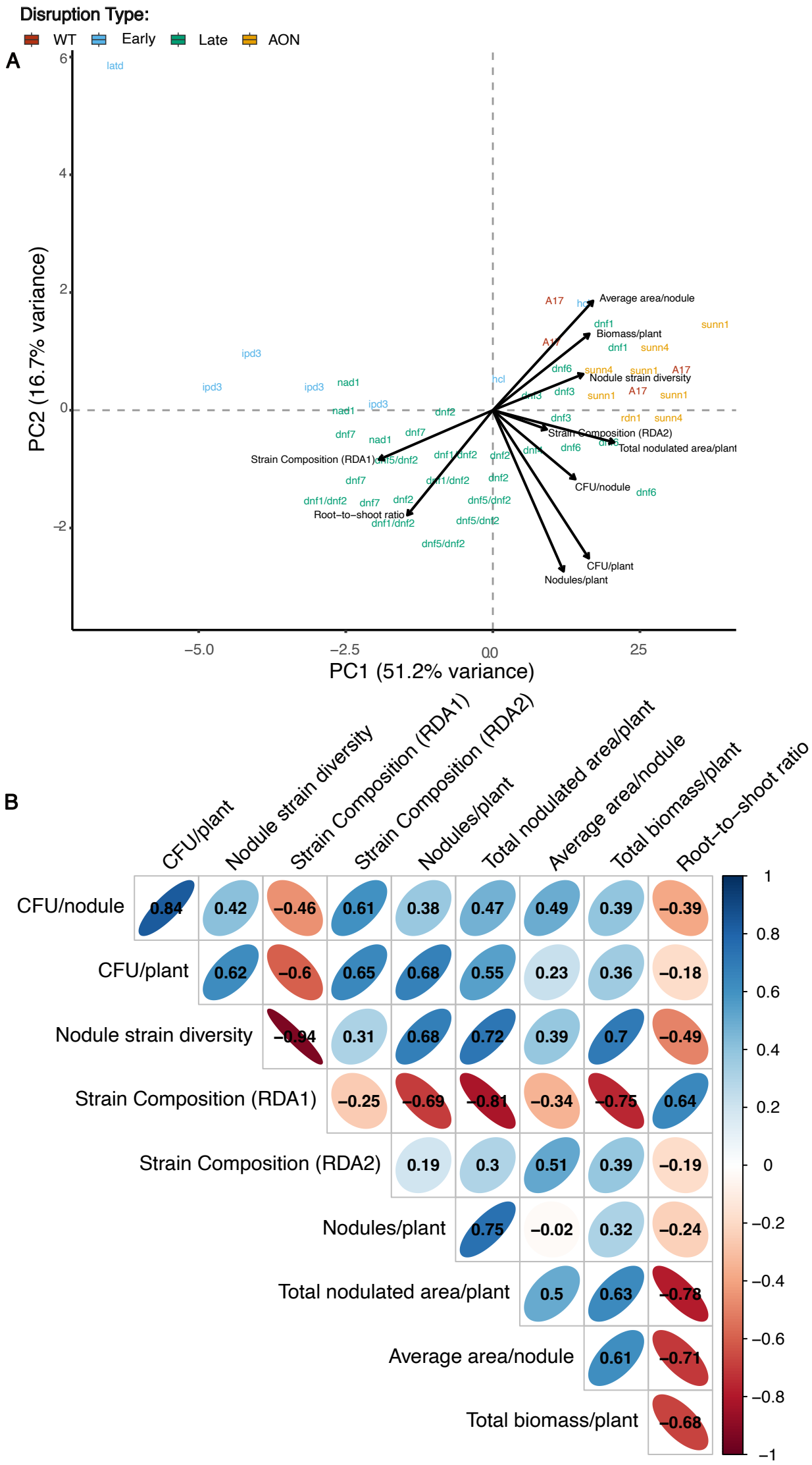


Low → High | Low → High | Low → High


Disruption Type:
WT
Early
Late
AON
A
B
C
D
Host genotype
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
Nodules/plant
CFU/nodule
CFU/plant
Nodule strain diversity
E
Relative Strain Fitness
-8
-4
0
4
8
M.truncatula mutants
dnf7
ipd3
dnf5/dnf2
nad1
dnf1/dnf2
dnf2
dnf1
rdn1
sunn1
sunn4
dnf3
dnf6
dnf4
hcl
A17
latd
S.meliloti isolates


A
Disruption Type:
Early
Late
AON
Strain fitness shift
-3
-1
1
3
M.truncatula mutants
hcl
dnf7
ipd3
dnf5dnf2
nad1
dnf1dnf2
dnf2
dnf4
dnf3
dnf6
sunn4
sunn1
dnf1
rdn1
latd
S.meliloti strains
B
A17
Δ dnf1
Δ dnf2.
Δ dnf3
Δ dnf4
Δ dnf6
Δ dnf7
Δ dnf5dnf2
Δ dnf1dnf2
Δ hcl
Δ ipd3
Δ latd
Δ nad1
Δ rdn1
Δ sunn1
Δ sunn4
−log10(p)
Chromosome
pSymA
pSymB


Disruption Type:
WT
Early
Late
AON
Pvalue Rank
1
2
3 4 5 9 90
Beta
Pervasiveness
12 CDO30_RS23155:MFS transporter (0.07)
14 CDO30_RS23160:NAD(P)H-binding protein (0.071)
9 CDO30_RS23125:SDR family oxidoreductase (pseudogene)(0.082)
13 CDO30_RS23065:SGNH/GDSL hydrolase family protein(0.075)
8 CDO30_RS22325:hypothetical protein (0.07)
12 CDO30_RS21460:malonyl−CoA synthase (0.07)
9 CDO30_RS18080:fixL, oxygen sensor histidine kinase (0.14)
3 CDO30_RS22055:class I SAM−dependent methyltransferase (0.093)
2 CDO30_RS22005:amidase (0.116)
3 CDO30_RS21460:malonyl−CoA synthase (0.07)
1 CDO30_RS19190:nodQ, bifunctional sulfate adenylyltransferase/adenylyl−sulfate kinase (0.075)
4 CDO30_RS19605:stc4, stachydrine N−demethylase reductase subunit Stc4(0.072)
1 CDO30_RS22335:site−specific integrase (0.188)
1 CDO30_RS20200:type II toxin−antitoxin system RelE/ParE family toxin(0.214)
1 CDO30_RS21100:DUF2200 domain−containing protein (0.081)
3 CDO30_RS21085:class I SAM−dependent methyltransferase (0.093)
2 CDO30_RS20205:type II toxin−antitoxin system ParD family antitoxin (0.198)
1 CDO30_RS21175:TRAP transporter small permease (0.081)
1 CDO30_RS18965:traA, Ti−type conjugative transfer relaxase (0.267)
2 CDO30_RS21490:S8 family serine peptidase (0.07)
1 CDO30_RS21375:PhzF family phenazine biosynthesis protein (0.198)
1 CDO30_RS21370:cold−shock protein (0.093)
2 CDO30_RS19120:glmS, glutamine−−fructose−6−phosphate transaminase (isomerizing) (0.372)
1 CDO30_RS24215:CopG family transcriptional regulator (pseudogene)(0.105)
1 CDO30_RS19555:ccoN, cytochrome−c oxidase2C cbb3−type subunit I (0.106)
1 CDO30_RS24970:type II toxin−antitoxin system RelE/ParE family toxin (0.128)
1 CDO30_RS22190:aldo/keto reductase(0.314)
latd
dnf3
hcl
sunn4
A17
dnf6
dnf1
nad1
dnf1dnf2
dnf5dnf2
rdn1
sunn1
dnf4
dnf2
dnf7
ipd3
A17
hcl
dnf3
dnf5dnf2
nad1
latd
dnf6
dnf4
sunn4
dnf2
dnf1dnf2
sunn1
rdn1
dnf1
dnf7
ipd3
* indicates the number of perturbations in which the LD group ranked within the top 10 candidates


Disruption Type:
WT
Early
Late
AON
A
PC2 (16.7% variance)
PC1 (51.2% variance)
Average area/nodule
Biomass/plant
Nodule strain diversity
Strain Composition (RDA2)
Total nodulated area/plant
CFU/nodule
CFU/plant
Nodules/plant
Root−to−shoot ratio
Strain Composition (RDA1)
B
CFU/plant
Nodule strain diversity
Strain Composition (RDA1)
Strain Composition (RDA2)
Nodules/plant
Total nodulated area/plant
Average area/nodule
Total biomass/plant
Root−to−shoot ratio
CFU/nodule 0.84 0.42 −0.46 0.61 0.38 0.47 0.49 0.39 −0.39
CFU/plant 0.62 −0.6 0.65 0.68 0.55 0.23 0.36 −0.18
Nodule strain diversity −0.94 0.31 0.68 0.72 0.39 0.7 −0.49
Strain Composition (RDA1) −0.25 −0.69 −0.81 −0.34 −0.75 0.64
Strain Composition (RDA2) 0.19 0.3 0.51 0.39 −0.19
Nodules/plant 0.75 −0.02 0.32 −0.24
Total nodulated area/plant 0.5 0.63 −0.78
Average area/nodule 0.61 −0.71
Total biomass/plant −0.68