Title: Mutations in legume genes that influence symbiosis 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], Vikram Venugopal[3], Siva Sankari[3], Alejandra Gil Polo[1], Garrett Levin[4], Barney Geddes[4] , Nevin D. Young[5], 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]Stowers Institute for Medical Research, Kansas City, MO, USA
[4]North Dakota State University, Department of Department of Microbiological Sciences, Fargo, ND 58102
[5]University of Minnesota, Department of Plant Pathology, St. Paul, Minnesota
[*]Corresponding author (liana.burghardt@psu.edu)

Author for correspondence: Liana T. Burghardt, Department of Plant Science, 109 Tyson Building, University Park, PA 16802, liana.burghardt@psu.edu,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 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 Nodule Cysteine-Rich (NCR) peptide production, plant defense, and nodule number regulation) with a mixture of 86 *Sinorhizobium meliloti* strains. Most mutations resulted in reduced host benefits, but the effects on rhizobial benefit (i.e., relative strain fitness) varied widely, revealing widespread host-by-strain fitness interactions. Genome-wide association analyses identified variants on rhizobial replicons pSymA and pSymB as important in mediating strain fitness responses to host mutations. Whereas most top variants 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 some 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. Competition assays across *Medicago* genotypes confirmed a pervasive role for one candidate (malonyl-CoA synthase), and AlphaFold multimer modelling suggests that many rhizobial top candidates could interact with host NCR peptides. Our results reveal how host genetic variation affects rhizobial fitness, setting the stage for improving rhizobial inoculants and breeding legume hosts better adapted to 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 (GxG) 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 (GxG) 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 (Genome Wide Association Studies) [5,6] suggest complex genetic architectures underlie these host-pathogen interactions [7,8]. More recently, studies of (GxG) interactions have expanded to mutualisms [9–12]. These studies suggest that (GxG) 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 (GxG) interactions by measuring how over a dozen host mutations alter the selective landscape experienced by scores of strains of mutualistic bacteria.

Legume hosts and rhizobium 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. Because 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, because all strains are already sequenced, 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, whereas 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 alter *relative strain fitness.* For instance, members of the NPD gene family in *Medicago truncatula* 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 genes alters selection on their nitrogen-fixing rhizobium symbionts *in planta*. To accomplish this goal, we chose eighteen existing *Medicago truncatula* mutant lines with disruptions in the known symbiotic genes. These genes

110 affect a wide range of host phenotypes and encompass the developmental trajectory of
111 nodulation. We inoculated these mutants with a mixture of 86 *Sinorhizobium meliloti* strains
112 (Fig. 1). We studied standard host investment and benefit traits, as well as multiple aspects of
113 rhizobial fitness (Table 1). We then used a S&R approach to measure the frequencies of rhizobial
114 strains (relative strain fitness) in nodules [25] and followed up with GWAS, to identify variants in
115 rhizobium genes that altered strain fitness when host genes were perturbed. We examined the
116 following questions: 1) Which host symbiosis genes alter rhizobial fitness? 2) Do altered host
117 genotypes have strain-specific effects on rhizobial fitness (i.e., is there evidence of genome x
118 genome interactions)? 3) If yes, which rhizobium genes harbour genetic variation underlying the
119 observed shifts in strain fitness, and are any of those variants selected in multiple host mutations?
120 And 4) Are any host investment or benefit traits correlated with rhizobial fitness traits? Whereas
121 few host genotypes altered rhizobial population sizes or nodule strain diversity, many hosts did
122 shift relative strain fitness. Most rhizobial genes associated with strain fitness shifts were found
123 to be important within specific hosts, although eighteen variants, (primarily on pSymA) were
124 important with six or more hosts, suggesting candidate genes that may play fundamental roles in
125 in-plant rhizobial fitness.
126
127 Materials and methods
128
129 ***Host symbiotic mutants:*** The development of symbiosis in nodules is complex, with each step
130 regulated by multiple hosts and rhizobium genes [27,36,37]. To encompass these steps, we chose 18
131 *M. truncatula* symbiosis mutants (Table S1), which we grouped into three categories: early-
132 acting, late-acting, and regulation of nodule number (Autoregulation of nodule number; AON) [38].
133 The early-acting genes (*hcl, ipd3*, *latd, dmi1-2*, *dmi2-3,* and *dmi3-1*) affect the initial processes,
134 such as Nod Factor recognition and nodule organogenesis [39–43]. The late-acting genes (*nad1,*
135 *dnf1, dnf2, dnf4, dnf2/dnf1, dnf5/dnf2, dnf6, dnf7*) influence nitrogen fixation [32,44], whereas the
136 regulation of nodule number genes are in the AON (Autoregulation Of Nodulation) pathway
137 (*sunn1, sunn4, rdn1*) [45,46]. Whereas these categorizations help set expectations for the impact of
138 the host mutations, each gene may have additional pleiotropic effects. Together, these mutants
139 provide a broad view of the effects of host symbiosis genes on their rhizobial partners in multi-
140 strain inoculations. Seeds for all host mutant lines were obtained from Dr. Dong Wang,
141 Department of Biochemistry and Molecular Biology, University of Massachusetts, Amherst. The
142 host mutant lines used here were generated through forward genetic screens and thus may carry
143 additional background mutations. Our aim is not to assign strict causality to individual host
144 mutations but rather to examine how alterations in the host genetic landscape broadly influence
145 rhizobial fitness.
146
147 ***Initial inoculum preparation:*** We created a mixed inoculum using 86 sequenced *Sinorhizobium*
148 *meliloti* (henceforth *S. meliloti*) strains, which were part of a collection isolated from soils from
149 21 locations in Spain and France, using the nodules of *Medicago* plants as a "trap" for capturing
150 *S. meliloti* isolates [47]. We grew each strain separately for 72 hours in Tryptone-Yeast extract

151 (TY) media at 28°C, shaking at 120 rpm, before mixing equal volumes of each culture to create
152 an initial inoculum with roughly equal proportions of each strain. To inoculate plants, we added
153 60 ml of the multi-strain mixture to 6 litres of sterile 0.85% NaCl solution and applied 10^7 cells
154 per pot. This density was chosen to reflect *S. meliloti* populations in field soil [48] and to avoid
155 triggering quorum-sensing mechanisms [49], which typically occur at higher cell densities. We
156 measured rhizobial density in the inoculum by plating a dilution series and counting the number
157 of colonies. We measured the strain frequencies in the inoculum by centrifuging eight 1 ml
158 aliquots at 18,000 x g for 2 minutes, removing the supernatant, and then freezing the pellet at -
159 20°C before extracting DNA. Initial communities were sequenced, and frequencies inferred (Fig.
160 S1) as outlined in the Rhizobial Fitness section below.
161
162 ***Greenhouse experiment to measure multi-strain symbiotic phenotypes:*** We grew five replicate
163 pots of each of the 18 *Medicago* mutants plus wildtype A17 (Fig. 1). Before planting, seeds were
164 scarified by gently abrading with sandpaper, surface sterilized by soaking in a 10% bleach
165 solution for 90 seconds and rinsed in de-ionized water. Seeds were stratified on wet filter paper
166 at 4°C in the dark for 2 days, then planted in pots containing steam-sterilized Sunshine Mix #5.
167 Germinated seedlings were thinned to three plants per pot. All pots were inoculated 48 hours
168 after seed planting with the initial 86-strain inoculum. Following inoculation, pots were
169 randomized into five spatial blocks with one representative of each genotype per block. Plants
170 were watered as needed and fertilized weekly using increasing concentrations of N-free Fårhaeus
171 medium [50]. Uninoculated plants were included to monitor cross-contamination. Pots were
172 harvested and traits measured 8 weeks post-inoculation. We measured nodule number per plant,
173 average area per nodule and total nodulated area per plant as measures of host investment, total
174 biomass per plant and root-to-shoot ratio as measures of host benefit, and population size of
175 rhizobia released per plant and per nodule, nodule strain diversity and relative strain fitness in
176 nodule pools as measures of rhizobial fitness. After 8 weeks, we harvested inoculated plants and
177 pooled nodules from ~ 3 plants per pot to ensure enough nodules (>100 /pot) to limit stochastic
178 variation and saturate diversity. None of the uninoculated plants formed nodules. We
179 photographed nodule pools per replicate to assess nodule number and size using ImageJ. Plant
180 biomass was measured by weighing the plant roots and shoots after oven-drying at 75 °C for 3
181 days. Because not all plants survived in every pot, the number of plants occasionally varied. To
182 control for this variation, all host traits were normalized per plant for analysis.
183
184 ***Rhizobial fitness:*** We surface-sterilized the nodules from each pot in 10% bleach for 1 minute,
185 rinsed the nodules in sterile water, and homogenized them for 90 seconds in 4 ml sterile 0.85%
186 sodium chloride solution using an Omni TH tissue homogenizer. We estimated the number of
187 rhizobia released from the nodule pools by dilution plating a 500-1000 ul aliquot of the
188 homogenate and counting the number of colonies formed. We then estimated the ***absolute fitness***
189 of the rhizobia on both a per-plant and per-nodule level by dividing the total number of colonies
190 by the number of plants or nodules harvested from each pot and plotting the $\log_{10}$ transformation
191 of this value. The remaining nodule homogenate was spun at 400 x g for 10 min at 10°C. to
192 pellet the bacteroids and the plant debris. We collected the supernatant enriched for

193 undifferentiated rhizobial cells, pelleted them at 18000 x g for 5 min, pipetted off the supernatant
194 before freezing the pellets at -20 °C until DNA extraction. DNA was extracted using a Qiagen
195 DNeasy® Plant extraction kit following the manufacturer's protocol, except we used 3 µL of
196 RNase in the first step and 50 µL of nuclease water for the final elution.
197
198 Samples were sequenced using Illumina short-read technology at the Huck Genomic Core
199 Facility at The Pennsylvania State University. Indexed libraries were generated using the
200 Illumina DNA Prep Kit. 12X libraries were pooled and sequenced on a 150 x 150 paired-end
201 NextSeq High Output run, resulting in an average read depth of ~75. We trimmed the reads using
202 TrimGalore! (v0.4.1, www.bioinformatics.babraham.ac.uk/projects/trim_galore/) with default
203 settings, except with minimum read length: 100 bp; quality threshold: 30 (−-quality); and
204 minimum adapter match: 3 (−stringency), used bwa mem (v0.7.17) with default settings to align
205 reads to the *S. meliloti* USDA1106 genome (NCBI BioProject: PRJNA388336) and estimated
206 strain frequencies using HARP[51]. Strain frequencies were used to 1) estimate ***nodule strain***
207 ***diversity*** using Shannon's H index (via the diversity function from the vegan package in R) and
208 2) calculate ***relative strain fitness*** of each strain in each replicate of a particular host mutant by
209 taking the $\log_2$ of the ratio of the final strain frequency in the nodules to the initial frequency in
210 the inoculum (median initial frequency = 0.0092, 5%-95% quantile: 0.006-0.018). The log
211 transformation puts increases and decreases in proportional representation on the same scale
212 (e.g., a fourfold decrease and increase will have a value of −2 and 2, respectively). Strains with
213 frequency estimates of 0 in the selected communities were assigned a fitness value of −8, less
214 than the lowest measurable reductions in other strains [25].
215
216 ***Statistical analyses:*** To assess the effects of the host mutants and the symbiotic stage on plant
217 traits and rhizobial fitness, we used linear models and analysis of variance (ANOVA) for the
218 univariate traits and redundancy analysis (RDA) for the multivariate trait of relative strain
219 fitness. All analyses were conducted in the R statistical environment (version 4.3.2). For each
220 univariate trait, we fit two linear models using the `lm()` function: one that includes host genotype
221 and block, and a second that includes the symbiotic disruption stage and block as additive
222 predictors. We compared variation explained by the two models using the adjusted R^2
223 (`anova()` function) to infer whether models that included mutant identity explained more
224 variance than the symbiotic stage alone. Because the genotypes are nested within the stage of
225 symbiosis, we ran two separate linear models to avoid introducing collinearity. We then
226 compared the variance explained by each model. To ensure this comparison was not biased by
227 differences in the number of levels (degrees of freedom) in each predictor, we used the adjusted
228 R^2, which corrects for model complexity and sample size. This approach allows us to assess the
229 relative explanatory power of broad functional categories (stage) versus specific genetic
230 backgrounds (genotypes) on rhizobial community structure. We $\log_{10}$-transformed average
231 nodule area/plant and total nodule area traits to reduce heteroscedasticity of variance. To assess
232 pairwise differences between A17 and each mutant for each trait, we used Tukey's Honest
233 Significant Difference (HSD) tests. Comparisons with an adjusted P-value < 0.05 were
234 considered statistically significant. To investigate how host genotype or symbiotic stage altered

235 the relative fitness of rhizobial strains (a multivariate trait), we performed an RDA (`rda()`
236 function of the vegan package, v2.6-4) using relative strain fitness as the response matrix and
237 host genotype or symbiotic stage as the explanatory variable. The significance of each model
238 was assessed by extracting the adjusted R^2 and via permutation-based ANOVA (1,000
239 permutations) using the `anova()` function with `by = "terms"`. To test whether the relative strain
240 fitness within each mutant differed from A17, we ran the model for each mutant-A17 pairwise
241 contrast. Predictors with $P < 0.05$ indicated significant changes in strain composition.
242
243 ***Genome-wide association study on relative strain fitness shifts in mutants vs. wildtype nodules***
244 The relative strain fitness shift of each strain for each host mutant was calculated by subtracting
245 the relative strain fitness in A17 from the relative strain fitness in the host mutants. We
246 conducted 15 GWAS (three mutants formed no nodules) to identify rhizobial variants underlying
247 host mutation-dependent ***strain fitness shifts*** for each nodulating mutant. To be thorough, we
248 also conducted a GWAS on strain fitness in each host genotype individually, which highlights
249 variants associated with strain success in host nodules generally, as well as the effect of the
250 perturbations. For each of these GWAS, we ran a linear mixed model (LMM) using GEMMA
251 (version 0.98.3). Analyses were run separately for the three replicons (chromosome, pSymA,
252 pSymB) to account for the different population structures. We used a published database of
253 30,076 SNPs representing previously defined linkage-disequilibrium (LD) groups within *S.*
254 *meliloti*[16]. SNP filtering was performed with bcftools-1.18 [52] and PLINK 1.9 [53] based on minor
255 allele frequency (0.05) and missingness (0.2). We selected the top five variant groups for each
256 mutation-driven fitness shift tested based on their *P*-value ranks, cross-compared in the other 15
257 GWAS. To control for multiple tests while enabling comparisons across replicons that differed in
258 variant group size from ~ 3,000 to 12,000, we adopted a Bonferroni-corrected *P*-value cutoff of
259 $P < 10^{-5}$ to approximate a nominal $P < 0.05$ across replicons. This choice of a consistent
260 intermediate threshold aligns with our goal of examining selection landscapes rather than
261 identifying individual causal variants. This allowed us to categorise a rhizobial candidate as
262 'pervasive' if it ranked in the top ten within six or more host mutants or 'limited effect' if it
263 appeared in five or fewer. We focused on ranks rather than *P*-values for this comparative
264 analysis, prioritizing variants with the most evidence, regardless of differences in power or
265 genetic architecture across host genetic perturbations.
266
267 ***Functional validation of malonyl CoA synthase gene through competition assays :*** To validate
268 the functional role of a candidate gene identified through GWAS, we mutated the malonyl-CoA
269 synthase gene within the genetically amenable *S. meliloti* Rm1021 background (mutant strain
270 named RmND1439). We competed the mutant against the corresponding wild type Rm1021,
271 across four host genotypes (A17, *dnf3*, *dnf7*, and *sunn4*). The mutant and wild-type strains were
272 cultured individually and mixed in equal proportions to create the initial inoculum. Strain
273 representation in the inoculum and post-inoculation nodule pools was quantified by performing a
274 dilution series and counting colony-forming units (CFU) on selective and non-selective plates.
275 Competitive fitness was expressed as the $\log_2$ fold change in the proportion of RmND1439

recovered from the nodules relative to its proportion in the initial inoculum. Detailed methods for strain construction and competition assays are provided in the Supplementary Methods.

In-silico prediction of the rhizobial candidate gene functions through protein-protein interaction using AlphaFold-Multimer and Foldseek: To obtain additional information about the potential for the protein products of the candidate rhizobial genes to interact with host proteins such as the NCR peptides directly, we performed in silico protein–protein interaction predictions using AlphaFold-Multimer [54,55]. The rhizobial candidate proteins were putatively annotated using Foldseek[56]. To infer candidate gene involvement during symbiosis, we also listed the nodule-enrichment scores of each rhizobial candidate from the Symbimics database (https://iant.toulouse.inra.fr/symbimics/) database[57]. These scores quantify the likelihood of the rhizobial candidate gene involvement during root nodule symbiosis, with higher values indicating stronger nodule-specific gene expression. Detailed procedures for structural modelling, interaction scoring, and data filtering are provided in the Supplementary Methods.

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 of rhizobia within nodules. This makes sense given that *nad1* host mutant is known to evoke a defense-like response in nodules that kills reproductively active bacteria, and *ipd3* host mutant has aborted infection threads, the part of the nodule in which reproductively competent rhizobia are found [43,58].

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 with significantly reduced *nodule strain diversity* (P = 0.009, Fig. 2C, Table S3). Second, we compared the relative fitness of each strain in the nodules of each host

318 mutant to A17 (Fig. 2E). Hierarchical clustering revealed that strain relative fitness did not
319 cluster solely by symbiotic stage and that the *latd* nodule strain fitness was quite different from
320 all other hosts, including wildtype. The *latd* host mutant results should be interpreted carefully—
321 *latd* uniquely had a single estimate for rhizobial fitness traits, because plants formed so few
322 nodules that we had to combine nodules across all pots to create a large enough sample for
323 sequencing. A redundancy analysis with PERMANOVAs reinforced this finding. Whereas the
324 symbiosis stage explained some variation in relative strain fitness, symbiotic genes had strong
325 individual effects (Fig. S2, Table S4, S5). Both early-acting (*ipd3* and *nad1)* and late-acting (all
326 *dnf* mutants except *dnf4,* which was marginal at $P = 0.06$, and *dnf1*) significantly altered
327 rhizobial relative fitness in nodules (Fig. 2E). The *dnf* mutants did not cluster together in the
328 RDA space, reinforcing that defects in nitrogen fixation did not produce uniform effects on
329 rhizobial enrichment. Some strains consistently exhibited high or low fitness across all host
330 genotypes (e.g., MAG322 vs. MAG27), but the fitness of most strains depended on host
331 symbiotic mutations and often changed the rank order of relative strain fitness of the strains
332 compared to wildtype (Fig. S3). , Strain shifts were often specific to individual host mutants; for
333 example, MAG726B, MAG701B, and MAG201 had high fitness in *dnf4* but low fitness in *latd*.
334 The host-dependent changes in strain fitness allow us to use GWAS to identify rhizobial allelic
335 variants associated with the observed strain fitness shifts caused by each mutation relative to
336 wildtype.
337
338 *Rhizobial genes underlying strain fitness shifts in response to host mutations were primarily*
339 *located on extrachromosomal replicons*
340 To determine genetic variants associated with shifts in strain fitness when the host genes were
341 disrupted, we performed 16 GWAS, one for the shifts in strain fitness caused by each of the 15
342 mutants (relative fitness in mutant-relative fitness in A17) and for comparison, one for relative
343 fitness in A17 (Fig.3, Fig. S4). We found 9 linkage groups on pSymA and 17 linkage groups on
344 pSymB that were strongly associated with strain fitness shifts ($P < 10^{-5}$) (Table 2, Table S6). In
345 contrast, linkage groups on the chromosome had weaker effects (none with $P < 10^{-5}$), except for
346 LD group 838 ($P = 6.38 \times 10^{-6}$), which was associated with increased strain fitness with *dnf6* and
347 *latd*. The *P* values of variants across different replicons may differ in part due to replicon-
348 specific patterns of relatedness among strains. However, the impact of population structure on
349 GWAS power is not always straightforward to predict. For instance, the chromosome may
350 exhibit long-range linkage disequilibrium, resulting in large linkage blocks that tag multiple
351 genes, which can both help or hinder association resolution. Additionally, we ran a second
352 GWAS to identify the genes underlying strain fitness in each host (Fig. S5, Table S7). As
353 expected, within the top five candidates, 11% were shared between A17 and the host mutants
354 (Table S8), the rest were unique for the mutants, indicating host genetic make-up influences
355 strain fitness (Table S9). Whereas there was some overlap in genetic variants across the two
356 GWAS (Table S10), the candidates for shifts in strain fitness in response to host genetic
357 perturbations were largely distinct (only 20 out of 113 variant groups of strain fitness shifts
358 overlapped with strain fitness) (Table S11). Below, we focus on candidates from the fitness shift

analyses because we are interested in rhizobial responses to changes in the host environment, rather than general fitness determinants.

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 S12). Most of the candidates on the chromosome were classified as 'limited effect' variant groups because they ranked among the top ten candidates, in five or fewer host mutations (Table S12). 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, peptidase T, and a hypothetical protein (Fig. S6). 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). We validated one pervasive candidate, the malonyl-CoA synthase, by knocking it out in *S. meliloti* Rm1021 genetic background (See Supplemental Methods) and competing the mutant (RmND1439) against the wild type strain across four host genotypes (*sunn4*, *dnf3*, *dnf7*, and A17). RmND1439 was almost completely absent from nodules, indicating a functional version of this gene is crucial for strain fitness in all hosts tested (Fig. S7). This result supports the gene's pervasive effect and confirms that our approach can identify rhizobial genes influencing strain performance. We further observed that 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 S12). 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.

Most limited-effect variant groups (82 of 94) were small (having fewer than three genes). The majority of these were on pSymB (Table S13). Those with strong statistical support (*P* value$<10^{-5}$) 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*). 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 the *latd* genotype*,* which encodes a nitrate transporter [59]. This was unexpected

because rhizobial glutamine synthetases are generally thought to be strongly downregulated in nodules where host-driven nitrogen assimilation processes dominate [57]. To further investigate the potential functional relevance of these candidates, we utilized AlphaFold Multimer, coupled with Foldseek, to predict possible protein–NCR peptide interactions. Host NCR peptides are a logical starting point for putative plant protein-rhizobial protein interactions because 1) several of our hosts are NCR mutants [27], 2) NCRs are known to have strain-specific effects [60] and 3) the large NCR family is derived from antimicrobial peptides, but we only know what a handful of them do[61–63]. We found that more than half of the top rhizobial candidate genes associated with strain fitness shifts (20 out of 35) had a high likelihood of physically interacting with one or more NCR peptides (Table S13, S14). These included rhizobial candidates associated with fitness shifts in NCR-compromised mutants whose annotations suggest that rhizobia might adjust their membrane transport, redox balance, and surface defences when host NCR activity changes (e.g., xanthine dehydrogenase family protein subunit M, a Teicoplanin resistance protein-like, and an ABC transporter ATP-binding protein, which could be an NCR efflux protein according to Foldseek). These findings provide additional support for the biological relevance of these candidates in mediating rhizobium–host interactions.

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 the number of 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). 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 a manner that reinforces 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 known host symbiotic genes create distinct environments that shape rhizobial strain success and identify rhizobial candidate genes underlying strain fitness shifts in response to host mutations. By applying a multi-strain inoculation approach here, we have created a more natural ecological setting to evaluate host-imposed selection on naturally occurring rhizobial genetic variants. Within this ecological framework, 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, which highlight a directional effect where host genetic variations reshape selection on rhizobial partners. 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 [28,48,64]. 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]. Whereas our results suggest that the effects of host genetic changes on rhizobial fitness may sometimes align with known gene functions,, we also observed broader positive correlations between rhizobia released per nodule and average nodule area 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 [65]. However, the host mutant lines used in this study were generated through forward genetic screens and, therefore, will likely carry additional background mutations. Thus, our aim is not to assign strict causality to individual mutations but rather to examine how alterations in the host genetic landscape broadly influence rhizobial fitness.

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 [63,66–68]. Although late-acting NCRs can be important determinants of host strain specificity in *M. truncatula*, our investigation reveals a more complex landscape [27,69]. Our analyses of two NCR mutations revealed differing effects on strain fitness within nodules. The *dnf7* mutant, which lacks functional NCR169 [62], was the only mutant background where we observedreduced nodule strain diversity, and it clustered with *ipd3,* which also supported smaller rhizobial population. In contrast, *dnf4*, which does not produce NCR211 and influences rhizobial differentiation during bacteroid formation [70], showed shifts in strain relative fitness similar to *dnf3* and *dnf6,* suggesting these uncharacterized mutants may also involve NCR-related

processes. A few rhizobia genes (*bacA, hrrP, yejABEF*) are known to affect the efficacy of NCRs [35,66,71]; however, none of those genes were top candidates in our GWAS, suggesting additional mechanisms of host restriction of rhizobia [61,72,73].

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 mutant genotypes (pervasive genes), and ~82% were associated with 5 or fewer host mutants(limited effect genes). , Pervasive effect alleles tended to exhibit fitness trade-offs; alleles linked to higher fitness in host mutants tended to be linked with 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 had little to no effect on strain fitness in the wild-type background. Most candidate genes with pervasive associations were highly expressed in the nitrogen fixation zone (Zone III) [57]. 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 [77], 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 [78]. These genes represent strong candidates for uncovering novel mechanisms that influence symbiont performance across diverse host contexts[79].

Beyond these candidates, most associations appeared host-specific, indicating that each mutated host background 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 is linked to enhanced strain fitness in *dnf4* mutants, possibly by protecting differentiating rhizobia [70]. Whereas these genes are candidates for altering strain fitness in some host mutant contexts and not others, our experiments do not pinpoint the stage at which 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, whereas our design allows us to detect which host symbiotic genes alter strain fitness, it does not allow us to 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]. Whereas data suggest the effects of microbe × microbe interactions on strain relative fitness in our system are weak [26], we 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 affects the fitness of rhizobial strains. 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 [80] or the cooccurrence of host genotypes across time and space [13,81].

Conclusion

Despite legumes being used for soil fertility improvement and actively managed for centuries, biological nitrogen fixation in agriculture remains inconsistent spatially and temporally [83–86]. This challenge has driven interest and investment in developing microbial inoculants to enhance crop yield, stress tolerance, and disease and pest management [20]. Rhizobia have been included in microbe-based bioinoculant formulations for centuries [79]. 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 that show promise as potential bioinoculants under controlled conditions often fail to reach densities high enough to influence crop yields in realistic agricultural systems [87]. Formulating ways to sustain rhizobial fitness in agricultural fields is a critical avenue for plants to continue acquiring nitrogen from their associated rhizobia [88].

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 [85,89]. 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 [90]. 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]. Whereas 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 the University of Massachusetts Amherst for generously providing the *M. truncatula* mutant seeds. We also appreciate the assistance of Lillian Cherry

and Jennifer E. Harris with the tedious task of picking and processing nodules. Prof. Katy Heath of the University of Illinois Urbana-Champaign and Dr. Mike Grillo (currently Assistant Professor at Loyola University, Chicago) for providing strains and sequencing data. This work was supported by the National Science Foundation grants IOS-1856744 to PT, NY, and LTB, and IOS-2243819, 1724993, 2243817, and 2243819 to PT and LTB. It was completed using computing resources provided by the Penn State Institute for Computational and Data Science, and sequencing was performed by the Huck Genomics Core Facility. LTB's work was supported by the USDA National Institute of Food and Agriculture and Hatch Appropriations under Project #PEN04760 and Accession #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; RB, GF, LTB data collection for main experiment; GL, AG, BG constructed the malonyl-CoA synthase mutant, SG did the competition assays; VV and SS did the Alphafold Multimer and Foldseek analysis; SG, LTB, RGB, BE, JS, S&R and GWAS analysis; LTB, SG, RB drafted the manuscript; SG, LTB, RB, JS, BE, PT, NY, GF, BG, GL,VV, AG and SS for the 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)

596 References

597 1. Friesen, M. L. *et al.* Microbially mediated plant functional traits. *Annu. Rev. Ecol. Evol. Syst.*
598 **42**, 23–46 (2011).
599 2. Percy, D. M. John N. Thompson, The Geographic Mosaic of Coevolution. *Syst. Biol.* **56**,
600 149–151 (2007).
601 3. Bartha, I. *et al.* A genome-to-genome analysis of associations between human genetic
602 variation, HIV-1 sequence diversity, and viral control. *eLife* **2**, e01123 (2013).
603 4. Jones, J. D. G. & Dangl, J. L. The plant immune system. *Nature* **444**, 323–329 (2006).
604 5. Amandine, C. *et al.* Unraveling coevolutionary dynamics using ecological genomics. *Trends*
605 *Genet. TIG* **38**, 1003–1012 (2022).
606 6. Demirjian, C., Vailleau, F., Berthomé, R. & Roux, F. Genome-wide association studies in
607 plant pathosystems: success or failure? *Trends Plant Sci.* **28**, 471–485 (2023).
608 7. Corwin, J. A. *et al.* The Quantitative Basis of the Arabidopsis Innate Immune System to
609 Endemic Pathogens Depends on Pathogen Genetics. *PLOS Genet.* **12**, e1005789 (2016).
610 8. Newberry, E. A., Minsavage, G. V., Holland, A., Jones, J. B. & Potnis, N. Genome-Wide
611 Association to Study the Host-Specificity Determinants of *Xanthomonas perforans.*
612 *Phytopathology®* **113**, 400–412 (2023).
613 9. Medina, M. & Sachs, J. L. Symbiont genomics, our new tangled bank. *Genomics* **95**, 129–
614 137 (2010).
615 10. Clare, E. L., Schiestl, F. P., Leitch, A. R. & Chittka, L. The promise of genomics in the study
616 of plant-pollinator interactions. *Genome Biol.* **14**, 207 (2013).
617 11. *Mutualism*. (Oxford University Press, Oxford, United Kingdom, 2015).

618 12. Batstone, R. T. Genomes within genomes: nested symbiosis and its implications for plant
619 evolution. *New Phytol.* **234**, 28–34 (2022).
620 13. Heath, K. D. & Stinchcombe, J. R. Explaining mutualism variation: a new evolutionary
621 paradox? *Evolution* **68**, 309–317 (2014).
622 14. Turkarslan, S. *et al.* Synergistic epistasis enhances the co-operativity of mutualistic
623 interspecies interactions. *ISME J.* **15**, 2233–2247 (2021).
624 15. Batstone, R. T. *et al.* Genome-wide association studies across environmental and genetic
625 contexts reveal complex genetic architecture of symbiotic extended phenotypes. *mBio* **13**,
626 e01823-22 (2022).
627 16. Epstein, B. *et al.* Combining GWAS and population genomic analyses to characterize
628 coevolution in a legume-rhizobia symbiosis. *Mol. Ecol.* **32**, 3798–3811 (2023).
629 17. Carlson, C. & Frederickson, M. E. An emerging view of coevolution in the legume–
630 rhizobium mutualism. *Mol. Ecol.* **32**, 3793–3797 (2023).
631 18. Gohar, D. *et al.* Genomic evidence of symbiotic adaptations in fungus-associated bacteria.
632 *iScience* **28**, 112253 (2025).
633 19. Denison, R. F., Kiers, E. T. & West, S. A. Darwinian Agriculture: When Can Humans Find
634 Solutions Beyond the Reach of Natural Selection? *Q. Rev. Biol.* **78**, 145–168 (2003).
635 20. Busby, P. E. *et al.* Research priorities for harnessing plant microbiomes in sustainable
636 agriculture. *PLOS Biol.* **15**, e2001793 (2017).

637 21. Burghardt, L. T. & diCenzo, G. C. The evolutionary ecology of rhizobia: multiple facets of
638 competition before, during, and after symbiosis with legumes. *Curr. Opin. Microbiol.* **72**,
639 102281 (2023).
640 22. Oldroyd, G. E. D., Murray, J. D., Poole, P. S. & Downie, J. A. The Rules of Engagement in
641 the Legume-Rhizobial Symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).
642 23. Poole, P., Ramachandran, V. & Terpolilli, J. Rhizobia: from saprophytes to endosymbionts.
643 *Nat. Rev. Microbiol.* **180**, 1691 (2018).
644 24. Heath, K. D. & Tiffin, P. Context dependence in the coevolution of plant and rhizobial
645 mutualists. *Proc. R. Soc. B Biol. Sci.* **274**, 1905–1912 (2007).
646 25. Burghardt, L. T. *et al.* Select and Resequence reveals relative fitness of bacteria in symbiotic
647 and free-living environments. *Proc. Natl. Acad. Sci.* **115**, 2425–2430 (2018).
648 26. Burghardt, L. T., Epstein, B., Hoge, M., Trujillo, D. I. & Tiffin, P. Host-associated rhizobial
649 fitness: dependence on nitrogen, density, community complexity, and legume genotype.
650 *Appl. Environ. Microbiol.* **88**, e00526-22 (2022).
651 27. Roy, S. *et al.* Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and
652 Symbiotic Nitrogen Fixation. *Plant Cell* **32**, 15–41 (2020).
653 28. Burghardt, L. T. Evolving together, evolving apart: measuring the fitness of rhizobial
654 bacteria in and out of symbiosis with leguminous plants. *New Phytol.* **228**, 28–34 (2020).
655 29. Ratcliff, W. C., Underbakke, K. & Denison, R. F. Measuring the fitness of symbiotic
656 rhizobia. *Symbiosis* **55**, 85–90 (2011).

657 30. Burghardt, L. T., Trujillo, D. I., Epstein, B., Tiffin, P. & Young, N. D. A select and
658 resequence approach reveals strain-specific effects of *Medicago* nodule-specific PLAT-
659 domain genes. *Plant Physiol.* **182**, 463–471 (2020).
660 31. Epstein, B. *et al.* Genome-wide association analyses in the model rhizobium *Ensifer meliloti*.
661 *mSphere* **3**, e00386-18 (2018).
662 32. Wang, C. *et al. NODULES WITH ACTIVATED DEFENSE 1* is required for maintenance of
663 rhizobial endosymbiosis in *Medicago truncatula*. *New Phytol.* **212**, 176–191 (2016).
664 33. Yang, S., Tang, F., Gao, M., Krishnan, H. B. & Zhu, H. *R* gene-controlled host specificity in
665 the legume–rhizobia symbiosis. *Proc. Natl. Acad. Sci.* **107**, 18735–18740 (2010).
666 34. Tang, F., Yang, S., Liu, J. & Zhu, H. *Rj4* , a Gene Controlling Nodulation Specificity in
667 Soybeans, Encodes a Thaumatin-Like Protein But Not the One Previously Reported. *Plant*
668 *Physiol.* **170**, 26–32 (2016).
669 35. Price, P. A. *et al.* Rhizobial peptidase HrrP cleaves host-encoded signaling peptides and
670 mediates symbiotic compatibility. *Proc. Natl. Acad. Sci.* **112**, 15244–15249 (2015).
671 36. Oldroyd, G. E. D. Speak, friend, and enter: signalling systems that promote beneficial
672 symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**, 252–263 (2013).
673 37. Wheatley, R. M. *et al.* Lifestyle adaptations of Rhizobium from rhizosphere to symbiosis.
674 *Proc. Natl. Acad. Sci. U. S. A.* **117**, 23823–23834 (2020).
675 38. Ferguson, B. J. *et al.* Legume nodulation: The host controls the party. *Plant Cell Environ.* **42**,
676 41–51 (2019).

677 39. Bright, L. J., Liang, Y., Mitchell, D. M. & Harris, J. M. The *LATD* Gene of *Medicago*
678 *truncatula* Is Required for Both Nodule and Root Development. *Mol. Plant-Microbe*
679 *Interactions®* **18**, 521–532 (2005).
680 40. Hogg, B. V., Cullimore, J. V., Ranjeva, R. & Bono, J.-J. The *DMI1* and *DMI2* Early
681 Symbiotic Genes of *Medicago truncatula* Are Required for a High-Affinity Nodulation
682 Factor-Binding Site Associated to a Particulate Fraction of Roots. *Plant Physiol.* **140**, 365–
683 373 (2006).
684 41. Messinese, E. *et al.* A Novel Nuclear Protein Interacts with the Symbiotic DMI3 Calcium-
685 and Calmodulin-Dependent Protein Kinase of *Medicago truncatula*. *Mol. Plant-Microbe*
686 *Interactions®* **20**, 912–921 (2007).
687 42. Smit, P. *et al. Medicago* LYK3, an Entry Receptor in Rhizobial Nodulation Factor Signaling.
688 *Plant Physiol.* **145**, 183–191 (2007).
689 43. Ovchinnikova, E. *et al.* IPD3 Controls the Formation of Nitrogen-Fixing Symbiosomes in
690 Pea and *Medicago* Spp. *Mol. Plant-Microbe Interactions®* **24**, 1333–1344 (2011).
691 44. Starker, C. G., Parra-Colmenares, A. L., Smith, L., Mitra, R. M. & Long, S. R. Nitrogen
692 Fixation Mutants of *Medicago truncatula* Fail to Support Plant and Bacterial Symbiotic
693 Gene Expression. *Plant Physiol.* **140**, 671–680 (2006).
694 45. Schnabel, E., Journet, E.-P., de Carvalho-Niebel, F., Duc, G. & Frugoli, J. The *Medicago*
695 *truncatula* SUNN Gene Encodes a CLV1-like Leucine-rich Repeat Receptor Kinase that
696 Regulates Nodule Number and Root Length. *Plant Mol. Biol.* **58**, 809–822 (2005).

697 46. Schnabel, E. L. *et al.* The *ROOT DETERMINED NODULATION1* Gene Regulates Nodule
698 Number in Roots of *Medicago truncatula* and Defines a Highly Conserved, Uncharacterized
699 Plant Gene Family. *Plant Physiol.* **157**, 328–340 (2011).
700 47. Riley, A. B., Grillo, M. A., Epstein, B., Tiffin, P. & Heath, K. D. Discordant population
701 structure among rhizobium divided genomes and their legume hosts. *Mol. Ecol.* **32**, 2646–
702 2659 (2023).
703 48. Denison, R. F. & Kiers, E. T. Life histories of symbiotic rhizobia and mycorrhizal fungi.
704 *Curr. Biol. CB* **21**, R775-85 (2011).
705 49. Darch, S. E., West, S. A., Winzer, K. & Diggle, S. P. Density-dependent fitness benefits in
706 quorum-sensing bacterial populations. *Proc. Natl. Acad. Sci.* **109**, 8259–8263 (2012).
707 50. Fåhraeus, G. The Infection of Clover Root Hairs by Nodule Bacteria Studied by a Simple
708 Glass Slide Technique. *Microbiology* **16**, 374–381 (1957).
709 51. Kessner, D., Turner, T. L. & Novembre, J. Maximum likelihood estimation of frequencies of
710 known haplotypes from pooled sequence data. *Mol. Biol. Evol.* **30**, 1145–1158 (2013).
711 52. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and
712 population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–
713 2993 (2011).
714 53. Purcell, S. *et al.* PLINK: A Tool Set for Whole-Genome Association and Population-Based
715 Linkage Analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
716 54. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**,
717 583–589 (2021).

718 55. Varadi, M. *et al.* AlphaFold Protein Structure Database in 2024: providing structure coverage
719 for over 214 million protein sequences. *Nucleic Acids Res.* **52**, D368–D375 (2024).
720 56. Van Kempen, M. *et al.* Fast and accurate protein structure search with Foldseek. *Nat.*
721 *Biotechnol.* **42**, 243–246 (2024).
722 57. Roux, B. *et al.* An integrated analysis of plant and bacterial gene expression in symbiotic
723 root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**,
724 817–837 (2014).
725 58. Domonkos, Á. *et al.* NAD1 Controls Defense-Like Responses in Medicago truncatula
726 Symbiotic Nitrogen Fixing Nodules Following Rhizobial Colonization in a BacA-
727 Independent Manner. *Genes* **8**, 387 (2017).
728 59. Harris, J. M. & Dickstein, R. Control of root architecture and nodulation by the *LATD/NIP*
729 transporter. *Plant Signal. Behav.* **5**, 1365–1369 (2010).
730 60. Wang, Q. *et al.* Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-
731 Fixing Symbiosis in a Strain-Specific Manner in *Medicago truncatula*. *Mol. Plant-Microbe*
732 *Interactions®* **31**, 240–248 (2018).
733 61. Guefrachi, I. *et al.* Extreme specificity of NCR gene expression in Medicago truncatula.
734 *BMC Genomics* **15**, 712 (2014).
735 62. Horváth, B. *et al.* Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes
736 symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant. *Proc. Natl. Acad. Sci.*
737 *U. S. A.* **112**, 15232–15237 (2015).

738 63. Montiel, J. *et al.* Morphotype of bacteroids in different legumes correlates with the number
739 and type of symbiotic NCR peptides. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 5041–5046 (2017).
740 64. Brewin, N. J. Development of the Legume Root Nodule. *Annu. Rev. Cell Biol.* **7**, 191–226
741 (1991).
742 65. Gopalakrishnan, S. *et al.* Plant growth promoting rhizobia: challenges and opportunities. *3*
743 *Biotech* **5**, 355–377 (2015).
744 66. Haag, A. F. *et al.* Protection of Sinorhizobium against Host Cysteine-Rich Antimicrobial
745 Peptides Is Critical for Symbiosis. *PLoS Biol.* **9**, e1001169 (2011).
746 67. Czernic, P. *et al.* Convergent Evolution of Endosymbiont Differentiation in Dalbergioid and
747 Inverted Repeat-Lacking Clade Legumes Mediated by Nodule-Specific Cysteine-Rich
748 Peptides. *Plant Physiol.* **169**, 1254–1265 (2015).
749 68. Mergaert, P. *et al.* Eukaryotic control on bacterial cell cycle and differentiation in the
750 *Rhizobium* –legume symbiosis. *Proc. Natl. Acad. Sci.* **103**, 5230–5235 (2006).
751 69. Walker, L., Lagunas, B. & Gifford, M. L. Determinants of host range specificity in legume-
752 rhizobia symbiosis. *Front. Microbiol.* **11**, (2020).
753 70. Kim, M. *et al.* An antimicrobial peptide essential for bacterial survival in the nitrogen-fixing
754 symbiosis. *Proc. Natl. Acad. Sci.* **112**, 15238–15243 (2015).
755 71. Nicoud, Q. *et al.* Sinorhizobium meliloti Functions Required for Resistance to Antimicrobial
756 NCR Peptides and Bacteroid Differentiation. *mBio* **12**, e00895-21 (2021).
757 72. Burghardt, L. T. *et al.* Transcriptomic basis of genome by genome variation in a legume-
758 rhizobia mutualism. *Mol. Ecol.* **26**, 6122–6135 (2017).

759 73. Lima, R. M., Kylarová, S., Mergaert, P. & Kondorosi, É. Unexplored Arsenals of Legume
760 Peptides With Potential for Their Applications in Medicine and Agriculture. *Front.*
761 *Microbiol.* **11**, 1307 (2020).
762 74. Savolainen, O., Lascoux, M. & Merilä, J. Ecological genomics of local adaptation. *Nat. Rev.*
763 *Genet.* **14**, 807–820 (2013).
764 75. Hoban, S. *et al.* Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions,
765 and Future Directions. *Am. Nat.* **188**, 379–397 (2016).
766 76. Elena, S. F. Local adaptation of plant viruses: lessons from experimental evolution. *Mol.*
767 *Ecol.* **26**, 1711–1719 (2017).
768 77. Karunakaran, R., East, A. K. & Poole, P. S. Malonate Catabolism Does Not Drive N_2
769 Fixation in Legume Nodules. *Appl. Environ. Microbiol.* **79**, 4496–4498 (2013).
770 78. Stammers, D. K. The structure of the negative transcriptional regulator NmrA reveals a
771 structural superfamily which includes the short-chain dehydrogenase/reductases. *EMBO J.*
772 **20**, 6619–6626 (2001).
773 79. Mus, F. *et al.* Symbiotic Nitrogen Fixation and the Challenges to Its Extension to
774 Nonlegumes. *Appl. Environ. Microbiol.* **82**, 3698–3710 (2016).
775 80. Burghardt, L. T., Epstein, B. & Tiffin, P. Legacy of prior host and soil selection on rhizobial
776 fitness in planta. *Evolution* **73**, 2013–2023 (2019).
777 81. Wippel, K. *et al.* Host preference and invasiveness of commensal bacteria in the Lotus and
778 Arabidopsis root microbiota. *Nat. Microbiol.* **6**, 1150–1162 (2021).

82. Sugawara, M. *et al.* Variation in bradyrhizobial NopP effector determines symbiotic incompatibility with Rj2-soybeans via effector-triggered immunity. *Nat. Commun.* **9**, 3139 (2018).

83. Kaminsky, L. M., Trexler, R. V., Malik, R. J., Hockett, K. L. & Bell, T. H. The inherent conflicts in developing soil microbial inoculants. *Trends Biotechnol.* **37**, 140–151 (2019).

84. Soumare, A. *et al.* Exploiting biological nitrogen fixation: a route towards a sustainable agriculture. *Plants* **9**, 1011 (2020).

85. Trivedi, P., Batista, B. D., Bazany, K. E. & Singh, B. K. Plant–microbiome interactions under a changing world: responses, consequences and perspectives. *New Phytol.* **234**, 1951–1959 (2022).

86. Batstone, R. T., Ibrahim, A. & MacLean, L. T. Microbiomes: Getting to the root of the rhizobial competition problem in agriculture. *Curr. Biol.* **33**, R777–R780 (2023).

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

88. Porter, S. S. & Sachs, J. L. Agriculture and the disruption of plant–microbial symbiosis. *Trends Ecol. Evol.* **35**, 426–439 (2020).

89. 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).

90. 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).

802 Figure Legends

803 **Fig. 1: Schematic overview of the experiment used to evaluate rhizobial fitness and host**
804 **traits across different host mutant backgrounds.** A set of 86 *Sinorhizobium* strains was
805 cultured and combined into a mixed inoculum. This inoculum was applied to *Medicago*
806 *truncatula* wild-type (A17) and host mutants with gene disruptions that act at different stages of
807 symbiosis: Early, Late, and Autoregulation of Nodulation (AON). The number of host mutants
808 per stage is indicated in parentheses. Host and rhizobial traits were measured at the nodule and
809 whole-plant scales. The experimental conditions are clarified under 'Experimental Details'. Each
810 treatment included five pots per host mutant with three plants per pot, grown in randomized
811 blocks under nitrogen (N)-free conditions for 8 weeks. Traits measured include indicators of
812 rhizobial benefit, host benefit, and host investment (Refer to Table 1 for more details). Different
813 coloured nodules indicate nodules formed by different rhizobial strains.
814
815 **Fig. 2: Host symbiotic mutations influence rhizobial fitness.** (A) Host mutations altered
816 nodule numbers in consistent and predictable ways. (B–C) Host genotype also influenced
817 rhizobial fitness, as measured by $\log_{10}$-transformed colony-forming units per nodule and per
818 plant (CFU/nodule) (B) and CFU /plant (C). (D) Nodule strain diversity was largely unaffected
819 by host mutations. Closed triangles indicate a significant increase or decrease relative to wild-
820 type A17 (based on Tukey-adjusted *P* values). Boxplots are color-coded by the symbiotic stage
821 disrupted by each host mutation (see Table S1). (E) Heatmap showing the median relative fitness
822 of each of 86 *S. meliloti* strains across the different host genotypes. Strains enriched relative to
823 their initial inoculum appear in shades of red; depleted strains appear in shades of blue. Host
824 genotypes are hierarchically clustered based on similarity in strain fitness profiles. Strains are
825 ranked by their relative fitness in A17. Mutants in bold differ significantly from A17 (see Table
826 S5 for RDA paired with the PERMANOVA results).
827
828 **Fig. 3: GWAS of strain fitness shifts in host mutants. (**A) The rhizobial strain fitness shifts for
829 each of the 86 strains, calculated by subtracting individual relative strain fitness measured in A17
830 from the relative strain fitness in the host mutants. The heatmap is hierarchically based on the
831 similarity of median 'shift' values across hosts. Shades of orange denote positive fitness shift
832 compared to A17, whereas shades of purple denote negative fitness shift compared to A17. (B)
833 Manhattan plot indicating the SNP (single-nucleotide polymorphism) positions, based on S.
834 *meliloti* USDA1106 association with the relative strain fitness shifts across the 15 host mutants
835 and strain fitness in A17. The dotted line indicates the *P*-value threshold used for identifying
836 candidate genes ($p < 10^{-5}$), and significant associations are color-coded according to host
837 mutation and A17. The GWAS was run separately on each replicon to account for different
838 population structures. The 'Δ' indicates 'strain fitness shift' in that host mutant. For comparison,
839 we also ran the GWAS on A17 relative strain fitness.
840
841 **Fig. 4: Rhizobial candidate genes underlying strain fitness shifts predominantly occur on**
842 **pSymA**. Heatmaps based on the inverse *P*-value ranks (red-yellow) and β values, which denote
843 the effect sizes (red-blue) for the top five variant groups on the pSymA across the 15 host

mutants and A17. Rows represent candidate variant groups, and columns represent host mutants. The host mutants are colour coded based on the stage of disruption they affect. Progressive dark shades (low *P*-value ranks) indicate a stronger association between the variant group and strain fitness shift in that host mutant. Heatmap based on the effect sizes (β values) for the same candidate variant groups was ordered to match the *p*-value ranking. Shades of red indicate host mutants where the corresponding variant has a positive effect on strain fitness shift, and shades of blue indicate negative effects as compared to A17. Colour intensity corresponds to effect magnitude. Alternate allele frequencies (i.e., not present in the reference genome USDA1106) for each candidate are also shown in parentheses next to the gene names. Pervasiveness of a variant group is defined as the 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 top 10 across six or more host genotypes. In contrast, a variant is classified as "limited effect" if it is in the top 10 candidates across five or fewer 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.

872 Tables

873

874 **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.

875

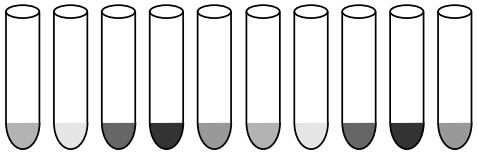
876 *CFU= Colony Forming Units

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.

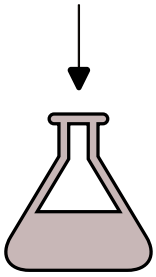
LD group[1]	Host mutants	Location	Alt. allele freq[2]	Eff. alt. vars in mutants	Eff alt. vars in WT	Variants	Genes ID's[3]
9034	*dnf1,dnf5/dnf2, dnf2, dnf4, dnf7, ipd3, nad1,rdn1 sunn1, sunn4*	pSymA	0.07	+	-	1	CDO30_RS23065:SGNH/GDSL hydrolase family protein
6978.alt[4]	*dnf1, dnf1/dnf2, dnf2, dnf4, 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 [1616] , [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 *S.meliloti* isolates and mix for inoculum

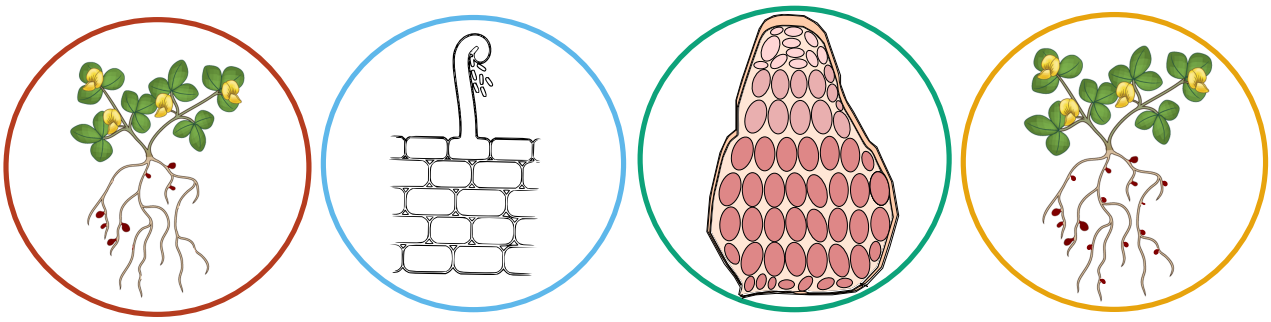


Experiment Details

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

2. Inoculate *Medicago* hosts mutants

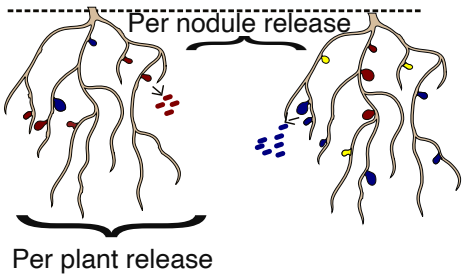
Wildtype | Host mutants spanning stage of symbiosis

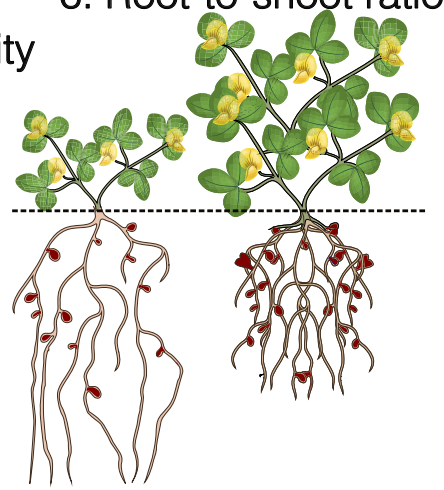


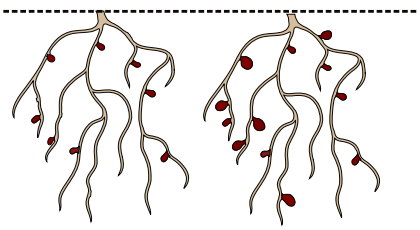
A17 | Early (6) | Late (9) | AON (3)

3. Measure host and rhizobial traits

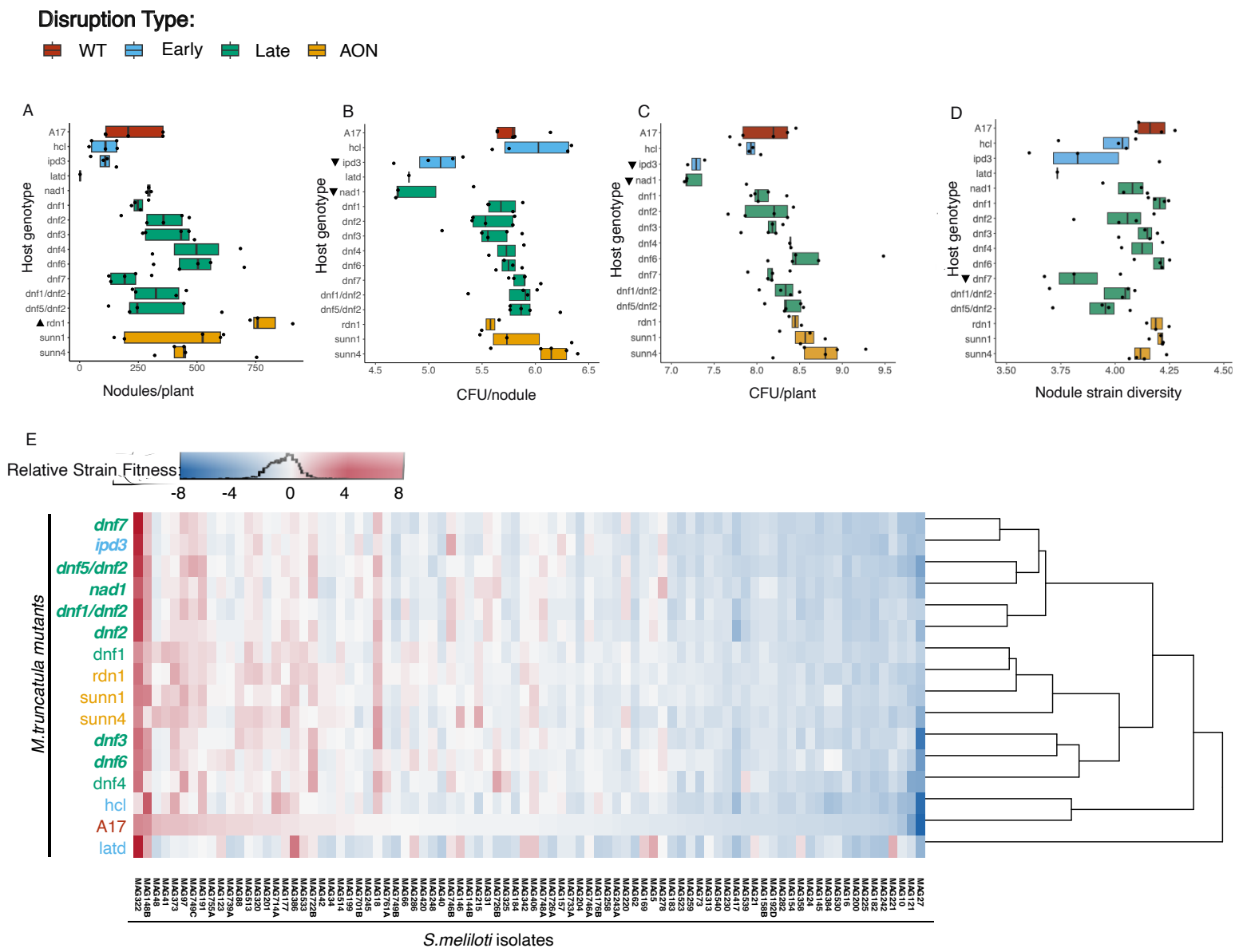
Rhizobial Benefits	Host Benefits	Host Investments
1. CFU/nodule	5. Total biomass/plant	7. Nodules/plant
2. CFU/plant	6. Root-to-shoot ratio	8. Average area/nodule
3. Nodule isolate diversity		9. Total nodulated area
4. Relative strain fitness		

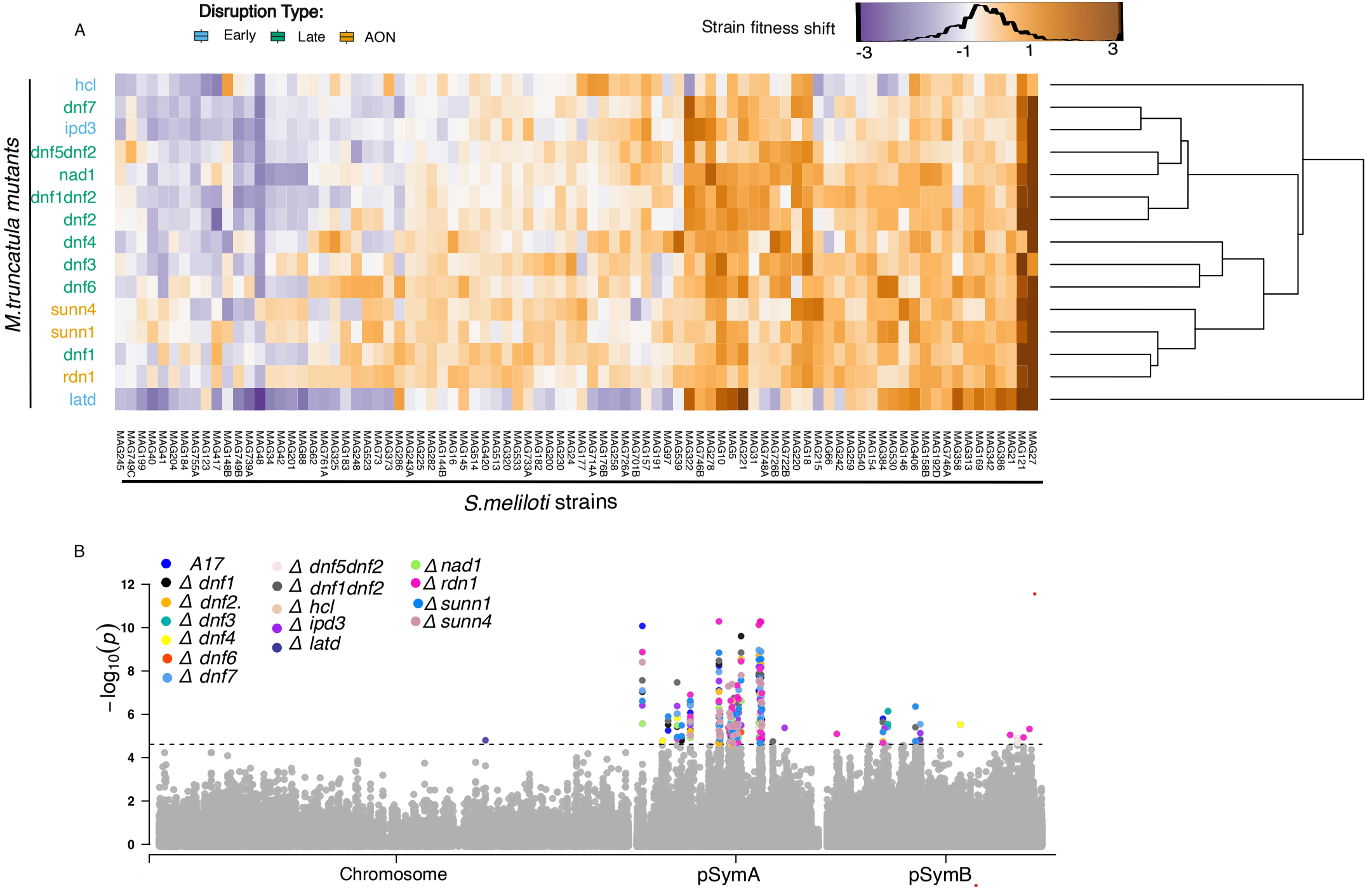


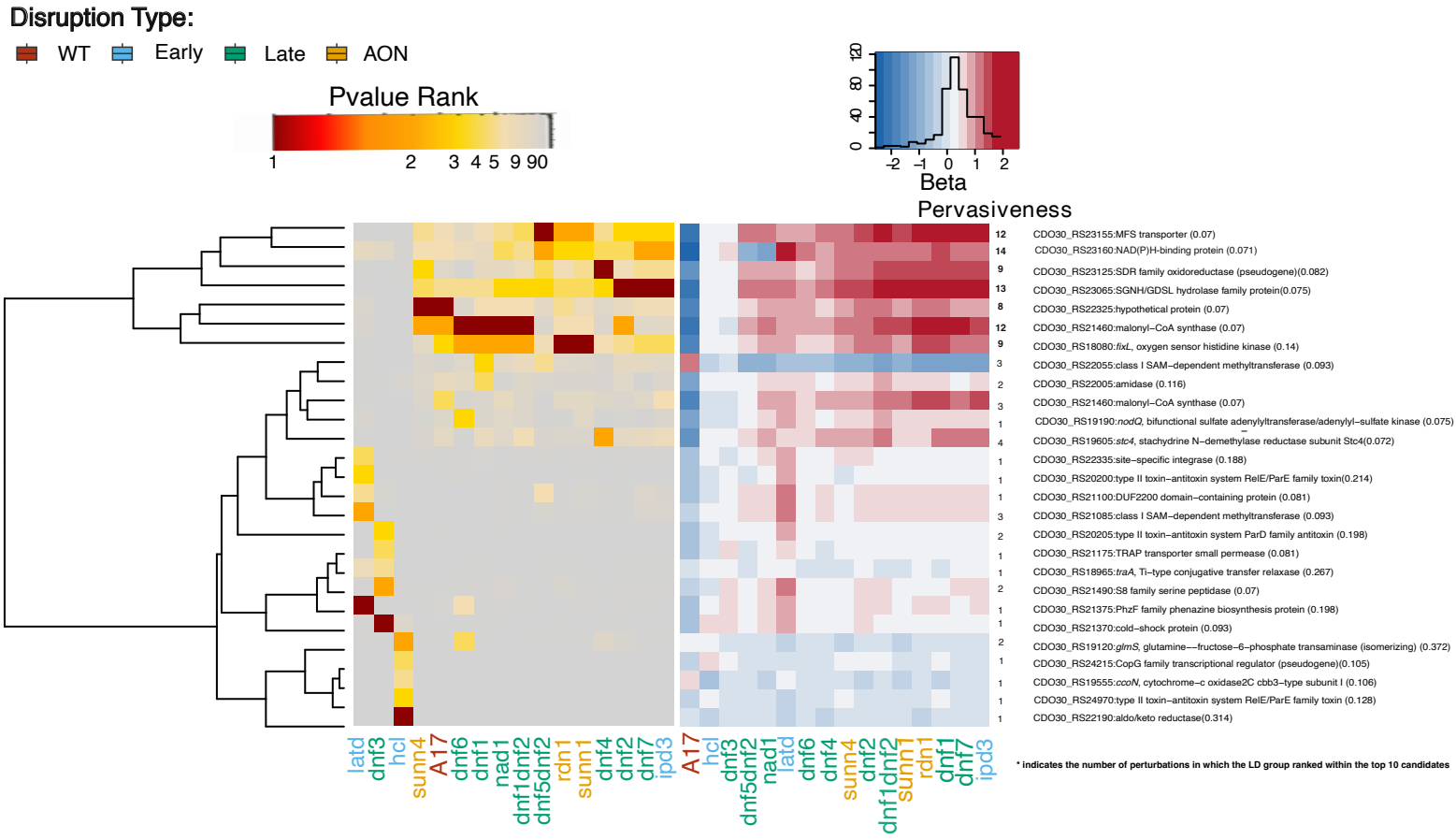


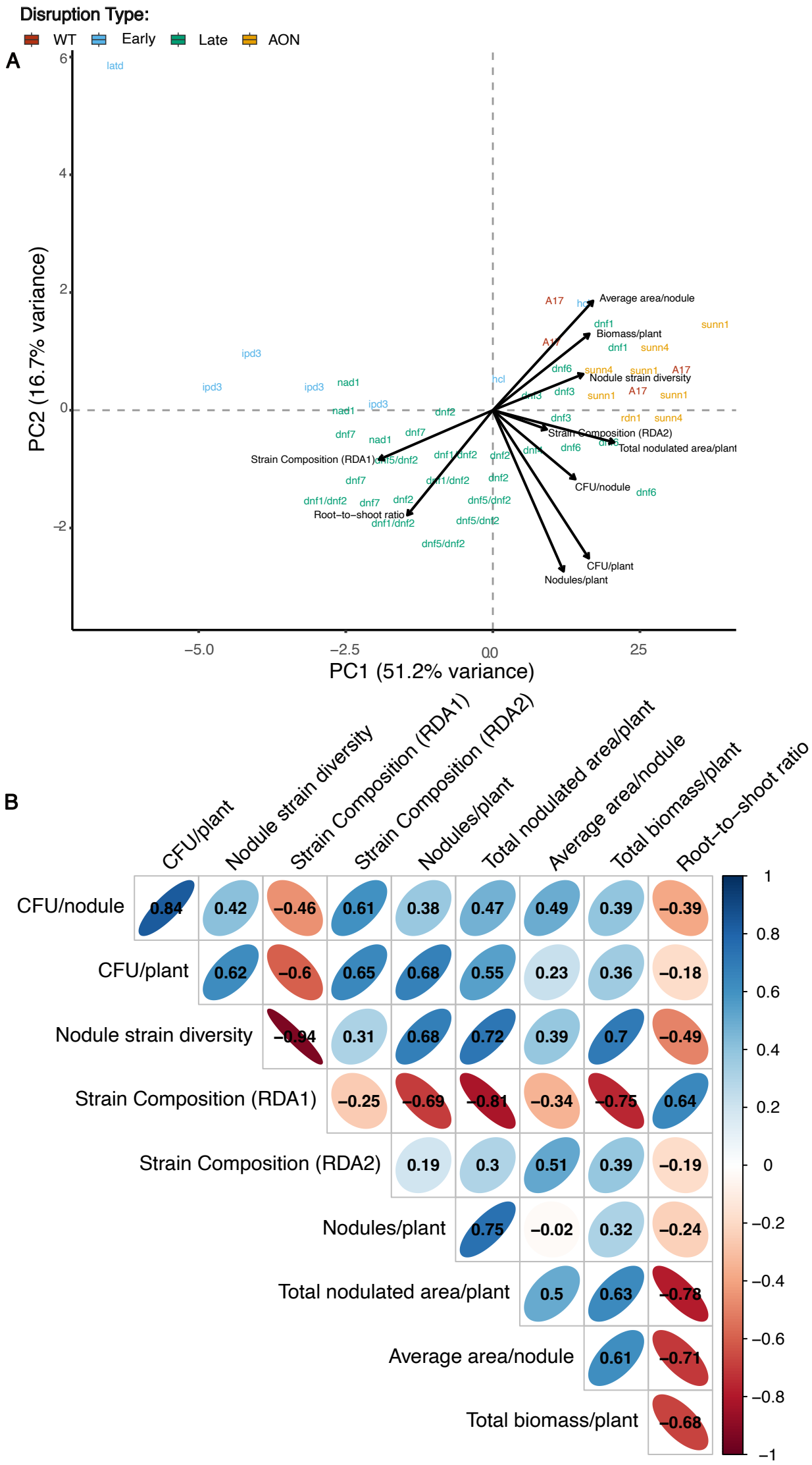


Low → High | Low → High | Low → High


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