Title: In multi-strain inoculations, mutation of *Medicago* symbiosis genes creates a complex selective landscape for naturally-occurring genetic variation in rhizobia

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. Single-strain screens have identified many host genes influencing symbiosis. However, it's unknown whether these genes influence which rhizobial strains colonize and thrive inside nodules during multi-strain inoculations. In this study, we inoculated 18 *Medicago truncatula* symbiotic mutants (including mutations that alter NCR peptide production, plant defence, and nodule number regulation) with a mixture of 86 *Sinorhizobium meliloti* strains. In multi-strain inoculations, most mutations led to reduced host benefits but widely varying effects on host investment and rhizobial benefit (i.e., strain relative fitness), revealing widespread host gene by strain fitness interactions. Genome-wide association studies identify 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 when one host gene was disrupted, we identified ten variant groups with pervasive effects across six or more host mutations. These 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 in mutants tended to adversely affect strain fitness in wild-type hosts. Our results reveal how host symbiosis genes perturb the selective landscape and symbiotic outcomes for rhizobia and set the stage for improving rhizobial inoculants and breeding legume hosts better adapted for multi-strain environments.

44 Introduction

45 The symbiosis between legume plants and rhizobia bacteria involves the formation of specialized
46 plant organs called root nodules, within which rhizobia convert atmospheric nitrogen into plant
47 assimilable forms and, in return, receive photosynthates for growth (1,2). This mutualism allows
48 legumes to grow in nitrogen-poor soil, providing a sustainable alternative to chemical fertilizer
49 application (3–5). The economic importance of this plant-microbe interaction has motivated
50 intensive functional genetics research (6). Screening of mutant populations of the annual diploid
51 legume *Medicago truncatula (M. truncatula)* for mutations in binary symbiotic traits like
52 nodulation (Nod +/-) or nitrogen fixation (Fix+/-) enabled the discovery of dozens of genes
53 underlying root nodule symbiosis (6,7). These genes were identified in controlled environments
54 where hosts were inoculated with a single rhizobia (8,9). However, the effects of at least four
55 host symbiosis genes on functional nitrogen fixation have been shown to depend on the rhizobial
56 strain (10–12), suggesting that single-strain results might not allow the inference of the effect of
57 host genes in natural and agricultural contexts, where a single legume can form nodules with
58 multiple strains or species of rhizobia (13). By examining the effects of over a dozen well-known
59 symbiosis genes on host and rhizobial fitness in mixed inoculations, we sought to identify
60 additional host genes with strain-specific effects and clarify the potential for host genes to shape
61 the rhizobial selective environment.

62 Affordable whole genome shotgun sequencing has recently enabled the high-throughput
63 assessment of rhizobial strain frequencies inside the nodules of *M. truncatula* when multiple
64 strains compete for nodulation (Burghardt et al., 2018; Batstone et al., 2022; Epstein et al.,
65 2022). Here, we take advantage of this method to evaluate the effects of 18 well-studied *M.*
66 *truncatula* mutants (knockouts or knockdowns) on the host symbiotic and the rhizobial traits
67 when *M. truncatula* is grown with a mixture of 86 *Sinorhizobium meliloti* (*S. meliloti*) strains.
68 Besides measuring standard symbiotic phenotypes, we estimated the relative fitness of each of
69 the rhizobial strains in the community within each host mutant. We also performed genome-wide
70 associations (GWA) on rhizobial strain fitness shifts in response to host mutations to identify the
71 rhizobia genes potentially responsible for rhizobia genome by host genome interactions
72 (Burghardt et al., 2018; Burghardt et al., 2020). This novel approach allowed us to examine how
73 mutations in host symbiotic genes may perturb the selection landscape for naturally occurring
74 genetic variation in rhizobia.

75 The formation and development of symbiosis in nodules is complex, with each step regulated by
76 multiple host and rhizobial genes (6,18,19). To encompass these steps, we chose 18 host
77 symbiosis mutants (Table S1). We grouped them into three categories based on their impact on
78 1) bacterial invasion (Early), 2) nodule organogenesis and nitrogen fixation (Late), or 3)
79 regulation of nodule number (Autoregulation of nodule number; AON). The early-stage genes
80 (*hcl, ipd3*, *latd, dmi1-2*, *dmi2-3,* and *dmi3-1*) affect the initial processes such as Nod Factor
81 recognition and nodule organogenesis (20–24). The late-stage genes (*nad1, dnf1, dnf2, dnf4,*
82 *dnf2/dnf1, dnf5/dnf2, dnf6, dnf7*) influence nitrogen fixation (25), while the AON pathway genes
83 (*sunn1, sunn4, rdn1*) regulate the overall number of nodules a plant forms (26,27). While these

84 loose categorizations help set baseline expectations for the impact of the host mutations on the
85 legume and the rhizobia involved, similar phenotypes can be caused by mutations in different
86 genetic pathways, and each gene may have additional pleiotropic effects or act in additional
87 contexts of nodule development when faced with more than one strain. Together, these focal
88 genotypes provide a broad view of the effects of host symbiosis genes on their rhizobial partners
89 in multi-strain inoculations.

90 We measured host benefit and host investment traits commonly used to characterize the
91 functional effect of host symbiosis genes (Table 1). Host benefit proxies include *total biomass*
92 *per plant* and the *root-to-shoot ratio*. Since nitrogen (N) is a growth-limiting resource,
93 comparing mutant dry *total biomass per plant* to wild type in N-limited conditions is a proxy for
94 the total nitrogen fixation in nodules and, thus, evidence of overall host benefit from the
95 symbiosis. Mutations within the genes *ipd3, latd, nad1, dnf1, dnf2, dnf4, dnf2/dnf1, dnf5/dnf2,*
96 *dnf6, dnf7* impair nitrogen fixation and prevent the host from benefiting from the symbiosis, as
97 evident from reduced biomass (20,24,25,28). A less common metric of host benefit from
98 rhizobia is the *root-to-shoot ratio*. When nitrogen is not limiting, plants shift to invest more in
99 above than belowground growth, resulting in lower root-to-shoot ratios (29). Thus, in nitrogen-
100 limited environments, the benefits of nitrogen fixation by rhizobia should be reflected in lower
101 root-to-shoot ratios and greater total biomass per plant. In contrast, the host investment traits
102 *nodules per plant*, *average area per nodule,* and *total nodule area per plant* reflect how many
103 costly organs/relationships a legume forms with the rhizobia and the host's average investment
104 into each nodule or overall investment in nodules, respectively. Disrupting *sunn, sunn4,* or *rdn1*
105 (AON genes) commonly causes shifts in host investment traits (hypernodulation) (26,27). Here,
106 we ask if these patterns recur during mixed-strain inoculations where hosts can choose partners.

107 While researchers often measure the effect of host mutations on host benefits and investment in
108 symbiosis, it is far less common to measure the effect of host mutations on rhizobia (30). The
109 benefits rhizobia obtain from hosts can be assessed by estimating the *population size of rhizobia*
110 released at the individual nodule and whole plant level (31). Host mutations that increase plant
111 defence responses (*nad1* and *dnf2*) will likely reduce the rhizobia released from each nodule,
112 while population responses at the host plant level could be compensated for by forming more
113 nodules (32). Host selectivity is another facet of rhizobial fitness that host genes can affect.
114 Some host genes can increase the number of strains able to form nodules, causing an increase in
115 *nodule strain diversity,* while others may reduce diversity. Lastly, host mutations could affect the
116 *relative strain fitness,* or which strains can best compete and thrive in host nodules. To our
117 knowledge, the only assessment of the effect of host mutations on rhizobial benefits during
118 mixed strain inoculations has been conducted on the *NPD* gene family (Nodule Plat Domain
119 encoding genes) of *M. truncatula* where knocking out all 5 NPD genes (Δ*npd1-5*) eliminated
120 nodulation when inoculated with only *S. meliloti* strain SM1021 but inoculating with a multi-
121 strain community of *S. meliloti* , partially restored functional nodulation despite nodules
122 supporting fewer rhizobial strains than WT (17). These results suggest the standard approach of
123 pairing single hosts to single strains and focusing only on plant traits may miss host by strain
124 interactions.

125 Here, we expanded the range of mutations from the *npds* to additional symbiotic genes with
126 wide-ranging phenotypic effects (Early, Late, and AON) and studied how they influence both

host and rhizobial fitness metrics when inoculated with a mixture of 86 *S. meliloti* strains (Figure 1). In addition to measuring univariate host and rhizobial traits, we used whole genome sequencing and the select and resequence (S&R) technique to measure the changes in relative rhizobial strain fitness diversity (Burghardt et al. 2018). Finally, using GWA, we identified natural genetic variation in rhizobial candidates for host genome x strain genome interaction. We examined the following key questions: 1) How much of the variation in host and rhizobial traits can be explained by symbiotic mutation categories (Early, Late, AON)? 2) Which host and rhizobial symbiotic traits are most sensitive to symbiotic gene mutation in multi-strain inoculations? 3) Do host symbiosis genes have strain-specific effects on rhizobial fitness (e.g., genome x genome interactions)? And 4) How does mutating these host symbiosis genes perturb the selective landscape for naturally-varying rhizobial genes? Overall, in a multi-strain inoculation, host symbiosis genes have varied effects on rhizobia and host benefits. Further, while many host mutations alter selection on rhizobia, the genomic basis of rhizobial fitness exhibits more overlap on the replicon with the symbiosis genes than the other two replicons.

Results

Symbiotic mutation type accounted for only a subset of the trait variation explained by disrupting specific symbiotic genes. To assess the contributions of symbiotic mutation type (Early, Late, AON) and the eighteen individual genetic mutations to each trait, we ran two linear models: one with mutation type as a predictor and the other with genotype. Models with genotype-specific predictors consistently had higher adjusted R-squared values (R_{adj} ranged from 0.22 to 0.88) than those with only symbiotic mutation types (R_{adj} ranged from 0.08 to 0.63) (Table 2). Although mutation type explained only a fraction of the phenotypic variance (Table S2), they explained a significant fraction of variance for all traits except *CFU per nodule* (Table S2). Notably, 61% of the variation in *total nodule area per plant* was explained by mutation type, with AON pathway mutations promoting more investment in nodules (Table 2, Table S3). Mutation type also explained 39% of the variation in the *root-to-shoot ratio* (Table 2), with early and late-stage disruptors increasing this ratio, indicating host nitrogen limitation (Table S3). Additionally, they accounted for 40% of the variation in *CFU per plant* (Table 2), with early-stage mutations reducing and AON mutations promoting the trait (Table S3). While some variation can be explained at the level of mutation types, the increased explanatory power provided by including specific genotypes suggests that moving beyond simple symbiotic mutation types is necessary to explain symbiotic trait variation in mixed inoculations. Thus, next, we focus on examining specific host mutations.

Some symbiotic traits are more sensitive to genetic mutations in mixed inoculations than others. To provide a more detailed assessment of the effect of each of the mutated genes, we studied the effects of each mutation on each of the nine focal traits (Fig 1, Fig S4). Note that host investment and rhizobial benefits could not be assessed for *dmi1-2*, *dmi2-3, and dmi3-1* mutants because they were non-nodulating even during multi-strain inoculation. Overall, we found that host benefit traits (*total biomass per plant* and *root-to-shoot ratio*) tended to be more sensitive to host symbiotic gene mutations than host investment or rhizobial benefit traits (Fig S4). Similarly, some genotypic mutations (e.g., *ipd3, nad1,* and *sunn4)* significantly influenced almost all host

169 and rhizobial traits, while others (e.g., *dnf1*) did not influence any of the traits (Fig S4). Below,
170 we highlight each trait category with a special focus on rhizobial benefit traits.

171 Host mutations differentially altered host benefit and investment traits. Perhaps unsurprisingly,
172 most genetic mutations influenced both biomass (Fig 2A, Fig S4) and the root-to-shoot ratio (Fig
173 2B, Fig S4). All mutations reduced biomass except *dnf1*, *rdn1,* and *sunn1* (Fig 2A). Similarly,
174 the *root-to-shoot ratio* increased relative to the wild type in 11 symbiotic mutants, indicating
175 nitrogen limitation (Fig 2B, Fig S4). In contrast, *sunn4* plants had a low *root-to-shoot ratio,*
176 presumably due to the very low biomass of these plants. Overall, our data suggest that most
177 symbiotic mutations identified in single-strain contexts are not strongly strain-dependent
178 regarding host benefit during multi-strain inoculations, with early- and late-stage mutants
179 reducing host benefits, but the degree of impairment varied among genes. On the other hand,
180 host investment traits were differentially affected by the mutants. Two out of the three host
181 investment traits (*total nodule area per plant* and *nodules per plant)* were higher than wild type
182 in AON mutants, *dnf4* and *dnf6*, but lower than wild type in six early and late stage mutants (Fig
183 2C, D; Fig S4). A single mutant (*hcl*) increased *average area per nodule* (Fig 2E, Fig S4).
184 Importantly, if we had only focused on the commonly used metric of *nodules per plant*, we
185 would have missed the effect of many early- and late-stage mutants on overall investment during
186 mixed inoculations. These observations highlight the complexity of symbiotic interactions and
187 underscore the importance of considering multiple host traits to understand the impacts of host
188 genetic mutations on plant performance.
189
190 The effect of genetic mutations on rhizobial benefits also depended on the trait. Three mutations
191 (*ipd3, latd,* and *nad1)* reduced both *CFU per plant* and *per nodule* (Fig 3A, B; Fig S4). In
192 contrast, the *CFU per plant* increased in four hosts (*dnf4*, *dnf6, sunn1,* and *sunn4*). Interestingly,
193 none of the mutations increased *CFU per nodule* (Table S4, Fig 3B, Fig S4), but rather, the
194 higher *CFU per plant* was due to forming more nodules. A quarter of genotypic mutations (*ipd3,*
195 *latd,* and *dnf7)* reduced the *nodule strain diversity*, while none increased diversity (Fig 3C, Fig
196 S4, Table S4). The reduced diversity of strains within *latd* mutants comes with the caveat that
197 *latd* formed very few nodules, so the observed reduced diversity could be due to a reduction in
198 the number of nodules rather than a shift in selectivity *per se*. These results suggest that
199 increased host investment positively influenced rhizobial benefits and that functioning versions
200 of the selected symbiosis genes tend to expand the range of strains with which hosts form
201 associations.
202
203 ***Mutations in early and late symbiosis genes, but not AON genes, lead to significant but subtle***
204 ***changes in strain relative fitness***. Interestingly, the eight host mutants affecting strain relative
205 fitness were all early- and late-stage mutants; none of the three AON genes (*sunn1, sunn4, rdn1*)
206 mutants significantly affected strain fitness (Fig S4, Fig S5, Table S5, Table S6). Hierarchical
207 clustering of the strain relative fitness revealed general trends across host genotypes (Fig 4). WT
208 grouped most closely with *hcl*, while the other early mutations showed varied effects: *latd* was
209 the most divergent, and *ipd3* clustered with late-stage disruptors *dnf7*, *dnf2*, *dnf1/dnf2*, *dnf5/dnf2,*
210 and *nad1*. Notably, *dnf3*, *dnf4*, and *dnf6* formed a distinct group of late mutants with the AON
211 mutants and *dnf1*. Mutants with similar effects on strain fitness also tended to have comparable
212 above-ground biomass, indicating that host resource allocation to nodules influences strain
213 fitness differences. While a few strains showed consistently high or low fitness across all

214 genotypes (e.g., MAG322 vs. MAG27), fitness variability was common across the host mutants.
215 Hierarchical clustering revealed ~71% of strains to be sensitive to host genotypic mutations.
216 Some strains strongly responded to specific genetic mutations, for example- MAG322,
217 MAG386, MAG5, and MAG221 becoming more common with *latd* mutation, and MAG148B
218 with *hcl* mutation. While these shifts are interesting, the overall pattern is of host mutations
219 causing slight fitness shifts across many strains. These subtle shifts give us the power to use
220 GWA to identify rhizobial allelic variants associated with shifts in response to host mutations.
221
222 ***GWA reveals variants on pSymA and pSymB, which are important for rhizobial fitness when***
223 ***symbiosis genes are disrupted***. We performed sixteen GWA, one for each of the shifts in strain
224 fitness caused by the fifteen mutants (WT minus mutant fitness) and one for WT fitness alone
225 (Fig 5A), to identify rhizobial variants associated with variation in fitness. The strongest
226 candidates for genome x genome associations based on *P* values were found on pSymA and
227 pSymB (Fig 5A). We found nine rhizobial linkage groups on pSymA and 17 linkage groups on
228 pSymB that were strongly associated with strain fitness shifts ($P < 10^{-5}$) (Fig 5B, Table 3). In
229 contrast, linkage groups on the chromosome had weaker effects (no P-values less than 10^{-5}),
230 except for LD group 838, which was associated with increased strain fitness within *latd*—a host
231 background with a limited sample size. (Fig 5B*)*. Some of these differences in power between
232 replicons are likely due to the divergent population structure, but the direction of impact is
233 difficult to predict (e.g., the chromosome has long-distance linkage among variants, creating
234 large linkage groups tagging many genes).
235
236 ***Mutating host symbiosis genes has widespread effects on the selective landscape for genetic***
237 ***variation in rhizobia.*** We were interested to discover if the genetic variants underlying rhizobial
238 fitness, when the symbiosis pathway was perturbed, had pervasive or limited effects. In other
239 words, we sought to summarize the genomic selective landscape that host gene mutation creates
240 for rhizobia, irrespective of arbitrary *P*-value cut-offs. For each replicon, we identified the five
241 most strongly associated candidates and then compared the P ranks (Fig 5B and Fig S6) and
242 effects sizes (Fig S7) of those variant groups across all the host mutations (Table S8 for a
243 complete list). This resulted in a list of 114 LD groups across the three replicons. To visualize
244 patterns, we clustered LD groups based on the inverse of P-value ranks (Fig 5B, Fig S6) and on
245 the effect sizes (beta values) (Fig S7). A vast majority of candidates on the chromosome and
246 pSymB affected strain fitness in fewer than three host mutants (Table S8, Fig S8). Exceptions on
247 the chromosome included a massive linkage group (901) with 144 variants in 65 genes. This LD
248 group was in the top ten rhizobial candidates when nine different host genes were mutated,
249 spanning all disruption types. On pSymB, four small LD groups were top-ranked in at least four
250 host mutants (ABC transporter and RKpR, a dihydrodipicolinate synthase family protein, pepT,
251 peptidase T, and a hypothetical protein) (Fig 5B). However, the strongest signals of shared
252 genomic selection were on pSymA (Fig S8), where we identified nine linkage groups with
253 pervasive effects on strain fitness shifts (Fig S6). The effect of these genetic variants on strain
254 fitness was not confined to a specific symbiotic stage, although the pervasive variants were not
255 important in two out of the three early-stage mutants (*hcl, latd*) (Fig S6). The non-reference
256 variant increased strain fitness within the host mutants, and interestingly, only in pervasive genes
257 did we see evidence of a fitness trade-off of the allele within WT (Fig S7). The pervasive
258 candidates included genes like malonyl COA synthase, SGNH/GDSL hydrolase family protein,

259 *fixL* which was an oxygen sensor histidine kinase, a short chain oxidoreductase, and a transport
260 protein belonging to the MFS (Major Facilitator Superfamily). Our results indicate that natural
261 genetic variation in metabolic genes plays a vital role in altering strain competitiveness within
262 nodules when symbiotic host genes are disrupted—a role challenging to detect using standard
263 reverse genetics approaches.
264
265 ***Most candidate variants were only important in one or two host symbiotic mutations,***
266 ***suggesting that each mutation creates unique environments for rhizobia***. Most of the limited-
267 effect LD groups that were small (less than three genes) and showed substantial statistical
268 support ($p<10^{-5}$) were on pSymB, and many were found with *nad1*, *dnf3*, *dnf4*, *rdn1*, *dnf5/dnf2*,
269 or *sunn1* hosts (Table 3). These genes included a couple of ABC transporters, each associated
270 with individual mutants (*sunn1, ipd3, dnf1/dnf2, rdn1*), glutamine synthetase (*latd*), genes
271 involved in purine metabolism (*dnf3*), methyltransferases (*nad1*), transcriptional regulators
272 (*rdn1*), and a couple of hypothetical proteins. It was interesting to note that for *latd,* which is
273 mutated in a nitrate transporter, the gene that is most associated with the observed strain fitness
274 shift is glutamine synthetase (33). Interestingly, when *dnf3* was disrupted, a LD group containing
275 45 SNPs in two genes (xanthine dehydrogenase family protein molybdopterin-binding subunit
276 and xanthine dehydrogenase family protein subunit M), increased strain fitness. Most of these
277 rhizobial genes did not significantly alter strain fitness in the WT host background, suggesting
278 this variation may be conditionally neutral. Overall, our results highlight the complexity of host
279 genome by strain genome interactions underlying rhizobial selection.
280
281 Discussion
282
283 This study enhances our understanding of how genetic changes in legume hosts affect the
284 environmental context and fitness of rhizobia within root nodules and the strain-specific nature
285 of these responses. While single-strain inoculations allow the direct measurement of phenotypic
286 outcomes, they limit understanding of relative strain responses and host gene effects. Using
287 multi-strain inoculations, we screened legume symbiotic mutants to find strain-specific effects of
288 symbiosis genes on host and rhizobial traits — thus exploring potential mechanisms that could
289 affect the population size and diversity of mutualistic partners (15,17,34,35). This study
290 expanded on our previous work, revealing the effects of the *NPD* gene family on rhizobial strain
291 fitness during multi-strain inoculation (17). We showed that host genetic mutations persistently
292 affect host benefits, with variable effects on host investment and rhizobial benefits. Finally, by
293 implementing a GWA, we map the selective landscape host gene mutations create for rhizobia
294 and identify candidate genes underlying the strain fitness shifts in response to each host
295 mutation. We find rhizobial genetic variants with both pervasive and limited effects across host
296 mutants and that symbiotic mutation categories provide little clarity about how rhizobial genes
297 influence strain fitness. Taken together, we present an overview of the selective landscape that
298 mutating symbiotic host genes creates for rhizobia, which can be leveraged to fine-tune this
299 mutualism.
300
301 As expected, most but not all of our symbiotic mutations showed reduced nitrogen fixation
302 and/or nodule investment during multi-strain inoculations similar to those observed in prior
303 single-strain studies. Because we conducted our experiments with N-free fertilizer, we inferred

304 nitrogen fixation efficacy in a multi-strain context by placing observed mutant biomass and root-
305 to-shoot ratios on a continuum from no nitrogen fixation (*dmi1-2*, *dmi2-3, and dmi3-1*) to wild-
306 type levels (*hcl*). Exceptions included *dnf1* and *dnf4,* which exhibited nitrogen fixation during
307 multi-strain inoculations despite failing to do so in single-strain experiments suggesting that they
308 can successfully form symbiosis with some strains (36,37). Partial overriding of single strain
309 phenotypes, in terms of host investment and host benefits, have been previously reported in
310 Δ*npd1-5* mutants, which are non-fixing with *Sinorhizobium* strain Sm1021, but regain 30%
311 nitrogen fixation during multi-strain inoculation, highlighting strain-specificity (17). Another
312 host mutant that did not exhibit the expected phenotype during multi-strain inoculations was the
313 *hcl* mutant, which, during single-strain inoculation with the strain Sm2011, formed significantly
314 fewer nodules, but during community inoculation, accumulated biomass similar to the WT with
315 comparable nodulation (23). Host mutation-dependent shifts in rhizobial entry mode have been
316 previously noted with gain of function *snf1* mutants *of Lotus japonicus* (encoding a T265I
317 calcium calmodulin-dependent kinase CCaMK *)*, which cause a shift from complex infection
318 thread mode to a simpler 'crack entry' mode of rhizobial entry (38). It would be interesting to
319 probe how rhizobia navigate the *hcl* mutation—indeed, our GWA reveals that all top variants on
320 all three replicons are unique to the *hcl* mutation. The host mutations that deviate from their
321 single-strain phenotypes during multi-strain inoculations create additional opportunities for
322 discovering novel rhizobia genes that ensure successful symbiosis in these hosts, which are
323 known to be poor or non-fixing with model *Sinorhizobium* strains.

324 Legumes releasing rhizobia from their senescing nodules into the soil, directly contribute to the
325 soil rhizobial population (39). Identifying host genes, like the *npd* family that amplify this effect
326 can lead to targeted soil microbiome interventions (17). A legume can increase the overall
327 amount of rhizobia released via two routes. First, they can release more rhizobia by increasing
328 nodule numbers—a pattern identified for two AON genes (*sunn1*, *sunn4)* and two late-stage
329 genes *(dnf4* and *dnf6*)—suggesting that these genes, when functional, reduce rhizobia released by
330 limiting nodule formation. This hypothesis seems plausible particularly for *sunn1(***Su**per
331 **N**umeric **N**odules1) and *sunn4 (***Su**per **N**umeric **N**odules4*)* which are known to control the
332 degree of nodulation, preventing the legume from overinvesting when nitrogen needs are met
333 (26). Alternatively, plants can have comparable nodulation but release more rhizobia per nodule,
334 as noted previously in *npd2,* which produced fewer nodules than WT during multi-strain
335 inoculations but released more rhizobia per nodule than WT (17). In contrast, here we found
336 most mutations had no effect on the number of rhizobia released from each nodule, and *ipd3,*
337 *latd* and *nad1* actually reduced the population size released. Thus, functional copies of these
338 genes support larger populations of rhizobia per nodule—indeed, disrupting *nad1* leads to
339 overactive plant defence responses in nodules, likely reducing rhizobial survival (32). Overall,
340 our results suggest that the effects of host genetic mutations on rhizobial fitness are gene-specific
341 and may differ from wild-type expectations (e.g., bigger nodules release more rhizobia).
342 Increasing the number of rhizobia the host supports represents one pathway to nurture robust
343 populations of beneficial microbes, and identifying the host genes that cause this amplification
344 sets the stage for targeted interventions aimed at bolstering soil rhizobial populations in
345 agricultural and conservation contexts (40)

346 Similar to our previous work on the *npd* genes (17), we also found subtle but significant relative
347 fitness shifts in specific strains within host mutants. For instance, while late-acting NCRs are

348 often cited as major determinants of host strain specificity in *M. truncatula*, our investigation
349 reveals a more complex landscape (6,42). Two NCR mutations had differing effects on strain
350 fitness within nodules. The *dnf7* mutant, lacking functional NCR169 expressed in the N-fixation
351 zone (43), grouped with the *ipd3* mutant. Surprisingly, *dnf4*, which does not produce NCR211
352 and influences rhizobial differentiation during bacteroid formation (37), showed similar effects
353 on strain fitness as *dnf3* and *dnf6*. There are a few ways, by which rhizobia are known to
354 overcome the inhibitory actions of the NCRs however none of these genes came up as top
355 candidates in our GWA suggesting additional mechanisms of host restriction of rhizobia. These
356 initial observations of NCR mutations further motivate exploring the functional roles and
357 mechanisms of the remaining ~600 NCRs expressed in nodules (44,45), especially since many
358 NCRs are differentially expressed in hosts depending on which strain inhabits the nodules and
359 thus could shape the rhizobial selective environment (46).

360 More broadly, though, our study shows that many other symbiotic genes beyond NCR's,
361 including AON genes, can significantly alter which rhizobial strain each plant enriches in mixed
362 inoculations. Mutations in host genes influencing different stages of symbiosis enriched similar
363 strains. This highlights the complex interplay between host genetics and microbial partners.
364 Activation of plant defence responses in nodules can also lead to differential strain fitness (32).
365 Given that *dnf2* and *nad1* both display elevated defence responses in nodules, all three genotypes
366 with mutations in *dnf2* and *nad1* enriched similar strains (9). The strain fitness pattern in *dnf1*
367 hosts was most similar to AON mutants, *dnf3, dnf4*, and *dnf6*. Although there are no known
368 overlapping roles for these mutants, their similarities may arise from functional nitrogen fixation,
369 allowing the host to reward some strains over others. Lastly, the genetic basis of some symbiosis
370 genes (e.g., *dnf3* and *dnf6*) remains a mystery-examination of candidate genes from our GWA
371 could provide insights(6).

372 Natural variation in symbiosis genes arises and segregates within legume populations, potentially
373 contributing to variation in selection on rhizobia (47,48). To measure the effect of this host
374 genetic variation, we explored how mutations in these genes affect rhizobial strain survival and
375 fitness, specifically asking if the same rhizobial genes were important for rhizobial fitness across
376 different host mutations or if unique host mutations create distinct selective pressures on
377 rhizobia. Our GWAS analyses identified rhizobial genes underlying fitness shifts in different
378 host mutations, revealing genes with both pervasive and limited effects. Variant groups with
379 widespread effects were primarily located on pSymA and pSymB. Interestingly, alternative
380 versions of these alleles increased fitness across many host mutations but strongly reduced
381 fitness in the wild-type. This suggests a potential cost to harbouring these variants, raising the
382 question of why they persist. These variations may be beneficial in other host accessions or
383 species (16). We also found many variants important in only one or a few host mutant
384 backgrounds, with the strongest associations on pSymB and, to some extent, pSymA. Unlike the
385 pervasive variants, most limited effect variants had little or no effect on strain fitness in WT
386 backgrounds, a pattern consistent with conditional neutrality (49–51). Together the data suggests
387 that the alternate versions of the pervasive genes might be maintained through balancing
388 selection due to host variation, while the unique genes may increase in frequency through neutral
389 processes and incidentally enhance fitness when specific symbiotic genes are disrupted.

390 Most candidate genes with pervasive effects are highly expressed in the nitrogen fixation zone
391 (Zone III), emphasizing the need for downstream validation assays to elucidate their role in
392 nitrogen fixation (52). For example, the *fixL* gene, an oxygen sensor, regulates oxygen levels
393 during nitrogen fixation, Malonyl-CoA synthase, induced selectively in the rhizospheres of
394 alfalfa, is responsible for malonate catabolism within root nodules (53), and a NAD(P)H-binding
395 protein from the conserved NmrA domain family shows high similarity to a transcriptional
396 regulator, involved in nitrogen metabolite repression in fungi (54). These genes represent low-
397 hanging fruit for uncovering novel mechanisms to broaden the host range of symbionts, aiming
398 to extend nitrogen fixation beyond legumes (55). Beyond these candidates, most genes only
399 altered strain fitness in specific symbiotic perturbations, indicating that each perturbation creates
400 a unique challenge for rhizobia survival and reproduction. Further experimentation will uncover
401 these novel mechanisms of strain survival and elucidate host by rhizobia genome by genome
402 interactions. Notable candidates include: A Class I SAM-dependent methyltransferase associated
403 with strain fitness in *nad1* mutants, suggesting the importance of epigenetic modifications for
404 rhizobial survival under host defences (32) and variants in *stc4*, encoding stachydrine N-
405 demethylase reductase, enhancing strain fitness in *dnf4* mutants, possibly by protecting
406 differentiating rhizobia (37). Our results highlight the potential of leveraging natural genetic
407 variation to explore undiscovered mechanisms in genome-by-genome interactions between hosts
408 and microbes.

409 *Conclusion:*
410 By identifying how genetic mutations affect symbiosis in multi-strain inoculations, we aimed to
411 understand the forces shaping rhizobial and legume evolution in ecologically realistic contexts
412 (13,30). Measuring the effects of host genetic variation on microbial fitness is challenging but
413 crucial for understanding the evolution of and maximizing the benefits of root nodule symbiosis.
414 By re-examining root nodule symbiosis in multi-strain contexts and using a GWA approach for
415 leveraging natural genetic variation, we identified novel rhizobial genetic factors influencing
416 strain fitness shifts, laying the groundwork for further investigations of legume genome by
417 rhizobia genome interactions (Burghardt et al., 2018; Burghardt et al., 2020). Validating these
418 rhizobia candidates in different contexts will help fine-tune existing mutualism and potentially
419 engineer new associations between rhizobia and novel hosts.
420
421 Materials and methods
422
423 ***Greenhouse experiment to measure multi-strain symbiotic phenotypes***
424 We created our mixed inoculum using 86 sequenced strains of *S. meliloti* originally isolated from
425 soils from across the native range of *M. truncatula* (57). We grew each strain in liquid tryptone
426 yeast media individually for three days to reach similar population sizes (FigS1) and then mixed
427 them in equal proportions. This community was used for three purposes: a) To inoculate host
428 plants, b) To measure rhizobial density in the initial community via dilution plating, and c) To
429 measure the strain frequencies in the initial inoculum via whole genome sequencing. Seeds from
430 the *M. truncatula* wild type (A17) and all 18 host mutants (Table S1) and were surface-sterilized,
431 scarified, cold-stratified, and sowed into pots as described in (14) *(see Supplemental Methods for*
432 *details)*. Plants were inoculated with 10^6 rhizobia per pot, three days after germination, fertilized
433 with increasing amounts of nitrogen-free fertilizer, and harvested 8 weeks post-inoculation to
434 measure the host and rhizobial traits (Table 1).

To assess plant investment, nodules collected from the roots were photographed on a green background, and nodule number and size were measured using a customized pipeline in ImageJ (imagej.net/ij/). To measure plant benefits, the roots and shoots were harvested separately, oven-dried, and weighed. To measure rhizobial benefits, we dissected and pooled all the nodules from the root systems in each pot. We surface sterilized and homogenized nodule pools and used dilution planting and colony counting to estimate the rhizobia released from nodules. To determine the frequency of strains in nodule pools, we enriched homogenized nodules for undifferentiated rhizobia using differential centrifugation and extracted DNA using a Qiagen Plant DNeasy kit. Illumina library preparations and whole genome, paired-end, short read sequencing on an Illumina NextSeq were conducted at the Huck Genome Core facility at Pennsylvania State University. We mapped reads to reference genome USDA1106 (58) according to (14) and used HARP (59) to estimate strain frequencies in nodules and the initial inoculum *(see Supplemental Methods for details)*.

Calculation of strain diversity in nodules and strain relative fitness

Strain frequencies were used to 1) estimate the *strain diversity* of the nodule pool as the exponent of Shannon diversity, using the "renyi" function in the vegan package of R. This data was used to interpret if the mutation type or the host genotype significantly altered the ability of the host to select for a wide variety of strains, indicated by a high Shannon diversity metric and 2) calculate *strain relative fitness,* which normalizes the data distribution and controls for small differences in initial strain frequency. For this, for each strain, we divided the final strain frequency by the initial frequency in the inoculum to get relative strain fitness (median initial frequency = 0.0092, 5% to 95% quantile: 0.006-0.018), then we calculated the $\log_2$ of the relative fitness to normalize the distribution. To allow for statistical analysis, strains with frequency estimates of 0 in the selected communities with negative infinity values after this transformation were assigned a relative fitness value of −8, a value less than the lowest measurable fitness in other strains (Burghardt et al., 2018).

Statistical Analyses

For each trait, we used a permanova ("ANOVA" function) to test for the effect of the symbiotic mutation types (trait ~ type + block) and the host genotypes (trait ~ genotype + block) on the traits. We compared the variance explained by mutation type and host genotype across models (Table S2). To determine which types or genotypes significantly affected the community compared to WT, we re-ran the models on WT and each type (or genotype) individually (Table S3, S4). We compared the fraction of variation explained by mutation type to the fraction explained by the genotype to determine if adding the genotype information improved model performance (Table 2). T-tests of individual model coefficients were used to compare each genotype to WT. To determine how the eight traits were related to each other across the genotypes, we used the "corrplot" function in R (Fig S2). To assess shifts in community composition across genotypes, we performed multivariate redundancy analysis ("RDA" function in the vegan package) (Table S7, Fig S5).

480 ***Genome-wide association study***
481 We conducted a genome-wide association study (Fig S3) to identify rhizobial variants potentially
482 responsible for host mutation-dependent strain fitness shifts. To obtain the fitness shifts, we
483 subtracted the mean strain fitness observed in nodule pools for each host mutant from strain
484 fitness measured in wild-type nodules. To perform the GWA, we used a linear mixed model
485 (LMM) implemented with the -lmm option in GEMMA (v0.98.3) (60). We used a published
486 SNP database comprising 30,076 SNPs representing previously defined linkage-disequilibrium
487 (LD) groups within *S. meliloti* for running the GWA (16). Initial filtering was done using
488 bcftools (61), followed by filtering based on a minor allele frequency of 0.05 and a missingness
489 threshold of 0.2 using PLINK (62). We estimated the proportion of phenotypic variance
490 explained (PVE) by genetic variants using the K-matrix - i.e., chip heritability. *P*-values for each
491 SNP were calculated with a likelihood ratio test. SNPs were assigned to gene annotations using a
492 custom pipeline in Python (16). We used a *P*-value cutoff of 10^{-5} to select the top five candidate
493 genes responsible for the strain fitness shifts for each host mutation. To compare their relative
494 importance, we queried the *P*-value ranks of these focal variant groups across the 14 host
495 mutations where nodules formed. Variant groups with *P* values in the top ten when six or more
496 host genes were perturbed were considered 'pervasive.' In contrast, those appearing when five or
497 fewer genes were mutated were considered 'limited effect' genes.
498
499 Acknowledgments
500 We thank Prof. Dong Wang of University of Massachusetts, Amherst for generously providing
501 the *M.truncatula* mutant seeds, Lillian Cherry, Maria Alejandra Gil-Polo, and Jennifer E. Harris
502 for help with the tedious task of picking and processing nodules. Prof. Katy Heath of University)
503 of Illinois Urbana- Champaign and Dr. Mike Grillo (currently Assistant Professor in Loyola
504 University, Chicago) for providing strains and sequencing data. This work was supported by the
505 National Science Foundation grant IOS-1856744 to PT, NY, and LTB and IOS-2243819 to PT
506 and LTB and was completed using computing resources provided by Penn State Institute for
507 Computational and Data Science and sequencing by the Huck Genomics Core Facility. LTB's
508 work was also supported by HATCH grant no. 1025611. The authors acknowledge the Huck
509 Institutes' Genomics Core Facility (RRID:SCR_023645) for use the Illumina NextSeq 2000.
510
511 Authorship Credits
512 LTB, NY, PT idea conception and funding, LTB experimental design, LTB, and RGB plant
513 care; LTB, SG, RB drafted the manuscript; RB, GF, LTB data collection, SG, LTB, RGB, BE,
514 JS analysis, SG, LTB, RB, JS, BE, PT, NY, GF manuscript revisions
515
516 Data Accessibility
517 All the raw reads from sequencing the rhizobial strains in the initial community and in the
518 rhizobial genomic DNA extracted from the nodule pools of the individual host mutants have
519 been deposited in NCBI under the Bioproject PRJNA401437 (Temporary Submission ID:
520 SUB14622715). All raw data and analyses code will be uploaded to the Github.
521
522 Figure Legends
523 **Fig.1: Experimental design schematic**. We grew 86 sequenced *Sinorhizobium meliloti s*trains
524 in liquid media and combined them in approximately equal proportions to generate a multi-strain
525 community. We used this community to inoculate the *M.truncatula* wildtype (A17) and 18

526 mutant hosts. We measured the traits which represent common proxies for host investment and
527 benefits. In addition, we measured rhizobial benefits including the population of rhizobia
528 released per nodule and per plant and applied whole genome shotgun sequencing to determine
529 the change in the relative fitness of the rhizobial strains prior to and after exposure to the host
530 mutants.
531
532 **Fig.2**: **Host symbiotic mutations reduce host benefits and differentially impact host**
533 **investments even in mixed inoculations**. Hosts benefit traits like (A) total biomass per plant
534 and (B) root-to-shoot ratio were reduced and increased respectively in almost all the host mutants
535 except *sunn4*. The host investment traits which include (C) total nodule area per plant, (D)
536 nodules per plant, and (E) average area per nodule were differentially affected by the host
537 mutants. Up/down triangles denote when a host mutant significantly increases/decreases a trait
538 with respect to WT. The boxplots are colour coded based on the stage of symbiosis (disruption
539 type) to which the individual host mutants belong (See Table S3 for details).
540
541 **Fig.3: Mutating host symbiosis genes affect rhizobial benefits.** Host mutations differentially
542 influenced rhizobial benefit traits like (A) CFU/plant and (B) Cfu/nodule while (C) nodule
543 isolate diversity (C) was largely unaffected by the majority of host mutants. Up/down triangles
544 denote when a host mutant significantly decreases/increases a trait with respect to WT
545 respectively. The boxplots are colour coded based on the stage of symbiosis (Mutation type) to
546 which the individual host mutants belong (see Table S3 for details).
547
548 **Fig.4: Heatmap visualizing variation in strain fitness in pools of wildtype and mutant**
549 **nodules**. The host mutants (left) are hierarchically clustered based on the similarity of median
550 strain fitness patterns (Ward D method). The strain fitness are ordered (bottom) by strain fitness
551 in A17 from high relative fitness (red) to low (blue). Red indicates strain enrichment relative to
552 the strain frequency in WT, white indicates no fitness change, and blue indicates depletion
553 relative to WT.
554
555 **Fig.5: Rhizobial candidate genes underlying strain fitness shifts predominantly occur on**
556 **pSymA and pSymB**. (A) Manhattan plot indicating SNP(single nucleotide polymorphism)
557 positions in *S. meliloti* USDA1106 with strong statistical association with strain fitness shifts
558 across the 14 host mutants when compared to wildtype A17 hosts. The dotted line indicates the P
559 value threshold used for identifying candidates genes and for clarity only significant associations
560 are color-coded according to the host gene mutation. GWAs were conducted separately for each
561 replicon to account for divergent population structure, but are merged here for ease of
562 comparison. (B) Heatmaps based on hierarchical clustering of the inverse P-value ranks of the
563 top five variant groups in each of the 14 host mutations. The heatmap shows how the P value
564 ranks of the variant groups change across the spectrum of the host mutations.
565
566 References
567 1. Oldroyd GED, Murray JD, Poole PS, Downie JA. The Rules of Engagement in the Legume-
568 Rhizobial Symbiosis. Annu Rev Genet. 2011 Dec 15;45(1):119–44.

2. Poole P, Ramachandran V, Terpolilli J. Rhizobia: from saprophytes to endosymbionts. Nat Rev Microbiol. 2018 Jan;180:1691.

3. Vance CP. Symbiotic Nitrogen Fixation and Phosphorus Acquisition. Plant Nutrition in a World of Declining Renewable Resources. Plant Physiol. 2001 Oct 1;127(2):390–7.

4. Udvardi M, Below FE, Castellano MJ, Eagle AJ, Giller KE, Ladha JK, et al. A Research Road Map for Responsible Use of Agricultural Nitrogen. Front Sustain Food Syst [Internet]. 2021 May 31 [cited 2024 Aug 9];5. Available from: https://www.frontiersin.org/journals/sustainable-food-systems/articles/10.3389/fsufs.2021.660155/full

5. Jhu MY, Oldroyd GED. Dancing to a different tune, can we switch from chemical to biological nitrogen fixation for sustainable food security? PLOS Biol. 2023 Mar 14;21(3):e3001982.

6. Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, et al. Celebrating 20 Years of Genetic Discoveries in Legume Nodulation and Symbiotic Nitrogen Fixation. Plant Cell. 2020 Jan;32(1):15–41.

7. Barker DG, Bianchi S, Blondon F, Dattée Y, Duc G, Essad S, et al. Medicago truncatula, a model plant for studying the molecular genetics of theRhizobium-legume symbiosis. Plant Mol Biol Report. 1990 Feb;8(1):40–9.

8. Mitra RM, Shaw SL, Long SR. Six nonnodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. Proc Natl Acad Sci. 2004 Jul 6;101(27):10217–22.

9. Domonkos A, Horvath B, Marsh JF, Halasz G, Ayaydin F, Oldroyd GE, et al. The identification of novel loci required for appropriate nodule development in Medicago truncatula. BMC Plant Biol. 2013 Dec;13(1):157.

10. Wang Q, Liu J, Li H, Yang S, Körmöczi P, Kereszt A, et al. Nodule-Specific Cysteine-Rich Peptides Negatively Regulate Nitrogen-Fixing Symbiosis in a Strain-Specific Manner in *Medicago truncatula*. Mol Plant-Microbe Interactions®. 2018 Feb;31(2):240–8.

11. Pislariu CI, Sinharoy S, Torres-Jerez I, Nakashima J, Blancaflor EB, Udvardi MK. The Nodule-Specific PLAT Domain Protein NPD1 Is Required for Nitrogen-Fixing Symbiosis. Plant Physiol. 2019 Jul;180(3):1480–97.

12. Jardinaud MF, Carrere S, Gourion B, Gamas P. Symbiotic Nodule Development and Efficiency in the Medicago truncatula Mtefd-1 Mutant Is Highly Dependent on Sinorhizobium Strains. Plant Cell Physiol. 2023 Feb 16;64(1):27–42.

13. Burghardt LT, diCenzo GC. The evolutionary ecology of rhizobia: multiple facets of competition before, during, and after symbiosis with legumes. Curr Opin Microbiol. 2023 Apr 1;72:102281.

605 14. Burghardt LT, Epstein B, Guhlin J, Nelson MS, Taylor MR, Young ND, et al. Select and
606 resequence reveals relative fitness of bacteria in symbiotic and free-living environments.
607 Proc Natl Acad Sci. 2018 Mar 6;115(10):2425–30.

608 15. Batstone RT, Burghardt LT, Heath KD. Phenotypic and genomic signatures of interspecies
609 cooperation and conflict in naturally occurring isolates of a model plant symbiont. Proc R
610 Soc B Biol Sci. 2022 Jul 13;289(1978):20220477.

611 16. Epstein B, Burghardt LT, Heath KD, Grillo MA, Kostanecki A, Hämälä T, et al. Combining
612 GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia
613 symbiosis. Mol Ecol. 2022 Jul 21;mec.16602.

614 17. Burghardt LT, Trujillo DI, Epstein B, Tiffin P, Young ND. A Select and Resequence
615 Approach Reveals Strain-Specific Effects of *Medicago* Nodule-Specific PLAT-Domain
616 Genes. Plant Physiol. 2020 Jan;182(1):463–71.

617 18. Oldroyd GED. Speak, friend, and enter: signalling systems that promote beneficial symbiotic
618 associations in plants. Nat Rev Microbiol. 2013 Apr;11(4):252–63.

619 19. Wheatley RM, Ford BL, Li L, Aroney STN, Knights HE, Ledermann R, et al. Lifestyle
620 adaptations of Rhizobium from rhizosphere to symbiosis. Proc Natl Acad Sci U S A. 2020
621 Sep 22;117(38):23823–34.

622 20. Bright LJ, Liang Y, Mitchell DM, Harris JM. The *LATD* Gene of *Medicago truncatula* Is
623 Required for Both Nodule and Root Development. Mol Plant-Microbe Interactions®. 2005
624 Jun;18(6):521–32.

625 21. Hogg BV, Cullimore JV, Ranjeva R, Bono JJ. The *DMI1* and *DMI2* Early Symbiotic Genes
626 of *Medicago truncatula* Are Required for a High-Affinity Nodulation Factor-Binding Site
627 Associated to a Particulate Fraction of Roots. Plant Physiol. 2006 Jan 1;140(1):365–73.

628 22. Messinese E, Mun JH, Yeun LH, Jayaraman D, Rougé P, Barre A, et al. A Novel Nuclear
629 Protein Interacts With the Symbiotic DMI3 Calcium- and Calmodulin-Dependent Protein
630 Kinase of *Medicago truncatula*. Mol Plant-Microbe Interactions®. 2007 Aug;20(8):912–21.

631 23. Smit P, Limpens E, Geurts R, Fedorova E, Dolgikh E, Gough C, et al. Medicago LYK3, an
632 Entry Receptor in Rhizobial Nodulation Factor Signaling. Plant Physiol. 2007 Sep
633 6;145(1):183–91.

634 24. Ovchinnikova E, Journet EP, Chabaud M, Cosson V, Ratet P, Duc G, et al. IPD3 Controls
635 the Formation of Nitrogen-Fixing Symbiosomes in Pea and *Medicago* Spp. Mol Plant-
636 Microbe Interactions®. 2011 Nov;24(11):1333–44.

637 25. Starker CG, Parra-Colmenares AL, Smith L, Mitra RM, Long SR. Nitrogen Fixation Mutants
638 of *Medicago truncatula* Fail to Support Plant and Bacterial Symbiotic Gene Expression.
639 Plant Physiol. 2006 Feb 1;140(2):671–80.

640 26. Schnabel E, Journet EP, de Carvalho-Niebel F, Duc G, Frugoli J. The Medicago truncatula
641 SUNN Gene Encodes a CLV1-like Leucine-rich Repeat Receptor Kinase that Regulates
642 Nodule Number and Root Length. Plant Mol Biol. 2005 Aug;58(6):809–22.

643 27. Schnabel EL, Kassaw TK, Smith LS, Marsh JF, Oldroyd GE, Long SR, et al. The *ROOT*
644 *DETERMINED NODULATION1* Gene Regulates Nodule Number in Roots of *Medicago*
645 *truncatula* and Defines a Highly Conserved, Uncharacterized Plant Gene Family. Plant
646 Physiol. 2011 Sep 3;157(1):328–40.

647 28. Domonkos Á, Kovács S, Gombár A, Kiss E, Horváth B, Kováts G, et al. NAD1 Controls
648 Defense-Like Responses in Medicago truncatula Symbiotic Nitrogen Fixing Nodules
649 Following Rhizobial Colonization in a BacA-Independent Manner. Genes. 2017 Dec
650 14;8(12):387.

651 29. Agren GI. Root : Shoot Ratios, Optimization and Nitrogen Productivity. Ann Bot. 2003 Oct
652 8;92(6):795–800.

653 30. Burghardt LT. Evolving together, evolving apart: measuring the fitness of rhizobial bacteria
654 in and out of symbiosis with leguminous plants. New Phytol. 2020 Oct;228(1):28–34.

655 31. Denison RF. Legume Sanctions and the Evolution of Symbiotic Cooperation by Rhizobia.
656 Am Nat. 2000 Dec;156(6):567–76.

657 32. Wang C, Yu H, Luo L, Duan L, Cai L, He X, et al. *NODULES WITH ACTIVATED*
658 *DEFENSE 1* is required for maintenance of rhizobial endosymbiosis in *Medicago truncatula*.
659 New Phytol. 2016 Oct;212(1):176–91.

660 33. Harris JM, Dickstein R. Control of root architecture and nodulation by the *LATD/NIP*
661 transporter. Plant Signal Behav. 2010 Nov;5(11):1365–9.

662 34. Heath KD, Stinchcombe JR. EXPLAINING MUTUALISM VARIATION: A NEW
663 EVOLUTIONARY PARADOX?: PERSPECTIVE. Evolution. 2014 Feb;68(2):309–17.

664 35. Carlson C, Frederickson ME. An emerging view of coevolution in the legume–rhizobium
665 mutualism. Mol Ecol. 2023;32(14):3793–7.

666 36. Wang D, Griffitts J, Starker C, Fedorova E, Limpens E, Ivanov S, et al. A Nodule-Specific
667 Protein Secretory Pathway Required for Nitrogen-Fixing Symbiosis. Science. 2010 Feb
668 26;327(5969):1126–9.

669 37. Kim M, Chen Y, Xi J, Waters C, Chen R, Wang D. An antimicrobial peptide essential for
670 bacterial survival in the nitrogen-fixing symbiosis. Proc Natl Acad Sci. 2015 Dec
671 8;112(49):15238–43.

672 38. Madsen LH, Tirichine L, Jurkiewicz A, Sullivan JT, Heckmann AB, Bek AS, et al. The
673 molecular network governing nodule organogenesis and infection in the model legume Lotus
674 japonicus. Nat Commun. 2010 Apr 12;1(1):10.

675 39. Denison RF, Kiers ET. Life histories of symbiotic rhizobia and mycorrhizal fungi. Curr Biol
676 CB. 2011 Sep;21(18):R775-85.

677 40. Gopalakrishnan S, Sathya A, Vijayabharathi R, Varshney RK, Gowda CLL, Krishnamurthy
678 L. Plant growth promoting rhizobia: challenges and opportunities. 3 Biotech. 2015
679 Aug;5(4):355–77.

680 41. Burghardt LT, Epstein B, Guhlin J, Nelson MS, Taylor MR, Young ND, et al. Select and
681 resequence reveals relative fitness of bacteria in symbiotic and free-living environments.
682 Proc Natl Acad Sci. 2018 Mar 6;115(10):2425–30.

683 42. Walker L, Lagunas B, Gifford ML. Determinants of host range specificity in legume-
684 rhizobia symbiosis. Front Microbiol [Internet]. 2020 [cited 2023 Jan 10];11. Available from:
685 https://www.frontiersin.org/articles/10.3389/fmicb.2020.585749

686 43. Horváth B, Domonkos Á, Kereszt A, Szűcs A, Ábrahám E, Ayaydin F, et al. Loss of the
687 nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the
688 Medicago truncatula dnf7 mutant. Proc Natl Acad Sci U S A. 2015 Dec 8;112(49):15232–7.

689 44. Guefrachi I, Nagymihaly M, Pislariu CI, Van de Velde W, Ratet P, Mars M, et al. Extreme
690 specificity of NCR gene expression in Medicago truncatula. BMC Genomics. 2014 Aug
691 25;15(1):712.

692 45. Lima RM, Kylarová S, Mergaert P, Kondorosi É. Unexplored Arsenals of Legume Peptides
693 With Potential for Their Applications in Medicine and Agriculture. Front Microbiol. 2020
694 Jun 18;11:1307.

695 46. Burghardt LT, Guhlin J, Chun CL, Liu J, Sadowsky MJ, Stupar RM, et al. Transcriptomic
696 basis of genome by genome variation in a legume-rhizobia mutualism. Mol Ecol. 2017
697 Nov;26(21):6122–35.

698 47. Stanton-Geddes J, Paape T, Epstein B, Briskine R, Yoder J, Mudge J, et al. Candidate Genes
699 and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by Whole-Genome,
700 Sequence-Based Association Genetics in Medicago truncatula. Lukens L, editor. PLoS ONE.
701 2013 May 31;8(5):e65688.

702 48. Curtin SJ, Tiffin P, Guhlin J, Trujillo DI, Burghardt LT, Atkins P, et al. Validating Genome-
703 Wide Association Candidates Controlling Quantitative Variation in Nodulation. Plant
704 Physiol. 2017 Feb;173(2):921–31.

705 49. Savolainen O, Lascoux M, Merilä J. Ecological genomics of local adaptation. Nat Rev
706 Genet. 2013 Nov;14(11):807–20.

707 50. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, et al. Finding the
708 Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. Am
709 Nat. 2016 Oct 1;188(4):379–97.

710 51. Elena SF. Local adaptation of plant viruses: lessons from experimental evolution. Mol Ecol.
711 2017;26(7):1711–9.

712 52. Roux B, Rodde N, Jardinaud M, Timmers T, Sauviac L, Cottret L, et al. An integrated
713 analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture
714 microdissection coupled to RNA sequencing. Plant J. 2014 Mar;77(6):817–37.

715 53. Karunakaran R, East AK, Poole PS. Malonate Catabolism Does Not Drive N_2 Fixation in
716 Legume Nodules. Appl Environ Microbiol. 2013 Jul 15;79(14):4496–8.

717 54. Stammers DK. The structure of the negative transcriptional regulator NmrA reveals a
718 structural superfamily which includes the short-chain dehydrogenase/reductases. EMBO J.
719 2001 Dec 3;20(23):6619–26.

720 55. Mus F, Crook MB, Garcia K, Garcia Costas A, Geddes BA, Kouri ED, et al. Symbiotic
721 Nitrogen Fixation and the Challenges to Its Extension to Nonlegumes. Appl Environ
722 Microbiol. 2016 Jul;82(13):3698–710.

723 56. Sugawara M, Takahashi S, Umehara Y, Iwano H, Tsurumaru H, Odake H, et al. Variation in
724 bradyrhizobial NopP effector determines symbiotic incompatibility with Rj2-soybeans via
725 effector-triggered immunity. Nat Commun. 2018 Aug 7;9(1):3139.

726 57. Riley AB, Grillo MA, Epstein B, Tiffin P, Heath KD. Partners in space: Discordant
727 population structure between legume hosts and rhizobium symbionts in their native range
728 [Internet]. Evolutionary Biology; 2021 Jul [cited 2023 Mar 27]. Available from:
729 http://biorxiv.org/lookup/doi/10.1101/2021.06.30.449460

730 58. Nelson M, Guhlin J, Epstein B, Tiffin P, Sadowsky MJ. The complete replicons of 16
731 Ensifer meliloti strains offer insights into intra- and inter-replicon gene transfer, transposon-
732 associated loci, and repeat elements. Microb Genomics [Internet]. 2018 May 1 [cited 2023
733 Jun 17];4(5). Available from:
734 https://www.microbiologyresearch.org/content/journal/mgen/10.1099/mgen.0.000174

735 59. Kessner D, Turner TL, Novembre J. Maximum Likelihood Estimation of Frequencies of
736 Known Haplotypes from Pooled Sequence Data. Mol Biol Evol. 2013 May 1;30(5):1145–58.

737 60. Zhou X, Stephens M. Genome-wide efficient mixed-model analysis for association studies.
738 Nat Genet. 2012 Jul;44(7):821–4.

739 61. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and
740 population genetical parameter estimation from sequencing data. Bioinformatics. 2011 Nov
741 1;27(21):2987–93.

742 62. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A
743 Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. Am J
744 Hum Genet. 2007 Sep;81(3):559–75.

745

746 Tables

747 Table 1: Definitions of the symbiotic traits examined in this study

Symbiotic Traits	Category	Interpretation
Nodules per plant	Host Investment	Indicates 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* 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 the nodule pool indicates host selectivity and strength of selection
Strain relative fitness	Rhizobial Benefit	Enrichment/depletion of specific strains in nodule pools relative to other strains

748 *CFU= Colony Forming Units

749 Table 2: Comparison of the variance in host and rhizobial traits explained by individual
750 symbiosis genes (Genotype) versus the stage of symbiosis (Type) in a multi-strain inoculation.

Trait	Geno[1]	Type [1]	Geno + Block R_{adj}^2	Type + Block R_{adj}^2
Nodule number per plant	66	42	0.59	0.37
Average area per nodule	61	24	0.50	0.24
Total nodule area per plant	85	61	0.85	0.63
Total biomass per plant per plant	80	25	0.78	0.21
Root-to-shoot ratio	89	39	0.88	0.36
CFU per nodule	61	13	0.54	0.14
CFU per plant	84	40	0.84	0.42
Nodule strain diversity	58	23	0.44	0.19
Strain relative fitness	36	13	0.22	0.09

751 **Notes**: All terms explained more variance than expected by chance (P_{value}<0.001) except Type for CFU per nodules
752 (P_{value} = 0.03). All results are from linear models except Strain relative fitness, which is from a multivariate RDA
753 analysis and permANOVA. [1]Percentage of Variance Explained, [2]R_{adj} = Adjusted R-squared; Block= Replicate
754

755 Table 3: Candidate rhizobial genes in small linkage groups (LD) (three genes or less) associated
756 with ($P < 10^{-5}$) variation in rhizobial relative strain fitness.

LD group[1]	Host mutants	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*	0.07	(+)	(-)	1	CDO30_RS23065:SGNH/GDSL hydrolase family protein
6978.alt[4]	*dnf1, dnf1/dnf2, dnf2, dnf4, dnf7, dnf5/2, ipd3, rdn1, sunn1*	0.07	(+)	none	1	CDO30_RS23160:NAD(P)H-binding protein
5279	*dnf1, dnf2, dnf7, ipd3, nad1,rdn1, sunn1*	0.07	(+)	(-)	1	CDO30_RS21460:malonyl-CoA synthase
7969	*dnf1, dnf2, dnf6, nad1, sunn4*	0.06	(+)	none	1	CDO30_RS22325:hypothetical protein
6968	*dnf4, dnf7,ipd3, sunn1, sunn4*	0.10	(+)	none	1	CDO30_RS23125:SDR family oxidoreductase (pseudogene)
10122	*dnf1, dnf1/dnf2, dnf2, sunn1*	0.07	(+)	(-)	1	CDO30_RS27255:hypothetical protein
12516	*dnf1dnf2, rdn1, sunn1*	0.06	(+)	none	2	CDO30_RS28530:ABC transporter ATP-binding protein; CDO30_RS28525:*RkpR*
10601	*dnf7, ipd3, ltd*	0.05	(+)	none	1	CDO30_RS28740:*pepT*, peptidase T
3770	*rdn1, sunn4*	0.14	(+)	(-)	1	CDO30_RS18080:*fixL*, oxygen sensor histidine kinase *fixL*
12013	*dnf1dnf2, ipd3*	0.05	(+)	none	1	CDO30_RS28675:hypothetical protein
10882	*dnf4*	0.05	(+)	none	1	CDO30_RS30150:glycosyltransferase
9671	*dnf3*	0.07	(+)	none	3	CDO30_RS27435:four-carbon acid sugar kinase family protein
10140	*dnf3*	0.08	(+)	none	1	CDO30_RS27435:four-carbon acid sugar kinase family protein
10913	*dnf3*	0.05	(+)	none	2	CDO30_RS27440:xanthine dehydrogenase family protein molybdopterin-binding subunit; CDO30_RS27445:xanthine dehydrogenase family protein subunit M
10915	*dnf3*	0.05	(+)	none	42	CDO30_RS27440:xanthine dehydrogenase family protein molybdopterin-binding subunit; CDO30_RS27445:xanthine dehydrogenase family protein subunit M
8294	*dnf4*	0.07	(+)	none	1	CDO30_RS19605:*stc4*, stachydrine N-demethylase reductase subunit Stc4
8754	*dnf5/dnf2*	0.1	(+)	none	2	CDO30_RS32445:hypothetical protein
9239	*dnf5/dnf2*	0.09	(+)	none	1	CDO30_RS32445:hypothetical protein
10903	*ipd3*	0.05	(+)	none	1	CDO30_RS27310:ABC transporter ATP-binding protein
838	*latd*	0.09	(+)	none	1	CDO30_RS12270:glutamine synthetase family protein
4107	*nad1*	0.11	(-)	none	1	CDO30_RS22005:amidase
4386	*nad1*	0.09	(-)	none	2	CDO30_RS22060:Atu4866 domain-containing protein;CDO30_RS22055:class I SAM-dependent methyltransferase;
9865	*rdn1*	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*	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*	0.07	(+)	none	1	CDO30_RS25485:pyridoxamine 5'-phosphate oxidase family protein
10728	*rdn1*	0.5	(+)	none	1	CDO30_RS32910:LysR family transcriptional regulator
12841	*sunn1*	0.07	(+)	none	1	CDO30_RS28530:ABC transporter ATP-binding protein

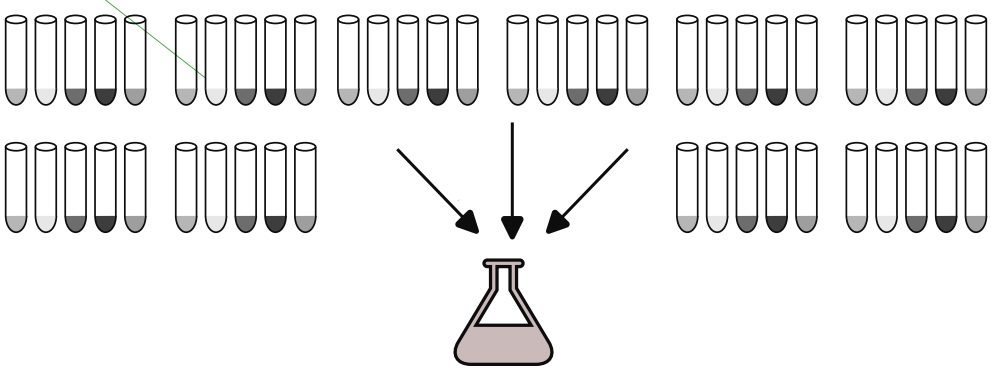
757 **Notes:** [1]LD group numbers were obtained from (16); [2]AF= non-reference (alternate) allele frequency. [3]Locus tags
758 are from the *Sinorhizobium meliloti* USDA1106 reference annotation available on MAGE
759 (http://www.genoscope.cns.fr/agc/microscope/home/index.php). [4]A tri allelic event

1. Grow 86 *S.meliloti* strains individually

2. Combine the cultures to create a mixed inoculum

3. Inoculate *M.truncatula* genotypes & grow in N-limited conditions

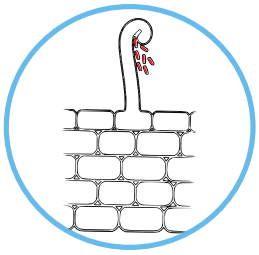
4. Measure nine symbiotic traits after eight weeks



WILDTYPE

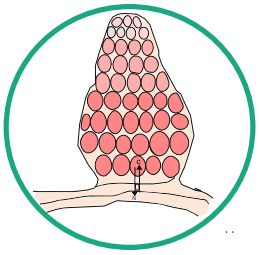
A17

EARLY



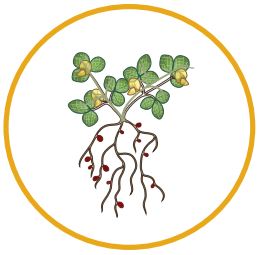
hcl, ipd3, dmi1-2, dmi2-3, dmi3-1, latd

LATE



dnf1, dnf2, dnf3, dnf4, dnf6, dnf7, dnf1/dnf2, dnf2/ dnf5, nad1

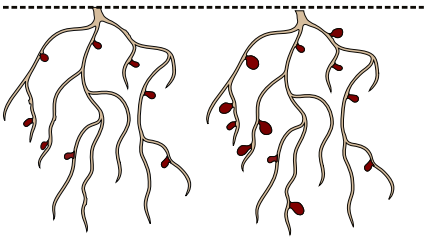
AON



sunn1, sunn4, rdn1

Host Investments

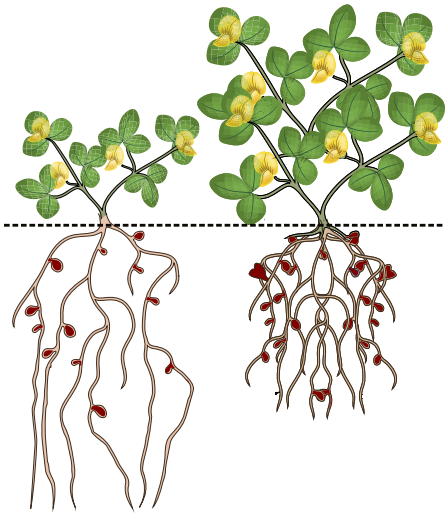
Low → High



- Nodules per plant
- Average area per nodule
- Total nodule area per plant

Host Benefits

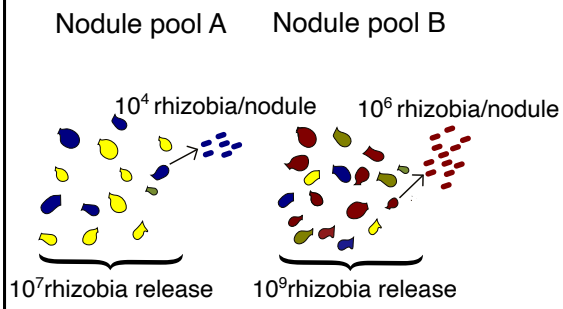
Low → High



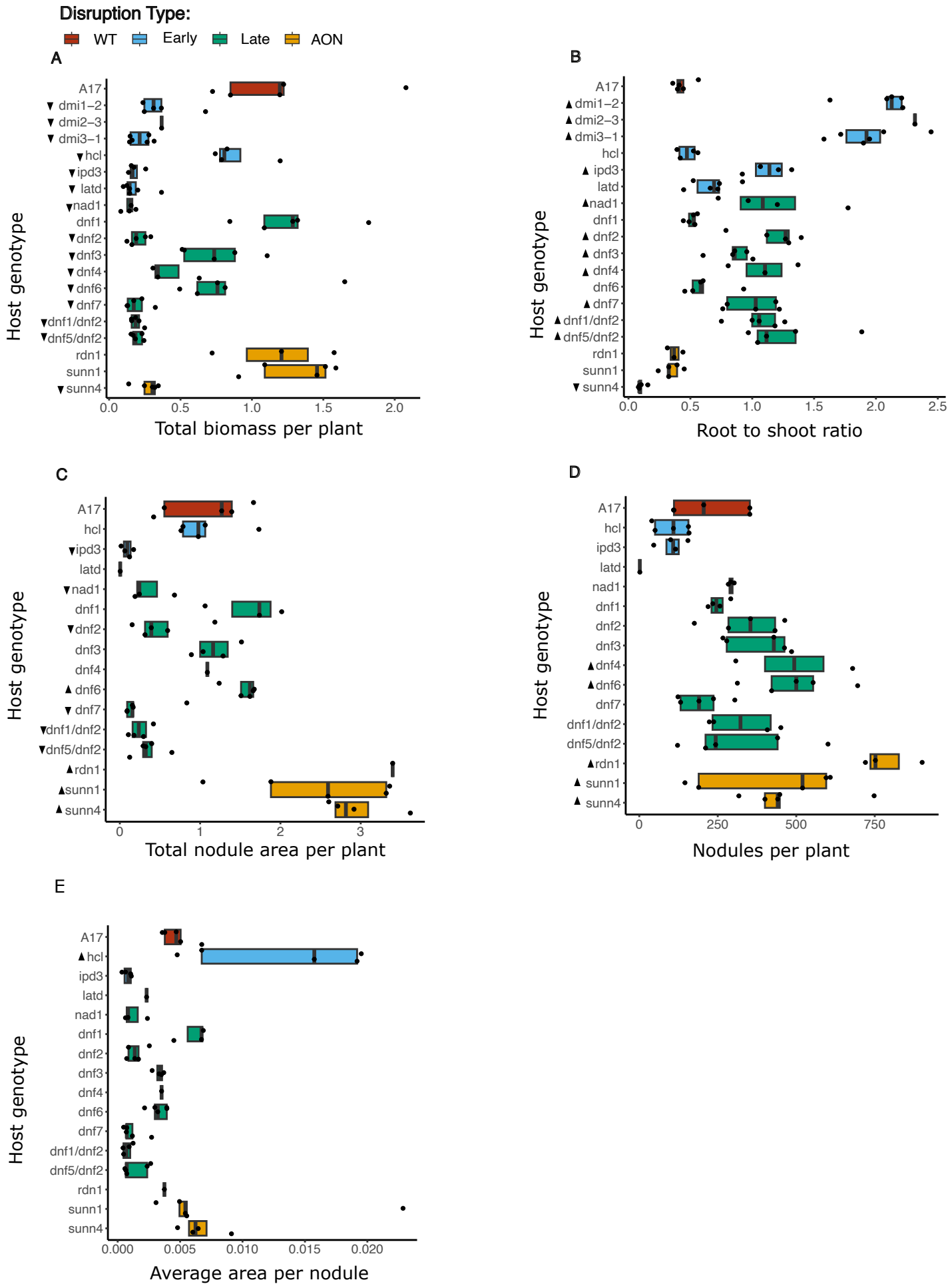
- Total biomass per plant
- Root -to-shoot ratio

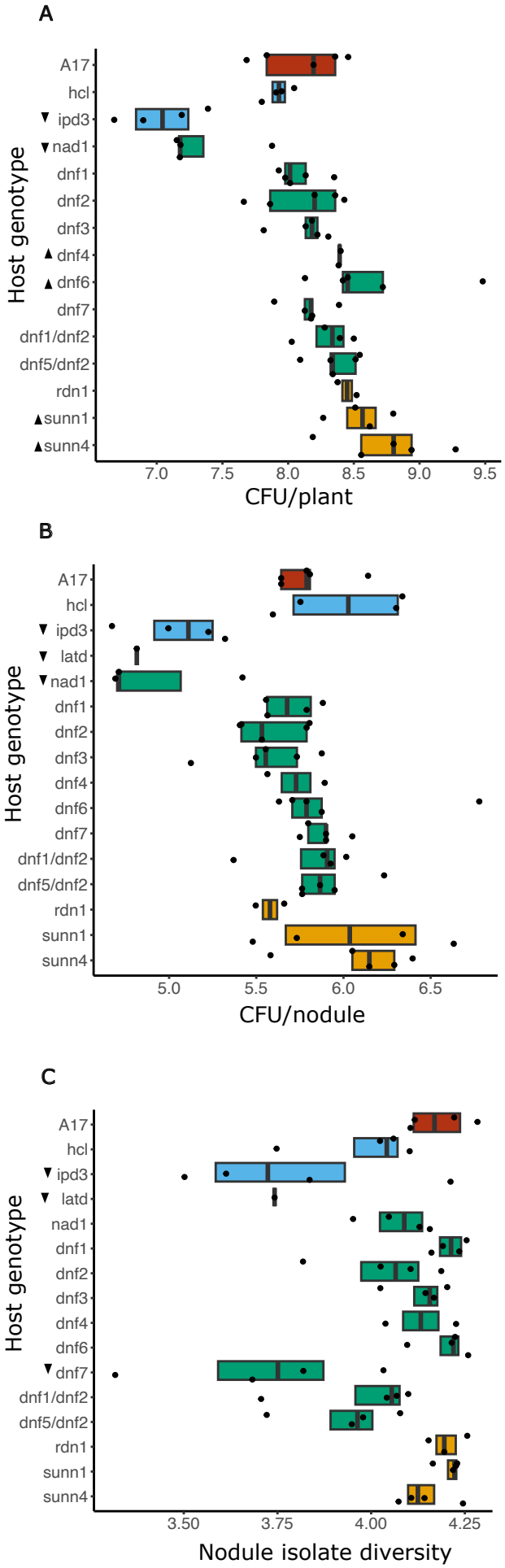
Rhizobial Benefits

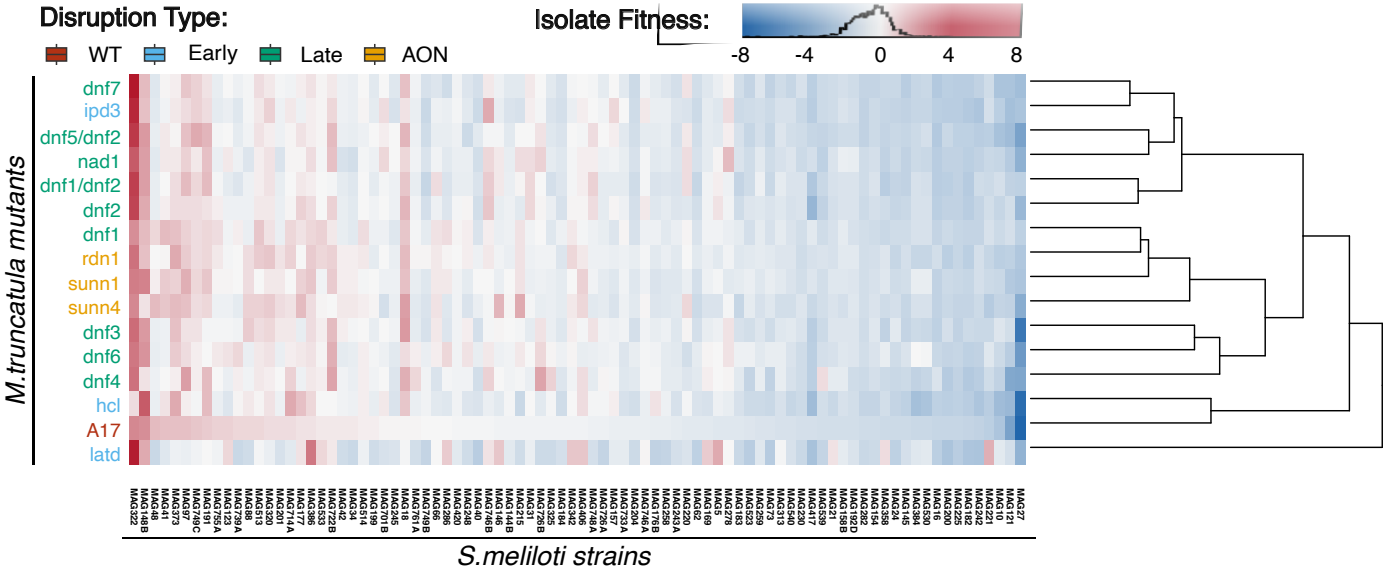
Low → High

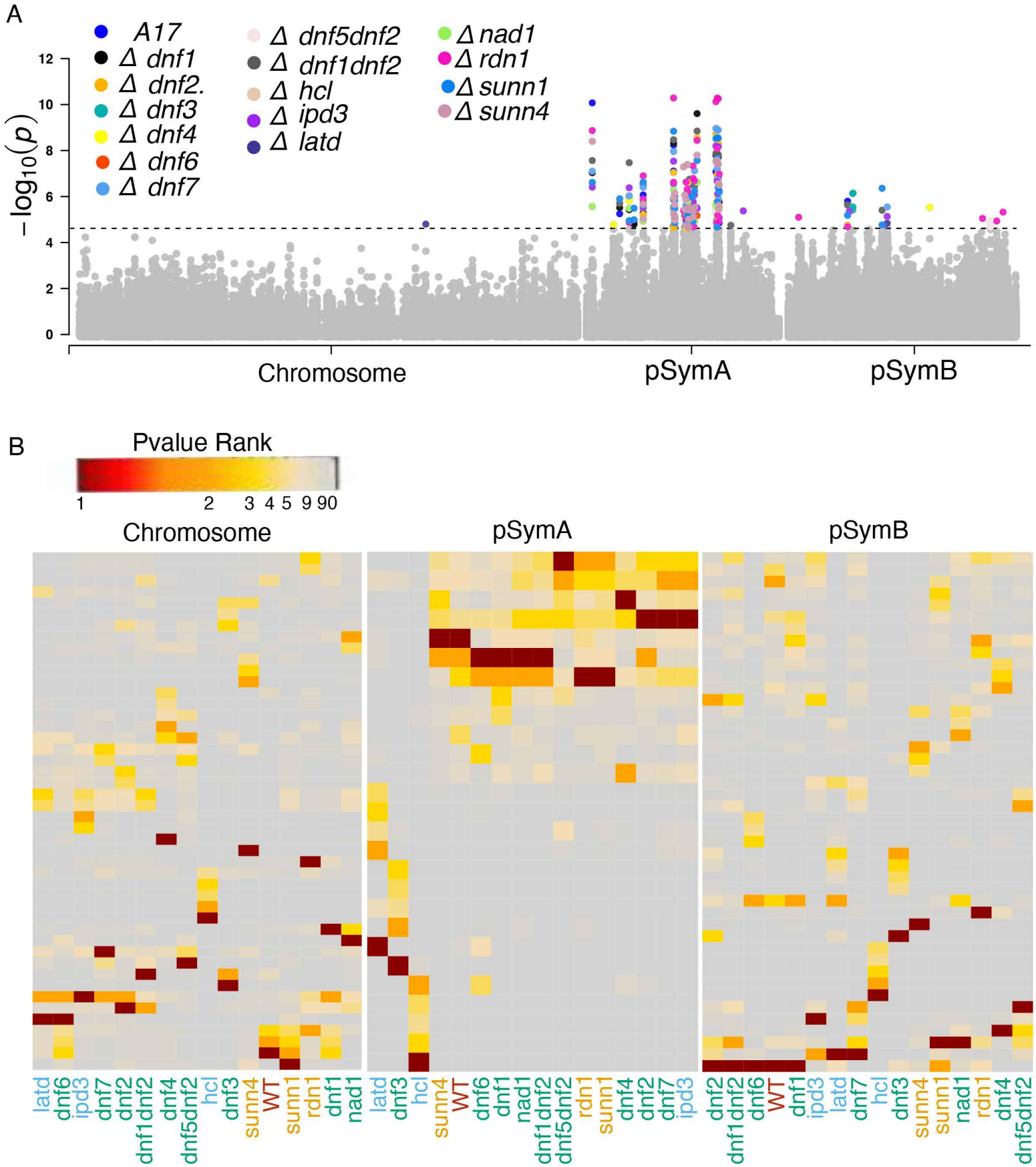


- Rhizobia released per nodule
- Rhizobia released per plant
- Nodule isolate diversity
- Nodule isolate fitness


Disruption Type:
WT
Early
Late
AON
A
Host genotype
A17
▼ dmi1−2
▼ dmi2−3
▼ dmi3−1
▼hcl
▼ipd3
▼ latd
▼nad1
dnf1
▼dnf2
▼dnf3
▼dnf4
▼dnf6
▼dnf7
▼dnf1/dnf2
▼dnf5/dnf2
rdn1
sunn1
▼sunn4
0.0
0.5
1.0
1.5
2.0
Total biomass per plant
B
Host genotype
A17
▲dmi1−2
▲dmi2−3
▲dmi3−1
hcl
▲ ipd3
latd
▲nad1
dnf1
▲ dnf2
▲ dnf3
▲ dnf4
dnf6
▲dnf7
▲dnf1/dnf2
▲dnf5/dnf2
rdn1
sunn1
▼ sunn4
0.0
0.5
1.0
1.5
2.0
2.5
Root to shoot ratio
C
Host genotype
A17
hcl
▼ipd3
latd
▼nad1
dnf1
▼dnf2
dnf3
dnf4
▲ dnf6
▼ dnf7
▼dnf1/dnf2
▼dnf5/dnf2
▲rdn1
▲sunn1
▲ sunn4
0
1
2
3
Total nodule area per plant
D
Host genotype
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
▲dnf4
▲dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
▲rdn1
▲ sunn1
▲ sunn4
0
250
500
750
Nodules per plant
E
Host genotype
A17
▲hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
0.000
0.005
0.010
0.015
0.020
Average area per nodule


A
A17
hcl
ipd3
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
Host genotype
7.0
7.5
8.0
8.5
9.0
9.5
CFU/plant
B
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
Host genotype
5.0
5.5
6.0
6.5
CFU/nodule
C
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
Host genotype
3.50
3.75
4.00
4.25
Nodule isolate diversity


Disruption Type:
WT
Early
Late
AON
Isolate 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 strains


A
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
B
Pvalue Rank
1
2
3
4
5
9
90
Chromosome
pSymA
pSymB
latd
dnf6
ipd3
dnf7
dnf2
dnf1dnf2
dnf4
dnf5dnf2
hcl
dnf3
sunn4
WT
sunn1
rdn1
dnf1
nad1
latd
dnf3
hcl
sunn4
WT
dnf6
dnf1
nad1
dnf1dnf2
dnf5dnf2
rdn1
sunn1
dnf4
dnf2
dnf7
ipd3
dnf2
dnf1dnf2
dnf6
WT
dnf1
ipd3
latd
dnf7
hcl
dnf3
sunn4
sunn1
nad1
rdn1
dnf4
dnf5dnf2