Title: Mutations in legume symbiosis genes create a complex selective landscape for natural genetic variation in rhizobia during multi-strain inoculations

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.

Introduction

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

Affordable whole genome shotgun sequencing has recently enabled the high-throughput assessment of rhizobial strain frequencies inside the nodules of *M. truncatula* when multiple strains compete for nodulation(14–16) . Here, we take advantage of this method to evaluate the effects of 18 well-studied *M. truncatula* mutants (knockouts or knockdowns) on the host

symbiotic and the rhizobial traits when *M. truncatula* is grown with a mixture of 86 *Sinorhizobium meliloti* (*S. meliloti*) strains. Besides measuring standard symbiotic phenotypes, we estimated the relative fitness of each of the rhizobial strains in the community within each host mutant. We also performed genome-wide associations (GWA) on rhizobial strain fitness shifts in response to host mutations to identify the rhizobia genes potentially responsible for rhizobia genome by host genome interactions (Burghardt et al., 2018; Burghardt et al., 2020). This novel approach allowed us to examine how mutations in host symbiotic genes may perturb the selection landscape for naturally occurring genetic variation in rhizobia.

The formation and development of symbiosis in nodules is complex, with each step regulated by multiple host and rhizobial genes (6,18,19). To encompass these steps, we chose 18 host symbiosis mutants (Table S1). We grouped them into three categories based on their impact on 1) bacterial invasion (Early), 2) nodule organogenesis and nitrogen fixation (Late), or 3) regulation of nodule number (Autoregulation of nodule number; AON). The early-stage genes (*hcl*, *ipd3*, *latd*, *dmi1-2*, *dmi2-3*, and *dmi3-1*) affect the initial processes such as Nod Factor recognition and nodule organogenesis (20–24). The late-stage genes (*nad1*, *dnf1*, *dnf2*, *dnf4*, *dnf2/dnf1*, *dnf5/dnf2*, *dnf6*, *dnf7*) influence nitrogen fixation (25), while the AON pathway genes (*sunn1*, *sunn4*, *rdn1*) regulate the overall number of nodules a plant forms (26,27). While these loose categorizations help set baseline expectations for the impact of the host mutations on the legume and the rhizobia involved, similar phenotypes can be caused by mutations in different genetic pathways, and each gene may have additional pleiotropic effects or act in additional contexts of nodule development when faced with more than one strain. Together, these focal genotypes provide a broad view of the effects of host symbiosis genes on their rhizobial partners in multi-strain inoculations.

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

While researchers often measure the effect of host mutations on host benefits and investment in symbiosis, it is far less common to measure the effect of host mutations on rhizobia (30). The benefits rhizobia obtain from hosts can be assessed by estimating the *population size of rhizobia* released at the individual nodule and whole plant level (31). Host mutations that increase plant defence responses (*nad1* and *dnf2*) will likely reduce the rhizobia released from each nodule, while population responses at the host plant level could be compensated for by forming more

nodules (32). Host selectivity is another facet of rhizobial fitness that host genes can affect. Some host genes can increase the number of strains able to form nodules, causing an increase in *nodule strain diversity,* while others may reduce diversity. Lastly, host mutations could affect the *relative strain fitness,* or which strains can best compete and thrive in host nodules. To our knowledge, the only assessment of the effect of host mutations on rhizobial benefits during mixed strain inoculations has been conducted on the *NPD* gene family (Nodule Plat Domain encoding genes) of *M. truncatula* where knocking out all 5 NPD genes (Δ*npd1-5*) eliminated nodulation when inoculated with only *S. meliloti* strain SM1021 but inoculating with a multi-strain community of *S. meliloti* , partially restored functional nodulation despite nodules supporting fewer rhizobial strains than WT (17). These results suggest the standard approach of pairing single hosts to single strains and focusing only on plant traits may miss host by strain interactions.

Here, we expanded the range of mutations from the *npds* to additional symbiotic genes with 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.

Materials and methods

Greenhouse experiment to measure multi-strain symbiotic phenotypes

We created a mixed inoculum using 86 sequenced *S. meliloti* strains isolated from soils across the native range of *M. truncatula* (33). Each strain was grown in liquid tryptone yeast media for three days to reach similar population sizes (Fig S1) and then mixed in equal proportions. This community was used to a) inoculate host plants, b) measure rhizobial density via dilution plating, and c) determine strain frequencies via whole genome sequencing. Seeds from *M. truncatula* WT and 18 host mutants (Table S1) were surface-sterilized, scarified, cold-stratified, and sown into pots as described in (14) (see Supplemental Methods for details). Plants were inoculated with 10^6 rhizobia per pot, fertilized with nitrogen-free fertilizer, and harvested 8 weeks post-inoculation to measure host and rhizobial traits (Table 1).

Nodules were photographed to assess plant investment, and their number and size were measured using ImageJ. Plant biomass was measured following oven drying of the roots and shoots. Nodules were pooled, sterilized, and homogenized to estimate rhizobia released via colony counting. Whole genome sequencing of the rhizobial DNA in the initial inoculum and extracted from the nodules using the Qiagen Plant DNeasy kit was done to map the resulting reads to reference genome USDA1106 (34), and calculate the strain frequency using HARP (35).

Calculation of strain diversity and relative fitness

Strain frequencies were used to 1) estimate nodule pool diversity using Shannon diversity (via the "renyi" function in the vegan package in R) and 2) calculate relative fitness by dividing final strain frequency by initial frequency (median initial frequency = 0.0092, 5%-95% quantile: 0.006-0.018) and calculating the $\log_2$ of the relative fitness to normalize the distribution. Strains with zero frequency were assigned a relative fitness of −8, a value less than the lowest measurable fitness, to allow for statistical analysis (36).

Statistical analyses

Permanova ("ANOVA" function) was used to test for the effect of mutant type (trait ~ type + block) and host genotypes (trait ~ genotype + block) on traits. We compared the variance explained by mutant type and host genotype (Table S2) and re-ran models individually on WT and each type or mutant (genotype)(Tables S3, S4). The fraction of variation explained by mutation type was compared to that explained by genotype (Table 2). T-tests were used to compare mutant effects to WT. Correlations between traits were assessed using "corrplot" in R (Fig S2), and community composition shifts were evaluated using redundancy analysis ("RDA" function, Table S7, Fig S5).

Genome-wide association study

We conducted a genome-wide association study (GWA) to identify rhizobial variants responsible for host mutation-dependent strain fitness shifts by subtracting mean strain fitness in mutant nodules from WT nodules. A linear mixed model (LMM) in GEMMA (v0.98.3) was used. We used a published database of 30,076 SNPs representing previously defined linkage-

disequilibrium (LD) groups within *S. meliloti* (15). SNP filtering was performed with bcftools (37) and PLINK (38) based on minor allele frequency (0.05) and missingness (0.2). The proportion of phenotypic variance explained (PVE) by genetic variants was estimated using the K-matrix. P-values were calculated via likelihood ratio tests, and top candidate genes were identified using a P-value cutoff of 10^-5. We ranked these variants across 14 host mutations to categorize them as 'pervasive' if they appeared in six or more perturbations or 'limited effect' if in five or fewer.

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

Host mutations differentially altered host benefit and investment traits. Perhaps unsurprisingly, most genetic mutations influenced both biomass (Fig 2A, Fig S4) and the root-to-shoot ratio (Fig 2B, Fig S4). All mutations reduced biomass except *dnf1, rdn1,* and *sunn1* (Fig 2A). Similarly, the *root-to-shoot ratio* increased relative to the wild type in 11 symbiotic mutants, indicating nitrogen limitation (Fig 2B, Fig S4). In contrast, *sunn4* plants had a low *root-to-shoot ratio,* presumably due to the very low biomass of these plants. Overall, our data suggest that most symbiotic mutations identified in single-strain contexts are not strongly strain-dependent regarding host benefit during multi-strain inoculations, with early- and late-stage mutants

reducing host benefits, but the degree of impairment varied among genes. On the other hand, host investment traits were differentially affected by the mutants. Two out of the three host investment traits (*total nodule area per plant* and *nodules per plant)* were higher than wild type in AON mutants, *dnf4* and *dnf6*, but lower than wild type in six early and late stage mutants (Fig 2C, D; Fig S4). A single mutant (*hcl*) increased *average area per nodule* (Fig 2E, Fig S4). Importantly, if we had only focused on the commonly used metric of *nodules per plant*, we would have missed the effect of many early- and late-stage mutants on overall investment during mixed inoculations. These observations highlight the complexity of symbiotic interactions and underscore the importance of considering multiple host traits to understand the impacts of host genetic mutations on plant performance.

The effect of genetic mutations on rhizobial benefits also depended on the trait. Three mutations (*ipd3, latd,* and *nad1)* reduced both *CFU per plant* and *per nodule* (Fig 3A, B; Fig S4). In contrast, the *CFU per plant* increased in four hosts (*dnf4*, *dnf6, sunn1,* and *sunn4*). Interestingly, none of the mutations increased *CFU per nodule* (Table S4, Fig 3B, Fig S4), but rather, the higher *CFU per plant* was due to forming more nodules. A quarter of genotypic mutations (*ipd3, latd,* and *dnf7)* reduced the *nodule strain diversity*, while none increased diversity (Fig 3C, Fig S4, Table S4). The reduced diversity of strains within *latd* mutants comes with the caveat that *latd* formed very few nodules, so the observed reduced diversity could be due to a reduction in the number of nodules rather than a shift in selectivity *per se*. These results suggest that increased host investment positively influenced rhizobial benefits and that functioning versions of the selected symbiosis genes tend to expand the range of strains with which hosts form associations.

Mutations in early and late symbiosis genes, but not AON genes, lead to significant but subtle changes in strain relative fitness. Interestingly, the eight host mutants affecting strain relative fitness were all early- and late-stage mutants; none of the three AON genes (*sunn1*, *sunn4*, *rdn1*) mutants significantly affected strain fitness (Fig S4, Fig S5, Table S5, Table S6). Hierarchical clustering of the strain relative fitness revealed general trends across host genotypes (Fig 4). WT grouped most closely with *hcl*, while the other early mutations showed varied effects: *latd* was the most divergent, and *ipd3* clustered with late-stage disruptors *dnf7*, *dnf2*, *dnf1/dnf2*, *dnf5/dnf2*, and *nad1*. Notably, *dnf3*, *dnf4*, and *dnf6* formed a distinct group of late mutants with the AON mutants and *dnf1*. Mutants with similar effects on strain fitness also tended to have comparable above-ground biomass, indicating that host resource allocation to nodules influences strain fitness differences. While a few strains showed consistently high or low fitness across all genotypes (e.g., MAG322 vs. MAG27), fitness variability was common across the host mutants. Hierarchical clustering revealed ~71% of strains to be sensitive to host genotypic mutations. Some strains strongly responded to specific genetic mutations, for example- MAG322, MAG386, MAG5, and MAG221 becoming more common with *latd* mutation, and MAG148B with *hcl* mutation. While these shifts are interesting, the overall pattern is of host mutations causing slight fitness shifts across many strains. These subtle shifts give us the power to use GWA to identify rhizobial allelic variants associated with shifts in response to host mutations.

GWA reveals variants on pSymA and pSymB, which are important for rhizobial fitness when symbiosis genes are disrupted. We performed sixteen GWA, one for each of the shifts in strain fitness caused by the fifteen mutants (WT minus mutant fitness) and one for WT fitness alone (Fig 5A), to identify rhizobial variants associated with variation in fitness. The strongest

candidates for genome x genome associations based on *P* values were found on pSymA and pSymB (Fig 5A). We found nine rhizobial linkage groups on pSymA and 17 linkage groups on pSymB that were strongly associated with strain fitness shifts ($P < 10^{-5}$) (Fig 5B, Table 3). In contrast, linkage groups on the chromosome had weaker effects (no P-values less than 10^{-5}), except for LD group 838, which was associated with increased strain fitness within *latd*—a host background with a limited sample size. (Fig 5B*)*. Some of these differences in power between replicons are likely due to the divergent population structure, but the direction of impact is difficult to predict (e.g., the chromosome has long-distance linkage among variants, creating large linkage groups tagging many genes).

Mutating host symbiosis genes has widespread effects on the selective landscape for genetic variation in rhizobia. We were interested to discover if the genetic variants underlying rhizobial fitness, when the symbiosis pathway was perturbed, had pervasive or limited effects. In other words, we sought to summarize the genomic selective landscape that host gene mutation creates for rhizobia, irrespective of arbitrary *P*-value cut-offs. For each replicon, we identified the five most strongly associated candidates and then compared the P ranks (Fig 5B and Fig S6) and effects sizes (Fig S7) of those variant groups across all the host mutations (Table S8 for a complete list). This resulted in a list of 114 LD groups across the three replicons. To visualize patterns, we clustered LD groups based on the inverse of P-value ranks (Fig 5B, Fig S6) and on the effect sizes (beta values) (Fig S7). A vast majority of candidates on the chromosome and pSymB affected strain fitness in fewer than three host mutants (Table S8, Fig S8). Exceptions on the chromosome included a massive linkage group (901) with 144 variants in 65 genes. This LD group was in the top ten rhizobial candidates when nine different host genes were mutated, spanning all disruption types. On pSymB, four small LD groups were top-ranked in at least four

host mutants (ABC transporter and RKpR, a dihydrodipicolinate synthase family protein, pepT, peptidase T, and a hypothetical protein) (Fig 5B). However, the strongest signals of shared genomic selection were on pSymA (Fig S8), where we identified nine linkage groups with pervasive effects on strain fitness shifts (Fig S6). The effect of these genetic variants on strain fitness was not confined to a specific symbiotic stage, although the pervasive variants were not important in two out of the three early-stage mutants (*hcl, latd*) (Fig S6). The non-reference variant increased strain fitness within the host mutants, and interestingly, only in pervasive genes did we see evidence of a fitness trade-off of the allele within WT (Fig S7). The pervasive candidates included genes like malonyl COA synthase, SGNH/GDSL hydrolase family protein, *fixL* which was an oxygen sensor histidine kinase, a short chain oxidoreductase, and a transport protein belonging to the MFS (Major Facilitator Superfamily). Our results indicate that natural genetic variation in metabolic genes plays a vital role in altering strain competitiveness within nodules when symbiotic host genes are disrupted—a role challenging to detect using standard reverse genetics approaches.

Most candidate variants were only important in one or two host symbiotic mutations, suggesting that each mutation creates unique environments for rhizobia. Most of the limited-effect LD groups that were small (less than three genes) and showed substantial statistical support ($p<10^{-5}$) were on pSymB, and many were found with *nad1*, *dnf3*, *dnf4*, *rdn1*, *dnf5/dnf2*, or *sunn1* hosts (Table 3). These genes included a couple of ABC transporters, each associated with individual mutants (*sunn1, ipd3, dnf1/dnf2, rdn1*), glutamine synthetase (*latd*), genes involved in purine metabolism (*dnf3*), methyltransferases (*nad1*), transcriptional regulators (*rdn1*), and a couple of hypothetical proteins. It was interesting to note that for *latd*, which is

mutated in a nitrate transporter, the gene that is most associated with the observed strain fitness shift is glutamine synthetase (39). Interestingly, when *dnf3* was disrupted, a LD group containing 45 SNPs in two genes (xanthine dehydrogenase family protein molybdopterin-binding subunit and xanthine dehydrogenase family protein subunit M), increased strain fitness. Most of these rhizobial genes did not significantly alter strain fitness in the WT host background, suggesting this variation may be conditionally neutral. Overall, our results highlight the complexity of host genome by strain genome interactions underlying rhizobial selection.

Discussion

This study enhances our understanding of how genetic changes in legume hosts affect the environmental context and fitness of rhizobia within root nodules and the strain-specific nature of these responses. While single-strain inoculations allow the direct measurement of phenotypic outcomes, they limit understanding of relative strain responses and host gene effects. Using multi-strain inoculations, we screened legume symbiotic mutants to find strain-specific effects of symbiosis genes on host and rhizobial traits -thus exploring potential mechanisms that could affect the population size and diversity of mutualistic partners (16,17,40,41). This study expanded on our previous work, revealing the effects of the *NPD* gene family on rhizobial strain fitness during multi-strain inoculation (17). We showed that host genetic mutations persistently affect host benefits, with variable effects on host investment and rhizobial benefits. Finally, by implementing a GWA, we map the selective landscape host gene mutations create for rhizobia and identify candidate genes underlying the strain fitness shifts in response to each host mutation. We find rhizobial genetic variants with both pervasive and limited effects across host mutants and that symbiotic mutation categories provide little clarity about how rhizobial genes

influence strain fitness. Taken together, we present an overview of the selective landscape that mutating symbiotic host genes creates for rhizobia, which can be leveraged to fine-tune this mutualism.

As expected, most but not all of our symbiotic mutations showed reduced nitrogen fixation and/or nodule investment during multi-strain inoculations similar to those observed in prior single-strain studies. Because we conducted our experiments with N-free fertilizer, we inferred nitrogen fixation efficacy in a multi-strain context by placing observed mutant biomass and root-to-shoot ratios on a continuum from no nitrogen fixation (*dmi1-2, dmi2-3, and dmi3-1*) to wild-type levels (*hcl*). Exceptions included *dnf1* and *dnf4,* which exhibited nitrogen fixation during multi-strain inoculations despite failing to do so in single-strain experiments suggesting that they can successfully form symbiosis with some strains (42,43). Partial overriding of single strain phenotypes, in terms of host investment and host benefits, have been previously reported in Δ*npd1-5* mutants, which are non-fixing with *Sinorhizobium* strain Sm1021, but regain 30% nitrogen fixation during multi-strain inoculation, highlighting strain-specificity (17). Another host mutant that did not exhibit the expected phenotype during multi-strain inoculations was the *hcl* mutant, which, during single-strain inoculation with the strain Sm2011, formed significantly fewer nodules, but during community inoculation, accumulated biomass similar to the WT with comparable nodulation (23). Host mutation-dependent shifts in rhizobial entry mode have been previously noted with gain of function *snf1* mutants *of Lotus japonicus* (encoding a T265I calcium calmodulin-dependent kinase CCaMK *),* which cause a shift from complex infection thread mode to a simpler ‘crack entry’ mode of rhizobial entry (44). It would be interesting to probe how rhizobia navigate the *hcl* mutation—indeed, our GWA reveals that all top variants on all three replicons are unique to the *hcl* mutation. The host mutations that deviate from their

single-strain phenotypes during multi-strain inoculations create additional opportunities for discovering novel rhizobia genes that ensure successful symbiosis in these hosts, which are known to be poor or non-fixing with model *Sinorhizobium* strains.

Legumes releasing rhizobia from their senescing nodules into the soil, directly contribute to the soil rhizobial population (45). Identifying host genes, like the *npd* family that amplify this effect can lead to targeted soil microbiome interventions (17). A legume can increase the overall amount of rhizobia released via two routes. First, they can release more rhizobia by increasing nodule numbers—a pattern identified for two AON genes (*sunn1*, *sunn4)* and two late-stage genes *(dnf4* and *dnf6*)—suggesting that these genes, when functional, reduce rhizobia released by limiting nodule formation. This hypothesis seems plausible particularly for *sunn1(***Su**per **N**umeric **N**odules1) and *sunn4 (***Su**per **N**umeric **N**odules4*)* which are known to control the degree of nodulation, preventing the legume from overinvesting when nitrogen needs are met (26). Alternatively, plants can have comparable nodulation but release more rhizobia per nodule, as noted previously in *npd2*, which produced fewer nodules than WT during multi-strain inoculations but released more rhizobia per nodule than WT (17). In contrast, here we found most mutations had no effect on the number of rhizobia released from each nodule, and *ipd3, latd* and *nad1* actually reduced the population size released. Thus, functional copies of these genes support larger populations of rhizobia per nodule—indeed, disrupting *nad1* leads to overactive plant defence responses in nodules, likely reducing rhizobial survival (32). Overall, our results suggest that the effects of host genetic mutations on rhizobial fitness are gene-specific and may differ from wild-type expectations (e.g., bigger nodules release more rhizobia). Increasing the number of rhizobia the host supports represents one pathway to nurture robust populations of beneficial microbes, and identifying the host genes that cause this amplification

sets the stage for targeted interventions aimed at bolstering soil rhizobial populations in agricultural and conservation contexts (46)

Similar to our previous work on the *npd* genes (17), we also found subtle but significant relative fitness shifts in specific strains within host mutants. For instance, while late-acting NCRs are often cited as major determinants of host strain specificity in *M. truncatula*, our investigation reveals a more complex landscape (6,47). Two NCR mutations had differing effects on strain fitness within nodules. The *dnf7* mutant, lacking functional NCR169 expressed in the N-fixation zone (48), grouped with the *ipd3* mutant. Surprisingly, *dnf4*, which does not produce NCR211 and influences rhizobial differentiation during bacteroid formation (43), showed similar effects on strain fitness as *dnf3* and *dnf6*. There are a few ways, by which rhizobia are known to overcome the inhibitory actions of the NCRs however none of these genes came up as top candidates in our GWA suggesting additional mechanisms of host restriction of rhizobia. These initial observations of NCR mutations further motivate exploring the functional roles and mechanisms of the remaining ~600 NCRs expressed in nodules (49,50), especially since many NCRs are differentially expressed in hosts depending on which strain inhabits the nodules and thus could shape the rhizobial selective environment (51).

More broadly, though, our study shows that many other symbiotic genes beyond NCR's, including AON genes, can significantly alter which rhizobial strain each plant enriches in mixed inoculations. Mutations in host genes influencing different stages of symbiosis enriched similar strains. This highlights the complex interplay between host genetics and microbial partners. Activation of plant defence responses in nodules can also lead to differential strain fitness (32). Given that *dnf2* and *nad1* both display elevated defence responses in nodules, all three genotypes

with mutations in *dnf2* and *nad1* enriched similar strains (9). The strain fitness pattern in *dnf1* hosts was most similar to AON mutants, *dnf3, dnf4*, and *dnf6*. Although there are no known overlapping roles for these mutants, their similarities may arise from functional nitrogen fixation, allowing the host to reward some strains over others. Lastly, the genetic basis of some symbiosis genes (e.g., *dnf3* and *dnf6*) remains a mystery-examination of candidate genes from our GWA could provide insights (6).

Natural variation in symbiosis genes arises and segregates within legume populations, potentially contributing to variation in selection on rhizobia (52,53). To measure the effect of this host genetic variation, we explored how mutations in these genes affect rhizobial strain survival and fitness, specifically asking if the same rhizobial genes were important for rhizobial fitness across different host mutations or if unique host mutations create distinct selective pressures on rhizobia. Our GWAS analyses identified rhizobial genes underlying fitness shifts in different host mutations, revealing genes with both pervasive and limited effects. Variant groups with widespread effects were primarily located on pSymA and pSymB. Interestingly, alternative versions of these alleles increased fitness across many host mutations but strongly reduced fitness in the wild-type. This suggests a potential cost to harbouring these variants, raising the question of why they persist. These variations may be beneficial in other host accessions or species (15). We also found many variants important in only one or a few host mutant backgrounds, with the strongest associations on pSymB and, to some extent, pSymA. Unlike the pervasive variants, most limited effect variants had little or no effect on strain fitness in WT backgrounds, a pattern consistent with conditional neutrality (54–56). Together the data suggests that the alternate versions of the pervasive genes might be maintained through balancing

selection due to host variation, while the unique genes may increase in frequency through neutral processes and incidentally enhance fitness when specific symbiotic genes are disrupted.

Most candidate genes with pervasive effects are highly expressed in the nitrogen fixation zone (Zone III), emphasizing the need for downstream validation assays to elucidate their role in nitrogen fixation (57). For example, the *fixL* gene, an oxygen sensor, regulates oxygen levels during nitrogen fixation, Malonyl-CoA synthase, induced selectively in the rhizospheres of alfalfa, is responsible for malonate catabolism within root nodules (58), and a NAD(P)H-binding protein from the conserved NmrA domain family shows high similarity to a transcriptional regulator, involved in nitrogen metabolite repression in fungi (59). These genes represent low-hanging fruit for uncovering novel mechanisms to broaden the host range of symbionts, aiming to extend nitrogen fixation beyond legumes (60). Beyond these candidates, most genes only altered strain fitness in specific symbiotic perturbations, indicating that each perturbation creates a unique challenge for rhizobia survival and reproduction. Further experimentation will uncover these novel mechanisms of strain survival and elucidate host by rhizobia genome by genome interactions. 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 variants in *stc4*, encoding stachydrine N-demethylase reductase, enhancing strain fitness in *dnf4* mutants, possibly by protecting differentiating rhizobia (43). Our results highlight the potential of leveraging natural genetic variation to explore undiscovered mechanisms in genome-by-genome interactions between hosts and microbes.

Conclusion:

By identifying how genetic mutations affect symbiosis in multi-strain inoculations, we aimed to understand the forces shaping rhizobial and legume evolution in ecologically realistic contexts (13, 30). Measuring the effects of host genetic variation on microbial fitness is challenging but crucial for understanding the evolution of and maximizing the benefits of root nodule symbiosis. By re-examining root nodule symbiosis in multi-strain contexts and using a GWA approach for leveraging natural genetic variation, we identified novel rhizobial genetic factors influencing strain fitness shifts, laying the groundwork for further investigations of legume genome by rhizobia genome interactions (30,36). Validating these rhizobia candidates in different contexts will help fine-tune existing mutualism and potentially engineer new associations between rhizobia and novel hosts.

Acknowledgments

We thank Prof. Dong Wang of University of Massachusetts, Amherst for generously providing the *M. truncatula* mutant seeds, Lillian Cherry, Maria Alejandra Gil-Polo, and Jennifer E. Harris for help with the tedious task of picking and processing nodules. Prof. Katy Heath of University) of Illinois Urbana- Champaign and Dr. Mike Grillo (currently Assistant Professor in Loyola University, Chicago) for providing strains and sequencing data. This work was supported by the National Science Foundation grant IOS-1856744 to PT, NY, and LTB and IOS-2243819 to PT and LTB and was completed using computing resources provided by Penn State Institute for Computational and Data Science and sequencing by the Huck Genomics Core Facility. LTB's work was also supported by HATCH grant no. 1025611. The authors acknowledge the Huck Institutes' Genomics Core Facility (RRID:SCR_023645) for use the Illumina NextSeq 2000.

Authorship Credits

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

Competing Interests:

The authors declare no competing interests.

Data Accessibility

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

References

1. Oldroyd GED, Murray JD, Poole PS, Downie JA. The Rules of Engagement in the Legume-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.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

33. Riley AB, Grillo MA, Epstein B, Tiffin P, Heath KD. Discordant population structure among rhizobium divided genomes and their legume hosts. Mol Ecol. 2023 May;32(10):2646–59.

34. Nelson M, Guhlin J, Epstein B, Tiffin P, Sadowsky MJ. The complete replicons of 16 Ensifer meliloti strains offer insights into intra- and inter-replicon gene transfer, transposon-associated loci, and repeat elements. Microb Genomics [Internet]. 2018 May 1 [cited 2024 Sep 24];4(5).

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Figure Legends

Fig.1: Experimental design schematic. We grew 86 sequenced *Sinorhizobium meliloti* strains in liquid media and combined them in approximately equal proportions to generate a multi-strain community. We used this community to inoculate the *M.truncatula* wildtype (A17) and 18 mutant hosts. We measured the traits which represent common proxies for host investment and benefits. In addition, we measured rhizobial benefits including the population of rhizobia released per nodule and per plant and applied whole genome shotgun sequencing to determine the change in the relative fitness of the rhizobial strains prior to and after exposure to the host mutants.

Fig.2: **Host symbiotic mutations reduce host benefits and differentially impact host investments even in mixed inoculations**. Hosts benefit traits like (A) total biomass per plant and (B) root-to-shoot ratio were reduced and increased respectively in almost all the host mutants except *sunn4*. The host investment traits which include (C) total nodule area per plant, (D) nodules per plant, and (E) average area per nodule were differentially affected by the host mutants. Up/down triangles denote when a host mutant significantly increases/decreases a trait with respect to WT. The boxplots are colour coded based on the stage of symbiosis (disruption type) to which the individual host mutants belong (See Table S3 for details).

Fig.3: Mutating host symbiosis genes affect rhizobial benefits. Host mutations differentially influenced rhizobial benefit traits like (A) CFU/plant and (B) Cfu/nodule while (C) nodule isolate diversity (C) was largely unaffected by the majority of host mutants. Up/down triangles denote when a host mutant significantly decreases/increases a trait with respect to WT

respectively. The boxplots are colour coded based on the stage of symbiosis (Mutation type) to which the individual host mutants belong (see Table S3 for details).

Fig.4: Heatmap visualizing variation in strain fitness in pools of wildtype and mutant nodules. The host mutants (left) are hierarchically clustered based on the similarity of median strain fitness patterns (Ward D method). The strain fitness are ordered (bottom) by strain fitness in A17 from high relative fitness (red) to low (blue). Red indicates strain enrichment relative to the strain frequency in WT, white indicates no fitness change, and blue indicates depletion relative to WT.

Fig.5: Rhizobial candidate genes underlying strain fitness shifts predominantly occur on pSymA and pSymB. (A) Manhattan plot indicating SNP(single nucleotide polymorphism) positions in *S. meliloti* USDA1106 with strong statistical association with strain fitness shifts across the 14 host mutants when compared to wildtype A17 hosts. The dotted line indicates the P value threshold used for identifying candidates genes and for clarity only significant associations are color-coded according to the host gene mutation. GWAs were conducted separately for each replicon to account for divergent population structure, but are merged here for ease of comparison. (B) Heatmaps based on hierarchical clustering of the inverse P-value ranks of the top five variant groups in each of the 14 host mutations. The heatmap shows how the P value ranks of the variant groups change across the spectrum of the host mutations.

Tables

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

*CFU= Colony Forming Units

Table 2: Comparison of the variance in host and rhizobial traits explained by individual 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

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

Table 3: Candidate rhizobial genes in small linkage groups (LD) (three genes or less) associated 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

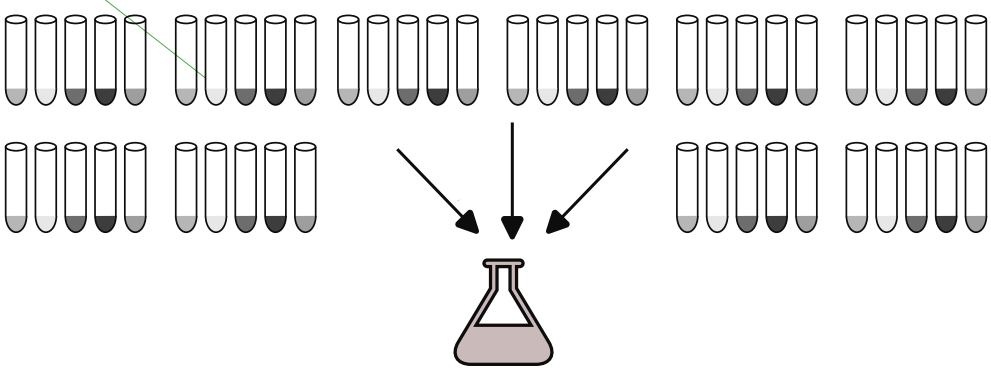
Notes: [1]LD group numbers were obtained from (15); [2]AF= non-reference (alternate) allele frequency. [3]Locus tags are from the *Sinorhizobium meliloti* USDA1106 reference annotation available on MAGE (http://www.genoscope.cns.fr/agc/microscope/home/index.php). [4]A tri allelic event

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

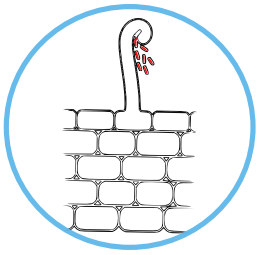
4. Measure nine symbiotic traits after eight weeks



WILDTYPE

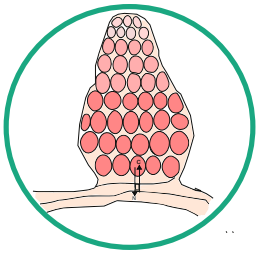
A17

EARLY



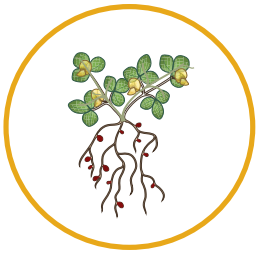
hcl, ipd3, dmi1-2, dmi2-3, dmi3-1, latd

LATE



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

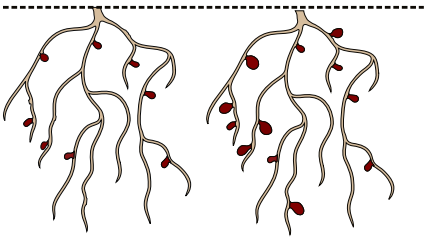
AON



sunn1, sunn4, rdn1

Host Investments

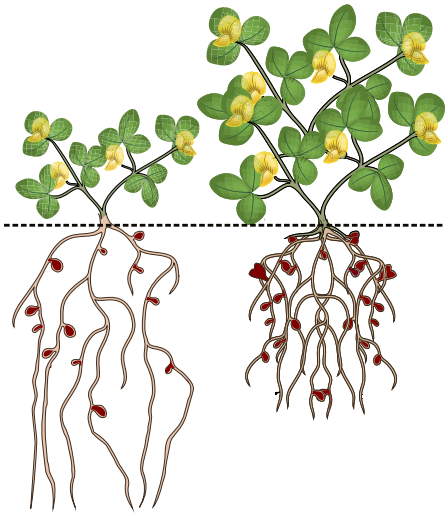
Low → High



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

Host Benefits

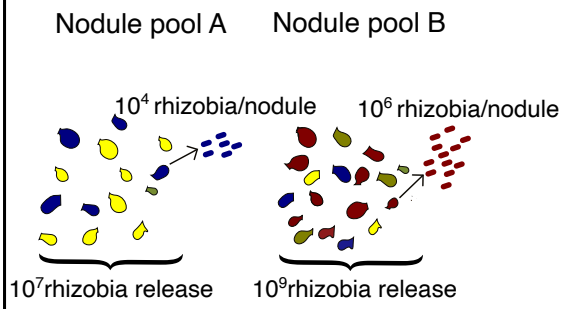
Low → High



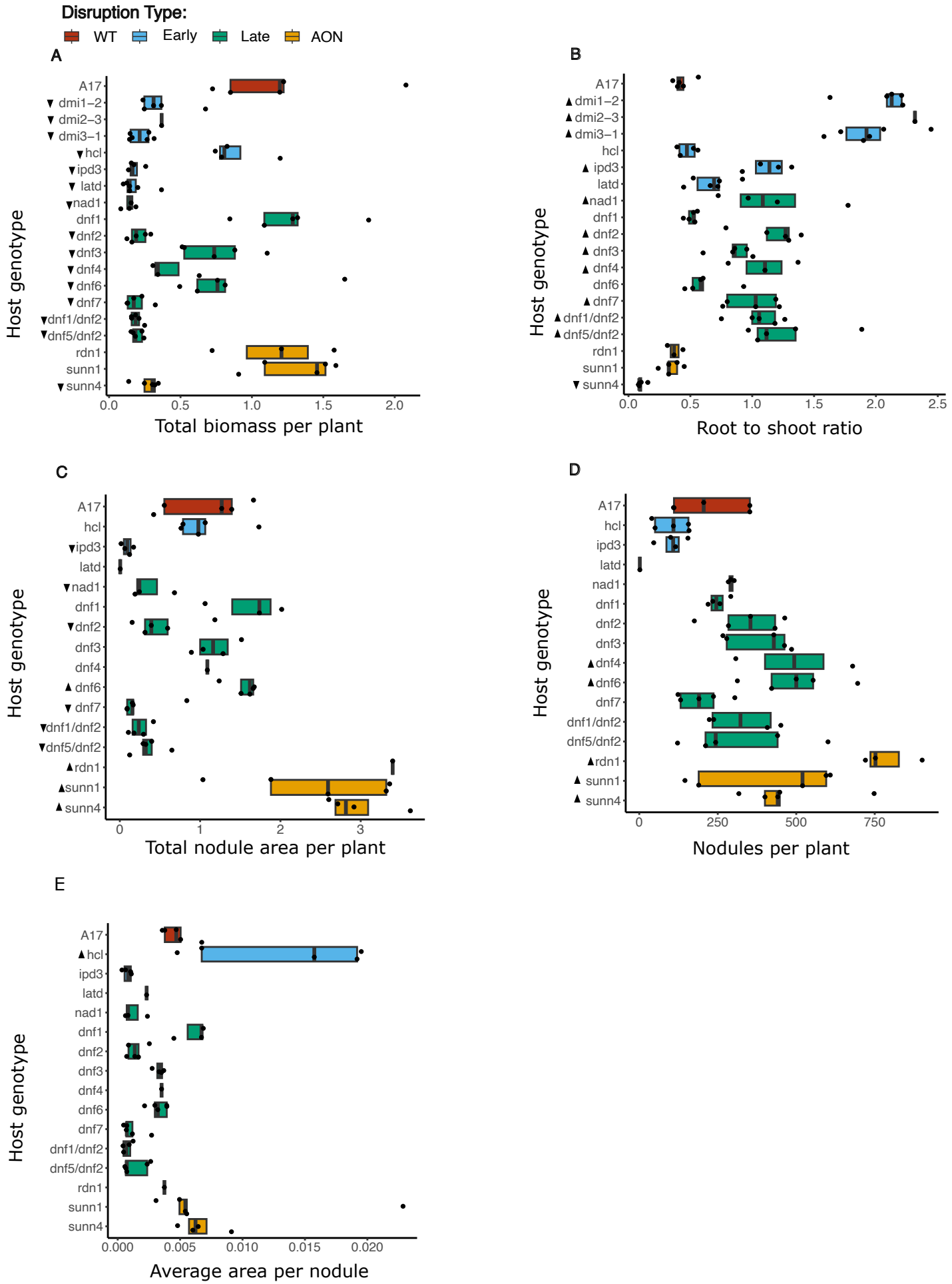
- Total biomass per plant
- Root -to-shoot ratio

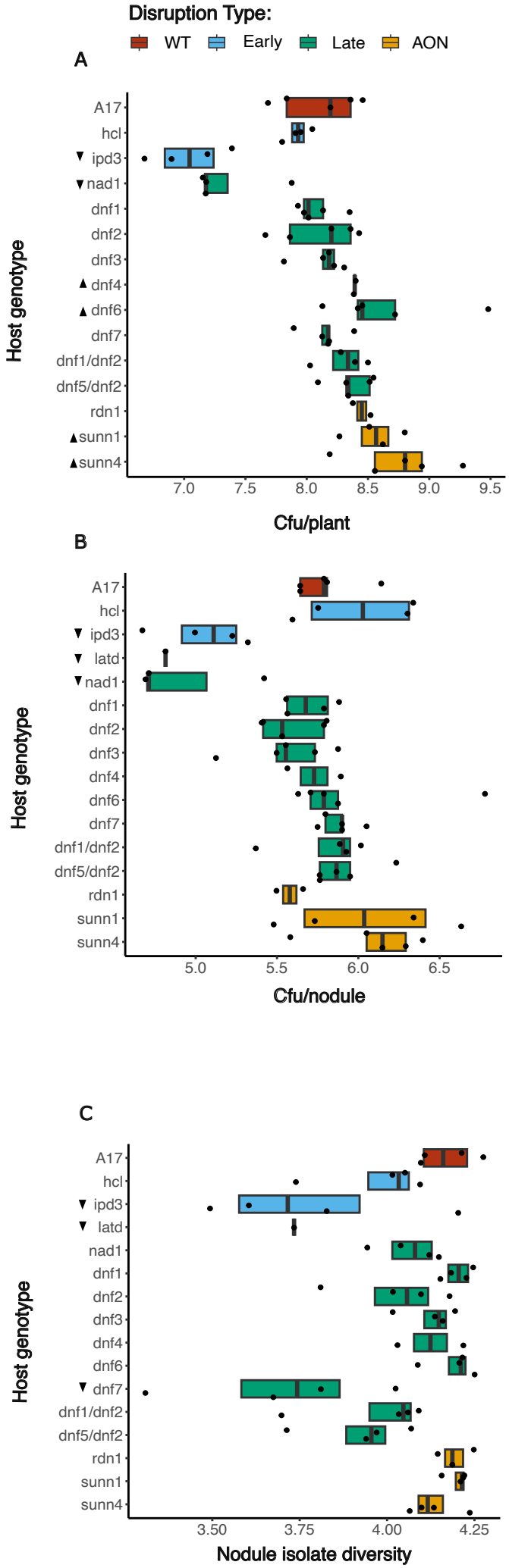
Rhizobial Benefits

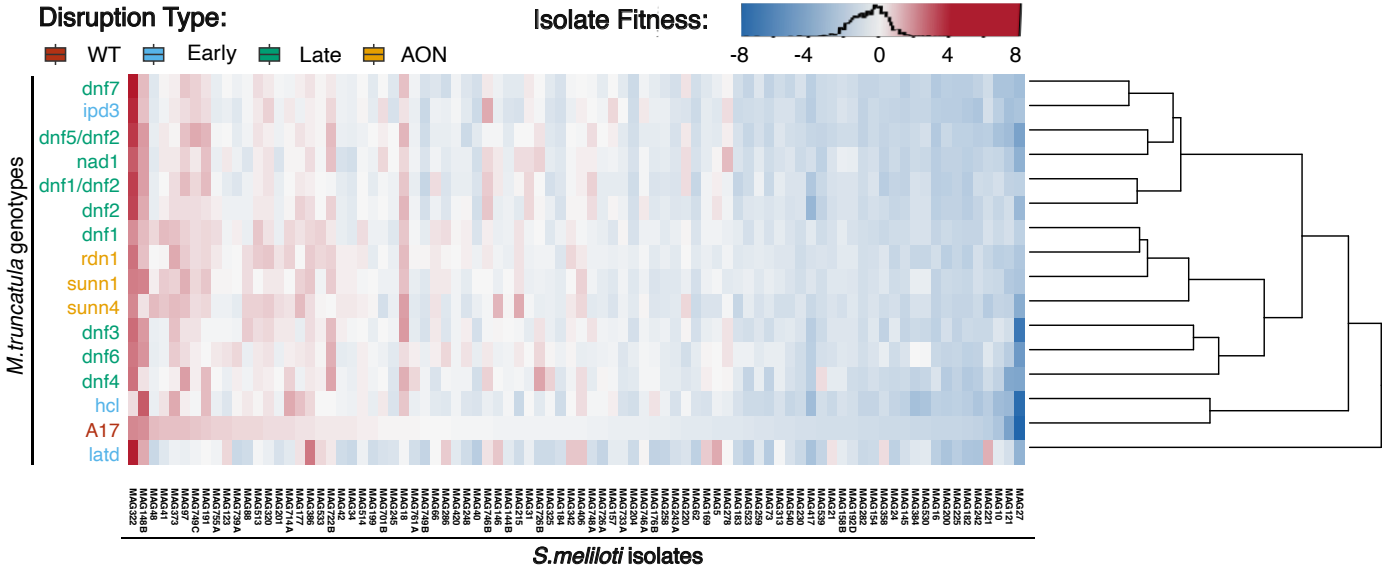
Low → High

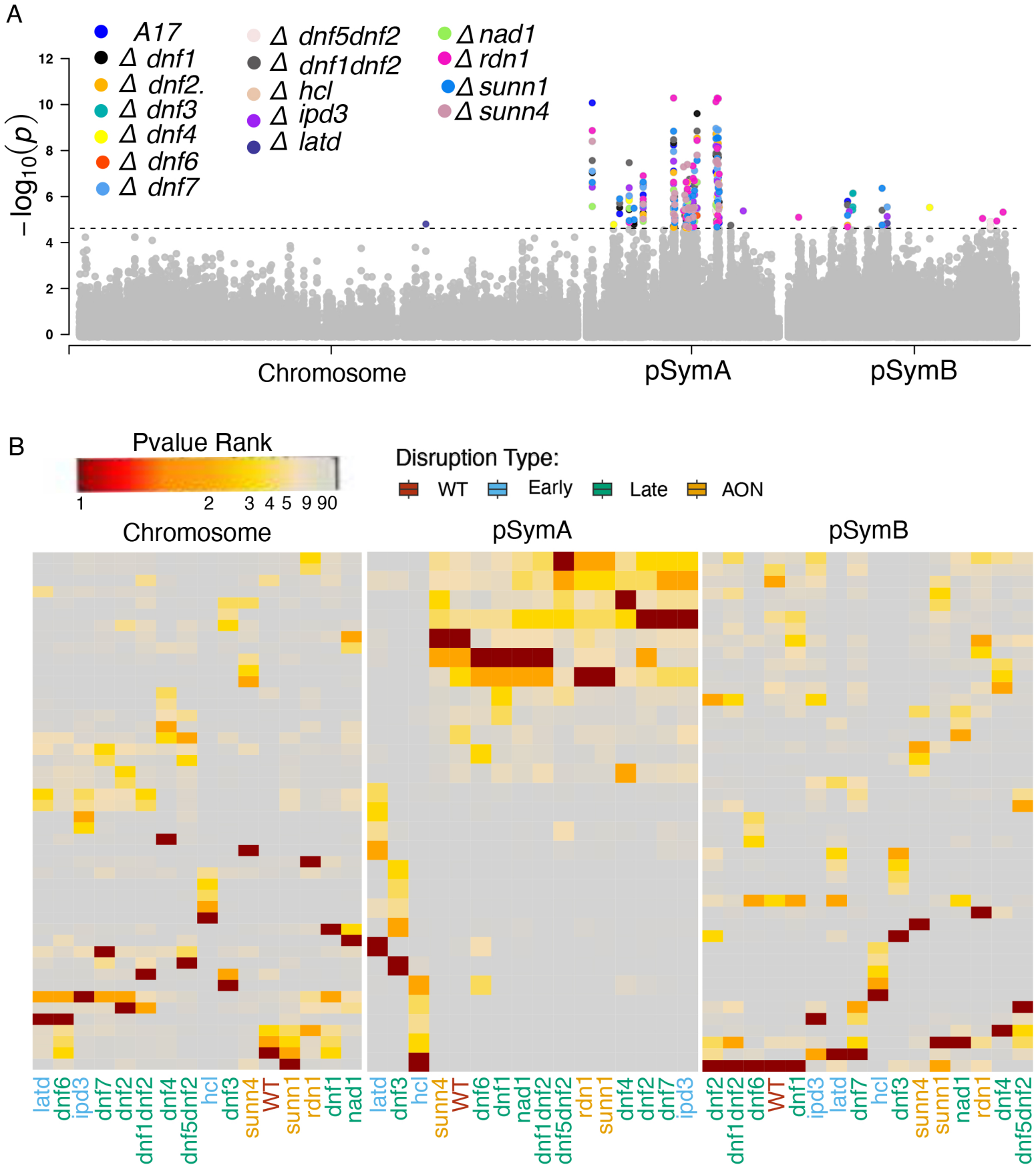


- Rhizobia released per nodule
- Rhizobia released per plant
- Nodule isolate diversity
- Nodule isolate fitness


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


Disruption Type:
WT
Early
Late
AON
A
Host genotype
A17
hcl
ipd3
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
7.0
7.5
8.0
8.5
9.0
9.5
Cfu/plant
B
Host genotype
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
5.0
5.5
6.0
6.5
Cfu/nodule
C
Host genotype
A17
hcl
ipd3
latd
nad1
dnf1
dnf2
dnf3
dnf4
dnf6
dnf7
dnf1/dnf2
dnf5/dnf2
rdn1
sunn1
sunn4
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 genotypes
dnf7
ipd3
dnf5/dnf2
nad1
dnf1/dnf2
dnf2
dnf1
rdn1
sunn1
sunn4
dnf3
dnf6
dnf4
hcl
A17
latd
S.meliloti isolates


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
Disruption Type:
WT
Early
Late
AON
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