The genetic basis of cooperation and conflict in natural populations of a model symbiont

Rebecca T. Batstone[a,1], **Liana T. Burghardt**[b], **and Katy D. Heath**[a,c]

[a]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801 USA; [b]Department of Plant Science, The Pennsylvania State University, 103 Tyson Building, University Park, PA, 16802 USA; [c]Department of Plant Biology, University of Illinois at Urbana-Champaign, 286 Morrill Hall, 505 South Goodwin Avenue, Urbana, IL 61801 USA

This manuscript was compiled on September 29, 2021

Although mutualisms are defined as net beneficial interactions among species, whether fitness conflict or alignment drive the evolution of these interactions is unclear. Examining the relationships between host and symbiont fitness proxies at both the organismal and genomic levels can illuminate the mechanisms underlying cooperation and conflict, while addressing the challenge of measuring microbial symbiont fitness. Here, we utilized data from several genome-wide association studies (GWAS) that involved 191 strains of the N-fixing rhizobium symbiont, *Ensifer meliloti*, collected from natural populations. Each strain was paired with one of two host *Medicago truncatula* genotypes in single- or mixed-strain inoculation experiments to determine how different proxies of symbiont fitness and host benefit were related to one another, and examine signatures of fitness conflict and alignment across scales. We found little evidence for fitness conflict; instead, loci tended to have concordant effects on both host and symbiont fitness and showed heightened nucleotide diversity and signatures of balancing selection compared to the rest of the genome. We additionally found that single versus competitive measures of rhizobium fitness are distinct, and that the latter should be used because they better reflect the ecological conditions rhizobia experience in nature. Our results suggest that although conflict appears to be largely resolved in natural populations of rhizobia, mutualistic coevolution between legumes and rhizobia can nonetheless maintain genetic diversity, potentially explaining why variation in symbiotic traits persists in nature.

conflict | cooperation | GWAS | mutualism | pleiotropy

A central goal of genetics research is to build a reliable map of genotype-to-phenotype relationships (G→P map) in order to uncover the genetic basis of important organismal phenotypes, including plant yield in agriculture or disease-related traits in humans (1–3). Recent advances in sequencing and microbiome research have revealed the ubiquity of symbioses, intimate interactions among species (*sensu* de Bary, 1878), meaning that many of these 'macrobial' host phenotypes (yield, disease resistance, et c.) are actually symbiotic extended phenotypes, influenced at least in part by loci present within interacting microbial symbionts in addition to the host (4–9). Understanding the genetic basis of symbiotic extended phenotypes and predicting how these complex traits evolve in nature therefore requires explicitly accounting for the intimate interaction of multiple genomes.

Studies of the G→P map in single organisms have highlighted the importance of pleiotropy, i.e., genes or mutations that influence more than one trait, with important consequences for development, aging, and disease (10, 11). In single organisms, pleiotropy can constrain adaptation when it underlies trade-offs among fitness-related traits (12–15). In the G→P map of symbiotic extended phenotypes, pleiotropic loci can also underlie traits of interacting individuals, which we henceforth refer to as symbiotic pleiotropy. In symbiotic mutualisms, wherein partners trade fitness benefits (16), symbiotic pleiotropy can couple the traits that benefit the individual (private traits) with traits that benefit the partner (cooperative traits), promoting cooperation (17–20). The directionality of symbiotic pleiotropy is critical because it determines the degree to which partners' fitnesses are aligned (i.e., same-sign or concordant effects) or in conflict (i.e., opposite-sign or discordant effects), and thus, defines the mutational landscape of mutualism evolution in nature. If symbiotic pleiotropy is extensive and its effects on fitness-related traits in the interacting partners tend to be discordant (i.e., increases in one partner's fitness come at the expense of the other partner), then conflict should underlie the evolution of mutualism, allowing for the possibility of cheating individuals that are competitively superior, as mutualism theory predicts (19–24). On the other hand, if pleiotropic effects are overwhelmingly concordant (i.e., effects on host and symbiont fitnesses are in the same direction), fitness alignment rather than conflict should be the null hypothesis in mutualism, and models relying on cheating genotypes as the main driver of mutualism evolution may not be suitable.

Genome-wide association studies (i.e., GWAS) can be used to reveal the genes, as well as specific variants, that underlie variation in mutualism traits in natural populations (25–31).

Significance Statement

Given the need to predict the outcomes of evolution in host-associated microbiomes, whether microbial and host fitnesses tend to trade-off remains a pressing question. We examined naturally-occurring genetic variation in 191 strains of the model microbial symbiont, *Ensifer meliloti*, and found that loci tended to have concordant effects on both microbe and host fitnesses, even in environments with multiple competing strains. Our results suggest that fitness alignment acts to maintain, rather than erode, genetic diversity in microbes, likely because the identities of concordant loci depend on the host genotype with which microbes interact. Our results additionally emphasize that quantifying measures of microbial competitive fitness is key for understanding how the microbiome can be harnessed to improve host fitness.

R.T.B., K.D.H, and L.T.B conceived of the work. R.T.B led the data analysis and drafted the article with guidance from K.D.H. and L.T.B., who also provided critical revisions and approved the final version for submission.

The authors declare no competing interests.

[1]To whom correspondence should be addressed. E-mail: rtbatstone@gmail.com

DRAFT

Because they provide an estimate of both the strength and direction of effects of particular alleles on the trait, GWAS are especially useful for identifying loci underlying symbiotic pleiotropy and facilitating the study of how historical selection has acted on those loci in the past (9, 32). For example, if loci underlying fitness conflict resemble those underlying antagonistic arms race dynamics, they might show signatures of rapid turnover and higher rates of evolutionary change as rare alleles are selectively favoured (33, 34), whereas loci underlying fitness alignment may exhibit signatures of evolutionary stability and lower rates of evolutionary change as more common alleles are selectively favoured (34).

Host-symbiont conflict may be measured at both the organismal-level (i.e., do low-quality individuals derive higher fitness from symbiosis?) or at the genomic-level (i.e., do alleles that increase one partner's fitness decrease the other partner's fitness?). Results at these two levels can differ; for instance, fitness alignment between the sexes prevails at the phenotypic-level, but sexually-antagonistic loci nonetheless exist and can have important consequences on the coevolution of sexual traits (35). While organismal-level genetic correlations reflect the action of all contributing alleles as they are packaged into genotypes, GWAS has the power to reveal individual allelic effects "unpackaged" from those multilocus genotypes - packaging in this case being due to linkage disequilibrium among genotypes arising from various processes including physical linkage, historical epistatic selection, or even neutral drift/population structure (36, 37). Thus GWAS can reveal variants underlying fundamental sources of conflict at the genomic-level even if fitness alignment is realized at the organismal-level.

Over the past ∼25 years, legume-rhizobium symbioses have been developed as models for understanding mutualism evolution (38–42). This interaction is one of the most ecologically and economically important symbioses, contributing upwards of 20 million tonnes of nitrogen (N) to the global N-cycle, and saving billions of dollars that would have otherwise been spent on synthetic N fertilizer production (43). Legumes house rhizobia within specialized root nodules and supply them with photosynthate, while rhizobia within the nodules convert atmospheric N to a plant-usable form, resulting in a beneficial exchange of resources. Key traits of this symbiosis, such as plant biomass and nodule number, are symbiotic extended phenotypes known to be influenced by variants in the genomes of both host and symbionts, as well as the epistatic interactions between them (Genotype-by-Genotype or G x G interactions; e.g., 4, 28, 38, 44–47).

Addressing patterns of conflict versus alignment over symbiotic traits requires robust proxies of fitness. Seed number, or close correlates such as plant biomass and leaf chlorophyll A, are well-established proxies for reproductive fitness in annual plants (25, 48, 49). Estimating rhizobium fitness has been an empirical and conceptual challenge, and thus has received much attention in the literature (39, 50–54). Nodule number and nodule size have both been used as fitness proxies because a rhizobium that establishes a nodule in symbiosis can gain a fitness benefit on the order of $10^5 - 10^7$ (55), larger nodules release more rhizobia (49, 50, 56–58), and intuitively strains that produce more nodules will release more rhizobia (4, 42, 58). While single inoculation experiments are required for simultaneously estimating genotypic means for both partners, they cannot address competition among strains in real populations. Until recently, competition could be studied with only a few strains, often using isogenic mutants (e.g., 42, 48, 59). Recent advances merging population genomics and metagenomics allow us to measure a strain's ability to compete for nodulation opportunities and for host investment in nodule growth (27, 60) and hint at a surprising lack of correlation between nodule number and competitive fitness, motivating our comprehensive analysis here.

A quantitative comparison of single-strain and multi-strain measures of rhizobium fitness, as well as their correlations with host fitness, is needed to decouple individual strain fitness from host growth and potentially reveal interspecies conflict. Early attempts to address conflict between legumes and rhizobia relied on single-strain inoculation experiments; here, negative correlations between rhizobium partner quality and rhizobium fitness (i.e., nodule number) were interpreted as evidence for conflict between host and symbiont fitnesses (4, 48, 58). In contrast, mutant and experimental evolution studies have suggested a lack of rhizobial mutations underlying conflict (31, 52). However, these experiments have been criticized for producing spurious positive correlations due to stronger fitness feedback between host plants and rhizobia in single strain inoculation, versus when multiple competing strains are present (53), particularly because competition is likely a critical component of fitness in natural and agricultural soils where many rhizobial strains coexist. Thus detecting conflict between hosts and rhizobia might require multi-strain inoculation experiments wherein the benefits to individual strains in a mixed population can be decoupled from overall plant growth.

Here we combine two datasets: first, phenotypic variation and underlying allelic effects from a GWAS of plant and rhizobium fitness proxies measured in single-strain inoculation across 191 strains of *Ensifer meliloti* from natural populations on each of two host genotypes, with a second dataset measuring competitive fitness of 89 of these strains on each of the two host genotypes. We combine these datasets to first ask: What are the relationships among rhizobium fitness components measured in both single- and mixed-strain inoculation experiments? Are these traits genetically distinct and thus likely to evolve independently, or are these linked through pleiotropy? We use this information to address the potential for genomic conflict in this symbiosis, asking: Do segregating mutations tend to have aligned or conflicting effects on host and symbiont fitnesses? Finally we ask whether there is any evidence that historical selection has differentially shaped alleles underlying alignment versus conflict, as we might predict under different models of mutualism evolution.

Results

Correlations amongst rhizobium fitness proxies differed between proxy combinations and host genotypes. Trade-offs at the organismal level, plus underlying antagonistic pleiotropy at the genomic level, suggest that the genetic mechanisms rhizobia use to form more nodules in single inoculation make them less able to proliferate within and compete for nodules in mixed inoculation (**Fig. 1**; see **SI Fig. S1** for all genetic correlations and **SI Fig. S2** for all pleiotropic variant counts). For some proxies, but not others, these patterns depended on the host genotype with which rhizobia interacted. First, in

DRAFT

single inoculation, when we regressed strain means for nodule number and nodule weight, we found significant negative correlations for both hosts (**Fig. 1A** left), indicating that strains creating larger nodules tended to form fewer total nodules on both host genotypes. At the genomic-level, most significant and all pleiotropic variants had discordant effects (**Fig. 1A** right), similarly indicating a trade-off whereby variants that are positively associated with nodule number tend to be negatively associated with nodule weight, or vice-versa.

Comparing single and mixed inoculation environments, when we regressed strain means for competitive fitness and nodule number, we found a weak but significant negative relationship for host line DZA only (**Fig. 1B** left), indicating that strains that were more common in mixed inoculation formed fewer nodules in single inoculation. At the genomic-level on DZA, most variants had discordant effects (**Fig. 1B** right), again indicating a trade-off between proxies at this scale. For host line A17, most pleiotropic variants also had discordant effects (**Fig. 1B** right), despite no relationship between these fitness proxies at the organismal-level (**Fig. 1B** right).

We found a significant, positive relationship between strain means for competitive fitness and nodule weight, again for DZA only (**Fig. 1C** left). This result indicates that strains that were more commonly found in nodules in mixed inoculation formed larger nodules in single inoculation. At the genomic-level, variant effects also tended to be concordant (**Fig. 1C** right), meaning that variants having positive or negative effects on nodule weight also tended to have consistent (i.e., positive or negative) effects on competitive fitness. For A17, while we found a lack of significant genetic correlation for these fitness proxies (**Fig. 1C** left), at the genomic level, a greater proportion of significant and all pleiotropic variants had discordant effects on nodule weight and competitive fitness (**Fig. 1C** right), suggesting a potential trade-off between these traits at the genomic-level. However, we found fewer total pleiotropic variants underlying the relationship between competitive fitness and nodule weight (**Fig. 1C**, right; see **SI Fig. S2** for pleiotropic variant counts) compared to the other two proxy pairs, suggesting that these proxies are largely governed by different molecular mechanisms and likely evolve independently, especially on A17.

Alignment between host fitness and rhizobial fitness proxies. We found mostly concordant relationships between competitive fitness and plant fitness at both the organismal and genomic levels (**Fig. 2**; see **SI Fig. S1** for all genetic correlations and **SI Fig. S2** for all pleiotropic variant counts), suggesting that conflict among these traits is largely resolved in natural rhizobium populations. We found no correlation, in either host, between strain means for nodule number and shoot biomass (**Fig. 2A** left), or nodule number and leaf chlorophyll A content on A17 (**Fig. 2D** left). However we saw a significantly negative correlation between nodule number and chlorophyll A for DZA (**Fig. 2D** left). Assuming chlorophyll A indicates the N-fixation efficiency of each strain (61–63), this suggests that strains forming more numerous (and smaller) nodules tended to fix less N. At the genomic-level, most variants had concordant effects for nodule number and shoot biomass on DZA, whereas the opposite was true for the same trait pair on A17 (**Fig. 2A** right). Most variants had discordant effects on nodule number and chlorophyll A content on both hosts as well (**Fig. 2D** right).

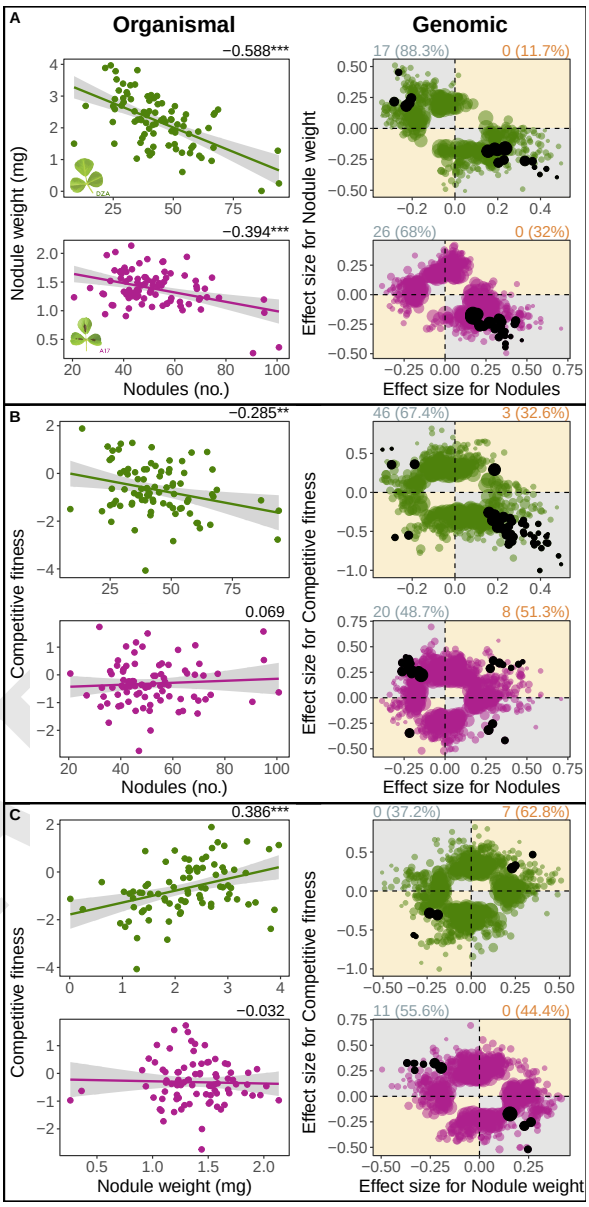

Fig. 1. Trade-offs among rhizobia (*Ensifer meliloti*) fitness proxies prevail at both the organismal (left panels) and genomic-levels (right panels). Organismal: Genetic correlations between pairwise fitness proxies correspond to traits measured on plant lines DZA (green, top rows) or A17 (pink, bottom rows), based on 89 *Ensifer meliloti* strains. Dots represent estimated marginal means for nodule number and nodule weight, both being measured in single inoculation experiments, or medians for competitive fitness measured in mixed-inoculation experiments. Numbers at top right of each correlation represent Pearson correlation coefficients, while asterisks represent significance: * = $p < 0.05$; ** = $p < 0.01$; *** = $p < 0.001$. **Genomic**: Dots represent variant effect sizes (i.e., beta values calculated in GEMMA), whereby effects were regressed for each pairwise trait combination. Dots falling along the diagonal (orange quadrants) represent variants with concordant (i.e., same-sign) effects, while those falling on the off-diagonal (grey quadrants) represent variants with discordant (i.e., opposite-sign) effects. Coloured dots represent variants that were significantly associated with one of the two proxies, while black dots represent pleiotropic variants, i.e., significantly associated with both proxies. Numbers outside of and percentages within parentheses at the top left and right of each plot represent the pleiotropic variant counts or proportion of total significant variants, respectively, that are discordant (left, in grey) or concordant (right, in orange).

Strain means for nodule weight and shoot biomass were significantly positively correlated for both hosts (**Fig. 2B** left). Nodule weight and chlorophyll A were significantly positively correlated on DZA, but not on A17 (**Fig. 2E** left). At the genomic-level, most significant and all pleiotropic variants had concordant effects for nodule weight and shoot biomass (**Fig. 2B** right), as well as nodule weight and chlorophyll A (**Fig. 2E** right), on both hosts.

Finally, competitive fitness and shoot biomass were significantly positively correlated for DZA but not for A17 (**Fig. 2C** left). In contrast, competitive fitness and chlorophyll A content were significantly positive on both hosts (**Fig. 2F** left). At the genomic-level, most variants had concordant effects on competitive fitness and shoot biomass (**Fig. 2C** right), and this pattern was even stronger between competitive fitness and chlorophyll A on DZA with all but one pleiotropic variant having concordant effects (**Fig. 2F** right). In contrast, variant effects were equally mixed on A17 for competitive fitness and chlorophyll A (**Fig. 2F** right).

Selection acts differently on genes underlying alignment versus conflict. Our GWAS approach not only allows us to identify the genes associated with fitness conflict or alignment, but to then go a step further and test whether these focal genes show signatures of selection or other processes that would cause patterns of molecular evolution to deviate significantly from neutrality. We *a priori* expected genes underlying fitness conflict to be higher in diversity; instead, our results using multiple diversity and neutrality metrics suggest the opposite – that rhizobium genes underlying alignment show stronger signatures of balancing selection. Specifically, genes underlying concordant symbiotic pleiotropy (i.e., solid orange lines in **Fig. 3**) had significantly elevated statistics relative to the null (i.e., distributions, based on all genes containing significant variants based on GWAS), for three of the four test statistics (π, p = 0.039; Fu and Li's F, p = 0.056; Fu and Li's D, p = 0.001), suggesting that genes underlying fitness alignment tend to be higher in nucleotide diversity and show stronger signatures of balancing selection. We did not see any significant deviations from the null for genes underlying discordant symbiotic pleiotropy (i.e., dotted grey lines in **Fig. 3**), or for genes underlying both concordant and discordant rhizobium fitness pleiotropy (**SI Fig. S3**; all comparisons > 0.1).

The genetic basis of rhizobium fitness and symbiotic pleiotropy. We identified a total of 168 variants underlying symbiotic pleiotropy, corresponding to 128 coding-genes (see **SI Dataset S2** for variant-level and **SI Table S2 & Dataset S3** for gene-level summaries). 60 and 93 of these variants were uniquely associated with fitness proxies measured on plant lines A17 and DZA, while only 15 variants were shared between hosts. For rhizobium fitness pleiotropy, we identified 130 variants, corresponding to 99 coding genes, and similarly found little overlap between hosts: 61 and 56 pleiotropic variants were uniquely associated with fitness proxies measured on A17 and DZA, respectively, while only 13 were shared, all of which also underlie symbiotic pleiotropy as well and can thus be considered "universally pleiotropic" (**SI Dataset S2**).

Regardless of pleiotropic category (i.e., rhizobium fitness vs symbiotic), most genes tagged by significant variants involved biological processes based on UNIPRO keywords including "amino acid transport" and "transcriptional regulation", as well as molecular functions including "oxidoreductase", "transferase", "DNA binding", "lyase", and "hydrolase", suggesting that the overlap in gene networks and functions may be more common across pleiotropic scales, which could be more formally tested in follow up transcriptomic studies and gene co-expression network analyses. We highlight some of the noteworthy genes identified in our analysis for each pleiotropic category in **SI Results**.

Discussion

Leveraging genomics to quantify the genetic architecture underlying symbiotic extended phenotypes gives us the power to address long-standing issues in mutualism evolution with genome-scale resolution. Here, we studied the relationships among multiple proxies of rhizobium and plant fitness to address critical questions about symbiont fitness and ask whether conflict between hosts and symbionts is prevalent at both the organismal and genomic levels. Our results overall suggest that: 1) fitness alignment between hosts and rhizobia is common at both the organismal- and genomic-levels, with genes underlying alignment showing elevated nucleotide diversity and signatures of balancing selection; and 2) the lack of a relationship or even trade-offs between rhizobium nodule number and competitive fitness mean that mixed-inoculation measures of rhizobium fitness should be prioritized if we want to predict rhizobium evolution. We discuss these main points in turn below.

Alignment of host and symbiont fitness. Many models of mutualism rely on the key assumption that cooperation is inherently costly and thus that selection should favour less-cooperative, potentially 'cheating', partners that forgo paying the costs whilst continuing to receive benefits – particularly in one-to-many symbioses where a 'tragedy of the commons' (64) should favour less-beneficial symbionts (21–23). Using genetic correlations, GWAS, and analyses of nucleotide variation to study the genetic basis of mutualism variation in natural populations of rhizobia, we find little evidence for trade-offs between rhizobium competitive fitness and plant fitness at the organismal- or genomic-levels, suggesting strong alignment between host and symbiont fitnesses in nature. Indeed, the lack of points in the bottom right quadrants of figures correlating partner quality with rhizobium competitive fitness, particularly for DZA, suggests little evidence of competitively superior 'cheating' strains that benefit at the expense of their host (65).

Fitness alignment is ultimately governed by the degree to which mutualistic trait optima are shared among partners, as well as the degree of fitness feedbacks that enforce alignment between hosts and symbionts (9, 21, 66, 67). With respect to nodule number, legume host plants have autoregulation of nodulation to limit the formation of costly nodules once sufficient N levels are achieved (68). No such constraint exists for rhizobia; every nodule formed is expected to lead to greater potential fitness benefits (69). Thus the optimum number of nodules for a plant is likely to be much lower than that for rhizobia. Indeed, the strongest evidence for host-symbiont conflict in our experiment comes from regressing plant fitness proxies with nodule number in single inoculation, suggesting conflict over a host's total investment in nodulation along

DRAFT

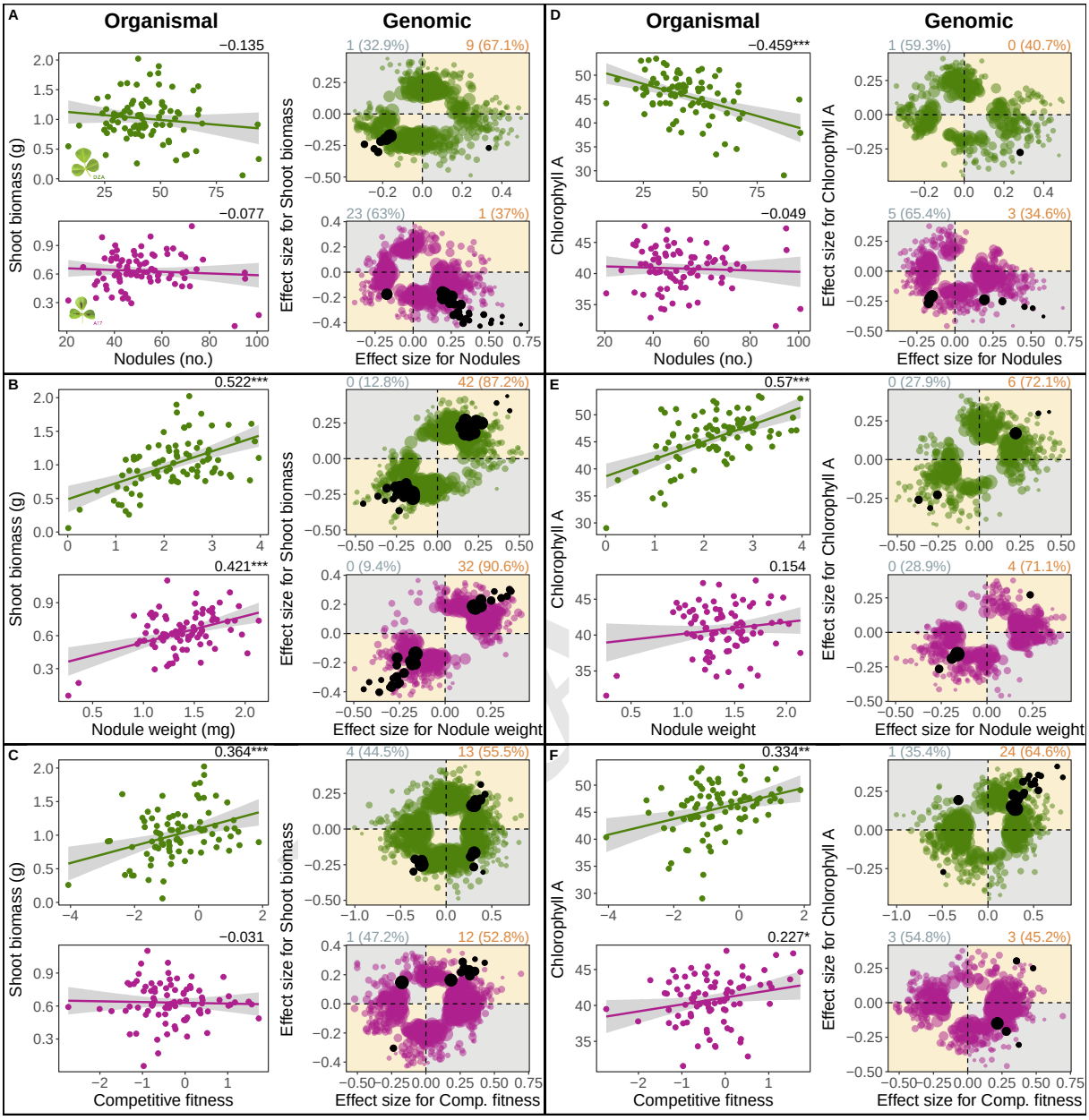

Fig. 2. Fitness alignment between rhizobia (*Ensifer meliloti*) and host (*Medicago truncatula*) prevails at both the organismal (left panels) and genomic-levels (right panels). Organismal: Genetic correlations between pairwise fitness proxies correspond to traits measured on plant lines DZA (green, top rows) or A17 (pink, bottom rows), based on 89 *Ensifer meliloti* strains. Dots represent estimated marginal means for shoot biomass (**A-C**) or leaf chlorophyll A content (**D-F**), as well as nodule number and nodule weight, all being measured in single inoculation experiments, or medians for competitive fitness measured in mixed-inoculation experiments. Numbers at top right of each correlation represent Pearson correlation coefficients, while asterisks represent significance: * = P < 0.05; ** = P < 0.01; *** = P < 0.001. **Genomic**: Dots represent variant effect sizes (i.e., beta values calculated in GEMMA), whereby effects were regressed for each pairwise trait combination. Dots falling along the diagonal (orange quadrants) represent variants with concordant (i.e., same-sign) effects, while those falling on the off-diagonal (grey quadrants) represent variants with discordant (i.e., opposite-sign) effects. Coloured dots represent variants that were significantly associated with one of the two proxies, while black dots represent pleiotropic variants, i.e., significantly associated with both proxies. Numbers outside of and percentages within parentheses at the top left and right of each plot represent the pleiotropic variant counts or proportion of total significant variants, respectively, that are discordant (left, in grey) or concordant (right, in orange).

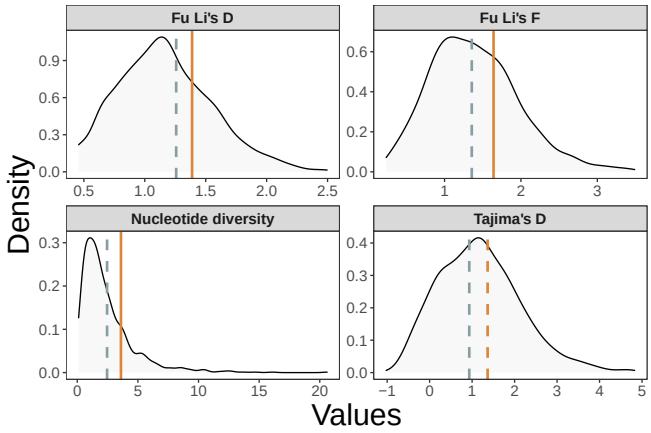

Fig. 3. Neutrality statistics show elevated values for genes underlying fitness alignment. Vertical lines represent the average values calculated for four separate statistics, grey for discordant and orange for concordant genes. Distributions represent the same statistics calculated for all genes containing significant variants based on GWAS. Dashed and solid lines represent non-significant ($p > 0.1$) and significant ($p < 0.1$) differences, respectively, between each focal gene category and all significant genes (i.e., distributions).

with previous studies (e.g., 4, 58, 70). For other traits, shared physiological constraints might result in shared fitness optima and align host and symbiont interests. For example, N_2 or O_2 permeability that scales with nodule surface area to volume ratio (71), such that nodules too large or small in size might be suboptimal for both partners. Determining whether nodule size and within-nodule fitness (i.e., cell counts per nodule) would be needed to test whether the relationship is indeed non-linear.

In one-to-many mutualisms like legumes and the diverse populations of rhizobia inhabiting their root nodules, fitness feedbacks operating on a per partner basis are predicted to align host and symbiont interests (21–24, 72). Empirical data supports the presence of such feedbacks via host sanctions wherein nodules fixing more N receive more plant C (40, 73–75), although the mechanisms are likely not perfectly accurate and have rarely been measured in the wild. Our observations of abundant alignment between host fitness proxies, nodule size, and rhizobium competitive fitness are consistent with fitness feedbacks operating to align host and symbiont fitness in nature, and are particularly striking given that neither host line co-occurs with the rhizobium strains studied, suggesting that feedbacks might operate in diverse communities irrespective of coevolutionary history.

Genomic resolution of conflict and alignment in symbiosis. Using genomics to dig below the phenotypic level can illuminate the mechanisms that favour cooperation over conflict by uncovering the genetic basis of traits central to mutualism evolution in nature, quantifying the extent of symbiotic pleiotropy, and suggesting patterns of historical selection acting on pleiotropic loci. We found little evidence for discordant pleiotropic variants underlying host and symbiont fitness, alongside evidence that selection has acted to maintain this genetic variation through time and/or space. An abundance of concordant pleiotropic variants, like we have found, may have resulted from physiological constraints that make discordant variants impossible (i.e., alleles associated with larger nodules and small host biomass are rare or non-existent), or because of correlational selection that removes discordant variants from the population altogether (76–78). Additionally, the action of feedbacks could prevent the expression of conflict loci (i.e., loci that benefit rhizobia at the expense of the host). In addition to direction, the extent of symbiotic pleiotropy (i.e., the number of pleiotropic variants) can inform whether traits are tightly coupled, or could evolve independently. Putting both the degree of concordance/discordance and number of pleiotropic variants together, strong alignment governed by many concordant pleiotropic variants should mean that proxies coevolve in the same direction, whereas conflict governed by many discordant pleiotropic variants should mean that proxies coevolve in opposite directions. Some traits with abundant pleiotropy (e.g., nodule size vs. shoot biomass) suggest a highly polygenic basis governed by many small-effect pleiotropic variants; other traits with little pleiotropy (e.g., nodule number vs. chlorophyll) are largely governed by different sets of genetic mechanisms, and are likely to evolve independently.

One challenge in mutualism research has been identifying patterns of selection acting on traits that determine fitness outcomes in natural populations, particularly for microbial symbioses (79, 80). One approach is to study patterns of molecular variation at candidate genes or loci identified through GWAS; key studies to date have often found stabilizing or purifying selection acting on key symbiosis genes for partner recognition or partner benefits, rather than patterns suggesting rapid turnover of alleles underlying conflict (81–84). In a recent in silico GWAS, O'Brien and colleagues (9) found that conflict over mutualistic trait optima between hosts and microbes tended to increase genetic diversity due to repeated sweeps, whereas alignment scenarios resulted in stabilizing selection and decreased genetic variation as trait optima were reached. In our empirical dataset, we found abundant fitness alignment, combined with evidence that the genes underlying this alignment harbour particularly high diversity and experience balancing selection. While the agents responsible for maintaining diversity remain to be determined, we found that concordant variants for both host and symbiont fitnesses were largely non-overlapping between the two host genotypes, suggesting that different host lines select for distinct rhizobium alleles, thus, maintaining rhizobia genetic diversity. That different host genotypes act as distinct selective environments for rhizobia is supported in a recent evolution experiment that paired rhizobia with five host genotypes (31), and has been found in other legume-rhizobium interactions (e.g., 85). More generally, balancing selection acting on genes underlying fitness alignment provides (at least in part) a solution to the 'market paradox', whereby host-driven selection for the 'best' symbiont genotype is expected to reduce overall symbiont diversity, yet quantitative variation in symbiotic outcomes remains (reviewed by 79, 80).

Despite abundant concordance in our data at both levels, our paired approach did reveal several instances of discordance at the variant-level despite alignment at the phenotypic-level (e.g., shoot biomass vs competitive fitness on DZA). These findings highlight the polygenic nature of partner quality variation in nature – where the collective action of individual mutations, their additive and nonadditive effects (47), and a history of selection shapes the trait variation currently present in natural populations (36, 37). Discordant alleles that are "packaged"

DRAFT

into concordant multilocus genotypes might be due to linkage disequilibrium resulting from multiple, non-mutually exclusive factors including epistatic interactions among individual variants that render discordant variants effectively neutral and/or past selection acting to favour allelic combinations that increase both host and symbiont fitness (and disfavour discordant combinations; 86, 87). We did not detect any evidence consistent with selection having acted on loci containing discordant variants; however, we note that this comparison was underpowered given the small number of variants and does not rule out correlational selection (88). Because our experiments were conducted in a single, low N environment, and external N directly reduces the dependence of hosts on rhizobia, we might be underestimating conflict in nature where N sources fluctuate over space and time. How selection on symbiotic mutualisms shifts in natural mutualist populations has rarely been measured (but see 70, 89–91), and should be a priority for future work.

Trade-offs among rhizobium fitness proxies and the rhizobium competition problem. Understanding how microbial symbioses, which are ubiquitous and important in nature, evolve (or coevolve) requires accurate estimates of symbiont fitness in ecologically realistic settings. Given our evidence at both the organismal- and genomic-levels for trade-offs among proxies, and because diverse strains of rhizobia co-occur in nature, competitive fitness proxies should be used whenever possible (27). Nevertheless these proxies are not replacements for single-strain inoculation studies because they cannot be used to assign individual genetic means for whole-plant measures of partner quality (plant biomass and seed number) to individual strains (e.g., 4, 28, 44, 47), necessitating that partner quality and rhizobium fitness be measured on separate individuals.

Comparing even the same fitness proxy across single and mixed inoculation experiments (e.g., nodule number) may be more like comparing apples to oranges if the physiological or signalling processes governing the proxy differ between the two types of experiments. In single-inoculation experiments, plant autoregulation of nodulation (68) means that plants singly inoculated with a less-effective strain will tend to form more total nodules than if they were inoculated with a more effective N-fixing strain (e.g., 4, 92). In mixed inoculation experiments, however, nodule number for a particular strain depends on their competitive ability due to strain-strain competition and/or host signaling interactions (93, 94), as well as any pre-infection partner choice, i.e., overrepresentation of more beneficial strains in nodules (42, 52, 95, 96). Therefore, the predominant processes that determine nodule number in mixed inoculation likely differ from those governing nodule number in single inoculation, the former being more representative of ecological conditions.

In contrast to nodule number, processes that govern nodule size may be more consistent across single- and mixed-inoculation experiments. For example, if post-infection host control mechanisms (aka host sanctions, *sensu* 42) represent the proportional allocation of plant C per unit N fixed by rhizobia within a nodule, then nodule size might be indicative of N-fixation efficiency of the strain, perhaps explaining why nodule size and number tend to trade-off in single inoculation (i.e., plants singly inoculated with less effective N-fixing strains form smaller, more numerous nodules). If nodule size is indeed more consistent across experiment types, we would predict *a priori* to see stronger positive genetic correlations (and a greater genetic overlap) between nodule size and competitive fitness, given that the latter represents not only a strain's ability to compete for nodule occupancy but also for plant C once nodules are formed (97). Yet, we found evidence for a positive correlation between these proxies only on one of our two host lines (DZA), and very little genetic overlap among variants underlying these proxies overall. Whether N-fixation or C cost (which presumably determine partner quality) also change when strains are in single versus mixed populations on plant roots remains unaddressed, but could be tested using within-nodule fitness measurements (e.g., cell counts per nodule).

Together our results suggest that the genetic architectures underlying rhizobium fitness proxies, and their relationships, are host-dependent, and thus that their evolutionary trajectories are influenced not only by variation in the rhizobium's genome, but also by variation in host traits. For example, host genotypes could differ in their ability to exert sanctions or partner choice, quantitative traits known to vary in host populations (48, 98). At the genomic-level, distinct variants underlying discordant pleiotropic effects on one host (A17) and concordant pleiotropic effects on the other (DZA) suggests that different genetic mechanisms (i.e., genes, pathways, metabolic processes) govern the relationship between these traits when rhizobia interact with different hosts. Such host genotype-dependent shifts in the rhizobium genetic architecture of these symbiotic extended phenotypes is supported by transcriptomics studies of G x G interactions (27, 45) and GWAS revealing distinct sets of candidate genes in different host backgrounds (28, 47); similar variation exists in hosts (99) and undoubtedly interacts with the variants we identify here, thus, should be accounted for if we are to build a reliable G→P map.

Conclusions. Combining multiple and single inoculation experiments in an association mapping framework using a panel of symbiont diversity from natural populations, we find a lack of conflict underlying mutualism benefits at the phenotypic- and genomic-levels and show that single-strain versus diverse multi-strain environments lead to the expression of distinct traits in symbiotic rhizobia. While we acknowledge the necessity of single strain estimates of partner quality, our findings highlight the need to study the fitness of symbionts in their population and community contexts to make predictions about the evolution of these bacterial symbionts. Our results suggest that fitness conflict between these hosts and symbionts might be largely resolved in natural populations, but paradoxically that symbiotic variation is actively maintained by selection. We emphasize the need for rhizobium population biology aimed at identifying the selective agents acting at ecologically relevant scales.

Materials and Methods

Study system. Full details are provided in Riley et al. (100) and Batstone et al. (47). *Ensifer* (formerly *Sinorhizobium*) *meliloti* is an Alphaproteobacteria that forms symbiosis with various *Medicago* spp., fixing atmospheric N in return for plant photosynthate. All 191 strains used here were collected from the native range of the host *Medicago truncatula*, spanning Spain to France (as detailed in (100)). *E. meliloti* has a multipartite genome, including a large (∼3.7 Mbp)

DRAFT

chromosome and two symbiosis plasmids (pSymA and pSymB, ∼1.4 Mbp and ∼1.7 Mbp, respectively), the latter known to contain many canonical genes for nodulation and N-fixation (101, 102). We used two lines of *M. truncatula* DZA 315.16 (hereafter DZA) and A17 in separate single inoculation experiments, and a mixed inoculation experiment detailed below.

Single-strain inoculation experiments. Full details are provided in Batstone et al. (47), and **SI Methods**. Briefly, we conducted two separate experiments in Sept and Nov 2018, one for each *M. truncatula* line, DZA and A17, respectively. At least three replicate seeds of each plant line were singly inoculated with one of 191 strains of *Ensifer meliloti* isolated from soil spanning *M. trunatula*'s native range (82, 100). Within each experiment, we measured proxies for both plant fitness (shoot biomass, leaf chlorophyll A content) and rhizobium fitness (nodule number, per nodule weight).

Mixed-strain inoculation experiment. Full details and explanation of the 'Select and Resequence' methodology is provided in Burghardt et al. (27), and **SI Methods**. Briefly, we grew A17 and DZA hosts with a mixed inocula composed of an 89 strain subset of those used in the previous experiment. We measured an additional fitness proxy, rhizobium competitive fitness, by sequencing pooled nodule samples from each plant and calculating each strain's fold change at the end of the experiment relative to its mean frequency at the beginning.

Phenotypic analyses. As described in Batstone et al. (47), we calculated the estimated marginal means for each trait in each experiment, correcting for the effect of rack, using the emmeans package (v1.4.1 103) in the R environment (104). We then conducted linear pairwise regressions (lm option in base R) for each trait against the other within each plant line, as well as the same trait between the two plant lines. We calculated the Pearson correlation coefficients using the "complete observation" option of the cor function in the corrplot package (v0.88, 105), and calculated adjusted p-values using the "fdr" (false discovery rate) method of the corr.test function of R's psych package (v1.8.12, 106). To visualize the resulting genetic correlations, we used a custom function that utilizes ggplot (v3.3.3, 107) to produce the paneled figures.

DNA isolation, whole-genome sequencing, and variant calling. Detailed methods are provided in Riley et al. (100), and **SI Methods**. We obtained DNA from each of the 191 rhizobium isolates, sequenced their full genomes including the chromosome and two symbiosis plasmids, used a common reference-based assembly pipeline to align sequences and call single nucleotide polymorphisms (SNPs), and filtered the resulting SNPs based on sequence quality, depth, minor allele frequency, and missingness, resulting in a total of 36,526 filtered SNPs.

Genome-wide association tests. Detailed methods are provided in **SI Methods**. We conducted linear mixed models (LMMs) implemented in GEMMA (v0.98.1, 108) to quantify the strength of the association between variants and fitness proxies after correcting for rhizobium population structure. We ran ten separate association analyses in GEMMA, one for each of the four fitness proxies measured in single-strain inoculation experiments (leaf chlorophyll A content, shoot biomass, nodule number, nodule weight), plus rhizobium competitive fitness measured in the mixed-strain inoculation experiment, and for both host plant lines (DZA, A17; five traits x two hosts = ten runs).

Genomic analyses. We identified pleiotropic variants as those that were significantly associated with more than one trait on the same host line. When variants were significantly associated with multiple rhizobium fitness proxies only, they were designated as underlying rhizobium fitness pleiotropy, whereas when they were associated with at least one rhizobium and plant fitness proxy, they were designated as underlying symbiotic pleiotropy.

In order to determine whether pleiotropic variants had discordant or concordant effects on pairwise trait combinations, we conducted correlations of effect sizes calculated in GEMMA for each pairwise trait combination measured on each plant line separately (e.g., nodule number versus nodule weight on DZA) using linear regression (lm function) in base R. When effects were the same in sign (either +,+ or -,-), we tagged them as being concordant, whereas when they were opposite in sign (either +,- or -,+), we tagged them as being discordant. We used a custom function in R to plot the resulting effect size correlations.

To test whether selection acted on genes underlying rhizobium fitness and symbiotic pleiotropy, we used the R package PopGenome (v2.7.5, 109) to compute several statistics that are commonly used to test for signatures of historical selection and/or departures of neutrality, namely, nucleotide diversity (i.e., π), Tajima's D, as well as Fu and Li's F and D (110, 111). Additional details appear in **SI Methods**.

Data Availability. Strains and plant lines are available upon request. All raw data and analysis code will be made available on GitHub upon publication. The Supplementary Information doc contains additional methods, results, figures and tables, as well as descriptions of datasets that have been uploaded along with this manuscript, and can be additionally accessed from bioRχiv (doi:https://doi.org/10.1101/2021.07.19.452989). Once raw sequence reads and assemblies are archived and made available on NCBI, accession numbers will be added to this manuscript.

ACKNOWLEDGMENTS. For greenhouse help, we thank Alex Riley, Hanna Lindgren, Cassandra Allsup, Laura Goralka, Mannix Burns, Brendan Epstein, and Amy-Marshall Colon, as well as the greenhouse staff, particularly Debbie Black. For feedback on the results and analyses, we thank Peter Tiffin, Brendan Epstein, and Robert Williamson. We acknowledge funding from the National Science Foundation (K.D. Heath: IOS-1645875 and NPGI-1401864; L.T. Burghardt: IOS-1724993 and IOS-1856744), and a Carl R. Woese Institute for Genomic Biology postdoctoral fellowship to R.T. Batstone, strain sequencing by Joint Genome Institute (CSP-1223795).

1. TF Mackay, EA Stone, JF Ayroles, The genetics of quantitative traits: challenges and prospects. *Nat. Rev. Genet.* **10**, 565–577 (2009).
2. M Pigliucci, Genotype–phenotype mapping and the end of the 'genes as blueprint' metaphor. *Philos. Transactions Royal Soc. B* **365**, 557–566 (2010).
3. B Lehner, Genotype to phenotype: lessons from model organisms for human genetics. *Nat. Rev. Genet.* **14**, 168–178 (2013).
4. KD Heath, Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution* **64**, 1446–1458 (2010).
5. ML Friesen, et al., Microbially mediated plant functional traits. *Annu. Rev. Ecol. Evol. Syst.* **42**, 23–46 (2011).
6. J Ferrari, F Vavre, Bacterial symbionts in insects or the story of communities affecting communities. *Philos. Transactions Royal Soc. B* **366**, 1389–1400 (2011).
7. AL Gould, et al., Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci.* **115**, E11951–E11960 (2018).
8. J Lynch, E Hsiao, Microbiomes as sources of emergent host phenotypes. *Science* **365**, 1405–1409 (2019).
9. AM O'Brien, CN Jack, ML Friesen, ME Frederickson, Whose trait is it anyways? Coevolution of joint phenotypes and genetic architecture in mutualisms. *Proc. Royal Soc. B* **288**, 20202483 (2021).
10. FW Stearns, One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767–773 (2010).
11. GP Wagner, J Zhang, The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213 (2011).
12. SC Stearns, Trade-offs in life-history evolution. *Funct. Ecol.* **3**, 259–268 (1989).
13. N Barton, Pleiotropic models of quantitative variation. *Genetics* **124**, 773–782 (1990).
14. D Roff, , et al., Trade-offs between growth and reproduction: an analysis of the quantitative genetic evidence. *J. Evol. Biol.* **13**, 434–445 (2000).
15. SP Otto, Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. Royal Soc. B* **271**, 705–714 (2004).
16. JL Bronstein, *Mutualism.* (Oxford University Press, USA), (2015).
17. KR Foster, G Shaulsky, JE Strassmann, DC Queller, CR Thompson, Pleiotropy as a mechanism to stabilize cooperation. *Nature* **431**, 693–696 (2004).
18. S Mitri, KR Foster, Pleiotropy and the low cost of individual traits promote cooperation. *Evolution* **70**, 488–494 (2016).
19. M Dos Santos, M Ghoul, SA West, Pleiotropy, cooperation, and the social evolution of genetic architecture. *PLoS Biol.* **16**, e2006671 (2018).
20. DC Queller, Pleiotropy and synergistic cooperation. *PLoS Biol.* **17**, e3000320 (2019).
21. JJ Bull, W Rice, Distinguishing mechanisms for the evolution of co-operation. *J. Theor. Biol.* **149**, 63–74 (1991).
22. S West, ET Kiers, I Pen, R Denison, Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15**, 830–837 (2002).
23. SA West, ET Kiers, EL Simms, RF Denison, Sanctions and mutualism stability: why do rhizobia fix nitrogen? *Proc. Royal Soc. B* **269**, 685–694 (2002).
24. KR Foster, H Kokko, Cheating can stabilize cooperation in mutualisms. *Proc. Royal Soc. B* **273**, 2233–2239 (2006).

DRAFT

25. ML Friesen, et al., The ecological genomic basis of salinity adaptation in tunisian *Medicago truncatula*. *BMC Genomics* **15**, 1160 (2014).
26. CR Klinger, JA Lau, KD Heath, Ecological genomics of mutualism decline in nitrogen-fixing bacteria. *Proc. Royal Soc. B* **283**, 20152563 (2016).
27. LT Burghardt, et al., Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proc. Natl. Acad. Sci.* **115**, 2425–2430 (2018).
28. B Epstein, et al., Genome-wide association analyses in the model rhizobium *Ensifer meliloti*. *MSphere* **3** (2018).
29. EJ Von Wettberg, et al., Ecology and genomics of an important crop wild relative as a prelude to agricultural innovation. *Nat. Commun.* **9**, 1–13 (2018).
30. SS Porter, J Faber-Hammond, AP Montoya, ML Friesen, C Sackos, Dynamic genomic architecture of mutualistic cooperation in a wild population of *Mesorhizobium*. *The ISME J.* **13**, 301–315 (2019).
31. RT Batstone, AM O'Brien, TL Harrison, ME Frederickson, Experimental evolution makes microbes more cooperative with their local host genotype. *Science* **370**, 476–478 (2020).
32. TF Mackay, W Huang, Charting the genotype–phenotype map: lessons from the *Drosophila melanogaster* genetic reference panel. *Wiley Interdiscip. Rev. Dev. Biol.* **7**, e289 (2018).
33. EA Herre, N Knowlton, UG Mueller, SA Rehner, The evolution of mutualisms: exploring the paths between conflict and cooperation. *Trends ecology & evolution* **14**, 49–53 (1999).
34. JL Sachs, CJ Essenberg, MM Turcotte, New paradigms for the evolution of beneficial infections. *Trends Ecol. & Evol.* **26**, 202–209 (2011).
35. P Innocenti, EH Morrow, The sexually antagonistic genes of *Drosophila melanogaster*. *PLoS Biol.* **8**, e1000335 (2010).
36. JK Conner, DL Hartl, , et al., *A Primer of Ecological Genetics*. (Sinauer Associates Incorporated), (2004).
37. M Lynch, B Walsh, *Genetics and Analysis of Quantitative Traits*. (Sinauer, Sunderland, MA) Vol. 1, (1998).
38. MA Parker, Plant fitness variation caused by different mutualist genotypes. *Ecology* **76**, 1525–1535 (1995).
39. RF Denison, Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *The Am. Nat.* **156**, 567–576 (2000).
40. ET Kiers, RA Rousseau, SA West, RF Denison, Host sanctions and the legume–rhizobium mutualism. *Nature* **425**, 78–81 (2003).
41. EL Simms, et al., An empirical test of partner choice mechanisms in a wild legume–rhizobium interaction. *Proc. Royal Soc. B* **273**, 77–81 (2006).
42. KD Heath, P Tiffin, Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution* **63**, 652–662 (2009).
43. CH Foyer, et al., Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* **2**, 1–10 (2016).
44. J Burdon, P Thrall, Spatial and temporal patterns in coevolving plant and pathogen associations. *The Am. Nat.* **153**, S15–S33 (1999).
45. KD Heath, PV Burke, JR Stinchcombe, Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol. Ecol.* **21**, 4735–4747 (2012).
46. LT Burghardt, et al., Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol. Ecol.* **26**, 6122–6135 (2017).
47. R Batstone, et al., The complex genetics of symbiotic extended phenotypes across multiple environments in a model mutualism (bioRχiv. https://www.biorxiv.org/content/10.1101/2021.08.03.454976v2) (2021).
48. AK Simonsen, JR Stinchcombe, Standing genetic variation in host preference for mutualist microbial symbionts. *Proc. Royal Soc. B* **281**, 20142036 (2014).
49. KD Heath, P Tiffin, Context dependence in the coevolution of plant and rhizobial mutualists. *Proc. Royal Soc. B* **274**, 1905–1912 (2007).
50. WC Ratcliff, K Underbakke, RF Denison, Measuring the fitness of symbiotic rhizobia. *Symbiosis* **55**, 85–90 (2011).
51. R Oono, CG Anderson, RF Denison, Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proc. Royal Soc. B* **278**, 2698–2703 (2011).
52. ML Friesen, Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytol.* **194**, 1096–1111 (2012).
53. ET Kiers, WC Ratcliff, RF Denison, Single-strain inoculation may create spurious correlations between legume fitness and rhizobial fitness. *New Phytol.* **198**, 4–6 (2013).
54. ML Friesen, KD Heath, One hundred years of solitude: integrating single-strain inoculations with community perspectives in the legume–rhizobium symbiosis. *New Phytol.* **198**, 7–9 (2013).
55. RF Denison, ET Kiers, Life histories of symbiotic rhizobia and mycorrhizal fungi. *Curr. Biol.* **21**, R775–R785 (2011).
56. H Moawad, W Ellis, E Schmidt, Rhizosphere response as a factor in competition among three serogroups of indigenous *Rhizobium japonicum* for nodulation of field-grown soybeans. *Appl. Environ. Microbiol.* **47**, 607–612 (1984).
57. R Oono, RF Denison, ET Kiers, Controlling the reproductive fate of rhizobia: how universal are legume sanctions? *New Phytol.* **183**, 967–979 (2009).
58. SS Porter, EL Simms, Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecol. Lett.* **17**, 1121–1129 (2014).
59. ML Friesen, Microbially mediated plant functional traits. *Mol. Microb. Ecol. Rhizosphere* **1**, 87–102 (2013).
60. LT Burghardt, B Epstein, P Tiffin, Legacy of prior host and soil selection on rhizobial fitness in planta. *Evolution* **73**, 2013–2023 (2019).
61. D Bullock, D Anderson, Evaluation of the minolta spad-502 chlorophyll meter for nitrogen management in corn. *J. Plant Nutr.* **21**, 741–755 (1998).
62. JM Swiader, A Moore, Spad-chlorophyll response to nitrogen fertilization and evaluation of nitrogen status in dryland and irrigated pumpkins. *J. Plant Nutr.* **25**, 1089–1100 (2002).
63. DJ Weese, KD Heath, BT Dentinger, JA Lau, Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* **69**, 631–642 (2015).
64. G Hardin, The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
65. EI Jones, et al., Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecol. Lett.* **18**, 1270–1284 (2015).
66. JL Sachs, UG Mueller, TP Wilcox, JJ Bull, The evolution of cooperation. *The Q. Rev. Biol.* **79**, 135–160 (2004).
67. ME Frederickson, Rethinking mutualism stability: cheaters and the evolution of sanctions. *The Q. Rev. Biol.* **88**, 269–295 (2013).
68. DE Reid, BJ Ferguson, S Hayashi, YH Lin, PM Gresshoff, Molecular mechanisms controlling legume autoregulation of nodulation. *Annals Bot.* **108**, 789–795 (2011).
69. JL Sachs, KW Quides, CE Wendlandt, Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *New Phytol.* **219**, 1199–1206 (2018).
70. JN Ossler, CA Zielinski, KD Heath, Tripartite mutualism: Facilitation or trade-offs between rhizobial and mycorrhizal symbionts of legume hosts. *Am. J. Bot.* **102**, 1332–1341 (2015).
71. CA King, LC Purcell, Soybean nodule size and relationship to nitrogen fixation response to water deficit. *Crop. Sci.* **41**, 1099–1107 (2001).
72. EG Weyl, ME Frederickson, WY Douglas, NE Pierce, Economic contract theory tests models of mutualism. *Proc. Natl. Acad. Sci.* **107**, 15712–15716 (2010).
73. G Laguerre, et al., Local and systemic N signaling are involved in *Medicago truncatula* preference for the most efficient *Sinorhizobium* symbiotic partners. *New Phytol.* **195**, 437–449 (2012).
74. JU Regus, et al., Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *Am. J. Bot.* **104**, 1299–1312 (2017).
75. A Westhoek, et al., Conditional sanctioning in a legume–rhizobium mutualism. *Proc. Natl. Acad. Sci.* **118** (2021).
76. DL Hartl, AG Clark, AG Clark, *Principles of population genetics*. (Sinauer associates Sunderland, MA) Vol. 116, (1997).
77. F Guillaume, MC Whitlock, Effects of migration on the genetic covariance matrix. *Evolution* **61**, 2398–2409 (2007).
78. JB Saltz, FC Hessel, MW Kelly, Trait correlations in the genomics era. *Trends Ecol. & Evol.* **32**, 279–290 (2017).
79. KD Heath, JR Stinchcombe, Explaining mutualism variation: a new evolutionary paradox? *Evolution* **68**, 309–317 (2014).
80. KS Stoy, AK Gibson, NM Gerardo, LT Morran, A need to consider the evolutionary genetics of host–symbiont mutualisms. *J. Evol. Biol.* **33**, 1656–1668 (2020).
81. JA Kimbrel, et al., Mutualistic co-evolution of typeIII effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLoS Pathog.* **9**, e1003204 (2013).
82. MA Grillo, S De Mita, PV Burke, KL Solórzano-Lowell, KD Heath, Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution* **70**, 2704–2717 (2016).
83. JB Yoder, Understanding the coevolutionary dynamics of mutualism with population genomics. *Am. J. Bot.* **103**, 1742–1752 (2016).
84. B Epstein, P Tiffin, Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. *Proc. Royal Soc. B* **288**, 20201804 (2021).
85. MA Parker, Legumes select symbiosis island sequence variants in *Bradyrhizobium*. *Mol. Ecol.* **21**, 1769–1778 (2012).
86. B Sinervo, E Svensson, Correlational selection and the evolution of genomic architecture. *Heredity* **89**, 329–338 (2002).
87. AG Jones, SJ Arnold, R Bürger, Stability of the G-matrix in a population experiencing pleiotropic mutation, stabilizing selection, and genetic drift. *Evolution* **57**, 1747–1760 (2003).
88. EI Svensson, et al., Correlational selection in the age of genomics. *Nat. Ecol. & Evol.* **5**, 562–573 (2021).
89. JA Lau, et al., Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. *Acta Oecologica* **39**, 80–86 (2012).
90. RT Batstone, MA Peters, AK Simonsen, JR Stinchcombe, ME Frederickson, Environmental variation impacts trait expression and selection in the legume–rhizobium symbiosis. *Am. J. Bot.* **107**, 195–208 (2020).
91. ME Afkhami, ML Friesen, JR Stinchcombe, Multiple mutualism effects generate synergistic selection and strengthen fitness alignment in the interaction between legumes, rhizobia and mycorrhizal fungi. *Ecol. Lett.* **24** (2021).
92. J Sachs, et al., Host control over infection and proliferation of a cheater symbiont. *J. Evol. Biol.* **23**, 1919–1927 (2010).
93. D Dowling, W Broughton, Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40**, 131–157 (1986).
94. Q Wang, J Liu, H Zhu, Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Front. Plant Sci.* **9**, 313 (2018).
95. BS Younginger, ML Friesen, Connecting signals and benefits through partner choice in plant–microbe interactions. *FEMS Microbiol. Lett.* **366**, fnz217 (2019).
96. X Zhang, L Wang, J Li, RT Batstone, ME Frederickson, *Medicago truncatula* adjusts root proliferation, nodule formation, and partner choice in response to local n heterogeneity. *Plant Soil* **450**, 417–428 (2020).
97. LT Burghardt, Evolving together, evolving apart: measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. *New Phytol.* **228**, 28–34 (2020).
98. RT Batstone, EM Dutton, D Wang, M Yang, ME Frederickson, The evolution of symbiont preference traits in the model legume *Medicago truncatula*. *New Phytol.* **213**, 1850–1861 (2017).
99. J Stanton-Geddes, et al., Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PloS One* **8**, e65688 (2013).
100. A Riley, M Grillo, B Epstein, P Tiffin, K Heath, Partners in space: Discordant population structure between legume hosts and rhizobium symbionts in their native range (bioRχiv. https://www.biorxiv.org/content/10.1101/2021.06.30.449460v1) (2021).
101. TM Finan, et al., The complete sequence of the 1,683-kb psymb megaplasmid from the n2-fixing endosymbiont *Sinorhizobium meliloti*. *Proc. Natl. Acad. Sci.* **98**, 9889–9894 (2001).
102. F Galibert, et al., The composite genome of the legume symbiont *Sinorhizobium meliloti*.

DRAFT

Science **293**, 668–672 (2001).
103. SR Searle, FM Speed, GA Milliken, Population marginal means in the linear model: an alternative to least squares means. *The Am. Stat.* **34**, 216–221 (1980).
104. RC Team, R foundation for statistical computing. *R: a language environment for statistical computing.* (2016).
105. W Taiyun, S Viliam, R package 'corrplot': Visualization of a Correlation Matrix (2017).
106. W Revelle, 'psych': Procedures for Personality and Psychological Research (2018).
107. H Wickham, ggplot2. *Wiley Interdiscip. Rev. Comput. Stat.* **3**, 180–185 (2011).
108. X Zhou, M Stephens, Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
109. B Pfeifer, U Wittelsbürger, SE Ramos-Onsins, MJ Lercher, Popgenome: an efficient swiss army knife for population genomic analyses in r. *Mol. Biol. Evol.* **31**, 1929–1936 (2014).
110. F Tajima, Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics* **123**, 585–595 (1989).
111. YX Fu, WH Li, Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).

DRAFT

bioRχiv
THE PREPRINT SERVER FOR BIOLOGY

Supplementary Information for

The genetic basis of cooperation and conflict in natural populations of a model symbiont

Rebecca T. Batstone, Liana T. Burghardt and Katy D. Heath

Corresponding Author: R. Batstone
E-mail: rtbatstone@gmail.com

This PDF file includes:

Supplementary text
Figs. S1 to S3
Tables S1 to S2
Legends for Dataset S1 to S3
SI References

Other supplementary materials for this manuscript include the following:

Datasets S1 to S3

Supporting Information Text

SI Methods

Single inoculation experiments. Seeds were prepared according to published protocols (1). Briefly, seeds were nick- or sandpaper-scarified, surface sterilized in EtOH for 30s, bleached (6% hypochlorite) for 7 min, rinsed with dH_2O until bleach was completely removed, and imbibed in dH_2O overnight in the dark at 4°C. For planting, we used SC10R Cone-tainers (Stuewe & Sons Inc., Tangent, OR), filling each with autoclaved potting media consisting of one part root wash (1:1:1 mix of field soil:torpedo sand:Turface) and one part Turface MVP calcined clay (Profile Products LLC, Buffalo Grove, IL, USA). A single seed was sown into each pot, and pots were completely randomized with respect to treatment across three greenhouse benches located at the University of Illinois at Urbana-Champaign (UIUC). We cultured rhizobia by loop-inoculating sterile liquid tryptone-yeast (TY) medium (2) with each strain individually, placing them into a shaking incubator set to 30°C and 200 rpm overnight. We measured the OD_{600} of each culture, and diluted to $\sim 10^6$ cells per ml using sterile TY media to ensure all plants were inoculated with the same initial starting densities. Plants were misted throughout the experiment using upright misting nozzles (1/2" M NPT, Senninger Irrigation Inc., Clermont FL), and harvested six to seven weeks post-inoculation.

Within each experiment, we measured proxies for both rhizobium quality (shoot biomass, leaf chlorophyll A content) and rhizobium fitness (nodule number, per nodule weight). We measured chlorophyll using a SPAD 502 Plus (Spectrum Technologies, Inc. Aurora, IL, USA), recording the mean of three separate measurements on the more recently emerged leaf. Once plants were harvested, all nodules were counted, and ten nodules were haphazardly chosen to be weighed to the nearest 0.001 mg, dividing this number by ten to obtain a per-nodule average for each plant. Shoots were placed in a drying oven set to 60°C and left for three days, being weighed to the nearest 0.01 g. Contamination was minimal in either experiment: we only found 24 nodules on six internal controls (range: 1-9 per plant), and no nodules on external controls.

Mixed inoculation experiment. We inoculated each strain into one of 89 culture tubes filled with 3 ml of TY media (6 g tryptone, 3 g yeast extract, 0.38 g $CaCl_2$ per L) (2) and grew cultures to stationary phase by placing the tubes in a 120 rpm shaker for 72 hours at 28°C. We combined all cultures, mixed well, and used this mixture to inoculate 3-day old seedlings with $\sim 10^7$ rhizobial cells diluted in 25 ml of N-free fertilizer. Three days prior to inoculation, we filled twelve Cone-tainers with steam-sterilized SUN GRO LP5 Sunshine mix and planted each with four seeds of either A17 or DZA. These seeds had been sterilized in 10% bleach, razor-blade scarified, and cold, dark stratified on petri plates for 2 days at 4°C. After two weeks plants were thinned to ∼3 plants per pot and were provided with nitrogen-free fertilizer (3) and sterile H_2O as needed. These plants spent the first month of life in a growth chamber and then were moved to the greenhouse. After twelve weeks we harvested and pooled all nodules from each replicate pot, sterilized them for 30 seconds in 10% bleach, rinsed them thoroughly, and placed the nodules in a 15 ml falcon tube with 3 ml of sterile 0.85% NaCl solution. We homogenized nodules for 1-2 min at the highest speed using a handheld tissue homogenizer (Omni International, Kennesaw, GA). We then enriched for undifferentiated rhizobia using centrifugation: first a 10 min slow spin at 400 x g to remove the heavy plant and bacteroid debri and then subjected the supernatant to a fast spin at 16,000 x g to pellet the remaining undifferentiated rhizobial cells. These pellets were frozen and we used a Plant DNeasy kit to extract DNA from both nodule pellets and the initial inoculum. Next we generated NexteraXT libraries obtained 250 bp paired-end reads on an Illumina NextSeq to ∼120x coverage. After trimming and cleaning, SNP variant frequencies were called by aligning reads to the USDA 1106 reference genome, finally HARP was used to infer the relative frequency of each strain in each pool (4). Sequencing revealed an unintended strain of *Ensifer medicae* in the community. This *E. medicae* strain comprised less than 0.3% of the original mixture and less than .01% of the final nodule community in either host. We omitted this strain from further analysis. Lastly, we calculated overall strain fitness as the fold change in the frequency of a strain (q_x) after selection relative to its mean frequency across the four sequencing replicates of the initial community $log2(q_x selected/q_{\bar{x}} initial)$. This transformation controls for slight differences in initial strain frequency and leads to a normally-distributed trait for downstream analysis. Raw read data can be found at NCBI BioProject: PRJNA388336.

DNA isolation, whole-genome sequencing, and variant calling. In order to obtain DNA from each of the 191 rhizobia isolates, we grew overnight cultures as described above, and used the Qiagen DNeasy kit (Hilden, Germany) to extract and purify DNA, as per the manufacturer's protocol. We then submitted DNA samples to be sequenced by the DOE Joint Genome Institute (JGI) for sequencing (Berkeley, CA, USA), using the Illumina HiSeq-2500 1TB platform (Illumina, Inc., San Diego, CA, USA) with 100 bp paired-end reads. We recovered only 166 whole-genome sequences that passed quality thresholds, and thus, we regrew 25 of the 191 strains, extracted their DNA using the Zymogen Quick-DNA kit for Fungi or Bacteria (Irvine, CA, USA), and then submitted these samples to be sequenced by the Roy J. Carver Biotechnology Center at the University of Illinois at Urbana-Champaign (USA) using the Novaseq 6000 platform (Illumina, Inc, San Diego, CA, USA) with 150 bp paired-end reads. We passed the resulting raw reads through a standard bioinformatics pipeline involving read trimming (TrimGalore!: https://github.com/FelixKrueger/TrimGalore; HT-Stream: https://github.com/s4hts/HTStream), alignment to the reference strain *Ensifer meliloti* UDSA 1106 (assembly ID: GCA_002197065.1) using the bwa mem algorithm with default settings (v0.7.17-r1188, 5), and variant calling using FreeBayes (v1.3.1-16, 6). We hard-filtered the resulting 491,277 variants using vcftools (v0.1.17, 7), reducing the set to 36,526 variants with quality scores above 20, minor allele frequencies (MAF) of $\leq 5\%$, minimum mean depth of 20 and a maximum mean depth of 230, successful genotype calls for at least 80% of the strains, and removed any sites that differed between our samples and the reference strain, but were invariant among our samples.

Genome-wide association tests. Because association analyses rely on variants being statistically distinguishable, we first had to identify variants in strong linkage disequilibrium (LD), picking a representative variant within each LD group using the method employed in Epstein et al. (8). Briefly, we calculated the pairwise LD among all quality-filtered variants (N = 36,526), grouping variants in strong LD (95%), and then picking the variant within each LD group exhibiting the highest minor allele frequency, and then the least missing data (if ties were present). This LD grouping analysis resulted in 8,646 representative variants that went on to be tested in association analyses: 757 on the chromosome, 3,277 on the pSymA, and 4,612 on the pSymB. We used the -lmm 4 option in GEMMA, including a standardized (-gk2 option) kinship matrix in order to control for population structure among our rhizobia strains. We used the LD-filtered variants to construct k-matrices for each genomic region separately (i.e., one for each of the chromosome, pSymA, and pSymB), given that each replicon exhibits its own distinct evolutionary history (9).

Although GEMMA calculates p-values for each association, these are likely to be highly conservative once corrected for multiple tests, and do not take into account differences in trait distributions or LD patterns unique to each dataset (8, 10). Thus, in order to determine the significance of each association, we conducted a permutation method similar to Epstein et al. (8), whereby we randomized phenotypes with respect to genotypes, re-ran the association analyses in GEMMA using the same parameters, and repeated this process 1000 times to produce a null distribution of effect sizes for each variant in the model. We then used the findinterval method in R to determine whether our observed effect sizes exceeded the 2.5% or 97.5% intervals of the null distribution, and if so, tagged these variants as being significantly associated with our traits of interest. After running permutation tests for all 10 host-trait combinations, 2,471 (∼30%) of the ∼8,650 LD-filtered variants were tagged as being significantly associated with one or more traits, in one or more host lines: in total, only 96 of these variants were found on the chromosome, while 998 and 1,377 were found on pSymA and pSymB, respectively. We then determined whether variants were synonymous versus nonsynonymous using SnpEff (v4.3t, 11), and extracted functional information for each gene tagged by significant variants using the intersect option in bedtools (v2.29.2, 12), including the genome annotation file available from NCBI for the reference sequence *Ensifer meliloti* USDA 1106 (assembly ID: GCA_002197065.1).

Genomic analyses. We calculated our four test statistics for several subsets of genes: all genes on the pSymA or pSymB plus one on the chromosome (N = 3,103; "all" genes), genes tagged by variants included in our GWAS (N = 1,726; "GWAS" genes), genes containing variants significantly associated with one or more traits (N = 1,051; "sig" genes), and our focal genes of interest, which contain at least one pleiotropic variant underlying rhizobium fitness or symbiotic pleiotropy (N = 198 total: for rhizobium fitness, N = 10 with concordant and 85 discordant effects; for symbiotic pleiotropy, 97 with concordant and 21 with discordant effects). Test statistics calculated for all 3,103 genes appear in **SI Dataset S1**.

We only included genes on the symbiosis plasmids given that only one of our focal pleiotropic genes was found on the chromosome, whereas 93 were found on pSymA and 104 on pSymB. In order to control for unequal sampling sizes within each focal gene category, we conducted t-tests comparing each focal gene category to a random subsample of equivalent size, p-values recorded, and then repeated 1000 times to calculate the proportion of tests in which our focal and randomized gene subsets were not significantly different (i.e., $P > 0.05$). If the proportion of non-significant tests was less than 10%, then we considered the focal gene category to be significantly different from the randomized set. Because we found that both "GWAS" and "sig" genes had elevated values of diversity compared to the entire genome (**SI Table S1**), likely because these genes by definition had to contain variants to be included in our GWAS, we decided to take a conservative approach by comparing our focal genes categories to the "sig" genes as our null.

SI Results

Genomic basis of rhizobium fitness and symbiotic pleiotropy. For genes underlying rhizobium fitness pleiotropy (**SI Table S2 & Dataset S3**), we found several variants of interest in some the more well-known symbiosis genes on pSymA, including *noe*A (nodulation protein, ORF/RefSeq ID: SMa0773/NP_435662.1) and *fix*N (SMa1220/NP_435911.1), underlying the tradeoff between nodule number and nodule weight or competitive fitness, respectively. We found a strong signal of tradeoffs between nodule number and nodule weight in genes on pSymA involved in catalytic activity (e.g., Fumarylacetoacetate hydrolase, SMa0247/NP_435377.1), cellular membrane integrity (e.g., *dct*Q TRAP permease, SMa0249/NP_435378.1), as well as genes on pSymB involved in oxidoreductase activity (e.g., *thu*B, SM_b20330/NP_436856.1). Several genes contained multiple variants that were associated both with discordant pleiotropy between nodule number and either nodule weight or competitive fitness, as well as concordant pleiotropy between nodule weight and competitive fitness, one of which was the only gene of interest found on the chromosome (i.e., *snd*H L-sorbosone dehydrogenase, SMc04291/NP_386155.1), while others were found on pSymB (e.g., methanol dehydrogenase protein, SM_b20173/NP_436713.1; uncharacterized domain-containing protein, SM_b20139/NP_436679.1).

We found several symbiotic pleiotropic variants in canonical symbiosis genes on pSymA, including *fix* genes (I2: SMa0621/NP_435571.1; K2: SMa0762/NP_435654.1; O3: SMa0615/NP_435567.1), *nod*H (SMa0851/NP_435710.1), and *nif*K (SMa0829/NP_435697.1), all of which underlie fitness alignment (**SI Table S2 & Dataset S3**). For example, the major allele of a variant located within the *fix*I2 gene was associated with an increase in both nodule weight and shoot biomass, while the major allele of a variant in *fix*K2 was associated with an increase in both shoot biomass and nodule nodule, both on DZA (**SI Dataset S2**). In general, we found a strong signal in branched-chain amino acid transport and many aminotransferases for compounds important in bacteroids and symbiosis including glutamine, aspartate, asparagine, proline, and putrescine, suggesting that there is lots of potential for alignment and conflict in amino acid cycling pathways, mutants of

which have been previously demonstrated to produce important symbiotic phenotypes (13). Genes underlying fitness conflict included a fatty acid desaturase (SMa0130/NP_435314.1), Pyrroline-5-carboxylate reductase (SM_b20003/NP_436546.1), and aldehyde dehydrogenase (SMa1844/NP_436261.1), all on A17, while the carbohydrate kinase (SM_b20489/NP_437011.1), DNA helicase (*uvr*D2: SMa2321/NP_436495.1), and esterase (SM_b21424/NP_437792.1) underlied conflict on DZA.

For genes containing variants that were "universally pleiotopic" on both hosts, including those underlying tradeoffs between rhizobium fitness proxies as well as fitness conflict or alignment, most (17/20 or 85%) were located on pSymB (**SI Dataset S3**). The major allele of these variants tended to be associated with a decrease in nodule number, and an increase in other rhizobium or plant fitness proxies. Universal genes of interest included a ferrodoxin reductase, *moc*F (SM_b20820/NP_437100.1); dihydroxy-acid dehydratase, *ilv*D4 (SM_b20115/NP_436655.1); and diaminobutyric acid deacetylase, *doe*B (SM0020_02920/WP_003525897.1), all on pSymB. The only characterized protein found on pSymA was an ornithine decarboxylase (Smed_5195/YP_001313905.1). Finally, we found many pleiotropic variants in genes involved in processes that are not intuitively linked to symbiosis, such as selenocysteine synthase, molybdoterin, conjugative transfer and plasmid stabilization, hemolysin, and siderophores.

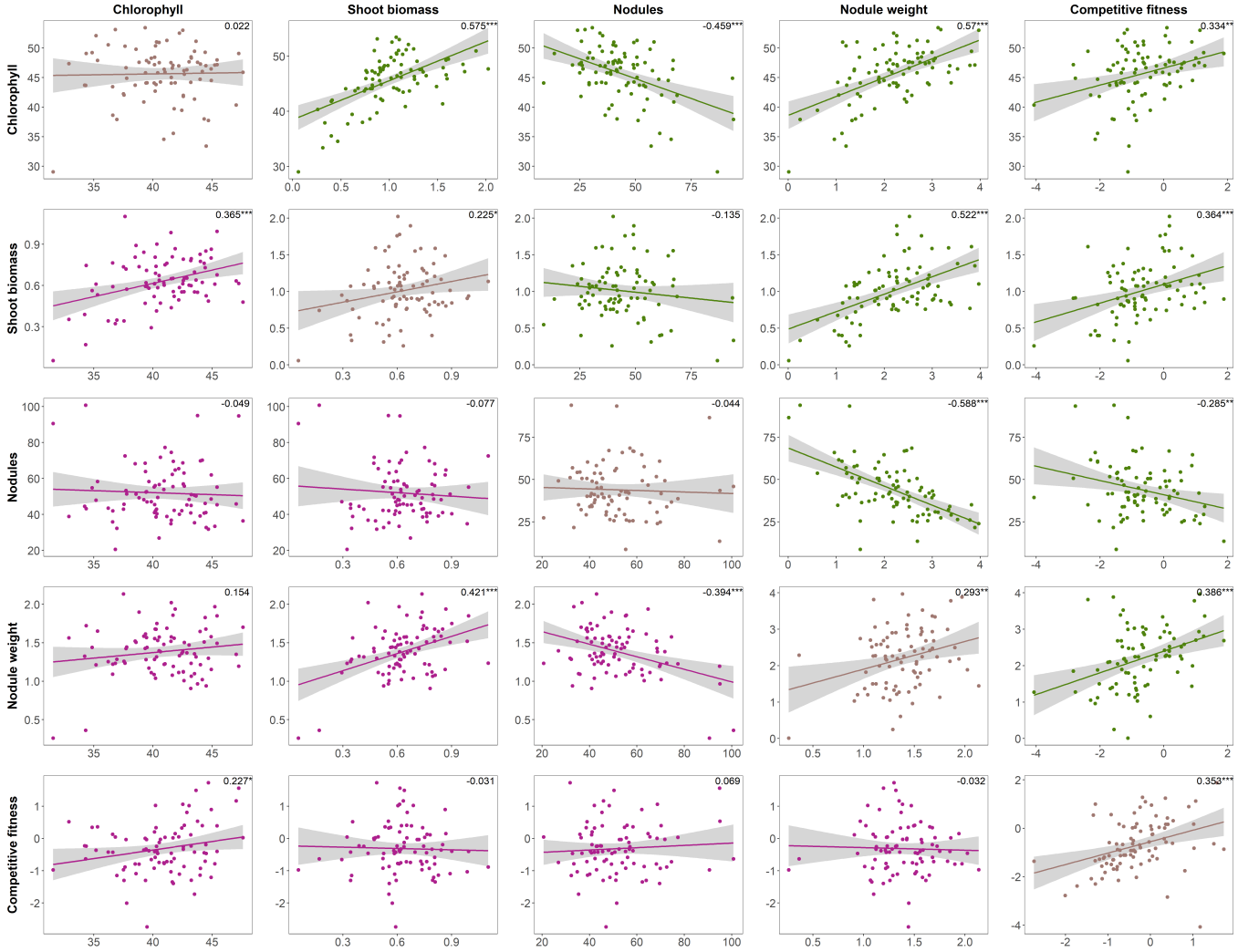

Fig. S1. Genetic correlations for traits measured on both host lines. Genetic correlations among traits measured on host lines DZA (above diagonal, in green) or A17 (below diagonal, in pink) as well as between the same trait across hosts (along diagonal). Correlations based on estimated marginal means of each rhizobia strain corrected for rack. Numbers in top right corners of each plot indicate Pearson correlation coefficients. Significance: $p < 0.001$ = '***'; $p < 0.01$ = '**'; $p < 0.05$ = '*'.

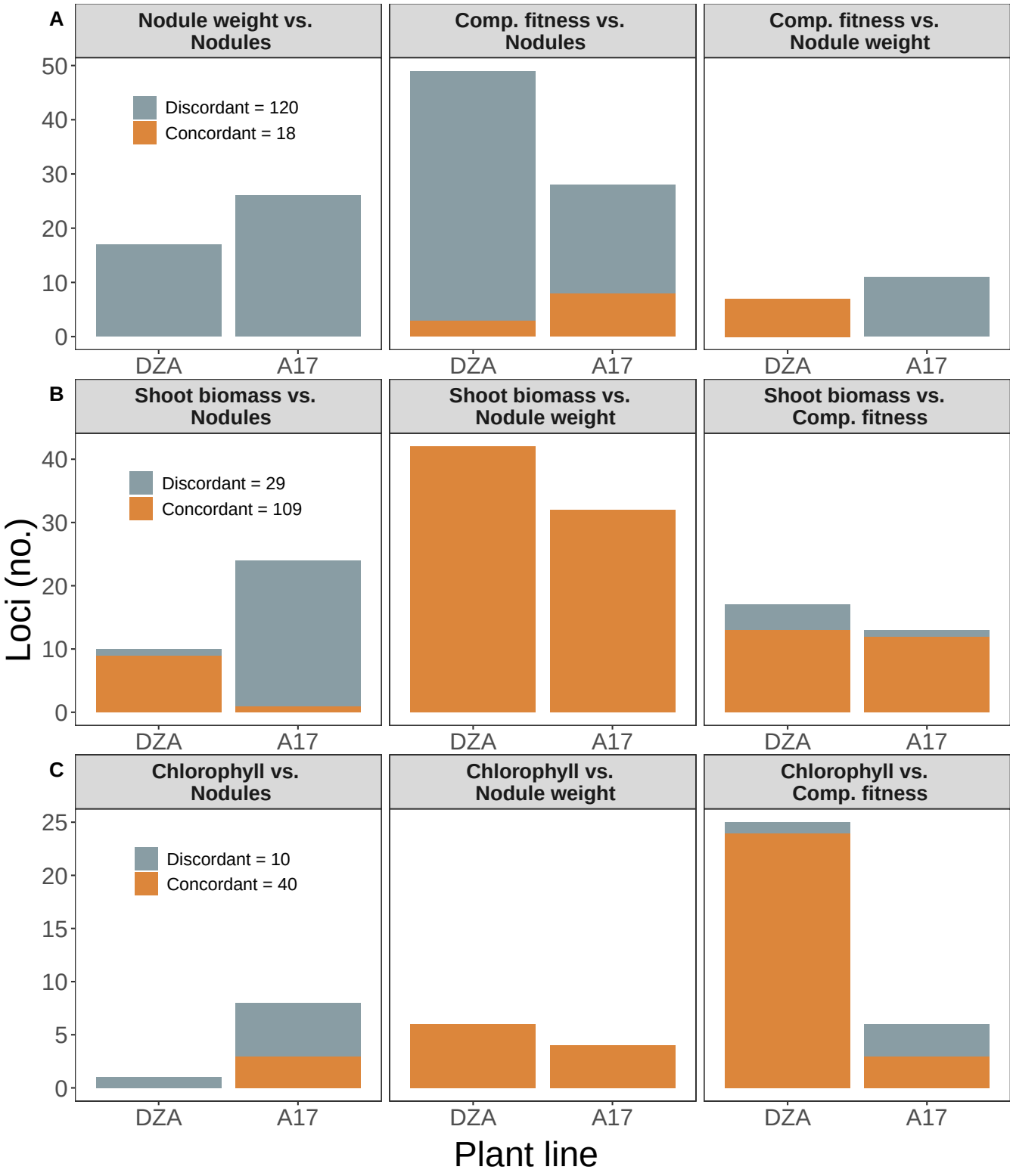

Fig. S2. Rhizobium fitness (A) and symbiotic pleiotropy (B,C) dominated by variants with discordant (grey) or concordant (orange) effects, respectively. Variants significantly associated with two or more rhizobium fitness proxies are considered to underlie rhizobium fitness pleiotropy while those associated with both plant and rhizobium fitness proxies underlying symbiotic pleiotropy. Pleiotropic variant counts with same sign (i.e., concordant, orange) or opposite sign (i.e., discordant, grey) effects on both traits appear in the leftmost panels in A-C.

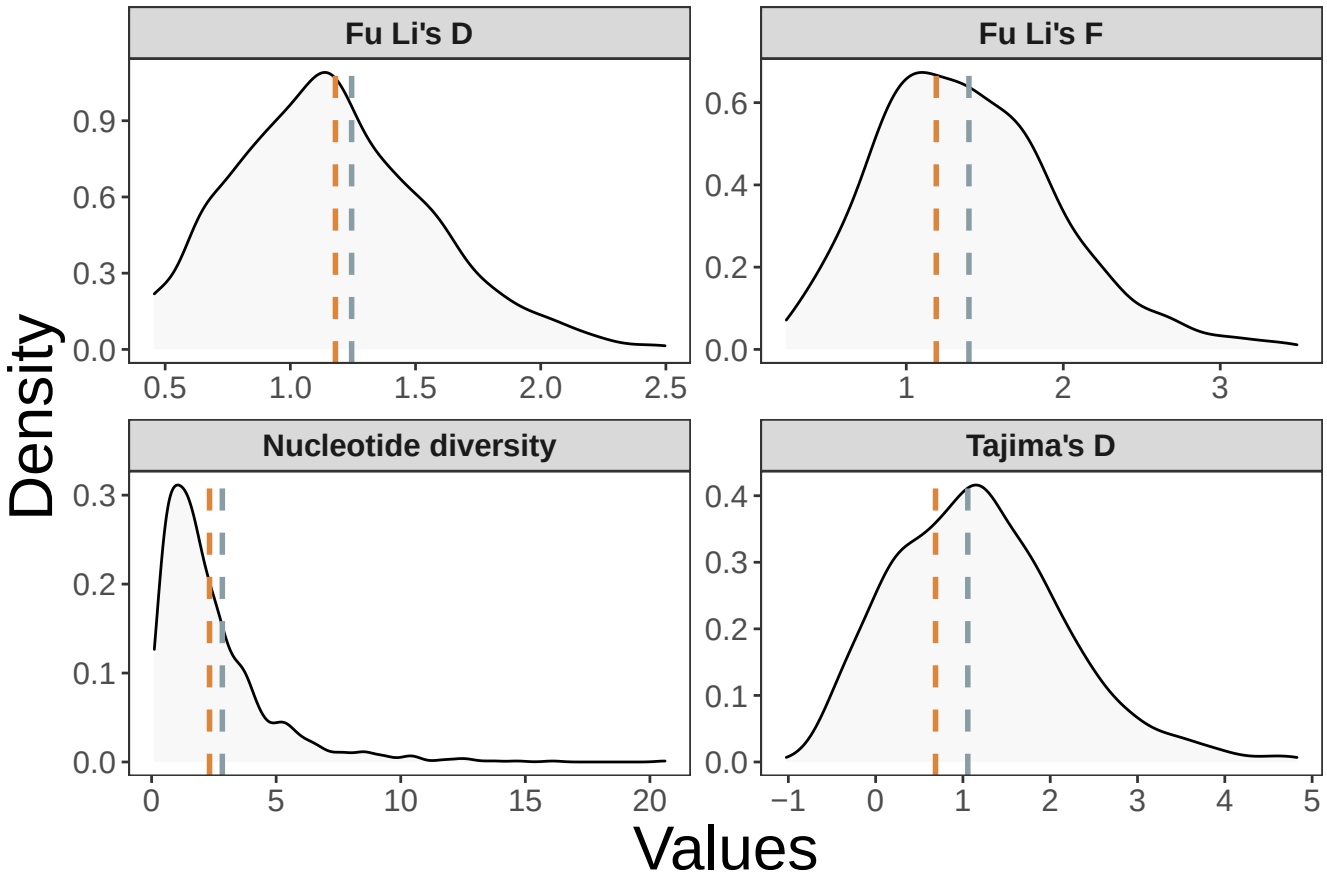

Fig. S3. Neutrality statistics do not significantly differ from null for genes underlying rhizobium fitness pleiotropy. Vertical lines represent the average values calculated for four separate statistics, grey for discordant and orange for concordant genes. Distributions represent the same statistics calculated for all genes containing significant variants based on GWAS. Dashed and solid lines represent non-significant ($p > 0.1$) and significant ($p < 0.1$) differences, respectively, between each focal gene category and all significant genes (i.e., distributions).

Table S1. Neutrality statistics are elevated only for genes underlying fitness alignment. Four different test statistics were calculated for six different gene categories: all genes on the pSymA or pSymB plus one on the chromosome (N = 3,103; “all” genes), genes tagged by variants included in our GWAS (N = 1726; “GWAS” genes), genes containing variants significantly associated with one or more traits (N = 1051; “sig” genes), and our focal genes of interest, which contain at least one pleiotropic variant underlying rhizobium fitness or symbiotic pleiotropy (N = 198 total: for rhizobium fitness, N = 10 with concordant and 85 discordant effects; for symbiotic pleiotropy, 97 with concordant and 21 with discordant effects). ‘Average’ represents the average value of the test statistic within each category. ‘Prop. NS’ represents the proportion of non-significant t-tests comparing each gene category to a random subset of genes in the "all", "GWAS", and "sig" categories. We use the last comparison (i.e., focal gene categories vs. all significant genes) as our null.

Pleiotropy	Category	Test statistic	Average	Prop. NS (all)	Prop. NS (GWAS)	Prop. NS (sig)	Sig. diff. from null?
NA	GWAS	Fu.Li.D	1.068	0	NA	NA	NA
NA	GWAS	Fu.Li.F	1.302	0.003	NA	NA	NA
NA	GWAS	π	2.069	0	NA	NA	NA
NA	GWAS	Tajima.D	1.148	0.515	NA	NA	NA
NA	sig	Fu.Li.D	1.164	0	NA	NA	NA
NA	sig	Fu.Li.F	1.394	0	NA	NA	NA
NA	sig	π	2.430	0	NA	NA	NA
NA	sig	Tajima.D	1.184	0.316	NA	NA	NA
Rhizobium fitness	concordant	Fu.Li.D	1.180	0.95	0.993	0.999	No
Rhizobium fitness	concordant	Fu.Li.F	1.191	0.984	0.988	0.949	No
Rhizobium fitness	concordant	π	2.325	0.873	1	0.999	No
Rhizobium fitness	concordant	Tajima.D	0.686	0.905	0.83	0.804	No
Rhizobium fitness	discordant	Fu.Li.D	1.245	0.032	0.113	0.871	No
Rhizobium fitness	discordant	Fu.Li.F	1.399	0.593	0.891	0.995	No
Rhizobium fitness	discordant	π	2.838	0.003	0.409	0.893	No
Rhizobium fitness	discordant	Tajima.D	1.056	0.983	0.964	0.941	No
Symbiosis	**concordant**	**Fu.Li.D**	**1.388**	**0**	**0**	**0.006**	**Yes**
Symbiosis	**concordant**	**Fu.Li.F**	**1.642**	**0.001**	**0.001**	**0.062**	**Yes**
Symbiosis	**concordant**	**π**	**3.603**	**0**	**0.003**	**0.025**	**Yes**
Symbiosis	concordant	Tajima.D	1.365	0.508	0.647	0.769	No
Symbiosis	discordant	Fu.Li.D	1.258	0.548	0.662	0.935	No
Symbiosis	discordant	Fu.Li.F	1.356	0.897	0.969	0.981	No
Symbiosis	discordant	π	2.458	0.382	0.885	0.979	No
Symbiosis	discordant	Tajima.D	0.933	0.973	0.963	0.929	No

Table S2. Gene-level summary of pleiotropic variants categorized as underlying rhizobium fitness (N = 99) or symbiotic (N = 128) pleiotropy, and whether such pleiotropy was concordant (i.e., same-sign effects on both proxies) or discordant (i.e., opposite-sign effects on both proxies). The number of variants and associations represents the number of unique pleiotropic variants, and the total number of associations (i.e., a single variant can have multiple associations), respectively. The 'Max effect size' represents the absolute maximum effect size of a variant tagged within that gene. 'Host genotype' refers to the host in which the variant effects were significant, either on plant line DZA, A17, or both.

Region	RefSeq ID	Putative function	Variants (no.)	Associations (no.)	Max effect size	Host genotype	Rhizobium fitness	Symbiosis
pSymB	NP_436880.1	hypothetical protein	2	5	0.308	A17	concordant	discordant
pSymB	NP_436905.1	aspartate aminotransferase family protein	2	4	0.330	A17	concordant	NA
pSymB	NP_437823.1	hypothetical protein	1	2	0.389	A17	concordant	NA
pSymB	NP_436558.1	ABC transporter permease	1	4	0.373	A17	discordant	mixed
pSymB	NP_436561.1	sugar ABC transporter substrate-binding protein	1	3	0.373	A17	discordant	mixed
pSymB	NP_436633.1	lysylphosphatidylglycerol synthetase family protein	2	6	0.361	A17	discordant	mixed
pSymB	NP_436643.1	FAD-dependent oxidoreductase	1	4	0.399	A17	discordant	mixed
pSymB	NP_437001.1	ABC transporter ATP-binding protein	1	3	0.345	A17	discordant	mixed
pSymA	NA	hypothetical protein	1	2	0.300	A17	discordant	NA
pSymA	NP_435272.1	AGE family eπmerase/isomerase	1	3	0.227	A17	discordant	NA
pSymA	NP_435292.1	BA14K family protein	1	2	0.392	A17	discordant	NA
pSymA	NP_435377.1	fumarylacetoacetate hydrolase	1	2	0.303	A17	discordant	NA
pSymA	NP_435378.1	TRAP transporter permease DctQ	1	2	0.337	A17	discordant	NA
pSymA	NP_435547.2	hypothetical protein	1	3	0.409	A17	discordant	NA
pSymA	NP_435911.1	cytochrome c oxidase subunit 1	1	3	0.271	A17	discordant	NA
pSymA	NP_435919.1	nitrate reductase	1	3	0.239	A17	discordant	NA
pSymA	NP_436056.1	oxidoreductase	1	2	0.291	A17	discordant	NA
pSymA	NP_436085.1	cation:proton antiporter	1	2	0.230	A17	discordant	NA
pSymA	NP_436089.2	GGDEF domain-containing protein	1	2	0.258	A17	discordant	NA
pSymA	NP_436397.2	hemolysin-type calcium-binding protein	2	6	0.335	A17	discordant	NA
pSymA	NP_436518.1	molybdopterin dehydrogenase	2	4	0.276	A17	discordant	NA
pSymA	NP_436519.2	oxidoreductase	1	2	0.347	A17	discordant	NA
pSymB	NP_436594.1	alpha/beta hydrolase	1	2	0.378	A17	discordant	NA
pSymB	NP_436599.1	SAM-dependent methyltransferase	1	2	0.292	A17	discordant	NA
pSymB	NP_436628.1	protein-L-isoaspartate O-methyltransferase	1	2	0.340	A17	discordant	NA
pSymB	NP_436773.1	glycosyl transferase	1	2	0.164	A17	discordant	NA
pSymB	NP_436856.1	gfo/Idh/MocA family oxidoreductase	2	4	0.178	A17	discordant	NA
pSymB	NP_437315.1	nucleotide-binding protein	1	2	0.247	A17	discordant	NA
pSymB	NP_437322.1	adenine deaminase	1	2	0.234	A17	discordant	NA
pSymB	NP_437324.1	phosphoribosyltransferase	1	2	0.221	A17	discordant	NA
pSymB	NP_437327.1	allantoicase	1	2	0.260	A17	discordant	NA
pSymB	NP_437330.1	xanthine dehydrogenase small subunit	1	2	0.303	A17	discordant	NA
pSymB	NP_437331.1	xanthine dehydrogenase molybdopterin binding subunit	1	2	0.270	A17	discordant	NA
pSymB	NP_437386.1	LysR family transcriptional regulator	2	4	0.302	A17	discordant	NA
pSymB	NP_437405.1	sigma-54-dependent Fis family transcriptional regulator	1	2	0.271	A17	discordant	NA
pSymB	NP_437417.1	phosphoethanolamine transferase	1	2	0.319	A17	discordant	NA
pSymB	NP_437418.1	hypothetical protein	1	2	0.280	A17	discordant	NA
pSymB	YP_001312350.1	transporter	1	2	0.184	A17	discordant	NA
pSymB	YP_001313347.1	NCS2 family permease	1	2	0.209	A17	discordant	NA
pSymA	YP_001314404.1	hypothetical protein	1	2	0.273	A17	discordant	NA
pSymA	YP_001314405.1	glycerol transporter	1	2	0.325	A17	discordant	NA
pSymA	YP_001314499.1	GntR family transcriptional regulator	1	2	0.252	A17	discordant	NA
pSymA	NP_435247.1	formate dehydrogenase-N subunit alpha	1	2	0.257	A17	NA	concordant
pSymA	NP_435250.1	formate dehydrogenase accessory protein FdhE	4	8	0.279	A17	NA	concordant
pSymA	NP_435251.1	L-selenocysteinyl-tRNA(Sec) synthase	3	6	0.291	A17	NA	concordant
pSymA	NP_435253.1	selenocysteine-specific translation elongation factor	1	2	0.356	A17	NA	concordant
pSymA	NP_435307.1	cytochrome C	1	2	0.214	A17	NA	concordant
pSymA	NP_435593.1	hypothetical protein	2	4	0.327	A17	NA	concordant
pSymA	NP_435594.1	transcriptional regulator	1	2	0.298	A17	NA	concordant
pSymA	NP_435599.1	HlyD family secretion protein	1	3	0.166	A17	NA	concordant
pSymA	NP_435610.2	hypothetical protein	1	2	0.191	A17	NA	concordant
pSymA	NP_435697.1	nitrogenase molybdenum-iron protein subunit beta	1	3	0.270	A17	NA	concordant
pSymA	NP_435710.1	nodulation protein H	1	2	0.186	A17	NA	concordant
pSymA	NP_436262.1	tartrate dehydrogenase	1	2	0.195	A17	NA	concordant
pSymA	NP_436462.1	integrase	2	4	0.333	A17	NA	concordant
pSymB	NP_436553.1	alpha/beta hydrolase	1	2	0.296	A17	NA	concordant
pSymB	NP_436569.1	tripartite tricarboxylate transporter TctB family protein	1	3	0.199	A17	NA	concordant
pSymB	NP_436640.1	LLM class F420-dependent oxidoreductase	1	3	0.237	A17	NA	concordant
pSymB	NP_436646.1	FadR family transcriptional regulator	1	3	0.304	A17	NA	concordant
pSymB	NP_436678.1	amidohydrolase	1	2	0.271	A17	NA	concordant
pSymB	NP_436774.1	CDP-paratose 2-eπmerase	1	2	0.165	A17	NA	concordant
pSymB	NP_436981.1	UDP-glucose 4-eπmerase GalE	1	3	0.276	A17	NA	concordant
pSymB	NP_437117.1	cell surface protein	1	2	0.254	A17	NA	concordant

Table S2 continued from previous page

Region	RefSeq ID	Putative function	Variants (no.)	Associations (no.)	Max effect size	Host genotype	Rhizobium fitness	Symbiosis
pSymB	NP_437160.1	hypothetical protein	1	2	0.190	A17	NA	concordant
pSymB	NP_437221.1	IclR family transcriptional regulator	1	4	0.163	A17	NA	concordant
pSymB	NP_437366.1	beta-N-acetylhexosaminidase	1	2	0.364	A17	NA	concordant
pSymB	NP_437398.1	fructose-bisphosphate aldolase class II	1	2	0.331	A17	NA	concordant
pSymB	NP_437637.1	ABC transporter ATP-binding protein	1	2	0.195	A17	NA	concordant
pSymB	NP_437662.1	hypothetical protein	1	3	0.213	A17	NA	concordant
pSymB	YP_001313261.1	two-component system response regulator	1	2	0.173	A17	NA	concordant
pSymA	NP_435314.1	fatty acid desaturase	1	2	0.176	A17	NA	discordant
pSymA	NP_435598.1	hypothetical protein	1	2	0.187	A17	NA	discordant
pSymA	NP_436261.1	aldehyde dehydrogenase	1	2	0.281	A17	NA	discordant
pSymB	NP_436546.1	pyrroline-5-carboxylate reductase	1	2	0.161	A17	NA	discordant
pSymB	NP_436984.1	beta-mannosidase	1	2	0.245	A17	NA	discordant
pSymB	NP_438036.2	sulfite oxidase-like oxidoreductase	1	3	0.321	A17	NA	discordant
pSymB	NP_436648.1	peptide ABC transporter substrate-binding protein	2	6	0.309	A17	NA	mixed
pSymB	NP_437804.1	Zn-dependent hydrolase	2	5	0.381	Both	concordant	discordant
pSymA	NP_435627.1	IclR family transcriptional regulator	2	6	0.208	Both	discordant	concordant
pSymB	NP_436566.1	5-oxoprolinase	2	4	0.444	Both	discordant	concordant
pSymB	NP_437100.1	ferredoxin reductase	2	4	0.275	Both	discordant	concordant
pSymB	NP_437305.1	peptide ABC transporter substrate-binding protein	2	4	0.329	Both	discordant	concordant
pSymB	YP_001313723.1	membrane assembly protein AsmA	2	4	0.335	Both	NA	mixed
pSymA	NP_435563.1	hypothetical protein	2	4	0.306	Both	discordant	concordant
pSymB	NP_437004.1	N-acetylglutaminylglutamine synthetase	4	10	0.470	Both	discordant	concordant
pSymA	YP_001313905.1	amino acid decarboxylase	3	7	0.385	Both	discordant	concordant
pSymB	NP_436652.1	EamA/RhaT family transporter	1	4	0.301	Both	discordant	discordant
pSymB	NP_436701.1	hypothetical protein	1	4	0.349	Both	discordant	discordant
pSymB	NP_436950.1	NAD-dependent succinate-semialdehyde dehydrogenase	1	4	0.559	Both	discordant	discordant
pSymB	NP_436967.1	L-idonate 5-dehydrogenase	1	4	0.370	Both	discordant	discordant
pSymB	NP_436974.1	hypothetical protein	1	4	0.452	Both	discordant	discordant
pSymB	NP_436975.1	gluconolactonase	2	9	0.423	Both	discordant	discordant
pSymB	NP_437017.1	hypothetical protein	2	6	0.566	Both	discordant	discordant
pSymB	WP_003525897.1	N-alpha-acetyl diaminobutyric acid deacetylase DoeB	1	4	0.636	Both	discordant	discordant
pSymB	NP_436767.1	hypothetical protein	2	5	0.489	Both	discordant	NA
pSymB	NP_437206.1	Ti-type conjugative transfer relaxase TraA	4	9	0.584	Both	mixed	concordant
pSymA	NP_435603.2	calcium/proton exchanger	3	11	0.210	Both	NA	concordant
pSymB	NP_436651.1	ABC transporter ATP-binding protein	2	7	0.307	Both	discordant	discordant
pSymB	NP_436650.1	ABC transporter permease	2	9	0.358	Both	discordant	mixed
pSymB	NP_436655.1	dihydroxy-acid dehydratase	2	8	0.339	Both	discordant	concordant
pSymA	NP_436524.1	amidase	2	5	0.288	DZA	concordant	concordant
pSymA	NP_436481.1	penicillin-binding protein 1A	1	2	0.263	DZA	concordant	NA
pSymB	NP_436738.1	pyrroloquinoline quinone biosynthesis protein B	1	3	0.227	DZA	concordant	NA
pSymB	NP_437079.1	copper chaperone	1	2	0.239	DZA	concordant	NA
pSymB	NP_437156.1	DNA ligase D	1	2	0.408	DZA	concordant	NA
pSymB	NP_437856.1	cytochrome ubiquinol oxidase subunit I	1	2	0.429	DZA	concordant	NA
pSymA	NP_435596.1	hypothetical protein	3	6	0.255	DZA	discordant	concordant
pSymA	NP_435597.2	AI-2E family transporter	4	9	0.208	DZA	discordant	concordant
pSymA	NP_435386.1	alcohol dehydrogenase	1	2	0.209	DZA	discordant	NA
pSymA	NP_435662.1	nodulation protein NoeA	1	2	0.204	DZA	discordant	NA
pSymA	NP_436112.1	adenylate cyclase	1	3	0.386	DZA	discordant	NA
pSymA	NP_436119.2	LysR family transcriptional regulator	1	2	0.358	DZA	discordant	NA
pSymA	NP_436135.2	nuclear transport factor 2 family protein	1	2	0.314	DZA	discordant	NA
pSymA	NP_436516.2	class I SAM-dependent methyltransferase	1	3	0.510	DZA	discordant	NA
pSymB	NP_436547.1	LysR family transcriptional regulator	1	3	0.399	DZA	discordant	NA
pSymB	NP_436564.1	Lrp/AsnC family transcriptional regulator	1	2	0.326	DZA	discordant	NA
pSymB	NP_436573.1	RNA polymerase sigma factor SigJ	1	3	0.263	DZA	discordant	NA
pSymB	NP_436656.1	amylo-alpha-1%2C6-glucosidase	2	6	0.412	DZA	discordant	NA
pSymB	NP_436680.1	dihydrodiπcolinate synthase family protein	1	3	0.454	DZA	discordant	NA
pSymB	NP_436709.1	hypothetical protein	1	2	0.323	DZA	discordant	NA
pSymB	NP_436710.1	S-(hydroxymethyl)glutathione dehydrogenase	1	2	0.331	DZA	discordant	NA
pSymB	NP_436766.1	ABC transporter ATP-binding protein	1	2	0.448	DZA	discordant	NA
pSymB	NP_436789.1	dihydrodiπcolinate synthase family protein	1	2	0.209	DZA	discordant	NA
pSymB	NP_436791.1	malate dehydrogenase	1	2	0.267	DZA	discordant	NA
pSymB	NP_436797.1	amino acid dehydrogenase	1	2	0.343	DZA	discordant	NA
pSymB	NP_436798.1	hydroxyproline-2-eπmerase	1	2	0.378	DZA	discordant	NA
pSymB	NP_437053.1	sugar phosphate isomerase/eπmerase	1	3	0.358	DZA	discordant	NA
pSymB	NP_437093.2	ABC transporter ATP-binding protein	1	3	0.256	DZA	discordant	NA
pSymB	NP_437094.1	hypothetical protein	1	2	0.397	DZA	discordant	NA
pSymB	NP_438092.1	2-(1%2C2-epoxy-1%2C2-dihydrophenyl)acetyl-CoA isomerase	1	2	0.250	DZA	discordant	NA
pSymA	WP_014990337.1	hypothetical protein	1	2	0.169	DZA	discordant	NA

Table S2 continued from previous page

Region	RefSeq ID	Putative function	Variants (no.)	Associations (no.)	Max effect size	Host genotype	Rhizobium fitness	Symbiosis
pSymB	YP_001312688.1	S-formylglutathione hydrolase	1	2	0.238	DZA	discordant	NA
pSymB	YP_002122330.1	DUF899 domain-containing protein	1	2	0.519	DZA	discordant	NA
Chromosome	NP_386155.1	sorbosone dehydrogenase	1	3	0.423	DZA	mixed	NA
pSymB	NP_436679.1	DUF2243 domain-containing protein	2	6	0.286	DZA	mixed	NA
pSymB	NP_436713.1	PQQ-dependent dehydrogenase%2C methanol/ethanol family	1	4	0.423	DZA	mixed	NA
pSymA	NP_435259.1	hypothetical protein	1	2	0.244	DZA	NA	concordant
pSymA	NP_435294.1	D-aminopeptidase	1	2	0.256	DZA	NA	concordant
pSymA	NP_435306.1	hypothetical protein	1	2	0.328	DZA	NA	concordant
pSymA	NP_435317.2	AI-2E family transporter	1	2	0.246	DZA	NA	concordant
pSymA	NP_435318.2	bifunctional diguanylate cyclase/phosphodiesterase	1	2	0.259	DZA	NA	concordant
pSymA	NP_435324.1	hypothetical protein	1	2	0.276	DZA	NA	concordant
pSymA	NP_435326.1	malonyl-CoA synthase	1	3	0.204	DZA	NA	concordant
pSymA	NP_435334.1	hypothetical protein	1	2	0.168	DZA	NA	concordant
pSymA	NP_435343.1	MFS transporter	1	2	0.185	DZA	NA	concordant
pSymA	NP_435368.1	glutamate dehydrogenase	1	2	0.237	DZA	NA	concordant
pSymA	NP_435372.1	dihydroxy-acid dehydratase	1	2	0.259	DZA	NA	concordant
pSymA	NP_435559.2	alpha/beta hydrolase	1	2	0.281	DZA	NA	concordant
pSymA	NP_435560.2	adenylate cyclase	1	2	0.299	DZA	NA	concordant
pSymA	NP_435567.1	cytochrome c oxidase%2C cbb3-type subunit II	1	3	0.185	DZA	NA	concordant
pSymA	NP_435571.1	copper-translocating P-type ATPase	1	2	0.237	DZA	NA	concordant
pSymA	NP_435604.2	amino acid ABC transporter substrate-binding protein	1	3	0.196	DZA	NA	concordant
pSymA	NP_435615.1	carbamate kinase	1	2	0.211	DZA	NA	concordant
pSymA	NP_435621.1	sugar ABC transporter substrate-binding protein	2	4	0.388	DZA	NA	concordant
pSymA	NP_435623.1	carbohydrate ABC transporter permease	2	4	0.224	DZA	NA	concordant
pSymA	NP_435628.1	NAD(P)-dependent oxidoreductase	1	2	0.308	DZA	NA	concordant
pSymA	NP_435641.1	molecular chaperone GroEL	1	2	0.189	DZA	NA	concordant
pSymA	NP_435654.1	transcriptional regulator	1	2	0.171	DZA	NA	concordant
pSymA	NP_435748.1	Ti-type conjugative transfer system protein TraG	1	2	0.206	DZA	NA	concordant
pSymA	NP_435752.1	mechanosensitive ion channel family protein	1	2	0.452	DZA	NA	concordant
pSymA	NP_435753.1	sensor histidine kinase	2	4	0.384	DZA	NA	concordant
pSymA	NP_435978.2	sugar ABC transporter permease	1	2	0.464	DZA	NA	concordant
pSymA	NP_435981.1	ABC transporter ATP-binding protein	1	3	0.291	DZA	NA	concordant
pSymA	NP_435985.1	sugar phosphate isomerase/eπmerase	1	2	0.404	DZA	NA	concordant
pSymA	NP_436006.1	LysR family transcriptional regulator	1	2	0.403	DZA	NA	concordant
pSymA	NP_436072.1	NADH-quinone oxidoreductase subunit I 2	2	5	0.314	DZA	NA	concordant
pSymA	NP_436073.1	histidine kinase	1	2	0.203	DZA	NA	concordant
pSymA	NP_436166.2	amidohydrolase family protein	1	3	0.319	DZA	NA	concordant
pSymA	NP_436209.1	TonB-dependent siderophore receptor	1	2	0.276	DZA	NA	concordant
pSymA	NP_436219.1	aminotransferase	1	2	0.384	DZA	NA	concordant
pSymA	NP_436244.1	MBL fold metallo-hydrolase	1	2	0.197	DZA	NA	concordant
pSymA	NP_436245.1	MFS transporter	1	2	0.201	DZA	NA	concordant
pSymA	NP_436271.1	ABC transporter permease	1	3	0.334	DZA	NA	concordant
pSymA	NP_436482.1	penicillin-binding protein 2	1	3	0.350	DZA	NA	concordant
pSymB	NP_436559.1	ABC transporter permease	1	3	0.284	DZA	NA	concordant
pSymB	NP_436799.1	hypothetical protein	1	2	0.209	DZA	NA	concordant
pSymB	NP_436969.1	GGDEF-domain containing protein	1	2	0.384	DZA	NA	concordant
pSymB	NP_437002.1	FadR family transcriptional regulator	2	6	0.215	DZA	NA	concordant
pSymB	NP_437003.1	N-acetylglutaminylglutamine amidotransferase	1	2	0.193	DZA	NA	concordant
pSymB	NP_437007.1	sugar ABC transporter ATP-binding protein	1	3	0.214	DZA	NA	concordant
pSymB	NP_437015.1	3-ketoacyl-ACP reductase	1	3	0.290	DZA	NA	concordant
pSymB	NP_437016.1	LysR family transcriptional regulator	1	3	0.228	DZA	NA	concordant
pSymB	NP_437030.1	arabinose ABC transporter substrate-binding protein	1	2	0.197	DZA	NA	concordant
pSymB	NP_437035.1	chemotaxis protein CheB	1	2	0.262	DZA	NA	concordant
pSymB	NP_437145.1	hypothetical protein	1	2	0.290	DZA	NA	concordant
pSymB	NP_437243.1	acetyl/proπonyl/methylcrotonyl-CoA carboxylase subunit alpha	1	2	0.249	DZA	NA	concordant
pSymB	NP_437510.1	sulfoacetaldehyde acetyltransferase	1	3	0.355	DZA	NA	concordant
pSymB	NP_437900.1	asparagine synthetase B	1	2	0.407	DZA	NA	concordant
pSymA	YP_001312338.1	sugar ABC transporter permease	1	2	0.206	DZA	NA	concordant
pSymA	YP_001313599.1	cell envelope protein	3	7	0.266	DZA	NA	concordant
pSymA	YP_001314274.1	N-acetyltransferase	2	6	0.345	DZA	NA	concordant
pSymA	YP_001314472.1	mandelate racemase	1	2	0.245	DZA	NA	concordant
pSymA	NP_436495.1	ATP-dependent helicase	1	2	0.279	DZA	NA	discordant
pSymB	NP_437005.1	transcriptional regulator	1	2	0.232	DZA	NA	discordant
pSymB	NP_437011.1	carbohydrate kinase	1	2	0.258	DZA	NA	discordant
pSymB	NP_437792.1	esterase	1	2	0.302	DZA	NA	discordant
pSymA	NP_435605.1	putrescine-ornithine antiporter	2	4	0.236	DZA	NA	mixed
pSymA	NP_435751.1	Ti-type conjugative transfer relaxase TraA	2	5	0.284	DZA	NA	mixed

SI Datasets

SI Dataset S1 (all_gene_stats.wREADME.xlsx)

Summary table for all genes on pSymA and pSymB (as well as one on the chromosome) in which neutrality statistics were calculated. The first tab "all_gene_stats" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.

SI Dataset S2 (pleio_vars.ps_wREADME.xlsx)

Summary table for all variants underlying rhizobium fitness (N = 130) or symbiotic (N = 168) pleiotropy. The first tab "pleio_vars.ps" provides variant-level information, while the "README" tab provides a brief description of each column in the first tab.

SI Dataset S3 (pleio_vars.gene_Uniprot_wREADME.xlsx)

Summary table for all genes underlying rhizobium fitness (N = 99) or symbiotic (N = 128) pleiotropy. The first tab "pleio_vars.gene" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.

References

1. DG Barker, et al., Growing *M . truncatula*: choice of substrates and growth conditions (2006).
2. P Somasegaran, HJ Hoben, Methods in Legume-Rhizobium Technology in *Handbook for Rhizobia.* (Springer, New York, NY, US), (1994).
3. B Bucciarelli, J Hanan, D Palmquist, CP Vance, A standardized method for analysis of *Medicago truncatula* phenotypic development. *Plant Physiol.* **142**, 207–219 (2006).
4. D Kessner, TL Turner, J Novembre, Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. *Mol. Biol. Evol.* **30**, 1145–1158 (2013).
5. H Li, R Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
6. E Garrison, G Marth, Haplotype-based variant detection from short-read sequencing. *arXiv preprint arXiv:1207.3907* (2012).
7. P Danecek, et al., The variant call format and vcftools. *Bioinformatics* **27**, 2156–2158 (2011).
8. B Epstein, et al., Genome-wide association analyses in the model rhizobium *Ensifer meliloti. MSphere* **3** (2018).
9. A Riley, M Grillo, B Epstein, P Tiffin, K Heath, Partners in space: Discordant population structure between legume hosts and rhizobium symbionts in their native range (bioRχiv. https://www.biorxiv.org/content/10.1101/2021.06.30.449460v1) (2021).
10. GA Churchill, RW Doerge, Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971 (1994).
11. P Cingolani, et al., A program for annotating and predicting the effects of single nucleotide polymorphisms, snpeff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* **6**, 80–92 (2012).
12. AR Quinlan, IM Hall, Bedtools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
13. MF Dunn, Key roles of microsymbiont amino acid metabolism in rhizobia-legume interactions. *Critical Rev. Microbiol.* **41**, 411–451 (2015).