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 October 27, 2021

Given the need to predict the outcomes of evolution in host-associated microbiomes, whether microbial and host fitnesses trade off, generating conflict, remains a pressing question. Examining the relationships between host and microbe fitness proxies at both the phenotypic and genomic levels can illuminate the mechanisms underlying cooperation and conflict, while addressing the challenge of measuring microbial fitness. We examined naturally-occurring genetic variation in 191 strains of the model microbial symbiont, *Ensifer meliloti*, paired with each of two host *Medicago truncatula* genotypes in single- or mixed-strain inoculation experiments to determine how multiple proxies of microbial and host fitness were related to one another and test key predictions about mutualism evolution at the genomic scale. We found little evidence for fitness conflict; loci tended to have concordant effects on both microbe and host fitnesses, even in environments with multiple competing microbes. Our results emphasize the importance of quantifying microbial competitive fitness for understanding microbiome evolution and thus harnessing microbiomes to improve host fitness. Additionally, we find that mutualistic coevolution between legumes and rhizobia can act to maintain, rather than erode, genetic diversity, potentially explaining why variation in mutualism 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). Studies of the G→P map in single organisms have highlighted the importance of pleiotropy, i.e., loci that influence more than one trait, with important consequences for development, aging, and disease (4, 5). Pleiotropy can constrain adaptation when it underlies trade-offs among fitness-related traits of the same organism (6–9). In symbiotic mutualisms, wherein partners trade fitness benefits (10), pleiotropic loci can also underlie traits of interacting individuals, which we henceforth refer to as **symbiotic pleiotropy**.

In the G→P map of symbiotic traits, 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 (11–16). In contrast, 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. Identifying loci underlying symbiotic pleiotropy is therefore key for illuminating the genetic basis of mutualism traits and predicting how they evolve in nature.

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 (17–23). 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 (24, 25). For example, loci underlying fitness conflict might resemble those underlying antagonistic arms race dynamics, showing signatures of rapid turnover and higher rates of evolutionary change as rare alleles are selectively favoured (26, 27). In contrast, loci underlying fitness alignment may exhibit signatures of evolutionary stability and lower rates of evolutionary change as more common alleles are selectively favoured (27). Because GWAS has the power to reveal individual allelic effects, it can be used to identify loci underlying fundamental sources of conflict at the genomic-level even if fitness alignment is realized at the phenotypic-level.

Over the past ∼25 years, legume-rhizobium symbioses have been developed as models for understanding mutualism evolution (28–32). 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 (33). 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 known to be influenced by variants in the genomes of both host and microbes, as well as the epistatic interactions between them (Genotype-by-Genotype or G x G interactions; e.g., 20, 28, 34–38), making this symbiosis an excellent model for understanding mutualistic coevolution.

A quantitative comparison of single-strain and multi-strain

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

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, or the fitness benefit a rhizobium genotype confers to its host (e.g., shoot biomass), and rhizobium fitness (e.g., nodule number) were interpreted as evidence for conflict between host and symbiont fitnesses (35, 39, 40). In contrast, mutant and experimental evolution studies have suggested a lack of rhizobial mutations underlying conflict (23, 41). 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 (42). Thus detecting conflict between hosts and rhizobia requires multi-strain inoculation experiments wherein the benefits to individual strains in a mixed population can be decoupled from overall plant growth. 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 (19, 43) and hint at a surprising lack of correlation between nodule number and competitive fitness, motivating our comprehensive analysis here.

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 same 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. For some proxies, but not others, these patterns depended on the host genotype with which rhizobia interacted (**Fig. 1**; see **SI Fig. S1** for all genetic correlations and **SI Fig. S2A** for all pleiotropic variant counts).

First, in 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. S2A** 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, suggesting that conflict among these traits is largely resolved in natural rhizobium populations (**Fig. 2** & **SI Fig. S3**; see **SI Fig. S1** for all genetic correlations, and **SI Fig. S2B,C** for all pleiotropic variant counts).

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 (**SI Fig. 3A** left). However we saw a significantly negative correlation between nodule number and chlorophyll A for DZA (**SI Fig. 3A** left). Assuming chlorophyll A indicates the N-fixation efficiency of each strain (44–46), 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 (**SI Fig. 3A** right).

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 (**SI Fig. 3B** left). At

DRAFT

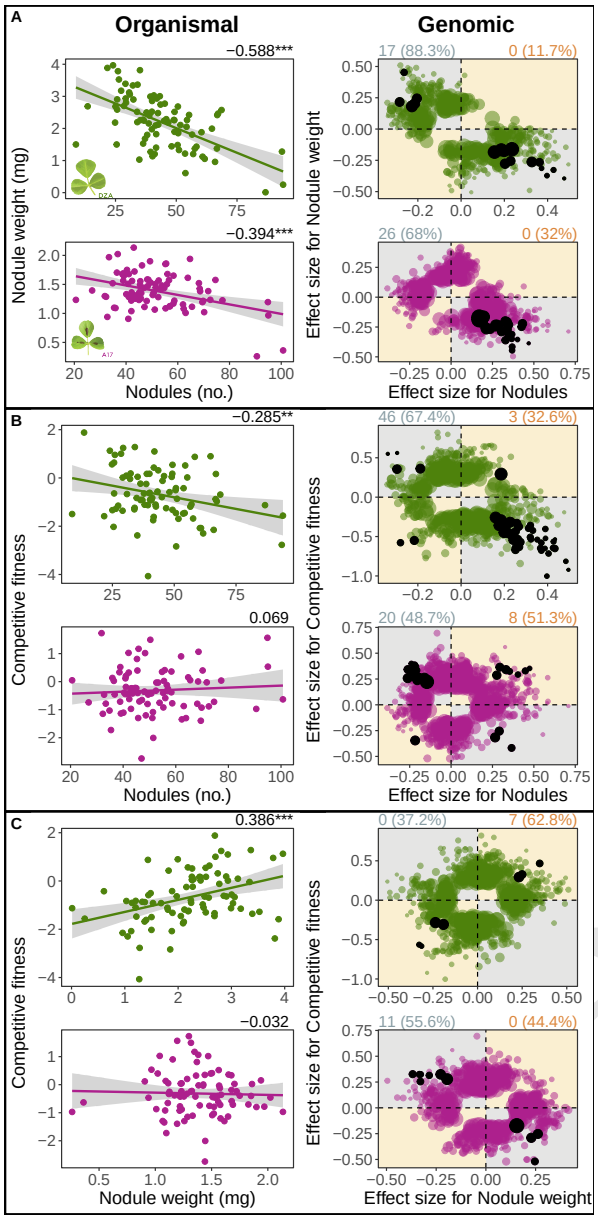

Fig. 1. Trade-offs among rhizobia (*Ensifer meliloti*) fitness proxies prevail at both the organismal (left panels) and genomic-levels (right panels). Organismal: Genetic correlations between pairwise fitness proxies correspond to traits measured on plant lines DZA (green, top rows) or A17 (pink, bottom 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).

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 (**SI Fig. 3B** right), on both hosts.

Finally, competitive fitness and shoot biomass were significantly positively correlated for DZA but not for A17 (**Fig. 2C** left), while competitive fitness and chlorophyll A content were significantly positive on both hosts (**SI Fig. 3C** left). At the genomic-level, most pleiotropic 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 (**SI Fig. 3C** right). Variant effects were equally mixed on A17 for competitive fitness and chlorophyll A (**SI Fig. 3C** right).

Selection acts differently on genes underlying alignment versus conflict. Using multiple diversity and neutrality metrics, we show that rhizobium genes underlying alignment exhibit higher nucleotide diversity and 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). 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. S4**, all comparisons > 0.1; **SI Dataset S1**).

The genetic basis of symbiotic pleiotropy is host-dependent. Comparing the identities and putative functions of variants underlying symbiotic pleiotropy on both host lines revealed little overlap - concordant variants giving rise to both high host and symbiont fitness largely differed between the two host genotypes. Specifically, 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, respectively, while only 15 variants were shared between hosts. We found a similar trend for rhizobium fitness pleiotropy: of 130 variants corresponding to 99 coding genes, 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**). 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 symbiosis traits 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

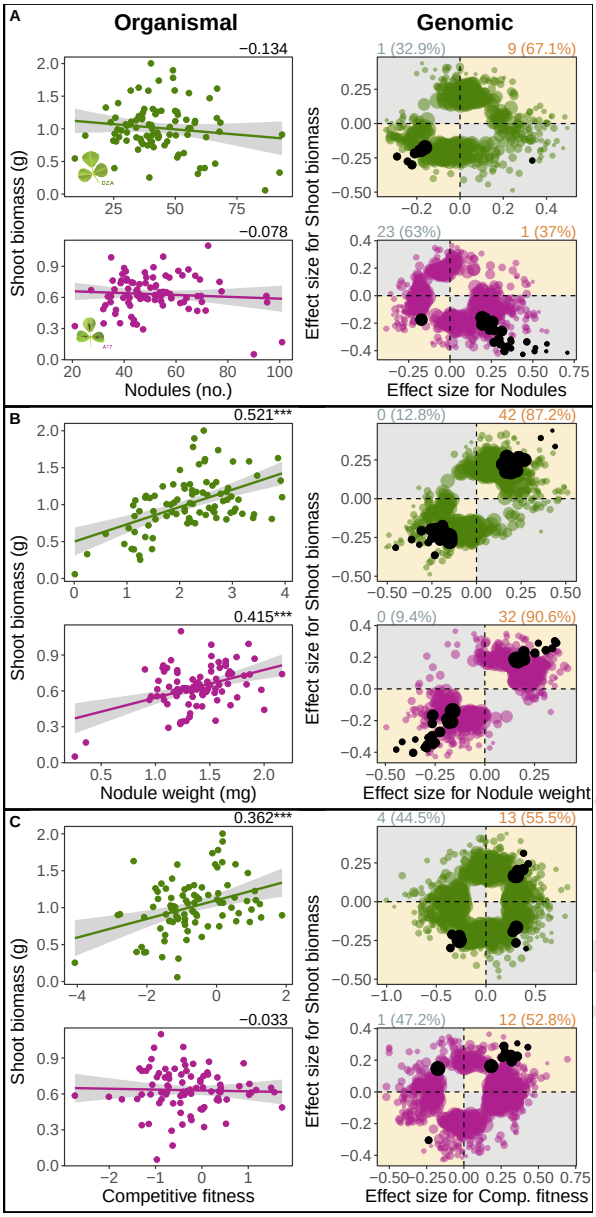

Fig. 2. Fitness alignment between rhizobia (*Ensifer meliloti*) and host (*Medicago truncatula*) prevails at both the organismal (left panels) and genomic-levels (right panels). Organismal: Genetic correlations between pairwise fitness proxies correspond to traits measured on plant lines DZA (green, top rows) or A17 (pink, bottom rows), based on 89 *Ensifer meliloti* strains. Dots represent estimated marginal means for shoot biomass 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).

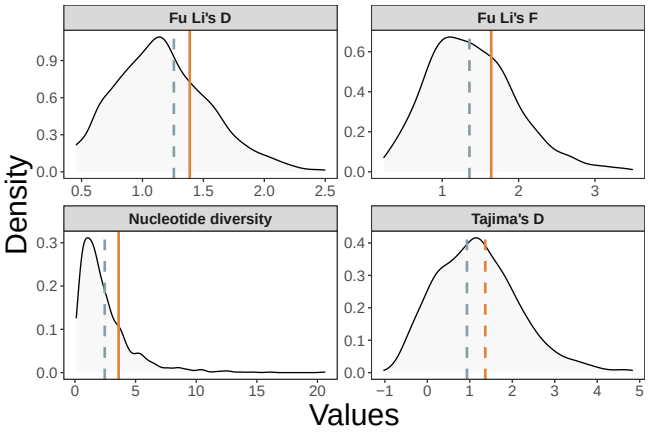

Fig. 3. Neutrality statistics show elevated values for genes underlying fitness alignment. Vertical lines represent the average values calculated for four separate statistics, grey for discordant and orange for concordant genes. Distributions represent the same statistics calculated for all genes containing 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).

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' (47) should favour less-beneficial symbionts (11–13). 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 (48).

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 (11, 25, 49, 50). For example, legume host plants have autoregulation of nodulation to limit the formation of costly nodules once sufficient N levels are achieved (51). No such constraint exists for rhizobia; every nodule formed is expected to lead to greater potential fitness benefits (52). 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 with previous studies (e.g., 35, 40, 53).

In one-to-many mutualisms, such as 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 (11–14, 54). 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 nonexistent), or because of correlational selection that removes discordant variants from the population altogether (55–57). 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. Trait pairs 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 thus 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 (58, 59). 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 (60–63). In a recent in silico GWAS, O'Brien and colleagues (25) 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 pairing rhizobia with five host genotypes (23), and has been found in other legume-rhizobium interactions (e.g., 64). 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 58, 59). If the genotype of the 'best' symbiont differs among host genotypes, then host selection should lead to greater, and not lower, symbiont diversity.

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 (38), and a history of selection shapes the trait variation currently present in natural populations (65, 66). Discordant alleles that are "packaged" 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; 67, 68). Despite the lack of evidence consistent with selection having acted on loci containing discordant variants our experiments were conducted in a single, low N environment and thus, 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 53, 69–71), 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 (19). 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., 20, 34, 35, 38), 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 (51) 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., 35, 72). In mixed inoculation

DRAFT

experiments, however, nodule number for a particular strain depends on their competitive ability due to strain-strain competition and/or host signaling interactions (73, 74), as well as any pre-infection partner choice, i.e., overrepresentation of more beneficial strains in nodules (32, 41, 75, 76). 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.

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 (39, 77). 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 (19, 36) and GWAS revealing distinct sets of candidate genes in different host backgrounds (20, 38); similar variation exists in hosts (78) 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, especially if we want to make predictions about microbiome evolution. 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. (79) and Batstone et al. (38). *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 (79)). *E. meliloti* has a multipartite genome, including a large (∼3.7 Mbp) 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 (80, 81). 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. (38), 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 (61, 79). 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. (19), 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. (38), 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 82) in the R environment (83). We then conducted linear pairwise regressions (lm option in base R) for each trait against the other within each plant line.

DNA isolation, whole-genome sequencing, and variant calling. Detailed methods are provided in Riley et al. (79), 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, 84) 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, 85) 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 (86, 87). Additional details appear in **SI Methods**.

DRAFT

Data Availability. Strains and plant lines are available upon request. All raw data and analysis code are available on GitHub (see "Genetics_conflict_cooperation" folder). 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 and GitHub. 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. FW Stearns, One hundred years of pleiotropy: a retrospective. *Genetics* **186**, 767–773 (2010).
5. GP Wagner, J Zhang, The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nat. Rev. Genet.* **12**, 204–213 (2011).
6. SC Stearns, Trade-offs in life-history evolution. *Funct. Ecol.* **3**, 259–268 (1989).
7. N Barton, Pleiotropic models of quantitative variation. *Genetics* **124**, 773–782 (1990).
8. D Roff, , et al., Trade-offs between growth and reproduction: an analysis of the quantitative genetic evidence. *J. Evol. Biol.* **13**, 434–445 (2000).
9. SP Otto, Two steps forward, one step back: the pleiotropic effects of favoured alleles. *Proc. Royal Soc. B* **271**, 705–714 (2004).
10. JL Bronstein, *Mutualism.* (Oxford University Press, USA), (2015).
11. JJ Bull, W Rice, Distinguishing mechanisms for the evolution of co-operation. *J. Theor. Biol.* **149**, 63–74 (1991).
12. 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).
13. 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).
14. KR Foster, H Kokko, Cheating can stabilize cooperation in mutualisms. *Proc. Royal Soc. B* **273**, 2233–2239 (2006).
15. M Dos Santos, M Ghoul, SA West, Pleiotropy, cooperation, and the social evolution of genetic architecture. *PLoS Biol.* **16**, e2006671 (2018).
16. DC Queller, Pleiotropy and synergistic cooperation. *PLoS Biol.* **17**, e3000320 (2019).
17. ML Friesen, et al., The ecological genomic basis of salinity adaptation in tunisian *Medicago truncatula*. *BMC Genomics* **15**, 1160 (2014).
18. CR Klinger, JA Lau, KD Heath, Ecological genomics of mutualism decline in nitrogen-fixing bacteria. *Proc. Royal Soc. B* **283**, 20152563 (2016).
19. 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).
20. B Epstein, et al., Genome-wide association analyses in the model rhizobium *Ensifer meliloti*. *MSphere* **3** (2018).
21. 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).
22. 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).
23. 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).
24. 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).
25. 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).
26. 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).
27. JL Sachs, CJ Essenberg, MM Turcotte, New paradigms for the evolution of beneficial infections. *Trends Ecol. & Evol.* **26**, 202–209 (2011).
28. MA Parker, Plant fitness variation caused by different mutualist genotypes. *Ecology* **76**, 1525–1535 (1995).
29. RF Denison, Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *The Am. Nat.* **156**, 567–576 (2000).
30. ET Kiers, RA Rousseau, SA West, RF Denison, Host sanctions and the legume–rhizobium mutualism. *Nature* **425**, 78–81 (2003).
31. 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).
32. KD Heath, P Tiffin, Stabilizing mechanisms in a legume–rhizobium mutualism. *Evolution* **63**, 652–662 (2009).
33. CH Foyer, et al., Neglecting legumes has compromised human health and sustainable food production. *Nat. Plants* **2**, 1–10 (2016).
34. J Burdon, P Thrall, Spatial and temporal patterns in coevolving plant and pathogen associations. *The Am. Nat.* **153**, S15–S33 (1999).
35. KD Heath, Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution* **64**, 1446–1458 (2010).
36. KD Heath, PV Burke, JR Stinchcombe, Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Mol. Ecol.* **21**, 4735–4747 (2012).
37. LT Burghardt, et al., Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol. Ecol.* **26**, 6122–6135 (2017).
38. 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).
39. AK Simonsen, JR Stinchcombe, Standing genetic variation in host preference for mutualist microbial symbionts. *Proc. Royal Soc. B* **281**, 20142036 (2014).
40. SS Porter, EL Simms, Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecol. Lett.* **17**, 1121–1129 (2014).
41. ML Friesen, Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytol.* **194**, 1096–1111 (2012).
42. 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).
43. LT Burghardt, B Epstein, P Tiffin, Legacy of prior host and soil selection on rhizobial fitness in planta. *Evolution* **73**, 2013–2023 (2019).
44. D Bullock, D Anderson, Evaluation of the minolta spad-502 chlorophyll meter for nitrogen management in corn. *J. Plant Nutr.* **21**, 741–755 (1998).
45. 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).
46. DJ Weese, KD Heath, BT Dentinger, JA Lau, Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* **69**, 631–642 (2015).
47. G Hardin, The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
48. EI Jones, et al., Cheaters must prosper: reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecol. Lett.* **18**, 1270–1284 (2015).
49. JL Sachs, UG Mueller, TP Wilcox, JJ Bull, The evolution of cooperation. *The Q. Rev. Biol.* **79**, 135–160 (2004).
50. ME Frederickson, Rethinking mutualism stability: cheaters and the evolution of sanctions. *The Q. Rev. Biol.* **88**, 269–295 (2013).
51. DE Reid, BJ Ferguson, S Hayashi, YH Lin, PM Gresshoff, Molecular mechanisms controlling legume autoregulation of nodulation. *Annals Bot.* **108**, 789–795 (2011).
52. JL Sachs, KW Quides, CE Wendlandt, Legumes versus rhizobia: a model for ongoing conflict in symbiosis. *New Phytol.* **219**, 1199–1206 (2018).
53. 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).
54. EG Weyl, ME Frederickson, WY Douglas, NE Pierce, Economic contract theory tests models of mutualism. *Proc. Natl. Acad. Sci.* **107**, 15712–15716 (2010).
55. DL Hartl, AG Clark, AG Clark, *Principles of population genetics.* (Sinauer associates Sunderland, MA) Vol. 116, (1997).
56. F Guillaume, MC Whitlock, Effects of migration on the genetic covariance matrix. *Evolution* **61**, 2398–2409 (2007).
57. JB Saltz, FC Hessel, MW Kelly, Trait correlations in the genomics era. *Trends Ecol. & Evol.* **32**, 279–290 (2017).
58. KD Heath, JR Stinchcombe, Explaining mutualism variation: a new evolutionary paradox? *Evolution* **68**, 309–317 (2014).
59. 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).
60. JA Kimbrel, et al., Mutualistic co-evolution of typeIII effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLoS Pathog.* **9**, e1003204 (2013).
61. 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).
62. JB Yoder, Understanding the coevolutionary dynamics of mutualism with population genomics. *Am. J. Bot.* **103**, 1742–1752 (2016).
63. 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).
64. MA Parker, Legumes select symbiosis island sequence variants in *Bradyrhizobium*. *Mol. Ecol.* **21**, 1769–1778 (2012).
65. JK Conner, DL Hartl, , et al., *A Primer of Ecological Genetics.* (Sinauer Associates Incorporated), (2004).
66. M Lynch, B Walsh, *Genetics and Analysis of Quantitative Traits.* (Sinauer, Sunderland, MA) Vol. 1, (1998).
67. B Sinervo, E Svensson, Correlational selection and the evolution of genomic architecture. *Heredity* **89**, 329–338 (2002).
68. 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).
69. JA Lau, et al., Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. *Acta Oecologica* **39**, 80–86 (2012).
70. 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).
71. ME Afkhami, ML Friesen, JR Stinchcombe, Multiple mutualism effects generate synergistic selection and strengthen fitness alignment in the interaction between legumes, rhizobia and

DRAFT

mycorrhizal fungi. *Ecol. Lett.* **24** (2021).
72. J Sachs, et al., Host control over infection and proliferation of a cheater symbiont. *J. Evol. Biol.* **23**, 1919–1927 (2010).
73. D Dowling, W Broughton, Competition for nodulation of legumes. *Annu. Rev. Microbiol.* **40**, 131–157 (1986).
74. Q Wang, J Liu, H Zhu, Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. *Front. Plant Sci.* **9**, 313 (2018).
75. BS Younginger, ML Friesen, Connecting signals and benefits through partner choice in plant–microbe interactions. *FEMS Microbiol. Lett.* **366**, fnz217 (2019).
76. 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).
77. 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).
78. 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).
79. 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).
80. 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).
81. F Galibert, et al., The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* **293**, 668–672 (2001).
82. 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).
83. RC Team, R foundation for statistical computing. *R: a language environment for statistical computing.* (2016).
84. X Zhou, M Stephens, Efficient multivariate linear mixed model algorithms for genome-wide association studies. *Nat. Methods* **11**, 407–409 (2014).
85. 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).
86. F Tajima, Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics* **123**, 585–595 (1989).
87. YX Fu, WH Li, Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709 (1993).

DRAFT