2022

Combining GWAS and Population Genomic Analyses to Characterize Coevolution in a Legume-rhizobia Symbiosis

Brendan Epstein
University of Minnesota, St Paul

Liana T. Burghardt
University of Pennsylvania

Katy D. Heath
University of Illinois at Urbana-Champaign

Michael A. Grillo
Loyola University Chicago, mgrillo1@luc.edu

Adam Kostanecki
University of Minnesota, St Paul

Received: 4 March 2022 | Revised: 3 June 2022 | Accepted: 4 July 2022

DOI: 10.1111/mec.16602

FROM THE COVER

MOLECULAR ECOLOGY WILEY

Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis

Brendan Epstein[1] | Liana T. Burghardt[2] | Katy D. Heath[3,4] | Michael A. Grillo[5] | Adam Kostanecki[1] | Tuomas Hämälä[1] | Nevin D. Young[1,6] | Peter Tiffin[1]

[1]Department of Plant and Microbial Biology, University of Minnesota, St. Paul, Minnesota, USA

[2]Department of Plant Sciences, The University of Pennsylvania, University Park, Pennsylvania, USA

[3]Department of Plant Biology, University of Illinois, Urbana, Illinois, USA

[4]Carl R. Woese Institute for Genomic Biology, University of Illinois, Urbana, Illinois, USA

[5]Department of Biology, Loyola University Chicago, Chicago, Illinois, USA

[6]Department of Plant Pathology, University of Minnesota, St. Paul, Minnesota, USA

Correspondence
Peter Tiffin, Department of Plant and Microbial Biology, University of Minnesota. St. Paul, MN 55108, USA.
Email: ptiffin@umn.edu

Present address
Tuomas Hämälä, School of Life Sciences, University of Nottingham, Nottingham, UK

Funding information
National Science Foundation, Grant/Award Number: IOS-1856744; U.S. Department of Energy, Grant/Award Number: CSP-503446

Handling Editor: Jeremy B Yoder

Abstract

The mutualism between legumes and rhizobia is clearly the product of past coevolution. However, the nature of ongoing evolution between these partners is less clear. To characterize the nature of recent coevolution between legumes and rhizobia, we used population genomic analysis to characterize selection on functionally annotated symbiosis genes as well as on symbiosis gene candidates identified through a two-species association analysis. For the association analysis, we inoculated each of 202 accessions of the legume host *Medicago truncatula* with a community of 88 *Sinorhizobia (Ensifer) meliloti* strains. Multistrain inoculation, which better reflects the ecological reality of rhizobial selection in nature than single-strain inoculation, allows strains to compete for nodulation opportunities and host resources and for hosts to preferentially form nodules and provide resources to some strains. We found extensive host by symbiont, that is, genotype-by-genotype, effects on rhizobial fitness and some annotated rhizobial genes bear signatures of recent positive selection. However, neither genes responsible for this variation nor annotated host symbiosis genes are enriched for signatures of either positive or balancing selection. This result suggests that stabilizing selection dominates selection acting on symbiotic traits and that variation in these traits is under mutation-selection balance. Consistent with the lack of positive selection acting on host genes, we found that among-host variation in growth was similar whether plants were grown with rhizobia or N-fertilizer, suggesting that the symbiosis may not be a major driver of variation in plant growth in multistrain contexts.

KEYWORDS
bacteria, Medicago, microbes, mutualism, population genomics

1 | INTRODUCTION

Rhizobial bacteria and legume hosts form a symbiosis that is clearly the result of selection imposed by symbiotic partners, that is, coevolution (Masson-Boivin & Sachs, 2018). Less clear is the nature of ongoing coevolution between these symbiotic partners. One possibility is that fitness conflicts between hosts and rhizobia drive selection and adaptation (Hoeksema, 2010; Queller &

Strassmann, 2018; Sachs et al., 2018). Under this scenario, symbiosis genes would be enriched for signatures of positive or balancing selection (Ebert & Fields, 2020). Alternatively, the fitness of hosts and mutualists may be largely aligned (Frederickson, 2013, 2017; Friesen, 2012) resulting in symbiotic partners imposing stabilizing selection on one another. Under this scenario, variation in the genes underlying symbiotic traits may reflect mutation-selection balance (Gano-Cohen et al., 2020; Heath & Stinchcombe, 2014; Kiester et al., 1984).

The symbiosis between rhizobial bacteria and legume hosts originated ~60 million years ago (Ma), and both legumes and rhizobia have adapted to each other (Doyle, 1998; Sprent, 2008). The symbiosis is formed when a rhizobium invades a plant root hair, and then the plant forms a symbiotic organ called a nodule. Within nodules, hosts and rhizobia exchange resources. The rhizobia convert atmospheric N_2 to a plant usable form, thereby enabling plant growth and reproduction in environments where N is limiting. In turn, rhizobia inside the nodule obtain both sugars and an environment that enables high rates of bacterial reproduction (Denison & Kiers, 2011)—a single rhizobium that forms a nodule can leave hundreds of thousands of offspring at the end of a plant's growing season (Oono et al., 2011; Ratcliff et al., 2011). Nevertheless, neither legumes nor rhizobia are fully dependent on their symbiotic partner. Legumes do not need rhizobia when soil N is available, and rhizobia, which are horizontally transmitted, cycle between living as mutualists inside of plants and as saprotrophs in the soil. In experimental settings, each plant host can form tens to hundreds of nodules, with each nodule potentially founded by a different rhizobium strain.

Genotype-by-genotype interactions (G×G) for partner fitness, the basis of reciprocal selection between hosts and symbionts (Wade, 2007), has been found in both host-associated rhizobial fitness (Burghardt et al., 2018; Heath & Tiffin, 2009; Mendoza-Suárez et al., 2020; Porter & Simms, 2014; Simonsen & Stinchcombe, 2014) and the benefits hosts obtain from specific rhizobial strains, at least in single-strain inoculation experiments (e.g., Bamba et al., 2020; Gano-Cohen et al., 2020; Heath, 2010; Porter & Simms, 2014; Regus et al., 2017; Sugawara et al., 2013; Torkamaneh et al., 2020). Plant genotypes also vary in symbiotic traits, such as the number and size of the nodules they produce, two proxies of the plant-associated reproductive success of rhizobia (Friesen, 2012; Gano-Cohen et al., 2020; Heath, 2010; Ossler & Heath, 2018). The G×G variation that these experiments reveal can result from a combination of some strains being more successful at forming nodules with certain hosts and from variation in either the rewards the symbiotic partners receive from each other after nodules are formed (Bohlool & Schmidt, 1974; Wang et al., 2012; Young & Johnston, 1989) or in costs of symbiosis (Quides et al., 2021). Importantly, however, G×G variation has been found primarily in experiments in which each host is inoculated with only a single strain of bacteria. Whether such variation is found when hosts are able to preferentially form a symbiosis with some strains or reward more beneficial strains after the symbiosis is established is less clear. Among-host variation in symbiont composition is also found in other host-symbiont, host-microbiome, and host-pathogen systems (e.g., Bálint et al., 2015; Goodrich et al., 2014; Zhao et al., 2016).

Given that symbionts can strongly affect their partner's reproductive success, among-host and G×G variation in host-associated reproductive success of rhizobia could reflect ongoing reciprocal selection and adaptation (Heath & Nuismer, 2014; Heath & Stinchcombe, 2014; Sachs et al., 2011; Thrall et al., 2007; Yoder, 2016). Alternatively, selection on symbionts by hosts may be weak because processes happening outside the host are far more important in shaping rhizobial evolution. For instance, in field soils where plants form nodules with multiple strains (Bailly et al., 2006), the performance of an individual strain might have little effect on host fitness (Burghardt et al., 2019; Wendlandt et al., 2019).

Analyses of genomic data from the model legume *Medicago truncatula* have revealed evidence for selection acting on a few functionally annotated symbiosis genes, but most symbiosis genes appear to be under purifying selection (De Mita et al., 2006, 2007; Grillo et al., 2016; Paape et al., 2013; Yoder, 2016), suggesting pervasive stabilizing selection and that rhizobia are not driving adaptation in the *Medicago* host. In rhizobia, there is some evidence of balancing selection on introgressed symbiosis genes, possibly due to frequency dependent selection (Ostrowski et al., 2015), and some genes affecting host-associated fitness in single-strain inoculations bear signatures of balancing selection (Batstone et al., 2021). Nevertheless, the among-species divergence of annotated symbiosis genes, primarily nodulation and N-fixation related genes, suggests a history of mostly purifying selection, with only a small number of genes showing signatures of positive selection (Epstein & Tiffin, 2021). Notably, these previous studies have focused on a limited set of functionally annotated symbiosis genes or relied on single-strain inoculation experiments to identify candidate genes contributing to variation.

The primary goal of this study was to investigate whether G×G variation in the legume-rhizobia symbiosis reflects ongoing reciprocal selection and adaptation. To address this goal, we first conducted a select and resequence (S&R) experiment in which each of 202 accessions of *M. truncatula* was inoculated with a synthetic community of 88 strains of *Sinorhizobium meliloti*. We used whole-genome sequence data to estimate the representation of each of the 88 strains on each host, enabling us to conduct association analyses to identify both plant and rhizobial genes that contribute to among-host variation in the rhizobial strain composition of the nodule community. We then used population genomic analyses to search for evidence that these genes, as well as genes that functional analyses have identified as being involved in symbiosis, bear signatures of selective sweeps or balancing selection. A secondary goal of this study was to evaluate the extent to which variation in host growth is driven by variation in the benefits hosts derive from symbiosis. If rhizobia-imposed selection on hosts is strong, then we expect among-host variation in growth will be much larger when plants are dependent on rhizobial symbionts rather than fertilizer for nitrogen. To address this second goal we compared the extent of among-host variation in plant size between plants growing with rhizobia versus N-fertilizer.

2 | MATERIALS AND METHODS

We grew eight replicates of each of 202 accessions of *Medicago truncatula*, which is the majority of the Medicago HapMap collection (Stanton-Geddes et al., 2013, Dataset S1). Five of these replicates were inoculated with a synthetic community of 88 strains of *Sinorhizobia meliloti*. To compare among-accession plant growth with rhizobia versus inorganic N we also grew three replicates of each accession with N-fertilizer and no rhizobia. Prior to planting, seeds were scarified by gently abrading with sandpaper, surface sterilized by soaking for 90 s in 10% bleach, washed in dH_20, then stratified on wet filter paper, in the dark, for 3 days at 4°C. After stratification, seeds were planted into 650 ml containers (Stuewe & Sons) filled with a steam sterilized mixture of five parts Sunshine Mix LP5 to one part Turface. Each replicate was distributed into one of eight randomized blocks across five greenhouse benches. After germination, we removed seedlings, leaving three plants per pot (each pot was a replicate–we left three plants to increase the number of nodules to pool for each genotypic replicate). Two days after planting, five replicates of each accession were inoculated with a mixture of 88 strains derived from a collection of 165 *Ensifer* (*Sinorhizobium*) *meliloti* strains (Dataset S2). These strains were collected by Riley et al. (2022) from 21 locations in the native range of *M. truncatula* in Spain and France. The 88 strains were chosen such that no pair of strains differed by <1000 of the ~160,000 variable sites that were genotyped in all strains. Details on the strain sequencing and variant calling are provided in Appendix S1: Methods.

Following Burghardt et al. (2018), we grew each of the 88 strains in TY media on a shaker set to 120 rpm for 72 h at 30°C, and then mixed 3 ml of each single-strain culture to form the multistrain inoculum with approximately equal representation of each strain (median: 1.1%, min: 0.7%, max: 2.1%). The inoculum was diluted 100-fold with 0.85% NaCl and 15 ml was added to each pot in five of the blocks (approximately 5×10^7 cells per pot, 80,000 cells/ml soil). Plants were fertilized once or twice weekly with N-free Fahraeus medium (0.5 mM $MgSO_4$, 0.7 mM KH_2PO_4, 50 uM Fe-EDTA, 0.8 mM Na_2HPO_4, 1.0 mM $CaCl_2$, and 1 mg/L of $MnCl_2$, $CuSO_4$, $ZnSO_4$, H_3BO_3, and Na_2MoO_4).

Nine weeks after planting, on a single day, we measured the length of the longest branch of each plant. At this time we also started harvesting plants. At the time of harvest, we separated above-ground growth from roots; removed, cleaned, and photographed nodules; and extracted rhizobia following the protocol in Burghardt et al. (2018) with minor modifications to improve yield. In brief, we placed nodules and 3 ml 0.85% NaCl in a 15 ml tube, homogenized for 90 s using an Omni TH tissue homogenizer, centrifuged the homogenized slurry at 400 *g* for 8 min at 10°C, then centrifuged the supernatant at 10,000 *g* for 5 min. The pellet from this spin, which is heavily enriched for undifferentiated rhizobia, was frozen until DNA was extracted using QIAgen DNEasy plant kits. Data on above ground biomass was collected from all five blocks, but due to time constraints (harvesting coincided with the onset of the COVID pandemic) we harvested nodules from only four of the five inoculated blocks. Biomass was measured after drying the above-ground growth at 60°C for a minimum of 48 h.

We estimated the number and mean area of nodules from each plant using ImageJ. To enhance contrast, we photographed nodules on a green background. Images were colour corrected in "Photos for Mac" and uploaded into FIJI. We batch processed images by splitting out the green channel and thresholding to a black and white image "SetThreshold(0,105)". All thresholded images were checked visually and rerun or manually cleaned if necessary. We selected the area of the image with nodules using the polygon tool and further cleaned images manually. Finally, we used Particle Analyser to batch assess nodule number, size, shape, and variance. We set the scale (393.5 pixels/cm), made the image binary, removed outliers <5 pixels, and thresholded particle size minimum (0.002 cm 2.4 cm^2 to remove media particles). Nodule number was manually counted on ~5% of images to check reliability of the pipeline.

Three replicates (blocks) of each plant accession, on a greenhouse bench adjacent to the five replicates of inoculated plants, were not inoculated with rhizobia. These uninoculated plants were grown using the same protocols as the inoculated plants but were given 25 ml of N-fertilizer once or twice each week (5 mM KNO_3 the week seeds were planted, 10 mM KNO_3 the second week, 20 mM KNO_3 twice during the third week, 80 mM NH_4NO_3 once a week for the next 4 weeks, and 80 mM NH_4NO_3 twice a week for the rest of the experiment). The root systems from approximately half of these plants were examined for nodules; none were found.

2.1 | Estimates of strain frequencies

We estimated the frequencies of rhizobial strains in the nodule community of each plant using the approach used in Burghardt et al. (2018). In brief, bacterial DNA extracted from the nodules of each plant was sequenced to a mean depth of 120 (range = 73–344) using paired end 150 bp reads on an Illumina Novaseq (library construction and sequencing was done by the University of Minnesota Genomics Center). We used TrimGalore! (version 0.5.0) to trim adapters and low quality bases, with the minimum adapter length set to three (--stringency), error rate set to 0.1 (-e), quality threshold set to 30 (--quality), and minimum read length set to 99. Trimmed reads were aligned to the *S. meliloti* USDA1106 genome (GCA_002197065.1) using bwa mem (version 0.7.17) with default settings. We then used HARP (Kessner et al., 2013) to estimate strain frequencies from these alignments, using 162,255 SNPs that were genotyped in each of the 88 strains and had variant quality scores >20 (501,228 invariant sites were also included in the analysis). Most of the subsequent analyses were performed on median relative fitness (Burghardt et al., 2018). We calculated relative fitness by dividing strain frequencies for each replicate pot by the mean frequency in the initial community, log_2-transforming those values, and then taking the median across replicates for each host-strain combination. Our analyses rely on relative fitness because HARP provides estimates of the relative frequencies of each strain in each sample

but does not provide estimates of the absolute reproductive success of strains on different hosts.

2.2 | Analyses of phenotypes, strain composition of nodule communities, and rhizobial fitness

We used the glm function in R to test for genetic variation in each of five symbiotic traits: length of longest branch, above-ground dry mass, nodule number, mean nodule area, and Shannon-Wiener's diversity (hereafter Shannon's) of the strains found in nodules. In these analyses we included block as a cofactor. For measures of above-ground plant growth we also included whether plants were on the edge of a bench (these plants received more sun). Plant location explained a very small portion of the variance in below-ground traits (all $p > .5$) and thus was not included as a cofactor in those analyses.

Host-specific strain composition of the nodule communities – that is, the relative frequencies of each of the 88 rhizobia strains in the nodules on each of 202 *M. truncatula* accessions – is a highly multidimensional phenotype. We used two principal component analyses (PCA), both based on a matrix of median relative fitness of each strain on each host, to identify different aspects of this complex phenotype. Both PCAs were implemented using the R function prcomp using data that were centred but not scaled. We did not scale the data because all values were measured in the same units and on the same scale. The first PCA differentiated among-host variation in the composition of rhizobial strains found in nodules (host accessions were rows, rhizobial strain identities were columns). We used each of the first two PCs from this analysis as a response variable in an association analysis to identify plant genes that potentially contribute to among-host variation in the strain composition of nodule communities. The second PCA focused on variation in the frequency of rhizobial strains among hosts (rhizobial strain identities were rows, host accessions were columns). We used the two major dimensions from this PCA as response variables in association analysis to identify rhizobial genes that potentially contribute to among-strain variation in host-associated rhizobium reproductive success. Lastly, we characterized the nodule community by calculating the Shannon's index on the median frequency of strains in each host genotype. The PCs, unlike Shannon's index, incorporate information on strain identities, for example, two host accessions, each of which form nodules with 44 strains but with no overlap in strain identities, would have equal Shannon values but would have very different PC values.

2.3 | Association analyses—*Medicago*

We conducted association analyses to identify *Medicago* variants associated with variation in nodule phenotype (nodule number, nodule area) and strain composition of the nodule community (Shannon's index and PC1 & PC2 differentiating variation among hosts). These analyses were conducted using 10,774,972 *M. truncatula* variants with MAF >5% (9.12 M SNPs, 1.66 M indels), identified by mapping previously collected sequence reads (Stanton-Geddes et al., 2013; accession numbers in Dataset S1) to the Mt version 5.0 reference genome (Pecrix et al., 2018) (details on read alignment and variant calling are in the Appendix S1: Methods). Gene-based analyses used the version 1.8 annotation available from INRA (accessed 26 August 2021) combined with the assembly from NCBI (GCA_003473485.1).

We conducted association analyses, both with and without a K-matrix of among-sample relatedness included as a covariate, using GEMMA (version 0.98.1, Zhou & Stephens, 2012). It is standard practice to include population structure covariates in association analysis to reduce potential bias arising from covariance between phenotypes and populations. However, including such covariates can prevent the identification of phenotypically important genes if causal variants are themselves geographically structured (Atwell et al., 2010). We calculated the K-matrix in GEMMA with the standardized K matrix option. Results from analyses with and without a K-matrix (linear-mixed model versus linear model) were quite similar (Figure S1), indicating that unequal relatedness among plant accessions or rhizobium strains did not greatly affect results. Thus, we focus our presentation on linear mixed model results. We considered the 1% of genomic windows harbouring the most strongly associated genomic variants, based on effect size, as candidate windows, and the genes in those windows were considered candidate genes. We also generated null expectations for association analyses by running 100 analyses, for each trait, on phenotypic data that had been permuted using the R package MVNPermute (Abney, 2015) following the method in Voichek and Weigel (2020), which maintains the mean and genetic covariance structure of the original data. In the main text we focus on the 1% of windows with strongest associations as candidates, but results are qualitatively similar when the 50 or 100 windows of strongest association are considered as candidates (Table S1), suggesting our results are not strongly dependent on the stringency applied for identifying candidate genes.

2.4 | Association analyses—*Sinorhizobia*

Sinorhizobia association analyses were conducted using GEMMA (version 0.98.1) on phenotype values (PC1 and PC2) transformed to a normal distribution using an inverse rank-based normal transformation. As with the *Medicago* analyses, we used MVNPermute to permute the GWA traits. Because of the nature of recombination in bacteria, linkage disequilibrium (LD) can be extensive, including among variants in discontiguous regions of the genome. Moreover, the extent of LD differs among the three *Sinorhizobia* replicons. We used two approaches to reduce the potential for variants in high LD with many other variants to bias association results by capturing not only their own effects but also the effects of linked variants (e.g., Bulik-Sullivan et al., 2015). First, following the approach used in Epstein et al. (2018), we included only one variant from each group

of variants in strong LD ($R^2 \geq 0.95$, MAF >0.05) with one another. Due to differences in allele frequencies, SNPs in the same coding region can be members of different LD groups. The full set of 17,626 LD groups contained 81,920 variants (MAF >0.05) distributed across 8780 genes. To limit the difficulties of interpreting results due to extensive LD, we considered only variants in the 16,324 LD groups that harboured variants in three or fewer genes ("small LD groups"). These groups contained 33,052 variants and tagged 4039 genes. Second, because LD differs among replicons, we conducted association analyses on each of the replicons separately.

We considered the 20 genes with the largest effect size on each Ensifer replicon as candidate genes. Twenty represents approximately 2% of chromosomal and 1.3% of both pSymA and pSymB genes included in the analyses. Although this cutoff is slightly more lenient than that used for the host, we selected it to provide some power to evaluate whether the empirical results deviated from random expectations. To calculate an FDR for the *Sinorhizo* association analyses, we analysed 1000 sets of permuted data, and counted the number of genes that had an effect size equal to or greater than the smallest effect size among the candidate genes (i.e., for the top 20, this would be the effect size of the 20th gene). As with *Medicago*, the set of candidates identified with and without a standardized K-matrix included as a covariate were similar (Figure S1), indicating that unequal relatedness among rhizobial strains did not greatly affect results.

2.5 | Population genetic analyses

We focused on two statistics, β and RAiSD μ to compare the selective history of candidate genes (identified through association analyses) and annotated symbiosis genes to genome-wide expectations. The β statistic is designed to detect evidence of balancing selection (Siewert & Voight, 2020). The μ statistic of RAiSD (Alachiotis & Pavlidis, 2018) is a composite measure that is based on the expected effects of selective sweeps on nucleotide diversity, LD, and the site frequency spectrum. These two statistics were chosen because of their reported power to detect evidence of selection (we also report Tajima's D in Datasets S3 and S4). We examined both candidate and annotated symbiosis genes because focusing only on candidate genes identified through genome-wide association studies (GWAS) may miss genes that have recently been driven to fixation through coevolution (MacPherson et al., 2018). We considered the *Medicago* genes that Roy et al. (2020) identified as playing important roles in the symbiosis to be annotated symbiosis genes.

Roy et al. (2020) identified 224 *Medicago* genes, which we mapped to the Mt5.0 annotation using a liftover file provided by INRA (https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/downloads/1.8/MtrunA17r5.0-ANR-EGN-r1.8.vs.JCVI-Mt4.0v2-gene.zip). Twelve of the 224 genes were not included in the Mt4-to-Mt5.0 annotation liftover; we identified nine of these genes (Medtr5g094210, Medtr7g063220, Medtr7g078700, Medtr8g067470, Medtr8g465280, MT35v5_contig_51603_1, MT35v5_contig_52215_1, MT35v5_contig_55897_1, MT4Noble_057132) in Mt5.0 using BLAST, but were unable to confidently locate three microRNAs and do not include those in our analyses (mtr-miR160a (miR160), mtr-miR167a (miR167), and mtr-miR172c (miR172c)). Four genes in Roy et al. (2020), (Medtr1g027020, Medtr4g073400, Medtr5g007630, Medtr5g026460) matched to two Mt5 genes, but in all cases those genes were adjacent to each other and because we conducted our analyses on 10 kb genomic windows, these have little effect on our results.

For *M. truncatula*, we calculated β and μ using SNP data from 266 accessions aligned to the Mtv5.0 reference genome; only non-imputed biallelic variants with <20% missing data were included. These 266 accessions include the 202 accessions of the GWA panel and 64 sequenced *M. truncatula* accessions that are part of the *Medicago* HapMap collection Stanton-Geddes et al., 2013. β was estimated for each variant with MAF >0.05, using the default window size of 1 Kb and the folded version of the statistic. RAiSD was run using the default window size of 50 SNPs. We then combined β and μ estimates within 10 Kb regions into nonoverlapping windows and characterized selective signals at each window using the maximum β and μ estimates. We compared the values for each statistic from candidate windows (the 1% of windows harbouring variants of largest effect sizes from the association analyses) to the genome-wide values of each statistic.

For each *Sinorhizobia* gene included in the association analysis, we calculated β and μ using genome sequence data from the 165 strains from which the 88 strains used for association analyses were chosen. We used the 165 strains because the 88 strains were chosen to increase diversity and thus may represent a biased sample of diversity in the population from which they were sampled. To avoid biases that might arise from few polymorphisms, β and μ were calculated only on biallelic SNPs genotyped in at least 80% of the strains, and we included only genes that were segregating ≥10 SNPs. Because of the potential for extensive genomic rearrangements as well as PAVs segregating within *Sinorhizobia* (Nelson et al., 2018) we calculated statistics for each coding region, rather than using the windows-based approach that we used for the *M. truncatula* data. We compared the mean and median value of each statistic calculated for candidate genes to the values on all other genes on each of the three replicons.

For each replicon, the 20 genes with the greatest effect size were considered candidate genes and the nodulation and N-fixation genes compiled by Epstein and Tiffin (2021) were considered annotated symbiosis genes. We used *t*-tests and Wilcoxon tests to estimate the probability of observing mean and median values obtained for candidates and annotated symbiosis genes different from the genome-wide values by chance. To account for potential bias in the identification of candidate genes, we also compared the means and medians of permutation candidates, which we identified through association analysis of 1000 data sets generated by MVNPermute, as described above. We compared the minor allele frequencies (MAF)

of the variant with the largest effect size in each candidate gene to the genome-wide MAF of nonsynonymous and presence-absence variants, which we chose so that candidate alleles were compared to other putatively functional variants.

3 | RESULTS

3.1 | Among-host variation in symbiosis traits and strain fitness

We characterized symbiotic traits of each of 202 accessions of *M. truncatula* hosts inoculated with a synthetic community of 88 rhizobial strains. This experiment revealed extensive among-accession variation in symbiotic traits and plant growth. Host genotype explained 42% of variation in nodule number, 67% of variation in average nodule area, 37% of variation in Shannon-Wiener's index of nodule community diversity, and approximately 50% of variation in plant biomass (all $p < .001$, Figure 1a, Table S2). Rhizobial fitness also varied among strains and host accessions (Figure 1b); strain identity explained 43% of variance in rhizobial fitness, whereas 49% of the variance was attributed to host × strain interactions (Table S2).

To test whether among-accession variation in biomass reflects variation in the net benefits of symbiosis, we compared the among-accession variation in biomass of plants in the rhizobia inoculation experiment to the variation in a parallel experiment in which plants were not inoculated, but were given N-fertilizer. This experiment was conducted at the same time and on adjacent benches in the same greenhouse room as the inoculation experiment. In the N-fertilizer conditions, host accession explained 50% of the total among-plant variance in plant size, approximately equal to the 49% of the total variance explained in the inoculated experiment. Because N-fertilized plants were grown on benches adjacent to the inoculated plants, we used a permutation analysis to evaluate the extent to which symbiosis-dependent plant growth was genotype-dependent. Specifically, we randomly assigned blocks of plants to either the inoculated or N-fertilizer treatment and evaluated the magnitude of the variation due to the accession X N-source interaction. Across the permutations, this interaction accounted for a mean of approximately 5% of the total variance – not greatly less than the 8% of variance in the empirical data (Figure S2). Taken together, these results suggest that the efficacy of symbiosis has a much smaller effect on host-genotype variation in plant growth than factors independent of N source.

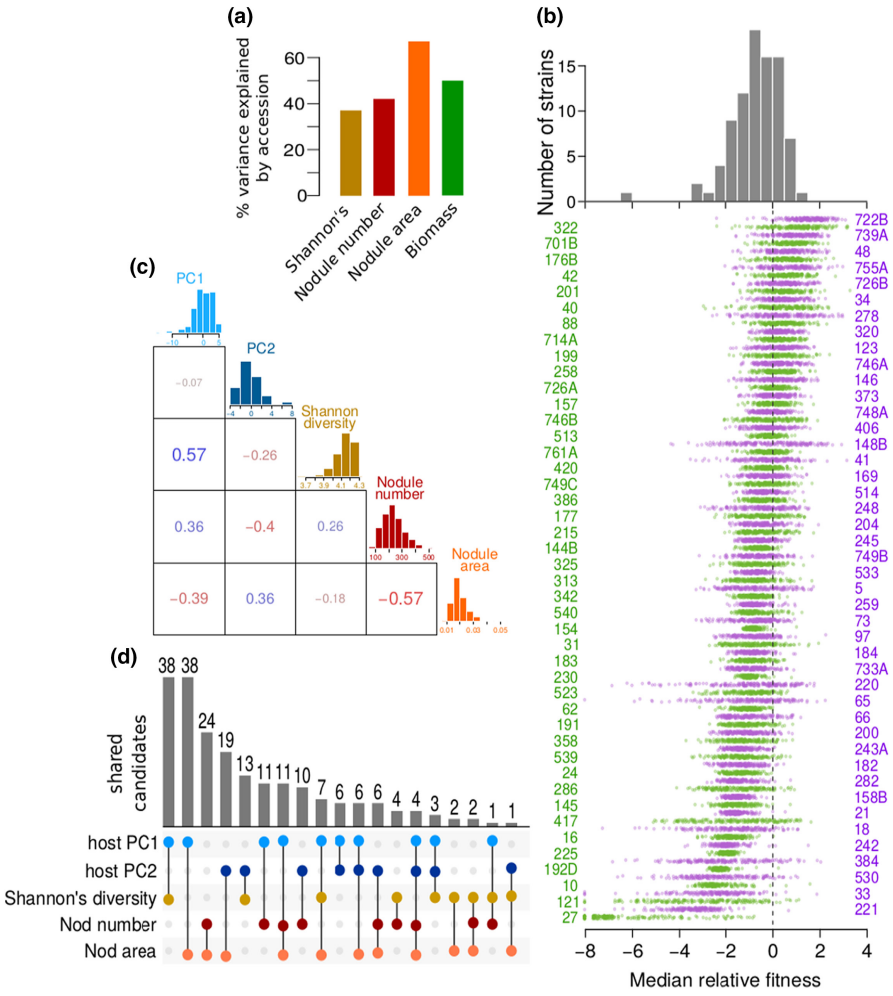

FIGURE 1 Both hosts and rhizobia exhibit genetic variation for symbiotic traits. (a) Percentage of host trait variance (PVE) explained by host genotype, (b) Distribution of relative fitness (mean across host genotypes, top) for each of the 88 rhizobial strains and strain relative fitness (rows) on each host accession (purple or green points), (c) distributions (diagonal) and genetic correlations (off diagonal) between host traits, and (d) number of association candidates shared among traits

3.2 | Association analysis - *Medicago*

To identify host genes that contribute to variation in rhizobial fitness, we conducted association analyses on nodule number, average nodule size, Shannon's diversity of the nodule community, and the first two PCs of a PCA that characterized the relative frequencies of the 88 rhizobial strains in the nodules of the 202 host accessions (Dataset S5). The first two PCs from this analysis explained 16% and 9% of the among-host variation in strain composition of the nodule community (Figure S3). Although both PC1 and PC2 are correlated with values of Shannon's diversity ($r = 0.57$ and -0.26 for PC1 and PC2, respectively, Figure 1c), the PCs incorporate information on the identity of each strain whereas Shannon's diversity does not (e.g., two hosts that each form nodules with a different subset of half of the strains would have equal values of Shannon's diversity but would have different PC values). Nodule community metrics (PC1, PC2, and Shannon's) also were correlated with the number of nodules plants produced (Figure 1c, all $p < .001$) and average nodule area ($p < .001$ except for Shannon's $p < .01$). While the 25% of variance (16% + 9%) is considerably more than expected if the variance was equally distributed among the 88 PC dimensions, we note that a considerable portion of the among-host variance is not captured by these two PCs.

We focused our interpretation of the *Medicago* association analysis on the 1% of the 42,867 genomic windows that harboured the variants of strongest effect size for each trait (Dataset S3, Figure S4). Two-hundred windows were identified as candidates for two or more traits, 37 were associated with three traits, and four with four traits (Figure 1d). Two functional annotations, both associated with metabolism of xenobiotic molecules, were strongly overrepresented among these candidates: 10 of 71 ABC-type xenobiotic transporters were associated with three traits and seven of 14 flavin-containing monooxygenases, which facilitate secretion of xeno-compounds through oxidation (Eswaramoorthy et al., 2006), were associated with two or more traits. The strongest association for both nodule number and nodule size was MtrunA17_Chr2g0300121, a putative NADH:ubiquinone reductase, that might be involved with energy provisioning (Lindström & Mousavi, 2020). We did not find evidence that the *Medicago* annotated symbiosis genes are major contributors to variation in symbiosis phenotypes. No more than five annotated symbiosis genes were among the top 1% of most strongly associated genes for any of the phenotypes (all $p > .10$, Dataset S6).

3.3 | Association analysis—*Sinorhizobia*

To identify rhizobial genes that contribute to variation in rhizobial fitness, we first conducted a GWAS on each of the first two dimensions (PCs) of a PCA of among-strain variation in fitness (Dataset S7). The first PC (PC1) captured 68% of variation in rhizobial fitness. This PC is effectively a measure of strains' across-host median fitness (correlation between PC1 and median fitness across hosts, $r = -0.99$, Figure S2). PC2 accounted for 12.5% of the remaining variance. We focused our interpretation of the association analyses on the 20 genes on each replicon harbouring the strongest effect size variants (Dataset S4). The 20 genes represent approximately 2% of chromosomal and 1.3% of both pSymA and pSymB genes included in the analyses. Although permutation-based FDR estimates suggest these candidates may contain a considerable number of false positives (Table S3), we did find substantial variance in rhizobial fitness and the composition of the nodule community, evidence that these traits are under genetic control (Figure 1, Table S2). Twenty candidates also provide some statistical power to test whether *Sinorhizobia* candidate genes harbour signals of selection that are different from noncandidates (see below).

The PC1 candidates on the *Sinorhizobia* pSymA replicon include multiple genes previously implicated as important for symbiosis (Table 1, Dataset S8): two are annotated as symbiosis genes, one was among the 24 pSymA genes identified by experimental evolution as increasing host-associated strain fitness or rhizobial partner quality (Batstone et al., 2020), two (*nifK*, and *fixP2*) were among the 20 genes most strongly associated with fitness variation among *Ensifer* strains sampled from across Europe and North America Burghardt et al., 2018 – completely different strains than those used in the current study – and three (*nifK* [CDO30_RS19300], a LysR family transcriptional regulator [CDO30_RS19615], and a NAD-dependent succinate-semialdehyde dehydrogenase [CDO30_RS19385]) were among the top 30 candidates influencing host biomass in single strain inoculations using that same global community (Epstein et al., 2018). The overlap between association candidates and annotated symbiosis genes is not limited to the 20 most strongly associated candidates; three additional annotated symbiosis genes (*fixL*, *nolF*, *nodQ*,) are among the 50 most strongly associated pSymA PC1-candidates ($p < .1$). Moreover, the pSymA symbiosis gene *nodM* (CDO30_RS19120/WP_010967462.1), which may affect the strength of rhizobial signalling (Baev et al., 1992), had the fourth strongest effect on variation along PC2.

3.4 | Selective history of candidate genes underlying symbiosis traits

Our primary motivation for identifying genes associated with variation in symbiosis traits was to look for signatures of selection consistent with ongoing coevolution. We focused on two population genetic statistics to search for evidence of recent selection: β and μ which are designed to detect histories of balancing and positive selection, respectively (Alachiotis & Pavlidis, 2018; Siewert & Voight, 2020). For *Medicago*, we calculated β and μ for each variable site and then characterized each 10 kb genomic window by the maximum value of each statistic. For *Sinorhizobia*, because of high levels of genomic rearrangement among strains (Nelson et al., 2018), we calculated β and μ for each gene.

The *Medicago* candidate windows had mean and median values of β that were slightly higher than the genome-wide values, but the differences were relatively small (Figure 2, Table S1; candidate

TABLE 1 *Sinorhizobia* pSymA candidate genes (20 genes with strongest effect size) underlying variation in rhizobial fitness PC1 that have been previously implicated as playing a role in symbiosis. The genes are listed both by the NCBI gene locus tag ("CD…"), which uniquely tags all genes, as well as by the NCBI Refseq ID, for protein-coding genes, in parentheses

Annotated function	pSymA candidate	Other support for role in symbiosis
Type II toxin-antitoxin system HigB family toxin	CDO30_RS18795 WP_003526630.1	Among top 20 candidates for PC2
fixP2 (*ccoP*), cytochrome-c oxidase, cbb3-type subunit III	CDO30_RS19540 WP_014989946.1	Annotated symbiosis, involved in respiration in microaerobic conditions in symbiosis (Delgado et al., 1998)
nifK, nitrogenase molybdenum- iron protein subunit beta	CDO30_RS19300 WP_003532776.1	Annotated symbiosis (Ruvkun et al., 1982), among top 30 candidates affecting host biomass in Epstein et al. (2018)
LysR family transcriptional regulator	CDO30_RS19615 WP_010967389.1	Experimental evolution candidate for both rhizobia fitness and host benefit (Batstone et al., 2020), among top 20 candidates affecting host biomass in Epstein et al. (2018)
EAL domain-containing protein	CDO30_RS21505 WP_010967073.1	Among top 20 candidates underlying variation in host-associated fitness in Burghardt et al. (2018)
Adenylate/guanylate cyclase domain-containing protein	CDO30_RS22020 WP_027988995.1	Among top 20 candidates underlying variation in host-associated fitness in Burghardt et al. (2018)
NAD-dependent succinate- semialdehyde dehydrogenase	CDO30_RS19385 WP_010967426.1	Among top 30 candidates affecting host biomass in Epstein et al. (2018)
VirB4 family type IV secretion system protein	CDO30_RS25135 WP_014528660.1	VirB genes affect *E.meliloti* competitiveness for nodulation and have host-genotype specific effects on nodule number (Nelson et al., 2017). Candidate variants in these two genes are in strong LD ($r^2 > 0.95$) with one another
TrbC/VirB2 family protein	CDO30_RS25125 WP_010967695.1	

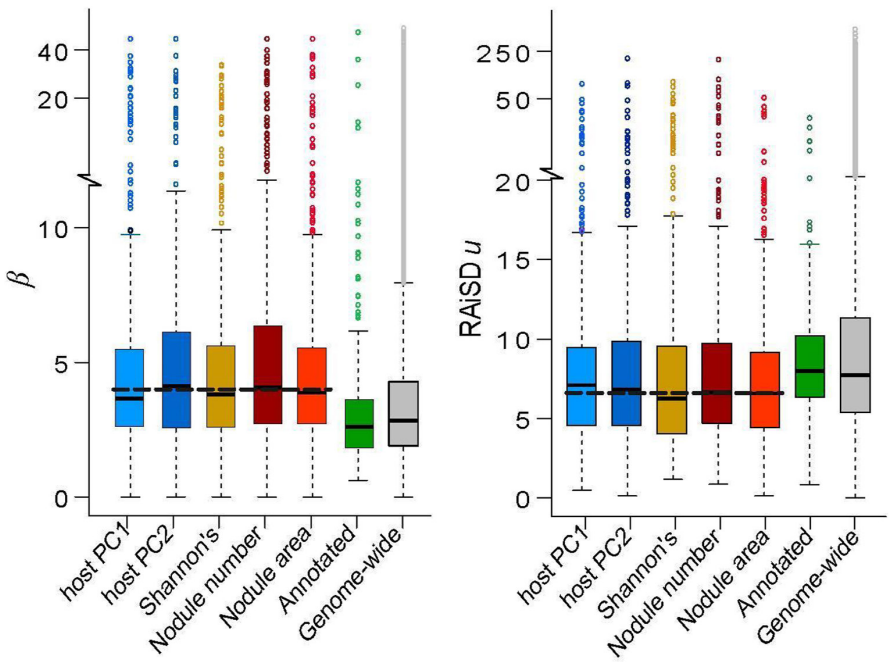

FIGURE 2 Distributions of β and RAiSD μ for the *Medicago* genomic windows harbouring association candidates and genome-wide values. The horizontal dashed line indicates the median values from the permutation analysis. Boxes indicate interquartile ranges with the horizontal solid line indicating the median. Whiskers indicate 1.5 times the interquartile range. The y-axis is on a linear scale below the break, a log2 scale above the break

averages = 5.09–5.78, medians = 3.66–4.12, genome wide average = 4.03, median 2.96). To test if these differences were due to association analyses being biased towards identifying alleles of intermediate frequency (Josephs et al., 2017; Myles et al., 2009), we repeated the association analyses on 100 permuted data sets, in which phenotypes were randomly assigned to accessions. The slight differences in β appears to be due to this bias; with the exception of the nodule number candidates, the mean and median β values of candidates were similar to or less than the averages from the permutation analyses (Figure 2, Table S1). Similarly, the mean and median values of RAiSD μ were slightly smaller than either genome-wide values or values from the permutation analysis (Figure 2, Table S1), and thus provide no evidence for the candidate genes having experienced recent selective sweeps. The *Medicago* annotated symbiosis

genes (Roy et al., 2020) had average values of β (3.52) and μ (2.62) that were slightly lower than genome-wide averages (Figure 2). The lack of evidence for either positive or balancing selection on *Medicago* candidates is not due to our decision to consider the 1% of genomic windows as candidates; results are similar if only the 50 or 100 windows harbouring the strongest associations are considered (Table S1).

Examining only the mean or median values of the selection statistics might fail to identify a subset of genes that are subject to strong selection. For this reason, we looked at the number of candidates with β or μ values in the upper 1% of genome-wide values. Candidates for nodule number, nodule area, and PC1 had more high values of β (highest 1%) than expected by chance (12, 9, and 9; $p < .001$, $p = .02$, and $p = .02$, respectively; Table S4). However, these probabilities fail to account for association analyses being biased towards identifying genes of intermediate allele frequencies. To account for this bias, we compared our empirical results to the β values of permutation candidates; the 1% of genes most strongly associated in a GWAS of phenotypes randomly assigned to genome sequences. Compared to the permuted data we found only a slight enrichment for genes with high β values; seven of 100 permuted data sets had ≥12 high β value windows for nodule number, nine had ≥9 for nodule area, and 51 had ≥9 for PC1 (Table S4). No more than two candidate windows were in the upper tail of μ values (all $p > 0.9$). Among the annotated symbiosis genes, only three were in windows with high values of β and none were in windows with high values of μ.

Although we do not find evidence for overall enrichment, the window with the fourth greatest β value contains NUP144, a nucleoporin (MtrunA17_Chr5g0447511) essential for Ca^{2+} signalling in nodule development (Kanamori et al., 2006). Moreover, three windows in the top 1% of β values were association candidates for three or more traits, including one containing two putative flavin-containing monooxygenases (MtrunA17_Chr2g0286671, MtrunA17_Chr2g0286681), and one containing a putative ABC-type xenobiotic transporter (MtrunA17_Chr8g035378), both functional annotations that were strongly enriched among *Medicago* association candidates. Variants in the chromosomal region harbouring the two flavin-containing monooxygenases were also strongly associated with symbiosis phenotypes in Stanton-Geddes et al. (2013).

We searched for evidence of selection on *Sinorhizobia* candidate and annotated symbiosis genes using the same approach we used for *Medicago* genes. On all three replicons, the candidate genes had mean and median values of both β and μ similar to those of the genomic background (Figure 3, Table S5, Dataset S4). The annotated symbiosis genes also had values of β and μ similar to the genomic background. Several of the genes with highest β or μ values on pSymA were candidate or annotated symbiosis genes (Figure 3). Among the 20 pSymA genes with highest values of mu, there were three annotated symbiosis genes: *fixK*, *fixN1*, and *fixN2*, which might be important in regulating N-fixation (Rutten et al., 2021); two PC2 candidates—one of which is annotated as being involved in transcriptional regulation, the other (*relE/parE*) involved in a toxin-antitoxin system (Fic, RelE/ParE) (Fiebig et al., 2010; Veyron et al., 2018)—and one PC1 candidate, *virB*, which encodes a type IV secretion system protein that is essential for infection of plants in at least some bacteria (e.g., Fullner et al., 1994). Although we did not include type IV secretion system proteins in our list of annotated symbiosis genes, these proteins may play important roles in host-specificity (Nelson et al., 2017). Of the four type IV secretion system proteins included in our population genomic analysis, three had a μ value among the top 10 pSymA μ values. These genes might be important in the adaptation of rhizobia to host environments; however, the probability of finding three annotated symbiosis genes among those with the twenty highest values of μ, by chance, is not improbable ($p = .12$).

4 | DISCUSSION

The symbiosis between legumes and rhizobia is clearly the result of coevolution. The effect these mutualists can have on one another's reproductive success, that hosts have evolved mechanisms to

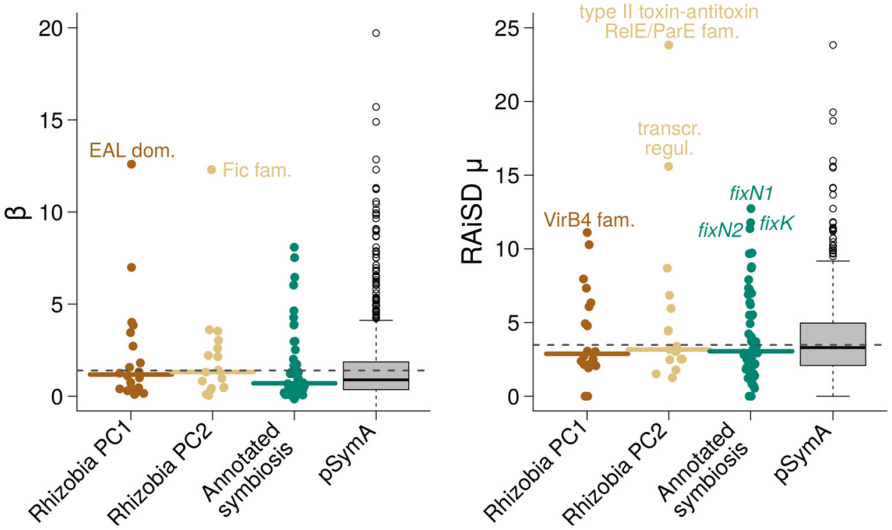

FIGURE 3 β and μ values for *Sinorhizobia* pSymA candidate genes (20 genes with strongest effect sizes) underlying variation in rhizobial fitness (PC1 and PC2) and annotated symbiosis genes. Candidate and symbiosis genes with values in the top 20 are labelled. Solid horizontal lines indicate the medians for each gene class, for the pSymA data the box indicates the interquartile range, the whiskers indicate 1.5 times the interquartile range. The horizontal dashed line indicates the mean of the median values from 1000 permutations. Similar figures for chromosome and pSymB genes are in Figure S5

control symbionts, and that rhizobia have evolved to evade these control mechanisms (Gano-Cohen et al., 2020; Sachs et al., 2018), make it likely that these symbionts continue to coevolve. Here, we used population genomic analyses to characterize the nature of ongoing coevolution between the legume *Medicago truncatula* and the rhizobia *Sinorhizobia meliloti*. The premise of this approach is that if hosts and symbionts are coevolving in a manner that is stable and strong enough to be driving adaptive change in either partner, then the genes that contribute to variation in the symbiosis should bear signatures of recent selection. We focus on two sets of genes in each species: (1) genes that association analyses we conducted identify as contributing to host or strain symbiotic performance in contemporary populations and (2) genes that functional analyses have previously identified as being involved in the formation or functioning of the symbiosis. The association candidates may include genes that functional analyses have not identified either because their effects are more subtle than might be identified through a genetic screen or their effects are only manifest in the multistrain inoculation we used in our experiments.

Our analysis revealed little evidence that either positive or balancing selection is a dominant force driving the evolution of symbiosis genes. The lack of evidence for strong positive or balancing selection on *Medicago* symbiosis genes is consistent with results from Yoder (2016), who searched for evidence of selection on a smaller set of *Medicago*, symbiosis genes as well as other genomic scans that have not revealed strong evidence for selection driving adaptation in *Medicago* symbiosis genes (Bonhomme et al., 2015; Branca et al., 2011; Grillo et al., 2016). Similarly, the lack of a strong signal of selection on *Ensifer* symbiosis-related genes is consistent with results from Epstein et al. (2012), who did not find obvious signals of adaptation among annotated symbiosis genes in a smaller, far more geographically diverse sample of *Sinorhizobia meliloti*. Similarly, Epstein and Tiffin (2021) found evidence for positive selection having driven among-rhizobia divergence of only four (*noeE*, *nolE*, a nitrogenase, and *fixK*) of 120 annotated symbiosis genes. The consistency of these results indicates that *Medicago* and *Ensifer* are not coevolving in a way that strongly drives divergence in the genes that underlie symbiotic traits. Note, however, that coevolution might be driving adaptation in some *Sinorhizobia* symbiosis genes – although not more than expected by chance, three annotated symbiosis genes have high values of μ, consistent with recent positive selection. These three genes include *fixK* which also was identified as a target of selection in a comparative genomic analysis (Epstein & Tiffin, 2021).

While our population genomic analysis reveals little evidence for ongoing coevolution driving the divergence of symbiosis genes, it is important to acknowledge the limits of gene-based population genetic statistics to search for evidence of selection. Although population genomic analyses can be powerful for identifying evidence of strong selection acting on genes that have a major phenotypic effect, their power to detect evidence of selection acting on phenotypes with a polygenic basis is limited (Chevin & Hospital, 2008; Pritchard et al., 2010). If adaptation in the *Medicago* and *Ensifer* symbiosis is primarily polygenic, then our gene-centric analysis may not have detected evidence for that adaptation, although the fact that the mean and median values of the symbiotic genes are largely consistent with random samples would argue against this.

The lack of evidence for rhizobia driving adaptation of host symbiosis genes may seem inconsistent with results from single-strain inoculation experiments that show that benefits legume hosts receive from symbiosis differ substantially, not only among rhizobial strains but also between specific combinations of hosts and rhizobia (Burdon et al., 1999; Burghardt et al., 2018; Heath, 2010; Parker, 1995). Our experiment differs from those experiments in that we inoculated plants with a multistrain community of rhizobia. A multistrain inoculum may more closely mimic what plants experience in nature, where strains compete for nodulation opportunities, plants are exposed to and form nodules with multiple strains, and plants can differentially reward strains after nodules are formed (Bailly et al., 2006; Denison & Kiers, 2011; Rangin et al., 2008; West et al., 2002). In fact, the majority of genetic variation segregating within the *Sinorhizobia* strains used here is found within discrete sampling locations (Riley et al., 2022). Plants that are grown in soil with multiple strains of rhizobia may preferentially associate with or reward strains that are more beneficial, and thereby may offset the negative effects of less beneficial strains that are observed when plants are inoculated with only one strain (Somasegaran & Bohlool, 1990; Wendlandt et al., 2019).

Although we do not find strong evidence for balancing or positive selection in either symbiotic partner, our results do not mean that symbiotic partners do not affect each other's fitness. Clearly, forming a symbiosis with rhizobia can greatly increase plant growth; and the reproductive success of rhizobia and some symbiotic traits, such as the number of nodules produced, can be environmentally dependent in at least some systems (e.g., Saturno et al., 2017). We also know that hosts can affect rhizobial strain fitness and drive adaptation in rhizobial populations (Batstone et al., 2020), at least when plants are fully dependent on rhizobia for nitrogen supply and the rhizobial communities are not complex. Rhizobial populations also can diverge as a result of soil N availability (Weese et al., 2015), and this divergence may be driven by positive selection, perhaps associated with the loss of a symbiotic gene (Klinger et al., 2016). These empirical results suggest that local conditions may shape selection acting on the legume-rhizobia symbiosis. In fact, the *Sinorhizobia* genes our selection analyses identify as having strongest signals of positive or balancing selection, in the geographically restricted sample we analysed here, showed little overlap with genes identified in a much smaller but geographically widespread sample (Epstein et al., 2012). Moreover, two studies (De Mita et al., 2007; Grillo et al., 2016) have found evidence for local adaptation at *DMI1*, an important symbiotic gene in *M. truncatula*. Further supporting a potential role in rhizobia having driven local adaptation of *DMI1*, Grillo et al. (2016) found that *DMI1* alleles were associated with preferential associations between hosts and two of three rhizobial strains. Here, however, in our range-wide sample of *Medicago*, we do not find evidence of strong selection on *DMI1*.

We also do not find that *DMI1* was an important contributor to the strain composition in the nodule community of hosts, perhaps due to the inoculum community we used, which was far more complex than that analysed by Grillo et al. (2016).

Mutualistic relationships, such as that between *Medicago* and *Sinorhizobia*, are important for the reproductive success of their symbiotic partners. The potential for symbionts to extract benefits without providing benefits to their hosts has raised the prospect that symbiotic relationships may coevolve in a manner similar to that in antagonistic symbiosis. This coevolution would be characterized by recurring bouts of adaptation associated with symbionts evolving to cheat their hosts and hosts evolving mechanisms to prevent such cheaters. Our analyses are not consistent with such evolution. Rather, our results are consistent with fitness alignment between *Medicago* hosts and *Ensifer* symbionts (Friesen, 2012). If this conclusion is correct, then the genetic variation found in the *Medicago-Ensifer* symbiosis might reflect mutation-selection balance (Heath & Stinchcombe, 2014; Kiester et al., 1984) or reflect complex frequency-dependent or environmentally-specific tradeoffs (Bever, 2002; Parker, 1999; Wendlandt et al., 2021) that are not reflected in the population genomic analyses conducted.

AUTHOR CONTRIBUTIONS

Brendan Epstein, Liana T. Burghardt, Nevin D. Young, and Peter Tiffin designed the research, Brendan Epstein, Liana T. Burghardt, Adam Kostanecki, Katy D. Heath, Nevin D. Young, and Peter Tiffin performed the research, Katy D. Heath and Michael A. Grillo provided strains, Brendan Epstein, Tuomas Hämälä and Peter Tiffin analysed data, Brendan Epstein and Peter Tiffin wrote the manuscript.

ACKNOWLEDGEMENTS

We thank Amanda Gorton, Molly Meath, Erica Houser, Reda Al Abou-Shanab, and Ivan Sosa Marquez for help with the tedious task of harvesting nodules. We also thank Jeremy Yoder and three anonymous reviewers for comments that improved the manuscript and helped us to clarify our thinking of coevolution. The Minnesota Supercomputing Institute (MSI) at the University of Minnesota provided resources that contributed to this research. This work was supported by the National Science Foundation under grant IOS-1856744 and Department of Energy JGI CSP-503446. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

CONFLICT OF INTEREST

The authors declare no competing interests.

OPEN RESEARCH BADGES

This article has earned an Open Data Badge for making publicly available the digitally-shareable data necessary to reproduce the reported results. The data is available at the datadryad.org/stash/share/XzkusnKvpTlCM4CpFK93OvwtGK7j75ZOqFfV8YwPr2M.

DATA AVAILABILITY STATEMENT

Sequence reads for *Medicago truncatula* (from Branca et al., 2011; Stanton-Geddes et al., 2013) have been made available under BioProject PRJNA256006 as listed in Dataset S1. Variants used for analyses are available at www.medicagohapmap2.org and from https://v1.legumefederation.org/data/v2/Medicago/truncatula/diversity/A17.gnm5.Epstein_Burghardt_2022/. *Sinorhizobia* reads from the select and resequence communities are available under BioProject PRJNA401437 and reads for the individual genome sequences are available from NCBI under the accession numbers listed in Dataset S2. Variant calls for *Ensifer* are available from Dryad. Phenotypic data are available in Dataset S5, and strain frequency estimates are available in Dataset S9. Dryad files are temporarily available from datadryad.org/stash/share/XzkusnKvpTlCM4CpFK93OvwtGK7j75ZOqFfV8YwPr2M for review.

ORCID

Brendan Epstein https://orcid.org/0000-0001-7083-1588
Liana T. Burghardt https://orcid.org/0000-0002-0239-1071
Katy D. Heath https://orcid.org/0000-0002-6368-744X
Nevin D. Young https://orcid.org/0000-0001-6463-4772
Peter Tiffin https://orcid.org/0000-0003-1975-610X

REFERENCES

Abney, M. (2015). Permutation testing in the presence of polygenic variation. *Genetic Epidemiology*, *39*(4), 249–258. https://doi.org/10.1002/gepi.21893

Alachiotis, N., & Pavlidis, P. (2018). RAiSD detects positive selection based on multiple signatures of a selective sweep and SNP vectors. *Communications Biology*, *1*(1), 79. https://doi.org/10.1038/s42003-018-0085-8

Atwell, S., Huang, Y. S., Vilhjálmsson, B. J., Willems, G., Horton, M., Li, Y., Meng, D., Platt, A., Tarone, A. M., Hu, T. T., Jiang, R., Muliyati, N. W., Zhang, X., Amer, M. A., Baxter, I., Brachi, B., Chory, J., Dean, C., Debieu, M., … Nordborg, M. (2010). Genome-wide association study of 107 phenotypes in Arabidopsis thaliana inbred lines. *Nature*, *465*, 627–631.

Baev, N., Schultze, M., Barlier, I., Ha, D. C., Virelizier, H., Kondorosi, E., & Kondorosi, A. (1992). Rhizobium *nodM* and *nodN* genes are common nod genes: *NodM* encodes functions for efficiency of nod signal production and bacteroid maturation. *Journal of Bacteriology*, *174*(23), 7555–7565. https://doi.org/10.1128/jb.174.23.7555-7565.1992

Bailly, X., Olivieri, I., De Mita, S., Cleyet-Marel, J.-C., & Béna, G. (2006). Recombination and selection shape the molecular diversity pattern of nitrogen-fixing *Sinorhizobium sp*. associated to *Medicago*. *Molecular Ecology*, *15*(10), 2719–2734. https://doi.org/10.1111/j.1365-294X.2006.02969.x

Bálint, M., Bartha, L., O'Hara, R. B., Olson, M. S., Otte, J., Pfenninger, M., Robertson, A. L., Tiffin, P., & Schmitt, I. (2015). Relocation, high-latitude warming and host genetic identity shape the foliar fungal microbiome of poplars. *Molecular Ecology*, *24*(1), 235–248. https://doi.org/10.1111/mec.13018

Bamba, M., Aoki, S., Kajita, T., Setoguchi, H., Watano, Y., Sato, S., & Tsuchimatsu, T. (2020). Massive rhizobial genomic variation associated with partner quality in *Lotus–Mesorhizobium* symbiosis. *FEMS*

Microbiology Ecology, *96*(12), fiaa202. https://doi.org/10.1093/femsec/fiaa202

Batstone, R. T., Burghardt, L. T., & Heath, K. D. (2021). The genetic basis of cooperation and conflict in natural populations of a model symbiont. *BioRxiv*. https://doi.org/10.1101/2021.07.19.452989

Batstone, R. T., O'Brien, A. M., Harrison, T. L., & Frederickson, M. E. (2020). Experimental evolution makes microbes more cooperative with their local host genotype. *Science*, *370*(6515), 476–478.

Bever, J. D. (2002). Dynamics within mutualism and the maintenance of diversity: Inference from a model of interguild frequency dependence. *Ecology Letters*, *2*(1), 52–61. https://doi.org/10.1046/j.1461-0248.1999.21050.x

Bohlool, B. B., & Schmidt, E. L. (1974). Lectins: A possible basis for specificity in the rhizobium—Legume root nodule symbiosis. *Science*, *185*(4147), 269–271. https://doi.org/10.1126/science.185.4147.269

Bonhomme, M., Boitard, S., San Clemente, H., Dumas, B., Young, N., & Jacquet, C. (2015). Genomic signature of selective sweeps illuminates adaptation of *Medicago truncatula* to root-associated microorganisms. *Molecular Biology and Evolution*, *32*(8), 2097–2110. https://doi.org/10.1093/molbev/msv092

Branca, A., Paape, T. D., Zhou, P., Briskine, R., Farmer, A. D., Mudge, J., Bharti, A. K., Woodward, J. E., May, G. D., Gentzbittel, L., Ben, C., Denny, R., Sadowsky, M. J., Ronfort, J., Bataillon, T., Young, N. D., & Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences*, *108*(42), E864–E870. https://doi.org/10.1073/pnas.1104032108

Bulik-Sullivan, B. K., Loh, P.-R., Finucane, H. K., Ripke, S., Yang, J., Schizophrenia Working Group of the Psychiatric Genomics Consortium, Patterson, N., Daly, M. J., Price, A. L., & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature Genetics*, *47*(3), 291–295.

Burdon, J. J., Gibson, A. H., Searle, S. D., Woods, M. J., & Brockwell, J. (1999). Variation in the effectiveness of symbiotic associations between native rhizobia and temperate Australian *Acacia*: Within-species interactions. *Journal of Applied Ecology*, *36*(3), 398–408. https://doi.org/10.1046/j.1365-2664.1999.00409.x

Burghardt, L. T., Epstein, B., Guhlin, J., Nelson, M. S., Taylor, M. R., Young, N. D., Sadowsky, M. J., & Tiffin, P. (2018). Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. *Proceedings of the National Academy of Sciences*, *115*(10), 2425–2430. https://doi.org/10.1073/pnas.1714246115

Burghardt, L. T., Epstein, B., & Tiffin, P. (2019). Legacy of prior host and soil selection on rhizobial fitness in planta. *Evolution*, *73*, 2013–2023.

Chevin, L.-M., & Hospital, F. (2008). Selective sweep at a quantitative trait locus in the presence of background genetic variation. *Genetics*, *180*(3), 1645–1660. https://doi.org/10.1534/genetics.108.093351

De Mita, S., Ronfort, J., McKhann, H. I., Poncet, C., El Malki, R., & Bataillon, T. (2007). Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in *Medicago truncatula*. *Genetics*, *177*(4), 2123–2133. https://doi.org/10.1534/genetics.107.076943

De Mita, S., Santoni, S., Hochu, I., Ronfort, J., & Bataillon, T. (2006). Molecular evolution and positive selection of the symbiotic gene NORK in *Medicago truncatula*. *Journal of Molecular Evolution*, *62*(2), 234–244. https://doi.org/10.1007/s00239-004-0367-2

Delgado, M. J., Bedmar, E. J., & Downie, J. A. (1998). Genes involved in the formation and assembly of rhizobial cytochromes and their role in symbiotic nitrogen fixation. In R. K. Poole (Ed.), *Advances in microbial physiology* (pp. 191–231). Academic Press. https://doi.org/10.1016/S0065-2911(08)60132-0

Denison, R. F., & Kiers, E. T. (2011). Life histories of symbiotic rhizobia and mycorrhizal fungi. *Current Biology*, *21*(18), R775–R785. https://doi.org/10.1016/j.cub.2011.06.018

Doyle, J. J. (1998). Phylogenetic perspectives on nodulation: Evolving views of plants and symbiotic bacteria. *Trends in Plant Science*, *3*(12), 473–478. https://doi.org/10.1016/S1360-1385(98)01340-5

Ebert, D., & Fields, P. D. (2020). Host-parasite co-evolution and its genomic signature. *Nature Reviews Genetics*, *21*, 754–768.

Epstein, B., Abou-Shanab, R. A. I., Shamseldin, A., Taylor, M. R., Guhlin, J., Burghardt, L. T., Nelson, M., Sadowsky, M. J., & Tiffin, P. (2018). Genome-wide association analyses in the model rhizobium *Ensifer meliloti*. *MSphere*, *3*(5), e00386–e00318. https://doi.org/10.1128/mSphere.00386-18

Epstein, B., Branca, A., Mudge, J., Bharti, A. K., Briskine, R., Farmer, A. D., Sugawara, M., Young, N. D., Sadowsky, M. J., & Tiffin, P. (2012). Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. *PLoS Genetics*, *8*(8), e1002868. https://doi.org/10.1371/journal.pgen.1002868

Epstein, B., & Tiffin, P. (2021). Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. *Proceedings of the Royal Society B*, *288*(1942), 20201804.

Eswaramoorthy, S., Bonanno, J. B., Burley, S. K., & Swaminathan, S. (2006). Mechanism of action of a flavin-containing monooxygenase. *Proceedings of the National Academy of Sciences*, *103*(26), 9832–9837. https://doi.org/10.1073/pnas.0602398103

Fiebig, A., Castro Rojas, C. M., Siegal-Gaskins, D., & Crosson, S. (2010). Interaction specificity, toxicity and regulation of a paralogous set of ParE/RelE-family toxin–antitoxin systems. *Molecular Microbiology*, *77*(1), 236–251. https://doi.org/10.1111/j.1365-2958.2010.07207.x

Frederickson, M. E. (2013). Rethinking mutualism stability: Cheaters and the evolution of sanctions. *The Quarterly Review of Biology*, *88*(4), 269–295. https://doi.org/10.1086/673757

Frederickson, M. E. (2017). Mutualisms are not on the verge of breakdown. *Trends in Ecology & Evolution*, *32*(10), 727–734. https://doi.org/10.1016/j.tree.2017.07.001

Friesen, M. L. (2012). Widespread fitness alignment in the legume–rhizobium symbiosis. *New Phytologist*, *194*(4), 1096–1111. https://doi.org/10.1111/j.1469-8137.2012.04099.x

Fullner, K. J., Stephens, K. M., & Nester, E. W. (1994). An essential virulence protein of Agrobacterium tumefaciens, VirB4, requires an intact mononucleotide binding domain to function in transfer of T-DNA. *Molecular & general genetics*, *245*, 704–715.

Gano-Cohen, K. A., Wendlandt, C. E., Al Moussawi, K., Stokes, P. J., Quides, K. W., Weisberg, A. J., Chang, J. H., & Sachs, J. L. (2020). Recurrent mutualism breakdown events in a legume rhizobia metapopulation. *Proceedings of the Royal Society B: Biological Sciences*, *287*(1919), 20192549. https://doi.org/10.1098/rspb.2019.2549

Goodrich, J. K., Waters, J. L., Poole, A. C., Sutter, J. L., Koren, O., Blekhman, R., … Ley, R. E. (2014). Human genetics shape the gut microbiome. *Cell*, *159*(4), 789–799. https://doi.org/10.1016/j.cell.2014.09.053

Grillo, M. A., De Mita, S., Burke, P. V., Solórzano-Lowell, K. L. S., & Heath, K. D. (2016). Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in *Medicago truncatula*. *Evolution*, *70*(12), 2704–2717. https://doi.org/10.1111/evo.13095

Heath, K. D. (2010). Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution*, *64*(5), 1446–1458. https://doi.org/10.1111/j.1558-5646.2009.00913.x

Heath, K. D., & Nuismer, S. L. (2014). Connecting functional and statistical definitions of genotype by genotype interactions in coevolutionary studies. *Frontiers in Genetics*, *5*, 77. https://doi.org/10.3389/fgene.2014.00077

Heath, K. D., & Stinchcombe, J. R. (2014). Explaining mutualism variation: A new evolutionary paradox? *Evolution*, *68*(2), 309–317.

Heath, K. D., & Tiffin, P. (2009). Stabilizing mechanisms in a legume-rhizobium mutualism. *Evolution*, *63*(3), 652–662. https://doi.org/10.1111/j.1558-5646.2008.00582.x

Hoeksema, J. D. (2010). Ongoing coevolution in mycorrhizal interactions. *New Phytologist*, *187*(2), 286–300. https://doi.org/10.1111/j.1469-8137.2010.03305.x

Josephs, E. B., Stinchcombe, J. R., & Wright, S. I. (2017). What can genome-wide association studies tell us about the evolutionary forces maintaining genetic variation for quantitative traits? *New Phytologist*, *214*(1), 21–33.

Kanamori, N., Madsen, L. H., Radutoiu, S., Frantescu, M., Quistgaard, E. M. H., Miwa, H., Downie, J. A., James, E. K., Felle, H. H., Haaning, L. L., Jensen, T. H., Sato, S., Nakamura, Y., Tabata, S., Sandal, N., & Stougaard, J. (2006). A nucleoporin is required for induction of Ca^{2+} spiking in legume nodule development and essential for rhizobial and fungal symbiosis. *Proceedings of the National Academy of Sciences*, *103*(2), 359–364. https://doi.org/10.1073/pnas.0508883103

Kessner, D., Turner, T. L., & Novembre, J. (2013). Maximum likelihood estimation of frequencies of known haplotypes from pooled sequence data. *Molecular Biology and Evolution*, *30*(5), 1145–1158. https://doi.org/10.1093/molbev/mst016

Kiester, A. R., Lande, R., & Schemski, D. W. (1984). Models of coevolution and speciation in plants and their pollinators. *American Naturalist*, *124*, 220–243.

Klinger, C. R., Lau, J. A., & Heath, K. D. (2016). Ecological genomics of mutualism decline in nitrogen-fixing bacteria. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1826), 20152563. https://doi.org/10.1098/rspb.2015.2563

Lindström, K., & Mousavi, S. A. (2020). Effectiveness of nitrogen fixation in rhizobia. *Microbial Biotechnology*, *13*(5), 1314–1335. https://doi.org/10.1111/1751-7915.13517

MacPherson, A., Otto, S. P., & Nuismer, S. L. (2018). Keeping pace with the red queen: Identifying the genetic basis of susceptibility to infectious disease. *Genetics*, *208*(2), 779–789. https://doi.org/10.1534/genetics.117.300481

Masson-Boivin, C., & Sachs, J. L. (2018). Symbiotic nitrogen fixation by rhizobia—The roots of a success story. *Biotic Interactions*, *44*, 7–15. https://doi.org/10.1016/j.pbi.2017.12.001

Mendoza-Suárez, M. A., Geddes, B. A., Sánchez-Cañizares, C., Ramírez-González, R. H., Kirchhelle, C., Jorrin, B., & Poole, P. S. (2020). Optimizing Rhizobium-legume symbioses by simultaneous measurement of rhizobial competitiveness and N_2 fixation in nodules. *Proceedings of the National Academy of Sciences*, *117*(18), 9822–9831. https://doi.org/10.1073/pnas.1921225117

Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S. (2009). Association mapping: Critical considerations shift from genotyping to experimental design. *The Plant Cell*, *21*(8), 2194–2202. https://doi.org/10.1105/tpc.109.068437

Nelson, M., Guhlin, J., Epstein, B., Tiffin, P., & Sadowsky, M. J. (2018). The complete replicons of 16 *Ensifer meliloti* strains offer insights into intra-and inter-replicon gene transfer, transposon-associated loci, and repeat elements. *Microbial Genomics*, *4*(5), e000174.

Nelson, M. S., Chun, C. L., & Sadowsky, M. J. (2017). Type IV effector proteins involved in the Medicago-Sinorhizobium symbiosis. *Molecular Plant-Microbe Interactions*, *30*(1), 28–34. https://doi.org/10.1094/MPMI-10-16-0211-R

Oono, R., Anderson, C. G., & Denison, R. F. (2011). Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1718), 2698–2703. https://doi.org/10.1098/rspb.2010.2193

Ossler, J. N., & Heath, K. D. (2018). Shared genes but not shared genetic variation: legume colonization by two belowground symbionts. *The American Naturalist*, *191*(3), 395–406. https://doi.org/10.1086/695829

Ostrowski, E. A., Shen, Y., Tian, X., Sucgang, R., Jiang, H., Qu, J., Katoh-Kurasawa, M., Brock, D. A., Dinh, C., Lara-Garduno, F., Lee, S. L., Kovar, C. L., Dinh, H. H., Korchina, V., Jackson, L., Patil, S., Han, Y., Chaboub, L., Shaulsky, G., … Queller, D. C. (2015). Genomic signatures of cooperation and conflict in the social amoeba. *Current Biology*, *25*(12), 1661–1665. https://doi.org/10.1016/j.cub.2015.04.059

Paape, T., Bataillon, T., Zhou, P., Kono, T. J. Y., Briskine, R., Young, N. D., & Tiffin, P. (2013). Selection, genome-wide fitness effects and evolutionary rates in the model legume *Medicago truncatula*. *Molecular Ecology*, *22*(13), 3525–3538. https://doi.org/10.1111/mec.12329

Parker, M. A. (1999). Mutualism in metapopulations of legumes and rhizobia. *The American Naturalist*, *153*(S5), S48–S60. https://doi.org/10.1086/303211

Parker, M. P. (1995). Plant fitness variation caused by different mutualist genotypes. *Ecology*, *76*(5), 1525–1535. https://doi.org/10.2307/1938154

Pecrix, Y., Staton, S. E., Sallet, E., Lelandais-Brière, C., Moreau, S., Carrère, S., Blein, T., Jardinaud, M., Latrsse, D., Zouine, M., Zahm, M., Chantry-Darmon, C., Lopez-Roquest, C., Bouchez, O., Bérard, A., Debelle, F., Muños, S., Bendahmane, A., Bergès, H., … Gamas, P. (2018). Whole-genome landscape of Medicago truncatula symbiotic genes. *Nature Plants*, *4*(12), 1017–1025. https://doi.org/10.1038/s41477-018-0286-7

Porter, S. S., & Simms, E. L. (2014). Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecology Letters*, *17*(9), 1121–1129. https://doi.org/10.1111/ele.12318

Pritchard, J. K., Pickrell, J. K., & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Current Biology*, *20*(4), R208–R215. https://doi.org/10.1016/j.cub.2009.11.055

Queller, D. C., & Strassmann, J. E. (2018). Evolutionary Conflict. *Annual Review of Ecology, Evolution, and Systematics*, *49*(1), 73–93. https://doi.org/10.1146/annurev-ecolsys-110617-062527

Quides, K. W., Weisberg, A. J., Trinh, J., Salaheldine, F., Cardenas, P., Lee, H.-H., Jariwata, R., Chang, J. H., & Sachs, J. L. (2021). Experimental evolution can enhance benefits of rhizobia to novel legume hosts. *Proceedings of the Royal Society B: Biological Sciences*, *288*(1951), 20210812. https://doi.org/10.1098/rspb.2021.0812

Rangin, C., Brunel, B., Cleyet-Marel, J.-C., Perrineau, M.-M., & Béna, G. (2008). Effects of *Medicago truncatula* genetic diversity, rhizobial competition, and strain effectiveness on the diversity of a natural Sinorhizobium species community. *Applied and Environmental Microbiology*, *74*(18), 5653–5661.

Ratcliff, W. C., Underbakke, K., & Denison, R. F. (2011). Measuring the fitness of symbiotic rhizobia. *Symbiosis*, *55*(2), 85–90. https://doi.org/10.1007/s13199-011-0150-2

Regus, J. U., Quides, K. W., O'Neill, M. R., Suzuki, R., Savory, E. A., Chang, J. H., & Sachs, J. L. (2017). Cell autonomous sanctions in legumes target ineffective rhizobia in nodules with mixed infections. *American Journal of Botany*, *104*(9), 1299–1312. https://doi.org/10.3732/ajb.1700165

Riley, A., Grillo, M., Epstein, B., Tiffin, P., & Heath, K. D. (2022). Discordant population structure within rhizobium divided genomes and between rhizobia and legume hosts in their native range. *Authorea*. https://doi.org/10.22541/au.164637769.91740255/v1

Roy, S., Liu, W., Nandety, R. S., Crook, A., Mysore, K. S., Pislariu, C. I., Frugoli, J., Dickstein, R., & Udvardi, M. K. (2020). Celebrating 20 Years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *The Plant Cell*, *32*(1), 15–41. https://doi.org/10.1105/tpc.19.00279

Rutten, P. J. H., Steel, G. A., Hood, V. K., Ramachandran, L., McMurtry, B., Geddes, B., Papachristodoulou, A., & Poole, P. S. (2021). Multiple sensors provide spatiotemporal oxygen regulation of gene expression in a Rhizobium-legume symbiosis. *PLoS Genetics.*, *17*, e1009099. https://doi.org/10.1371/journal.pgen.1009099

Ruvkun, G. B., Sundaresan, V., & Ausubel, F. M. (1982). Directed transposon Tn5 mutagenesis and complementation analysis of rhizobium meliloti symbiotic nitrogen fixation genes. *Cell*, *29*(2), 551–559. https://doi.org/10.1016/0092-8674(82)90171-4

Sachs, J. L., Essenberg, C. J., & Turcotte, M. M. (2011). New paradigms for the evolution of beneficial infections. *Trends in Ecology & Evolution*, *26*(4), 202–209. https://doi.org/10.1016/j.tree.2011.01.010

Sachs, J. L., Quides, K. W., & Wendlandt, C. E. (2018). Legumes versus rhizobia: A model for ongoing conflict in symbiosis. *New Phytologist*, *219*(4), 1199–1206. https://doi.org/10.1111/nph.15222

Saturno, D. F., Cerezini, P., Moreira da Silva, P., Oliveira, A. B., Oliveira, M. C. N., Hungria, M., & Nogueira, M. A. (2017). Mineral nitrogen impairs the biological nitrogen fixation in soybean of determinate and indeterminate growth types. *Journal of Plant Nutrition*, *40*(12), 1690–1701. https://doi.org/10.1080/01904167.2017.1310890

Siewert, K. M., & Voight, B. F. (2020). BetaScan2: Standardized statistics to detect balancing selection utilizing substitution data. *Genome Biology and Evolution*, *12*(2), 3873–3877. https://doi.org/10.1093/gbe/evaa013

Simonsen, A. K., & Stinchcombe, J. R. (2014). Standing genetic variation in host preference for mutualist microbial symbionts. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1797), 20142036. https://doi.org/10.1098/rspb.2014.2036

Somasegaran, P., & Bohlool, B. B. (1990). Single-strain versus multistrain inoculation: effect of soil mineral n availability on rhizobial strain effectiveness and competition for nodulation on chick-pea, soybean, and dry bean. *Applied and Environmental Microbiology*, *56*(11), 3298–3303. https://doi.org/10.1128/aem.56.11.3298-3303.1990

Sprent, J. I. (2008). 60Ma of legume nodulation. What's new? What's changing? *Journal of Experimental Botany*, *59*(5), 1081–1084. https://doi.org/10.1093/jxb/erm286

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., Bharti, A. K., Farmer, A. D., Zhou, P., Cenny, R., May, G. D., Erlandson, S., Yakub, M., Sugawara, M., Sadkowsky, M. J., Young, N. D., & Tiffin, P. (2013). Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS One*, *8*(5), e65688. https://doi.org/10.1371/journal.pone.0065688

Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., Gyaneshwar, P., Denny, R., Mudge, J., Bharti, A. K., Parmer, A. D., May, G. D., Woodward, J. E., Médigue, C.,Vallenet, D., Laius, A. l., Rouy, Z., Martinez-Vaz, B., Tiffin, P., … Sadwosky, M. J. (2013). Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. *Genome Biology*, *14*(2), R17.

Thrall, P. H., Hochberg, M. E., Burdon, J. J., & Bever, J. D. (2007). Coevolution of symbiotic mutualists and parasites in a community context. *Trends in Ecology & Evolution*, *22*(3), 120–126. https://doi.org/10.1016/j.tree.2006.11.007

Torkamaneh, D., Chalifour, F.-P., Beauchamp, C. J., Agrama, H., Boahen, S., Maaroufi, H., Rajcan, I., & Belzile, F. (2020). Genome-wide association analyses reveal the genetic basis of biomass accumulation under symbiotic nitrogen fixation in African soybean. *Theoretical and Applied Genetics*, *133*(2), 665–676. https://doi.org/10.1007/s00122-019-03499-7

Veyron, S., Peyroche, G., & Cherfils, J. (2018). FIC proteins: From bacteria to humans and back again. *Pathogens and Disease*, *76*(2), fty012. https://doi.org/10.1093/femspd/fty012

Voichek, Y., & Weigel, D. (2020). Identifying genetic variants underlying phenotypic variation in plants without complete genomes. *Nature Genetics*, *52*(5), 534–540. https://doi.org/10.1038/s41588-020-0612-7

Wade, M. J. (2007). The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics*, *8*(3), 185–195. https://doi.org/10.1038/nrg2031

Wang, D., Yang, S., Tang, F., & Zhu, H. (2012). Symbiosis specificity in the legume – rhizobial mutualism. *Cellular Microbiology*, *14*(3), 334–342. https://doi.org/10.1111/j.1462-5822.2011.01736.x

Weese, D. J., Heath, K. D., Dentinger, B. T. M., & Lau, J. A. (2015). Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution*, *69*(3), 631–642. https://doi.org/10.1111/evo.12594

Wendlandt, C. E., Helliwell, E., Roberts, M., Nguyen, K. T., Friesen, M. L., von Wettberg, E., Price, P., Griffitts, J. S., & Porter, S. S. (2021). Decreased coevolutionary potential and increased symbiont fecundity during the biological invasion of a legume-rhizobium mutualism. *Evolution*, *75*(3), 731–747. https://doi.org/10.1111/evo.14164

Wendlandt, C. E., Regus, J. U., Gano-Cohen, K. A., Hollowell, A. C., Quides, K. W., Lyu, J. Y., Adinata, E. S., & Sachs, J. L. (2019). Host investment into symbiosis varies among genotypes of the legume *Acmispon strigosus*, but host sanctions are uniform. *New Phytologist*, *221*(1), 446–458. https://doi.org/10.1111/nph.15378

West, S. A., Kiers, E. T., Simms, E. L., & Denison, R. F. (2002). Sanctions and mutualism stability: Why do rhizobia fix nitrogen? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1492), 685–694. https://doi.org/10.1098/rspb.2001.1878

Yoder, J. B. (2016). Understanding the coevolutionary dynamics of mutualism with population genomics. *American Journal of Botany*, *103*(10), 1742–1752. https://doi.org/10.3732/ajb.1600154

Young, J. P. W., & Johnston, A. W. B. (1989). The evolution of specificity in the legume-rhizobium symbiosis. *Trends in Ecology & Evolution*, *4*(11), 341–349. https://doi.org/10.1016/0169-5347(89)90089-X

Zhao, Y., Huang, J., Wang, Z., Jing, S., Wang, Y., Ouyang, Y., Cai, B., Zin, X-F., Lieu, X., Zhang, C., Pan, Y., Ma, R., Li, O., Jiang, W., Zeng, Y., Shangguan, X., Wang, H., Du, B., Zhu, L., … He, G. (2016). Allelic diversity in an NLR gene BPH9 enables rice to combat planthopper variation. *Proceedings of the National Academy of Sciences*, *113*(45), 12850–12855. https://doi.org/10.1073/pnas.1614862113

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, *44*(7), 821–824.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Epstein, B., Burghardt, L. T., Heath, K. D., Grillo, M. A., Kostanecki, A., Hämälä, T., Young, N. D., & Tiffin, P. (2022). Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis. *Molecular Ecology*, *00*, 1–14. https://doi.org/10.1111/mec.16602