



Mobile gene clusters and coexpressed plant–rhizobium pathways drive partner quality variation in symbiosis

Muhammad Rizwan Riaz^{a,1} , Ivan Sosa Marquez^{b,c,1} , Hanna Lindgren^a , Garrett Levin^d , Rebecca Doyle^{a,e}, Mario Cerón Romero^a, Julia C. Paoli^g , Jenny Drnevich^f , Christopher J. Fields^f , Barney A. Geddes^d , Amy Marshall-Colón^{a,c,2} , and Katy D. Heath^{a,c,2}

Affiliations are included on p. 9.

Edited by Éva Kondorosi, Hungarian Academy of Sciences, Biological Research Centre, Szeged, Hungary; received June 13, 2024; accepted June 18, 2025

Plant–microbe symbioses such as the legume–rhizobium mutualism are vital in the web of ecological relationships within both natural and managed ecosystems, influencing primary productivity, crop yield, and ecosystem services. The outcome of these interactions for plant hosts varies quantitatively and can range from highly beneficial to even detrimental depending on natural genetic variation in microbial symbionts. Here, we take a systems genetics approach, harnessing the genetic diversity present in wild rhizobial populations to predict genes and molecular pathways crucial in determining partner quality, i.e., the benefits of symbiosis for legume hosts. We combine traits, dual-RNAseq of both partners from active nodules, pangenomics/pantranscriptomics, and Weighted Gene Co-expression Network Analysis (WGCNA) for a panel of 20 *Sinorhizobium meliloti* strains that vary in symbiotic partner quality. We find that genetic variation in the nodule transcriptome predicts host plant biomass, and WGCNA reveals networks of genes in plants and rhizobia that are coexpressed and associated with high-quality symbiosis. Presence–absence variation of gene clusters on the symbiosis plasmid (pSymA), validated in planta, is associated with high or low-quality symbiosis and is found within important coexpression modules. Functionally our results point to management of oxidative stress, amino acid and carbohydrate transport, and NCR peptide signaling mechanisms in driving symbiotic outcomes. Our integrative approach highlights the complex genetic architecture of microbial partner quality and raises hypotheses about the genetic mechanisms and evolutionary dynamics of symbiosis.

symbiosis | systems genetics | dual-RNAseq | horizontal gene transfer | partner quality

Plant–microbe symbioses form the foundation of a complex network of ecological interactions that drive primary productivity in both natural and managed ecosystems. These interactions can have positive, negative, or little impact on one or both species involved (1, 2) with mutualistic symbioses resulting in fitness benefits to both partners (3). Despite these categorical distinctions, the fitness benefits to partners in any mutualistic interaction vary quantitatively (4, 5) along a continuum depending on the environmental context (6–8) and the genotype of one or both partners (5, 9–15). Because the fitness-determining traits in symbiosis are often microscopic or biochemical, a predictive understanding of symbiosis in ecology, evolution, and agriculture requires a mechanistic approach that is grounded in functional genetics but leverages natural genetic variation (14).

Quantitative genetic trait variation, the focus of targeted breeding programs and selection in natural populations (16), is underlain by SNP variants, presence–absence variation (PAV) of genes and gene regions, and dynamic gene expression (17, 18). Despite the dominance of the single-locus central dogma viewpoint in bacterial genetics to date, this quantitative framework is salient for bacterial traits for several reasons. In contrast to sexual eukaryotes, for which genotype–phenotype mapping studies were developed (19, 20), bacterial recombination and genetic variation are not driven by reproduction but occur horizontally (21) as a result of both homologous and nonhomologous mechanisms of recombination, often involving mobile genetic elements (MGEs) (22). As a result, bacterial genomes are often highly dynamic and require reference-quality genome assemblies for optimum variant calling, MGE detection, and annotation of the full pangenome (all genes in a population) (23–27). What's more, many symbiotic bacteria have multipartite genomes, which consist of two or more large replicons [i.e., chromosome, megaplasmids, or chromids (28, 29)], which can have vastly different demographic processes, genetic variation, recombination, and rates of PAV (5, 30–33). The expression of these multipartite genomes is highly plastic, with bacteria leveraging different genes or even whole replicons depending on the environment (34–36). Thus, a thorough understanding of ecologically and agriculturally relevant symbiosis requires that we quantify how pangenomic variation

Significance

Addressing challenges from sustainable agriculture to ecosystem resilience is only possible by integrating symbiotic microorganisms into our biological understanding. The effects of symbionts on hosts percolate up from genomes through expression and metabolism to result in complex traits including nitrogen fixation and plant health. Here, we leverage a model nitrogen-fixing symbiosis, integrating genomes, traits, and gene expression to discover novel genes, gene clusters, and associated metabolic and regulatory processes that are important for the variable effects of bacteria on plant growth. We provide functional validation and insight into variation in symbiotic functioning and foundational information on the inheritance of important traits in microbial symbiosis. This raises exciting questions about how these keystone symbionts evolve in nature.

Author contributions: M.R.R., I.S.M., H.L., B.A.G., A.M.-C., and K.D.H. designed research; M.R.R., I.S.M., H.L., G.L., R.D., M.C.R., J.C.P., B.A.G., A.M.-C., and K.D.H. performed research; G.L., M.C.R., and B.A.G. contributed new reagents/analytic tools; M.R.R., I.S.M., R.D., M.C.R., J.D., C.J.F., A.M.-C., and K.D.H. analyzed data; and M.R.R., I.S.M., H.L., G.L., R.D., J.C.P., J.D., C.J.F., B.A.G., A.M.-C., and K.D.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under [Creative Commons Attribution License 4.0 \(CC BY\)](https://creativecommons.org/licenses/by/4.0/).

¹M.R.R. and I.S.M. contributed equally to this work.

²To whom correspondence may be addressed. Email: amymc@illinois.edu or kheath@illinois.edu.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2411831122/-/DCSupplemental>.

Published July 29, 2025.

percolates up through gene expression networks and metabolic pathways to generate phenotypic variation.

Like all eukaryotes, plants require nitrogen (N) in accessible forms (nitrates and ammonia), which is often limiting in the soil, restricting growth. Leguminous plants overcome this limitation by forming symbiosis with a polyphyletic group of bacteria known as rhizobia, which inhabit root nodules and therein convert atmospheric N (N_2) into plant-usable forms (3, 37–39). The legume–rhizobium symbiosis begins with symbiont recognition and infection, leading to nodule formation (40, 41), differentiation of rhizobia into bacteroids (37, 42), and finally resource exchange between partners (35, 43). The resulting mutualism is a key contributor to the nonanthropogenic N cycle, controlling N availability for legume crops, nonlegume crops through rotation, and primary producers in natural terrestrial ecosystems (44, 45).

The interaction between *Medicago truncatula* and *Sinorhizobium meliloti* is a critical model for studying symbiotic nitrogen fixation. Decades of work on this model legume–rhizobium symbiosis have revealed much about the genetics and biochemistry governing the interaction (46–49), and the effects of both host and symbiont genetic variation present in natural populations on host growth and fitness, sometimes called partner quality (5, 50, 51). Functional genetic and –omics studies have culminated in hundreds of genes known to be required for symbiosis (52, 53) and an increasingly detailed understanding of symbiotic metabolism (54–57). Natural populations harbor abundant variation in symbiotic compatibility and efficiency maintained via largely unknown genetic and evolutionary processes occurring in recognition and signaling genes, metabolic pathways, and other molecular mechanisms (36, 38, 58). Our understanding of these genetic underpinnings of partner quality variation in nature remains nascent, though this knowledge holds promise for manipulating symbiotic relationships to enhance agricultural productivity and for predicting how environmental changes will impact nitrogen cycling. Thus, in addition to molecular genetic approaches that typically focus on one or a handful of well-characterized genotypes, studying the genetic variation in the symbiotic transcriptome across genotypes from natural populations grown in a controlled, common garden design can provide a handle with which to resolve novel pathways involved in high-quality symbiosis, while also highlighting which known pathways are variable and thus likely to be a source of evolutionary innovation in natural populations.

Here, we take a systems genetics approach (59–62) leveraging standing genetic variation present in wild populations of rhizobia, transcriptomics using dual-RNAseq of host and symbiont gene expression from active symbiotic nodules (34), and pangenomes for a panel of 20 wild rhizobial isolates known to vary in partner quality (5), in order to predict the underlying molecular pathways/genes most related to partner quality in nitrogen-fixing symbiosis. We show how plant and symbiont gene expression networks interact with the presence–absence of functional gene clusters to drive mutualistic partner quality variation.

Results

Variation in the Nodule Transcriptome and Presence–Absence of Gene Clusters on pSymA Predict Host Plant Biomass. To examine the molecular mechanisms by which rhizobial genetic diversity influences partner quality with its host plant, we inoculated *M. truncatula* cv. DZA 315.16 (hereafter DZA) plants with 20 individual rhizobia strains (*SI Appendix, Fig. S1*) from natural populations of *S. meliloti*; these strains were previously shown to induce a diverse range of phenotypes on this host (5). In the present study, we measured 10 phenotypes from host plants

(*Dataset S1*) (10 biological replicates per *S. meliloti* strain). A trait correlation network of the phenotypes showed that plant shoot biomass was positively correlated with root biomass, chlorophyll, leaf number, and height, and negatively correlated with nodule number (*SI Appendix, Fig. S2*). Plant shoot biomass was highly variable across strains ($\sigma^2 = 0.024$; CV = 0.60; $P < 0.0001$; *SI Appendix, Fig. S3*). Hereafter, we refer to strains that confer high to low biomass accumulation as high-, intermediate-, and low-quality partners, respectively (48, 50).

We obtained symbiotic transcriptomes by extracting nodule RNA from a subset of four biological replicates per strain for RNA-seq analysis (Illumina paired-end reads ranging from fifty-eight to hundred million reads). We first mapped transcripts to *M. truncatula* [A17 *r5.0* downloaded from the French National Research Institute for Agriculture, Food and Environment Medicago Bioinformatics Resources (63)] and *S. meliloti* (reference strain 1021 assembly GCF_000006965.1 ASM696v1) (64, 65). Approximately 44 to 87% of total reads were mapped with this reference-based analysis (includes coding and noncoding genes), where 12.4 to 96.3% of those reads mapped to *Medicago*, depending on the sample (*Dataset S2*). To uncover how natural genetic variation across rhizobium strains influences the expressed genome, we first used principal component analysis (PCA) to find that partner quality variation (same replicates as RNAseq; Fig. 1A) was, overall, associated with transcriptomic variation. Importantly, PCA on strain means revealed that *S. meliloti* PC1 explains 16.5% of the variability in the transcriptome and separates strains by partner quality (Fig. 1B). Similarly, Medicago PC1 and PC2 together explain 48.1% of transcriptome variability and separate strains by partner quality (Fig. 1C).

Most annotated genes (77.7% of the *S. meliloti* 1021 genome) are commonly expressed (counts ≥ 10 in all replicates) across strains, while 3% of genes are not expressed in any strain (counts < 10 in all replicates). Examination of the remaining genes (19.3%) revealed on/off patterns of expression where certain genes are expressed in some strains and not expressed in others (*SI Appendix, Fig. S4*). The binary nature of these genes affords us the opportunity to perform a discrete analysis of functional expression variation across strains. Thus, we imposed a filtering criterion to obtain a refined list of genes that are only expressed in at least three but not more than 17 strains, resulting in 574 genes with on/off expression patterns. Of these, 82 genes significantly correlate with shoot biomass ($|r| > 0.5$, $\text{fdr} < 0.15$).

To determine whether the on/off expression patterns were caused by PAV in the bacterial genomes, whereby the lack of expression would be the result of not having the gene, we generated high-quality de novo assemblies of all 20 strains using PacBio long-read sequencing. *BLAST+* sequence search of these closed, reference-quality genomes confirmed that 54 of the 82 genes with on/off expression patterns are also PAVs based on the 1021 reference (*Dataset S3*; see *Materials and Methods* for gene selection criteria). Next, we performed a pangenome analysis to establish the full set of core and variable gene content (in our 20 strains, plus the 1021 reference strain and the newly released reference strain MABNR56 assembly GCF_037023865.1 ASM3702386v1). The fraction of the core (present in all strains), shell (variable presence across strains), and cloud (rare) gene groups varied among the main genome elements; pSymA contained a much smaller fraction (19.9%) of core genes compared to either the chromosome (56.1%) or pSymB (67.9%; *SI Appendix, Fig. S5*). In addition, eight strains feature a smaller, fourth accessory plasmid (BioProject PRJNA1009820). We found 60 additional open reading frames (ORFs) that were present at intermediate frequency across the 20 wild strains, absent in the 1021 reference, and significantly correlated with shoot biomass ($|r| > 0.5$, $\text{fdr} < 0.15$).

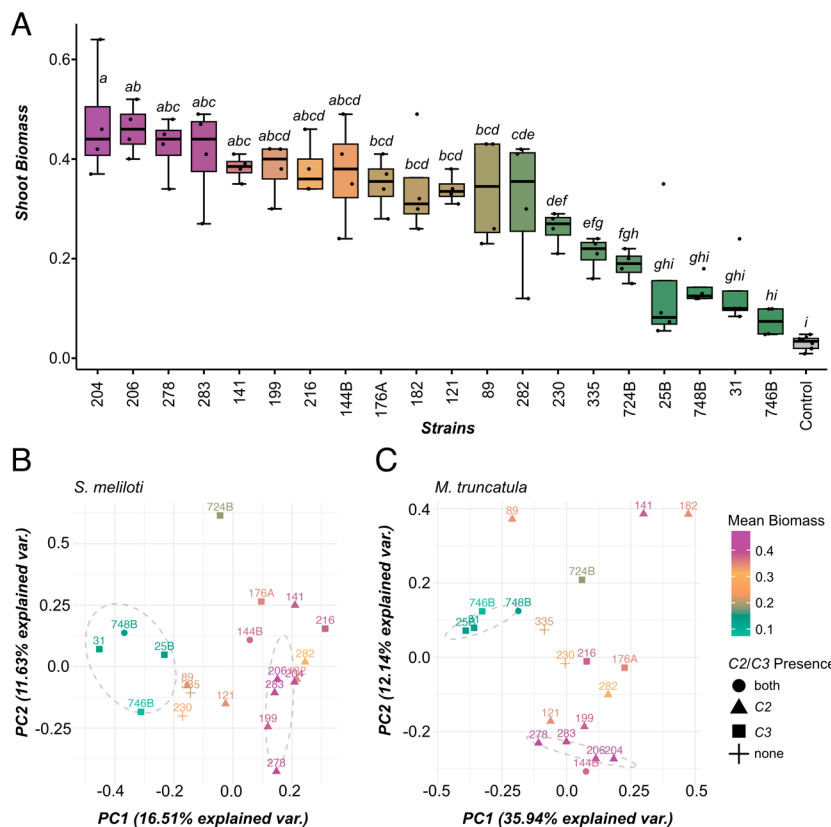


Fig. 1. Partner quality of 20 *S. meliloti* strains and PCA of nodule transcriptomes. (A) Boxplots present the distribution of plant biomass when *M. truncatula* host plants were inoculated with one of 20 strains (four replicates each) or remained uninoculated (control). Letter-based groupings of means were assigned by ANOVA followed by Duncan's test ($\alpha = 0.05$). (B) PCA plot derived from the expression of 5,993 *S. meliloti* genes (mapped to *S. meliloti* 1021 reference). (C) PCA plot derived from the expression of the 5,000 most highly variable genes of *M. truncatula*. The transcriptome profiles are quantified using average \log_{CPM} values of expressed genes. Each data point on the PCA plot corresponds to a strain; data points closer together are more similar in terms of gene expression patterns than those further away. Each data point is color-coded according to the mean biomass value of the respective strain. Data point shape represents the presence or absence of distinct gene clusters (C2, C3) in strains.

The 54 “reference-based” and additional 60 “pangenome” PAV loci of interest were largely found on pSymA, versus pSymB or the chromosome. Fifty-one “reference-based” pSymA genes are clustered into six syntenic groups, henceforth called clusters C1 (2 genes), C2 (4 genes), C3 (10 genes), C4 (14 genes), C5 (17 genes), and C6 (4 genes) (Fig. 2A and Dataset S3)—plus one single IS3-type transposase. An additional two transposases that form part of these 54 genes are located in pSymB (IS630-like element) and the chromosome (IS256-like element). Thirty-three “pangenome” PAV loci were also clustered into six groups (defined as two or more ORFs within 3 kb of each other) on pSymA, henceforth called clusters pan-C1 (2 ORFs), pan-C2 (8 ORFs), pan-C3 (3 ORFs), pan-C4 (2 ORFs), pan-C5 (11 ORFs), pan-C6 (7 ORFs) (Fig. 2A and Dataset S3). Clusters C1, C2, C4, C6, pan-C4, and pan-C6 are present in high-quality strains, while clusters C3, C5, pan-C1, pan-C2, pan-C3, and pan-C5 are mostly present in low-quality strains (Fig. 2B and Dataset S3). Strains that show intermediate phenotypes have different combinations of high- and low-quality associated gene clusters (strains in the center of Fig. 2B).

Reference-Based Clusters for Validation and Functional Analysis. Focusing on the clusters found using the 1021 annotation, since this strain can be readily manipulated (53, 66), the presence/absence patterns of C2 and C3 were most associated with host DZA shoot biomass across the 20 *S. meliloti* strains (Fig. 2C). We confirmed the relevance of these gene clusters in partner quality

variation by deleting clusters C2 (1021- Δ C2) and C3 (1021- Δ C3) in reference strain 1021. Consistent with our prediction based on our 20 strains, *Medicago* cv. DZA inoculated with 1021- Δ C2 had a 42% decrease in plant biomass compared to DZA inoculated with 1021 ($P = 0.024$; Fig. 2D). While we expected loss of Cluster 3 to increase DZA plant biomass, given its association with low-quality strains in this natural population, we found that plants inoculated with 1021- Δ C3 were not statistically different from those inoculated with 1021 ($P = 0.124$; Fig. 2D).

Cluster C2 (286,919 to 288,703 in ref. *S. meliloti* 1021) has functions related to N export, flavonoid signal transduction, and nodulation and includes an *Lrp/AsnC* family transcriptional regulator (*WP_234826226.1*). This TF family has been shown in *Escherichia coli* and *Bacillus subtilis* to have global and/or specific regulatory functions; in *E. coli*, AsnC is known to regulate asparagine synthetase (*asnB*). Asparagine has been reported as an important form of N produced in nodules and used by host legumes, and thus could be critical for host benefits. Cluster C2 also contains two SnoL-like domain-containing proteins (*WP_010967276.1*/UniProt: Q930B7, *WP_010967277.1*) of the NTF2 superfamily (IPR032710) that can reportedly interact with proteins binding to flavonoid molecules involved in the communication between a close relative to *S. meliloti*, *Rhizobium etli*, and its host plant. There is also a class I SAM-dependent methyltransferase protein (*WP_010967274.1*/UniProt: B3KLX3/GO:0008757) that is related to NodS (SAM) proteins involved in oligosaccharide biosynthetic processes, although the specific function of this gene in *Medicago-Sinorhizobium* symbiosis is unknown.

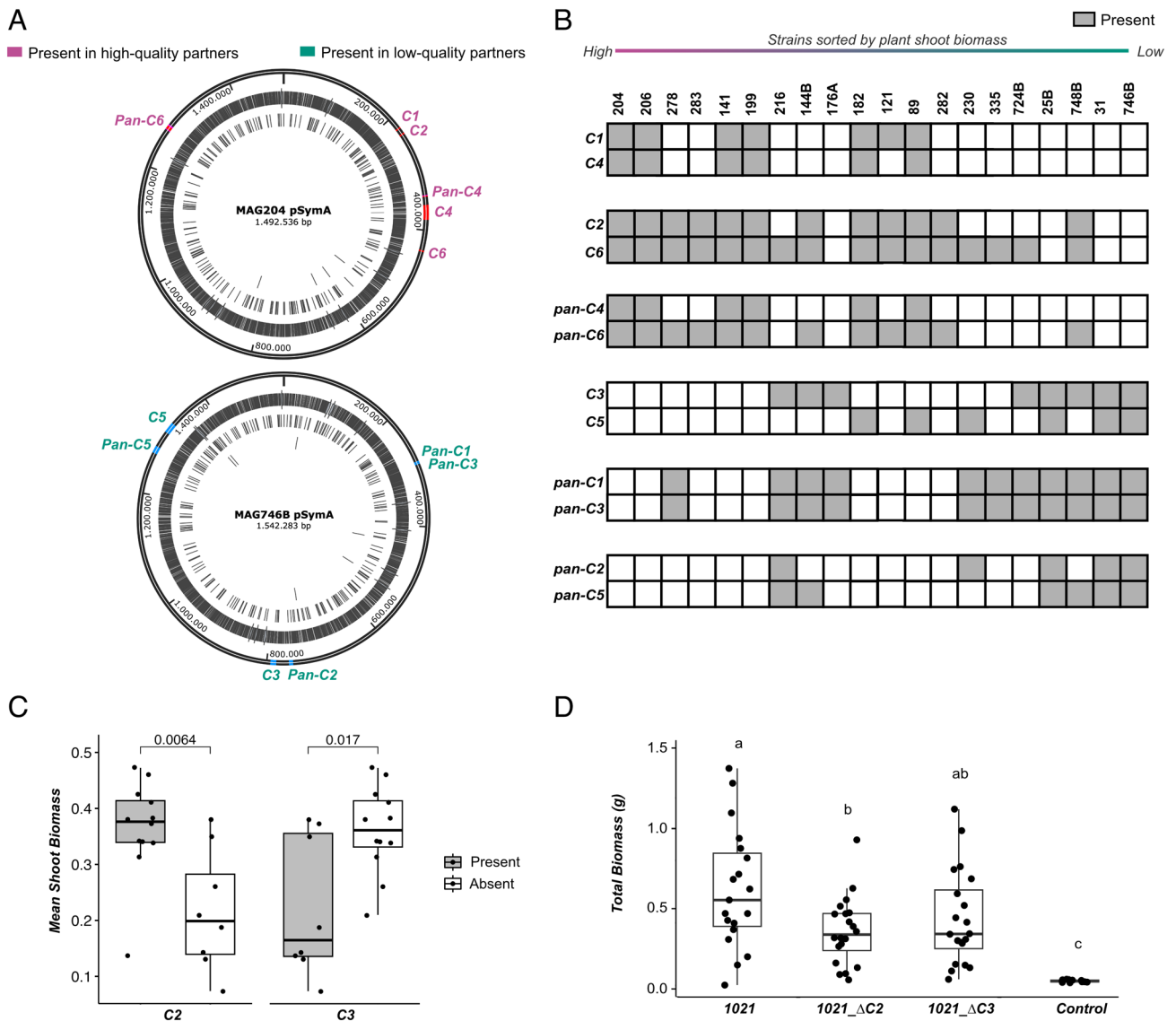


Fig. 2. Gene cluster PAV associated with partner quality. (A) Circular map of the symbiotic plasmid (pSymA) of MAG204 and MAG746B showing the location of 1021-based (C1-C6) and pangenome-based (*panC1-panC6*) gene clusters. (B) Matrix displaying the presence or absence of gene clusters. Rows represent cluster names (C1-C6, *panC1-panC6*), and columns represent strains, sorted from high to low shoot biomass. Gray boxes indicate the presence of the cluster in the strain. (C) Mean shoot biomass of *M. truncatula* in symbiosis with strains in which clusters C2 and C3 are either present (gray) or absent (white). Each data point represents the mean shoot biomass of a strain (20 strains from the current experiment), and group means are compared through an independent *t* test. Significance levels are shown for each comparison. (D) Total biomass of *M. truncatula* inoculated with 1021, deletion mutants of clusters C2 (1021_ΔC2) and C3 (1021_ΔC3), or mock inoculated (control). Significantly different means are based on Tukey post hoc test.

PAV on the pSymA Phylogeny Indicates Mobility of Gene Clusters. Heterogeneity in the PAV of gene clusters within well-defined clades of the core pSymA phylogeny (composed of the superset of 191 strains; *SI Appendix, Fig. S6A*) indicates the gain and loss of clusters through time, suggesting mobilization of these genetic elements via unidentified mechanisms. For example, a closer inspection of the pattern of both PAV and copy number of Cluster C2, even within our 20-strain subset, suggests cluster duplication and gene gain/loss through the evolutionary history of pSymA. First, the presence of C2 is scattered across the core strain phylogeny (strains in bold in *SI Appendix, Fig. S6B*). Second, four strains have two copies, suggesting recent acquisition of a second cluster. Third, drilling down even further by creating a phylogeny of the Cluster C2 itself (using an alignment of C2 gene sequences from strains where the cluster is present; *SI Appendix, Fig. S6C*) reveals two things: that the topology of the C2 phylogeny is distinct from the pSymA core genome topology, and that the C2

copies found in the same strains are not recent duplicates – again indicating independent movement of C2 relative to pSymA.

The Nodule Pantranscriptome Reveals Coexpression Modules That Correlate with Gene Clusters on pSymA and Partner Quality. Next, we explored pantranscriptomic variation by realigning our RNAseq data to the *S. meliloti* pangenome and *M. truncatula* reference and employed coexpression analyses to study how variation in bacterial gene expression is related to patterns of PAV in gene clusters and strain partner quality. Of all 12,107 gene groups identified in the pangenome, 39.6% were commonly expressed (counts ≥ 10 in all replicates), 4.7% were not expressed, and 55.7% were variably expressed across strains (*Dataset S4*). While ORFs in the core genome were more highly expressed than those in the accessory (shell and cloud) genome, there was higher dispersion (variance in gene expression among strains) in the accessory genome (*SI Appendix, Fig. S7*).

Our weighted gene coexpression network analysis (WGCNA) grouped genes into 31 modules representing the genetic variation in the bacterial gene network (Fig. 3A and *SI Appendix*, Fig. S8)

(67). Modules vary in size from 10 to 2,145 genes, and some are enriched for either the chromosome, pSymA, or pSymB (Fig. 3A). Next, we correlated these modules with partner quality traits. The

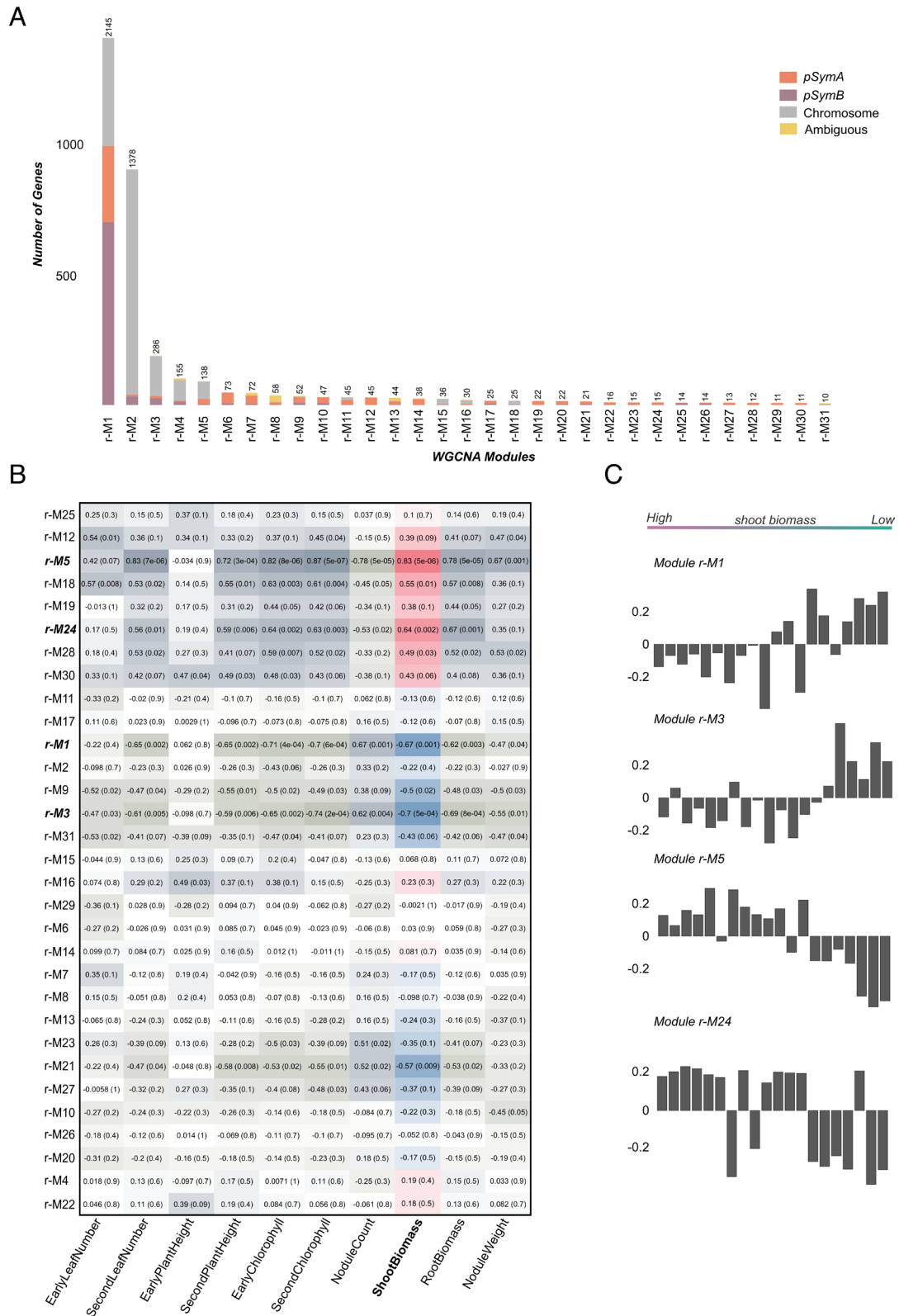


Fig. 3. WGCNA of *S. meliloti* transcriptome. (A) Barplot showing the number of genes per WGCNA module, ordered by module size. The colors indicate the proportion of chromosomal and plasmid genes in each module. (B) Heatmap displaying the correlation coefficient (and *P*-value) between module eigengenes (ME) (rows) and plant traits (columns). The modules are sorted based on their similarity to each other. Highly significant modules ($|r| > 0.6$) are bolded and italicized. Correlations between modules and host plant biomass are highlighted (pink: positive; blue: negative). (C) The bar plots display eigengene values for significantly correlated modules for each strain, sorted from high to low shoot biomass along the *X* axis. Expression of genes in modules M5 and M24 are positively correlated with shoot biomass, while those in M1 and M3 are negatively correlated.

largest modules were dominated by chromosomal and pSymB genes and tended to be negatively correlated with biomass (Fig. 3 *A* and *B*); of 31 modules, only two were significantly positively associated with shoot biomass, while two were negatively correlated with shoot biomass ($|r| > 0.6$, $P < 0.05$) (Fig. 3 *B* and *C*). Genes were labeled as ambiguous when they were found on different genomic elements (e.g., chromosome, pSymA, pSymB, or accessory plasmids) across the strains in our study. Interestingly, two modules (r-M8 and r-M13) feature many of these potentially mobile genes, though neither module was significantly associated with plant biomass (Fig. 3 *A* and *B*). Three pangenome PAV clusters, pan-C2, pan-C4, and pan-C5 were filtered out from WGCNA (median expression < 10 counts); however, while these three clusters were present in only 5 to 6 strains, they were expressed when present (Dataset S5).

The two modules positively correlated with host shoot biomass include r-M5 ($|r| = 0.83$, P -value $4e-06$), and r-M24 ($|r| = 0.64$, P -value 0.002) (Fig. 3*B*). These modules were closely related in hierarchical clustering of WGCNA modules (SI Appendix, Fig. S8); r-M5 is enriched with chromosomal genes, and r-M24 is enriched with pSymA genes (Fig. 3*A* and Dataset S6). Notably, r-M24 includes many of the transcriptional regulation related PAV genes of clusters C2 and pan-C6, which are coexpressed with genes encoding the type II toxin-antitoxin (TA) system. This includes *RelB/E* family loci [WP_010967242.1 (A), WP_010967243.1 (T)] located on pSymA. Deletion of these two genes in a different symbiotic context (RmP110 with congeneric *Medicago sativa*) was not found to significantly affect N fixation (68), though other TA systems are known to influence symbiosis (69, 70), and we found that knocking out C2 in Sm1021 did decrease plant biomass. Apart from TA systems, r-M24 also includes Clusters of Orthologous Groups (COG) terms for transcriptional regulation such as *Lrp/AsnC* (WP_014532001.1) and arabinose regulatory mechanisms (*GlxA*; WP_014989851.1, and *AraF*; WP_010975369.1), and an SRPBCC-domain containing protein (WP_127535573), an effector protein shown to promote symbiosis (71). However, no COG terms were significantly enriched in r-M24. Module r-M5 is enriched in COG terms for energy production and conversion ($fdr < 0.0001$) and coenzyme transport and metabolism ($fdr < 0.0001$), and includes many important N fixation genes on pSymA including *fixABCX* and *nifABDEHKNX* (SI Appendix, Fig. S9 and Datasets S6 and S7) (53, 72, 73).

The two modules negatively correlated with host shoot biomass include r-M1 ($|r| = 0.67$, P -value 0.001), and r-M3 ($|r| = 0.7$, P -value 0.0005) (Fig. 3*B* and Dataset S6). Module r-M1 features mostly (~80%) pSymB and chromosomal genes (Fig. 3*A*) and is enriched ($fdr < 0.0001$) in 13 COG categories including carbohydrate transport (by ABC transporters) and metabolism, transcription, and lipid transport and metabolism (Dataset S7). Module r-M3 includes mostly genes on the chromosome (Fig. 3*A*) and is enriched ($fdr < 0.0001$) in posttranslational modification, turnover, chaperones and transcription (Dataset S7). Although they were not found in the top modules, clusters pan-C1 and pan-C3 were both found in r-M21, which was marginally negatively correlated with shoot biomass ($r = -0.57$; Dataset S6).

Plant-Bacteria Coexpression Reveals Host and Bacterial Modules That Are Correlated with Each Other and with Plant Biomass.

We also studied how plant transcriptomes varied across partner rhizobium strains and identified plant gene modules (Dataset S8) significantly correlated with shoot biomass (more details in the SI Appendix, Supplemental Text). To examine the coexpression of plant and rhizobium genes within the nodule, we correlated modules across both networks (plant and rhizobia) using Module Eigengenes (MEs). Two plant (p) modules (p-M2 and p-M10) were significantly correlated with both rhizobium modules (cutoff: $|r| > 0.7$) and with plant shoot biomass ($|r| > 0.6$) (Fig. 4 and Dataset S9). We investigated these and closely related modules to uncover the relationships among legume and rhizobial biological processes regulating biomass accumulation.

The plant module p-M2 and its close relatives p-M5 and p-M14 are positively correlated with shoot biomass and negatively correlated with r-M1 and r-M3 (Fig. 4 and Dataset S9). The representative Gene Ontology (GO) terms for the p-M2 module are release and recycling of phosphorus, the formation and function of nitrogenase via PPIases and Fe binding, and systemic resistance (genes listed in Dataset S10). Module p-M5 is enriched in GO terms related to protein and messenger RNA (mRNA) turnover, while p-M14 genes are enriched in GO terms for the synthesis of protein and mRNA (Dataset S10); together, these terms suggest transcriptional and translational reprogramming toward the benefit of plant biomass accumulation.

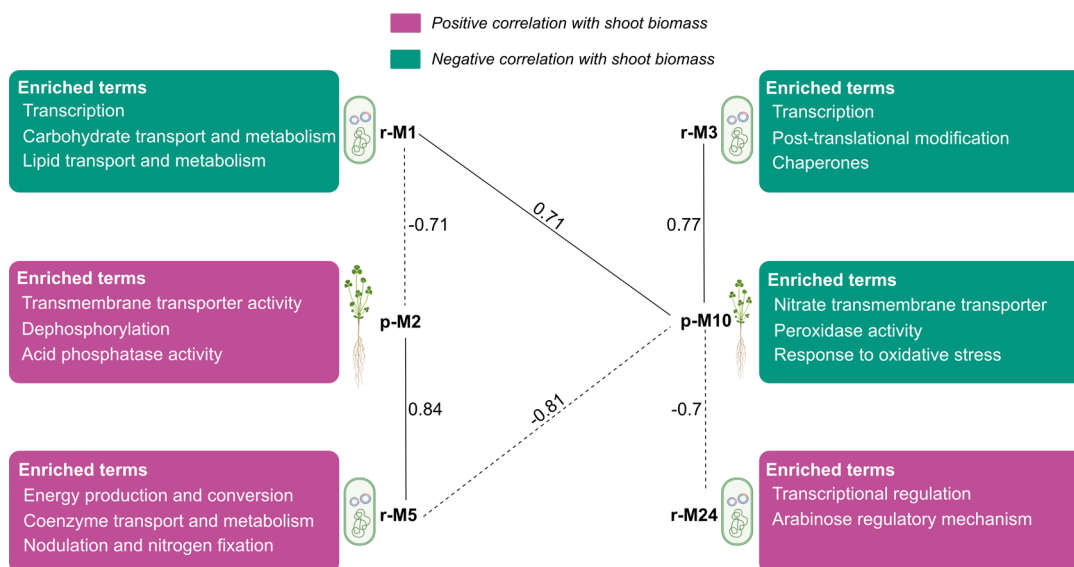


Fig. 4. Relationships among important coexpressed plant and rhizobia modules. Pearson correlations between MEs of plant and rhizobia modules are shown. Enriched COG/GO terms are shown for each module. Negative correlations are represented as dotted lines, positive correlations as solid lines.

The second plant module p-M10, along with close relatives p-M4 and p-M35, are negatively correlated with shoot biomass and are positively correlated with rhizobia modules r-M1 and rM3 while negatively correlated with r-M5 and r-M24 (Fig. 4 and Dataset S9). Representative GO terms for module p-M10 include oxidative stress and nitrate transmembrane transport (Fig. 4 and Dataset S10); module p-M4 is enriched in GO terms for O₂ binding and transport, killing of cells of another organism, and mRNA processing. The p-M4 module is dominated by late nodulin genes, a category of nodule-specific cysteine-rich (NCR) peptides that are frequently expressed in ineffective nodules (Dataset S8) (74), these late nodulin genes are also expressed in module p-M35 that is enriched in GO terms related to the import of proteins into the inner membrane of the mitochondria.

Putting it all together, we find that cluster PAV underlies expression patterns between plant and bacterial gene modules associated with partner quality. Specifically, in higher-quality strains (containing the transcriptional regulators in C2 and pan-C6), the expression of genes in modules r-M5–r-M24–p-M2 (Fig. 4) is coordinated such that genes involved in defense against toxins and oxidative stress (r-M24) and canonical nitrogen-fixation (*nif*fix) genes (r-M5) are coexpressed with plant genes (p-M2) involved in phosphorus turnover (needed to enhance N-fixation and nodule respiration), ferric and ferrous iron binding, and protein peptidyl-prolyl isomerization, all necessary components for the proper folding and function of nitrogenase in effective nodules (75–79). Moreover, in these high-quality strains, the coordinated down-regulation of bacterial genes involved in carbohydrate and lipid transport and metabolism, along with plant genes encoding NCR peptides and oxidative stress responses, are associated with increased biomass accumulation. As expected, opposite expression patterns are observed in lower-quality strains (containing C3 and not C2) that result in less biomass accumulation. These interrelationships – gene cluster PAV, coordinated host-bacterial coexpression, and plant benefits—emphasize the complexity of the genetic and regulatory underpinnings of natural variation in N-fixing symbiosis.

Discussion

Managing legume–rhizobium symbiosis for sustainability and agricultural improvement and understanding symbiosis evolution in natural populations both require a systems-level analysis grounded in standing genetic variation (62). Recent strides in genomics, transcriptomics, and computational biology, along with abundant natural diversity within the *M. truncatula*–*S. meliloti* symbiosis have created opportunities for large-scale analyses of active processes within established nodules (34, 49, 80). These processes drive the trade of resources that ultimately determines the degree to which plants benefit from rhizobial symbiosis (54, 63, 81–83). Leveraging the naturally occurring genetic diversity across 20 *S. meliloti* strains, we employ a systems genetics approach that combines dual-RNAseq transcriptomics of host and symbiont with long-read, reference-quality genomes and phenotypic data to 1) highlight bacterial genes and pathways that contribute to partner quality variation available to natural and artificial selection for symbiotic improvement, 2) discover reveal variation in pSymA gene presence–absence that contributes to plant benefits, and 3) resolve coordinated host-symbiont coexpression in the nodule pantranscriptome associated with plant benefits.

Genetic Architecture and Inheritance of Rhizobium Partner Quality. Our research supports growing evidence that partner quality variation in nature depends on complex genetic architecture, including the modular inheritance of genes on symbiotic plasmids

(84–87), particularly pSymA in *S. meliloti*. Nucleotide variations in the core genes of *S. meliloti* and their response to the environment (G × E) have previously been linked with quantitative natural variation in host benefits, as highlighted in the recent Genome-wide association study (GWAS) study by Batstone et al. (2022) and others (88). Here, we refine our understanding of the genetic architecture of partner quality in *S. meliloti* by adding the presence/absence of pSymA gene clusters, as well as correlated gene expression changes in thousands of genes in both host and symbiont. The association between PAV in functionally relevant gene clusters and partner quality in our dataset, along with incongruent evolutionary histories of clusters compared to the pSymA phylogeny, suggest that horizontal mobility of symbiosis genes, and their modularity in linked gene clusters, are likely major forces that generate variation in partner quality (36, 67, 85). This is particularly noticeable with cluster C2, which was found in high-quality strains in our population and was even found in duplicate in a subset of our strains including (but not limited to) some of the highest-quality strains (204, 199, 141). Whether copy number or simply presence of these putative symbiosis gene clusters is more important for symbiotic function remains to be determined; both are known to be important in *Mesorhizobium* (21). Knocking out C2 in the laboratory reference strain 1021 caused a major reduction in plant biomass, validating the association between C2 and partner quality. Integrating across diverse studies that draw from quantitative and population genetic approaches (33, 89), mapping (5, 21, 51), experimental evolution (4, 90–92), and transcriptomics (9, 49, 80, 93) suggests a model of symbiotic partner quality in which hundreds of diverse genetic underpinnings, and thus diverse modes of transmission and patterns of selection, determine the benefits of symbiosis for hosts in nature.

Our experimental design, in which a single genotype of hosts was inoculated with one of 20 strains in a common garden, allowed us to quantify how genetic variation in *Sinorhizobium* contributes to expression variation in both plant and rhizobium genes. Here, changes in plant gene expression are plastic responses of the plant transcriptome to genetic variation residing in the bacterial genome (i.e., symbiotic extended phenotypes; 4). Variation in *Sinorhizobium* gene expression, on the other hand, could be due to variants within the expressed genes themselves or variants elsewhere in the rhizobium genome and might involve multigene interaction (44, 66), epigenetic modifications (94, 95), or even structural variants/rearrangements that influence expression but not gene content (96). Remarkably, 56 out of 60 “universal” candidate genes (those mapped in at least three experiments) from our previous GWAS using a much larger set of strains (5) were actively expressed in nodules, suggesting that natural variation at the loci themselves, as well as expression of these loci, play a role in partner quality. We found 54 of these genes in modules significantly negatively correlated with shoot biomass (43 genes in r-M1 and 11 genes in r-M10; Dataset S11). Together these approaches provide robust targets for novel gene discovery and spotlight the role of natural variation in amino acid, polyamine, and sugar transporters in optimal symbiosis (*WP_010967350.1*; *SM_RS26550*, *WP_010967351.1*; *SM_RS26555*, *WP_010975345.1*; *SM_RS19440*; Dataset S11). They also highlight the importance of misregulation and miscommunication in poor-quality symbiosis, since a regulator and solute transport were associated with smaller shoot biomass (*WP_010975343.1*; *SM_RS19430*, *WP_010975557.1*; *SM_RS20515*; Dataset S11). Though these genes are novel, these functional hypotheses fit with our current understanding of bacterial metabolism, which is largely based on studies of a single laboratory strain (1021) and its isogenic mutants (29, 66, 97). Additional functional studies would be necessary to resolve the molecular mechanisms by which variation at these

particular loci influence symbiosis, but these are tantalizing candidates given that transport functions typically lack resolution in metabolic models (54–57).

One implication of our findings is that clusters of genes associated with decreased partner quality in our study appear to be horizontally transmitted and maintained at intermediate frequency in natural *S. meliloti* populations. There are several hypotheses that might explain these results. First, selection during other phases of the bacterial lifestyle, beyond the nodules, might increase the frequency of these clusters if they improve bacterial fitness in other environments. For example, Burghardt et al. (2018) found little relationship between the relative fitness of strains in symbiosis versus soil, indicating environment-dependent selection on rhizobia. Second, it is possible that these same clusters are beneficial in other host genotypes; indeed, the partner quality of different *Sinorhizobium* strains is well known to depend on the host genotype (4, 5, 93, 98), implying that the phenotypic effect of alleles in those strains interact with genes in the host [i.e., intergenomic epistasis (98)]. Similarly, the effects of clusters on host and symbiont fitness might depend on epistatic interactions with other genes in the rhizobium genome; indeed, while our initial analysis indicated that cluster C3 was associated with low-quality partners, knocking out this cluster in the reference led to nonsignificant effects on partner quality—suggesting epistasis. Finally, fitness in multiscale symbiosis (including, for example, plasmids within bacteria and bacteria within hosts) (99) might play a role in maintaining seemingly unfit genes (86). There is a growing recognition that horizontally transmissible elements including plasmids, transposons, and other types of MGEs [are known to be abundant on pSymA (100)] have their own fitness interests and agency in natural populations (67). If gene clusters that decrease partner quality actually increase the frequency with which the MGE that carries them is duplicated, transmitted, or permitted entry, they might persist even in the face of plant-mediated selection for high quality rhizobia (67, 99, 101).

Cross-Membrane Transport Systems and Signaling Pathway Heterogeneity Are Major Drivers in Plant Cell and Bacteroid Communication and Mutualistic Coordination. Our dual-RNAseq of both plant and rhizobium genes actively expressed in nodule tissue enables us to identify functional processes in both partners that are strongly correlated with each other and with plant biomass. Modules positively associated with plant biomass (r-M5, r-M24, and p-M2) were mutually enriched in oxidative protection and energy mechanisms for the synthesis of N fixation genes along with the necessary plant processes and cofactors needed for the formation, stabilization, and function of the nitrogenase complex. Examining processes negatively correlated with plant biomass (r-M1, r-M3, and p-M10, p-M4, p-M35) suggests a role for plant NCR signaling genes expressed in low-quality interactions (102, 103) (Dataset S8). NCRs are activated in successive waves during nodule organogenesis and exhibit specific spatial localization of their transcripts from the apical to the proximal nodule zones (104, 105). Diverse NCR molecules have been associated with both infection response mechanisms and positive signaling for mutualism in compatible interactions, but also the terminal differentiation of bacteroids and ongoing maintenance of mutualism (102, 103, 106, 107). Given that we studied the genetic variation in gene expression across rhizobium strains, plant NCR expression is plastically responding (directly or indirectly) to genetic factors in the genomes of low-quality strains (104, 108). The roles of bacterial effectors, NCRs, and their coordinated

bacterial responses are still being worked out using functional genetic approaches (109); nevertheless, it is known that NCRs can result in arrested development of bacteroids in symbiosomes (110, 111) and that strains have different susceptibilities to NCRs depending on both the presence of particular loci [e.g., BacA-like proteins (105)] as well as the genetic variation in unknown genes (110, 111). Importantly, all of our strains improve plant growth over control (Fig. 1A); thus, it is possible that NCR overexpression in our study signals plant response to nonterminally differentiated bacteroids due to incompatibilities with strains, rather than a defense response to rhizobia that would decrease plant fitness. It is also important to note that some NCRs in our study and in others promote high quality symbiosis (15).

It is important to note that there are ~200 annotated *ABC transporters* in *S. meliloti*, significantly more genes than in other similar alpha-proteobacteria (81, 112); nevertheless, our dataset is enriched in these functions compared to expectation based on their frequency in the genome. ABC transporters are heterogenic transmembrane protein systems that consist of the adenosine triphosphate-binding cassette-transporter, whose function is to actively mobilize solutes across the lipid membrane of the cell (113). Our study raises questions about communication dynamics between both partners in symbiosis, particularly involving ABC transporters and membrane permeability systems. Molecule transportation systems are thought to be determinants of mutualistic symbiosis between rhizobia and legumes and other organisms such as arbuscular mycorrhizae (114) and are particularly important for the communication between bacteroid and plant cells in the symbiosome as a means of nutrient mobilization and metabolic integration (97). The positive association of certain bacterial modules (e.g., r-M5 and r-M24) that contain various transporters (i.e. lipid, protein, AA, carbohydrate, etc.) with plant modules (p-M2) for membrane functions and amino acid metabolism (sulfur, sugar, nitrogen, metal ion, etc), but negative association with plant nitrate transmembrane transport (p-M10), suggests microbial membrane transport influences gene regulation in the plant cell and ultimately plant growth.

Conclusions. Systems genetics approaches have the potential to resolve the genetic mechanisms underlying ecologically and economically important variation in complex traits such as partner quality in nitrogen-fixing symbiosis. The contribution of host-strain interactions, G x E interactions, epistasis, and modular gene clusters suggest complex genetic architecture and Horizontal gene transfer underlying partner quality variation, shedding light on the fundamental nature of trait variation and inheritance in microbial populations. Identifying novel loci involved in oxidative stress, respiration, cross-membrane transport, carbon and nitrogen metabolism, and particular plant NCRs for future genetic studies will add resolution to mechanistic models of symbiosis, while providing a model for how host-symbiont coexpression analysis can generate hypotheses about coordinated functions in symbiosis. Our results also raise new evolutionary questions about how functional gene cluster PAV is maintained despite longstanding and abundant evidence for fitness alignment (via plant partner choice and/or sanctions) that increases the fitness of high-quality rhizobia (115–120). Our integrated approach thus provides valuable insights into the genomic and molecular intricacies of symbiotic interactions but also sets the stage for future research directions in understanding the evolution and sustainability of legume–rhizobium partnerships for agricultural improvement.

Materials and Methods

Experiment Design and Plant Measurements. We grew host plants, *M. truncatula* (DZA), in a factorial, single inoculation experiment with 20 *S. meliloti* strains; 10 plants per strain were phenotyped (200 plants, plus controls), and four replicates were used for nodule RNA extraction (80 plants total). Plants and strains were chosen to maximize partner quality variation based on previous observations and were grown in the greenhouse as previously described (5). We took data on leaf number and height early in the experiment (~2 and 4 wk after planting). During harvest (6 wk postplanting), we measured several additional indicators of rhizobium partner quality: shoot and root dry biomass, chlorophyll content using a SPAD 502DL plus Chlorophyll meter (Spectrum technologies, inc. Bridgend, UK) and total number of nodules. For RNA, mature nodules (47 per plant on average) were collected from *Medicago* roots (mean nodule weight = 77 mg) into steel-bead prefilled tubes, immediately flash-frozen in liquid nitrogen, and stored at -80°C . Frozen nodules were homogenized using a prechilled bead beater machine and stored back at -80°C after total tissue disruption. We used the Plant RNeasy kit from QIAGEN[®] (MD, USA), adding an additional bead beater step while tissue was in lysis buffer, to obtain an average of 348 ng/ μL (45.3–1282.1) of pure RNA.

Sequencing, Data Preprocessing, and Reference-Based Transcriptome. RNAseq libraries were prepared with Illumina's TruSeq Stranded mRNAseq Sample Prep kit (Illumina, CA, USA), then pooled, quantitated by qPCR, and sequenced on one S4 lane for 300 cycles from both ends of the fragment on a NovaSeq 6000 sequencing system (Illumina). Fastq files were generated, demultiplexed with the bcl2fastq v2.20 Conversion Software (Illumina), and adaptors were trimmed. *M. truncatula* (A17 r5.0) reference files downloaded from INRAE/CNRS Medicago Bioinformatics Resources (63), and *S. meliloti* (strain 1021 assembly GCF_000006965.1 ASM696v1) reference files were downloaded from NCBI (64, 65). Salmon (121) (v 1.4.0) was used to index the combined *M. truncatula* (mRNA + ncRNA + rRNA files) and *S. meliloti* (cDNA + rRNA files) transcriptomes using the decoy-aware method with the entire *M. truncatula* + *S. meliloti* genomes as the decoy sequence. Then quasi-mapping was performed to map reads to the combined transcriptomes with additional arguments `--seqBias`, `--gcBias`, `--numBootstraps=30`, `--validateMappings`, and `--recoverOrphans` to help improve the accuracy of mappings. Sequence data can be found in the Gene Expression Omnibus under accession number GSE212235. Additional statistical analyses were done in R (122) (v4.1.0) using packages as indicated in *SI Appendix*.

Identification of 1021-Based PAV in *S. meliloti*. The PAV of genes in each strain was determined by establishing counts per million (CPM) cutoffs for each strain. Genes with expression below the cutoff in three out of four replicates are labeled as "absent" in that strain, following a criterion adapted from the `filterByExpr` function in the `edgeR` package (123, 124) in R (122) (v4.1.0). Further filtering based on gene presence-absence patterns resulted in a set of 574 genes (*Dataset S11*) that were present in at least three and a maximum of 17 strains; of those, genes highly correlated with plant shoot biomass were selected based on cutoff ($|r| > 0.5$). PAV of these candidates were further verified with the `BLAST+` (125) search of gene sequences (cutoff: at least 98% identity and query coverage to call presence) against long-read assemblies of 20 strains (BioProject PRJNA1009820). We kept only candidate genes that had a 100% agreement between the predicted presence/absence calls based on the expression pattern and the sequence search results. Additional details can be found in the *SI Appendix*.

Long-Read Assemblies of 20 *S. meliloti* Strains and Pangenome Assembly. DNA extraction was performed using the PacBio Nanobind CBB kit[®] (PacBio, USA) on cultivated bacterial cells obtained from 20 strains of *S. meliloti*, then DNA extracts processed for PacBio Sequel II HiFi sequencing (126). Consensus long-read assemblies for the 20 *S. meliloti* strains were constructed using the `Tracycler`[®] tool (127), using default parameters, setting a minimum of 100 \times reads. To create neighbor-joining trees of the core genomes of the *S. meliloti* chromosome, pSymA, and pSymB, we aligned genomes and created trees using the `SPINE-NUCMER-SNPS` pipeline (128), then edited trees in `FigTree` (v1.4.4) and `ggtree` in R (v4.1.0).

Core, Shell, and Cloud categories across the pangenome were based on their presence across strains. Genes were grouped by their genomic element (chromosome, plasmid pSymB, plasmid pSymA, or eight nondefined extra chromosomal elements). Genes were then classified as "Core" (>99% of strains), "Shell" (15 to 95% of strains), or "Cloud" (<15% of strains) (129). The data were summarized

by counts of gene groups and proportions calculated relative to the total of annotated genes for each element. Genome assemblies for all *S. meliloti* samples, including the 20 strains plus two reference genomes (Strain 1021 assembly GCF_000006965.1 ASM696v1; MABNR56 assembly GCF_037023865.1 ASM3702386v1), along with relevant annotation, were processed using Panaroo v1.5.1 (130) to generate a pangenome network for all protein-coding genes from the 22 *S. meliloti* strains. Full methods can be found in the *SI Appendix*.

Pantranscriptome Gene Expression Quantification and Summary Statistics. Salmon v1.10.0 (121) was used to generate a combined *M. truncatula* + *S. meliloti* pantranscriptome reference database. First, we combined MtrunA17r5.0 mRNA, ncRNA, and rRNA transcript FASTA files [downloaded from INRAE/CNRS Medicago Bioinformatics Resources (63)] with the *S. meliloti* pangenome "combined_DNA_CDS.fasta" transcripts, along with decoys consisting of the full MtrunA17r5.0 genome FASTA and the two NCBI *S. meliloti* reference genomes, ASM696v1 and ASM3702386v1. Salmon was then used to quantify the counts per transcript. Depending on the sample, the percentage of reads aligned to the combined *Medicago* + *S. meliloti* pangenome was 3.1 to 35.9%, of which 20.8 to 95.7% mapped to *Medicago* (*Dataset S2*). While this total alignment percentage appears much lower than the reference-based transcriptome, it does not include the ~7 to 75% rRNA per sample, only protein-coding transcripts. Full methods can be found in the *SI Appendix*.

Genetic Manipulation of 1021-Based Gene Clusters C2 and C3. Gene cluster deletions in *S. meliloti* Rmp110 were generated via an Flippase-Flippase Recognition Target recombination system: Plasmids with flanking FRT sites were constructed using Golden Gate cloning (131), introduced into *E. coli*, and then mobilized into *S. meliloti* by triparental mating. Next we performed a common garden experiment in the greenhouse to assess the effect of modified strains on *M. truncatula*. Plants were inoculated with mock control, wild-type strain 1021, or deletion mutants (1021_ΔC2, 1021_ΔC3), then grown for 8 wk. Aboveground tissue was harvested for biomass measurements, and we used one-way ANOVA and Tukey post hoc tests of variation among inoculation treatment. Full methods including a full list of bacterial strains and plasmids can be found in the *SI Appendix*.

WGCNA. We used the strain means of gene expression (average *logCPM*) to construct the coexpression networks in the WGCNA (132) package in R (122) (v4.1.0). The networks were constructed separately for *M. truncatula* and *S. meliloti* using the *Medicago* reference genome and pangenome of *S. meliloti* along with two reference strains.

Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (133) was used for assigning KEGG Orthology numbers (*Sinorhizobium* and *Medicago*) and mapping of genes to metabolic processes and BRITE categories. The COG annotation of reference *Sinorhizobium* genes was obtained from `eggNOG-mapper` (134). KEGG and COG categories assignments were used to evaluate the enrichment of gene sets using `ClusterProfiler` (135) package in R (122) (v4.1.0). Full methods can be found in the *SI Appendix*.

Data, Materials, and Software Availability. R scripts of the analyses are available on GitHub (<https://github.com/mrizwanriaz/symbiosis>) (136). Long-read assemblies can be accessed at NCBI BioProject PRJNA1009820 (137). Bulk RNAseq raw reads of *M. truncatula* and *S. meliloti* are available on GEO accession GSE212235 (138).

ACKNOWLEDGMENTS. We are grateful for funding from the Carl R. Woese Institute for Genomic Biology, NSF (IOS-1645875, DBI-2022049, IOS-2243821, and IOS-2243818), and Consejo Nacional de Ciencia y Tecnología, Mexico. We want to acknowledge the valuable technical support and guidance of the Roy J. Carver Biotechnology Center, particularly from Alvaro G. Hernandez and Chris L. Wright. We are also grateful for the valuable discussion and feedback from Caroline Oldstone-Jackson, Sierra Bedwell, and Chase Schwarz. We thank two anonymous reviewers for valuable feedback that improved this paper.

Author affiliations: ^aDepartment of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801; ^bDepartment of Microbiology, University of Illinois Urbana-Champaign, Urbana, IL 61801; ^cCarl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL 61801; ^dDepartment of Microbiological Sciences, North Dakota State University, Fargo, ND 58102; ^eDepartment of Biology, McMaster University, Hamilton, ON L8S 4K1, Canada; and ^fUniversity of Illinois Urbana-Champaign, Roy J. Carver Biotechnology Center, Urbana, IL 61801

1. K. Faust, J. Raes, Microbial interactions: From networks to models. *Nat. Rev. Microbiol.* **10**, 538–550 (2012).
2. A. L. Gould *et al.*, Microbiome interactions shape host fitness. *Proc. Natl. Acad. Sci. U.S.A.* **115**, E11951–E11960 (2018).
3. D. Thoms, Y. Liang, C. H. Haney, Maintaining symbiotic homeostasis: How do plants engage with beneficial microorganisms while at the same time restricting pathogens? *Mol. Plant Microbe Interact.* **34**, 462–469 (2021).
4. R. T. Batstone, A. M. O'Brien, T. L. Harrison, M. E. Frederickson, Experimental evolution makes microbes more cooperative with their local host genotype. *Science* **370**, 476–478 (2020).
5. R. T. Batstone *et al.*, Genome-wide association studies across environmental and genetic contexts reveal complex genetic architecture of symbiotic extended phenotypes. *mBio* **13**, e0182322 (2022).
6. N. C. Johnson, J. H. Graham, The continuum concept remains a useful framework for studying mycorrhizal functioning. *Plant Soil* **363**, 411–419 (2012).
7. N. C. Johnson, J. H. Graham, F. A. Smith, Functioning of mycorrhizal associations along the mutualism–parasitism continuum. *New Phytol.* **135**, 575–586 (1997).
8. E. T. Kiers *et al.*, Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science* **333**, 880–882 (2011).
9. K. D. Heath, P. V. Burke, J. R. Stinchcombe, Coevolutionary genetic variation in the legume–rhizobium transcriptome. *Mol. Ecol.* **21**, 4735–4747 (2012).
10. S. A. West, E. T. Kiers, I. Pen, R. F. Denison, Sanctions and mutualism stability: When should less beneficial mutualists be tolerated? *J. Evol. Biol.* **15**, 830–837 (2002).
11. J. K. Ardley *et al.*, Nodule morphology, symbiotic specificity and association with unusual rhizobia are distinguishing features of the genus *Listia* within the Southern African crotalariaid clade *Listia* s.l.. *Ann. Bot.* **112**, 1–15 (2013).
12. R. Bonneau, Learning biological networks: From modules to dynamics. *Nat. Chem. Biol.* **4**, 658–664 (2008).
13. L. Salvaudon, T. Giraud, J. A. Shykoff, Genetic diversity in natural populations: A fundamental component of plant–microbe interactions. *Curr. Opin. Plant Biol.* **11**, 135–143 (2008).
14. K. D. Heath, J. R. Stinchcombe, Explaining mutualism variation: A new evolutionary paradox? *Evolution* **68**, 309–317 (2014).
15. D. Wang, S. Yang, F. Tang, H. Zhu, Symbiosis specificity in the legume: Rhizobial mutualism. *Cell Microbiol.* **14**, 334–342 (2012).
16. N. H. Barton, P. D. Keightley, Understanding quantitative genetic variation. *Nat. Rev. Genet.* **3**, 11–21 (2002).
17. R. Della Coletta, Y. Qiu, S. Ou, M. B. Hufford, C. N. Hirsch, How the pan-genome is changing crop genomics and improvement. *Genome Biol.* **22**, 3 (2021).
18. M. Bamba *et al.*, Massive rhizobial genomic variation associated with partner quality in *Lotus-mesorhizobium* symbiosis. *FEMS Microbiol. Ecol.* **96**, fiae202 (2020).
19. C. R. Cooney, J. E. Mank, A. E. Wright, Constraint and divergence in the evolution of male and female recombination rates in fishes. *Evolution* **75**, 2857–2866 (2021).
19. H. Wang *et al.*, QTL analysis of salt tolerance in Sorghum bicolor during whole-plant growth stages. *Plant Breeding* **139**, 455–465 (2020).
21. H. Kehlet-Delgado *et al.*, The evolutionary genomics of adaptation to stress in wild rhizobium bacteria. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2311127121 (2024).
22. M. Ashok Saini *et al.*, *An Introduction to Microbial Genomic Islands for Evolutionary Adaptation and Pathogenicity* (Springer, Singapore, 2023), pp. 1–15.
23. M. Sugawara *et al.*, Comparative genomics of the core and accessory genomes of 48 Sinorhizobium strains comprising five genospecies. *Genome Biol* **14**, R17 (2013).
24. G. Orazi, A. J. Collins, R. J. Whitaker, Prediction of prophages and their host ranges in pathogenic and commensal *Neisseria* species. *mSystems* **7**, e0008322 (2022).
25. M. I. A. Cavassim *et al.*, Symbiosis genes show a unique pattern of introgression and selection within a *Rhizobium leguminosarum* species complex. *Microb. Genom.* **6**, e000351 (2020).
26. F. Rosconi *et al.*, A bacterial pan-genome makes gene essentiality strain-dependent and evolvable. *Nat. Microbiol.* **7**, 1580–1592 (2022).
27. M. Nelson, J. Guhlin, B. Epstein, P. Tiffin, M. J. Sadovsky, The complete replicons of 16 *Ensifer meliloti* strains offer insights into intra- and inter-replicon gene transfer, transposon-associated loci, and repeat elements. *Microb. Genom.* **4** (2018).
28. P. W. Harrison, R. P. Lower, N. K. Kim, J. P. Young, Introducing the bacterial “chromid”: Not a chromosome, not a plasmid. *Trends Microbiol.* **18**, 141–148 (2010).
29. G. C. diCenzo, A. M. MacLean, B. Milunovic, G. B. Golding, T. M. Finan, Examination of prokaryotic multipartite genome evolution through experimental genome reduction. *PLoS Genet.* **10**, e1004742 (2014).
30. B. Epstein, P. Tiffin, Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. *Proc. Biol. Sci.* **288**, 20201804 (2021).
31. S. J. Bush *et al.*, Presence-absence variation in *A. thaliana* is primarily associated with genomic signatures consistent with relaxed selective constraints. *Mol. Biol. Evol.* **31**, 59–69 (2014).
32. L. M. Zhang *et al.*, Genome-wide patterns of large-size presence/absence variants in sorghum. *J. Integr. Plant Biol.* **56**, 24–37 (2014).
33. A. B. Riley, M. A. Grillo, B. Epstein, P. Tiffin, K. D. Heath, Discordant population structure among rhizobium divided genomes and their legume hosts. *Mol. Ecol.* **32**, 2646–2659 (2023).
34. B. Roux *et al.*, An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant J.* **77**, 817–837 (2014).
35. R. Lederemann, C. C. M. Schulte, P. S. Poole, How rhizobia adapt to the nodule environment. *J. Bacteriol.* **203**, e0053920 (2021).
36. S. Liu, J. Jiao, C. F. Tian, Adaptive evolution of rhizobial symbiosis beyond horizontal gene transfer: From genome innovation to regulation reconstruction. *Genes* **14**, 274 (2023).
37. G. E. Oldroyd, J. D. Murray, P. S. Poole, J. A. Downie, The rules of engagement in the legume–rhizobial symbiosis. *Annu. Rev. Genet.* **45**, 119–144 (2011).
38. T. Coba de la Pena, E. Fedorova, J. J. Pueyo, M. M. Lucas, The symbiosome: Legume and rhizobia co-evolution toward a nitrogen-fixing organelle? *Front. Plant Sci.* **8**, 2229 (2017).
39. W. F. Chen, E. T. Wang, Z. J. Ji, J. J. Zhang, Recent development and new insight of diversification and symbiosis specificity of legume rhizobia: Mechanism and application. *J. Appl. Microbiol.* **131**, 553–563 (2021).
40. J. Unay, X. Perret, A minimal genetic passkey to unlock many legume doors to root nodulation by rhizobia. *Genes* **11**, 521 (2020).
41. A. A. Alemneh, Y. Zhou, M. H. Ryder, M. D. Denton, Mechanisms in plant growth-promoting rhizobacteria that enhance legume–rhizobial symbioses. *J. Appl. Microbiol.* **129**, 1133–1156 (2020).
42. Q. Wang, J. Liu, H. Zhu, Genetic and molecular mechanisms underlying symbiotic specificity in legume–rhizobium interactions. *Front. Plant Sci.* **9**, 313 (2018).
43. Y. Ma, C. Xiao, J. Liu, G. Ren, Nutrient-dependent regulation of symbiotic nitrogen fixation in legumes. *Hortic Res* **12**, uhac321 (2025).
44. R. K. Goyal, A. K. Mattoo, M. A. Schmidt, Rhizobial–host interactions and symbiotic nitrogen fixation in legume crops toward agriculture sustainability. *Front. Microbiol.* **12**, 669404 (2021).
45. J. Zhao *et al.*, Global systematic review with meta-analysis reveals yield advantage of legume-based rotations and its drivers. *Nat. Commun.* **13**, 4926 (2022).
46. S. R. Long, W. J. Buikema, F. M. Ausubel, Cloning of *Rhizobium-meliloti* nodulation genes by direct complementation of Nod- mutants. *Nature* **298**, 485–488 (1982).
47. J. M. Ane *et al.*, Medicago truncatula DMI1 required for bacterial and fungal symbioses in legumes. *Science* **303**, 1364–1367 (2004).
48. N. Zhang *et al.*, Metabolomic profiling reveals suppression of oxylipin biosynthesis during the early stages of legume–rhizobia symbiosis. *FEBS Lett.* **586**, 3150–3158 (2012).
49. B. S. Watson *et al.*, Integrated metabolomics and transcriptomics reveal enhanced specialized metabolism in Medicago truncatula root border cells. *Plant Physiol.* **167**, 1699–1716 (2015).
50. L. T. Burghardt, B. Epstein, P. Tiffin, Legacy of prior host and soil selection on rhizobial fitness in planta. *Evolution* **73**, 2013–2023 (2019).
51. B. Epstein *et al.*, Combining GWAS and population genomic analyses to characterize coevolution in a legume–rhizobia symbiosis. *Mol. Ecol.* **32**, 3798–3811 (2023).
52. S. Roy *et al.*, Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. *Plant Cell* **32**, 15–41 (2020).
53. B. A. Geddes *et al.*, Minimal gene set from Sinorhizobium (Ensifer) meliloti pSymA required for efficient symbiosis with Medicago. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2018015118 (2021).
54. G. C. diCenzo, M. Tesi, T. Pfau, A. Mengoni, M. Fondi, Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* **11**, 2574 (2020).
55. G. C. diCenzo *et al.*, Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in Sinorhizobium meliloti. *Nat. Commun.* **7**, 12219 (2016).
56. T. Pfau *et al.*, The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. *Sci. Rep.* **8**, 12504 (2018).
57. C. C. M. Schulte *et al.*, Metabolic control of nitrogen fixation in rhizobium–legume symbioses. *Sci. Adv.* **7**, eab2433 (2021).
58. N. D. Young *et al.*, The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature* **480**, 520–524 (2011).
59. F. A. Feltus, Systems genetics: A paradigm to improve discovery of candidate genes and mechanisms underlying complex traits. *Plant Sci.* **223**, 45–48 (2014).
60. M. Civelek, A. J. Lusic, Systems genetics approaches to understand complex traits. *Nat. Rev. Genet.* **15**, 34–48 (2014).
61. D. Kliebenstein, Quantitative genomics: Analyzing intraspecific variation using global gene expression polymorphisms or eQTLs. *Annu. Rev. Plant Biol.* **60**, 93–114 (2009).
62. J. K. Daniel, Systems biology uncovers the foundation of natural genetic diversity. *Plant Physiol.* **152**, 480–486 (2010).
63. Y. Pexric *et al.*, Whole-genome landscape of Medicago truncatula symbiotic genes. *Nat. Plants* **4**, 1017–1025 (2018).
64. E. W. Sayers *et al.*, Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **50**, D20–D26 (2022).
65. F. Galibert *et al.*, The composite genome of the legume symbiont Sinorhizobium meliloti. *Science* **293**, 668–672 (2001).
66. G. C. diCenzo *et al.*, Multidisciplinary approaches for studying rhizobium–legume symbioses. *Can. J. Microbiol.* **65**, 1–33 (2019).
67. G. E. Wardell, M. F. Hynes, P. J. Young, E. Harrison, Why are rhizobial symbiosis genes mobile? *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200471 (2022).
68. B. Milunovic, G. C. diCenzo, R. A. Morton, T. M. Finan, Cell growth inhibition upon deletion of four toxin-antitoxin loci from the megaplasmids of Sinorhizobium meliloti. *J. Bacteriol.* **196**, 811–824 (2014).
69. X. Chen *et al.*, The Mesorhizobium huakuii transcriptional regulator *AbiEi* plays a critical role in nodulation and is important for bacterial stress response. *BMC Microbiol.* **21**, 245 (2021).
70. J. Lipuma *et al.*, A vapBC-type toxin-antitoxin module of Sinorhizobium meliloti influences symbiotic efficiency and nodule senescence of Medicago sativa. *Environ. Microbiol.* **16**, 3714–3729 (2014).
71. S. T. N. Ratu *et al.*, Rhizobia use a pathogenic-like effector to hijack leguminous nodulation signalling. *Sci. Rep.* **11**, 2034 (2021).
72. F. Ampe, E. Kiss, F. Sabourdy, J. Batut, Transcriptome analysis of Sinorhizobium meliloti during symbiosis. *Genome Biol.* **4**, R15 (2003).
73. L. Xiaofang, L. Zhangqun, What determines symbiotic nitrogen fixation efficiency in rhizobium: Recent insights into Rhizobium leguminosarum. *Arch. Microbiol.* **205**, 300 (2023).
74. V. Lullien, D. G. Barker, P. de Lajudie, T. Huguet, Plant gene expression in effective and ineffective root nodules of alfalfa (Medicago sativa). *Plant Mol. Biol.* **9**, 469–478 (1987).
75. A. Bargaz, C. Ghoulam, J. J. Drevon, Specific expression and activity of acid phosphatases in common bean nodules. *Plant Signal. Behav.* **8**, e25022 (2013).
76. N. Gavini, S. Tungtur, L. Pulakat, Peptidyl-prolyl cis/trans isomerase-independent functional NifH mutant of Azotobacter vinelandii. *J. Bacteriol.* **188**, 6020–6025 (2006).
77. Y. Li *et al.*, Metal nutrition and transport in the process of symbiotic nitrogen fixation. *Plant Commun.* **5**, 100829 (2024).
78. S. A.-N. Thahir, Phosphate role in the rhizobium–legume symbiosis—A review. *Int. J. Sci. Basic Appl. Res.* **36**, 178–191 (2017).
79. Y. Yu *et al.*, Induced systemic resistance for improving plant immunity by beneficial microbes. *Plants* **11**, 386 (2022).
80. L. T. Burghardt *et al.*, Transcriptomic basis of genome by genome variation in a legume–rhizobia mutualism. *Mol. Ecol.* **26**, 6122–6135 (2017).
81. M. J. Barnett *et al.*, Nucleotide sequence and predicted functions of the entire Sinorhizobium meliloti pSymA megaplasmid. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9883–9888 (2001).
82. D. Capela *et al.*, Analysis of the chromosome sequence of the legume symbiont Sinorhizobium meliloti strain 1021. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 9877–9882 (2001).
83. M. J. Barnett, C. J. Toman, R. F. Fisher, S. R. Long, A dual-genome Symbiosis Chip for coordinate study of signal exchange and development in a prokaryote–host interaction. *Proc. Natl. Acad. Sci. U.S.A.* **101**, 16636–16641 (2004).

84. A. M. O'Brien, C. N. Jack, M. L. Friesen, M. E. Frederickson, Whose trait is it anyways? Coevolution of joint phenotypes and genetic architecture in mutualisms. *Proc. Biol. Sci.* **288**, 20202483 (2021).
85. A. J. Weisberg *et al.*, Pangenome evolution reconciles robustness and instability of rhizobial symbiosis. *mBio* **13**, e0007422 (2022).
86. R. T. Batstone, Genomes within genomes: Nested symbiosis and its implications for plant evolution. *New Phytol.* **234**, 28–34 (2022).
87. S. S. Porter, J. Faber-Hammond, A. P. Montoya, M. L. Friesen, C. Sackos, Dynamic genomic architecture of mutualistic cooperation in a wild population of Mesorhizobium. *ISME J.* **13**, 301–315 (2019).
88. B. Epstein *et al.*, Genome-wide association analyses in the model Rhizobium *Ensifer meliloti*. *mSphere* **3**, 00386–18 (2018).
89. D. J. Weese, K. D. Heath, B. T. Dentinger, J. A. Lau, Long-term nitrogen addition causes the evolution of less-cooperative mutualists. *Evolution* **69**, 631–642 (2015).
90. M. Tang, O. Bouchez, S. Cruveiller, C. Masson-Boivin, D. Capela, Modulation of quorum sensing as an adaptation to nodule cell infection during experimental evolution of legume symbionts. *mBio* **11**, e03129–19 (2020).
91. K. W. Quides *et al.*, Experimental evolution can enhance benefits of rhizobia to novel legume hosts. *Proc. Biol. Sci.* **288**, 20210812 (2021).
92. Z. J. Ji *et al.*, Physiological and symbiotic variation of a long-term evolved *Rhizobium* strain under alkaline condition. *Syst. Appl. Microbiol.* **43**, 126125 (2020).
93. C. Fagorzi *et al.*, Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-Rhizobium symbiosis. *mSystems* **6**, 00974–20 (2021), 10.1128/mSystems.00974–00920.
94. I. Passeri, F. Vaccaro, A. Mengoni, C. Fagorzi, Moving toward the inclusion of epigenomics in bacterial genome evolution: Perspectives and challenges. *Int. J. Mol. Sci.* **25**, 4425 (2024).
95. M. A. Sanchez-Romero, J. Casadesus, The bacterial epigenome. *Nat. Rev. Microbiol.* **18**, 7–20 (2020).
96. V. Periwal, V. Scaria, Insights into structural variations and genome rearrangements in prokaryotic genomes. *Bioinformatics* **31**, 1–9 (2015).
97. M. Udvardi, P. S. Poole, Transport and metabolism in legume-rhizobia symbioses. *Annu. Rev. Plant Biol.* **64**, 781–805 (2013).
98. K. D. Heath, Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution* **64**, 1446–1458 (2010).
99. K. D. Heath, R. T. Batstone, M. Ceron Romero, J. G. McMullen II, MGEs as the MVPs of partner quality variation in legume-rhizobium symbiosis. *mBio* **13**, e0088822 (2022).
100. E. Giuntini *et al.*, Large-scale genetic variation of the symbiosis-required megaplasmid pSymA revealed by comparative genomic analysis of Sinorhizobium meliloti natural strains. *BMC Genomics* **6**, 158 (2005).
101. J. P. J. Hall, E. Harrison, D. A. Baltrus, Introduction: The secret lives of microbial mobile genetic elements. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **377**, 20200460 (2022).
102. H. Tiriz *et al.*, Antimicrobial nodule-specific cysteine-rich peptides induce membrane depolarization-associated changes in the transcriptome of Sinorhizobium meliloti. *Appl. Environ. Microbiol.* **79**, 6737–6746 (2013).
103. B. Lagunas *et al.*, Rhizobial nitrogen fixation efficiency shapes endosphere bacterial communities and Medicago truncatula host growth. *Microbiome* **11**, 146 (2023).
104. I. Guefrachi *et al.*, Extreme specificity of NCR gene expression in Medicago truncatula. *BMC Genomics* **15**, 712 (2014).
105. Q. Nicoud *et al.*, Sinorhizobium meliloti functions required for resistance to antimicrobial NCR peptides and bacteroid differentiation. *mBio* **12**, e0089521 (2021).
106. B. Horvath *et al.*, Loss of the nodule-specific cysteine rich peptide, NCR169, abolishes symbiotic nitrogen fixation in the Medicago truncatula dnf7 mutant. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15232–15237 (2015).
107. B. Horvath *et al.*, The Medicago truncatula nodule-specific cysteine-rich peptides, NCR343 and NCR-new35 are required for the maintenance of rhizobia in nitrogen-fixing nodules. *New Phytol.* **239**, 1974–1988 (2023).
108. S. Zhang *et al.*, Widely conserved AHL transcription factors are essential for NCR gene expression and nodule development in Medicago. *Nat. Plants* **9**, 280–288 (2023).
109. S. Sankari *et al.*, A haem-sequestering plant peptide promotes iron uptake in symbiotic bacteria. *Nat. Microbiol.* **7**, 1453–1465 (2022).
110. Q. Wang *et al.*, Nodule-specific cysteine-rich peptides negatively regulate nitrogen-fixing symbiosis in a strain-specific manner in Medicago truncatula. *Mol. Plant Microbe Interact.* **31**, 240–248 (2018).
111. S. Yang *et al.*, Microsymbiont discrimination mediated by a host-secreted peptide in Medicago truncatula. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 6848–6853 (2017).
112. T. H. Mauchline *et al.*, Mapping the Sinorhizobium meliloti 1021 solute-binding protein-dependent transcriptome. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17933–17938 (2006).
113. W. Stephan, Structure and mechanism of ABC transporters. *F1000Prime Rep.* **7**, P7:14 (2015).
114. J. Banasiak, T. Jamruszka, J. D. Murray, M. Jasinski, A roadmap of plant membrane transporters in arbuscular mycorrhizal and legume-rhizobium symbioses. *Plant Physiol.* **187**, 2071–2091 (2021).
115. R. F. Denison, Legume sanctions and the evolution of symbiotic cooperation by rhizobia. *Am. Nat.* **156**, 567–576 (2000).
116. E. T. Kiers, R. A. Rousseau, S. A. West, R. F. Denison, Host sanctions and the legume-rhizobium mutualism. *Nature* **425**, 78–81 (2003).
117. K. D. Heath, P. Tiffin, Stabilizing mechanisms in a legume-rhizobium mutualism. *Evolution* **63**, 652–662 (2009).
118. V. J. Puhua *et al.*, Fitness variation among host species and the paradox of ineffective rhizobia. *J. Evol. Biol.* **31**, 599–610 (2018).
119. A. Westhoek *et al.*, Conditional sanctioning in a legume-Rhizobium mutualism. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2025760118 (2021).
120. R. T. Batstone, L. T. Burghardt, K. D. Heath, Phenotypic and genomic signatures of interspecies cooperation and conflict in naturally occurring isolates of a model plant symbiont. *Proc. Biol. Sci.* **289**, 20220477 (2022).
121. R. Patro, G. Duggal, M. I. Love, R. A. Irizarry, C. Kingsford, Salmon provides fast and bias-aware quantification of transcript expression. *Nat. Methods* **14**, 417–419 (2017).
122. R Core Team, R: A Language and Environment for Statistical Computing (Version 4.1, R Foundation for Statistical Computing, Vienna, Austria, 2024). <https://www.R-project.org/>. Accessed 15 February 2021.
123. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
124. C. W. Law *et al.*, RNA-seq analysis is easy as 1–2–3 with limma, Glimma and edgeR. *F1000Res* **5**, 1408 (2016).
125. C. Camacho *et al.*, BLAST+: Architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
126. A. Rhoads, K. F. Au, PacBio sequencing and its applications. *Genomics Proteomics Bioinformatics* **13**, 278–289 (2015).
127. R. R. Wick *et al.*, Tricycler: Consensus long-read assemblies for bacterial genomes. *Genome Biol.* **22**, 266 (2021).
128. A. J. Collins, R. J. Whitaker, CRISPR comparison toolkit: Rapid identification, visualization, and analysis of CRISPR array diversity. *CRISPR J.* **6**, 386–400 (2023).
129. P. G. Livingstone, R. M. Morphew, D. E. Whitworth, Genome sequencing and pan-genome analysis of 23 *Coralloccoccus* spp. strains reveal unexpected diversity, with particular plasticity of predatory gene sets. *Front. Microbiol.* **9**, 3187 (2018).
130. G. Tonkin-Hill *et al.*, Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol.* **21**, 180 (2020).
131. B. A. Geddes *et al.*, BEVA2.0: Modular assembly of golden gate-compatible vectors with expanded utility for genetic engineering. *Can. J. Microbiol.* **71**, 1–13 (2025).
132. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
133. Y. Moriya, M. Itoh, S. Okuda, A. C. Yoshizawa, M. Kanehisa, KAAS: An automatic genome annotation and pathway reconstruction server. *Nucleic Acids Res.* **35**, W182–W185 (2007).
134. C. P. Cantalapiedra, A. Hernandez-Plaza, I. Letunic, P. Bork, J. Huerta-Cepas, eggNOG-mapper v2: Functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol. Biol. Evol.* **38**, 5825–5829 (2021).
135. T. Wu *et al.*, Clusterprofiler 4.0: A universal enrichment tool for interpreting omics data. *Innov* **2**, 100141 (2021).
136. M. R. Riaz, symbiosis. GitHub. <https://github.com/mrizwanriaz/symbiosis>. Deposited 15 May 2024.
137. M. R. Riaz, Sinorhizobium meliloti natural isolates genome assemblies. BioProject. <https://www.ncbi.nlm.nih.gov/bioproject?term=PRJNA1009820>. Deposited 28 August 2023.
138. M. R. Riaz, Systems genetics of partner quality: Quantitative variation in gene expression and qualitative variation in plasmid-borne gene clusters underlies symbiosis variation. *Gene Expression Omnibus*. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE212235>. Deposited 29 August 2022.