SUPPLEMENTAL MATERIALS

SUPP. METHODS

Study system. *Medicago truncatula* Gaertn. is an annual legume native to the Mediterranean region in Southern Europe and Northern Africa, and it also occurs in Asia. It lives in a symbiotic association with nitrogen-fixing bacteria in the genus *Ensifer* (formerly *Sinorhizobium*) and has been used as a model-organism to study symbiotic interactions. We used two lines of *Medicago truncatula* in this experiment: A17 (Australia) and DZA 315.16 (Algeria; hereafter abbreviated as DZA). These lines and have been leveraged in studies establishing the role of variable plant-encoded nodule-specific cysteine rich peptides (NCRs, specifically those encoded by *nfs*1 and *nfs*2) in governing the level of nitrogen fixed (Fix+, Fix–, or intermediate) in particular host and strain combinations (1, 2, 3, 4); these lines are included in the *Medicago* HapMap panel of re-sequenced GWAS lines and RIL parents (http://www.medicagohapmap.org). We used 191 strains of the rhizobium *E. meliloti* in this study, which were isolated from the nodules of *M. truncatula* plants grown in soils from 24 sites in Corsica, France and Spain, as part of a larger effort to understand coevolution in this legume-rhizobium mutualism (5, 6).

Experimental design. Seeds were scarified with either a razor blade or sandpaper, then sterilized by first rinsing with 95% EtOH followed by soaking in commercial bleach (6% hypochlorite) for 7 min, rinsed thoroughly with water, and imbibed in ddH_2O overnight at 4°C in the dark. Seeds were then transplanted into sterilized 107 ml SC7 Cone-tainers (Stuewe & Sons Inc., Tangent, OR), each containing an autoclave-sterilized mixture of Turface MVP calcined clay (Profile Products LLC, Buffalo Grove, IL) and the UIUC greenhouse's root wash mix (1:1:1 soil–calcined clay–torpedo sand) in equal volumes, for a final mixture of 1:4:1 soil–calcined clay–torpedo sand. Pots were randomized into racks in the greenhouse, with 10 pots per rack, and racks were evenly arranged across three benches.

E. meliloti cultures were grown in liquid tryptone-yeast (TY) medium (7) for 18-20 hrs at 30°C. Before inoculation, the cell density of each culture was measured with NanoDrop 2000c spectrophotometer (Thermo Fisher Scientific) and adjusted to ~ 10^6 cells/ml (OD_{600} = 0.1) by diluting the cultures with sterile TY medium, when necessary. Each plant was inoculated with 500 ml of liquid culture 10-12 days after seeds were planted, and the soil surface received a thin (~0.5 cm) layer of sterile sand after inoculation to minimize cross-contamination. Plants were misted (1/2" M NPT upright misting nozzle, Senninger Irrigation Inc., Clermont, FL) four times per day for 45 min at a time for the first two weeks after transplant, and 30 min at a time thereafter. Plants were given supplemental lighting up to to 14 hr day length and were not fertilized throughout the experiments.

Data collection. For each experiment, we measured plant height, leaf number, and chlorophyll content at four weeks after planting. Chlorophyll was measured using a SPAD 502 Plus (Spectrum Technologies, Inc. Aurora, IL); we recorded the mean of

three measurements on the most recently emerged leaf. At the time of harvest, we counted total root nodules and collected 10 nodules from each plant to estimate per-nodule fresh weight. Shoots and roots were dried and weighed to determine the dry biomass of each plant.

Rhizobium genomic data. Full details are available in Riley et al. (6). Briefly, strains were grown as described above. Rhizobium genomic DNA was extracted and sequenced, followed by quality control and variant calling as described in Riley et al. (6). Variant data were hard-filtered using vcftools (v0.1.17, 8) to include variants with quality scores above 20. Multi-nucleotide polymorphisms were retained as single variants and sites that lacked genotype calls for more than 20% of the strains were removed. We included in the analyses only variants that had a minor allele frequency (MAF) of $\leq$ 5%. We found 491,227 variants represented in our genomes. After filtering for quality and frequency, 36,526 variants remained.

Phenotypic analyses. Within the R environment (9), we implemented linear mixed models (LMMs) using the R package lme4 (v1.1-27.1, 10) to test for G $\times$ E between 191 strains of rhizobia (G) and the two greenhouse experiments (E) for each plant line separately. We additionally partitioned G $\times$ E interactions into variance versus crossing effects using Cockerham's method (11, 12, 13). We square root-transformed all phenotypic variables to improve the normality of the data. For all phenotypic traits (shoot biomass, leaf number, plant height, chlorophyll content), the model included experiment and strain, and their interaction as fixed effects. Rack was included as an additional random effect. From these models, we calculated estimate marginal means for each strain using the emmeans package (v1.4.1, 14).

Our experimental design, which featured two experiments nested within each host genotype, allows robust statistical estimation of G × E, but is underpowered to test for G × G interactions for partner quality phenotypes using ANOVA. Nevertheless a separate study (K. Heath, unpublished data) featuring a subset (N = 20) of the strains from the current study provides strong statistical support for G × G interactions for partner quality in this system (e.g., plant x strain G × G; χ^2 = 83.3; p < 0.0001 for plant aboveground biomass; **Supp. Fig. S11**), similar to several other studies in this system (e.g., 15, 16, 17, 18, 19). We additionally calculated the broad-sense heritability ($H^2 = \frac{V_G}{V_P}$) for rhizobium strains within each experiment using LMMs in which response variables were square root-transformed, and both strain and rack were included as random effects. The significance of V_G was assessed by comparing the full model to one in which the strain term was excluded, and a log-likelihood ratio test was performed between the two models to assess whether the proportion of variation explained (PVE) by the strain term was significant. Given the complex multivariate nature of our data at both the phenotypic and genomic levels, we focus the main text on shoot biomass, our core metric of partner quality; results for all other traits (i.e., leaf chlorophyll A, plant height, leaf number) can be found in the Supplementary Materials.

Genome-wide association studies. We performed multiple genome-wide association studies (GWAS) to identify rhizobium genomic variants associated with symbiotic partner quality. We conducted association tests for partner quality traits measured on plant hosts using a LMM approach to GWAS as implemented in the program GEMMA (v0.98.1, 20). Mapping analyses were performed separately for each of the four experiments, on standardized emmeans that corrected for the effects of rack (see above). GWAS relies on variants that are statistically distinguishable from one another (i.e., are not closely linked), and so we first identified genetic variants in strong linkage disequilibrium by splitting variants from all three rhizobia genomic elements into linkage groups based on the LD threshold $r_2 \geq 0.95$ using a customized script available at

the Dryad repository associated with Epstein et al. (21) (https://datadryad.org/stash/dataset/doi:10.5061/dryad.tn6652t), and picked a representative variant from each linkage group with the highest minor allele frequency and least missing data when ties were present (21). We ended up with 6,512 unlinked variants, 600 of which were on the chromosome, 1,797 were on pSymA, and 4,115 were on pSymB. We used selected variants to compute standardized kinship (k) matrices in GEMMA using the - gk 2 option, one for each genomic element, and performed the GWAS mapping using the lmm -4 option. In order to ensure our methods of calculating the k-matrix did not influence the results, we also conducted additional association tests for shoot biomass using k-matrices constructed for: 1) the entire genome (all three replicons together), and 2) for all variants, rather than those filtered by LD. However, our results were comparable regardless of the method used for calculating the k-matrix (**Supp. Fig. S12**), and thus, we only present results based on the initial method.

We assigned significance to particular variants using a permutation test which randomizes genotypes with respect to phenotypes, runs the resulting LMM in GEMMA 1000 times, and then tags loci from the non-randomized run that fell above the 95% false discovery rate cut off (21, 22). Based on these significance tests, we first summarized variants to the gene-level, in which we identified the genes closest to or encompassing our significant variants using the intersect option in bedtools (v2.29.2, 23), and excluded any intergenic regions.

Finally, we were interested in the genetic basis of environmental-dependency (i.e., "G $\times$ E" genes), where the allelic effects varied among experiments within a particular host genotype due to either conditional neutrality (i.e., significant effects on the phenotype in one environment but not the other) or antagonistic allelic effects (i.e., significant effects in both environments, but in the opposite direction). Because mapping experiments suffer from false negatives (24), the lack of a significant association in one experiment does not rigorously identify patterns of conditional neutrality at the individual locus or at the global (whole genome) level (24, 25, 26). Thus we took multiple approaches to inspecting G $\times$ E in our association analyses. First, we used cross-environment correlations between the estimated effects from each experiment (computed independently, see above) to visually assess the degree to which allelic effects were consistent across experiments (within a host genotype) and identify antagonistic allelic effects (loci with significant, but opposite, effects in the two experiments). Next, to test the global null hypothesis that the direction and magnitude of the estimated rhizobium allelic effects were consistent across the two experiments for a particular host (i.e., that allelic effects estimated in the two environments fell along the 1:1 line), we used a permutation test wherein we first resampled the estimated effects from the first experiment (i.e., DZA experiment I) with experimental error (standard deviation) and calculated the slope of their correlation, repeated this for 1000 permutations, then computed the probability of our observed slope given this simulated null distribution. Beyond this global test, to rigorously identify individual loci that contribute to the environmental response, we calculated plasticity for each strain (27, 28); plasticity was calculated as the natural log of the response ratio of the phenotype across experiments (e.g., $log(\frac{\text{shoot biomass exp. 3 (DZA)}}{\text{shoot bioamss exp. 1 (DZA)}})$; Lau et al. 29, Heath et al. 30) and mapped this trait separately for each of the two host genotypes.

Candidate gene functional analyses. To explore the biological interpretation of our various gene lists (e.g., A17-only, Experiment 1, Universal; see results), we used DAVID Bioinformatics Resources (v.6.8, 31) to test for significantly overrepresented UNIPROT keywords, GO terms, and KEGG pathways, as described in Sherman et al. (32). We report the results of enrichment analyses as "marginal" when the pre-FDR p-value

associated with the EASE modified Fisher exact test was < 0.05 and "significant" when FDR-corrected p-value was < 0.05. We explored key pathways and genes implicated in particular gene sets using BioCyc (33, 34).

SUPP. RESULTS

We interrogated the gene sets from two key studies that have associated natural variation in *E. meliloti* genomes with symbiotic partner quality (see **Supp. Dataset S2**, "overlap" column; Epstein et al. 21, Batstone et al. 22). Our nearly-universal gene set contained eight loci that overlapped with the top 100 associations with A17 biomass from Epstein et al. (21). Most notable is the fructose-6-phosphate aminotransferase *nod*M/*glm*S (SMa0878/NP_435728.1) that catalyzes a precursor of both peptidoglycan and Nod factor in the glucosamine biosynthesis pathway. This locus is located in the symbiosis gene region of pSymA, though a paralog exists on the chromosome (SMc00231/NP_385762.1; Barnett and Long 35). Knockout mutants of *nod*M are known to decrease N-fixation of *E. meliloti* on alfalfa (36) and *Rhizobium leguminosarum* (37); together with Epstein et al. (21), our studies highlight the role of natural variation in bacterial glucosamine metabolism in determining plant health. We also found six genes in this nearly-universal set that were also associated with symbiotic partner quality, rhizobium fitness, or both in the experimental evolution study of Batstone et al. (22). Most notable are two *tra* (transfer) loci (*tra*A2 on pSymB and *tra*G on pSymA), potentially part of a Type IV Secretion System (T4SS) responsible for targeting proteins to host cells (38, 39). While the variants we found in these loci are segregating in natural populations in the native range of *E. meliloti*, these loci also evolved *de novo* in response to passaging through the same host for multiple generations (22), making them strong candidates for a consistent role in symbiosis.

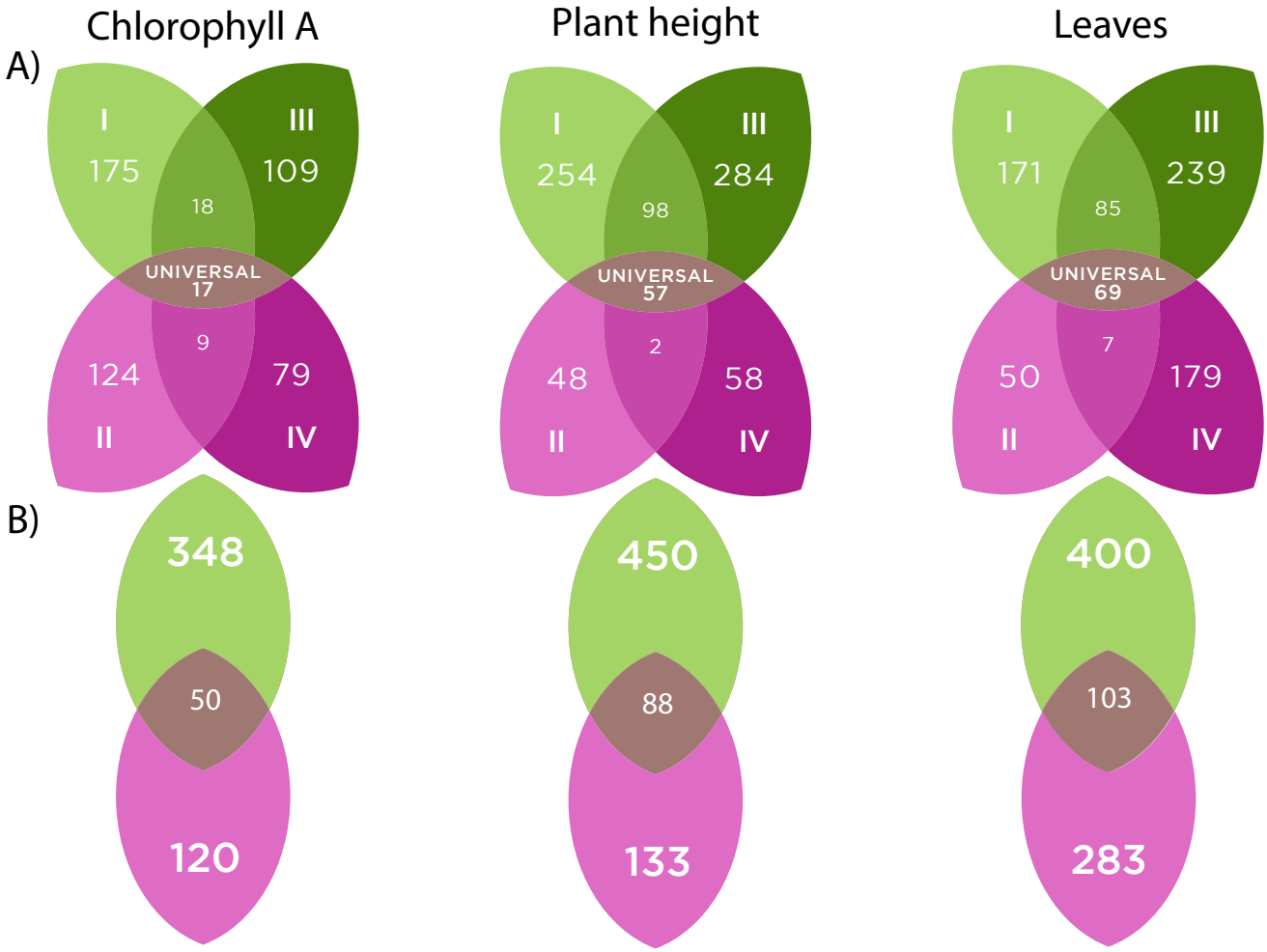
SUPP. DATASETS

- Dataset S1: **"SNPs_ann_ps_shoot.wREADME.xlsx"**
  Summary table for all variants significantly associated with shoot biomass in any of the four experiments. The first tab "SNPs_ann_ps_shoot" provides variant-level information, while the "README" tab provides a brief description of each column in the first tab.
- Dataset S2: **"genes_shoot_uniprot.wREADME.xlsx"**
  Summary table for all genes containing variants significantly associated with shoot biomass in any of the four experiments. The first tab "genes_shoot_uniprot" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.
- Dataset S3: **"SNPs_ann_ps_all.wREADME.xlsx"**
  Summary table for all variants significantly associated with one or more partner quality traits in any of the four experiments. The first tab "SNPs_ann_ps_all" provides variant-level information, while the "README" tab provides a brief description of each column in the first tab.
- Dataset S4: **"SNPs_ann_gene_all.wREADME.xlsx"**
  Summary table for all genes containing variants significantly associated with one or more partner quality traits. The first tab "SNPs_ann_gene_all" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.
- Dataset S5: **"DAVID_outputs_combined.wREADME.xlsx"**

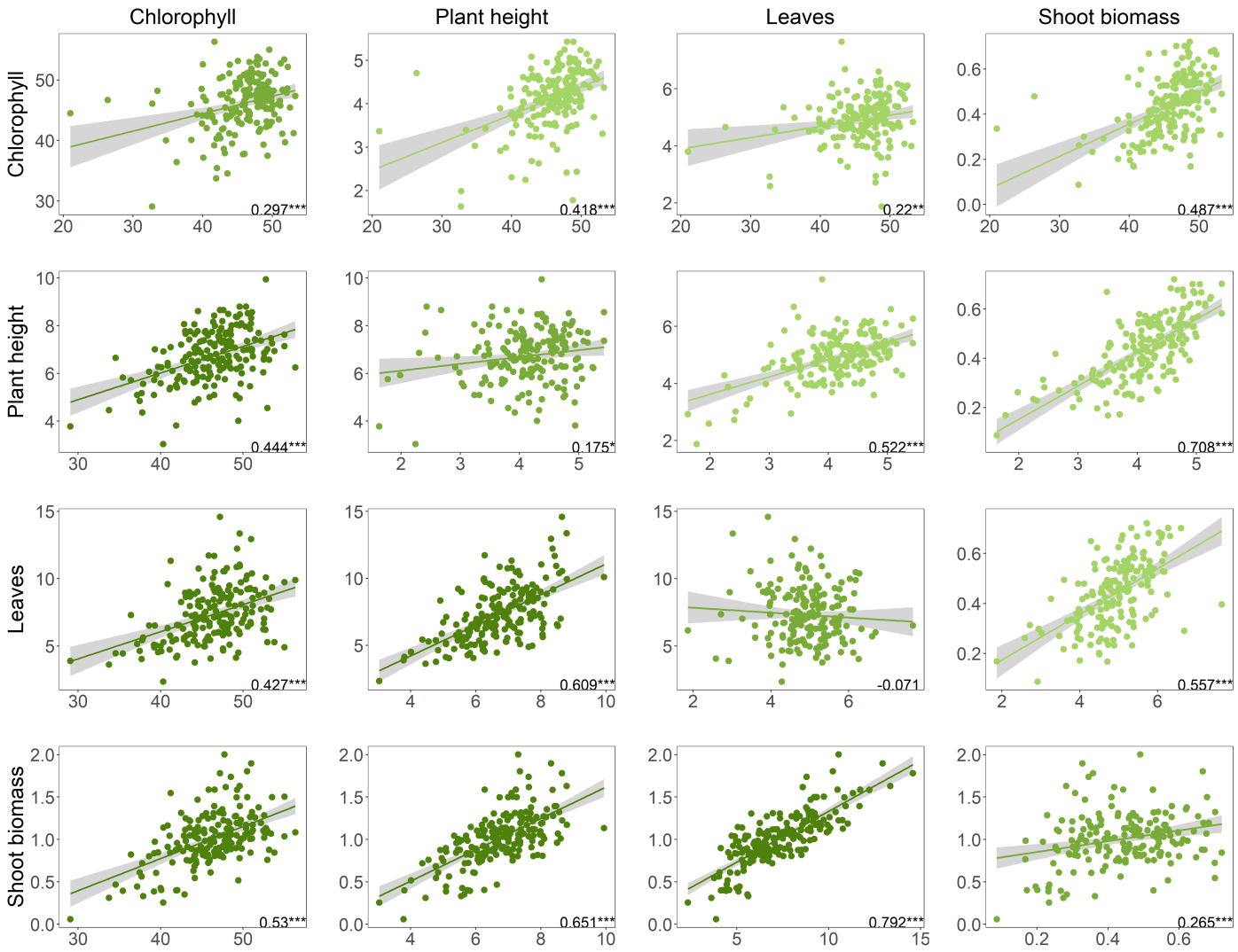
Summary table for all genes run through DAVID for which terms (GO, UNIPROT, KEGG pathways, et c.) were significantly (or marginally) enriched. The first tab "DAVID_outputs_combined" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.

- Dataset S6: **"plast_overlap_shoot.wREADME.xlsx"**
  Summary table for all genes containing variants significantly associated with either shoot biomass or plasticity for shoot biomass or both (i.e., overlapping). The first tab "plast_overlap_shoot" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.
- Dataset S7: **"genes_shoot.plast.wREADME.xlsx"**
  Summary table for all genes containing variants significantly associated with plasticity based on shoot biomass. The first tab "genes_shoot.plast" provides gene-level information, while the "README" tab provides a brief description of each column in the first tab.

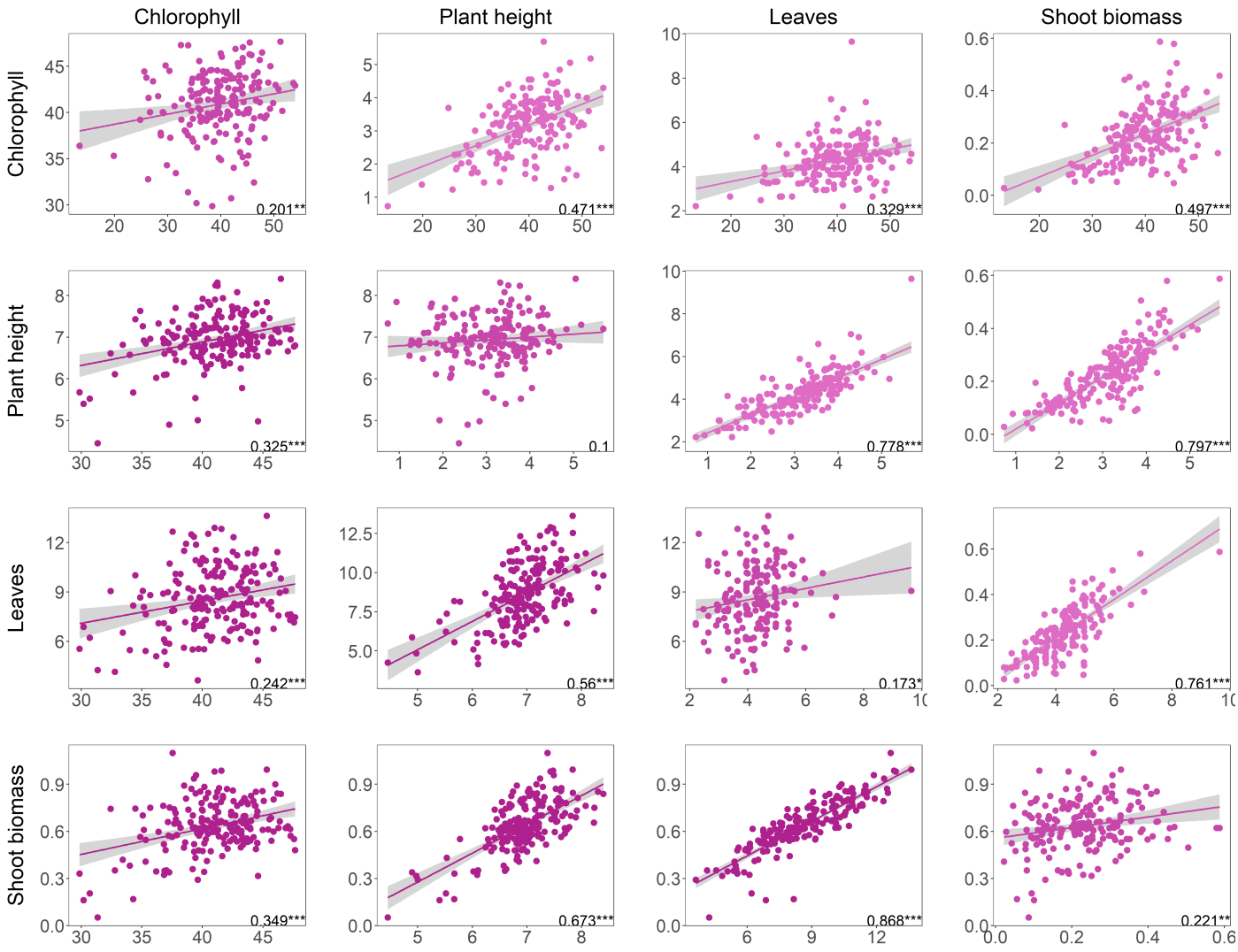
SUPP. FIGURES



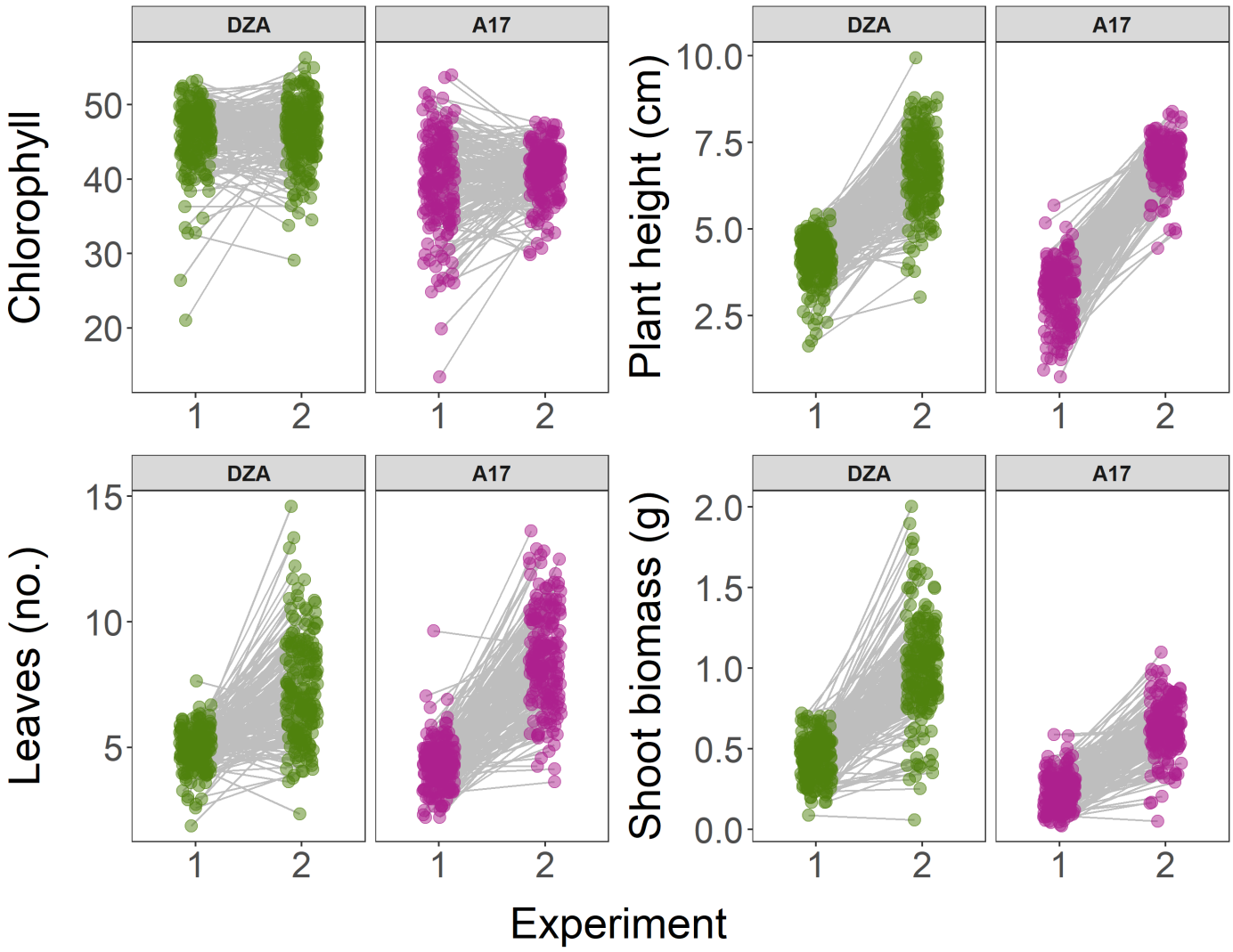
Supp. Figure S1 **Genes underlying G × E are prevalent across partner quality traits. Venn diagrams showing number of rhizobium (*E. meliloti*) genes significantly associated with three other partner quality phenotypes for A) each of four separate mapping experiments or B) cross-experiment plasticity for either host genotype DZA in green (experiments I and III, in green) or A17 (experiments II and IV in pink). The mauve oval in the center represents either A) universal genes that contribute to trait variation in at least three of the four experiments or B) genes associated with cross-experiment trait plasticity in both host genotypes.**



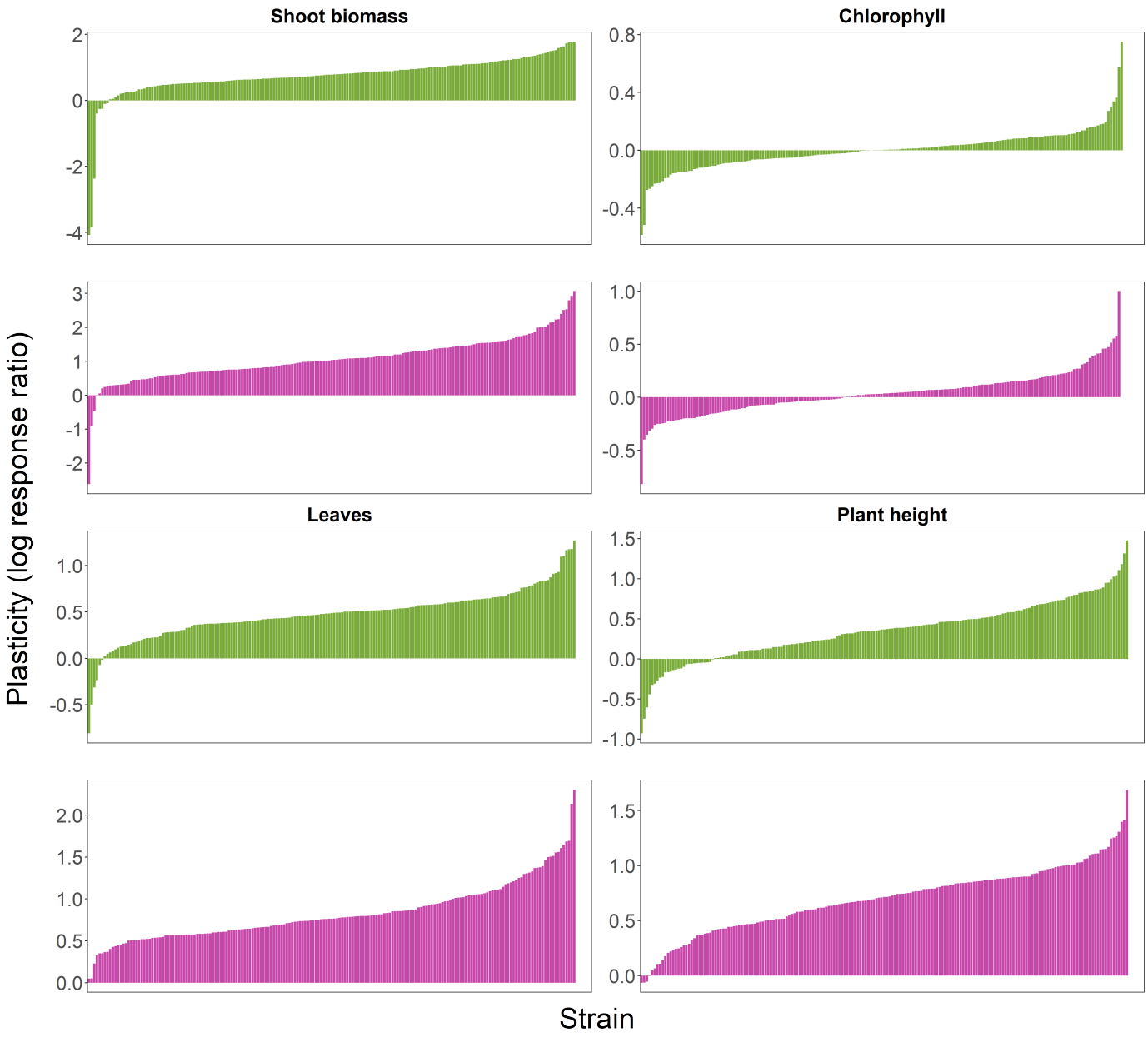
Supp. Figure S2 **Genetic correlations for traits measured on DZA. Genetic correlations among traits within experiments (above or below diagonal) or between the same trait across experiments (along diagonal). Correlations based on estimated marginal means of each rhizobia strain corrected for rack on plant line DZA. Numbers in bottom right corners of each plot indicate Pearson correlation coefficients. Plots above and below the diagonal are for traits measured in experiments 1 and 3, respectively. Significance: $p < 0.001$ = '***'; $p < 0.01$ = '**'; $p < 0.05$ = '*'.**



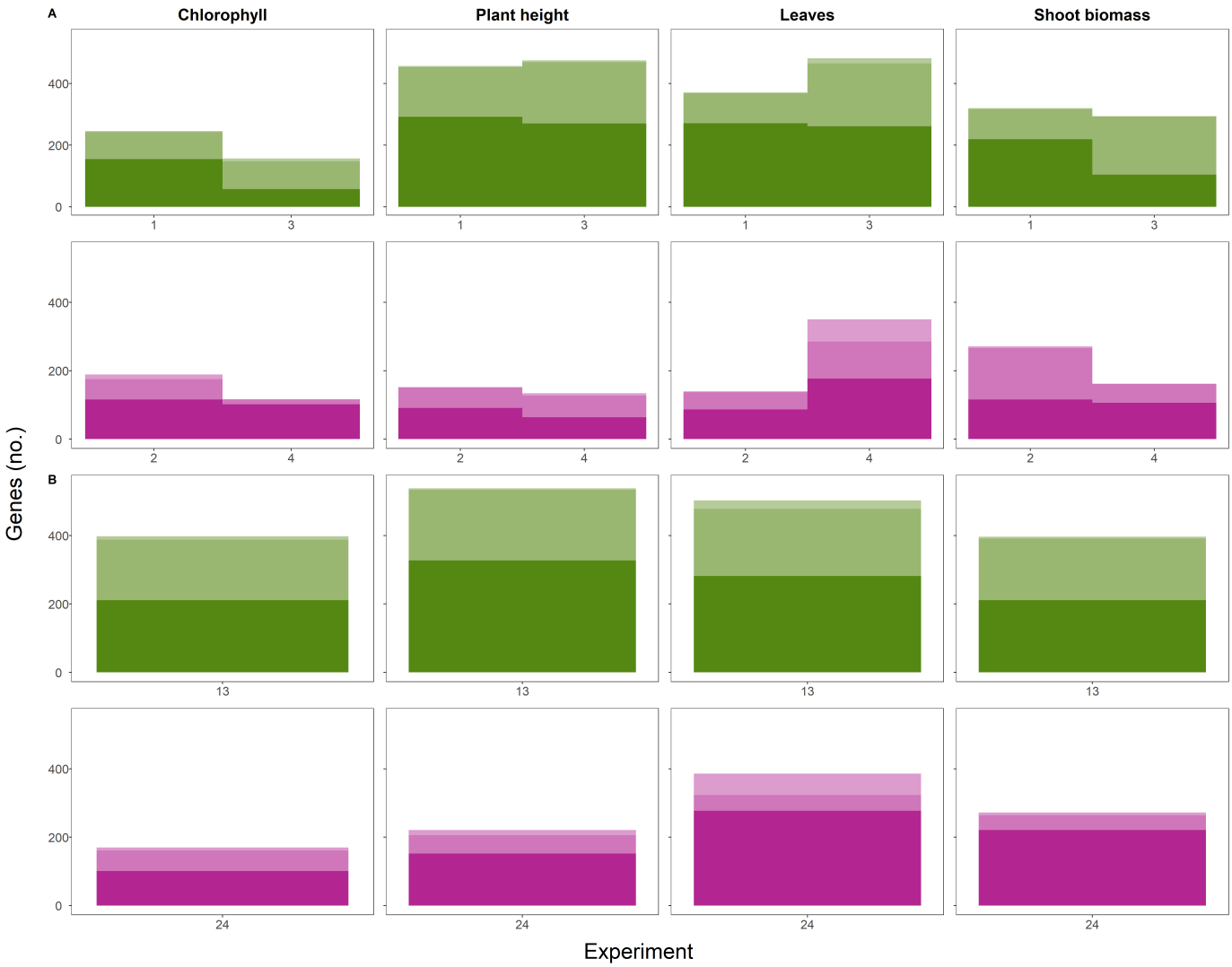
Supp. Figure S3 **Genetic correlations for traits measured on A17. Genetic correlations among traits within experiments (above or below diagonal) or between the same trait across experiments (along diagonal). Correlations based on estimated marginal means of each rhizobia strain corrected for rack on plant line A17. Numbers in bottom right corners of each plot indicate Pearson correlation coefficients. Plots above and below the diagonal are for traits measured in experiments 2 and 4, respectively. Significance: $p < 0.001$ = '***'; $p < 0.01$ = '**'; $p < 0.05$ = '*'.**



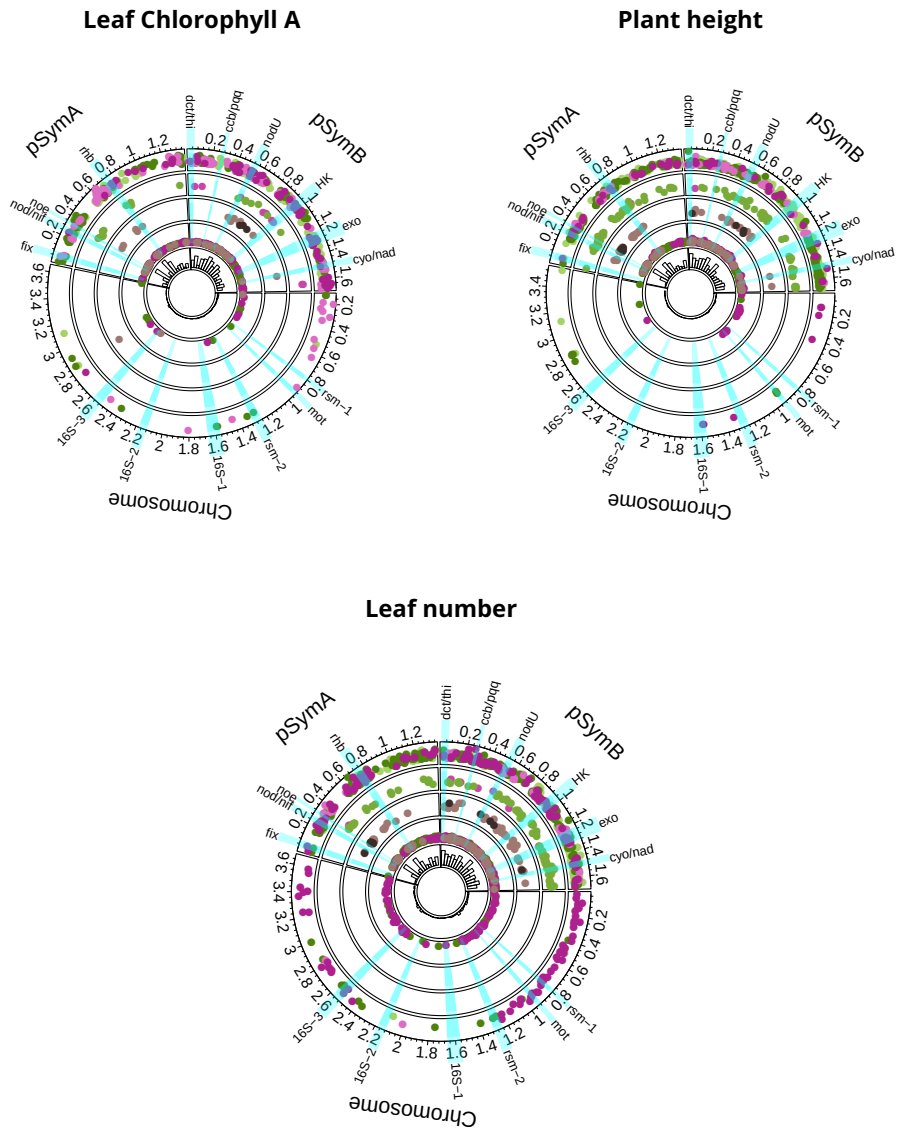
Supp. Figure S4 **Extensive G × E between experiments. Reaction norms for partner quality traits across experiments. Data points represent estimated marginal means corrected for rack within each experiment. Points in green for means estimated on DZA, purple for A17.**



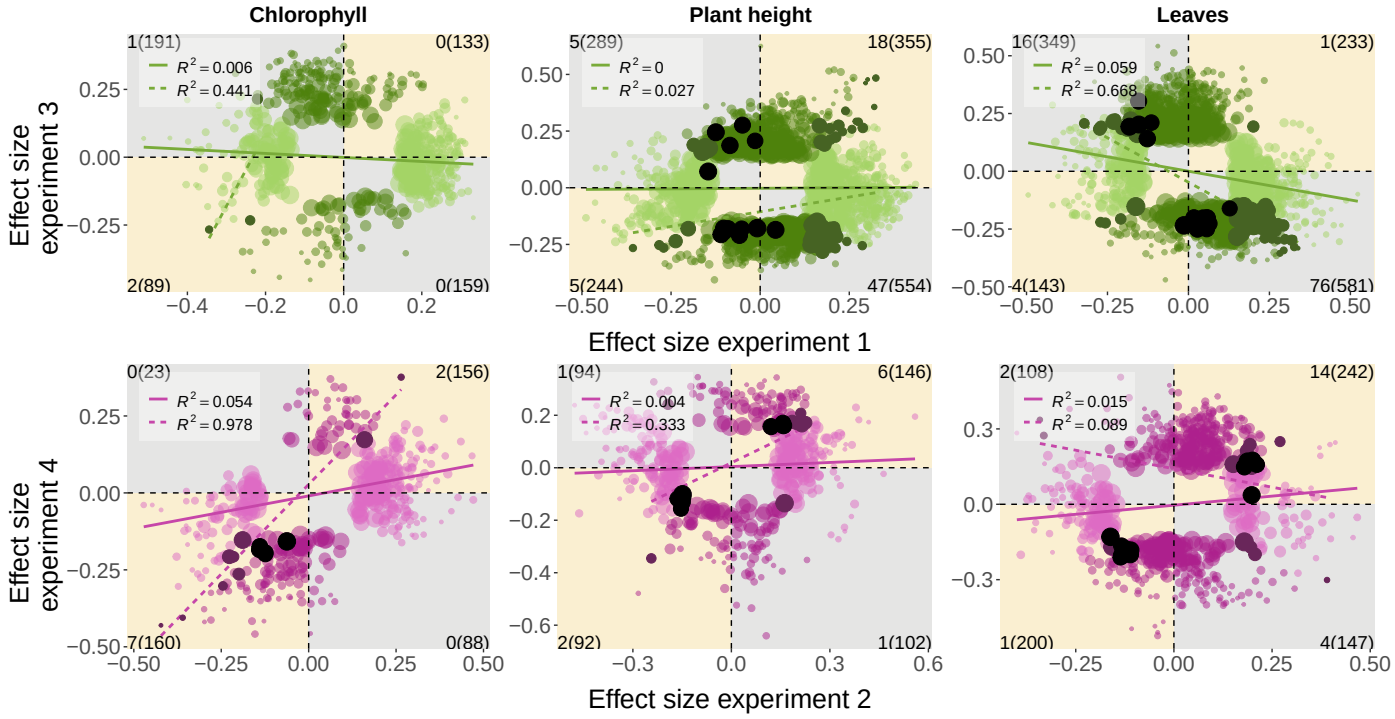
Supp. Figure S5 **Strains significantly varied in their response to different experiments. Variation among *E. meliloti* strains in plasticity, calculated as the log response ratio for each trait between the two experiments with each host genotype (experiments 1 & 3 with DZA in green; experiments 2 & 4 with A17 in pink).**



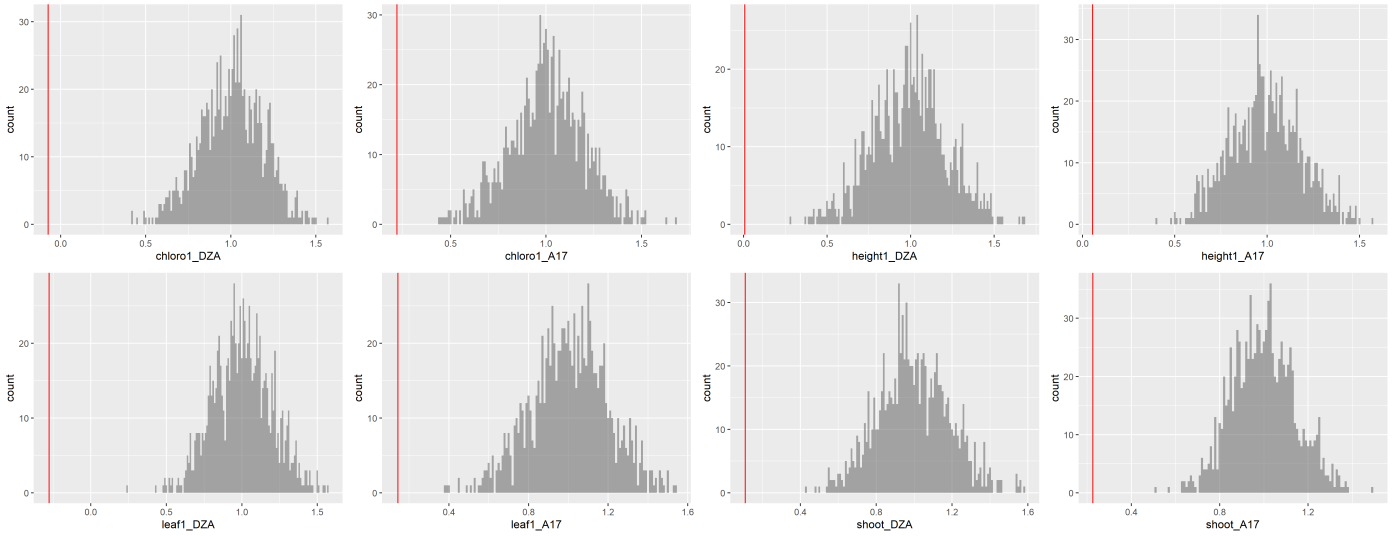
Supp. Figure S6 **Small-scale shifts of loci across genomic regions that contribute to partner quality variation. Distribution of genomic locations (chromosome in lightest shade, pSymA in medium shade, and pSymB in darkest shade) for the *E. meliloti* loci significantly associated with partner quality phenotypes in each of four mapping experiments with either host genotype DZA (green) or host genotype A17 (pink). The plasticity panels (B) represent the genomic locations of rhizobium loci associated with the response of each trait across the two experiments for each host line (1-3 for DZA and 2-4 for A17).**



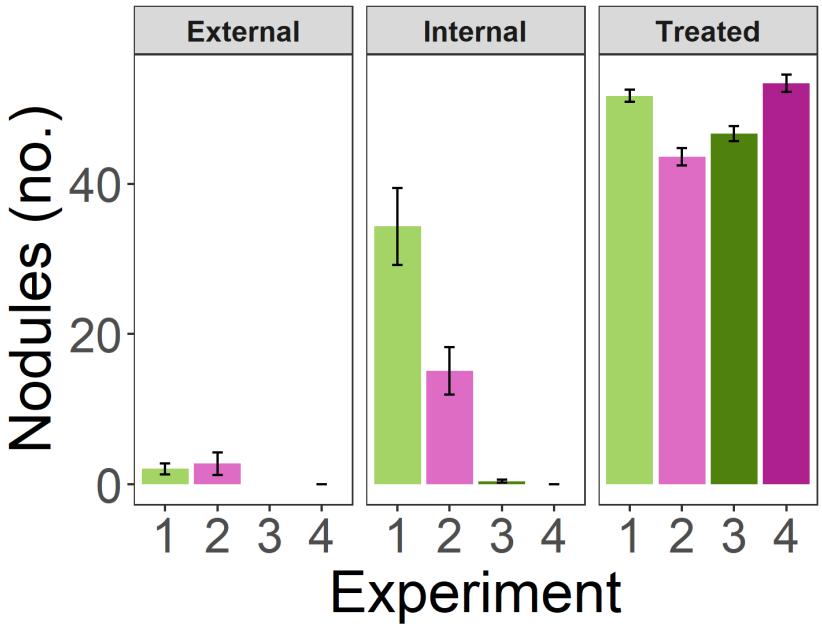
Supp. Figure S7 **Loci associated with partner quality are mostly limited to the symbiosis plasmids. Circos plots showing positions of genes (dots) significantly associated with three additional partner quality traits. Each ring represents a different gene category, outermost to innermost: 1) G x E, 2) G x G, 3) partially universal and universal, 4) plasticity, while 5) depicts a histogram based on the total number of significant genes across 100 kbp-sized windows. The x- and y-axes for rings 1-4 represent genomic position (Mbp) and average absolute effect sizes of variants within each gene, respectively. The colours reflect categories in the Venn Diagrams: for rings 1, 2, and 4, genes associated with DZA-only traits are represented by shades of green, on A17-only with shades of purple, and both hosts in mauve (ring 4). For ring 3, genes associated with both hosts in more than three environments are represented in mauve (i.e., "partially universal"), and universal genes in black. Relevant loci are highlighted in blue, with abbreviations for clusters on the outer circle as follows: rsm-1: rsmD,E; ribF; groL; hisG,Z. mot: fliF,I,N,P,Q,R; flgB,D,F,G; motA,B; flhB. rsm-2: sppA; lptB; rpoN; raiA; ptsN; hrcA; rph; rdgB; ubiB; coaBC; iolB-E; cysK; rmsI; pyrF; queG, corA. 16S-1: metB; rrf (5S rRNA); 16S rRNA; hrpB; hisA,H; addA,B; trxA; trpB; hpcH; gyrB; rho. 16S-2: 16S rRNA; rrf (5S rRNA); oppD; glnQ; hppD; hmgA; maiA; purU; lpdA; modC; cobG,H,M. 16S-3: grxC; ptsP; prmC; clpB; 16S rRNA; rrf (5S rRNA); tkt; deoC. fix: nnrU; norD,E; hemN; nirK; napA,E,F; fixG-L,P,Q,S; ccoN-Q; ric; nosR. nod/nif: nolF,G; nodA-C,D,D3,E,F,H,I,J,N; nifA,B,D,E,H,K,N,T,X; fabG; syrA; fdxB; fixA-C. noe: noeA,B; nodL; ccoN2,O2,P2,Q2; nodD2; groL. rhb: selB; fdhE; fdxH; fdnG; repA,B; katG; rhbC,D,F; basC; kdpB,F. dct/thi: urtA-C; dctD; thiC,O,S; mtnA; nspC; paaB,J,I,X. ccb/pqq: fghA; moxF; gfa; cbbX; rbcL; fba; tkt; pqqA,D,E. nodU: ugpC; ehuA-D; eutA,B; doeA,B; uxuA; galE; nodU. HK (housekeeping): alc; uraH; xdhA-C; guaD; pbpC; doeC; hutG,H; ltrA; phnC-E,N; gabD,T. exo: exsH; cueC-E; galE; exoA,F,H,I,K,L,M,O,P,Q,U,V,W,Y; thiD; nirD; nfeD, der; nnrU; bacA; map; glpK; xdhA; guaD; lldD. cyo/nad: cyoA-D; rfaL; minC,D; nadE; asnB.**



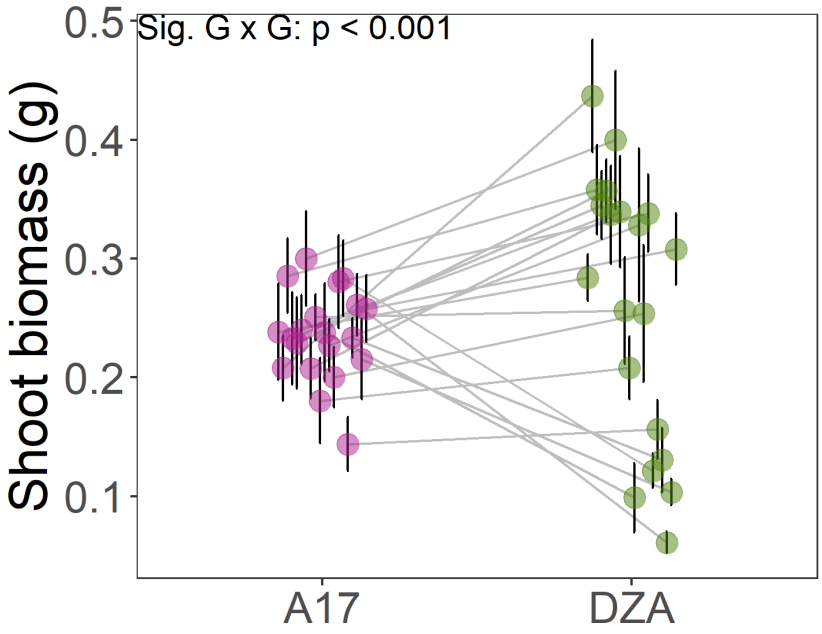
Supp. Figure S8 **Extensive G × E revealed at the variant-level. Variant-level G × E for partner quality loci. Shown are correlations between the estimated effects of individual *E. meliloti* loci on three difference partner quality metrics (from GWAS) in each of two experiments for either host DZA (green) and A17 (pink). Only allelic effects that were significant in one (lighter colours) or both (dark points) environments are shown, while black dots represent nearly universal variants, i.e., associated with the same trait in three experiments. Linear relationships and R^2 values are depicted for all significant variants (solid coloured line) or variants significant in both experiments (dotted coloured line). Variant counts for each quadrant are shown in the corners of each plot (variants significant in both experiments, followed by all variants in parentheses).**



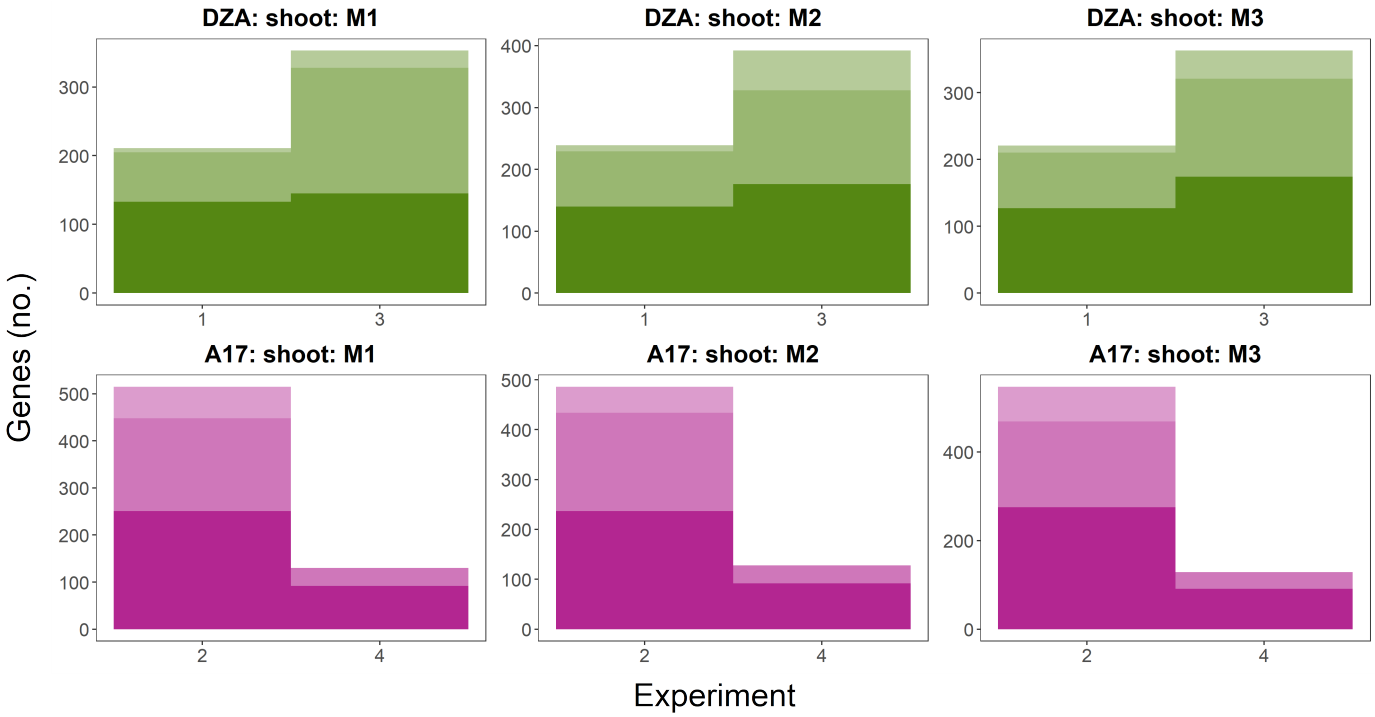
Supp. Figure S9 **Global analyses show that allelic effects are significantly different across experiments. Distributions of slopes calculated by resampling the estimated effects from one experiment (e.g., DZA experiment I) with experimental error (standard deviation), regressing against the observed effects in that experiment, and repeating 1000 times. Red lines depict observed slopes when estimated effects were regressed between experiments.**



Supp. Figure S10 **Contamination was minimal across experiments. Nodule number for inoculated (i.e., Treated) versus Internal and External uninoculated control plants in four experiments: Experiments 1 and 3 used host genotype DZA (shades of green) while 2 and 4 used genotype A17 (shades of pink).**



Supp. Figure S11 **Significant Genotype-by-genotype (G × G) interactions between two host genotypes (A17 and DZA) and 20 *E. meliloti* strains. All 20 strains were included in the 191 strains used in current paper. Both hosts were grown together in a single experiment (Heath et al., unpublished data). Type III ANOVA based on a linear mixed model that corrected for rack and researcher was used to test for the main effects of rhizobium strain (χ^2 = 94.891, p < 0.001), host genotype (χ^2 = 0.553, p = 0.457), and strain-by-host (G x G) interaction (χ^2 = 83.272, p < 0.001) on plant shoot biomass.**



Supp. Figure S12 **Minimal differences in the number of significant genes across three different methods. To ensure the significant variants identified in our study did not depend on the computational methods used (i.e., how the k-matrix was computed), we compared the number of genes tagged by variants significantly associated with shoot biomass in all four experiments and both plant lines (top row in shades of green = DZA; bottom row in shades of pink = A17) for three separate methods: M1) k-matrix calculated for each genomic region separately, only unlinked variants included as input; M2) k-matrix calculated for the whole genome, only unlinked variants included as input; and M3) k-matrix calculated for the whole genome, linked variants included as input. The different shades represent whether genes were located on the chromosome (lightest), pSymA (medium), and pSymB (darkest).**

SUPP. REFERENCES

1. **Tirichine L, de Billy F, Huguet T**. 2000. Mtsym6, a gene conditioning *Sinorhizobium* strain-specific nitrogen fixation in *Medicago truncatula*. Plant Physiol 123 (3):845–852.
2. **Wang Q, Yang S, Liu J, Terecskei K, Ábrahám E, Gombár A, Domonkos Á, Szűcs A, Körmöczi P, Wang T, et al.**. 2017. Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. Proc Natl Acad Sci 114 (26):6854–6859.
3. **Yang S, Wang Q, Fedorova E, Liu J, Qin Q, Zheng Q, Price PA, Pan H, Wang D, Griffitts JS, et al.**. 2017. Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. Proc Natl Acad Sci 114 (26):6848–6853.
4. **Wang Q, Liu J, Zhu H**. 2018. Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. Front Plant Sci 9:313.
5. **Grillo MA, De Mita S, Burke PV, Solórzano-Lowell KL, Heath KD**. 2016. Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in *Medicago truncatula*. Evolution 70 (12):2704–2717.
6. **Riley A, Grillo M, Epstein B, Tiffin P, Heath K**. 2021. Partners in space: Discordant population structure between legume hosts and rhizobium symbionts in their native range .
7. **Somasegaran P, Hoben HJ**. 1994. Methods in Legume-Rhizobium Technology. *In* Handbook for Rhizobia. Springer, New York, NY, US.
8. **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al.**. 2011. The variant call format and VCFtools. Bioinformatics 27 (15):2156–2158.
9. **Team RC**. 2016. R Foundation for statistical computing. R: a language environment for statistical computing .
10. **Bates D, Sarkar D, Bates MD, Matrix L**. 2007. The lme4 package. R Package Version 2 (1):74.
11. **Cockerham CC**. 1963. Estimation of genetic variances. Stat Genet Plant Breed 982:53–94.
12. **Muir W, Nyquist W, Xu S**. 1992. Alternative partitioning of the genotype-by-environment interaction. Theor Appl Genet 84 (1-2):193–200.
13. **Batstone RT, Peters MA, Simonsen AK, Stinchcombe JR, Frederickson ME**. 2020. Environmental variation impacts trait expression and selection in the legume–rhizobium symbiosis. Am J Bot 107 (2):195–208.
14. **Searle SR, Speed FM, Milliken GA**. 1980. Population marginal means in the linear model: an alternative to least squares means. The Am Stat 34 (4):216–221.
15. **Heath KD**. 2010. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. Evolution 64 (5):1446–1458.
16. **Heath KD, Burke PV, Stinchcombe JR**. 2012. Coevolutionary genetic variation in the legume-rhizobium transcriptome. Mol Ecol 21 (19):4735–4747.
17. **Burghardt LT, Guhlin J, Chun CL, Liu J, Sadowsky MJ, Stupar RM, Young ND, Tiffin P**. 2017. Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. Mol Ecol 26 (21):6122–6135.
18. **Burghardt LT, Epstein B, Tiffin P**. 2019. Legacy of prior host and soil selection on rhizobial fitness in planta. Evolution 73 (9):2013–2023.
19. **Fagorzi C, Bacci G, Huang R, Cangioli L, Checcucci A, Fini M, Perrin E, Natali C, diCenzo GC, Mengoni A**. 2021. Nonadditive Transcriptomic Signatures of Genotype-by-Genotype Interactions during the Initiation of Plant-Rhizobium Symbiosis. MSystems 6 (1):e00974–20.
20. **Zhou X, Stephens M**. 2014. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat Methods 11 (4):407–409.
21. **Epstein B, Abou-Shanab RA, Shamseldin A, Taylor MR, Guhlin J, Burghardt LT, Nelson M, Sadowsky MJ, Tiffin P**. 2018. Genome-wide association analyses in the model Rhizobium *Ensifer meliloti*. MSphere 3 (5).
22. **Batstone RT, O'Brien AM, Harrison TL, Frederickson ME**. 2020. Experimental evolution makes microbes more cooperative with their local host genotype. Science 370 (6515):476–478.
23. **Quinlan AR, Hall IM**. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26 (6):841–842.
24. **Rockman MV**. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66 (1):1–17.
25. **Korte A, Farlow A**. 2013. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods 9 (1):1–9.
26. **Tibbs Cortes L, Zhang Z, Yu J**. 2021. Status and prospects of genome-wide association studies in plants. The Plant Genome 14 (1):e20077.
27. **Lasky JR, Forester BR, Reimherr M**. 2018. Coherent synthesis of genomic associations with phenotypes and home environments. Mol Ecol Resour 18 (1):91–106.
28. **Lorts CM, Lasky JR**. 2020. Competition× drought interactions change phenotypic plasticity and the direction of selection on Arabidopsis traits. New Phytol 227 (4):1060–1072.
29. **Lau JA, Bowling EJ, Gentry LE, Glasser PA, Monarch EA, Olesen WM, Waxmonsky J, Young RT**. 2012. Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. Acta Oecologica 39:80–86.
30. **Heath KD, Podowski JC, Heniff S, Klinger CR, Burke PV, Weese DJ, Yang WH, Lau JA**. 2020. Light availability and rhizobium variation interactively mediate the outcomes of legume–rhizobium symbiosis. Am J Bot 107 (2):229–238.
31. **Huang DW, Sherman BT, Tan Q, Kir J, Liu D, Bryant D, Guo Y, Stephens R, Baseler MW, Lane HC, et al.**. 2007. DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Res 35 (suppl_2):W169–W175.
32. **Sherman BT, Lempicki RA, et al.**. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4 (1):44–57.
33. **Karp PD, Billington R, Caspi R, Fulcher CA, Latendresse M, Kothari A, Keseler IM, Krummenacker M, Midford PE, Ong Q, et al.**. 2019. The BioCyc collection of microbial genomes and metabolic pathways. Briefings Bioinform 20 (4):1085–1093.
34. **Caspi R, Billington R, Keseler IM, Kothari A, Krummenacker M, Midford PE, Ong WK, Paley S, Subhraveti P, Karp PD**. 2020. The MetaCyc database of metabolic pathways and enzymes- a 2019 update. Nucleic Acids Res 48 (D1):D445–D453.
35. **Barnett MJ, Long SR**. 2018. Novel genes and regulators that influence production of cell surface exopolysaccharides in Sinorhizobium meliloti. J Bacteriol 200 (3):e00501–17.
36. **Baev N, Schultze M, Barlier I, Ha DC, Virelizier H, Kondorosi E, Kondorosi A**. 1992. *Rhizobium nod*M and *nod*N genes are common nod genes: *nod*M encodes functions for efficiency of nod signal production and bacteroid maturation. J Bacteriol 174 (23):7555–7565.
37. **Marie C, Barny MA, Downie J**. 1992. *Rhizobium leguminosarum* has two glucosamine syntheses, *gim*S and *nod*M, required for nodulation and development of nitrogen-fixing nodules. Mol Microbiol 6 (7):843–851.

38. **Cao Y, Miller SS, Dornbusch MR, Castle SS, Lenz P, Ferguson J, Sadowsky MJ, Nelson MS, Klatt C, Samac DA**. 2018. Widespread occurrence of Sinorhizobium meliloti strains with a type IV secretion system. Symbiosis 75 (2):81–91.

39. **Paço A, da Silva J, Eliziário F, Brígido C, Oliveira S, Alexandre A**. 2019. *tra*G Gene Is Conserved across *Mesorhizobium* spp. Able to Nodulate the Same Host Plant and Expressed in Response to Root Exudates. BioMed Res Int 2019.