The complex genetics of symbiotic extended phenotypes across environments in a model mutualism

Rebecca T. Batstone[*,1,2], **Hanna Lindgren**[†,1], **Cassandra M. Allsup**[†], **Laura A. Goralka**[†], **Alex B. Riley**[†], **Michael A. Grillo**[‡], **Amy Marshall-Colon**[*,†] **and Katy D. Heath**[*,†,3]

[*]Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801 USA, [†]Department of Plant Biology, University of Illinois at Urbana-Champaign, 286 Morrill Hall, 505 South Goodwin Avenue, Urbana, IL 61801 USA, [‡]Department of Biology, Loyola University Chicago, 1032 W. Sheridan Road, Chicago, IL 60618 USA

ABSTRACT A goal of modern biology is to develop the genotype-phenotype (G→P) map, a predictive understanding of how genomic information generates the organismal trait variation that forms the basis of both natural and managed communities. As microbiome research advances, however, it has become clear that many of these traits are governed by genetic variation encoded not only by the host's own genome, but also by the genomes of myriad cryptic symbionts. Thus many ecologically-important traits are likely symbiotic extended phenotypes, and this recognition adds even more complexity to our conceptions of the G→P map. In model symbioses such as the legume-rhizobium mutualism, host growth and fitness often depend on genetic variation in symbiont partner quality, and our ability to manipulate host and symbiont genotype combinations, combined with increasingly precise sequencing and mapping approaches, provides an opportunity to characterize the genetic nature of these symbiotic extended phenotypes. Here we use naturally-occurring genetic variation in 191 strains of the nitrogen-fixing symbiont *Ensifer meliloti* in four mapping experiments to study the genomic architecture of symbiotic partner quality within and across environmental contexts and host genotypes. We demonstrate the quantitative genetic nature of symbiotic extended phenotypes, including extensive context-dependency in both the identity and functions of symbiont loci that control host growth. We additionally resolve a core set of universal loci from populations in the native range that are likely important in all or most environments, and thus, serve as excellent targets both for genetic engineering and future coevolutionary studies of symbiosis.

KEYWORDS GWAS; mapping; *Medicago truncatula*; mutualism; partner quality; rhizobium; GxE; GxG; genotype-phenotype map, G→P map, root nodule, nodulation, symbiotic nitrogen fixation

Introduction

We live in a symbiotic world. It is increasingly recognized that many important traits in macrobes, including human metabolism, insect diet and defense, and plant nutrient foraging are actually *symbiotic extended phenotypes* governed (at least in part) by cryptic variation in their microbial symbionts (e.g., Kironomos 2003; Oliver *et al.* 2005; Ridaura *et al.* 2013; Brock *et al.* 2016; Gilbert and Lynch 2019; O'Brien *et al.* 2021; Afkhami *et al.* 2021; reviewed by Brundrett 2002; Friesen *et al.* 2011; Hurst 2017; Hawkes *et al.* 2020; Moeller and Sanders 2020). Incorporating the complex genetics arising from these interactions, however, has lagged behind the ecological effects. Characterizing the genetic basis of these cross-domain relationships represents a symbiotic extension of the genotype-phenotype (G→P) map (Hansen 2006; Orgogozo *et al.* 2015; Mackay and Huang 2018), wherein genetic variation present in symbionts (G) presents as phenotypic variation (P) expressed by the host. In symbiosis, as in all organisms, the G→P map is a crucial step towards a predictive understanding of evolution. Rates and trajectories of the phenotypic response to selection depend on the underlying genetic architecture (Hansen 2006; McGuigan 2006), conflict and cooperation in symbiosis are governed by causal genomic variants

doi: https://doi.org/10.1101/2021.08.03.454976
Manuscript compiled: Monday 9th August, 2021, using the Overleaf template available at https://www.overleaf.com/gallery/tagged/gsa

[1]These authors contributed equally to this work.

[2]Co-corresponding author: Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 West Gregory Drive, Urbana, IL 61801 USA. E-mail: rtbatstone@gmail.com

[3]Co-corresponding author: Department of Plant Biology, University of Illinois at Urbana-Champaign, 286 Morrill Hall, 505 South Goodwin Avenue, Urbana, IL 61801 USA. E-mail: kheath@illinois.edu

with pleiotropic effects on host and symbiont fitness (Queller and Strassmann 2018; O'Brien *et al.* 2021; Batstone *et al.* 2021), while a fuller understanding of the regulatory and metabolic basis of symbiosis can illuminate how key evolutionary innovations arise (e.g., Douglas 1994; Otero-Bravo *et al.* 2018; Bublitz *et al.* 2019; Thairu and Hansen 2019). The G→P is also crucial for bioengineering, where we must know not only which genes to target and exactly how to edit them, but also anticipate how these changes will affect non-target phenotypes and how consistent these effects will be across diverse environments and/or genetic backgrounds (Checcucci *et al.* 2018; Marshall-Colón and Kliebenstein 2019).

A key lesson from studies of the G→P map within single organisms is that phenotypic variation fundamentally arises in non-additive ways as genomic variation is filtered through "layer(s) of context-dependence" (Orgogozo *et al.* 2015) including genotype-by-environment interactions, epistasis, and pleiotropy (Cheverud 1982, 1996; Hansen 2013; Lee *et al.* 2013; Mackay and Huang 2018; Campbell-Staton *et al.* 2017; Saltz *et al.* 2017). Given the importance of symbiotic extended phenotypes to natural and managed ecosystems, understanding the complex genetics of how two (or more) genomes interact with each other and with their environment to generate trait variation is an important step to better understanding how these traits (co)evolve (O'Brien *et al.* 2021) and how to better manipulate them in the future to address societal challenges (Batstone 2021).

The role of context-dependency in the G→P is particularly relevant for extended symbiotic phenotypes, given the potential for higher-order interactive genetic effects. In addition to epistasis or G × E for a single genome, interacting genomes generate genotype-by-genotype interactions between host and symbiont (intergenomic epistasis, or G × G: Frank 1992; Parker 1995; Lambrechts *et al.* 2006; Wade 2007; Heath 2010; Parker *et al.* 2017 and even G × G × E whereby trait values depend not only on the interaction of alleles in both partners, but on the environmental context (Piculell *et al.* 2008; Heath *et al.* 2010; Wendling *et al.* 2017). Add to this the fascinating genetics of bacteria: diverse, often multipartite, genomes driven by mobility (Harrison *et al.* 2010; Dicenzo and Finan 2017; Misra *et al.* 2018), which generate additional within-genome intergenomic epistasis between chromosome and mobile genetic elements (MGEs) or among MGEs (e.g., plasmids; Silva *et al.* 2011; Harrison and Brockhurst 2012; San Millan *et al.* 2014). On the one hand, these complex context-dependent effects are critical for evolution; for example, G × G × E for fitness is the statistical expression of the coevolutionary selection mosaic (Thompson 2005, 2009; Wade 2007). On the other hand, genes that consistently affect symbiotic extended phenotypes arising in plant-microbe interactions, independent of these layers of context dependency, are important targets for breeding and bioengineering (Friesen *et al.* 2011; Busby *et al.* 2017; Checcucci *et al.* 2018; Porter and Sachs 2020; Batstone 2021). Addressing both types of genetic effects requires fine-scale mapping of the loci contributing to symbiotic extended phenotypes across partner genotype and environmental contexts to generate a more holistic picture of the genes that universally contribute and those that are context-dependent, providing insights both for natural (co)evolutionary processes and sustainable agriculture.

The symbiosis between nitrogen-fixing rhizobia and leguminous plants has been a key model for the genetics of symbiosis (Cook 1999; Heath and Grillo 2016; Sachs *et al.* 2018; diCenzo *et al.* 2019; Burghardt 2020), with resequencing projects, germplasm collections, and many genetic and genomic tools, particularly in *Medicago-Ensifer* interactions (Young *et al.* 2005; Epstein *et al.* 2012, 2018; Stanton-Geddes *et al.* 2013; Burghardt *et al.* 2018). These symbioses are also major drivers of the terrestrial nitrogen (N) cycle, contributing up to 21 million tonnes of fixed N each year (Graham and Vance 2003; Foyer *et al.* 2016). In their specialized root nodules, legumes trade photosynthetically-derived fixed carbon in exchange for N fixed by the rhizobia; because of this symbiosis, legumes are keystone members of natural plant communities and also contribute to sustainable agriculture as food and forage crops and in rotation with non-legume crops (reviewed by Fustec *et al.* 2010; Duchene *et al.* 2017). Despite the ubiquity and importance of the symbiosis in N cycling, there is immense variability in the outcome of individual interactions at the intraspecific level, with interactions ranging from highly beneficial to even parasitic depending on the genotype of the rhizobium and (at least sometimes) its interaction with the host plant genotype (Burdon and Thrall 1999; Thrall *et al.* 2000; Heath 2010; Barrett *et al.* 2012; Porter and Simms 2014; Wendlandt *et al.* 2019), making this symbiosis a key model for both additive and interactive genetic effects in symbiosis (i.e., G × G, G × E).

Rhizobial genomes themselves have been studied as models of bacterial genome evolution, reflecting the dynamic tension between the core set of genes, often on the main chromosome, and the more flexible genes, which are often mobile and found on plasmids, islands, and other elements (Harrison *et al.* 2010; Young *et al.* 2016; Dicenzo and Finan 2017; Misra *et al.* 2018; Cavassim *et al.* 2020). While bacteria vary considerably in the rate of recombination relative to mutation (Didelot and Maiden 2010; Power *et al.* 2017; González-Torres *et al.* 2019), rhizobia and many other bacteria show abundant recombination, allowing genome-wide association studies (GWAS) to detect associations between individual genomic variants and traits of interest (Klinger *et al.* 2016; Epstein *et al.* 2018; Porter *et al.* 2017; Power *et al.* 2017; Cavassim *et al.* 2020). Symbiosis plasmids, in particular, recombine readily - allowing particularly fine resolution of genes located there (Epstein *et al.* 2018; Nelson *et al.* 2018; Epstein and Tiffin 2021). Decades of functional genetic studies have resolved a number of symbiosis genes, mostly those that disrupt symbiosis when knocked out (reviewed by Wang *et al.* 2012; Roy *et al.* 2020), enabling increasingly well-resolved models for how interactions establish and how trade is coordinated (diCenzo *et al.* 2020). Nevertheless, mapping efforts in legume-rhizobium symbiosis, which leverage standing genetic variation, indicate that a much broader set of genes and genetic pathways contribute to the quantitative variation in partner quality that we know to be important in natural and managed systems (Stanton-Geddes *et al.* 2013; Klinger *et al.* 2016; Epstein *et al.* 2018; Porter *et al.* 2019; Batstone *et al.* 2020a). Thus quantitative genetic mapping efforts can generate novel candidate genes for symbiosis while shedding light on the G→P map of symbiotic extended phenotypes in nature.

Here we study the complex genetic basis and layers of context-dependency in the G→P map of an ecologically-important symbiotic extended phenotype: partner quality in the legume-rhizobium mutualism. Using a panel of 191 strains of the model rhizobium *Ensifer meliloti* isolated from natural populations in the native range, we ask whether the genetic architecture of rhizobium partner quality differs across environments and across host genotypes, or alternatively if variation at a core set of rhizobium genes control partner quality in all contexts; study how the three genomic elements of *E. meliloti*

contribute to partner quality variation across two host genotypes and in multiple environments, validating the functional division of labor of these elements in nature; and finally suggest candidate metabolic functions and genetic loci that are responsible for both context-dependency as well as universally-beneficial effects that are consistent across environments.

Materials and 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 level of nitrogen fixed (Fix+, Fix–, or intermediate) in particular host and strain combinations (Tirichine *et al.* 2000; Wang *et al.* 2017; Yang *et al.* 2017; Wang *et al.* 2018); 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 (Grillo *et al.* 2016; Riley *et al.* 2021).

Experimental design

We performed four greenhouse experiments to estimate partner quality phenotypes in *E. meliloti*. In each experiment, plants from one of two host lines (either A17 or DZA) were grown in single inoculation with each of 191 strains, with three replicates per strain per experiment (six total replicates for each plant line x strain combination). Uninoculated controls (40 per experiment) were included to gauge contamination. Experiments were planted in 2018, in March (I: DZA), May (II: A17), Sept (III: DZA), and Nov (IV: A17). Contamination in these experiments was minimal overall, and limited to the first two experiments (**Supp. Fig. S1**).

Seeds were scarified with either a razor blade or sandpaper, then sterilized by first rinsing with EtOH followed by soaking in commercial bleach 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 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 $\sim 10^6$ 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.* (2021). 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.* (2021). Variant data were hard-filtered using vcftools (v0.1.17, Danecek *et al.* 2011) 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 (Team 2016), we implemented linear mixed models (LMMs) using the R package lme4 (v1.1-27.1, Bates *et al.* 2007) 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 (Cockerham 1963; Muir *et al.* 1992; Batstone *et al.* 2020b). 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, Searle *et al.* 1980).

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; $F_{df=19}$ = 83.3; $p < 0.0001$ for plant aboveground biomass; **Supp. Fig. S2**), similarly to several other studies in this system (e.g., Heath 2010; Heath *et al.* 2012; Burghardt *et al.* 2017, 2019; Fagorzi *et al.* 2021). 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 a 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, Zhou and Stephens 2014). 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.* (2018) (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 (Epstein *et al.* 2018). 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. S3**), 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 (Epstein *et al.* 2018; Batstone *et al.* 2020a). 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, Quinlan and Hall 2010), and excluded any intergenic regions. Next, we binned these genes into three categories based on the context-dependency of their phenotypic effects and thus their contribution to the layers of the G→P map for each of our symbiotic extended phenotypes. First, "nearly-universal genes" were those found to have significant effects in at least three of the four experiments for a particular trait ("universal genes" were mapped in all four experiments). We used a conservative approach to call "G × G genes" as those mapped in both experiments for one host genotype but neither of the experiments for the other host genotype (i.e., "DZA G × G" genes were significant in both experiments I and III with DZA but neither II nor IV with host A17).

Finally, we were interested in the genetic basis of environmental-dependency (i.e., "G × 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 (Rockman 2012), 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 (Rockman 2012; Korte and Farlow 2013; Tibbs Cortes *et al.* 2021). Thus we took multiple approaches to inspecting G × 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 (Lasky *et al.* 2018; Lorts and Lasky 2020); 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.* 2012; Heath *et al.* 2020) 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, Huang *et al.* 2007) to test for significantly overrepresented UNIPROT keywords, GO terms, and KEGG pathways, as described in Sherman *et al.* (2009). 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 (Karp *et al.* 2019; Caspi *et al.* 2020).

Data Availability

Strains and plant lines are available upon request. All code will be made available on GitHub upon publication. The Supplementary Materials doc contains additional figures and tables, as well as descriptions of files that can be downloaded from BioRxiv directly. Once raw sequence reads and assemblies will be archived and made available on NCBI, accession numbers to be added here as well.

Results

Our phenotypic and genetic analyses suggest important roles for context-dependency in the genetic control of symbiotic extended phenotypes. Overall, we found 1,453 variants (**Supp. File S1**) in 746 (**Supp. File S2**) unique coding genes associated with shoot biomass, our focal rhizobium partner quality phenotype, while 5,402 variants (**Supp. File S3**) and 1,770 coding genes (**Supp. File S4**) were associated with all partner quality phenotypes analyzed in this study; an additional 572 variants fell into 369 distinct non-coding regions (**Supp. File S3**). Three categories of G→P associations emerged from our analyses (**Fig. 1**; see **Supp. Fig. S4** for other traits). First were conditional associations that depended on the environment (G × E: comparing experiments within host lines, **Fig. 1A**); 510 (68%) were found in only one of the four experiments. Second were "G × G genes", the 51 (7%) genes that were associated with partner quality in both experiments with one host genotype but were found in neither experiment with the other host genotype (**Fig. 1A**). Specifically, 18 A17-specific genes were found in both experiments with this host (but never with DZA), while 33 DZA-specific genes were found in both experiments with this host (but never with A17). Last, but not least, were "universal" genes, the 60 (8%) genes that were found to be associated with rhizobium partner quality independent of host genotype or environment (union in the center of **Fig. 1A**). Specifically, while only five of these genes were associated with partner quality in all four experiments, 55 additional genes were found in three of the four experiments and thus are strong candidates for contributing to partner quality in many conditions. We discuss these three types of genetic effects in turn, at the phenotypic, genomic (three elements in the tripartite genome), gene, and variant level, then conclude with the general implications of our findings for plant-microbe interactions and symbiosis evolution more broadly.

Genotype-by-environment interactions in the G→P map of symbiotic partner quality

Environmental dependence was a pervasive pattern at both the phenotypic, genomic, and individual gene levels, suggesting that different sets of rhizobium genes contribute to symbiotic partner quality under different environmental conditions. At the phenotypic level, ANOVA indicated abundant genetic variance in partner quality traits, including both significant main effects of strain as well as strong strain-by-experiment (G × E) interactions (see **Table 1** and within-experiment heritabilities in **Table 2**). Cross-environment genetic correlations (of strain means between experiments within a host line) were generally significant (**Table 2**), consistent with strain main effects (some strains had consistently higher partner quality than others; see **Supp. Figs. S5** and **S6** for all within and between experiment trait correlations). Nevertheless, for most traits in both hosts, G × E was driven by changes in the rank order of strain means across environments (**Table 2** crossing (%); **Supp. Fig. S8**). Rhizobium strains varied considerably in cross-environment plasticity for shoot biomass, with some strains having consistently low/high partner quality (i.e., resulting in large or small plant biomass) independent of the environment, while others responded strongly to the environmental differences between experiments (**Supp. Fig. S7**).

At the genomic-level, our evidence indicates that G × E was driven by small-scale shifts in the identity, and estimated allelic effects, of individual loci across environments rather than large shifts in which genomic elements contributed to partner

Table 1 Linear mixed model genotype by environment (G x E) ANOVAs for square-root transformed traits measured on plant lines A17 (left) and DZA (right). Rack included as a random effect, while all other terms are fixed. Significance of rack determined by calculating the log likelihood ratio between models with and without the random effect of rack.

Trait	Model term	A17	DZA
Chlorophyll	Intercept	619.4(1)***	1447.65(1)***
	Strain	252.6(190)**	361.67(190)***
	Experiment	0.15(1)	0(1)
	Strain X Experiment	191.38(182)	205.16(183)
	Rack	30.86(1)***	23.91(1)***
Plant height	Intercept	373.58(1)***	355.55(1)***
	Strain	340.68(190)***	268.05(190)***
	Experiment	20.24(1)***	2.5(1)
	Strain X Experiment	308.63(185)***	177.71(183)
	Rack	38.35(1)***	54.12(1)***
Leaves	Intercept	207.69(1)***	164.09(1)***
	Strain	355.37(190)***	207.58(190)
	Experiment	7.85(1)**	0.45(1)
	Strain X Experiment	243.88(185)**	218.54(183)*
	Rack	27.51(1)***	41.67(1)***
Shoot biomass	Intercept	97.65(1)***	179.32(1)***
	Strain	544.81(190)***	499.67(190)***
	Experiment	10.08(1)**	6.93(1)**
	Strain X Experiment	330.06(185)***	262.2(183)***
	Rack	11.32(1)***	155.67(1)***

Significance: $p < 0.001$ = '***'; $p < 0.01$ = '**'; $p < 0.05$ = '*'

Table 2 Broad-sense heritabilities (H^2) and cross-environment genetic correlations (*r*) of traits calculated within each experiment and for each plant line (i.e., I and III for DZA, and II and IV for A17). G × E interactions were partitioned into changes in variance versus rank order (i.e., crossing), and the percent due to crossing is presented.

Trait	Plant line	H^2 (exps. 1/2)	H^2 (exps. 3/4)	*r*	Crossing (%)
Shoot biomass	A17	0.302***	0.262***	0.221**	99.83
Chlorophyll		0	0.093*	0.201**	0
Plant height		0.181***	0.173***	0.1	73.37
Leaves		0.173***	0.127***	0.173*	94.13
Shoot biomass	DZA	0.248***	0.253***	0.265***	87.19
Chlorophyll		0.118*	0.141***	0.297***	99.06
Plant height		0.131**	0.065*	0.175*	99.97
Leaves		0.05	0.084*	-0.071	63.97

Significance: $p < 0.001$ = '***'; $p < 0.01$ = '**'; $p < 0.05$ = '*'

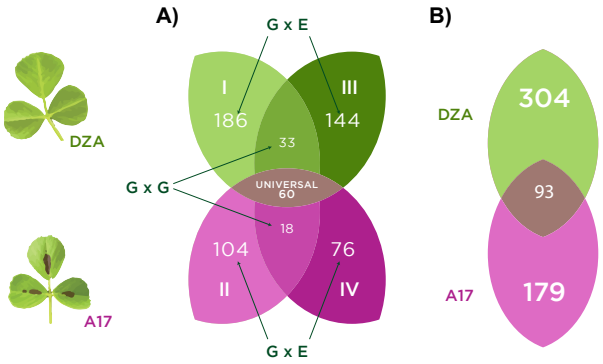


Figure 1 The identities of loci associated with partner quality mostly depend on the environment. Venn diagram showing number of rhizobium (*E. meliloti*) genes significantly associated with plant (*M. truncatula*) aboveground biomass (one metric of symbiotic partner quality) in **A)** each of four separate mapping experiments with either host line DZA (I and III in green) or host line A17 (II and IV in pink). Green and pink unions represent genes contributing to plant biomass in both experiments with either DZA or A17, respectively, while the mauve oval in the center represents universal genes found to contribute to plant aboveground biomass in at least three of the four experiments. **B)** Genes significantly associated with cross-experiment plasticity in plant aboveground biomass for host genotype DZA in green (experiments I vs. III), A17 in pink (experiments II vs. IV), or independently associated with plasticity in both hosts (central mauve triangle).

quality variation. The vast majority of genes contributing to partner quality variation were located on the symbiotic elements (megaplasmid pSymA or chromid pSymB; **Fig. 2**; **Supp. File S1**; see **Supp. Fig. S9** for other traits), although some loci were found on the chromosome but only for G × E and plasticity. Within host lines, 330 total genes were mapped with host line DZA (186 and 144 in experiments 1 and 3, respectively), while only 180 total genes mapped to host line A17 (104 and 76 in experiments 2 and 4, respectively; **Fig. 1A**).

At the gene-level, for experiment 1 in host line DZA, the INTERPRO terms "Transcription regulator *hth*, *lac*I" and "Lambda repressor-like, DNA-binding domain" were marginally significantly enriched (p = 0.00101 and 0.00482, respectively), as was the "Oxidative phosphorylation" KEGG pathway (p = 0.00959) (**Supp. File S5**), whereas for experiment 3 in host line DZA, the "Benzoate degradation" KEGG pathway was significantly enriched (FDR-corrected p = 0.00880), while the GO terms "3,4-dihydroxybenzoate catabolic process" and "transmembrane transport" were marginally significantly enriched (p = 0.00204 and 0.00376, respectively; **Supp. File S5**). For experiment 2 in host line A17, the GO terms "transcription, DNA-templated" were significantly enriched (FDR corrected p = 0.00143), as were the related UNIPROT keywords for "transcription regulation" (FDR corrected p = 0.00310) and "DNA-binding" (p = 0.00310) (**Supp. File S5**), whereas for experiment 4 in host line A17, the INTERPRO term "Amidohydrolase 1" was marginally significantly enriched (p = 0.0307) (**Supp. File S5**). Overall, many of the underlying molecular processes important for driving variation in partner quality appear to be environmentally-dependent, making it harder to predict which genes and pathways will be important for determining symbiotic outcomes under different conditions.

At the variant-level, for each host (DZA or A17), we used correlations of the estimated allelic effects to assess the degree to which the individual effects of rhizobium alleles on partner quality were consistent in direction and/or magnitude, or whether they depended on the environment. When variants were significantly associated with partner quality in both experiments, they tended to have particularly consistent effects for host line A17, but less so for DZA (dotted lines in **Fig. 3B,A**, respectively). In fact, on A17, all of the nearly-universal variants (**Fig. 3B** black dots) had concordant (i.e., same-sign) effects on shoot biomass between experiments, whereas the opposite was true for DZA (**Fig. 3A** black dots): all nearly-universal variants had discordant (i.e., opposite-sign) effects between experiments. Regardless of host line, however, the vast majority of variants in our studies had conditionally-neutral effects, even those with large magnitude (**Fig. 3**; **Supp. Fig. S10**), and for both hosts, we rejected the global null hypothesis that allelic effects were the same across experiments (both $p < 0.0001$; **Supp. Fig. S11**). For host line DZA, 12 particularly interesting variants had significant but opposing effects on plant biomass across the two experiments; specifically, 11 of these estimated allelic effects were positive in experiment I but negative in experiment III (bottom right quadrant; **Fig. 3B**). Such associations might point to interesting environmentally-dependent genes, though only two of these particular variants were predicted to be non-synonymous, one located in the glycan metabolism gene *kdul*I1 (SMa0214/NP_435359.2), and the other in the ABC transporter gene (SMa0799/NP_435679.2; **Supp. File S1**).

Next we mapped the among-strain differences in plasticity (**Supp. Fig. S7**) to identify specific loci contributing to G × E. At the genomic-level, most (63%) plasticity loci were found on pSymB (**Fig. 2**; **Supp. Fig. S12** for shoot biomass). Of 576 genes, 400 (69%) were also associated with shoot biomass variation *within* at least one of the four experiments when each was mapped independently – indicating abundant overlap in the genetic architecture of both within- and among-environment partner quality variation (**Supp. File S6**). Plasticity loci were marginally significantly enriched for the INTERPRO terms "Ti-type conjugative transfer relaxase *tra*A" (p = 0.0292) and "*mob*A/*mob*L protein" (p = 0.0292), as well as several INTRO-PRO terms containing "tetratricopeptide-repeat" (p range: 0.006 to 0.0310), among others (**Supp. File S5**). Of 576 total plasticity genes, only 93 (16%) were mapped in both host genotypes (**Fig. 1B**, **Fig. 2**; **Supp. File S7**). Perhaps most interesting are the 14 of these genes that were *not* associated with variation within any of the four experiments when mapped separately (**Supp. File S6**); these loci are particularly strong candidates for understanding how G x E in rhizobia scales up to alter host growth (**Supp. File S1**), and none to our knowledge have known functional roles in symbiosis.

The genetic architecture of rhizobium partner quality depends on host genotype

Given our four-experiment design, we used the conservative approach of only considering G × G genes to be those that were associated with partner quality in *both* experiments with one host genotype and *neither* of the other. We found 33 genes that contributed to partner quality but only in DZA, and 18 that contributed but only in A17 (**Fig. 1**; **Supp. File S2**). These genes split evenly across the two symbiosis plasmids of the genome (pSymA: N = 24; pSymB: N = 27). At the gene-level, in the list of A17-only G x G genes, analyses of UNIPROT keywords indicated that "Selenium" was marginally signifi-

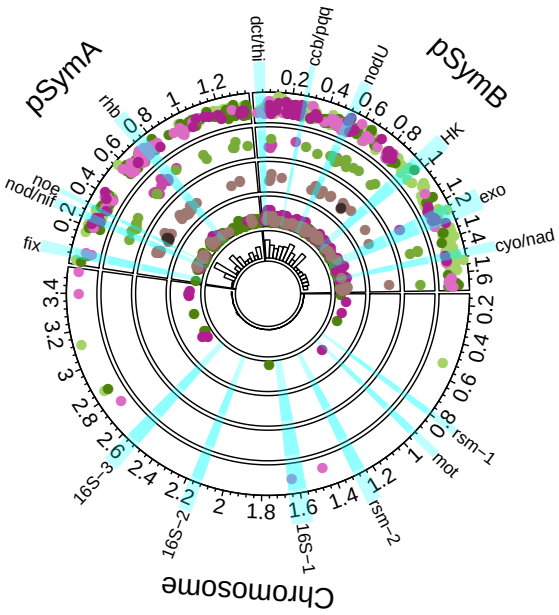


Figure 2 Loci associated with partner quality mostly limited to the symbiosis plasmids. Circos plot showing positions of genes (dots) significantly associated with shoot biomass. Each ring represents a different gene category, outermost to innermost: 1) G x E, 2) G x G, 3) partially universal and universal, 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 colors 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 specified in **Supp. Fig. S9**.

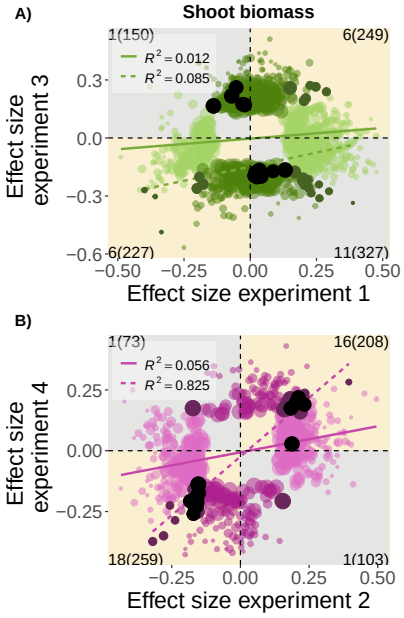


Figure 3 Allelic effects are more consistent between experiments for A17, but not for DZA. Correlations between the estimated allelic effects of individual *E. meliloti* variants on plant shoot biomass (from GWAS) in each of two experiments for either **A)** host DZA (green) or **B)** A17 (pink). Only allelic effects that were significant in one (lighter colors) 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 variants (solid line) or variants significant in both experiments (dotted line). Counts of significant variants for one or both environments appear in the corners of each quadrant within or outside the parentheses, respectively.

cantly overrepresented (p = 0.00771), while GO terms for "transferase activity, transferring glycosyl groups" were marginally overrepresented (p = 0.0417) (**Supp. File S5**). Notably, these genes include both *exo*U (SMb20948/NP_437611.1) and *exo*W (SMb21690/NP_437613.1; **Supp. File S2**), two glucosyltransferases critical for adding the 6th and 7th sugars (respectively) on the succinoglycan molecule, an exopolysaccharide required for infection thread formation and thus host plant invasion for effective symbiosis (Becker *et al.* 1993; Reuber and Walker 1993; Nogales *et al.* 2006). Non-functional *exo* genes are known to result in less-efficient, though sometimes not entirely deficient, symbiosis (Mendis *et al.* 2016). Together with this prior knowledge, our finding that variants in two succinoglycan biosynthesis enzymes affected host biomass production in host A17 suggests a strong role for succinoglycan-mediated host invasion in mediating symbiotic benefits in natural populations. Future studies of coevolution between succinoglycan structure and host plant detection would likely be fruitful. Also notable is *sel*A (SMa0011/NP_435251.1; **Supp. File S2**), a transferase required for the biosynthesis of selenocysteine, a less-used amino acid incorporated into select proteins in only about 20% of bacterial genomes (Copeland 2005; Peng *et al.* 2016; Wells and Stolz 2020). To our knowledge, nothing is known about the role of selenocysteine in rhizobia or in legume-rhizobium symbiosis.

By contrast, the list of DZA-only G × G genes was dominated by GO terms "transcription factor activity, sequence-specific DNA binding" (p = 0.0247) and "cellular amino acid metabolic process" (p = 0.0425) (**Supp. File S5**). Notably this list contains three *lys*R transcriptional regulators (of at least 90 in the genome; (**Supp. File S2**)): SMa2287 (NP_436477.2) on pSymA plus SM_b20494 (NP_437016.1) and SM_b21434 (NP_437803.1) on pSymB are not well-studied, and had weak (SM_b20494) or

undetectable effects in a plasmid insertion mutagenesis screen compared to the named *lsr*A and *lsr*B (Luo *et al.* 2005). This class of loci is well-known to control expression of genes for symbiosis, both mutualistic and pathogenic (Schell 1993). Beyond those discovered in knockdown studies, our results suggest additional symbiotic roles for natural variation at these *lys*R regulators in *E. meliloti*. Interestingly, one metabolic process gene (*gdh*A: SMa0228/NP_435368.1 on pSymA; **Supp. File S2**), part of one bacterial pathway for assimilation of ammonium into glutamate (Merrick and Edwards 1995), was identified in a comparative genomics study of five *Ensifer* (formerly *Sinorhizobium*) species as specific to *E. meliloti* (Sugawara *et al.* 2013). The role of *gdh*A in rhizobial symbiosis is not well-known, though interestingly it was found to increase ammonium assimilation when the *E. coli* copy was expressed transgenically in tobacco (Ameziane *et al.* 2000).

Universal associations highlight transport functions and secretion systems

Despite these layers of context-dependency, we did find significant main effects of strain for all phenotypes (**Table 1**). Concomitantly we resolved a number of loci that were consistently associated with shoot biomass, being mapped in either three experiments (55 nearly-universal genes) or even all four experiments (5 universal genes) (**Fig. 1**). At the genomic and gene-levels, the set of 60 universal/nearly-universal genes were split between pSymA and pSymB (38 and 22 loci, respectively; **Fig. 2** mauve dots; **Supp. File S2**) and featured marginally-enriched INTERPRO terms for "Tetratricopeptide-like helical" ($p < 0.001$) and "Tetratricopeptide repeat-containing domain" ($p = 0.00160$) as well as UNIPROT keywords "Transmembrane" ($p = 0.0190$) and "Transmembrane helix" ($p = 0.0407$), though none of these terms were significant after FDR-correction (**Supp. File S5**).

Of the five truly universal genes (mapped in all 4 experiments; **Supp. File S1**), most do not have known functions in symbiosis. This includes a *cax* gene putatively involved in calcium/proton exchange (SMa0675/NP_435603.2; **Supp. File S2**). Though *cax* genes are widespread in bacteria and eukaryotes (Shigaki *et al.* 2006), and the importance of calcium both in nodule establishment and trade of benefits is known (Oldroyd and Downie 2004; Clarke *et al.* 2014), it is difficult to hypothesize on the function of this particular gene or its genetic variants in this symbiosis at this time. Given their as-yet unknown functions and lack of context-dependency in our studies, the five universal candidates might hold the most potential for novel functional information and consistent phenotypic effects, which might make them ideal candidates both for validation and for symbiosis improvement.

Like the five universal genes, the "nearly-universal" set of genes associated with shoot biomass in three (of four) experiments highlights the existence of segregating natural variation in a number of interesting metabolic pathways, only some of which have established roles in symbiosis. We found five loci annotated as involved in transmembrane transport (**Supp. File S2**), including *kdp*A (SMa2333/NP_436501.1), *pot*E (SMa0678/NP_435605.1), and *msb*A1 (SM_b20813/NP_437093.2) – in addition to the nodulation protein *glm*S/*nod*M (SMa0878/NP_435728.1) and the universally-associated *cax* transporter (discussed above). The role of potassium transporters such as *kdp*A in osmoregulation during symbiosis is not well understood (Domínguez-Ferreras *et al.* 2009). The *pot*E locus codes for a putrescine/ornithine antiporter, while another nearby nearly-universal locus (SMa0682/NP_435607.4) is a predicted amino acid decarboxylase in the putrescine biosynthesis pathway - potentially suggesting a role for variation in putrescine metabolism (known to vary in *E. meliloti*; Epstein *et al.* 2018) in symbiotic partner quality (Epstein *et al.* 2018; Becerra-Rivera and Dunn 2019).

We interrogated the gene sets from two key studies that have associated natural variation in *E. meliloti* genome with symbiotic partner quality (see **Supp. File S2**, "overlap" column; Epstein *et al.* 2018; Batstone *et al.* 2020a). Our nearly-universal gene set contained eight loci that overlapped with the top 100 associations with A17 biomass from Epstein *et al.* (2018). 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 2018). Knockout mutants of *nod*M are known to decrease N-fixation of *E. meliloti* on alfalfa (Baev *et al.* 1992) and *Rhizobium leguminosarum* (Marie *et al.* 1992); together with Epstein *et al.* (2018), 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.* (2020a). 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 (Cao *et al.* 2018; Paço *et al.* 2019). 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 (Batstone *et al.* 2020a), making them strong candidates for a consistent role in symbiosis.

Discussion

Understanding our symbiotic world requires a genetically accurate understanding of the symbiotic extended phenotypes upon which selection acts. While evolutionary ecology has long recognized the importance of genetic variation in symbiosis while ignoring the underlying mechanisms, functional geneticists have traditionally resolved mechanisms without taking into account natural variation. Synthesis of these two perspectives has started to resolve ecologically-relevant quantitative variation at the nucleotide level (Branca *et al.* 2011; Stanton-Geddes *et al.* 2013; Friesen *et al.* 2014; Porter *et al.* 2017; Epstein *et al.* 2018; Burghardt *et al.* 2018). Here we quantify multiple symbiotic extended phenotypes in four mapping experiments using a model plant-microbe symbiosis and find that the genetic architecture of symbiotic partner quality is complex, with layers of context-dependency typical of quantitative traits governed by networks shaped by many interacting loci and environmental-dependence. We find that some loci in the microbial symbiont have strong and consistent effects on host growth across environments, contributing to the overall differences in mutualistic partner quality that have been the focus of many empirical and theoretical studies of mutualism to date (reviewed by Denison 2000; Simms and Taylor 2002; Heath and Stinchcombe 2014; Sachs *et al.* 2018; Stoy *et al.* 2020). Nonetheless, the vast majority of loci identified in our study were significantly associated with variation in partner quality in specific environments (G × E effects), or with specific host genotypes (G × G effects). We first discuss the

roles of environmentally-dependent loci, versus universal loci that are found consistently. Next we discuss the coevolutionary implications of genotype-dependence (G x G interactions), then wrap up with a call for further synthesis with metabolic network models towards systems genetics of symbiosis.

Environmental context-dependency in the G→P map of symbiosis

Ecological effects of context-dependency in mutualism have been recognized for a long time, i.e., traits and mutualism benefits shift across environmental conditions such as nutrient availability or light environments (Chamberlain *et al.* 2014; Bronstein 2015; Hoeksema and Bruna 2015). More recent studies have begun to document the evolutionary changes that can result from these ecological effects, e.g., divergence of host and/or symbiont symbiosis traits across strong ecological gradients (Porter *et al.* 2011; Weese *et al.* 2015; Klinger *et al.* 2016; Rúa *et al.* 2016; Piculell *et al.* 2019). Evolutionary change in response to environments implies that the loci underlying selected traits have differential effects on fitness across environments. Here we identify the loci that generate important trait variation both within and among environments, the trait variation upon which selection acts in nature.

The majority of context-dependent partner quality genes we identified might be viewed much like the conditionally-neutral variation so often found in studies of local adaptation (Anderson *et al.* 2011, 2013; Des Marais *et al.* 2013; Wadgymar *et al.* 2017; Mee and Yeaman 2019), contributing strongly to variation in partner quality in some contexts but having a range of weaker effects (both in the same or opposing direction) in another context. For example, by experimentally-evolving strains of *E. meliloti* on five *M. truncatula* lines, Batstone *et al.* (2020a) found that local adaptation was largely governed by conditional neutrality (beneficial on local host, neutral on non-local hosts) or mildly deleterious effects on non-local hosts, likely due to drift in the local context. Yet, in natural populations in which rhizobia population sizes are much larger and more diverse, and gene flow is present, the extent to which local adaptation occurs has rarely been tested but might be unlikely (Heath 2010; Harrison *et al.* 2017; Riley *et al.* 2021), with the exception of populations differentiated by strong ecological gradients (e.g., Porter *et al.* 2011). Moreover, the host genotypes used in our study, DZA and A17, are unlikely to share an evolutionary history with our strains, and so it remains unclear whether stronger trade-offs (and less conditional neutrality) would be present if our strains shared an evolutionary history with the host lines being tested.

Despite widespread conditional neutrality, a handful of interesting variants had strong effects on host growth, but in *opposing directions* across experiments within a single host genetic background. These sorts of antagonistic effects can favor different variants in different environments, and thus, have potential to help explain the maintenance of mutualism variation in nature (Frederickson 2013; Heath and Stinchcombe 2014; Pahua *et al.* 2018; Vaidya and Stinchcombe 2020). Nevertheless we note that these sorts of G × E variants were rare in our study, despite strong rank-order effects among strain means at the organismal level (**Table 2**); moreover, because our "environments" were simply different greenhouse experiments, relating such antagonistic effects to adaptation in the wild will require the type of *in situ* studies that are common in plants (e.g., Anderson *et al.* 2013; Rúa *et al.* 2016; Popovic and Lowry 2020) but more difficult in soil microbes.

We nevertheless found several genes and pathways that are ideal for functional follow-up studies aimed at identifying consistent associations, or loci which might benefit from "fine-tuning" symbiotic benefits towards improving plant health. The strengths of GWAS (capturing genetic variation, ecological relevance) also lead to weaknesses (imprecision due to confounding population structure, false negatives and false positives). Thus our strongest recommendations for functional hypotheses for validation in follow up studies are those emerging from comparing our consistent (universal and nearly-universal) candidate gene lists with GWAS associations from other research groups in different genomic backgrounds (e.g., Epstein *et al.* 2018; Batstone *et al.* 2020a). These efforts suggest important roles for natural variation in rhizobial glucosamine processing, T4SS transferase loci (*tra*A and *tra*G), and a handful of others (**Supp. File S2**) in determining symbiotic partner quality, and thus, host quality in nature.

Overall we find that the loci underlying quantitative variation in symbiotic extended phenotypes often (but not always) depend on the environment, and therefore that a nuanced understanding of how complex traits interact with environmental variables will be necessary for many of the lofty goals in plant microbiome and symbiosis research (Friesen *et al.* 2011; Busby *et al.* 2017). This point has been made before (Mackay *et al.* 2009), as the presence of G × E as studied at the phenotypic level has been recognized for decades; what is novel here is our ability to interrogate this variation at the genomic level and for multiple host lines (see below).

Co-evolutionary implications of genotype-dependence

Previous mapping efforts in the legume-rhizobium system have focused on the HapMap collections, which maximize diversity using a range-wide sample. Our study focuses on the segregating natural variation at a smaller geographic scale, a scale at which pSymA and pSymB segregate (Riley *et al.* 2021); thus the G × G-driven variation we find here would be available to local evolutionary and/or coevolutionary processes. G × G interactions for fitness outcomes have long been of interest in the legume-rhizobium symbiosis (Parker 1995; Burdon and Thrall 1999; Heath 2010; Porter *et al.* 2011; Barrett *et al.* 2012) and other interactions (Vale and Little 2010; Hoeksema 2010; Chong and Moran 2016) because such statistical interactions generate the fitness variation that drives coevolution, but also have implications for breeding and agricultural production because the functional effects of symbiont variation depends on the crop genotype (Wallace *et al.* 2018; Clouse and Wagner 2021). Recent molecular genetic and transcriptomic approaches on a handful of genotypes have begun to resolve the mechanistic underpinnings of G × G (Heath *et al.* 2012; Guefrachi *et al.* 2014; Burghardt *et al.* 2017; Wang *et al.* 2018; Walker *et al.* 2020; Fagorzi *et al.* 2021), while biparental or GWAS mapping approaches (Gorton *et al.* 2012; Stanton-Geddes *et al.* 2013; Epstein *et al.* 2018) provide broader insight into the genetic architecture underlying G × G (i.e., the number, average effect size, and consistency of loci). The picture emerging from ours and others' studies is that G × G, like symbiotic partner quality itself, has a complex, polygenic basis and will require both statistically sophisticated and metabolically-informed models to unravel.

Symbiotic extended phenotypes as quantitative traits

In the age of rapid microbiome sequencing and expanding efforts to characterize the loci in plant genomes that contribute

to microbiome variation among cultivars or genotypes (e.g., rice, *Lotus*, *Medicago*), our results present an important juxtaposition, as the abundant and context-dependent genetic variation characterized in detail here occurs within a set of 191 strains with 98-100% average nucleotide identity (well above the typical threshold for delineating and enumerating operational taxonomic units, or OTUs). At the same time, functional genetic studies have made much progress identifying loci critical to symbiosis establishment and downstream processes by creating knock-out or knock-down mutants and comparing their associated symbiotic phenotypes to wildtype strain (e.g., *dnf* mutants in legumes, Ané *et al.* 2004; fix+/fix- mutants of rhizobia, Wheatley *et al.* 2020), but these loci are often viewed as on-or-off switches for symbiosis more generally, or cooperation more specifically in models of mutualism theory (Bull and Rice 1991; West *et al.* 2002a,b; Foster and Kokko 2006). Our study demonstrates that the vast majority of loci within the symbiont genome act more like dials than on-and-off switches, generating the quantitative variation in symbiotic extended phenotypes observed in nature. While our approach allows us to zero-in on particular loci underlying G $\times$ E and G $\times$ G interactions, *in silico* modelling of both plant (Pfau *et al.* 2018) and rhizobium (Checcucci *et al.* 2016) metabolism that links together genetic information with metabolic pathways could then be used to simulate key symbiotic processes under a wide range of conditions (diCenzo and Finan 2018). Ultimately combining transcriptomic, genomic, and metabolomic datasets will be required for a synthetic understanding of how nucleotide variation percolates up through shared symbiotic metabolic networks (Civelek and Lusis 2014; Kliebenstein 2004; Marshall-Colón and Kliebenstein 2019). Efforts to reintegrate research on symbiosis genetics and (co)evolution with plant-microbiome work (reviewed in Wagner 2021; Batstone 2021) will be fruitful in revealing additional intraspecific variation driving patterns of genotype-dependence and coevolution and resolving mechanisms of host control of the microbiome.

Acknowledgements

We thank greenhouse staff (esp. Debbie Black). We acknowledge funding from the National Science Foundation (IOS-1645875 and NPGI-1401864), a Carl R. Woese Institute for Genomic Biology postdoctoral fellowship to R.T. Batstone, strain sequencing by Joint Genome Institute (CSP-1223795), as well as analytical advice from B. Epstein.

Authorship Statement

KDH and AMC conceived of the work, acquired funding, and designed the experiments. MAG collected soil, performed trapping experiments, and isolated strains. ABR performed preliminary experiments and extracted DNA. HL and CMA led experimental work and data collection with assistance from LAG, ABR, KDH, RTB, and AMC. RTB, HL, and KDH performed data analysis and interpretation. KDH, RTB, and HL drafted the article. All authors participated in critical revisions and approved the final version for submission.

Literature Cited

Afkhami, M. E., M. L. Friesen, and J. R. Stinchcombe, 2021 Multiple mutualism effects generate synergistic selection and strengthen fitness alignment in the interaction between legumes, rhizobia and mycorrhizal fungi. Ecology Letters .

Ameziane, R., K. Bernhard, and D. Lightfoot, 2000 Expression of the bacterial *gdh*a gene encoding a nadph glutamate dehydrogenase in tobacco affects plant growth and development. Plant and Soil **221**: 47–57.

Anderson, J. T., C.-R. Lee, C. A. Rushworth, R. I. Colautti, and T. Mitchell-Olds, 2013 Genetic trade-offs and conditional neutrality contribute to local adaptation. Molecular Ecology **22**: 699–708.

Anderson, J. T., J. H. Willis, and T. Mitchell-Olds, 2011 Evolutionary genetics of plant adaptation. Trends in Genetics **27**: 258–266.

Ané, J.-M., G. B. Kiss, B. K. Riely, R. V. Penmetsa, G. E. Oldroyd, *et al.*, 2004 *Medicago truncatula dmi*1 required for bacterial and fungal symbioses in legumes. Science **303**: 1364–1367.

Baev, N., M. Schultze, I. Barlier, D. C. Ha, H. Virelizier, *et al.*, 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. Journal of Bacteriology **174**: 7555–7565.

Barnett, M. J. and S. R. Long, 2018 Novel genes and regulators that influence production of cell surface exopolysaccharides in sinorhizobium meliloti. Journal of Bacteriology **200**: e00501–17.

Barrett, L. G., L. M. Broadhurst, and P. H. Thrall, 2012 Geographic adaptation in plant–soil mutualisms: tests using *Acacia* spp. and rhizobial bacteria. Functional Ecology **26**: 457–468.

Bates, D., D. Sarkar, M. D. Bates, and L. Matrix, 2007 The lme4 package. R Package Version **2**: 74.

Batstone, R. T., 2021 Genomes within genomes: nested symbiosis and its implications for plant evolution. New Phytologist (in review).

Batstone, R. T., L. T. Burghardt, and K. D. Heath, 2021 The genetic basis of cooperation and conflict in natural populations of a model symbiont. bioRχiv. https://www.biorxiv.org/content/10.1101/2021.07.19.452989v2.

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

Batstone, R. T., M. A. Peters, A. K. Simonsen, J. R. Stinchcombe, and M. E. Frederickson, 2020b Environmental variation impacts trait expression and selection in the legume–rhizobium symbiosis. American Journal of Botany **107**: 195–208.

Becerra-Rivera, V. A. and M. F. Dunn, 2019 Polyamine biosynthesis and biological roles in rhizobia. FEMS Microbiology Letters **366**: fnz084.

Becker, A., A. Kleickmann, H. Küster, M. Keller, W. Arnold, *et al.*, 1993 Analysis of the *Rhizobium meliloti* genes exou, exov, exow, exot, and exoi involved in exopolysaccharide biosynthesis and nodule invasion: exou and exow probably encode glucosyltransferases. Molecular Plant-Microbe Interactions: MPMI **6**: 735–744.

Branca, A., T. D. Paape, P. Zhou, R. Briskine, A. D. Farmer, *et al.*, 2011 Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. Proceedings of the National Academy of Sciences **108**: E864–E870.

Brock, D. A., K. Jones, D. C. Queller, and J. E. Strassmann, 2016 Which phenotypic traits of dictyostelium discoideum farmers are conferred by their bacterial symbionts? Symbiosis **68**: 39–48.

Bronstein, J. L., 2015 *Mutualism*. Oxford University Press, USA.

Brundrett, M. C., 2002 Coevolution of roots and mycorrhizas of

land plants. New phytologist **154**: 275–304.
Bublitz, D. C., G. L. Chadwick, J. S. Magyar, K. M. Sandoz, D. M. Brooks, *et al.*, 2019 Peptidoglycan production by an insect-bacterial mosaic. Cell **179**: 703–712.
Bull, J. J. and W. Rice, 1991 Distinguishing mechanisms for the evolution of co-operation. Journal of Theoretical Biology **149**: 63–74.
Burdon, J. and P. Thrall, 1999 Spatial and temporal patterns in coevolving plant and pathogen associations. The American Naturalist **153**: S15–S33.
Burghardt, L. T., 2020 Evolving together, evolving apart: measuring the fitness of rhizobial bacteria in and out of symbiosis with leguminous plants. New Phytologist **228**: 28–34.
Burghardt, L. T., B. Epstein, J. Guhlin, M. S. Nelson, M. R. Taylor, *et al.*, 2018 Select and resequence reveals relative fitness of bacteria in symbiotic and free-living environments. Proceedings of the National Academy of Sciences **115**: 2425–2430.
Burghardt, L. T., B. Epstein, and P. Tiffin, 2019 Legacy of prior host and soil selection on rhizobial fitness in planta. Evolution **73**: 2013–2023.
Burghardt, L. T., J. Guhlin, C. L. Chun, J. Liu, M. J. Sadowsky, *et al.*, 2017 Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. Molecular Ecology **26**: 6122–6135.
Busby, P. E., C. Soman, M. R. Wagner, M. L. Friesen, J. Kremer, *et al.*, 2017 Research priorities for harnessing plant microbiomes in sustainable agriculture. PLoS Biology **15**: e2001793.
Campbell-Staton, S. C., Z. A. Cheviron, N. Rochette, J. Catchen, J. B. Losos, *et al.*, 2017 Winter storms drive rapid phenotypic, regulatory, and genomic shifts in the green anole lizard. Science **357**: 495–498.
Cao, Y., S. S. Miller, M. R. Dornbusch, S. S. Castle, P. Lenz, *et al.*, 2018 Widespread occurrence of sinorhizobium meliloti strains with a type iv secretion system. Symbiosis **75**: 81–91.
Caspi, R., R. Billington, I. M. Keseler, A. Kothari, M. Krummenacker, *et al.*, 2020 The MetaCyc database of metabolic pathways and enzymes- a 2019 update. Nucleic Acids Research **48**: D445–D453.
Cavassim, M. I. A., S. Moeskjær, C. Moslemi, B. Fields, A. Bachmann, *et al.*, 2020 Symbiosis genes show a unique pattern of introgression and selection within a rhizobium leguminosarum species complex. Microbial Genomics **6**.
Chamberlain, S. A., J. L. Bronstein, and J. A. Rudgers, 2014 How context dependent are species interactions? Ecology Letters **17**: 881–890.
Checcucci, A., M. Bazzicalupo, A. Mengoni, C. Viti, L. Dziewit, *et al.*, 2016 Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nature Communications **7**: 1–10.
Checcucci, A., G. C. diCenzo, V. Ghini, M. Bazzicalupo, A. Becker, *et al.*, 2018 Creation and characterization of a genomically hybrid strain in the nitrogen-fixing symbiotic bacterium *Sinorhizobium meliloti*. ACS Synthetic Biology **7**: 2365–2378.
Cheverud, J. M., 1982 Phenotypic, genetic, and environmental morphological integration in the cranium. Evolution pp. 499–516.
Cheverud, J. M., 1996 Developmental integration and the evolution of pleiotropy. American Zoologist **36**: 44–50.
Chong, R. A. and N. A. Moran, 2016 Intraspecific genetic variation in hosts affects regulation of obligate heritable symbionts. Proceedings of the National Academy of Sciences **113**: 13114–13119.
Civelek, M. and A. J. Lusis, 2014 Systems genetics approaches to understand complex traits. Nature Reviews Genetics **15**: 34–48.
Clarke, V. C., P. C. Loughlin, D. A. Day, and P. Smith, 2014 Transport processes of the legume symbiosome membrane. Frontiers in Plant Science **5**: 699.
Clouse, K. M. and M. R. Wagner, 2021 Plant genetics as a tool for manipulating crop microbiomes: Opportunities and challenges. Frontiers in Bioengineering and Biotechnology **9**.
Cockerham, C. C., 1963 Estimation of genetic variances. Statistical genetics and plant breeding **982**: 53–94.
Cook, D. R., 1999 *Medicago truncatula*–a model in the making! Current Opinion in Plant Biology **2**: 301–304.
Copeland, P. R., 2005 Making sense of nonsense: the evolution of selenocysteine usage in proteins. Genome Biology **6**: 1–5.
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, *et al.*, 2011 The variant call format and vcftools. Bioinformatics **27**: 2156–2158.
Denison, R. F., 2000 Legume sanctions and the evolution of symbiotic cooperation by rhizobia. The American Naturalist **156**: 567–576.
Des Marais, D. L., K. M. Hernandez, and T. E. Juenger, 2013 Genotype-by-environment interaction and plasticity: exploring genomic responses of plants to the abiotic environment. Annual Review of Ecology, Evolution, and Systematics **44**: 5–29.
Dicenzo, G. C. and T. M. Finan, 2017 The divided bacterial genome: structure, function, and evolution. Microbiology and Molecular Biology Reviews **81**: e00019–17.
diCenzo, G. C. and T. M. Finan, 2018 Techniques for large-scale bacterial genome manipulation and characterization of the mutants with respect to in silico metabolic reconstructions. In *Metabolic Network Reconstruction and Modeling*, pp. 291–314, Springer.
diCenzo, G. C., M. Tesi, T. Pfau, A. Mengoni, and M. Fondi, 2020 Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. Nature Communications **11**: 1–11.
diCenzo, G. C., M. Zamani, A. Checcucci, M. Fondi, J. S. Griffitts, *et al.*, 2019 Multidisciplinary approaches for studying rhizobium–legume symbioses. Canadian Journal of Microbiology **65**: 1–33.
Didelot, X. and M. C. Maiden, 2010 Impact of recombination on bacterial evolution. Trends in Microbiology **18**: 315–322.
Domínguez-Ferreras, A., S. Muñoz, J. Olivares, M. J. Soto, and J. Sanjuán, 2009 Role of potassium uptake systems in *Sinorhizobium meliloti* osmoadaptation and symbiotic performance. Journal of Bacteriology **191**: 2133–2143.
Douglas, A., 1994 Symbiosis as a source of novel metabolic capabilities. Symbiotic Interactions Oxford: Oxford UP pp. 12–40.
Duchene, O., J.-F. Vian, and F. Celette, 2017 Intercropping with legume for agroecological cropping systems: Complementarity and facilitation processes and the importance of soil microorganisms. a review. Agriculture, Ecosystems & Environment **240**: 148–161.
Epstein, B., R. A. Abou-Shanab, A. Shamseldin, M. R. Taylor, J. Guhlin, *et al.*, 2018 Genome-wide association analyses in the model rhizobium *Ensifer meliloti*. MSphere **3**.
Epstein, B., A. Branca, J. Mudge, A. K. Bharti, R. Briskine, *et al.*, 2012 Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. PLoS Genetics **8**:

1–10.
Epstein, B. and P. Tiffin, 2021 Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. Proceedings of the Royal Society B **288**: 20201804.
Fagorzi, C., G. Bacci, R. Huang, L. Cangioli, A. Checcucci, *et al.*, 2021 Nonadditive transcriptomic signatures of genotype-by-genotype interactions during the initiation of plant-rhizobium symbiosis. Msystems **6**: e00974–20.
Foster, K. R. and H. Kokko, 2006 Cheating can stabilize cooperation in mutualisms. Proceedings of the Royal Society B **273**: 2233–2239.
Foyer, C. H., H.-M. Lam, H. T. Nguyen, K. H. Siddique, R. K. Varshney, *et al.*, 2016 Neglecting legumes has compromised human health and sustainable food production. Nature Plants **2**: 1–10.
Frank, S., 1992 Models of plant-pathogen coevolution. Trends in Genetics **8**: 213–219.
Frederickson, M. E., 2013 Rethinking mutualism stability: cheaters and the evolution of sanctions. The Quarterly Review of Biology **88**: 269–295.
Friesen, M. L., S. S. Porter, S. C. Stark, E. J. von Wettberg, J. L. Sachs, *et al.*, 2011 Microbially mediated plant functional traits. Annual Review of Ecology, Evolution, and Systematics **42**: 23–46.
Friesen, M. L., E. J. von Wettberg, M. Badri, K. S. Moriuchi, F. Barhoumi, *et al.*, 2014 The ecological genomic basis of salinity adaptation in tunisian *Medicago truncatula*. BMC Genomics **15**: 1160.
Fustec, J., F. Lesuffleur, S. Mahieu, and J.-B. Cliquet, 2010 Nitrogen rhizodeposition of legumes. a review. Agronomy for Sustainable Development **30**: 57–66.
Gilbert, J. A. and S. V. Lynch, 2019 Community ecology as a framework for human microbiome research. Nature Medicine **25**: 884–889.
González-Torres, P., F. Rodríguez-Mateos, J. Antón, and T. Gabaldón, 2019 Impact of homologous recombination on the evolution of prokaryotic core genomes. MBio **10**: e02494–18.
Gorton, A. J., K. D. Heath, M.-L. Pilet-Nayel, A. Baranger, and J. R. Stinchcombe, 2012 Mapping the genetic basis of symbiotic variation in legume-rhizobium interactions in *Medicago truncatula*. G3: Genes | Genomes | Genetics **2**: 1291–1303.
Graham, P. H. and C. P. Vance, 2003 Legumes: importance and constraints to greater use. Plant Physiology **131**: 872–877.
Grillo, M. A., S. De Mita, P. V. Burke, K. L. Solórzano-Lowell, and K. D. Heath, 2016 Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in *Medicago truncatula*. Evolution **70**: 2704–2717.
Guefrachi, I., M. Nagymihaly, C. I. Pislariu, W. Van de Velde, P. Ratet, *et al.*, 2014 Extreme specificity of NCR gene expression in *Medicago truncatula*. BMC Genomics **15**: 1–16.
Hansen, T. F., 2006 The evolution of genetic architecture. Annu. Rev. Ecol. Evol. Syst. **37**: 123–157.
Hansen, T. F., 2013 Why epistasis is important for selection and adaptation. Evolution **67**: 3501–3511.
Harrison, E. and M. A. Brockhurst, 2012 Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends in Microbiology **20**: 262–267.
Harrison, P. W., R. P. Lower, N. K. Kim, and J. P. W. Young, 2010 Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends in Microbiology **18**: 141–148.
Harrison, T. L., C. W. Wood, I. L. Borges, and J. R. Stinchcombe, 2017 No evidence for adaptation to local rhizobial mutualists in the legume *Medicago lupulina*. Ecology and Evolution **7**: 4367–4376.
Hawkes, C. V., J. J. Bull, and J. A. Lau, 2020 Symbiosis and stress: how plant microbiomes affect host evolution. Philosophical Transactions of the Royal Society B **375**: 20190590.
Heath, K. D., 2010 Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. Evolution **64**: 1446–1458.
Heath, K. D., P. V. Burke, and J. R. Stinchcombe, 2012 Coevolutionary genetic variation in the legume-rhizobium transcriptome. Molecular Ecology **21**: 4735–4747.
Heath, K. D. and M. A. Grillo, 2016 Rhizobia: tractable models for bacterial evolutionary ecology: Models for bacterial evolutionary ecology. Environmental Microbiology **18**.
Heath, K. D., J. C. Podowski, S. Heniff, C. R. Klinger, P. V. Burke, *et al.*, 2020 Light availability and rhizobium variation interactively mediate the outcomes of legume–rhizobium symbiosis. American Journal of Botany **107**: 229–238.
Heath, K. D. and J. R. Stinchcombe, 2014 Explaining mutualism variation: a new evolutionary paradox? Evolution **68**: 309–317.
Heath, K. D., A. Stock, and J. Stinchcombe, 2010 Mutualism variation in the nodulation response to nitrate. Journal of Evolutionary Biology **23**: 2494–2500.
Hoeksema, J. D., 2010 Ongoing coevolution in mycorrhizal interactions. New Phytologist **187**: 286–300.
Hoeksema, J. D. and E. Bruna, 2015 Context-Dependent Outcomes of Mutualistic Interactions. In *Mutualism*, edited by J. L. Bronstein, pp. 181–202, Oxford University Press, Oxford, UK.
Huang, D. W., B. T. Sherman, Q. Tan, J. Kir, D. Liu, *et al.*, 2007 David bioinformatics resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. Nucleic Acids Research **35**: W169–W175.
Hurst, G. D., 2017 Extended genomes: symbiosis and evolution. Interface Focus **7**: 20170001.
Karp, P. D., R. Billington, R. Caspi, C. A. Fulcher, M. Latendresse, *et al.*, 2019 The BioCyc collection of microbial genomes and metabolic pathways. Briefings in Bioinformatics **20**: 1085–1093.
Kliebenstein, D., 2004 Secondary metabolites and plant/environment interactions: a view through *Arabidopsis thaliana* tinged glasses. Plant, Cell & Environment **27**: 675–684.
Klinger, C. R., J. A. Lau, and K. D. Heath, 2016 Ecological genomics of mutualism decline in nitrogen-fixing bacteria. Proceedings of the Royal Society B **283**: 20152563.
Klironomos, J. N., 2003 Variation in plant response to native and exotic arbuscular mycorrhizal fungi. Ecology **84**: 2292–2301.
Korte, A. and A. Farlow, 2013 The advantages and limitations of trait analysis with GWAS: a review. Plant Methods **9**: 1–9.
Lambrechts, L., S. Fellous, and J. C. Koella, 2006 Coevolutionary interactions between host and parasite genotypes. Trends in Parasitology **22**: 12–16.
Lasky, J. R., B. R. Forester, and M. Reimherr, 2018 Coherent synthesis of genomic associations with phenotypes and home environments. Molecular Ecology Resources **18**: 91–106.
Lau, J. A., E. J. Bowling, L. E. Gentry, P. A. Glasser, E. A. Monarch, *et al.*, 2012 Direct and interactive effects of light and nutrients on the legume-rhizobia mutualism. Acta Oecologica **39**: 80–86.
Lee, S. H., S. Ripke, B. M. Neale, S. V. Faraone, S. M. Purcell, *et al.*, 2013 Genetic relationship between five psychiatric disorders

estimated from genome-wide snps. Nature Genetics **45**: 984–995.

Lorts, C. M. and J. R. Lasky, 2020 Competition× drought interactions change phenotypic plasticity and the direction of selection on arabidopsis traits. New Phytologist **227**: 1060–1072.

Luo, L., S.-Y. Yao, A. Becker, S. Rüberg, G.-Q. Yu, *et al.*, 2005 Two new *Sinorhizobium meliloti Lys*r-type transcriptional regulators required for nodulation. Journal of Bacteriology **187**: 4562–4572.

Mackay, T. F. and W. Huang, 2018 Charting the genotype–phenotype map: lessons from the drosophila melanogaster genetic reference panel. Wiley Interdisciplinary Reviews: Developmental Biology **7**: e289.

Mackay, T. F., E. A. Stone, and J. F. Ayroles, 2009 The genetics of quantitative traits: challenges and prospects. Nature Reviews Genetics **10**: 565–577.

Marie, C., M.-A. Barny, and J. Downie, 1992 *Rhizobium leguminosarum* has two glucosamine syntheses, *gim*s and *nod*m, required for nodulation and development of nitrogen-fixing nodules. Molecular Microbiology **6**: 843–851.

Marshall-Colón, A. and D. J. Kliebenstein, 2019 Plant networks as traits and hypotheses: moving beyond description. Trends in Plant Science **24**: 840–852.

McGuigan, K., 2006 Studying phenotypic evolution using multivariate quantitative genetics. Molecular Ecology **15**: 883–896.

Mee, J. A. and S. Yeaman, 2019 Unpacking conditional neutrality: genomic signatures of selection on conditionally beneficial and conditionally deleterious mutations. The American Naturalist **194**: 529–540.

Mendis, H. C., T. F. Madzima, C. Queiroux, and K. M. Jones, 2016 Function of succinoglycan polysaccharide in *Sinorhizobium meliloti* host plant invasion depends on succinylation, not molecular weight. MBio **7**: e00606–16.

Merrick, M. and R. Edwards, 1995 Nitrogen control in bacteria. Microbiological Reviews **59**: 604–622.

Misra, H. S., G. K. Maurya, S. Kota, and V. K. Charaka, 2018 Maintenance of multipartite genome system and its functional significance in bacteria. Journal of Genetics **97**: 1013–1038.

Moeller, A. H. and J. G. Sanders, 2020 Roles of the gut microbiota in the adaptive evolution of mammalian species. Philosophical Transactions of the Royal Society B **375**: 20190597.

Muir, W., W. Nyquist, and S. Xu, 1992 Alternative partitioning of the genotype-by-environment interaction. Theoretical and Applied Genetics **84**: 193–200.

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

Nogales, J., S. Muñoz, J. Olivares, and J. Sanjuán, 2006 *Sinorhizobium meliloti* genes involved in tolerance to the antimicrobial peptide protamine. FEMS Microbiology Letters **264**: 160–167.

Oldroyd, G. E. and J. A. Downie, 2004 Calcium, kinases and nodulation signalling in legumes. Nature Reviews Molecular Cell Biology **5**: 566–576.

Oliver, K. M., N. A. Moran, and M. S. Hunter, 2005 Variation in resistance to parasitism in aphids is due to symbionts not host genotype. Proceedings of the National Academy of Sciences **102**: 12795–12800.

Orgogozo, V., B. Morizot, and A. Martin, 2015 The differential view of genotype–phenotype relationships. Frontiers in Genetics **6**: 179.

Otero-Bravo, A., S. Goffredi, and Z. L. Sabree, 2018 Cladogenesis and genomic streamlining in extracellular endosymbionts of tropical stink bugs. Genome Biology and Evolution **10**: 680–693.

O'Brien, A. M., C. N. Jack, M. L. Friesen, and M. E. Frederickson, 2021 Whose trait is it anyways? Coevolution of joint phenotypes and genetic architecture in mutualisms. Proceedings of the Royal Society B **288**: 20202483.

Paço, A., J. da Silva, F. Eliziário, C. Brígido, S. Oliveira, *et al.*, 2019 *tra*G gene is conserved across *Mesorhizobium* spp. able to nodulate the same host plant and expressed in response to root exudates. BioMed Research International **2019**.

Pahua, V., P. Stokes, A. Hollowell, J. Regus, K. Gano-Cohen, *et al.*, 2018 Fitness variation among host species and the paradox of ineffective rhizobia. Journal of Evolutionary Biology **31**: 599–610.

Parker, B. J., J. Hrček, A. H. McLean, and H. C. J. Godfray, 2017 Genotype specificity among hosts, pathogens, and beneficial microbes influences the strength of symbiont-mediated protection. Evolution **71**: 1222–1231.

Parker, M. P., 1995 Plant fitness variation caused by different mutualist genotypes. Ecology **76**: 1525–1535.

Peng, T., J. Lin, Y.-Z. Xu, and Y. Zhang, 2016 Comparative genomics reveals new evolutionary and ecological patterns of selenium utilization in bacteria. The ISME Journal **10**: 2048–2059.

Pfau, T., N. Christian, S. K. Masakapalli, L. J. Sweetlove, M. G. Poolman, *et al.*, 2018 The intertwined metabolism during symbiotic nitrogen fixation elucidated by metabolic modelling. Scientific Reports **8**: 1–11.

Piculell, B. J., J. D. Hoeksema, and J. N. Thompson, 2008 Interactions of biotic and abiotic environmental factors in an ectomycorrhizal symbiosis, and the potential for selection mosaics. BMC Biology **6**: 1–11.

Piculell, B. J., P. José Martínez-García, C. D. Nelson, and J. D. Hoeksema, 2019 Association mapping of ectomycorrhizal traits in loblolly pine (pinus taeda l.). Molecular Ecology **28**: 2088–2099.

Popovic, D. and D. B. Lowry, 2020 Contrasting environmental factors drive local adaptation at opposite ends of an environmental gradient in the yellow monkeyflower (mimulus guttatus). American Journal of Botany **107**: 298–307.

Porter, S. S., P. L. Chang, C. A. Conow, J. P. Dunham, and M. L. Friesen, 2017 Association mapping reveals novel serpentine adaptation gene clusters in a population of symbiotic mesorhizobium. The ISME Journal **11**: 248–262.

Porter, S. S., J. Faber-Hammond, A. P. Montoya, M. L. Friesen, and C. Sackos, 2019 Dynamic genomic architecture of mutualistic cooperation in a wild population of mesorhizobium. The ISME Journal **13**: 301–315.

Porter, S. S. and J. L. Sachs, 2020 Agriculture and the disruption of plant–microbial symbiosis. Trends in Ecology & Evolution **35**: 426–439.

Porter, S. S. and E. L. Simms, 2014 Selection for cheating across disparate environments in the legume-rhizobium mutualism. Ecology Letters **17**: 1121–1129.

Porter, S. S., M. L. Stanton, and K. J. Rice, 2011 Mutualism and adaptive divergence: co-invasion of a heterogeneous grassland by an exotic legume-rhizobium symbiosis. PloS One **6**: e27935.

Power, R. A., J. Parkhill, and T. de Oliveira, 2017 Microbial genome-wide association studies: lessons from human gwas.

Nature Reviews Genetics **18**: 41–50.

Queller, D. C. and J. E. Strassmann, 2018 Evolutionary conflict. Annual Review of Ecology, Evolution, and Systematics **49**: 73–93.

Quinlan, A. R. and I. M. Hall, 2010 Bedtools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26**: 841–842.

Reuber, T. L. and G. C. Walker, 1993 Biosynthesis of succinoglycan, a symbiotically important exopolysaccharide of *Rhizobium meliloti*. Cell **74**: 269–280.

Ridaura, V. K., J. J. Faith, F. E. Rey, J. Cheng, A. E. Duncan, *et al.*, 2013 Gut microbiota from twins discordant for obesity modulate metabolism in mice. Science **341**.

Riley, A., M. Grillo, B. Epstein, P. Tiffin, and K. Heath, 2021 Partners in space: Discordant population structure between legume hosts and rhizobium symbionts in their native range. bioRχiv. https://www.biorxiv.org/content/10.1101/2021.06.30.449460v1.

Rockman, M. V., 2012 The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution: International Journal of Organic Evolution **66**: 1–17.

Roy, S., W. Liu, R. S. Nandety, A. Crook, K. S. Mysore, *et al.*, 2020 Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. The Plant Cell **32**: 15–41.

Rúa, M. A., A. Antoninka, P. M. Antunes, V. B. Chaudhary, C. Gehring, *et al.*, 2016 Home-field advantage? evidence of local adaptation among plants, soil, and arbuscular mycorrhizal fungi through meta-analysis. BMC Evolutionary Biology **16**: 1–15.

Sachs, J. L., K. W. Quides, and C. E. Wendlandt, 2018 Legumes versus rhizobia: a model for ongoing conflict in symbiosis. New Phytologist **219**: 1199–1206.

Saltz, J. B., F. C. Hessel, and M. W. Kelly, 2017 Trait correlations in the genomics era. Trends in Ecology & Evolution **32**: 279–290.

San Millan, A., K. Heilbron, and R. C. MacLean, 2014 Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. The ISME Journal **8**: 601–612.

Schell, M. A., 1993 Molecular biology of the *Lys*r family of transcriptional regulators. Annual Review of Microbiology **47**: 597–626.

Searle, S. R., F. M. Speed, and G. A. Milliken, 1980 Population marginal means in the linear model: an alternative to least squares means. The American Statistician **34**: 216–221.

Sherman, B. T., R. A. Lempicki, *et al.*, 2009 Systematic and integrative analysis of large gene lists using david bioinformatics resources. Nature Protocols **4**: 44–57.

Shigaki, T., I. Rees, L. Nakhleh, and K. Hirschi, 2006 Identification of three distinct phylogenetic groups of CAX cation/proton antiporters. Journal of Molecular Evolution **63**: 815–825.

Silva, R. F., S. C. Mendonça, L. M. Carvalho, A. M. Reis, I. Gordo, *et al.*, 2011 Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. PLoS Genetics **7**: e1002181.

Simms, E. L. and D. L. Taylor, 2002 Partner choice in nitrogen-fixation mutualisms of legumes and rhizobia. Integrative and Comparative Biology **42**: 369–380.

Stanton-Geddes, J., T. Paape, B. Epstein, R. Briskine, J. Yoder, *et al.*, 2013 Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. PloS One **8**: e65688.

Stoy, K. S., A. K. Gibson, N. M. Gerardo, and L. T. Morran, 2020 A need to consider the evolutionary genetics of host–symbiont mutualisms. Journal of Evolutionary Biology **33**: 1656–1668.

Sugawara, M., B. Epstein, B. D. Badgley, T. Unno, L. Xu, *et al.*, 2013 Comparative genomics of the core and accessory genomes of 48 *Sinorhizobium* strains comprising five genospecies. Genome Biology **14**: 1–20.

Team, R. C., 2016 R foundation for statistical computing. R: a language and environment for statistical computing .

Thairu, M. W. and A. K. Hansen, 2019 It's a small, small world: unravelling the role and evolution of small rnas in organelle and endosymbiont genomes. FEMS Microbiology Letters **366**: fnz049.

Thompson, J. N., 2005 *The geographic mosaic of coevolution*. University of Chicago Press.

Thompson, J. N., 2009 *The coevolutionary process*. University of Chicago press.

Thrall, P. H., J. Burdon, and M. J. Woods, 2000 Variation in the effectiveness of symbiotic associations between native rhizobia and temperate australian legumes: interactions within and between genera. Journal of Applied Ecology **37**: 52–65.

Tibbs Cortes, L., Z. Zhang, and J. Yu, 2021 Status and prospects of genome-wide association studies in plants. The Plant Genome **14**: e20077.

Tirichine, L., F. de Billy, and T. Huguet, 2000 Mtsym6, a gene conditioning *Sinorhizobium* strain-specific nitrogen fixation in *Medicago truncatula*. Plant Physiology **123**: 845–852.

Vaidya, P. and J. R. Stinchcombe, 2020 The potential for genotype-by-environment interactions to maintain genetic variation in a model legume–rhizobia mutualism. Plant Communications **1**: 100114.

Vale, P. F. and T. J. Little, 2010 Crispr-mediated phage resistance and the ghost of coevolution past. Proceedings of the Royal Society B **277**: 2097–2103.

Wade, M. J., 2007 The co-evolutionary genetics of ecological communities. Nature Reviews Genetics **8**: 185–195.

Wadgymar, S. M., D. B. Lowry, B. A. Gould, C. N. Byron, R. M. Mactavish, *et al.*, 2017 Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. Methods in Ecology and Evolution **8**: 738–749.

Wagner, M. R., 2021 Prioritizing host phenotype to understand microbiome heritability in plants. New Phytologist (Accepted article).

Walker, L., B. Lagunas, and M. L. Gifford, 2020 Determinants of host range specificity in legume-rhizobia symbiosis. Frontiers in Microbiology **11**: 3028.

Wallace, J. G., E. Rodgers-Melnick, and E. S. Buckler, 2018 On the road to breeding 4.0: unraveling the good, the bad, and the boring of crop quantitative genomics. Annual Review of Genetics **52**: 421–444.

Wang, D., S. Yang, F. Tang, and H. Zhu, 2012 Symbiosis specificity in the legume–rhizobial mutualism. Cellular Microbiology **14**: 334–342.

Wang, Q., J. Liu, and H. Zhu, 2018 Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. Frontiers in Plant Science **9**: 313.

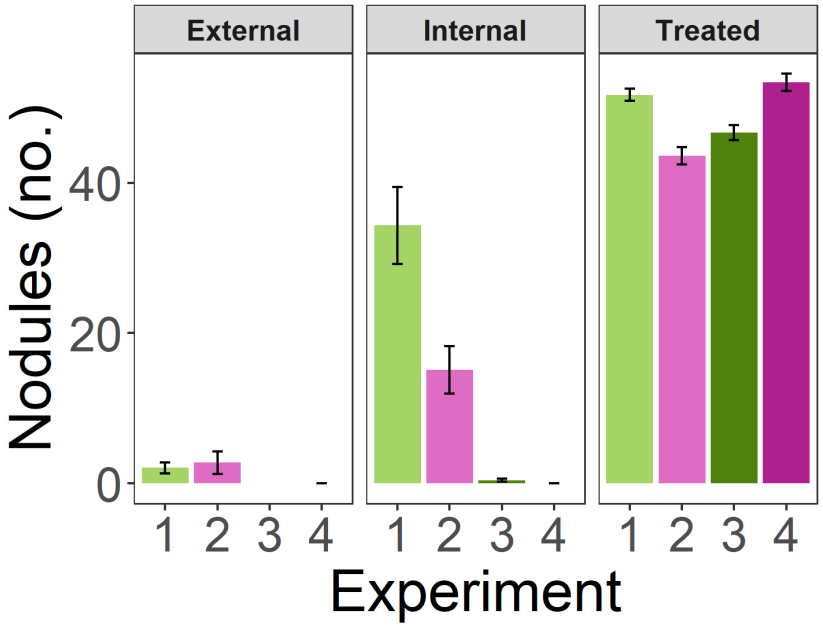
Wang, Q., S. Yang, J. Liu, K. Terecskei, E. Ábrahám, *et al.*, 2017 Host-secreted antimicrobial peptide enforces symbiotic selectivity in *Medicago truncatula*. Proceedings of the National

Academy of Sciences **114**: 6854–6859.
Weese, D. J., K. D. Heath, B. T. Dentinger, and J. A. Lau, 2015 Long-term nitrogen addition causes the evolution of less-cooperative mutualists. Evolution **69**: 631–642.
Wells, M. and J. F. Stolz, 2020 Microbial selenium metabolism: a brief history, biogeochemistry and ecophysiology. FEMS Microbiology Ecology **96**: fiaa209.
Wendlandt, C. E., J. U. Regus, K. A. Gano-Cohen, A. C. Hollowell, K. W. Quides, *et al.*, 2019 Host investment into symbiosis varies among genotypes of the legume acmispon strigosus, but host sanctions are uniform. New Phytologist **221**: 446–458.
Wendling, C. C., A. G. Fabritzek, and K. M. Wegner, 2017 Population-specific genotype x genotype x environment interactions in bacterial disease of early life stages of pacific oyster larvae. Evolutionary Applications **10**: 338–347.
West, S., E. T. Kiers, I. Pen, and R. Denison, 2002a Sanctions and mutualism stability: when should less beneficial mutualists be tolerated? Journal of Evolutionary Biology **15**: 830–837.
West, S. A., E. T. Kiers, E. L. Simms, and R. F. Denison, 2002b Sanctions and mutualism stability: why do rhizobia fix nitrogen? Proceedings of the Royal Society B **269**: 685–694.
Wheatley, R. M., B. L. Ford, L. Li, S. T. Aroney, H. E. Knights, *et al.*, 2020 Lifestyle adaptations of rhizobium from rhizosphere to symbiosis. Proceedings of the National Academy of Sciences **117**: 23823–23834.
Yang, S., Q. Wang, E. Fedorova, J. Liu, Q. Qin, *et al.*, 2017 Microsymbiont discrimination mediated by a host-secreted peptide in *Medicago truncatula*. Proceedings of the National Academy of Sciences **114**: 6848–6853.
Young, N. D., S. B. Cannon, S. Sato, D. Kim, D. R. Cook, *et al.*, 2005 Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiology **137**: 1174–1181.
Young, N. D., P. Zhou, and K. A. Silverstein, 2016 Exploring structural variants in environmentally sensitive gene families. Current Opinion in Plant Biology **30**: 19–24.
Zhou, X. and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods **11**: 407–409.

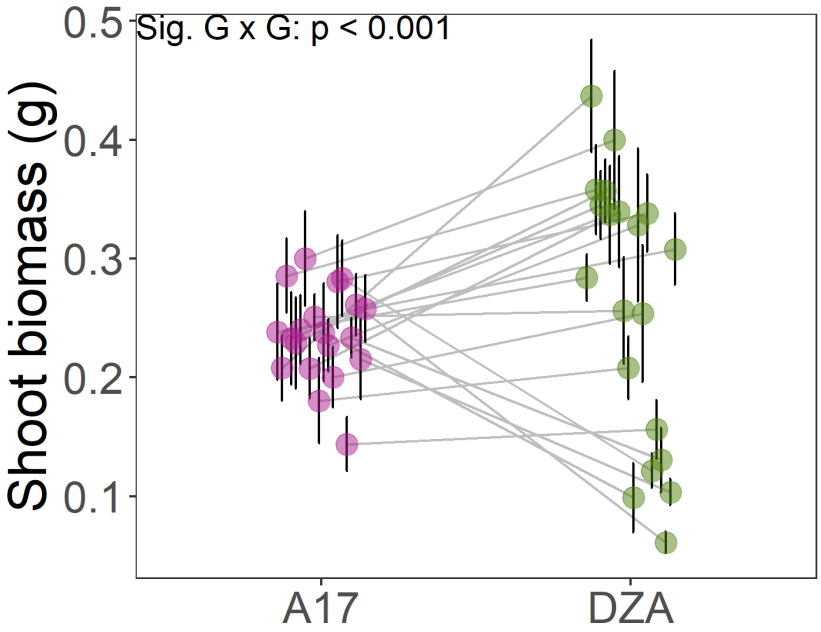
Supplementary Materials for Batstone *et al.*, The complex genetics of symbiotic extended phenotypes

Figs. S1 to S12
Files S1 to S7

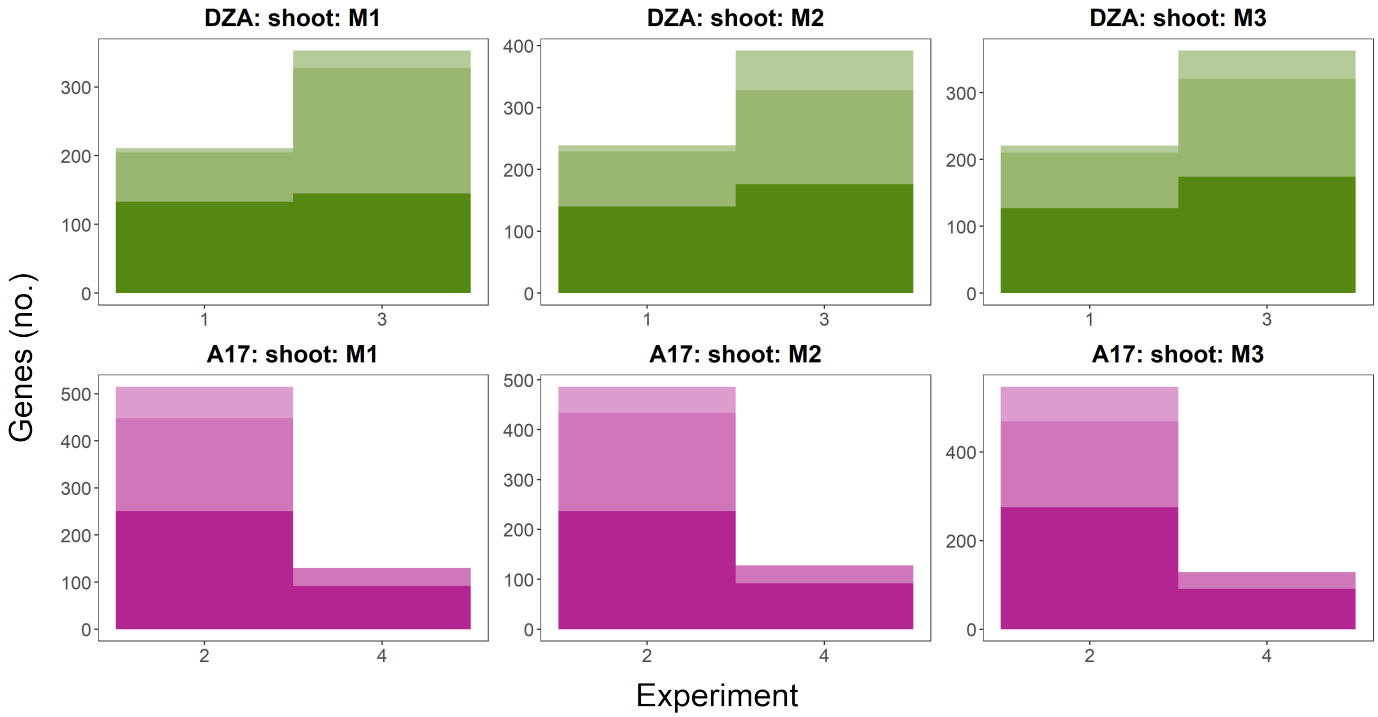
Supp. Figures



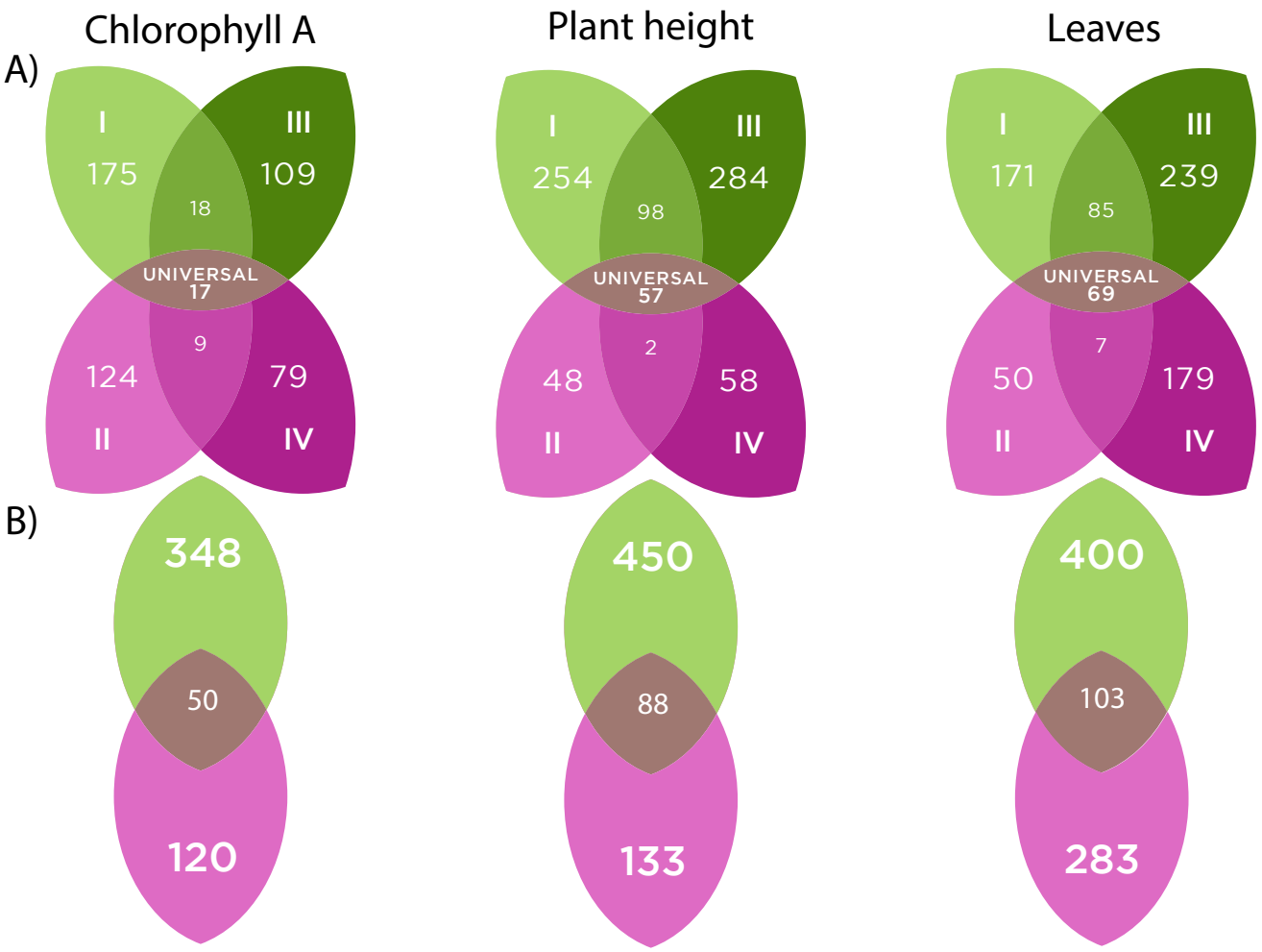
Supp. Figure S1 Contamination was minimal across experiments. Nodule number for inoculated versus internal and external control plants in four experiments: Experiments 1 and 3 used host genotype DZA while 2 and 4 used genotype A17.



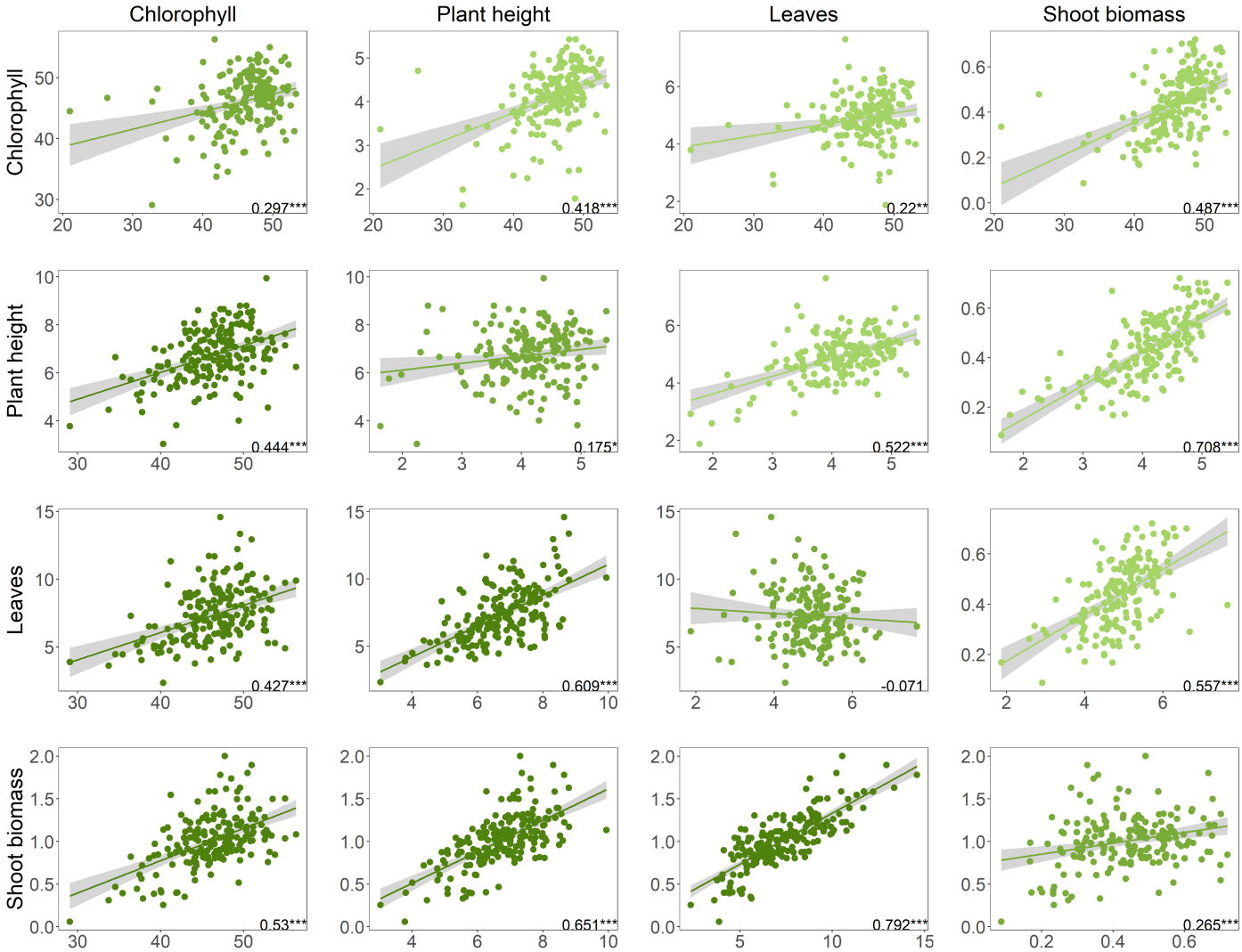
Supp. Figure S2 Significant Genotype-by-genotype (G $\times$ G) interactions between two host genotypes (A17 and DZA) and 20 *E. meliloti* strains. All 20 strains were included in the 191 strains used in current paper. Both hosts were grown together in a single experiment (Heath et al., unpublished data). 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.



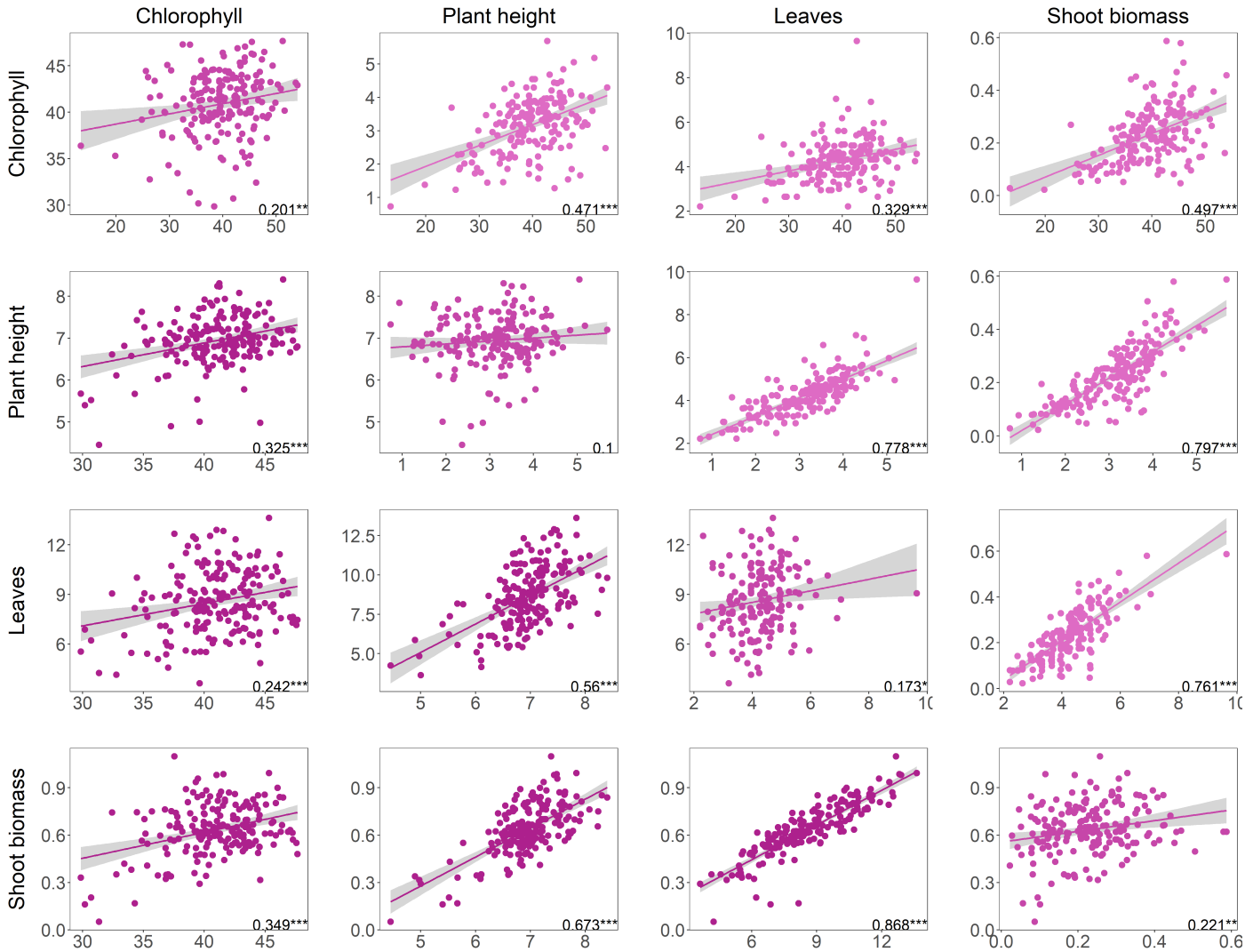
Supp. Figure S3 Minimal differences in the number of significant genes across three different methods. To ensure the variants identified in our study did not depend on the computational methods used (i.e., how the k-matrix was computed, how variants were selected for LD), we compared the number of genes 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 purple = 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).



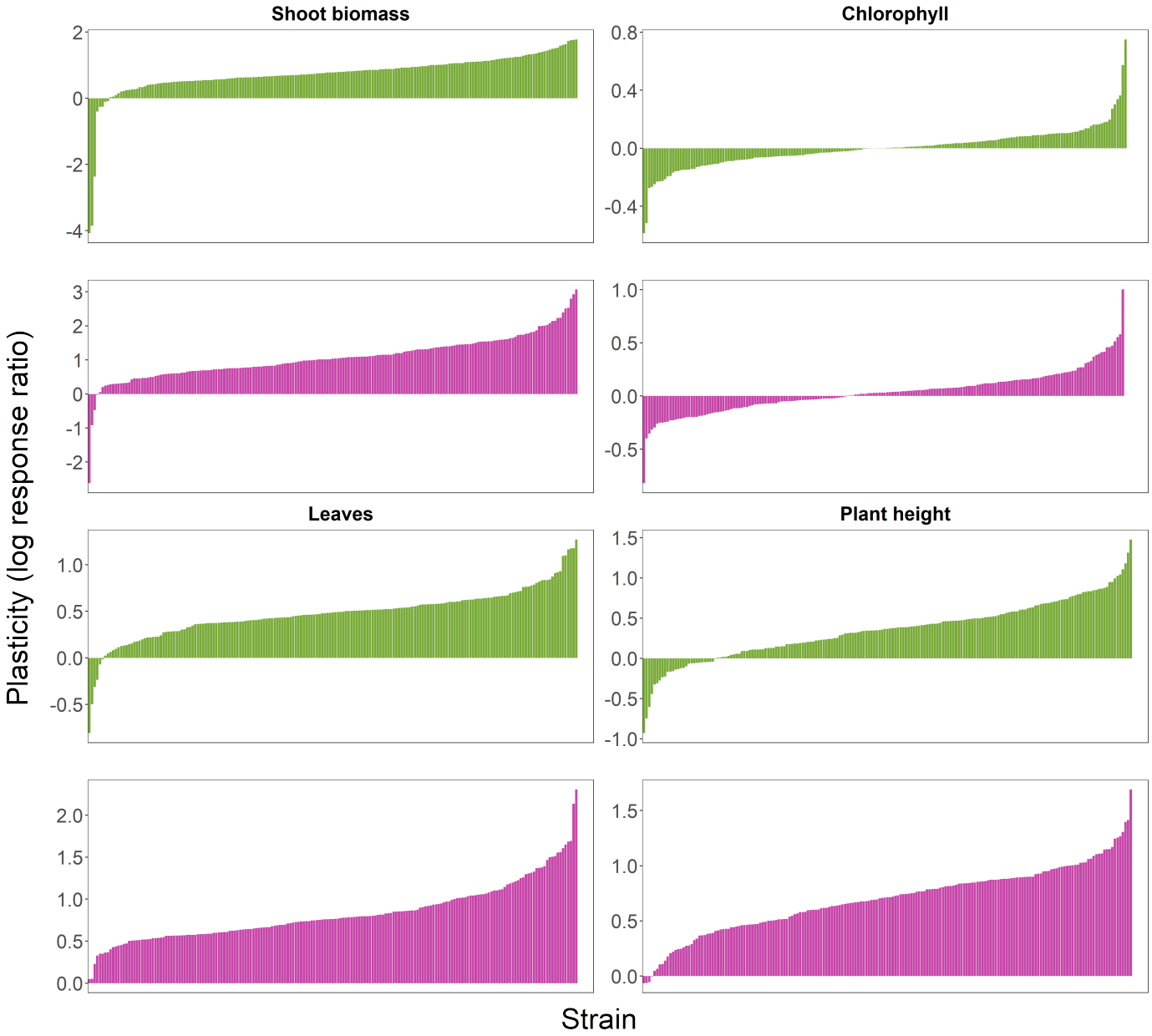
Supp. Figure S4 Genes underlying G × E are prevalent across partner quality traits. Venn diagrams showing number of rhizobium (*E. meliloti*) genes significantly associated with three other partner quality phenotypes for **A)** each of four separate mapping experiments or **B)** cross-experiment plasticity 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.



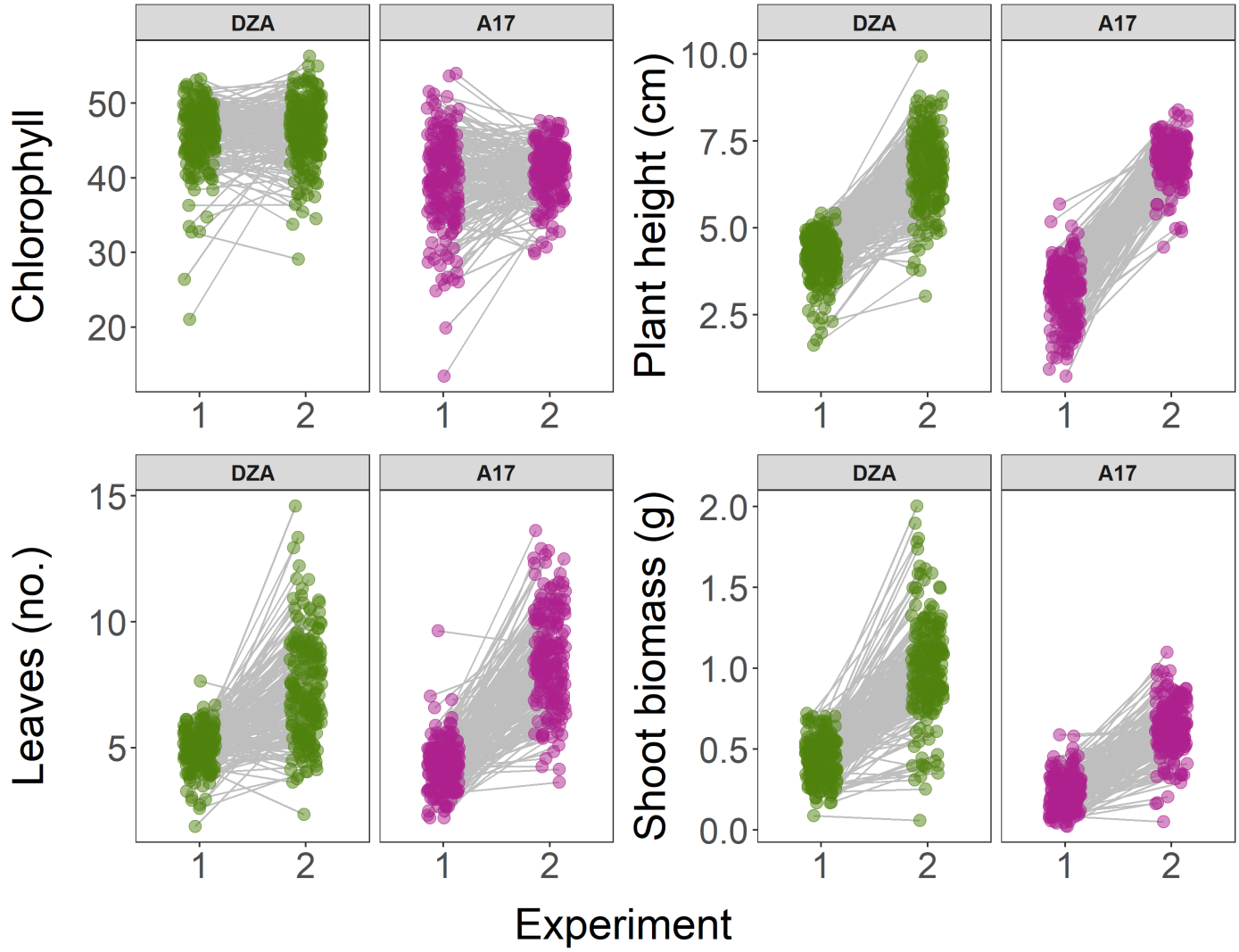
Supp. Figure S5 Genetic correlations for traits measured on DZA. Genetic correlations among traits within experiments (above or below diagonal) or between the same trait across experiments (along diagonal). Correlations based on estimated marginal means of each rhizobia strain corrected for rack on plant 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. Bottom right plot represents histogram for strain plasticity measured across the two experiments.



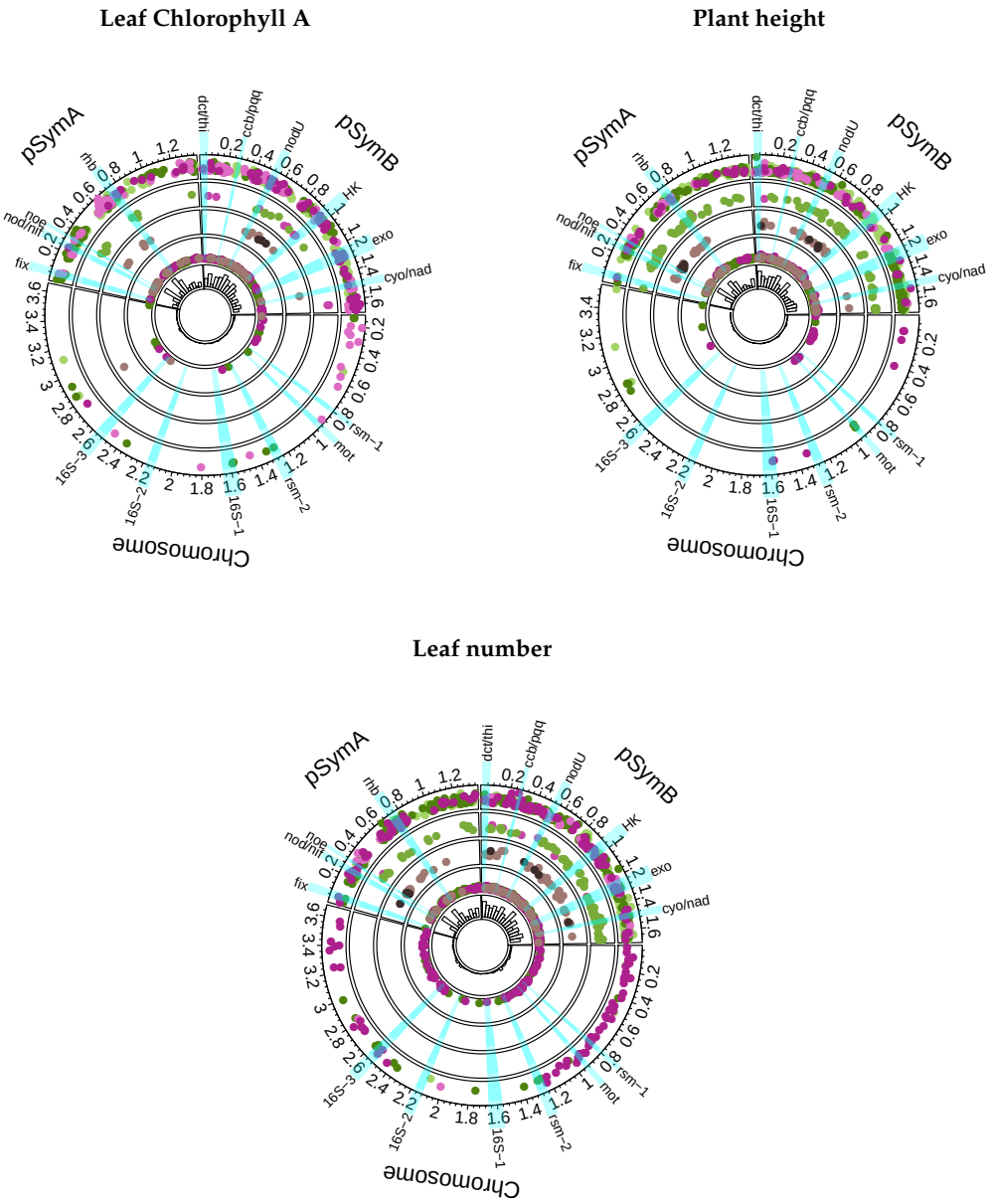
Supp. Figure S6 Genetic correlations for traits measured on A17. Genetic correlations among traits within experiments (above or below diagonal) or between the same trait across experiments (along diagonal). Correlations based on estimated marginal means of each rhizobia strain corrected for rack on plant 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. Bottom right plot represents histogram for strain plasticity measured across the two experiments.



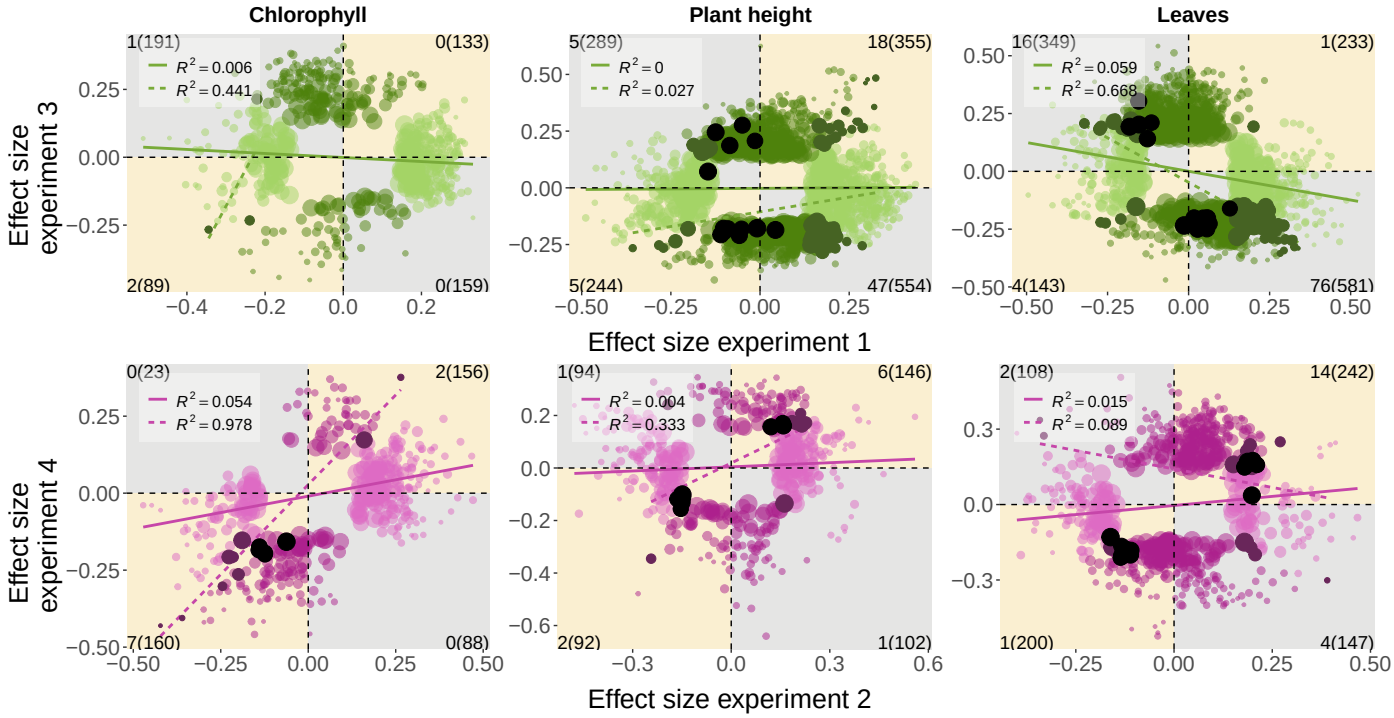
Supp. Figure S7 Strains significantly varied in their response to different experiments. Variation among *E. meliloti* strains in plasticity, calculated as the log response ratio for each trait between the two experiments with each host genotype (experiments 1 & 3 with DZA in green; experiments 2 & 4 with A17 in pink).



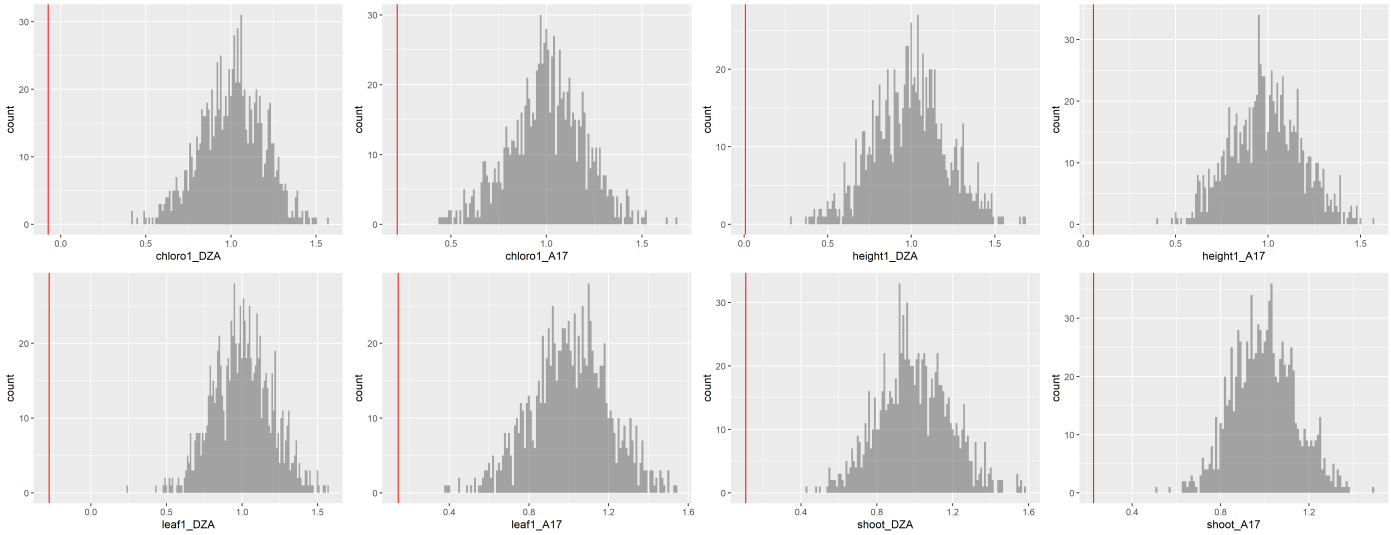
Supp. Figure S8 Extensive G × E between experiments. Reaction norms for partner quality traits across experiments. Data points represent estimated marginal means corrected for rack within each experiment. Points in green for means estimated on DZA, purple for A17.



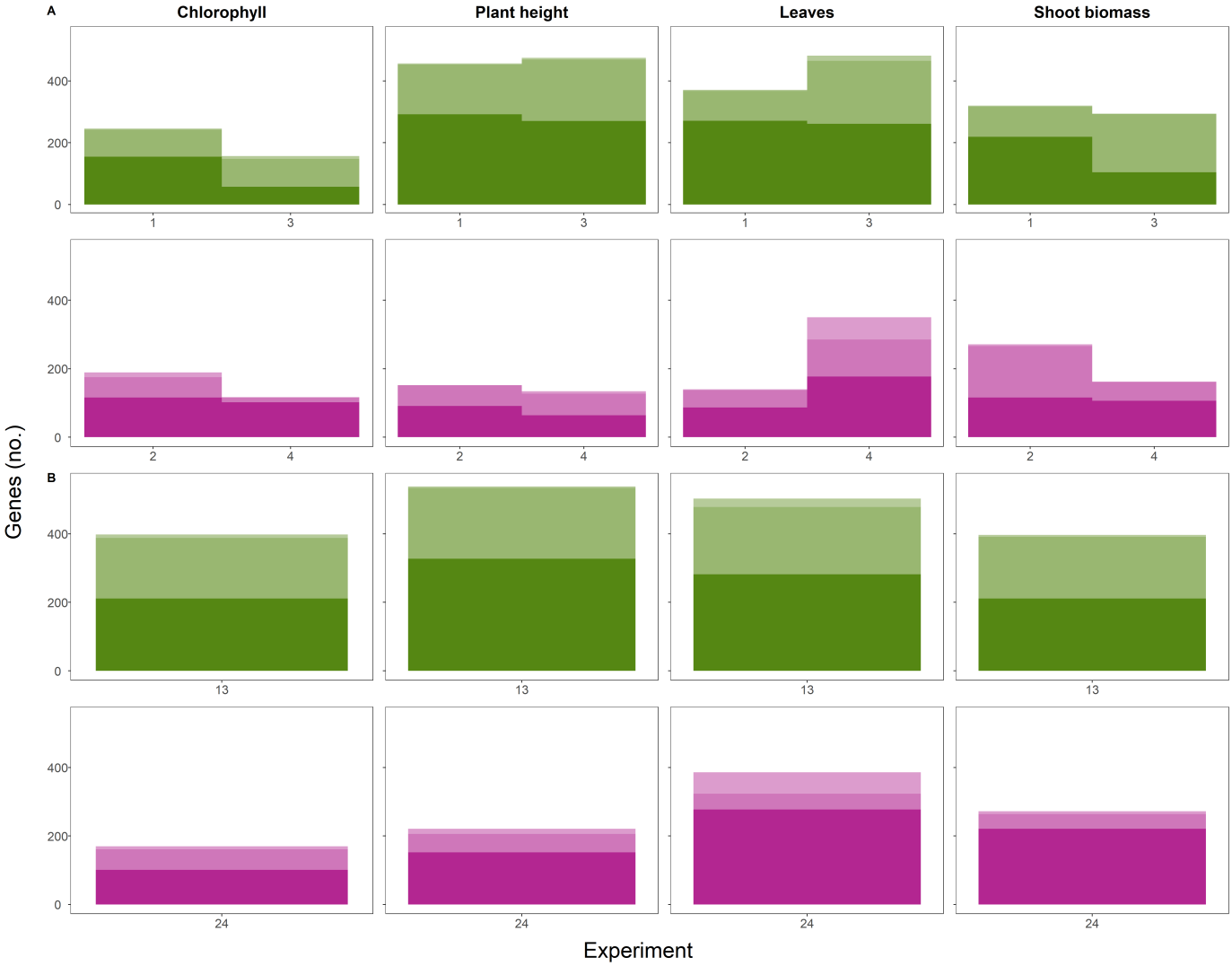
Supp. Figure S9 Loci associated with partner quality are mostly limited to the symbiosis plasmids. Circos plots showing positions of genes (dots) significantly associated with three additional partner quality traits. Each ring represents a different gene category, outermost to innermost: 1) G x E, 2) 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 colors 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 S10 Extensive G × E between experiments. Genome-level G × E for partner quality loci. Shown are correlations between the estimated effects of individual *E. meliloti* loci on plant shoot biomass (from GWAS) in each of two experiments for either host DZA (green) and A17 (pink). Only allelic effects that were significant in one (lighter colors) 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 line) or variants significant in both experiments (dotted 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 S11 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 S12 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. Files

- File 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.
- File 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.
- File 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.
- File 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.
- File 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.
- File 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.
- File 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.