Received: 15 March 2017 | Accepted: 10 July 2017

DOI: 10.1111/mec.14285

ORIGINAL ARTICLE

WILEY MOLECULAR ECOLOGY

Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism

Liana T. Burghardt[1] | Joseph Guhlin[1] | Chan Lan Chun[2] | Junqi Liu[3] | Michael J. Sadowsky[2] | Robert M. Stupar[3] | Nevin D. Young[1,4] | Peter Tiffin[1]

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

[2]BioTechnology Institute, University of Minnesota, St. Paul, MN, USA

[3]Department of Agronomy and Plant Genetics, University of Minnesota, St. Paul, MN, USA

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

Correspondence
Liana T. Burghardt, Department of Plant and Microbial Biology, University of Minnesota, St. Paul, MN, USA.
Email: liana.burghardt@gmail.com

Present address
Chan Lan Chun, Natural Resources Research Institute and Department of Civil Engineering, University of Minnesota Duluth, Duluth, MN, USA

Funding information
NSF, Grant/Award Number: IOS-1237993

Abstract

In the legume-rhizobia mutualism, the benefit each partner derives from the other depends on the genetic identity of both host and rhizobial symbiont. To gain insight into the extent of genome × genome interactions on hosts at the molecular level and to identify potential mechanisms responsible for the variation, we examined host gene expression within nodules (the plant organ where the symbiosis occurs) of four genotypes of *Medicago truncatula* grown with either *Ensifer meliloti* or *E. medicae* symbionts. These host × symbiont combinations show significant variation in nodule and biomass phenotypes. Likewise, combinations differ in their transcriptomes: host, symbiont and host × symbiont affected the expression of 70%, 27% and 21%, respectively, of the approximately 27,000 host genes expressed in nodules. Genes with the highest levels of expression often varied between hosts and/or symbiont strain and include leghemoglobins that modulate oxygen availability and hundreds of Nodule Cysteine-Rich (NCR) peptides involved in symbiont differentiation and viability in nodules. Genes with host × symbiont-dependent expression were enriched for functions related to resource exchange between partners (sulphate/iron/amino acid transport and dicarboxylate/amino acid synthesis). These enrichments suggest mechanisms for host control of the currencies of the mutualism. The transcriptome of *M. truncatula* accession HM101 (A17), the reference genome used for most molecular research, was less affected by symbiont identity than the other hosts. These findings underscore the importance of assessing the molecular basis of variation in ecologically important traits, particularly those involved in biotic interactions, in multiple genetic contexts.

KEYWORDS
Ensifer (*Sinorhizobium*), genome × genome interaction, *Medicago truncatula*, nitrogen fixation, Nodule Cysteine-Rich peptides, transcriptome plasticity

1 | INTRODUCTION

When growing in symbiosis with legume plants, rhizobium bacteria convert atmospheric nitrogen (N) into a form accessible to plants (Herridge, Peoples, & Boddey, 2008). This symbiosis serves as a model of resource mutualisms where each partner provides a key resource to the other; the plant provides energy in the form of dicarboxylates (primarily malate, a central component of the citric acid cycle) to the rhizobium, and the rhizobium provide the vital nutrient nitrogen to the plant in a usable form (ammonium/amino acids) (Haag et al., 2013; Udvardi & Day, 1997). This exchange occurs inside specialized plant organs called nodules. Much research on interactions between rhizobia and legumes has focused on the genes and signalling necessary for nodule formation and nitrogen

fixation (Kang, Li, Sinharoy, & Verdier, 2016; Oldroyd, Murray, Poole, & Downie, 2011). However, relatively little is known about the variation that exists in the costs/benefits traded between partners in nodules. A better understanding of this trade-off may provide insight into fundamental evolutionary processes, such as the stability of mutualisms and the presence of genetic variation within mutualisms (Garcia & Gerardo, 2014; Gilbert, Bosch, & Ledón-Rettig, 2015; Heath & Stinchcombe, 2014). Here, we use a well-studied symbiosis between the legume *Medicago trucatula* and the rhizobia *Ensifer* (formerly *Sinorhizobium*) species to investigate the extent to which hosts alter gene expression when paired with different symbiont partners, identify potential physiological mechanisms underlying host-specific and genome × genome interactions, and interpret them in the context of mutualism evolution.

Early modelling suggested mutualism should be difficult to maintain due to the potential for "cheater" genotypes to evolve (i.e., those that extract benefits from the partner without providing resources in return). However, extensive empirical work has failed to uncover evidence for rampant cheating, and many mutualisms appear to be stable (Friesen, 2012; Sachs, 2015; Sachs, Skophammer, & Regus, 2011). Recent models demonstrate that many mechanisms can create fitness alignment between partners (e.g., sanctions/negotiations, partner screening, population structure) thus alleviating conflicts of interest (Akçay, 2015; Jones et al., 2015). Each of these models posit specific mechanisms acting at different stages of the interaction, yet in most empirical systems the mechanisms that contribute to the evolution and stability of interspecific relationships are unknown (Baskett & Schemske, 2015). Mechanistic knowledge is particularly limited when one partner is a microbe because direct measurement of subcellular processes is technically challenging. Gene expression, an intermediate step between genotypic variation and phenotypic variation, can be particularly valuable for inferring subcellular traits that modify interactions with microbes and thus provide insight into the mechanistic basis of host genome × symbiont genome interactions (Simms & Porter, 2012).

As in many systems, phenotypic studies in legumes and rhizobia have shown that even within a species, symbionts are not all equivalent in the fitness benefits they confer upon their hosts. Similarly, not all hosts are equal in the fitness benefits they confer upon symbionts (Gorton, Heath, Pilet- Nayel, Baranger, & Stinchcombe, 2012; Heath & Tiffin, 2009; Larrainzar, Gil-Quintana, Seminario, Arrese-Igor, & González, 2014; Simonsen & Stinchcombe, 2014; Sugawara et al., 2013). Empirical data suggest that legume hosts provide fewer resources to nodules that provide little benefit to the plant (Gubry-Rangin, Garcia, & Bena, 2010; Heath & Tiffin, 2009; Oono, Anderson, & Denison, 2011). If hosts alter resources provided to nodules based on the benefits the plant obtains, we should see differences in host transcription related to the "success" of the mutualism (i.e., exchange of dicarboxylates and amino acids) and the creation of a favourable environment for rhizobial survival and/or nitrogen fixation in nodules.

There are extensive changes in the plant gene expression during nodule development (Benedito et al., 2008) and between different regions of mature nodules (Roux et al., 2014). For example, Benedito et al. found that 26,000 genes were expressed in developing and mature *M. truncatula—Ensifer* nodules. Thirty per cent of those genes showed a twofold change in expression at some point during development (Benedito et al., 2008). Likewise, more than 14,000 genes were differentially expressed among nodule regions (Roux et al., 2014). In the nitrogen fixation region, transporters that putatively mediate exchanges of resources among partners were upregulated as well as antimicrobial peptides related to defensins called Nodule Cysteine Repeat peptides (NCRs). In the context of legume/rhizobia nodulation, the NCRs are regulators of bacterial (symbiont) differentiation and viability (Pan & Wang, 2017) some of which are required for functional symbiosis (Shabab et al., 2016). When host genes necessary for early signalling between partners are mutated, the nodule transcriptome exhibits extensive alterations relative to wild-type plants (Larrainzar et al., 2015). Similarly, when host genes necessary for late nodule development and functional nitrogen fixation are mutated, nodules arrest at different stages of development and vary extensively in terms of symbiont and host transcriptomes (Lang & Long, 2015; Starker, Parra-Colmenares, Smith, Mitra, & Long, 2006). Taken together, these data suggest that nodule transcriptomes differ based on developmental stage and the amount of nitrogen the rhizobia inside are converting into a usable form.

However, little is known about the extent of naturally occurring variation in the transcriptome of functioning nodules and the degree host transcriptomes can be altered by symbiont strain. To our knowledge, only a single study has examined how gene expression is affected by interactions between natural genetic variants of legume hosts and rhizobial symbionts. Using microarrays, Heath, Burke, and Stinchcombe (2012) measured both rhizobial and host transcriptomes for two host genotypes when combined with two different rhizobial strains. They found that the expression of approximately 600 of ~10,000 assayed *Medicago truncatula* genes differed between the two hosts. In contrast, when they compared the expression state of the same host paired with a different symbiont strain only five host genes showed differential expression, and the expression of only one gene was significantly affected by the interaction between host genotype and symbiont strain (genome × genome). The almost complete lack of symbiont-dependent differences in host gene expression is surprising given the large observed influence of host genotype on rhizobial gene expression and the repeatedly identified symbiont-dependent effects on plant phenotypes. The transcriptome profiles in Heath et al. (2012) are, however, limited not only by the use of two closely related host genotypes and two closely related symbiont strains but also by the fact that the microarray used to assay host gene expression contained <20% of the 50,894 annotated genes in the current *M. truncatula* genome assembly (Tang et al., 2014).

To gain a broader and more complete picture of the symbiont-dependent transcriptome differences in *M. truncatula* nodules, and thereby gain greater insight into the molecular features of mutualism, we used RNAseq to measure nodule transcriptome responses of four host accessions to two symbiont species. These hosts by

symbiont combinations harbour significant variation in nodulation phenotypes and strain-dependent plant growth. Our overall goal was to identify mechanisms potentially responsible for variation in host-dependent response to rhizobial strains. First, we identified plant genes that are differentially expressed (i) among plant hosts but are unaffected by symbiont strain, (ii) in response to symbiont strain but shows consistent responses across plant hosts and (iii) due to the combined effects of both plant hosts and symbiont strain. After identifying genes with strong evidence for each expression pattern, we looked for enrichments in Gene Ontology (GO) terms and Protein Family (Pfam) domains. These enrichments provide insight into the functional roles of each category of differentially expressed genes. Importantly, we conducted these enrichment analyses using genes expressed in the nodules as the baseline, not the entire *Medicago* genome. Thus, the enrichments we find point towards functions that tend to vary in response to host genotype and/or symbiont strain specifically within nodules rather than functions that differentiate tissues types.

2 | MATERIALS AND METHODS

We collected host phenotype and transcriptome data on host-strain pairs using two complimentary experiments. In both experiments, we used four *Medicago truncatula* accessions (HM101 (A17), HM056, and HM034 and HM340 (R108)). Accession HM101 was used for the *M. truncatula* reference genome (Young et al., 2011). HM340 has been previously used to generate a large collection of tnt1 mutations for reverse genetic assays (Tadege et al., 2008). We inoculated plants with one of two *Ensifer* species, either *E. medicae* strain WSM419 (Reeve et al., 2010) or *E. meliloti* strain KH46c (Epstein et al., 2012). *E. meliloti* and *E. medicae* strains are the primary symbionts of *M. truncatula*.

2.1 | Quantification of symbiotic phenotypes

To measure the phenotypic effects each rhizobial strain had on each host genotype, we grew 10 replicates of each host in combination with each of the two rhizobial strains. We inoculated three-day-old seedlings, growing in Sunshine Mix #5 growth medium, with ~10^6 rhizobia in 1 ml. 0.85% NaCl solution. We grew plants for 2 months and watered and fertilized them as needed with sterile, N-free nutrient solution (Bucciarelli, Hanan, Palmquist, & Vance, 2006). We then measured dry vegetative mass, the average dry mass of the largest 10–20 nodules, and counted the number of nodules. We analysed the effect of host genotype and rhizobial strain on these phenotypes using an ANOVA with type III Sums of Squares ($phenotype = \beta_0 + \beta_1 Host + \beta_2 Sym + \beta_3 Host \times Sym + \beta_4 Tray + \beta_5 Edge$), with the last two terms included to account for variation due to positional effects. To specifically test the effect of rhizobial strain on each phenotype, we subdivided each phenotype by host and tested the effect of rhizobia on the phenotype while controlling for location effects as outlined above ($phenotype = \beta_0 + \beta_1 Sym + \beta_2 Tray + \beta_3 Edge$).

2.2 | Measurement of nodule (and root) gene expression

Plants were grown from seeds that were scarified in sulphuric acid for 5–8 min, sterilized with 10% bleach (0.525% sodium hypochlorite) for 90 s, rinsed eight times with sterile water then left on moist paper at 4°C for 3 days prior to germination. Plants were grown in a 1:1 Sunshine Mix #5: Turface MVP mixture in Leonard Jars in a growth chamber at 25°C with a 16 hrs (light) and 21°C 8 h (dark) cycle. Plants were watered with sterile N-free nutrient solution (Bucciarelli et al., 2006) as needed. Three days after emergence, at least 12 individuals of each accession were inoculated with one of the two symbiont strains, and three plants were left uninoculated. Seedlings were inoculated with ~10^7 rhizobial cells diluted in 1 ml of 0.85% NaCl. Rhizobia were initially grown to stationary phase in tryptone yeast (TY) media. After inoculation, soil was covered with sterilized, paraffin-coated sand to prevent contamination and water evaporation. Nodules were collected 14-day post-inoculation and 0.8–1.2 g of nodules from 4 to 7 plants from each treatment were pooled for RNA extraction. Roots were washed to remove soil, and nodules were collected by cutting the root approximately 0.5 cm on each side of the nodule. These intact nodules were rinsed with sterile water, frozen in liquid nitrogen and stored at −80°C until RNA extraction. Roots from uninoculated plants were sampled in a similar manner. All extractions were completed within 5 days of harvest.

2.3 | RNA extraction, sequencing and processing

RNA was isolated from the flash-frozen nodules using Qiagen RNeasy mini kit. Barcoded libraries were generated using TruSeq Stranded total RNA kits and sequenced on an Illumina HiSeq 2000 to produce 2 × 100 bp, paired-end sequences. All samples generated more than 10 million reads with quality score >30. Library preparation and sequencing were performed by the University of Minnesota Genomics Center.

Reads were aligned to the HM101 reference genome (Tang et al., 2014; Young et al., 2011) using TopHat (maximum intron length of 8,000, minimum intron length of 40, b2-very sensitive) and Cufflinks (-u multi read correct, frag bias correct using multi FASTA file) (Trapnell, Pachter, & Salzberg, 2009; Trapnell et al., 2012). For all hosts, 3–4% of reads aligned to multiple locations, and the percentage of reads mapping was inversely related to how closely related each accession was to HM101, the accession used for the reference genome (mapping rates: HM101 = 93.9%, HM056 = 90%, HM034 = 87.5%, and HM340 = 72.8%). Despite these differences, approximately the same number of genes were identified as being expressed in each of the four of the hosts and HM340 showed no evidence of having lower expression levels than HM101 either in the entire expressed gene set or in the H × S focal set (see "Using pairwise models and fold change criterion to generate a focal list"; Fig. S1), and mapping rate variation should have little effect on measurement of symbiont effects within hosts.

We filtered alignments with Bamtools by requiring a mapping quality >30 and then used HTseq (Anders, Pyl, & Huber, 2015) to count the number of best-matched, aligned reads via the union method. In all replicates, >99% of all best match reads aligned unambiguously resulting in a median of 12.4 million reads per sample (ranging from 9.95 to 15.8 million). Counts were then normalized across samples using the standard "median of ratios" method implemented in the R package DESeq2 (Love, Huber, & Anders, 2014). In brief, the normalization factor for each sample was calculated by first determining, for each gene in each sample, the ratio of the focal sample to the geometric mean of that gene across all samples. Next, the median ratio for each sample was determined across all genes. The "median of ratios" method prevents highly expressed genes from biasing normalization. We calculated this size factor once across all genes in each sample and used it for the following purposes: (i) to normalize count data prior to filtering down to expressed genes, visualizing expression patterns, and determining fold changes; (ii) as a sample-specific covariate (*s*) that accounts for library size within all statistical models (see "Analysis of differential expression" for details). Note that after normalization, counts are no longer restricted to integer values.

2.4 | Identifying expressed/nodule-specific genes and quality control

We focused analyses on genes for which two replicates in at least one of the eight treatments had normalized read counts >5. A little over half (27,021) of the genes annotated in the Mtv4.0 reference genome (Tang et al., 2014) met this criterion and were thus designated as "expressed genes." Approximately 66% of these genes had read counts normalized by library >5 in all replicates in all eight treatments and 94% had read counts >5 in two or more treatments. Using these same expression criteria we found 26,288 genes to be expressed in the roots.

We conducted a variance stabilized Principal Component (PC) analysis on the normalized expression of the nodule-expressed genes using DESeq2 (Love et al., 2014). Given a variance-mean dependence, a variance-stabilizing transformation (VST) transforms the values so the variance is approximately independent of the mean. Preliminary PC analysis revealed three samples with patterns of gene expression that were suggestive of possible contamination: one HM056 sample inoculated with *E. meliloti* had an expression profile that was more similar to the HM056 plants inoculated with *E. medicae* than with the other *E. meliloti* replicates, and two HM101 plants, one inoculated with *E. meliloti* and one inoculated with *E. medicae* grouped more closely with each other than they did with the other HM101 samples. For these samples, we checked the identity of the symbiont by mapping RNAseq reads that did not align to the *M. truncatula* genome, to the *E. meliloti* (Galibert et al., 2001) and *E. medicae* (Reeve et al., 2010) reference genomes. This mapping revealed that the two HM101 samples contained evidence of having been inoculated with both *E. meliloti* and *E. medicae*, and thus, those two samples were discarded from subsequent analysis. However, there was no evidence for contamination of the HM056 sample and therefore to be conservative we retained this sample in the statistical analyses but removed it from visualizations and clustering of treatment similarities to avoid the outlier from biasing effect size estimates.

2.5 | Analysis of differential expression

To assess the effect of host genotype (Host), symbiont strain (Sym), and interactions between host and symbiont (Host × Sym) on gene expression, we used DESeq2. DESeq2 uses generalized linear models (glm) and dispersion priors estimated from the data to model raw counts of sequence reads (Love et al., 2014). For each gene (*i*) in sample (*j*), the read count (K_{ij}) is described with a generalized linear model (glm) of the negative binomial family with a log link function.

$$K_{ij} \sim \mathrm{NB}(\mathrm{mean} = \mu_{ij}, \mathrm{dispersion} = \alpha_i)$$

$$\mu_{ij} = s_j\, q_{ij}$$

$$\log_2(q_{ij}) = \beta_0 + \beta_1 \mathrm{Host} + \beta_2 \mathrm{Sym} + \beta_3 \mathrm{Host} \times \mathrm{Sym}$$

The mean of the distribution is equal to a quantity (q_{ij}) proportional to the concentration of cDNA fragments from the gene in the sample scaled by a size factor specific to that sample (s_j) estimated on an experiment-wide basis. In these statistical models, we are comparing the expression of the same gene across treatments, and thus, there is no need to standardize counts by gene length. However, in downstream analyses where we compare expression levels of genes of different lengths to each other (such as in Table 1), we standardized counts by coding sequence length.

For each gene, we used nested likelihood ratio tests (LRT) to evaluate the statistical significance of Host, Sym and Host × Sym on gene expression. We first tested if including the Host × Sym interaction improved the likelihood of the model by comparing the full model to a model with the interaction term removed using a chi-square test. If the interaction term was not significant (at FDR < 0.05), we used a LRT to ascertain the significance of the Host and Sym terms individually while controlling for the other term. Both Host and Sym terms often significantly improve the model thus providing evidence for an additive effect on expression (Host + Sym). This means a single gene can demonstrate both host and symbiont-dependent expression. To be conservative in our testing, we did not allow DESeq2 to optimize a minimum count filter that maximizes the number of significant statistical tests (independent filtering=FALSE).

2.6 | Using pairwise models and fold change criterion to generate a focal list

The full model analysis does not provide an easy means to impose a minimum fold change on expression and therefore might contain genes with only very small changes in expression. Thus, we tested each of the six pairwise comparisons between host genotypes using the same statistical framework outlined above. As above, we used the library scalars calculated across all nodule samples for each

TABLE 1 The 10 genes with the highest expression in nodules. To make expression comparable across genes we divided the mean normalized read counts by the coding sequence length before ranking genes

Gene (MT v4.0)	Mean counts (norm)	Exp. Cat (full model)	Annotation
Medtr5 g066070	196,613.7	H + S	leghemoglobin Lb120-1
Medtr3 g055450	121,778.2	H + S	NMS32/34 protein, putative
Medtr3 g055440	155,853.8	H + S	nodulin-25 protein
Medtr6 g463200	28,052.6	H	Nodule Cysteine-Rich (NCR) secreted peptide
Medtr7 g027180	27,723.4	H + S	Nodule Cysteine-Rich (NCR) secreted peptide
Medtr5 g081000	56,398.6	H	leghemoglobin Lb120-1
Medtr1 g011540	51,914.6	H + S	leghemoglobin Lb120-1
Medtr6 g445080	22,582.7	H + S	Nodule Cysteine-Rich (NCR) secreted peptide
Medtr5 g081030	41,665.3	H + S	leghemoglobin Lb120-1
Medtr4 g066070	38,789.4	H × S	Nodule-specific Glycine-Rich Peptide

comparison and all expressed genes were tested for every combination. To focus on genes likely to have the greatest effect, we instituted the following criteria for membership in a given expression category (i.e., affected by host, affected by symbiont, affected by both host and symbiont, or interactions between the effect of host and symbiont): the gene had to be significant at a Benjamin-Hochberg FDR $p < .05$, and there had be a doubling (or halving) of expression across the respective symbiont or host contrast. For the Host × Sym terms, we calculated the absolute value of the difference between hosts for each symbiont. This calculation accounts for the possibility of crossing reaction norms in which average expression levels may not change due to rank order reversals of gene expression.

2.7 | Analysis of functional enrichment

To identify biological functions that were overrepresented in each differential expression (DE) category, we compared the representation of GO-terms and Protein Family (Pfam) domains (Finn et al., 2016) of the DE genes to the 27,021 genes expressed in the nodules: 11,113 of these genes have been assigned level-one biological processes GO annotations and 20,922 have one or more Pfam domains. To create these assignments, we used INTERPROSCAN v5.16-55.0 (Jones et al., 2014) run on the complete *Medicago truncatula* 4.0v1 predicted protein set from JCVI (Tang et al., 2014). While most genes had 1–2 Pfam domains associated with them, 602 genes had ≥10 Pfam domains, often due to repeat regions. These genes can skew the enrichment analysis and thus we removed them, leaving 20,320 genes associated with 44,424 Pfam domains.

To quantify the extent to which a term was overrepresented in a group we calculated its enrichment compared to the list of all expressed genes in nodules and tested the statistical significance of the enrichment using a Fisher's exact test.

$$\text{Enrichment} = \frac{\#\text{ times term is in Focal Genes}}{\text{total }\#\text{ terms in Focal Genes}} \Big/ \frac{\#\text{ times term in Expressed Genes}}{\text{total }\#\text{ terms in Expressed Genes}}$$

3 | RESULTS

3.1 | Symbiotic phenotypes were strongly altered by specific combinations of host and symbiont

Plant biomass varied significantly among the combinations of host genotype and rhizobial strain (Figure 1a, $p_{host \times strain\ df=3}$ = 2.7e-13); HM101 and HM056 were relatively unaffected by symbiont identity, HM034 accumulated ~15% more biomass with *E. medicae* and HM340 accumulated twice as much biomass with *E. meliloti*. The number of nodules produced, average nodule weight and host biomass per nodule all exhibited strongly significant host by symbiont interactions (Figure 1b–d, $p_{host \times strain\ df=3}$ = 6.7e-9, 3.4e-10, 5.4e-12, respectively) with HM340 forming far fewer yet larger and more beneficial nodules (as assessed based on vegetative biomass) with *E. meliloti* (see Fig. S2 for picture of plant and nodule phenotypes), whereas the other three hosts formed more, smaller nodules with *E. meliloti* (Figure 1c,d). For all plant hosts, the strain with which the plant formed more nodules was the strain that provided less benefit in terms of vegetative biomass.

3.2 | NCRs, leghemoglobin, sugar and amino acid synthases and transporters were highly expressed

The transcriptomes of the eight hosts by symbiont combinations included a total 27,021 expressed genes. After library normalization, the mean read counts of these genes across all samples were 108 reads with the bottom 2.5% having ≤2 reads and the top 2.5% having ≥2,830 reads. The three most highly expressed genes each accounted for ~1% of total reads: Medtr5 g066070 (leghemoglobin Lb120-1), Medtr3 g055440 (nodulin-25 protein) and its neighbour Medtr3 g055450 (NMS 32/34 protein putative) (Table 1 for the top 10). As expected, the 1,000 genes with highest expression—after normalizing for gene length—were enriched for GO annotations

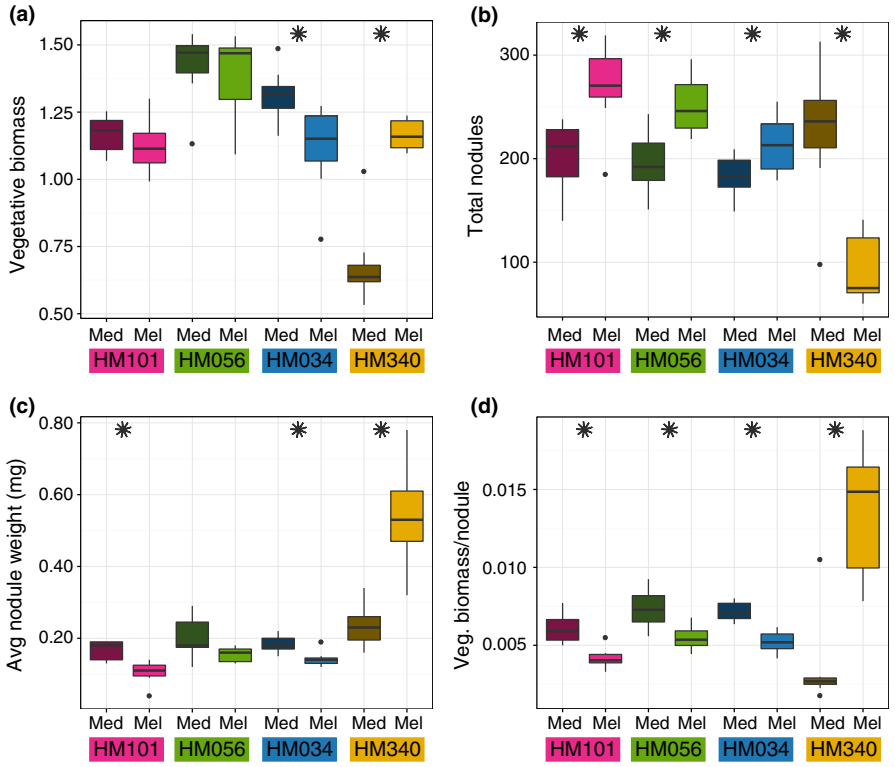

FIGURE 1 Effect of rhizobial strain on host phenotypes: vegetative biomass (a), total nodule number (b), average nodule weight (c) and vegetative biomass per nodule (d). Host genotypes are shown in different colours, and symbiont strain is indicated by intensity (Med: *Ensifer medicae*-darker shade or Mel: *E. meliloti*-lighter shade). Host and strain had an interactive effect on all phenotypes based on ANOVA testing (all phenotypes $p << .001$). Asterisks indicate within host contrasts where rhizobial strain significantly influenced phenotypes ($p < .05$)

related to nodule morphogenesis, translation, oxygen and iron transport, polyamine/glutamine/asparagine biosynthesis, glucose metabolism and stress response (Table S1). The expression of the vast majority of these genes varied significantly among hosts (414), among symbionts (25), host and symbiont (230), or the interaction between host and symbiont (265). Strikingly, 299 of the 1,000 most highly expressed genes were NCRs. Expression of these NCR's strongly differed among hosts (Figure 2), especially HM101, and also differed based on the identity of the symbiont in the nodule, a result consistent with trends found in Heath et al. (2012).

3.3 | Host genotype and symbiont strain structure nodule transcriptome

To examine transcriptome-wide differences in expression, we used a PC analysis to describe the variance in transcript levels among the eight hosts by symbiont combinations. The first two PCs together accounted for >62% of the variance in expression (Figure 3) with the among-host differences along the first PC mirroring the relatedness among accessions. HM101 and HM056, which are most closely related (Stanton-Geddes et al., 2013), had the most similar expression profiles. HM340 (R108) is the most distantly related to HM101 (Branca et al., 2011) and has the most dissimilar expression profile. Symbiont identity did not affect the expression profiles of all plant accessions equally; with the reference accession HM101 having only very small differences in expression between symbiont strains.

We classified the 27,201 expressed genes into each of five categories based on whether expression varied among hosts (12,335 genes), symbionts (826 genes), both hosts and symbionts (6,463 genes), interactions between host and symbiont (5,785 genes), or were unaffected by either host or symbiont (1,612 genes) (Table 2, Figure 4). The read count and coding sequence length distributions of genes that fell into each of the categories did not strongly differ from the distribution of all expressed genes (Fig. S3) except, as expected, we had less power to detect differences in genes with low numbers of reads.

Pairwise contrasts between each of the four hosts allowed us to identify how expression was affected by specific host by symbiont combinations and to impose a minimum fold change criterion between treatments to lessen the number of false positives among the differentially expressed genes. This filtering tended to exclude genes with low numbers of reads but did not select by coding sequence length (Fig. S4). Across the six contrasts, we found 4,149 genes with significant pairwise Host × Sym and a doubling of expression. More differences were detected between pairs of more distantly related hosts than closely related hosts (Table 2). A total of 1,727 genes met Host × Sym criteria in two or more of the pairwise contrasts, and we considered these as focal H × S genes. The vast majority (1,644) of these focal genes were also identified as Host × Sym using the full model. Approximately 5% of the focal H × S genes (85) showed no evidence of being expressed in roots (Table S2).

To further characterize the focal H × S genes we clustered them based on their symbiont-dependent change in expression (Figure 5 and Fig. S5). This analysis revealed four major categories of gene response: upregulated in HM101 inoculated with *E. medicae* and in HM034 inoculated with *E. meliloti* (cluster 1: 526 genes), upregulated in HM340 inoculated with *E. meliloti* and in HM034 inoculated with *E. medicae* (cluster 2: 517 genes), strongly

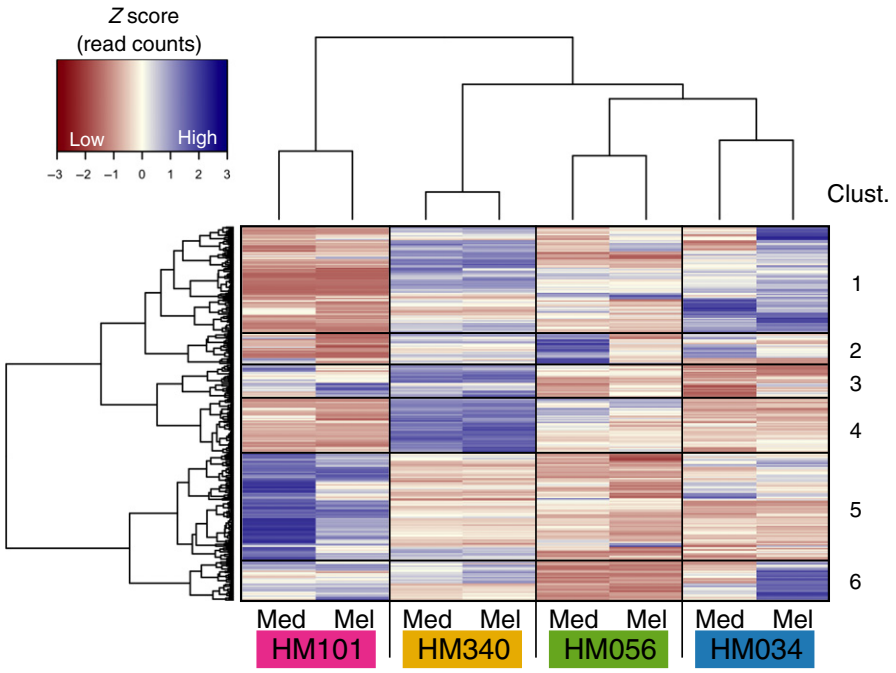

FIGURE 2 Clustering of the highly expressed Nodule Cysteine-Rich (NCR) secreted peptides *Z* scores scale differences in expression between treatments in terms of standard deviations from the mean. Each of the horizontal lines represents a different gene (299), and the vertical bars represent each of eight host × symbiont combinations. Deeper red values indicate expression lower than the mean for that gene in that treatment, and deeper blue values indicate expression higher than the mean. The outlier HM056 *medicae* replicate was removed before calculation of *Z* scores for each treatment. Clustering was conducted in R using the euclidean distance method (euclidean in dist function) paired with Ward's clustering method (ward.D2 in hclust function)

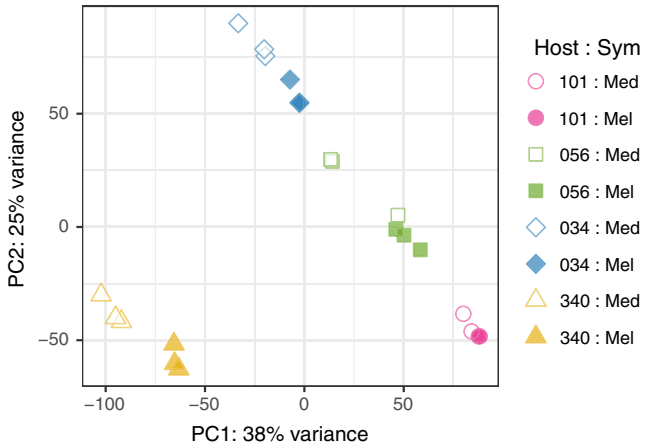

FIGURE 3 Visualization of overall variation in gene expression for all genes expressed in nodules. Each data point indicates a biological replicate with host genotype indicated by colour and shape and symbiont indicated by shading (*Ensifer meliloti*-filled and *E. medicae*-open). Note that two replicates of 101:Mel and 034:Mel graph directly on top of each other. The empty square grouping with the filled squares indicates the HM056 outlier that was retained in statistical and principle component calculations but removed from visualizations. Variance stabilized principal components analysis was conducted using the DESeq2 package in R

upregulated by *E. meliloti* (cluster 3: 222 genes) or by *E. medicae* (cluster 4: 502 genes) in all hosts but HM101. The pattern in cluster 2 is the same pattern seen in the phenotypic data; HM340 and HM034 biomass accumulations were positively affected by *E. meliloti* and *E. medicae*, respectively (Figure 1a). This analysis also confirmed that HM101 is less responsive to symbiont identity than the other three hosts.

TABLE 2 Number of expressed genes that showed statistical evidence of each type of expression pattern: variation among Hosts, Symbionts (Sym), both Hosts + Symbionts, additive influence of symbiont and host (Host + Sym) or interaction between host and symbiont (Host × Sym). The top line notes the number of genes based on a full model that included all hosts and is followed by results for each pairwise host comparison. In all cases, significance was based on an FDR-corrected *p*-value < .05. For the pairwise comparisons, genes also had to meet fold change requirements (see methods). Each gene was assigned to a single category with a significant interaction pre-empting any lower order effects

	None	Sym	Host	Host + Sym	Host × Sym
All hosts	1,612	826	12,335	6,463	5,785
Pairwise host contrast					
HM101 vs. HM056	19,588	70	6,777	220	366
HM101 vs. HM034	16,862	155	8,327	495	1,182
HM101 vs. HM340	14,926	243	8,831	997	2,024
HM056 vs. HM034	18,834	1,449	5,157	1,372	209
HM056 vs. HM340	16,087	1,559	6,846	1,264	1,265
HM034 vs. HM340	15,825	1,934	6,284	1,513	1,465

3.4 | Functional enrichment analyses reveal downstream and upstream candidates

Gene ontology (GO) enrichment analyses suggest that expression categories are functionally distinct groups of genes (Table 3). Not surprisingly, the genes that showed no evidence of host or symbiont variation were primarily involved in basic cellular processes. Genes that varied among hosts but not symbionts were enriched for nodule morphogenesis and defence responses, while genes that were affected by symbiont only were primarily involved in cellular

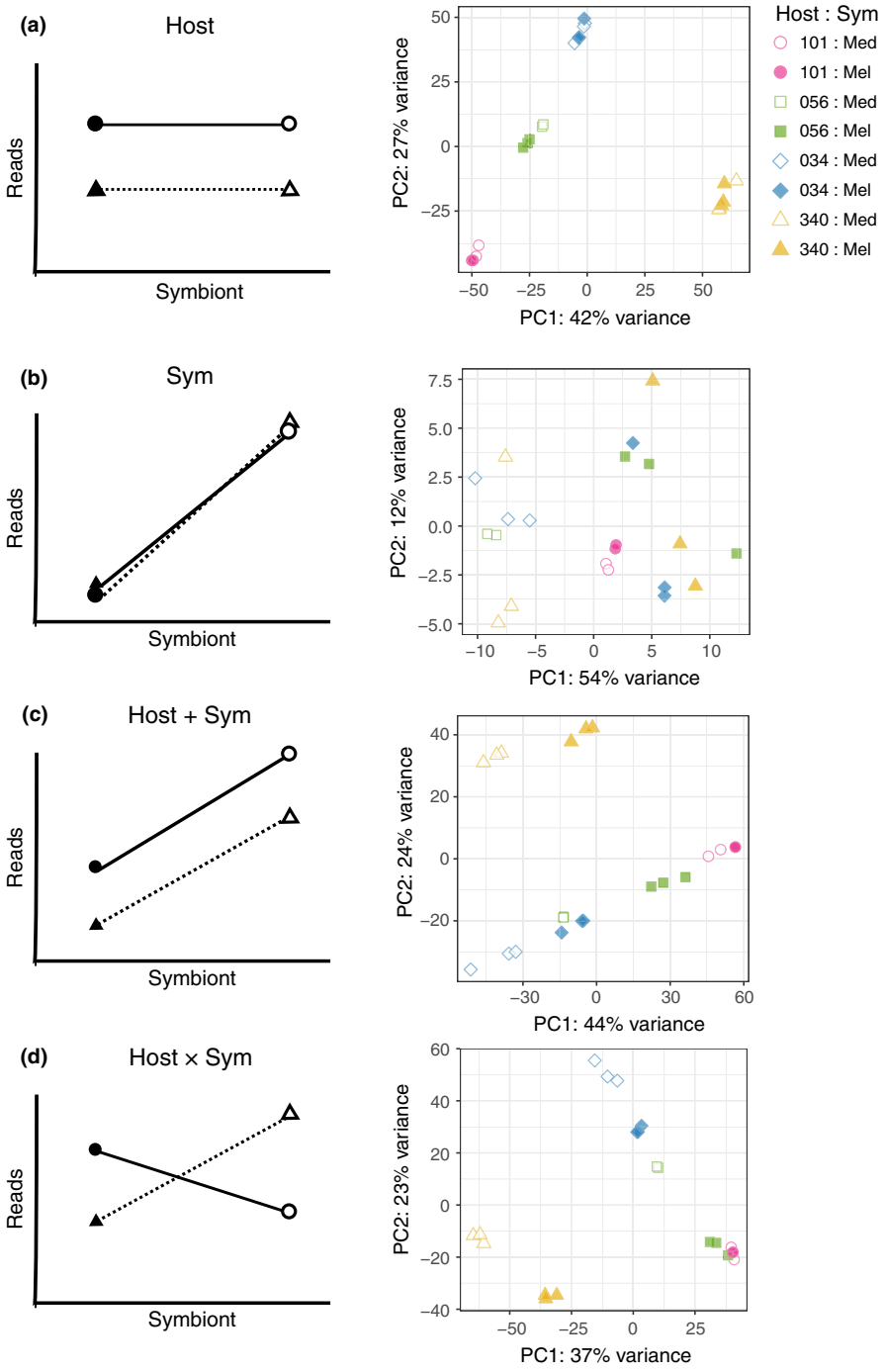

FIGURE 4 Example schematics of gene expression patterns for each of our categories (left) and visualization of the expression variation in genes that fall into each of these categories based on the full statistical model (right). Each data point indicates a biological replicate with host genotype indicated by colour and shape and symbiont indicated by shading (*Ensifer meliloti*-filled and *E. medicae*-empty). (a) Genes that fall into the host category (group by host: colour/shape), (b) Genes that fall into the symbiont category (group by symbiont: filled vs. open), (c) Genes with additive host and symbiont effects (filled shape is always to the upper right of open shapes) and (d) Genes with interactions between host and symbiont. Sometimes replicates of 101:Mel and 034:Mel plotted directly on top of each other. Variance stabilized principal components analysis was conducted using the DESEQ2 package in R. Fig. S5 shows PCA for focal H × S genes which are quite similar to the pattern shown here in d

development (microtubule-based movement, mitotic nuclear division and regulation of cyclin-dependent serine/threonine kinase activity) suggesting that the rate of nodule development might differ between the *E. meliloti* and *E. medicae* strains. Genes affected additively by both host and symbiont were enriched for functions related to redox, defence and pathogenesis. Lastly, the Host × Sym (Table 3) and focal H × S (Table S3) genes were enriched for redox processes implicated in the regulation of nodule metabolism (Dalton, Becana, Matamoros, & Udvardi, 2010), glutamate biosynthesis, trehalose biosynthesis and carbohydrate metabolism which may alter rhizobial performance (Sugawara, Cytryn, & Sadowsky, 2010; Vellosillo et al., 2007), and transporters of phosphate, iron and sulphate which are key cofactors of the symbiosis (Brear, Day, & Smith, 2013; Krusell et al., 2005; Udvardi & Poole, 2013; Wipf, Mongelard, van Tuinen, Gutierrez, & Casieri, 2014). Direct comparisons of enrichment between groups (e.g., Host vs. Host × Sym) corroborated and strengthened the signal of fold changes of top enrichment categories (data not shown).

Pfam enrichment analyses, which can provide insight into the function of genes lacking GO annotation, suggest that the focal

H × S genes are enriched for genes with the potential for direct effects on the host-symbiont interaction: malate biosynthesis, defence response to abiotic/biotic stressors (lipoxygenases with PLAT domains), amino acid transport (EamA-like transporters) and transcription domains AP2, TIFY and CCT (Table S3). Both AP2 and TIFY transcription factors have been implicated in regulating gene expression in response to environmental stimuli (Aparicio-Fabre et al., 2013; Bai, Meng, Huang, Qi, & Chen, 2011; Licausi, Ohme-Takagi, & Perata, 2013) and some AP2 genes control expression of nodule genes (Nova-Franco et al. 2015). These families of transcriptional regulators may control the differences in downstream expression we detect. Interestingly, when we conduct analyses on each of the expression clusters (Figure 5) separately, we find stronger statistical support for enrichment, despite the fact that group sizes are smaller (Table S4). For instance, cluster 2 where expression patterns mimic phenotype is enriched for domains involved in dicarboxylate biosynthesis (20× increase in glycosyltransferase family 20) and amino acid synthesis (20× increase for glutamine synthetase and 17× for aminotransferases).

4 | DISCUSSION

The phenotypic and fitness consequences of biotic interactions often depend on the genotype of each of the interacting partners. The mechanisms underlying these genome by genome interactions are perhaps best described in host pathogen systems in which the interaction outcomes can depend on intraspecific variation in host resistance genes and pathogen virulence (Didelot, Walker, Peto, Crook, & Wilson, 2016; Guidot et al., 2014; Toft & Andersson, 2010; Yue et al., 2015). While empirical data clearly demonstrate that natural genetic variation in partners shapes the outcomes of resource mutualisms (Heath & Stinchcombe, 2014), little is known about the specific host genes or the physiological mechanisms that are either responsible for or accompany these interactions (Baskett & Schemske, 2015).

Our study revealed extensive host-, symbiont- and host × symbiont-specific responses in nodule gene expression. Specifically, the expression of nearly 70% of nodule-expressed genes differed among hosts, approximately 27% of genes were affected by the symbiont strain identity, and the expression of more than 20% of genes depended on the interaction between the host and strain. Importantly, the extensive transcriptome variation we found is not variation among tissues or between inoculated or uninoculated plants, that is between developmental states. Rather, these are differences in the same tissue of different host genotypes regardless of symbiont identity, in response to different symbionts regardless of host identity, or due to interactions between hosts and symbionts (i.e., G × G). The importance of between-host variation is consistent with results from Heath et al. (2012) who also found widespread differences in nodule gene expression between hosts. However, unlike Heath et al. (2012) we also detected statistically and likely biologically significant effects of rhizobial strain and host by rhizobia combinations on the host nodule transcriptome. The reasons for the differences between the two studies are presumably due, in part, to the use of more

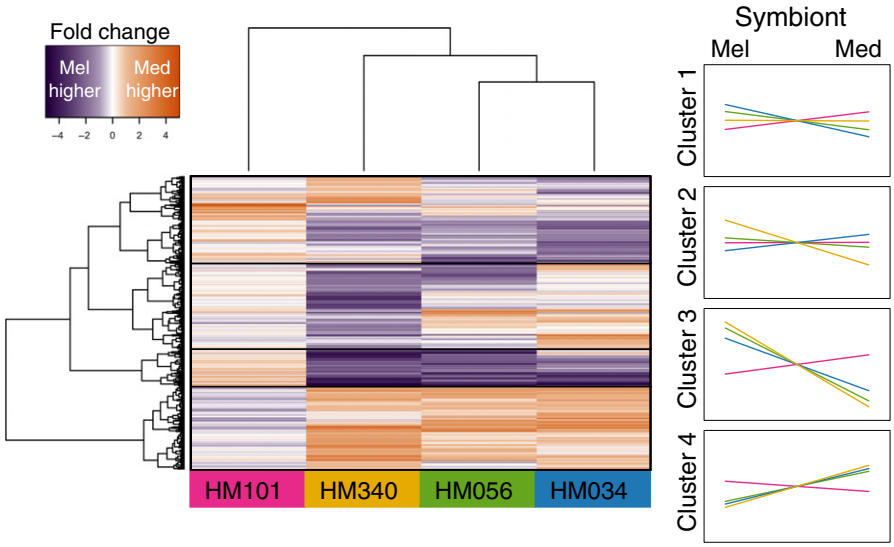

FIGURE 5 Clustering and heatmap of fold change patterns (log_2) of the 1,767 focal *Medicago truncatula* genes that show strong and consistent host by symbiont (Host × Sym) interactions. This metric assesses gene expression changes due to symbiont within each host. Each horizontal line represents a different gene, and the vertical bars represent each of the four hosts. Purple values indicate genes that have higher expression when *Ensifer meliloti* is in the nodule (Mel. higher), and orange values indicate genes that have higher expression when *E. medicae* is in the nodule (Med. higher). The right panels show the average slope of host response (lines) to symbiont in each cluster. See Table S4 for Pfam enrichments for each cluster. The outlier HM056 *medicae* replicate was removed before calculation of fold changes. Clustering was conducted in R using the euclidean distance method (euclidean in dist function) paired with Ward's clustering method (ward.D2 in hclust function)

TABLE 3 Summary of the five most strongly enriched biological processes (GO-terms, level 1) for each expression category in Table 2. The 1,000 genes with the lowest *p*-values within each category were used for analyses, except for the Symbiont only class for which there were only 826 genes. For genes with no evidence of Sym or Host effect, we used the 1,000 genes with the highest Host *p*-values. The *p*-value reported for the enrichment tests are from a Fisher's exact test

GO Biological processes	Fold change	*p*-Value
No H or S 1,000 genes		
Translation	3.24	.000
ATP hydrolysis coupled proton transport	12.08	.000
Ubiquitin-dependent protein catabolic process	3.19	.002
Mitochondrial electron transport, ubiquinol to cytochrome c	21.47	.009
Gene silencing by RNA	7.43	.011
Top Sym (826 genes)		
Microtubule-based movement	5.99	6.69E-06
Mitotic nuclear division	9.95	.001
Regulation of cyclin-dependent protein serine/threonine kinase activity	5.62	.003
Transmembrane transport	2.35	.010
Pseudouridine synthesis	5.16	.010
Top Host 1,000 genes		
Defence response	3.35	2.34E-18
Nodule morphogenesis	2.50	3.08E-08
Nucleosome assembly	4.07	4.54E-05
L-arabinose metabolic process	9.21	.007
Formaldehyde catabolic process	20.47	.009
Top H + S 1,000 genes		
Defence response	2.24	9.95E-9
Pathogenesis	5.86	.003
Oxidation–reduction process	1.47	.007
Purine ribonucleoside salvage	15.63	.015
Exocytosis	3.75	.029
Top H × S 1,000 genes		
Oxidation-reduction process	1.77	2.74E-05
Transmembrane transport	2.58	.0003
Trehalose biosynthetic process	6.50	.002
Carbohydrate metabolic process	1.69	.004
Sulphate transport	4.65	.016

sensitive RNAseq rather than microarrays to assay gene expression, substantially increasing the number of genes examined as well as the use of a broader range of host and rhizobial strains. Further, Heath et al. (2012) focused on plants and strains collected from geographically proximate locations in southern France, whereas our experiments involved rhizobia from the two major lineages (*E. meliloti and E. medicae*) that often co-occur and form symbiosis with *M. truncatula*.

4.1 | Phenotypic genome × genome interactions

Consistent with previous work (e.g., Batstone, Dutton, Wang, Yang, & Frederickson, 2017; Heath & Tiffin, 2007; Larrainzar et al., 2014; Simms & Porter, 2012; Sugawara et al., 2013; Terpolilli, O'Hara, Tiwari, Dilworth, & Howieson, 2008) nodule number, nodule size and plant biomass were all significantly affected by the genotype of the host, the symbiont and their interaction. Interestingly, within the four hosts greater plant biomass, a trait associated with greater nitrogen fixation (Friesen, 2012) was always associated with greater average nodule mass and fewer total nodules (Figure 1). This result suggests that nodule mass, or the ability to obtain resources from a host once symbiosis is established, rather than nodule number, is a better proxy for the comparative benefits plants derive from symbiosis with particular rhizobial strains (Laguerre et al., 2012).

Although our data set is too small to rigorously test whether symbiont-related plant growth is correlated with relatedness among hosts, the most phylogenetically distinct host (HM340/R108) benefited far more from *E. meliloti*, whereas the other three hosts obtained equal or greater benefits from *E. medicae*. While we would not necessarily expect early transcriptional profiles (at 3 weeks) to predict the long-term phenotypic outcomes of the interaction, 517 of the focal genes with host by symbiont variation do match the phenotypic pattern of we measured (see below for a discussion of the enriched functions in this group). It is of course possible that the Host × Sym genes that do not match the phenotypic outcomes are also directly involved but represent more complex gene interactions than we tested for (i.e., epistasis) (Chou, Delaney, Draghi, & Marx, 2014; Phillips, 2008) or encode proteins with functionally important amino acid variation among hosts. Alternatively, these Host × Sym genes may reflect either expression changes associated with phenotypes not measured or variation without important phenotypic consequences (Ghalambor et al., 2015).

4.2 | Functions of the host- and symbiont-dependent transcriptome

One class of genes widely affected by host genotype (and symbiont strain) was those associated with defence response (e.g., NBS-LRR's). Although legumes suppress some defence mechanisms during nodule formation, other defence mechanisms are induced, and plant immunity plays an important role in the establishment and maintenance of functional symbiosis (Gourion, Berrabah, Ratet, & Stacey, 2015). Our results suggest not only that the expression of defence genes is altered by rhizobia inoculation (e.g., Kouchi et al., 2004; Lohar et al., 2006) but also that hosts differ strikingly in the extent to which they induce defence genes in nodules.

One aspect of the *Medicago-Ensifer* symbiosis is interesting here. Non-differentiated, reproductively capable rhizobia are only found in a small portion of the nodule near the meristem. The

majority of the nodule is filled with morphologically differentiated bacteria enclosed in plant cells, bacteroids; these bacteria are no longer reproductively competent and thus represent an evolutionary dead end (Denison & Kiers, 2011). Assuming the nodule transcriptomes we collected primarily reflect expression in the large portion of nodules filled with bacteroids, then the changes in defence genes we observe likely reflect changes in the bacteroid environment as opposed to the environment experienced by the reproductive rhizobia population. Differential expression of host genes that alter the bacteroid environment may represent mechanisms used by hosts to control differentiation and nitrogen fixation in bacteroids (Oono et al., 2011). Although not classified as "defence genes" in our GO analysis, recent work on a group of genes called NCR's is particularly fascinating in this context. Originally identified as defensins or antimicrobial peptides, these genes have undergone a massive expansion and diversification in *Medicago* and recent work suggests they regulate rhizobial differentiation into bacteroids and viability in the host (Guefrachi et al., 2014; Montiel et al., 2017; Pan & Wang, 2017; Shabab et al., 2016). These genes tend to be upregulated in the fixation zone (Roux et al., 2014), and here we found that hosts differ strongly in which suite of NCR genes they express (Figure 2).

Whereas defence genes were strongly overrepresented among genes with expression affected by the identity of host and the symbiont, genes with expression that was significantly affected by the interactions between host and symbiont (i.e., genome × genome) were enriched for processes that play important roles in the exchange of nitrogen and carbohydrates (i.e., the currency of the legume-rhizobium symbiosis) between the plant and the bacteria. That these focal genes may disproportionately be involved in acquisition of nitrogen is also supported by finding that these genes are overrepresented among genes upregulated in the nodule N-fixation zone (Table S5: based on data from Roux et al. (2014)) and were disproportionality upregulated in actively fixing nodules (Table S6: Lang & Long, 2015).

The cluster of H × S focal genes whose expression mimics the long-term phenotypic outcomes of the interaction (cluster 2 in Figure 4) provides an intriguing starting place for future exploration. For instance, glutamine synthetases are 17 times more likely to occur in this cluster than in the nodule transcriptome as a whole. This enzyme performs the first step in plant assimilation of fixed nitrogen (Betti et al., 2012; Lardi et al., 2016) and was also found to be upregulated by *M. truncatula* in tripartite interactions between rhizobia and mycorrhizal fungi (Afkhami & Stinchcombe, 2016). That this is the first gene in the pathway may not be coincidental. Mathematical models of flux control in metabolic pathways (Wright & Rausher, 2010) and empirical tests of these results (Olson-Manning, Lee, Rausher, & Mitchell-Olds, 2013; Olson-Manning, Strock, & Mitchell-Olds, 2015) suggest that the first step in a given pathway has a disproportionate influence on the flux through a pathway. The enrichment of malate (dicarboxylate) synthesis—the form of energy delivered to bacteroids to power fixation (Udvardi & Poole, 2013)—provides another obvious candidate.

4.3 | Do transcript data advance our understanding of the evolutionary ecology of symbiosis?

Our results reveal the extent to which nodule transcriptomes are affected by natural genetic variation in their rhizobial partners. However, non-molecular phenotypic data had already provided extensive evidence for host, symbiont and host × symbiont variation (Mytton, El-Sherbeeny, & Lawes, 1977; Simonsen & Stinchcombe, 2014), and many interesting aspects of the evolutionary ecology of mutualisms can clearly be investigated without molecular techniques. Further, transcriptome data alone cannot identify the genetic variants that alter regulatory cascades that cause the downstream functional differences that are observed; expression QTL analysis would be necessary to attain that level of resolution (Josephs, Lee, Stinchcombe, & Wright, 2015; Lowry et al., 2013; Majewski & Pastinen, 2011). It is worth asking, therefore, under what circumstances transcriptome data can provide new insight into ecological or evolutionary processes.

First, transcriptomes can generate a high-dimensional symbiotic phenotype that provides a glimpse into subcellular processes contributing to the long-term fitness outcomes of the interaction between specific genotypes. Progress in the identification of naturally variable genes and mechanisms involved in symbiotic resource mutualisms tends to lag behind other types of mutualisms (e.g., plant/pollinator) due in part to the challenge of identifying variation in processes occurring inside host tissues (Baskett & Schemske, 2015). Our analysis suggest that natural variation in symbiosis outcomes are related both to direct exchange of the currency of the symbiosis (nitrogen being transferred to the plant, dicarboxylates to the rhizobia) as well as putative antimicrobial peptides (NCRs and glycine-rich proteins) and redox environment in the nodule. This result supports the idea that hosts can alter resource allocation in response to symbiont identity (Heath & Tiffin, 2009; Oono et al., 2011) and sets the stage for future experiments directly manipulating these functions. Further, the finding that host defence genes vary in expression in response to symbiont identity raises the possibility that host genotype and symbiont strain might alter interactions with species that are not part of the symbiosis (Afkhami & Stinchcombe, 2016; Ossler, Zielinski, & Heath, 2015).

A second finding emphasized by our transcriptome results and others (Heath et al., 2012; Zhang et al., 2014) is the potential for the addition of biotic, abiotic and genetic context to inform gene function (Chamberlain, Bronstein, & Rudgers, 2014; Hoeksema & Bruna, 2015). Much progress in the field of molecular ecology has been driven by evolutionary ecologists borrowing tools and information (i.e., gene annotations and results from functional analyses) derived from a single model genotype to gain insight into ecological and evolutionary pattern and process. However, if the genomic basis of how an organism interacts with its environment varies among genotypes, then annotation based on function in a single context will likely miss important aspects of biology. For example, if we had only used A17 (HM101; the standard *M. truncatula* genotype used in molecular genetic analyses) to assess host

response to symbiont, we would have concluded that symbiont strain had a limited effect on nodule transcription. Our current results clearly show that the response is more complex than this. In this light, host genotype- and symbiont strain-associated variation in expression can complement gene expression atlases based on developmental stages, such as those available through the *Medicago* gene expression atlas (Benedito et al., 2008; He et al., 2009).

ACKNOWLEDGEMENTS

We thank N. Springer, A. Waters and K.D. Heath for helpful comments and B. Epstein and D. Nedveck for phenotyping assistance. Funding for this study was provided by NSF grant IOS-1237993.

DATA ACCESSIBILITY

Raw data have been deposited on NCBI sequence read archive: http://www.ncbi.nlm.nih.gov/sra/SRP077692. Additional data available from the Dryad Digital Repository: https://doi.org/10.5061/dryad.348k2 which contains the following: (i) the raw phenotype data and R code for the statistical models of that data, (ii) the processed read counts and the R source code for the DESEQ2 statistical analysis categorizing gene expression patterns, (iii) a tab-delimited text file that lists statistical and cluster category membership and normalized read counts for each sample for the ~27,000 genes expressed in the nodules, (iv) tab-delimited files containing statistical results for the full model run on both nodule and root samples and pairwise model results for each host contrast.

AUTHOR CONTRIBUTIONS

P.T./N.Y./M.S./R.S. designed the expression experiment, J.L. conducted it and C.L.C. extracted the RNA from the samples. L.T.B. conducted the phenotypic experiment L.T.B./J.G. conducted analyses and L.T.B./P.T. wrote the manuscript with feedback from the other authors.

ORCID

Liana T. Burghardt http://orcid.org/0000-0002-0239-1071

REFERENCES

Afkhami, M. E., & Stinchcombe, J. R. (2016). Multiple mutualist effects on genomewide expression in the tripartite association between *Medicago truncatula*, nitrogen-fixing bacteria and mycorrhizal fungi. *Molecular Ecology*, *25*, 4946–4962.

Akçay, E. (2015). Evolutionary models of mutualism. In J. L. Bronstein (Ed.), *Mutualism* (pp. 57–76). New York, NY: Oxford University Press.

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. *Bioinformatics*, *31*, 166–169.

Aparicio-Fabre, R., Guillén, G., Loredo, M., Arellano, J., Valdés-López, O., Ramírez, M., … Hernández, G. (2013). Common bean (*Phaseolus vulgaris L.*) PvTIFY orchestrates global changes in transcript profile response to jasmonate and phosphorus deficiency. *BMC Plant Biology*, *13*, 26.

Bai, Y., Meng, Y., Huang, D., Qi, Y., & Chen, M. (2011). Origin and evolutionary analysis of the plant-specific TIFY transcription factor family. *Genomics*, *98*, 128–136.

Baskett, C. A., & Schemske, D. W. (2015). Evolution and genetics of mutualism. In J. L. Bronstein (Ed.), *Mutualism* (pp. 77–92). Oxford, UK: Oxford University Press.

Batstone, R. T., Dutton, E. M., Wang, D., Yang, M., & Frederickson, M. E. (2017). The evolution of symbiont preference traits in the model legume *Medicago truncatula*. *New Phytologist*, *213*, 1850–1861.

Benedito, V. A., Torres-Jerez, I., Murray, J. D., Andriankaja, A., Allen, S., Kakar, K., … Udvardi, M. K. (2008). A gene expression atlas of the model legume *Medicago truncatula*. *The Plant Journal*, *55*, 504–513.

Betti, M., García-Calderón, M., Pérez-Delgado, C. M., Credali, A., Estivill, G., Galván, F., … Márquez, A. J. (2012). Glutamine synthetase in legumes: Recent advances in enzyme structure and functional genomics. *International Journal of Molecular Sciences*, *13*, 7994–8024.

Branca, A., Paape, T. D., Zhou, P., Briskinec, R., Farmer, A. D., Mudge, J., … Tiffin, P. (2011). Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume *Medicago truncatula*. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, E864–E870.

Brear, E. M., Day, D. A., & Smith, P. M. C. (2013). Iron: An essential micronutrient for the legume-rhizobium symbiosis. *Frontiers in Plant Science*, *4*, 359.

Bucciarelli, B., Hanan, J., Palmquist, D., & Vance, C. (2006). A standardized method for analysis of *Medicago truncatula* phenotypic development. *Plant Physiology*, *142*, 207–219.

Chamberlain, S. A., Bronstein, J. L., & Rudgers, J. A. (2014). How context dependent are species interactions? *Ecology Letters*, *17*, 881–890.

Chou, H. H., Delaney, N. F., Draghi, J. A., & Marx, C. J. (2014). Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLoS Genetics*, *10*(2), e1004149. https://doi.org/10.1371/journal.pgen.1004149

Dalton, D. A., Becana, M., Matamoros, M. A., & Udvardi, M. (2010). Recent insights into antioxidant defenses of legume root nodules. *New Phytologist*, *188*, 960–976.

Denison, R. F., & Kiers, E. T. (2011). Life histories of symbiotic rhizobia and mycorrhizal fungi. *Current Biology*, *21*, R775–R785.

Didelot, X., Walker, A. S., Peto, T. E., Crook, D. W., & Wilson, D. J. (2016). Within-host evolution of bacterial pathogens. *Nature Reviews Microbiology*, *14*, 150–162.

Epstein, B., Branca, A., Mudge, J., Bharti, A. K., Farmer, A. D., Mudge, J., … Tiffin, P. (2012). Population genomics of the facultatively mutualistic bacteria *Sinorhizobium meliloti* and *S. medicae*. *PLoS Genetics*, *8*, e1002868.

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., … Bateman, A. (2016). The Pfam protein families database: Towards a more sustainable future. *Nucleic Acids Research*, *44*, D279–D285.

Friesen, M. L. (2012). Widespread fitness alignment in the legume-rhizobium symbiosis. *New Phytologist*, *194*, 1096–1111.

Galibert, F., Finan, T. M., Long, S. R., Puhler, A., Abola, P., Ampe, F., … Batut, J. (2001). The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science*, *293*, 668–672.

Garcia, J. R., & Gerardo, N. M. (2014). The symbiont side of symbiosis: Do microbes really benefit? *Frontiers in Microbiology*, *5*, 1–6.

Ghalambor, C. K., Hoke, K. L., Ruell, E. W., Fischer, E. K., Reznick, D. N., & Hughes, K. A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature*, *525*, 372–375.

Gilbert, S. F., Bosch, T. C. G., & Ledón-Rettig, C. (2015). Eco-Evo-Devo: Developmental symbiosis and developmental plasticity as evolutionary agents. *Nature Reviews Genetics*, *16*, 611–622.

Gorton, A. J., Heath, K. D., Pilet- Nayel, M.-L., Baranger, A., & Stinchcombe, J. R. (2012). Mapping the genetic basis of symbiotic variation in legume-rhizobium interactions in *Medicago truncatula*. *G3 (Bethesda)*, *2*, 1291–1303.

Gourion, B., Berrabah, F., Ratet, P., & Stacey, G. (2015). Rhizobium-legume symbioses: The crucial role of plant immunity. *Trends in Plant Science*, *20*, 186–194.

Gubry-Rangin, C., Garcia, M., & Bena, G. (2010). Partner choice in *Medicago truncatula-Sinorhizobium* symbiosis. *Proceedings of the Royal Society B: Biological Sciences*, *277*, 1947–1951.

Guefrachi, I., Nagymihaly, M., Pislariu, C. I., Van de Velde, W., Ratet, P., Mars, M., … Alunni, B. (2014). Extreme specificity of NCR gene expression in *Medicago truncatula*. *BMC Genomics*, *15*, 712.

Guidot, A., Jiang, W., Ferdy, J. B., Thébaud, C., Barberis, P., Gouzy, J., & Genin, S. (2014). Multihost experimental evolution of the pathogen *Ralstonia solanacearum* unveils genes involved in adaptation to plants. *Molecular Biology and Evolution*, *31*, 2913–2928.

Haag, A. F., Arnold, M. F. F., Myka, K. K., Kerscher, B., Dall'Angelo, S., Zanda, M., … Ferguson, G. P. (2013). Molecular insights into bacteroid development during Rhizobium-legume symbiosis. *FEMS Microbiology Reviews*, *37*, 364–383.

He, J., Benedito, V. A., Wang, M., Murray, J. D., Zhao, P. X., Tang, Y., & Udvardi, M. K. (2009). The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, *10*, 441.

Heath, K. D., Burke, P. V., & Stinchcombe, J. R. (2012). Coevolutionary genetic variation in the legume-rhizobium transcriptome. *Molecular Ecology*, *21*, 4735–4747.

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

Heath, K. D., & Tiffin, P. (2007). Context dependence in the coevolution of plant and rhizobial mutualists. *Proceedings of the Royal Society of London B: Biological Sciences*, *274*, 1905–1912.

Heath, K. D., & Tiffin, P. (2009). Stabilizing mechanisms in a legume-rhizobium mutualism. *Evolution*, *63*, 652–662.

Herridge, D.F., Peoples, M.B., Boddey, R.M. (2008). Global inputs of biological nitrogen fixation in agricultural systems. *Plant and Soil*, *311*, 1–18.

Hoeksema, J. D., & Bruna, E. M. (2015). Context-dependent outcomes of mutualistic interactions. In J. L. Bronstein (Ed.), *Mutualism* (pp. 181–203). Oxford, UK: Oxford University Press.

Jones, E. I., Afkhami, M. E., Akçay, E., Bronstein, J. L., Bshary, R., Frederickson, M. E., … Friesen, M. L. (2015). Cheaters must prosper: Reconciling theoretical and empirical perspectives on cheating in mutualism. *Ecology Letters*, *18*, 1270–1284.

Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., … Hunter, S. (2014). InterProScan 5: Genome-scale protein function classification. *Bioinformatics*, *30*, 1236–1240.

Josephs, E. B., Lee, Y. W., Stinchcombe, J. R., & Wright, S. I. (2015). Association mapping reveals the role of purifying selection in the maintenance of genomic variation in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 15390–15395.

Kang, Y., Li, M., Sinharoy, S., & Verdier, J. (2016). A snapshot of functional genetic studies in *Medicago truncatula*. *Frontiers in Plant Science*, *7*, 1175.

Kouchi, H., Shimomura, K., Hata, S., Hirota, A., Wu, G. J., Kumagai, H., … Tabata, S. (2004). Large-scale analysis of gene expression profiles during early stages of root nodule formation in a model legume, *Lotus japonicus*. *DNA Research*, *11*, 263–274.

Krusell, L., Krause, K., Ott, T., Desbrosses, G., Krämer, U., Sato, S., … Udvardi, M. K. (2005). The sulfate transporter SST1 is crucial for symbiotic nitrogen fixation in *Lotus japonicus* root nodules. *The Plant Cell*, *17*, 1625–1636.

Laguerre, G., Heulin-Gotty, K., Brunel, B., Klonowska, A., Le Quéré, A., Tillard, P., … Lepetit, M. (2012). Local and systemic N signaling are involved in *Medicago truncatula* preference for the most efficient *Sinorhizobium* symbiotic partners. *New Phytologist*, *195*, 437–449.

Lang, C., & Long, S. R. (2015). Transcriptomic analysis of *Sinorhizobium meliloti* and *Medicago truncatula* symbiosis using nitrogen fixation deficient nodules. *Molecular Plant-Microbe Interactions*, *5*, 1–43.

Lardi, M., Murset, V., Fischer, H. M., Mesa, S., Ahrens, C. H., Zamboni, N., & Pessi, G. (2016). Metabolomic profiling of *Bradyrhizobium diazoefficiens*-induced root nodules reveals both host plant-specific and developmental signatures. *International Journal of Molecular Sciences*, *17*, 1–19.

Larrainzar, E., Gil-Quintana, E., Seminario, A., Arrese-Igor, C., & González, E. M. (2014). Nodule carbohydrate catabolism is enhanced in the *Medicago truncatula* A17-*Sinorhizobium medicae* WSM419 symbiosis. *Frontiers in Microbiology*, *5*, 1–7.

Larrainzar, E., Riely, B. K., Kim, S. C., Carrasquilla-Garcia, N., Yu, H. J., Hwang, H. J., … Cook, D. R. (2015). Deep sequencing of the *Medicago truncatula* root transcriptome reveals a massive and early interaction between Nod factor and ethylene signals. *Plant Physiology*, *169*(1), 233–265.

Licausi, F., Ohme-Takagi, M., & Perata, P. (2013). APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: Mediators of stress responses and developmental programs. *New Phytologist*, *199*, 639–649.

Lohar, D. P., Sharopova, N., Endre, G., Peñuela, S., Samac, D., Town, C., … VandenBosch, K. A. (2006). Transcript analysis of early nodulation events in *Medicago truncatula*. *Plant Physiology*, *140*, 221–234.

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*, 550.

Lowry, D. B., Logan, T. L., Santuari, L., Hardtke, C. S., Richards, J. H., DeRose-Wilson, L. J., … Juenger, T. E. (2013). Expression quantitative trait locus mapping across water availability environments reveals contrasting associations with genomic features in Arabidopsis. *The Plant Cell*, *25*, 3266–3279.

Majewski, J., & Pastinen, T. (2011). The study of eQTL variations by RNA-seq: From SNPs to phenotypes. *Trends in Genetics*, *27*, 72–79.

Montiel, J., Downie, J. A., Farkas, A., Bihari, P., Herczeg, R., Bálint, B., … Kondorosi, É. (2017). Morphotype of bacteroids in different legumes correlates with the number and type of symbiotic NCR peptides. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(19), 5041–5046.

Mytton, L. R., El-Sherbeeny, M. H., & Lawes, D. A. (1977). Symbiotic variability in *Vicia faba*. 3. Genetic effects of host plant, rhizobium strain and of host × strain interaction. *Euphytica*, *26*, 785–791.

Nova-Franco, B., Íñiguez, L. P., Valdés-López, O., Alvarado-Affrantranger, X., Leija, A., Fuentes, S. I., … Hernández, G. (2015). The micro-RNA172c – APETALA 2–1 node as a key regulator of the common bean- *Rhizobium* etli nitrogen fixation symbiosis. *Plant Physiology.*, *168*, 273–291.

Oldroyd, G. E. D., Murray, J. D., Poole, P. S., & Downie, J. A. (2011). The rules of engagement in the legume-rhizobial symbiosis. *Annual Review of Genetics*, *45*, 119–144.

Olson-Manning, C. F., Lee, C. R., Rausher, M. D., & Mitchell-Olds, T. (2013). Evolution of flux control in the glucosinolate pathway in *Arabidopsis thaliana*. *Molecular Biology and Evolution*, *30*, 14–23.

Olson-Manning, C. F., Strock, C. F., & Mitchell-Olds, T. (2015). Flux control in a defense pathway in *Arabidopsis thaliana* Is robust to environmental perturbations and controls variation in adaptive traits. *G3 (Bethesda)*, *5*, 2421–2427.

Oono, R., Anderson, C. G., & Denison, R. F. (2011). Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings. Biological Sciences/The Royal Society*, *278*, 2698–2703.

Ossler, J. N., Zielinski, C. A., & Heath, K. D. (2015). Tripartite mutualism: Facilitation or trade-offs between rhizobial and mycorrhizal symbionts of legume hosts. *American Journal of Botany*, *102*, 1332–1341.

Pan, H., & Wang, D. (2017). Nodule cysteine-rich peptides maintain a working balance during nitrogen-fixing symbiosis. *Nature Plants*, *3*, 17048.

Phillips, P. C. (2008). Epistasis - the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Reviews Genetics*, *9*, 855–867.

Reeve, W., Chain, P., O'Hara, G., Ardley, J., Nandesena, K., Bräu, L., … Howieson, J. (2010). Complete genome sequence of the *Medicago* microsymbiont *Ensifer (Sinorhizobium) medicae* strain WSM419. *Standards in Genomic Sciences*, *2*, 77–86.

Roux, B., Rodde, N., Jardinaud, M. F., Timmers, T., Sauviac, L., Cottret, L., … Gamas, P. (2014). An integrated analysis of plant and bacterial gene expression in symbiotic root nodules using laser-capture microdissection coupled to RNA sequencing. *Plant Journal*, *77*, 817–837.

Sachs, J. L. (2015). The exploitation of mutualisms. In J. L. Bronstein (Ed.), *Mutualism* (pp. 93–106). Oxford, UK: Oxford University Press.

Sachs, J. L., Skophammer, R. G., & Regus, J. U. (2011). Colloquium Paper: Evolutionary transitions in bacterial symbiosis. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 10800–10807.

Shabab, M., Arnold, M. F. F., Penterman, J., Wommack, A. J., Bocker, H. T., Price, P. A., … Walker, G. C. (2016). Disulfide cross-linking influences symbiotic activities of nodule peptide NCR247. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 10157–10162.

Simms, E. L., & Porter, S. S. (2012). Transcriptomic insights into mechanisms of symbiotic cooperation. *Molecular Ecology*, *21*, 4665–4668.

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

Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., … Tiffin, P. (2013). Candidate genes and genetic architecture of symbiotic and agronomic traits revealed by whole-genome, sequence-based association genetics in *Medicago truncatula*. *PLoS ONE*, *8*, e65688.

Starker, C. G., Parra-Colmenares, A. L., Smith, L., Mitra, R. M., & Long, S. R. (2006). Nitrogen fixation mutants of *Medicago truncatula* fail to support plant and bacterial symbiotic gene expression. *Plant Physiology*, *140*, 671–680.

Sugawara, M., Cytryn, E. J., & Sadowsky, M. J. (2010). Functional role of *Bradyrhizobium japonicum* trehalose biosynthesis and metabolism genes during physiological stress and nodulation. *Applied and Environmental Microbiology*, *76*, 1071–1081.

Sugawara, M., Epstein, B., Badgley, B. D., Unno, T., Xu, L., Reese, J., … Sadowsky, M. J. (2013). Comparative genomics of the core and accessory genomes of 48 Sinorhizobium strains comprising five genospecies. *Genome Biology*, *14*, R17.

Tadege, M., Wen, J., He, J., Tu, H., Kwak, Y., Eschstruth, A., … Mysore, K. S. (2008). Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. *The Plant Journal*, *54*, 335–347.

Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., … Town, C. D. (2014). An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics*, *15*, 312.

Terpolilli, J. J., O'Hara, G. W., Tiwari, R. P., Dilworth, M. J., & Howieson, J. G. (2008). The model legume *Medicago truncatula* A17 is poorly matched for N2 fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytologist*, *179*, 62–66.

Toft, C., & Andersson, S. G. E. (2010). Evolutionary microbial genomics: Insights into bacterial host adaptation. *Nature Reviews Genetics*, *11*, 465–475.

Trapnell, C., Pachter, L., Salzberg, S. L. (2009) TopHat: Discovering splice junctions with RNA-seq. *Bioinformatics*, *25*, 1105–1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., … Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols*, *7*, 562–578.

Udvardi, M. K., & Day, D. A. (1997). Metabolite transport across symbiotic membranes of legume nodules. *Annual Review of Plant Physiology and Plant Molecular Biology*, *48*, 493–523.

Udvardi, M., & Poole, P. S. (2013). Transport and metabolism in legume-rhizobia symbioses. *Annual Review of Plant Biology*, *64*, 781–805.

Vellosillo, T., Martínez, M., López, M. A., Vicente, J., Cascón, T., Dolan, L., … Castresana, C. (2007). Oxylipins produced by the 9-lipoxygenase pathway in Arabidopsis regulate lateral root development and defense responses through a specific signaling cascade. *The Plant Cell*, *19*, 831–846.

Wipf, D. D. D. D., Mongelard, G. G. G. G., van Tuinen, D. D., Gutierrez, L. L., & Casieri, L. L. (2014). Transcriptional responses of *Medicago truncatula* upon sulfur deficiency stress and arbuscular mycorrhizal symbiosis. *Frontiers in Plant Science*, *5*, 680.

Wright, K. M., & Rausher, M. D. (2010). The evolution of control and distribution of adaptive mutations in a metabolic pathway. *Genetics*, *184*, 483–502.

Young, N. D., Debellé, F., Oldroyd, G. E. D., Geurts, R., Cannon, S. B., Udvardi, M. K., … Roe, B. A. (2011). The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, *480*, 520–524.

Yue, M., Han, X., De Masi, L., Zhu, C., Ma, X., Zhang, J., … Schifferli, D. M. (2015). Allelic variation contributes to bacterial host specificity. *Nature Communications*, *6*, 8754.

Zhang, J.-Y., Cruz Carvalho MH, D. E., Torres-Jerez, I., Kang, Y., Allen, S. N., Huhman, D. V., … Udvardi, M. K. (2014). Global reprogramming of transcription and metabolism in *Medicago truncatula* during progressive drought and after rewatering. *Plant, Cell & Environment*, *37*, 2553–2576.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Burghardt LT, Guhlin J, Chun CL, et al. Transcriptomic basis of genome by genome variation in a legume-rhizobia mutualism. *Mol Ecol*. 2017;00:1–14. https://doi.org/10.1111/mec.14285