MOLECULAR ECOLOGY

Molecular Ecology (2012) **21**, 4735–4747 doi: 10.1111/j.1365-294X.2012.05629.x

Coevolutionary genetic variation in the legume-rhizobium transcriptome

KATY D. HEATH,* PATRICIA V. BURKE* and JOHN R. STINCHCOMBE†

Department of Plant Biology, University of Illinois, 250 Morrill Hall, 505 S. Goodwin Ave., Urbana, IL 61801, USA, †Department of Ecology and Evolutionary Biology and the Centre for the Analysis of Genome Evolution and Function, 25 Willcocks St., Toronto, ON, Canada M5S 3B2

Abstract

Coevolutionary change requires reciprocal selection between interacting species, where the partner genotypes that are favoured in one species depend on the genetic composition of the interacting species. Coevolutionary genetic variation is manifested as genotype × genotype (G × G) interactions for fitness in interspecific interactions. Although quantitative genetic approaches have revealed abundant evidence for G × G interactions in symbioses, the molecular basis of this variation remains unclear. Here we study the molecular basis of G × G interactions in a model legume-rhizobium mutualism using gene expression microarrays. We find that, like quantitative traits such as fitness, variation in the symbiotic transcriptome may be partitioned into additive and interactive genetic components. Our results suggest that plant genetic variation had the largest influence on nodule gene expression and that plant genotype and the plant genotype × rhizobium genotype interaction determine global shifts in rhizobium gene expression that in turn feedback to influence plant fitness benefits. Moreover, the transcriptomic variation we uncover implicates regulatory changes in both species as drivers of symbiotic gene expression variation. Our study is the first to partition genetic variation in a symbiotic transcriptome and illuminates potential molecular routes of coevolutionary change.

Keywords: coevolution, ecological genetics, genetic variation, genotype-by-genotype interaction, G × G, interaction, legume, mutualism, rhizobium, symbiosis

Received 14 December 2011; revision received 23 February 2012; accepted 1 March 2012

Introduction

Coevolution in species interactions is defined as reciprocal evolutionary change that results from partner-imposed natural selection (Janzen 1980). The driving force behind reciprocal selection is the genotype-by-genotype (G × G) interaction for fitness, i.e., when the genotypes favoured by selection in one species depend on the genotypes present in its partner population (Parker 1995; Heath & Tiffin 2007; Wade 2007). With G × G, the link between genotype and phenotype is not fixed, but instead depends on the genes present in an interacting partner of a different species. In practice, G × G for fitness may be detected by measuring traits across a range of partner genotype combinations and modelling the contributions of additive and interactive genetic effects. Although the quantitative genetic approach has successfully detected G × G interactions for fitness, particularly in host-symbiont interactions (Salvaudon *et al*. 2005; Hoeksema & Thompson 2007; Vale & Little 2009; Heath 2010), it is largely silent on the mechanistic basis of these interactions.

Little information exists on how metabolic or regulatory processes actually generate ecologically important G × G interactions, although this information can illuminate the genetic architecture of important symbiotic traits (Heath 2010). Assays of gene expression responses to laboratory-generated mutant mutualists, vs. wild type, have elucidated the downstream targets of particular

Correspondence: Katy D. Heath, Fax: 217 244 1224; E-mail: kheath@life.illinois.edu

© 2012 Blackwell Publishing Ltd

symbiosis genes (Barnett *et al.* 2004; Mitra *et al.* 2004). To our knowledge, however, no study has attempted to partition naturally occurring genetic variation in a symbiotic transcriptome among different combinations of host and symbiont genotypes—a necessary first step in determining how much transcriptional variation is owing to genetic variation in one of the partners, or to G × G interactions between interacting partners. Many previous studies attempting to compare gene expression between genotypes or lineages have unfortunately resulted in confounded environmental and genetic effects on expression variation (reviewed by Hodgins-Davis & Townsend 2009). Classical quantitative genetics demonstrates that a rigorous test for genetic variation requires controlled common garden experiments, in which genotypes are randomized to avoid unintentional covariance between genotypes and unmeasured micro-environmental factors (Falconer & Mackay 1996; Gibson 2008). The utility of this experimental approach for understanding the genetic basis of quantitative traits has been demonstrated by recent studies uncovering the genes and gene networks that underlie genetic variation for plasticity in response to salicylic acid in the model plant *Arabidopsis thaliana* (i.e., G × E interaction; van Leeuwen *et al.* 2007).

In the symbiotic mutualism between leguminous plants and rhizobia, partners exchange commodities that increase their individual reproductive fitness and also result in ecosystem-level impacts in both natural and managed communities (Graham & Vance 2003). In the symbiotic organs (nodules) on legume roots, rhizobia differentiate within plant cells into symbiotic bacteroids, which are capable of converting atmospheric N_2 to plant-usable forms. Fixed nitrogen (N), in the form of ammonium and amino acids, is exported across both bacterial membranes and the plant-derived peribacteroid membrane, which surrounds the symbiotic cells and serves as the location of metabolite transport between host and symbiont. In return, fixed carbon is transported to rhizobia in the form of amino acids, which are utilized by bacteroids to fuel the energy-intensive process of N fixation (Lodwig *et al.* 2003).

Legume-rhizobium interactions have recently attracted much evolutionary interest because they are ecologically and economically important and because genetic and physiological models of signalling, establishment and ongoing metabolite transfer already exist (Udvardi & Day 1997; Lodwig & Poole 2003; Riely *et al.* 2004; Ferguson *et al.* 2010). Evolutionary and ecological work has begun to reveal the selective forces acting on these interactions (Simms *et al.* 2006; De Mita *et al.* 2007; Heath & Tiffin 2009; Sachs *et al.* 2010; Oono *et al.* 2011), as well as how the ecological context can influence the trade of symbiotic benefits (Heath *et al.* 2010; Heath & Lau 2011). Despite the availability of numerous molecular and genetic resources (e.g., sequenced genomes, transformation methods, a HapMap project, mapping populations), the molecular basis of naturally occurring symbiotic variation in legume and rhizobium populations remains a major void in our understanding.

Here we study variation in the symbiotic transcriptome across a known G × G interaction between two legume hosts and two nitrogen-fixing rhizobium strains (Heath 2010), to evaluate how the symbiotic value of a given individual can shift depending on the partner genotype with which it interacts. Because we focus on genotypes from contemporary natural populations in their native range, our investigation sheds light on the genetic basis of symbiotic variation—the variation upon which coevolutionary selection is likely to act. We identify plant and rhizobium genes whose expression changed in response to either the G × G interaction or to partner genotype. Our results indicate that plant genotype had the largest impact on the nodule transcriptome and implicate genome-wide shifts in rhizobium gene expression in altering the trade of symbiotic benefits to feedback on partner fitness. We also find evidence that expression of regulatory loci and their downstream gene networks shift across these partner genotypes—consistent with a regulatory basis of natural symbiotic variation.

Materials and methods

Study system

For this experiment, we used families of *Medicago truncatula* (*Naut 1* and *Sals 4*) and *Sinorhizobium meliloti* (*Sals b* and *Sals c*) from populations *Nautique* and *Salses*, both on the Mediterranean coast of France, in the native range of the symbiosis (collecting and locations detailed in Heath 2010). These genotypes were chosen because they differed significantly in their effects on fitness benefits to both plant (growth, fruit production) and rhizobium (nodule number, nodule size) partners, and moreover because the plant G × rhizobium G interaction significantly impacted the benefits of symbiosis to plants, measured as both early leaf number and fruit number in a previous survey of 108 genotype combinations (Heath 2010). Although both rhizobium strains were sampled from the same population as plant genotype *Sals 4*, past work has demonstrated a lack of local adaptation of *Sals* rhizobia to *Sals* host plants (Heath 2010); therefore, the variation in gene expression described here is unlikely to reflect symbiotic variation arising from maladaptation of an allopatric host-symbiont combination.

© 2012 Blackwell Publishing Ltd

Greenhouse experiment and phenotypic analyses

The greenhouse experiment consisted of two plant and two rhizobium genotypes, grown in a fully factorial experiment with 24 replicates of each genotype combination, for a total of 96 plants. Seedlings and rhizobial inoculum were prepared as previously described (Heath 2010). Briefly, *M. truncatula* seeds were scarified, surface-sterilized and cold-stratified at 4 °C for 2 weeks before transplant to the greenhouse. Seedlings were transplanted into 6″ standard pots containing a 3:1 mix of Turface MVP (Profile Products LLC, Buffalo Grove, IL, USA) and Sunshine Mix #2 (Sun Gro, Bellevue, WA, USA) and were arranged in a randomized complete block design, with six blocks, in the Earth Sciences rooftop greenhouse facility at the University of Toronto. Rhizobium inoculum was prepared by growing rhizobium strains in tryptone yeast (TY) media (Somasegaran & Hoben 1994) for 2 days at 28 °C, and each plant was given 10^5 cells by pipetting. Plants were given adequate water throughout the experiment and were watered weekly with N-free 2X Fahreus nutrient solution (Somasegaran & Hoben 1994). We recorded early leaf number for each plant at 2 weeks after transplant to the greenhouse, which is known from previous work to be positively correlated with plant fruit number (Heath 2010). The complete phenotypic data set is available in Table S1a (Supporting information).

On the basis of preliminary estimates of per-plant RNA/cDNA yield, we harvested a subset of experimental plants (blocks 1–4, 64 plants) at 10 weeks post-transplant. Upon harvest, roots were shaken free of loose soil, and nodules were counted as they were removed from the root with forceps and plunged into liquid nitrogen for storage at −80 °C until RNA extraction. Aboveground biomass for all experimental plants was air-dried and weighed. For analysis of phenotypic data, we used ANOVA implemented in PROC MIXED (SAS Version 9.2; SAS Institute, Cary, NC, USA) to test for the fixed effects of plant genotype, rhizobium genotype, their interaction and block on dependent variables (early leaf number, nodule number and aboveground plant biomass). The general form of the statistical model was:

$$Y = \mu + \text{block} + \text{plant genotype} + \text{rhizobium genotype} + \text{plant genotype} \times \text{rhizobium genotype} + \text{error} \quad (1)$$

where Y is the phenotype of interest, μ is the model intercept (fit by default in most statistics packages), block describes the spatial position in the greenhouse, plant genotype is either *Naut 1* or *Sals 4*, rhizobium genotype is either *Sals b* and *Sals c*, and the interaction term describes how the effects of the plant genotype depend on rhizobial genotype, and vice versa.

Microarray assay and statistical analysis

We assayed gene expression using three biological replicates for each plant genotype × rhizobium genotype combination (four combinations) for a total of 12 chips. Detailed experimental protocols are available in the Supporting information (modified from Barnett *et al.* 2004; see Appendix S3, Supporting information). Using similar methods, Barnett *et al.* (2004) found close agreement of significant gene expression changes when comparing the Affymetrix method to qPCR of individual genes; where discrepancies were found, the results from microarray assays were more conservative. For each replicate, nodules from two plants in the same plant × rhizobium genotype combination (paired randomly) were pooled for RNA extraction using TRI-Reagent (Molecular Research Center Inc., Cincinnati, OH, USA) according to the manufacturer's suggested protocol. One cDNA library for each RNA extraction was constructed using SuperScript III (Invitrogen, Carlsbad, CA, USA). Resulting cDNA was fragmented with DNAse I (FisherThermo Scientific, Waltham, MA, USA), labelled and hybridized to Affymetrix Medicago Genechips according to Affymetrix kit procedures (Affymetrix, Santa Clara, CA, USA). Quality control, terminal labelling of cDNA, GeneChip hybridization and washing/staining, and initial analyses in Affymetrix GCOS software were performed by the Centre for the Analysis of Genome Evolution and Function at the University of Toronto (Appendix S3, Supporting information). Microarrays are well known to be sensitive to hybridization differences caused by nucleotide variation among genotypes (Kliebenstein *et al.* 2006); however, we found no evidence for overall differences in hybridization between the two rhizobium strains ($P = 0.34$) or between the two plant families ($P = 0.38$). Moreover, partner genotype-dependent changes in gene expression (i.e., plant genes that differ between rhizobium strains and rhizobium genes that differ between plant families) cannot be confounded with hybridization because the DNA-level variation causing the expression shift of interest lies in the genome of another species and thus cannot influence hybridization of the probe in question.

Statistical analyses were performed using JMP Genomics software (Version 5; SAS Institute). Data import, quality control and analysis were performed separately for *M. truncatula* and *S. meliloti* probe sets. Both data sets were subjected to median polish summarization, log2 transformation, RMA background correction and quantile normalization of raw probe intensity values present in .CEL files (Bolstad *et al.* 2003; Irizarry *et al.* 2003). Because it is standard to analyse array-based expression on a log2 scale (to stabilize variances between groups, account for mean-variance relationships

© 2012 Blackwell Publishing Ltd

for individual genes/probes and ensure symmetry in measures of up- and down-regulation: Quackenbush 2002; Gibson & Wolfinger 2004), our tests for interaction effects might be conservative as some interaction effects on the raw scale will be additive on the log scale (Berrington de González & Cox 2007).

Because we were interested in first understanding the overall impact of partner genetic variation on the nodule transcriptome, we used principal variance component analysis (PVCA; for details see Boedigheimer *et al.* 2008) to gain information on the total amount of expression variance explained by plant, vs. rhizobium, genotypes and the G × G interaction. PVCA was implemented in the correlation and principal components dialogue in JMP Genomics. Briefly, PVCA first fits principal components (PCs) to an expression data set. For each PC necessary to explain at least 90% of the total variance in the data set (six PCs for *S. meliloti* and eight for *M. truncatula*), a random effects ANOVA implemented in PROC MIXED is then used to estimate variance components for each factor of interest (i.e., family, strain, family × strain interaction, block). Finally, the overall contribution of each factor to the expression data set is estimated by weighting and summing each variance component across the relevant principal components.

We next used ANOVA to ask which individual probe sets (transcripts) responded significantly to plant genotype, rhizobium genotype, the plant genotype × rhizobium genotype interaction and block (all fixed). We did not include interactions with block because we were not interested in modelling fine-scale spatial variation in gene expression in the greenhouse environment. We accounted for multiple tests by controlling the false discovery rate (FDR) at 5% to balance rates of type I and type II error (Benjamini & Hochberg 1995; Verhoeven *et al.* 2005). For interaction-responsive genes, further pattern analysis of least-square means across plant and rhizobium genotype combinations was achieved by first scaling phenotypic and expression values to a mean of zero and variance of one using PROC STANDARD implemented in SAS software (Version 9.2; SAS Institute) and then performing hierarchical clustering in JMP Genomics. To further explore patterns of co-expression, Pearson correlations of means between significant genes and/or phenotypes were computed in the JMP Genomics cross-correlation dialogue. To test for over-representation of functional categories within our set of significant genes or within particular sets of responsive genes, relative to the entire set of genes on the chip, we used Fisher's exact tests implemented in the gene set enrichment dialogue in JMP Genomics (with FDR = 0.05). The full data set is deposited at GEO (GSE29027; http://www.ncbi.nlm.nih.gov), and all statistically significant probe sets are available in Appendix S1 (Supporting information).

Gene identification and annotation

Differentially expressed *S. meliloti* probe sets were matched to locus tags or systematic names by blasting the Affymetrix probe sequences (11 25-mer probes per probe set) against the *S. meliloti* gene coding sequences available from the RhizoGATE website (Becker *et al.* 2009) (http://www.cebitec.uni-bielefeld.de/CeBiTec/rhizogate/; downloaded 7 September 2010). Our BLAST criterion (e-value < 0.01) was parsed for two categories of matched probes: perfect (25 nucleotides per probe matched), or less than perfect (17–24 matched). For most (1899/2534) probe sets, all 11 probes per probe set matched perfectly to a single gene; the remaining data were filtered to include only those probe sets with at least nine perfectly matched probes. When probes from a single probe set matched two sequential genes or two genes within a gene family, or vice versa when a single gene matched two probe sets, a single match corresponding to the greater number of perfectly matched probe sets was chosen to facilitate downstream functional assignments and associated statistical analyses. If two matches were equal, the match that preserved synteny between the genes and probe sets was chosen. Our pipeline resulted in matching with high confidence 81% (2061/2534) of significant *S. meliloti* probe sets to locus tags. Annotations and COG functional category assignments for matched probe sets were taken from the RhizoGATE IGetDB database. The remaining significant probe sets all map perfectly to the genome, but not to the present 6225 genes, as probe sets were originally designed for 8305 predicted genes and now include partial hits to coding sequences and intergenic regions.

Our matching criteria were less stringent for the plant than for the bacterium because the genome is not finished, and probe sets were designed for an early version. Differentially expressed *M. truncatula* probe sets were matched to version 3.5 of the *M. truncatula* genome (Mt 3.5; IMGAG) by blasting probes for each probe set against the gene coding sequences available from the Medicago HapMap project website (http://www.medicagohapmap.org/?genome; downloaded 27 December 2010). We initially matched 44% of probe sets using BLAST criterion (e-value < 0.02) and filtering the list of probe sets to include only those with a minimum of eight perfectly matched 25-nucleotide probes, while tabulating for each probe set the number of perfectly matched probes, as well as the number of probes with 24- or 23-nucleotide matches for each gene. An additional 54% of probe sets were then matched to release 10 of the *M. truncatula* Gene Index tentative consensus

© 2012 Blackwell Publishing Ltd

sequences (MtGI TCs) using mapping files kindly supplied by He *et al.* (2009) using the same initial criterion. Many probe sets perfectly matched MT3.5 genes (36%) or MTGI10 TCs (80%) using this pipeline. As with ambiguous cases in *S. meliloti,* the highest-scoring gene or TC (based on either the number of probes with perfect hits or the best score for imperfectly matched probes) was selected to match one gene or TC to one probe set for downstream analysis. For equivalent matches, Mt3.5 genes were favoured over TCs. Of the initial 615 significant *M. truncatula* probe sets, 591 (96%) were matched to either a Mt3.5 gene (247 or 40%) or to a MTGI10 TC (344 or 56%). Plant genes were grouped into MapMan plant functional categories using the Mercator tool (http://www.gabipd.de/projects/MapMan).

Results

Genetic variation modulates the nodule transcriptome

We compared gene expression in each of four combinations of *Medicago truncatula* families (inbred lines; Heath 2010) and *Sinorhizobium meliloti* strains using Affymetrix Medicago GeneChips. We used both principal variance components analysis (PVCA) and ANOVA to understand how the entire transcriptome and individual genes responded to differences between plant families, between rhizobium strains and to the plant family × rhizobium strain (G × G) interaction. We were successful in minimizing uncontrolled environmental variance and detecting significant genetic and G × G variation in the nodule transcriptome (Table 1) and in symbiotic phenotypes (Table S1b, Supporting information). The percent of variance in gene expression accounted for by terms in our ANOVA model differed between the two species (Table 1). Interestingly, while rhizobium gene expression was sensitive to genetic differences in the plant host, and the effect of G × G was significant, plant gene expression responded much less to rhizobium strain and the family × strain interaction (Table 1). It should be noted that the sets of strain-responsive rhizobium probes and family-responsive plant probes (Table 1) should be interpreted with caution as these hybridization differences might result from nucleotide variation at probe sites.

Multiple lines of evidence indicate that modelling the G × G interaction significantly improved our understanding of the causes of rhizobium gene expression variation in this experiment. The distribution of *P*-values for the plant family × rhizobium strain interaction term across all rhizobium genes (Fig. S1, Supporting information) reveals an excess of low values, which is further supported by the number of tests that remain significant after FDR correction (see Table 1 and results below). Moreover, likelihood ratio tests of treatment effects in the PVCA analysis (see Appendix S4, Supporting information) indicated that plant, strain and their interaction all significantly affected rhizobium gene expression, whereas the interaction term did not significantly affect plant gene expression. All significant probe sets and annotations are presented in Appendix S1 (Supporting information).

The interaction-responsive transcriptome implicates plasmid-borne rhizobium gene expression and energy generation in beneficial interactions

Because plant and rhizobium fitness benefits in symbiosis often depend on the interaction of partner genotypes, and such G × G for fitness is the force driving coevolutionary change, we were most interested in the particular genes whose expression changed depending on the plant family × rhizobium strain interaction. In *S. meliloti,* 155 genes (or ~2.5% of the genome) responded to the G × G interaction (i.e., were interaction-responsive). Functional analysis did not reveal any

Table 1 The percent of expression variance (from PVCA), number of significant probe sets (from ANOVA) and number of matched genes for the effect of plant family, rhizobium strain, family × strain (G × G) interaction, and block on *Sinorhizobium meliloti* and *Medicago truncatula* gene expression

	Rhizobium gene expression			Plant gene expression		
Effect	% variance	No. of probe sets	No. of genes	% variance	No. of probe sets	No. of genes
Plant family	25.6	1484	1244	29.7	612	588
Rhizobium strain	50.9	1558	1216	3.7	5	5
Family × Strain	3.2	194	154	0.2	1	1
Block	10.4	128		30.5	124	
Residual	20.3			31.3		
Total*		2534	2061		719	591

PVCA, principal variance component analysis.

*Many probe sets responded to more than one model effect; these are represented once in the total.

© 2012 Blackwell Publishing Ltd

COG categories to be significantly under- or over-represented in the set of significant interaction genes, relative to the entire set of *S. meliloti* genes on the chip (Appendix S2, Supporting information).

To identify patterns of co-expression, we used hierarchical clustering to group the interaction-responsive genes according to their mean expression level in each of the four genotype combinations—resulting in six well-defined clusters (see Appendix S1, Supporting information; Fig. 1A). Furthermore, cluster analysis combined with cross-correlation analyses with early leaf number across four genotype combinations indicated that the genes positively associated with plant growth were located primarily in cluster I (Fig. 1B, C; Appendix S1, Supporting information). By contrast, genes with expression levels negatively correlated with plant growth were located in clusters IV and V (Fig. 1C; Appendix S1, Supporting information). Because these clusters contain genes whose expression is most closely correlated with plant symbiotic benefits, they likely contain the best candidates for mediating the trade of symbiotic benefits between these partners.

Interestingly, most genes in cluster I were located on the symbiotic plasmids (pSymA and pSymB), whereas genes in clusters IV and V were shifted away from pSymB and towards the chromosome ($\chi^2_{\text{d.f.}=2} = 17.2$, $P = 0.0002$; Fig. 2). Previous research has implicated the down-regulation of chromosomal gene expression in N-fixing bacteroids upon switching to the symbiotic lifestyle (Barnett *et al.* 2004). Our results indicate that the degree to which this genome-wide shift occurs is both genetically variable among compatible partners in nature and correlated with potential fitness benefits for plants. Both suggest that this shift in genome-wide rhizobium expression is ecologically and evolutionarily important in contemporary legume-rhizobium communities. Moreover, the set of differentially expressed genes we identified is largely unique from those identified in previous studies: the vast majority (80%) do not overlap with genes known to differentiate the free living from symbiotic lifestyles or to differ in mutant nodules (see annotations in Appendix S1, Supporting information). In addition, the majority of the genes that we have newly associated with symbiotic variation have

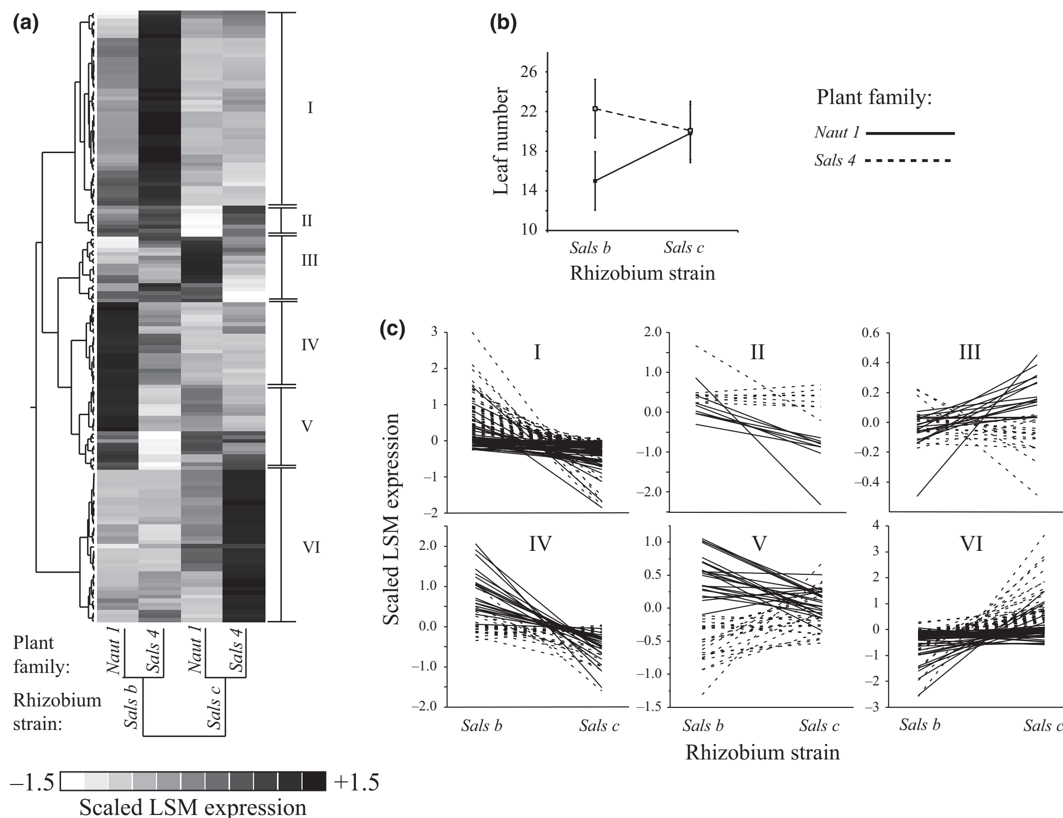

Fig. 1 Phenotypic and transcriptomic responses of *Sinorhizobium meliloti* to the plant genotype × rhizobium genotype interaction: (a) Hierarchical clustering of interaction-responsive rhizobium genes according to their scaled lsmeans expression profiles in four possible pairwise combinations of two plant families (*Naut 1*, *Sals 4*) and two rhizobium strains (*Sals b*, *Sals c*), (b) reaction norms (lsmeans with SE) of early plant performance (leaf number at 2 weeks) of the two plant families with each of two rhizobium strains and (c) plotted reaction norms of scaled expression lsmeans for genes from clusters I–VI.

© 2012 Blackwell Publishing Ltd

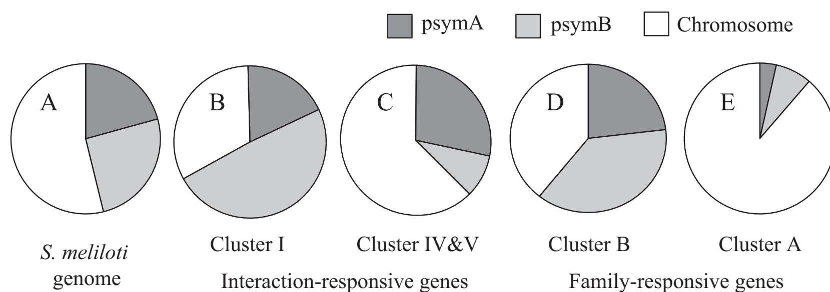

Fig. 2 Proportion of genes located on the chromosome, vs. symbiotic mega-plasmids pSymA and pSymB, for the entire set of *Sinorhizobium meliloti* genes on the chip (pie A), vs. those associated with either increased plant benefits (pie B: interaction-responsive cluster I and pie D: family-responsive cluster B), or decreased plant benefits (pie C: interaction-responsive clusters IV and V and pie E: family-responsive cluster A.

not been studied and have only minimal functional annotation.

Given the large number of interaction-responsive genes with minimal annotation, what changes in known genes or pathways are likely to mediate G × G interactions? Our results, combined with previous studies on the molecular regulation of nodule formation and root structure, suggest a role for 2-phenylacetic acid (PAA) metabolism, although other unknown pathways could of course also contribute. In particular, Cluster I contains all the genes involved in PAA degradation. PAA is a plant auxin that stimulates formation of lateral root primordia and lateral root growth more efficiently than indole-3-acetic acid (IAA; Wightman *et al.* 1980). Ding & Oldroyd (2009) argue that a balance of auxin (IAA and others), cytokinins and abscisic acid determines the development of nodules vs. lateral roots, and overproduction of IAA by transgenic *S. meliloti* led to increased numbers of functional nodules and increased root mass (Pii *et al.* 2007). Therefore, PAA metabolism has the potential to influence symbiotic benefits via alterations to nodule numbers, nodule size or both.

A pSymA-located aerobic form of formate dehydrogenase (*fdoH*) was also found in interaction-responsive cluster I. This enzyme is distinct from the chromosomal formate dehydrogenase (*fds*), which is a major energy source for anaerobic respiration in bacteria and was present in our family-responsive gene list. While *fdoH* is constitutively expressed in other bacteria (Abaibou *et al.* 1995), we found that its expression by *S. meliloti* is genetically variable in symbiosis and is positively correlated with plant performance (Appendix S1, Supporting information), suggesting that this enzyme may be important for microaerobic respiration in conjunction with N fixation. While nitrogenase itself was not present in our set of interaction-responsive genes, these results might suggest that naturally occurring symbiotic variation in the mechanisms generating energy in support of N fixation is important for altering the trade of benefits in symbiosis.

By contrast, interaction-responsive clusters IV and V consist of genes whose expression is negatively correlated with plant growth and contain chromosomal genes involved in basic metabolism and growth. Clusters IV and V include genes of the citric acid (TCA) cycle (*sucA, sdhC, mdh*) and amino acid synthesis (*asd, aroB*), which were found to be down-regulated in symbiosis in previous studies (Barnett *et al.* 2004; Becker *et al.* 2004). Most intriguing may be *exsI*, a putative transcriptional regulator from the *exs* gene cluster involved in production of low-molecular-weight succinoglycan (York & Walker 1997), which is a cell wall exopolysaccharide important in the establishment and maintenance of symbiosis (Gibson *et al.* 2008; Rinaude & Gonzalez 2009).

Despite the interactive effects of rhizobium genotypes on plant performance (Fig. 1), only one *M. truncatula* gene responded significantly to the G × G interaction. This gene (Medtr6g012630.1) appears to be one of 43 *M. truncatula* proteins that contain K homology (KH) domains, common RNA-binding motifs. The pattern of expression at Medtr6g012630.1 was negatively correlated (although not significantly; $r_{\text{d.f.}=4} = -0.83$; $P = 0.1697$) with plant growth and positively correlated with genes in cluster IV (Appendix S1, Supporting information), corroborating a potential role in less-beneficial interactions. Although the function of this gene in *M. truncatula* is unknown, KH proteins are known to play important regulatory roles as RNA-binding proteins in plants (Cheng *et al.* 2003; Lorkovic 2009). Furthermore, co-expression patterns mined from the *M. truncatula* Gene Expression Atlas (MtGEA; He *et al.* 2009) indicate that a MAP kinase (Medtr4g087620.1) is co-expressed and highly correlated with our interaction-responsive gene Medtr6g012630.1 across root nodule time-course experiments. Although Medtr4g087620.1 did not respond significantly in our study, MAP kinases are involved in regulatory signal cascades in all eukaryotes (Rodriguez *et al.* 2010). These data suggest the interaction-responsive KH domain gene Medtr6g012630.1 as

© 2012 Blackwell Publishing Ltd

an interesting candidate for regulating symbiotic benefit exchange in this interaction.

Plant-responsive rhizobium expression implicates plant regulatory variation

Our phenotypic data indicated that plant family *Sals 4* produced significantly fewer nodules and had significantly increased early growth compared to *Naut 1* (Appendix S1, Supporting information; Fig. 3A, B), suggesting that *Sals 4* nodules were more beneficial. We then examined how changes in rhizobium gene expression in response to different plant families might mediate these phenotypic differences. We separated the 1244 family-responsive rhizobium genes into those genes upregulated with family *Naut 1* (cluster A), vs. those upregulated with *Sals 4* (cluster B; Fig. 3C). Genes upregulated with *Naut 1* were enriched for COG categories O, E, L, C, J, U and F (i.e., housekeeping functions; Appendix S2, Supporting information), while those upregulated with *Sals 4* were significantly enriched for categories X (unknown/no COG) and K (transcription)—suggesting that a large portion of genes upregulated in more beneficial interactions remains unknown. KEGG pathway analysis (http://www.genome.jp/kegg/) corroborated this result: only 41 (8.2%) of cluster B genes, which are up-regulated in symbiotic interactions that improve plant performance, play known roles in metabolic pathways, compared to 213 (28.7%) of cluster A genes ($\chi^2_{d.f.=1}$ = 36.0, $P < 0.0001$).

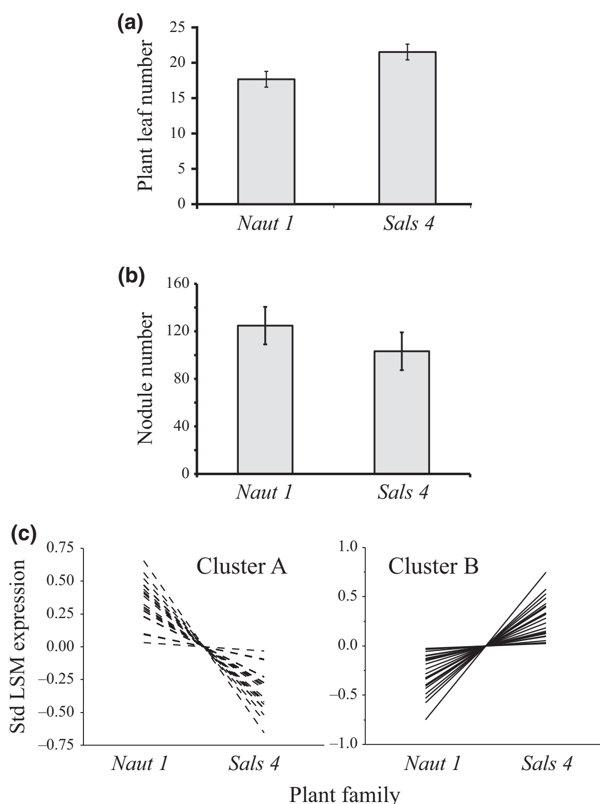

Fig. 3 Phenotypic and transcriptomic responses of *Sinorhizobium meliloti* to two different plant genotypes: (a) least-square means (+SE) of early plant performance (leaf number at 2 weeks) for two plant families (*Naut 1*, *Sals 4*) in our experiment, (b) number of symbiotic nodules at harvest for two plant families and (c) reaction norms of standardized expression lsmeans for plant family-responsive genes whose expression increased with *Naut 1* (left; cluster A) or *Sals 4* (right; cluster B).

The plant-responsive rhizobium transcriptome contained large numbers of putative or known rhizobium regulatory loci, although the functions of most of these loci and how they might regulate plant symbiotic benefits remain unclear. Notably, cluster B contained *aniA*, a regulator of rhizobial carbon flux necessary for efficient N fixation in symbiosis (Povolo & Casella 2000), *nfeD*, necessary for nodulation competitiveness (Garcia-Rodriguez & Toro 2000), and *selD* for selenocysteine of the active site of formate dehydrogenase (*fdoH*, *fdoI*, *fdhE*, and *fdsB*, *fdsG*) mentioned above. The putative extracytoplasmic function sigma factors *rpoE1*, *rpoE7* and *rpoE8* may be involved in plant-rhizobium communication and regulation. Cluster A, by contrast, contained *sigA*, a master regulator of housekeeping functions, along with numerous regulators of exopolysaccharide production (*expR*, *exoS*, *exoI*, *feuQ*; McIntosh *et al.* 2008; Griffitts *et al.* 2008; Chen *et al.* 2009). Cluster A also contained the chromosomal *rpoN* (*ntrA*) gene, a second master regulator involved in diverse housekeeping and symbiotic functions in a number of rhizobium species (Ronson *et al.* 1987; Barnett *et al.* 1998; Salazar *et al.* 2010).

Coordinate with the regulatory changes described above, rhizobia in the nodules of plant genotype *Sals 4* appeared to be performing more functions in support of N fixation, more transcription of genes on the symbiotic plasmids and less of numerous other housekeeping functions. For example, rhizobia in *Sals 4* nodules also upregulated a number of transporters that might affect the efficiency of symbiosis, including the *kup2* potassium uptake system (Dominguez-Ferraras *et al.* 2009) and the *sma1755* and *smb20356* cation channels (Appendix S1, Supporting information). Like interaction-responsive cluster I, most genes upregulated with *Sals 4* were located on the symbiotic plasmids, whereas cluster A contained an overwhelming majority of chromosomal genes ($\chi^2_{d.f.=2}$ = 342.0; $P < 0.0001$; Fig. 2). Again this suggests that the degree to which bacterial transcription shifts from the chromosome to the symbiotic plasmids is genetically variable, strongly influenced by the plant genome and important for plant benefits in symbiosis.

Interestingly, the aforementioned shifts in rhizobium gene expression corresponded to increased detection of seven *M. truncatula* nodule-specific cysteine-rich peptides

© 2012 Blackwell Publishing Ltd

(NRCs) in plant family *Sals 4*, compared to *Naut 1*. NRCs have recently been shown to control the terminal differentiation of rhizobia into bacteroids (Mergaert *et al.* 2003; Alunni *et al.* 2007; Van de Velde *et al.* 2010); therefore, these results might suggest that naturally occurring plant variation at NRCs effects major shifts in rhizobium gene expression. Importantly, any such variation (plant genes that differ between plant genotypes) might be caused by either differences in transcript abundances or by differential hybridization owing to SNPs. Either explanation offers interesting candidates for further exploration; nevertheless, a larger set of plant genotypes would also be necessary to rigorously correlate variation at plant NRCs with shifts in rhizobial gene expression.

Strain-responsive M. truncatula *genes also implicate regulatory variation*

Five *M. truncatula* genes responded significantly to rhizobium strain (including the interaction-responsive Medtr6g012630.1 discussed above), and we were able to assign putative functions to three of the four remaining genes. First, BG648909.1 is one of 11 *HAP3* genes in *M. truncatula* and is, therefore, a member of the CCAAT-binding heterotrimeric transcriptional regulator family commonly found in eukaryotes (Chodosh *et al.* 1988; Siefers *et al.* 2009). Other genes in the *HAP* family have been implicated in symbiotic regulation, including *MtHAP2-1,* a key transcriptional regulator of nodule meristem development (Combier *et al.* 2006; Siefers *et al.* 2009; Zanetti *et al.* 2010). Second, Medtr1g050520.1 is highly similar to a protein kinase/protein phosphatase 2C in *Arabidopsis thaliana* (AT2G40860)—suggesting that it too may be involved in signalling cascades. Third, Medtr5g011980.1 contains a plant lipid transfer protein with a bifunctional inhibitor domain. Such proteins may be involved in defence, pollination and germination (Lin *et al.* 2005), although their function in symbiosis is not understood. Our sample of significant strain-responsive plant genes, while too small to reach a definitive conclusion, suggests that upstream changes in plant regulatory networks might underlie symbiotic variation.

Discussion

Using genome-wide expression analysis to simultaneously assay plant and rhizobium partners in symbiosis, we found evidence of a major reorganization of the legume-rhizobium transcriptome in response to natural genetic variation in both partner species. We identified major shifts from chromosome-dominated to symbiotic plasmid-dominated rhizobial gene expression across genotype combinations exhibiting a strong $G \times G$ interaction, corresponding to decreased housekeeping functions and increased energy metabolism in support of N fixation and correlated with increased symbiotic benefits for plants. Partner-dependent plant gene expression also suggested an important role for regulatory variation in mediating symbiotic benefits. These results represent the first attempt to partition naturally occurring coevolutionary genetic variation in a symbiotic transcriptome.

Genetic control of the symbiotic transcriptome

We found that plant genotype exerted the largest overall influence on nodule gene expression, at least for the four host-symbiont genotype combinations included in our experiment. Most notably, the genetic identity of the plant host altered the expression of nearly a quarter of the rhizobium genome. Because the underlying DNA-level differences exist in the plant genome, such rhizobium expression changes are not attributable to differential hybridization among rhizobial strains owing to sequence variation. Many of these changes in rhizobium gene expression likely reflect the differential influence of plant genotypes on the developmental switch from undifferentiated cells in the infection thread to N-fixing bacteroids, as evidenced by a large shift between chromosomal and symbiotic plasmid expression across plant families and owing to the plant family $\times$ rhizobium strain interaction. Because this shift is genetically variable among naturally occurring genotypes and is correlated with plant symbiotic benefits, it may be subject to coevolutionary selection.

Genetic variation for major metabolic shifts exists in natural legume-rhizobium communities

For *Sinorhizobium meliloti*, the shift from a free-living, saprophytic to an intracellular symbiotic lifestyle involves shutting down many basic metabolic processes on the chromosome, including nitrate assimilation, growth and cell division, while switching on genes for micro-aerobic respiration and N fixation on the symbiotic plasmids (e.g., Barnett *et al.* 2004). The shifts in plant and rhizobium gene expression we detected across even the small set of genotype combinations included in our study suggest that one mechanism underpinning natural genetic variation for symbiotic benefits is the degree to which nodule rhizobia successfully transition to the symbiotic lifestyle. Because these genome-wide shifts arise from the $G \times G$ interaction, they are the result of heritable, nucleotide-level variation in the genomes of both partners and are, therefore, available to coevolution. The precise locations of these

© 2012 Blackwell Publishing Ltd

nucleotide changes remain unknown, although our results suggest an important role for upstream, regulatory variation (see below).

Phenotypic and transcriptomic shifts in response to plant family (i.e., nodule size, nodule number, gene expression) are likely to have important impacts on rhizobium fitness. Nodule size (weight or length) is positively correlated with the number of reproductive rhizobium cells within the nodule (Heath & Tiffin 2007; Oono *et al.* 2011). A previous microcosm study found that nodule number predicted rhizobium genotype frequencies in future generations (Heath & Tiffin 2009), perhaps not surprising as infection of each nodule (regardless of size) results in at least 10^5 rhizobium offspring (Oono *et al.* 2011). Thus, selection on rhizobia would be expected to favour *both* numerous and large nodules; however, past work on a large set of genotype combinations suggests a strong genetic trade-off between these two traits ($r_g = -0.8379$, $n = 106$, $P < 0.0001$; data from Heath *et al.* 2010). Data from the two plant families used here are consistent with this trade-off: *Naut 1* produced more nodules (Fig. 3B; Table S1b, Supporting information; $P = 0.0346$), and in previous work *Naut 1* produced significantly shorter nodules than *Sals 4* (mean length ± SE: 2.4 ± 1.6 mm for *Naut 1* vs. 3.0 ± 1.7 mm for *Sals 4*; $t_{d.f. = 18} = 2.58$, $P = 0.0187$; data from Heath *et al.* 2010). If rhizobial fitness is more strongly determined by nodule number, then our results suggest that expression of genes in Cluster A (Fig. 3C) is most associated with rhizobial fitness and that these genes differ from those whose expression is associated with plant performance and fitness (Cluster B; Fig. 3C). By contrast, if rhizobial fitness is more strongly determined by nodule size (and rhizobium offspring per nodule), then shared genes are associated with both plant and rhizobial fitness (Cluster B; Fig. 3C). Determining whether rhizobial fitness in natural populations is determined more strongly by nodule number or nodule size is an important challenge, with implications for the genetic basis of rhizobial adaptation and plant-rhizobium genetic conflict in the symbiotic interaction (Heath *et al.* 2010).

The regulatory nature of coevolutionary genetic variation

Kliebenstein (2010) has recently argued for two competing hypotheses to explain conditional genetic variation (e.g., G × E or G × G interactions). In the first, natural variation exists for the response of regulatory networks among environments, such that the same phenotype-encoding (structural or enzymatic) genes are regulated differently under different conditions. In the second hypothesis, the phenotype-encoding genes themselves differ between environments, with different genes expressed in different environments to cause a conditional phenotype. These alternative hypotheses have different implications for trait evolution. Because the same regulatory genes underlie the phenotype in multiple environmental conditions, variation at those genes has the potential to generate pleiotropy and thus genetic correlations across environments (Des Marais & Juenger 2010). By contrast, nucleotide variation in the conditionally expressed genes themselves would mean that genetic variation is expressed independently within each condition. In reality, these hypotheses are not mutually exclusive, and most genetic variation is probably attributable to both regulatory and structural gene changes. G × G interactions complicate the already-complicated plasticity (G × E) worldview, as here the environment itself evolves (Wolf *et al.* 2004; Wade 2007).

Our study explicitly tested for conditional variation in the nodule transcriptome, with symbiotic partner genotype as the environment in our experiment. Because of their low expression levels, regulatory loci are notoriously difficult to detect in global expression studies (Janga & Contreras-Moreira 2010). Nevertheless, we found evidence for numerous major regulatory shifts in our experiment contributing to partner- and interaction-dependent phenotypic variation. Our results might, therefore, support the gene regulation hypothesis, which would suggest that upstream variation in plant and rhizobium regulatory networks should generate genetic correlations across partner genotypes. Indeed, a previous study estimating such genetic correlations across partner genotypes found an abundance of positive correlations, but that their significance and magnitude depended on the particular genotypes (Heath 2010). The genetic correlation of plant fitness across the two strains included in the current study was significantly less than one ($r_g = 0.18$, $P < 0.001$; data from Heath 2010)—suggesting that magnitude or direction of selection on plants would differ between these two strains (e.g., Fig. 1). Because our study is the first of its kind, it remains to be determined whether major regulatory shifts like we found underlie other cases of G × G for this symbiosis.

Conclusions

Integrating evolutionary models of cooperation (Trivers 1971; Bull & Rice 1991) and the geographic mosaic of coevolution (Thompson 2005) with the genetic and physiological dynamics from experimental settings presents a great challenge in coevolutionary research. Understanding the molecular underpinnings of natural variation for partner fitness benefits is an important goal, as this is the variation upon which selection acts.

© 2012 Blackwell Publishing Ltd

Perhaps the best design for addressing both the amount of genetic variation in expression and the underlying genetic basis of it would be a nested association mapping population (McMullen *et al*. 2009; Scoville *et al*. 2009), although expression studies on that scale remain prohibitively expensive. We expect that combined quantitative genetic, genomic and transcriptomic approaches utilizing a larger set of symbiotic genotypes will further illuminate the sources of natural genetic variation and evolutionary trajectories of these ecologically important interactions.

Acknowledgements

KDH was supported by a postdoctoral fellowship from the Department of EEB at the University of Toronto. Our work was also supported by NSERC Canada, the Canadian Foundation for Innovation and the School of Integrative Biology at the University of Illinois Urbana-Champaign. We thank K. Samis, A. Simonsen, B. Campitelli, B. McGoey, A. Stock and A. Gorton for assistance with phenotypic data and nodule harvests, T. Nyugen for technical assistance with GeneChips, J. He and M. Barnett for advice on protocols and bioinformatics, and Richard Abbott and two anonymous reviewers for helpful comments on the manuscript.

References

Abaibou H, Pommier J, Benoit S, Giordano MA (1995) Expression and characterization of the *Escherichia coli fdo* locus and a possible physiological role for aerobic formate dehydrogenase. *Journal of Bacteriology*, **177**, 7141–7149.

Alunni B, Kevei Z, Redondo-Nieto M *et al*. (2007) Genomic organization and evolutionary insights on GRP and NCR genes, two large nodule-specific gene families in *Medicago truncatula*. *Molecular Plant-Microbe Interactions*, **20**, 1138–1148.

Barnett M, Swanson J, Long S (1998) Multiple genetic controls on Rhizobium meliloti syrA, a regulator of exopolysaccharide abundance. *Genetics*, **148**, 19–32.

Barnett M, Tolman C, Fisher R, Long S (2004) A dual-genome symbiosis chip for coordinate study of signal exchange and development in a prokaryote-host interaction. *Proceedings of the National Academy of Sciences of USA*, **101**, 16636–16641.

Becker A, Berges H, Krol E *et al*. (2004) Global changes in gene expression in *Sinorhizobium meliloti* 1021 under microoxic and symbiotic conditions. *Molecular Plant-microbe Interactions*, **17**, 292–303.

Becker A, Barnett MJ, Capela D *et al*. (2009) A portal for rhizobial genomes: RhizoGATE integrates a *Sinorhizobium meliloti* genome annotation update with postgenome data. *Journal of Biotechnology*, **140**, 45–50.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, **57**, 289–300.

Berrington de González A, Cox DR (2007) Interpretation of interaction: a review. *The Annals of Applied Statistics*, **1**, 371–385.

Boedigheimer MJ, Wolfinger RD, Bass MB *et al*. (2008) Sources of variation in baseline gene expression levels from toxicogenomics study control animals across multiple laboratories. *BMC Genomics*, **9**, 285.

Bolstad BM, Irizarry RA, Astrand M, Speed TP (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*, **19**, 185–193.

Bull J, Rice W (1991) Distinguishing mechanisms for the evolution of co-operation. *Journal of Theoretical Biology*, **149**, 63–74.

Chen EJ, Fisher RF, Perovich VM, Sabio EA, Long SR (2009) Identification of direct transcriptional target genes of ExoS/ChvI Two- component signaling in *Sinorhizobium meliloti*. *Journal of Bacteriology*, **191**, 6833–6842.

Cheng Y, Kato N, Wang W, Li J, Chen X (2003) Two RNA binding proteins, HEN4 and HUA1, act in the processing of AGAMOUS pre-mRNA in *Arabidopsis thaliana*. *Developmental Cell*, **4**, 53–66.

Chodosh LA, Olesen J, Hahn S *et al*. (1988) A yeast and a human CCAAT-binding protein have heterologous subunits that are functionally interchangeable. *Cell*, **53**, 25–35.

Combier J-P, Frugier F, de Billy F *et al*. (2006) MtHAP2-1 is a key transcriptional regulator of symbiotic nodule development regulated by microRNA169 in *Medicago truncatula*. *Genes & Development*, **20**, 3084–3088.

De Mita S, Ronfort J, McKhann HI *et al*. (2007) Investigation of the demographic and selective forces shaping the nucleotide diversity of genes involved in nod factor signaling in *Medicago truncatula*. *Genetics*, **177**, 2123–2133.

Des Marais DL, Juenger TE (2010) Pleiotropy, plasticity, and the evolution of plant abiotic stress tolerance. *Annals of the New York Academy of Sciences*, **1206**, 56–79.

Ding Y, Oldroyd GED (2009) Positioning the nodule, the hormone dictum. *Plant Signaling & Behavior*, **4**, 89–93.

Dominguez-Ferraras A, Munoz S, Olivares J, Soto MJ, Sanjuan J (2009) Role of potassium uptake systems in *Sinorhizobium meliloti* osmoadaptation and symbiotic performance. *Journal of Bacteriology*, **191**, 2133–2143.

Falconer DS, Mackay TFC (1996) *Introduction to Quantitative Genetics*. Longman Limited, Harlow.

Ferguson BJ, Indrasumunar A, Hayashi S *et al*. (2010) Molecular analysis of legume nodule development and autoregulation. *Journal of Integrative Plant Biology*, **52**, 61–76.

Garcia-Rodriguez FM, Toro N (2000) *Sinorhizobium meliloti nfe* (nodulation formation efficiency) genes exhibit temporal and spatial expression patterns similar to those of genes involved in symbiotic nitrogen fixation. *Molecular Plant-microbe Interactions*, **13**, 583–591.

Gibson G (2008) The environmental contribution to gene expression profiles. *Nature Reviews Genetics*, **9**, 575–581.

Gibson G, Wolfinger RD (2004) Gene expression profiling using mixed models. In: *Genetic Analysis of Complex Traits Using SAS* (ed. Saxton AM), pp. 251–278. SAS Publishing, Cary, North Carolina.

Gibson KE, Kobayashi H, Walker GC (2008) Molecular determinants of a symbiotic chronic infection. *Annual Review of Genetics*, **42**, 413–441.

Graham P, Vance C (2003) Legumes: importance and constraints to greater use. *Plant Physiology*, **131**, 872–877.

Griffitts JS, Carlyon RE, Erickson JH, Moulton JL, Barnett MJ *et al*. (2008) A *Sinorhizobium meliloti* osmosensory two-component

© 2012 Blackwell Publishing Ltd

system required for cyclic glucan export and symbiosis. *Molecular Microbiology*, **69**, 479–490.

He J, Benedito VA, Wang M *et al.* (2009) The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*, **10**, 441.

Heath KD (2010) Intergenomic epistasis and coevolutionary constraint in plants and rhizobia. *Evolution*, **64**, 1446–1458.

Heath KD, Lau JA (2011) Herbivores alter the fitness benefits of a plant-rhizobium mutualism. *Acta Oecologica*, **37**, 87–92.

Heath KD, Tiffin P (2007) Context dependence in the coevolution of plant and rhizobial mutualists. *Proceedings of the Royal Society of London. Series B*, **274**, 1905–1912.

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

Heath KD, Stock AJ, Stinchcombe JR (2010) Mutualism variation in the nodulation response to nitrate. *Journal of Evolutionary Biology*, **23**, 2494–2500.

Hodgins-Davis A, Townsend JP (2009) Evolving gene expression: from G to E to G x E. *Trends in Ecology & Evolution*, **24**, 649–658.

Hoeksema JD, Thompson JN (2007) Geographic structure in a widespread plant-mycorrhizal interaction: pines and false truffles. *Journal of Evolutionary Biology*, **20**, 1148–1163.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B *et al.* (2003) Summaries of affymetrix GeneChip probe level data. *Nucleic Acids Research*, **31**, e15.

Janga SC, Contreras-Moreira B (2010) Dissecting the expression patterns of transcription factors across conditions using an integrated network-based approach. *Nucleic Acids Research*, **38**, 6841–6856.

Janzen D (1980) When is it coevolution. *Evolution*, **34**, 611–612.

Kliebenstein DJ (2010) Systems biology uncovers the foundation of natural genetic diversity. *Plant Physiology*, **152**, 480–486.

Kliebenstein D, West M, van Leeuwen H *et al.* (2006) Genomic survey of gene expression diversity in *Arabidopsis thaliana*. *Genetics*, **172**, 1179–1189.

van Leeuwen H, Kliebenstein DJ, West MAL *et al.* (2007) Natural variation among *Arabidopsis thaliana* accessions for transcriptome response to exogenous salicylic acid. *Plant Cell*, **19**, 2099–2110.

Lin K-F, Liu Y-N, Hsu S-TD *et al.* (2005) Characterization and structural analyses of nonspecific lipid transfer protein 1 from mung bean. *Biochemistry*, **44**, 5703–5712.

Lodwig E, Poole P (2003) Metabolism of Rhizobium bacteroids. *Critical Reviews in Plant Sciences*, **22**, 37–78.

Lodwig E, Hosie A, Bourdes A *et al.* (2003) Amino-acid cycling drives nitrogen fixation in the legume-*Rhizobium* symbiosis. *Nature*, **422**, 722–726.

Lorkovic ZJ (2009) Role of plant RNA-binding proteins in development, stress response and genome organization. *Trends in Plant Science*, **14**, 229–236.

McIntosh M, Krol E, Becker A (2008) Competitive and cooperative effects in quorum-sensing-regulated galactoglucan biosynthesis in *Sinorhizobium meliloti*. *Journal of Bacteriology*, **190**, 5308–5317.

McMullen MD, Kresovich S, Villeda HS, Bradbury P, Li H (2009) Genetic properties of the maize nested association mapping population. *Science*, **325**, 737–740.

Mergaert P, Nikovics K, Kelemen Z *et al.* (2003) A novel family in *Medicago truncatula* consisting of more than 300 nodule-specific genes coding for small, secreted polypeptides with conserved cysteine motifs. *Plant Physiology*, **132**, 161–173.

Mitra R, Shaw S, Long S (2004) Six non-nodulating plant mutants defective for Nod factor-induced transcriptional changes associated with the legume-rhizobia symbiosis. *Proceedings of the National Academy of Sciences of USA*, **101**, 10217–10222.

Oono R, Anderson CG, Denison RF (2011) Failure to fix nitrogen by non-reproductive symbiotic rhizobia triggers host sanctions that reduce fitness of their reproductive clonemates. *Proceedings of the Royal Society of London. Series B*, **278**, 2698–2703.

Parker MA (1995) Plant fitness variation caused by different mutualist genotypes. *Ecology*, **76**, 1525–1535.

Pii Y, Crimi M, Cremonese G, Spena A, Pandolfini T (2007) Auxin and nitric oxide control indeterminate nodule formation. *BMC Plant Biology*, **7**, 21.

Povolo S, Casella S (2000) A critical role for *aniA* in energy-carbon flux and symbiotic nitrogen fixation in *Sinorhizobium meliloti*. *Archives of Microbiology*, **174**, 42–49.

Quackenbush J (2002) Microarray data normalization and transformation. *Nature Genetics*, **32**, 496–501.

Riely B, Ane J, Penmetsa R, Cook D (2004) Genetic and genomic analysis in model legumes bring Nod-factor signaling to center stage. *Current Opinion in Plant Biology*, **7**, 408–413.

Rinaude LV, Gonzalez JE (2009) The low-molecular-weight fraction of exopolysaccharide II from *Sinorhizobium meliloti* is a crucial determinant of biofilm formation. *Journal of Bacteriology*, **191**, 7216–7224.

Rodriguez MCS, Petersen M, Mundy J (2010) Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology*, **61**, 621–649.

Ronson CW, Nixon BT, Albright LM, Ausubel FM (1987) *Rhizobium meliloti ntrA* (*rpoN*) gene is required for diverse metabolic functions. *Journal of Bacteriology*, **169**, 2424–2431.

Sachs JL, Ehinger MO, Simms EL (2010) Origins of cheating and loss of symbiosis in wild Bradyrhizobium. *Journal of Evolutionary Biology*, **23**, 1075–1089.

Salazar E, Diaz-Mejia JJ, Moreno-Hagelsieb G, Martinez-Batallar G, Mora Y *et al.* (2010) Characterization of the NifA-RpoN regulon in *Rhizobium etli* in free life and in symbiosis with *Phaseolus vulgaris*. *Applied and Environment Microbiology*, **76**, 4510–4520.

Salvaudon L, Héraudet V, Shykoff JA (2005) Parasite-host fitness trade-offs change with parasite identity: genotype-specific interactions in a plant-pathogen system. *Evolution*, **59**, 2518–2524.

Scoville A, Lee YW, Willis JH, Kelly JK (2009) Contribution of chromosomal polymorphisms to the G-matrix of *Mimulus guttatus*. *New Phytologist*, **183**, 803–815.

Siefers N, Dang KK, Kumimoto RW *et al.* (2009) Tissue-specific expression patterns of Arabidopsis NF-Y transcription factors suggest potential for extensive combinatorial complexity. *Plant Physiology*, **149**, 625–641.

Simms E, Taylor D, Povich J *et al.* (2006) An empirical test of partner choice mechanisms in a wild legume-rhizobium interaction. *Proceedings of the Royal Society of London. Series B*, **273**, 77–81.

Somasegaran P, Hoben H (1994) *Handbook for Rhizobia: Methods in Legume-rhizobium Technology*. Springer-Verlag, New York.

© 2012 Blackwell Publishing Ltd

Thompson JN (2005) Coevolution: the geographic mosaic of coevolutionary arms races. *Current Biology*, **15**, R992–R994.

Trivers R (1971) The evolution of reciprocal altruism. *The Quarterly Review of Biology*, **46**, 35–57.

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

Vale PF, Little TJ (2009) Measuring parasite fitness under genetic and thermal variation. *Heredity*, **103**, 102–109.

Van de Velde W, Zehirov G, Szatmari A *et al*. (2010) Plant peptides govern terminal differentiation of bacteria in symbiosis. *Science*, **327**, 1122–1126.

Verhoeven K, Simonsen K, McIntyre L (2005) Implementing false discovery rate control: increasing your power. *Oikos*, **108**, 643–647.

Wade MJ (2007) The co-evolutionary genetics of ecological communities. *Nature Reviews Genetics*, **8**, 185–195.

Wightman F, Schneider EA, Thimann KV (1980) Hormonal factors controlling the initiation and development of lateral roots II. Effects of exogenous growth-factors on lateral root-formation in pea roots. *Physiologia Plantarum*, **49**, 304–314.

Wolf JB, Brodie III ED, Wade MJ (2004) The genotype-environment interaction and evolution when the environment contains genes. In: *Phenotypic Plasticity. Functional and Conceptual Approaches* (eds DeWitt TJ and Scheiner SM), pp. 173–190. Oxford University Press, Oxford, UK.

York GM, Walker GC (1997) The *Rhizobium meliloti exoK* gene and *prsD/prsE/exsH* genes are components of independent degradative pathways which contribute to production of low-molecular-weight succinoglycan. *Molecular Microbiology*, **25**, 117–134.

Zanetti ME, Blanco FA, Beker MP, Battaglia M, Aguilar OM (2010) A C subunit of the plant nuclear factor NF-Y required for rhizobial infection and nodule development affects partner selection in the common bean-Rhizobium etli symbiosis. *Plant Cell*, **22**, 4142–4157.

K.D.H. studies the evolutionary genetics of symbiosis with emphasis on mutualism coevolution. P.V.B. studies genetic regulation with a focus on model legume-rhizobium interactions. J.R.S. studies plant evolutionary genetics with a focus on ecologically important traits.

Data accessibility

Raw expression data are available at GEO (GSE29027).

Significant gene sets and phenotypic data uploaded as Supporting information.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 (a) Phenotypes and (b) Mixed model ANOVA.

Fig. S1 Distributions of ANOVA *P*-values for model effects (plant family, rhizobium strain, family × strain interaction, and block) for all (A) *Sinorhizobium meliloti*, or (B) *Medicago truncatula* probesets on the chip.

Appendix S1 List of all significantly differentially-expressed genes from *M. truncatula* and *S. meliloti* with annotations.

Appendix S2 Enrichment analyses using Fisher exact tests.

Appendix S3 Medicago-Sinorhizobium GeneChip procedures.

Appendix S4 Likelihood ratio tests of significance in PVCA analysis.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2012 Blackwell Publishing Ltd