# No evidence for adaptation to local rhizobial mutualists in the legume *Medicago lupulina*

Tia L. Harrison[1]*, Corlett W. Wood[1]*, Isabela L. Borges[1], and John R. Stinchcombe[1,2]

[1] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S3B2, Canada

[2] Centre for Genome Evolution and Function, University of Toronto, Toronto, Ontario, M5S3B2, Canada

* These authors contributed equally to this work.

*Running title*: Symbiont adaptation in *Medicago lupulina*

*Corresponding author*: Corlett W. Wood

*Address:*

Department of Ecology & Evolutionary Biology

University of Toronto

25 Willcocks Street, Room 3055

Toronto, ON

Canada M5S 3B2

*Fax*: 1 (416) 978-5878

*Phone:* 1 (647) 936-0565

*Email: corlett.wood@utoronto.ca*

**Keywords**

Coevolution, legume, rhizobia, reciprocal inoculation, genome scan, mutualism

**Abstract**

Local adaptation is a common but not ubiquitous feature of species interactions, and understanding the circumstances under which it evolves illuminates the factors that influence adaptive population divergence. Antagonistic species interactions dominate the local adaptation literature relative to mutualistic ones, preventing an overall assessment of adaptation within interspecific interactions. Here, we tested whether the legume *Medicago lupulina* is adapted to the locally abundant species of mutualistic nitrogen-fixing bacteria ("rhizobia"), which vary in frequency across its eastern North American range. We reciprocally inoculated northern and southern *M. lupulina* genotypes with the northern (*Ensifer medicae*) or southern bacterium (*E. meliloti*) in a greenhouse experiment. Neither northern nor southern plants produced more seeds, flowered earlier, or were more likely to flower when inoculated with their local rhizobium species, although plants produced more root nodules (the structures that house the bacteria) with their local rhizobia. We used a pre-existing dataset to perform a genome scan for loci that showed elevated differentiation between field-collected plants that hosted different bacteria. None of the loci we identified belonged to the well-characterized suite of legume-rhizobia symbiosis genes, suggesting that the rhizobia do not drive genetic divergence between *M. lupulina* populations. Our results demonstrate that symbiont local adaptation is weak in this mutualism despite large-scale geographic variation in the identity of the interacting species.

## Introduction

Characterizing the circumstances under which local adaptation evolves informs our understanding of the relative importance of gene flow and selection, and thereby the extent and limitations of adaptive evolution (Antonovics, 1976; Bridle & Vines, 2007; Hereford, 2009; Savolainen *et al.*, 2013; Whitlock, 2015). However, existing tests of local adaptation to the biotic environment focus disproportionately on antagonistic interactions (but see Anderson et al. 2004, Hoeksema and Thompson 2007, Barrett et al. 2012), limiting our understanding of adaptation within the broad suite of interspecific interactions that occur in nature. Here we performed a reciprocal inoculation experiment to test for local adaptation in a classic mutualism: the symbiosis between legumes and nitrogen-fixing bacteria.

Local adaptation—when native genotypes outperform foreign genotypes in their home environment (Hereford 2009)—is driven by differences in selection in alternative environments, and is reflected in divergent phenotypes and genotypes between populations. The literature on local adaptation between interacting species is dominated by antagonistic species interactions such as those between hosts and their parasites, pathogens, or prey (Brodie *et al.*, 2002; Kawecki & Ebert, 2004; Hoeksema & Forde, 2008; Koskella *et al.*, 2012). Direct tests of symbiont local adaptation in mutualisms are rare (Hoeksema & Forde, 2008; Brockhurst & Koskella, 2013). Nevertheless, several lines of evidence suggest that adaptation to the local mutualist is a common feature of positive species interactions. Phenotype matching between local plant and pollinator communities is pervasive (Anderson *et al.*, 2009; Gómez *et al.*, 2009; Koski & Ashman, 2015), and a recent reciprocal translocation experiment showed that a plant's reproductive success is highest in its local pollinator community (Newman *et al.*, 2015). In the classic mutualism

between leguminous plants and nitrogen-fixing bacteria, genotype-by-genotype interactions—when fitness depends jointly on the genotypes of both partners—account for a substantial proportion of genetic variation in fitness-related traits within plant populations (Heath, 2010; Heath *et al.*, 2012; Ehinger *et al.*, 2014). On a broad geographic scale, these interactions are predicted to manifest as symbiont local adaptation when coupled with population differences in symbiont genotype frequencies (Heath & Nuismer, 2014).

Ultimately, though, directly testing for symbiont local adaptation in mutualisms requires assaying the fitness consequences of sympatric and allopatric symbionts in a reciprocal inoculation experiment (Heath & Stinchcombe, 2014). The diagnostic signature of symbiont local adaptation in these experiments is a genotype-by-genotype interaction for fitness, indicating that the fitness of one partner depends on the identity of its symbiont (Clausen *et al.*, 1940; Clausen & Hiesey, 1958; Kawecki & Ebert, 2004). Although these experiments are frequently used to test for parasite local adaptation in antagonistic interactions (reviewed in Hoeksema and Forde 2008), they are less commonly used to test for mutualist local adaptation (but see (Hoeksema & Thompson, 2007; Johnson *et al.*, 2010; Barrett *et al.*, 2012; Newman *et al.*, 2015)).

The economically and ecologically important mutualism between legumes in the genus *Medicago* and nitrogen-fixing bacteria ("rhizobia") is well suited to testing for adaptation to the local mutualist (Cook *et al.*, 1997; Cook, 1999; Young *et al.*, 2011). In the facultative *Medicago*-rhizobia symbiosis, soil bacteria in the genus *Ensifer* (formerly *Sinorhizobium*) (Young, 2010) fix atmospheric nitrogen for their plant hosts in exchange for carbohydrates and housing in specialized root organs called nodules (Mylona *et al.*, 1995; van Rhijn & Vanderleyden, 1995).

In eastern North America the relative frequencies of two principal symbionts (*Ensifer medicae* and *E. meliloti*) (Béna *et al.*, 2005) vary along a latitudinal cline (Figure S1) (Harrison *et al.*, 2017), which may generate strong selection on *Medicago* populations to adapt to their local *Ensifer* species. The bacteria are essential for plant growth in nitrogen-poor edaphic environments (Simonsen & Stinchcombe, 2014a), and genes mediating the association experience strong selection in both *Medicago* and *Ensifer* (Bailly *et al.*, 2006; De Mita *et al.*, 2007; Epstein *et al.*, 2012; Bonhomme *et al.*, 2015). Finally, there is substantial evidence for genotype-by-genotype interactions for fitness traits between *Medicago truncatula* and its *Ensifer* symbionts (Heath, 2010; Gorton *et al.*, 2012; Heath *et al.*, 2012), and suggestive evidence for some degree of co-speciation in the two genera (Béna *et al.*, 2005).

In the present study, we performed a reciprocal inoculation experiment to test for adaptation to the local rhizobia species in the black medic (*Medicago lupulina*). We tested the effect of sympatric and allopatric rhizobia on plant fitness in a greenhouse experiment. We measured traits closely associated with plant fitness (seed production) and rhizobial fitness (number of nodules formed with the plant host) to examine plant adaptation to their local rhizobial species and rhizobial adaptation to their local plant populations, respectively. Second, we took advantage of an existing genomic data set and performed a genome scan for loci that exhibited elevated differentiation between field-collected plants associated with different bacterial species in natural populations. Genome scans identify loci that exhibit heightened differentiation between populations inhabiting alternative environments, which are presumed to constitute the genetic basis of local adaptation (Coop *et al.*, 2010; Günther & Coop, 2013; Savolainen *et al.*, 2013; Tiffin & Ross-Ibarra, 2014). Unlike reciprocal inoculation experiments,

these tests integrate across generations and ancillary environmental variation, capturing the cumulative effects of long-term selection in alternative environments (Tiffin & Ross-Ibarra, 2014; de Villemereuil *et al.*, 2015; Jensen *et al.*, 2016).

Neither the phenotypic nor genomic approaches revealed strong evidence of adaptation to the local rhizobia in *M. lupulina*. We found suggestive evidence that the rhizobia may be adapted to their local plant populations; plants formed more nodules with rhizobia from the same geographic region. However, overall our data suggest that symbiont local adaptation appears to be weak or absent in this mutualism's North American range.

## Materials and Methods

### *Study system*

*Medicago lupulina* is an annual, highly self-fertilizing legume native to Eurasia (Turkington & Cavers, 1979; Yan *et al.*, 2009). After its introduction to North America in the 1700s, *M. lupulina* expanded its range to occupy nitrogen-poor areas of the continent's temperate and subtropical regions (Turkington & Cavers, 1979). *Medicago* has a short generation time (Turkington & Cavers, 1979), its rhizobia are easily manipulated (Heath & Tiffin, 2007), an annotated genome is available in the genus (Young *et al.*, 2011), and the genes involved in the rhizobial mutualism are extensively characterized (Mylona *et al.*, 1995; Cook *et al.*, 1997; Young *et al.*, 2011).

The relative frequencies of *M. lupulina*'s two symbiotic rhizobia species (*Ensifer medicae* and *E. meliloti*) vary along a northwest-to-southeast cline in eastern North America (Figure S1) (Harrison *et al.*, 2017). Moreover, there is very little genetic diversity within each *Ensifer* species

at neutral sites and at known symbiosis genes (Harrison *et al.*, 2017). Whole-genome sequencing of *Ensifer* strains sampled from the same populations used in the present study found that neutral diversity extremely low in *E. medicae* ($\pi_{synonymous}$ = 0.0006 and in *E. meliloti* ($\pi_{synonymous}$ = 0.0001). Crucially, there was also very little diversity in genes known to be involved in the legume-rhizobia symbiosis. There were few or no SNPs in any nodulation genes (nodA, -B, or -C), nitrogen fixation genes (nifA, -B, -D, -E, -H, -N, or -X), or type III effector genes in either *E. medicae* ($\pi \leq 0.0004$ for these genes) or *E. meliloti* ($\pi \leq 0.0002$). Taken together, this evidence suggests that there is very little functional diversity within *Ensifer* species in North America, so we only tested a single strain per species in the reciprocal inoculation experiment described below.

*Reciprocal inoculation experiment*

To test for adaptation to the local rhizobia, we inoculated *M. lupulina* genotypes from the northern and southern portions of the plant's eastern North American range with either the locally abundant rhizobium species in the north (*E. medicae*) or in the south (*E. meliloti*). From a total of 39 *M. lupulina* populations sampled between Delaware and Ontario in September-October of 2013 (Harrison *et al.*, 2017), we selected 7 southern and 7 northern plant populations in which Harrison et al. (2017) detected only a single *Ensifer* species (Figure 2, Table S1; see Figure S1 for a complete map with all 39 sampled populations). Within each population, seeds and root nodules were collected from 2-10 randomly chosen *M. lupulina* individuals. All sampled plants were at least 0.5m apart. Nodules were stored at 4°C in plastic bags until they were processed. Field-collected seeds from these populations were grown in the greenhouse for

146 one generation to reduce maternal and environmental effects from the field, and we performed
147 our experiments using the progeny of these greenhouse-grown plants.

148 We planted $F_1$ greenhouse-derived seeds of 43 maternal families (27 from the north and
149 16 from the south) in a split-plot randomized complete block design in the greenhouse at the
150 University of Toronto. Seedlings were planted with sterile forceps into cone-tainers filled with
151 sand (autoclaved twice at 121°C). *Medicago lupulina* naturally colonizes sandy soils (Turkington
152 and Cavers 1979), and the plants rely heavily on nitrogen supplied by their rhizobial symbionts
153 under these conditions because sand is a nitrogen-poor substrate (Heath et al. 2010). Each block
154 was divided into two bacterial treatments, each containing 15 northern and 11 southern plants,
155 the locations of which were randomized within blocks. Populations were split across blocks. Due
156 to seed limitations, not all families were represented in every block, but within a block both
157 bacterial treatments comprised the same 26 families. We replicated this design across six blocks,
158 for a total of 312 plants (6-13 replicates per family for 37 families; 1-4 replicates per family for 6
159 families). An additional block containing 42 plants (33 from the north and 9 from the south)
160 served as an inoculation control, and a means for estimating plant performance and fitness in the
161 absence of either bacterial species. Prior to planting, seeds were scarified with a razor blade,
162 sterilized with ethanol and bleach, and stratified on 8% water agar plates at 4˚C for 7 days to
163 germinate. We misted seedlings with water daily and fertilized with 5mL of nitrogen-free
164 Fahraeus medium (noble.org/medicagohandbook) twice before inoculation with rhizobia.

165 The *Ensifer* strains used for inoculation were recovered from frozen samples collected by
166 Harrison et al. (2017) from two of the populations used in our experiment. The strains were
167 originally cultured from field-collected root nodules by sterilizing one nodule per plant in

ethanol and bleach, and crushing and plating it onto a 2% tryptone yeast (TY) agar plate. Strains were re-streaked onto TY agar four times to reduce contamination and grown at 30°C for 48 hours, after which they were transferred to liquid TY media and cultured for two days at 30°C. To identify each strain to species (*E. medicae* or *E. meliloti*), DNA was extracted from liquid cultures (cell density: 8 x $10^8$ cells/ml) using the MoBio UltraClean Microbial DNA Isolation Kit, whole-genome sequenced at SickKids Hospital (Toronto, Ontario), and genotyped using GATK (McKenna *et al.*, 2010). We used alignment scores and the *Ensifer* 16S locus (Rome *et al.*, 1997) to determine species identity of rhizobia strains associated with the sampled plants.

We selected one *E. medicae* strain from the northernmost population in Ontario and one *E. meliloti* strain from the southernmost population in Delaware for our experiment ("SEG" and "DE" in Figure 2). Genetic diversity is very low among strains within *Ensifer* species across North America (see "Study system" above) (Harrison *et al.*, 2017), so the specific strains used are not likely to influence our results. Prior to inoculation, these strains were cultured as described above from samples stored at -80°C. Liquid cultures were diluted with sterile TY media to an OD600 reading of 0.1 (a concentration of ~$10^6$ cells per mL) (Simonsen & Stinchcombe, 2014b). Each plant received 1 mL of inoculate 13 days after planting, and 1 mL again 10 days later. Controls were also inoculated twice with sterile TY media 10 days apart, and were used to assess rhizobia contamination across treatments. Throughout the remainder of the experiment, all plants were bottom-watered three times a week. We used two bottom-watering trays per block, such that all plants in a given bacterial treatment had the same tray, while those from the alternative bacterial treatment had a different tray.

We scored mortality weekly throughout the experiment, counted the number of leaves every 4 weeks, recorded the date of first flower, and collected seeds. After five months, which approximates the length of the April-October growing season in southern Ontario (Turkington & Cavers, 1979), we harvested all plants and collected any remaining unripe seeds. We dried and weighed aboveground tissue from each plant to the nearest 0.1 mg, and counted all seeds and root nodules (symbiotic organs housing the rhizobia).

We analyzed five traits to test whether northern and southern *M. lupulina* plants were adapted to their local rhizobium: number of seeds, aboveground biomass, flowering time (excluding plants that did not flower), probability of flowering, and number of nodules. All analyzes were performed in R v.3.2.4 with sum-to-zero contrasts ("contr.sum") (R Core Team, 2016), and we tested significance using type III sums of squares in the function Anova in the *car* package (Fox & Weisberg, 2011). Log-transformed aboveground biomass and flowering time were analyzed with general linear mixed models using the function lmer in the *lme4* package (Bates *et al.*, 2015). Probability of flowering and number of nodules were analyzed with generalized linear mixed models with binomial and Poisson error distributions, respectively, using the function glmer in the *lme4* package (Bates *et al.*, 2015). We verified that all dependent variables met the assumptions of linearity, normality, and homoscedasticity through visual inspection of quantile-quantile plots, plots of the residuals versus fitted values, and scale-location plots. Seed number was severely zero-inflated (42% of plants did not produce seeds), so we analyzed it using a mixture model (see below).

Each of the above models included rhizobia treatment (*E. medicae* or *E. meliloti*), region (north or south), and the rhizobia-by-region interaction as fixed effects. A significant rhizobia-

211 by-region interaction, in which northern plants have higher fitness when inoculated with *E.*
212 *medicae* and southern plants have higher fitness with *E. meliloti*, would be evidence for symbiont
213 local adaptation. We included a fixed effect of researcher in our analysis of nodule counts.
214 Block, population, and family nested within population were included as random effects. We
215 also included the block-by-treatment interaction as a random effect because the rhizobia
216 treatment was applied at the half-block rather than at the plant level (Altman & Krzywinski,
217 2015). While this design provides a weaker test of the rhizobia main effect, it is sensitive to the
218 detection of rhizobia-by-region interactions, the main goal of our experiment (Altman &
219 Krzywinski, 2015). We tested the significance of the population and family random effects in our
220 analyses of biomass, number of nodules, and flowering time using likelihood ratio tests in which
221 we compare the full model to reduced models without the random effects. Because these
222 likelihood ratio tests test a null hypothesis located at the boundary of parameter space (i.e., a
223 variance of zero), the p-values produced by these tests are twice what they should be, so we
224 divided these p-values in half (Bolker *et al.*, 2009).

225 We analyzed seed number with a zero-inflated Poisson model implemented with the
226 function MCMCglmm in the package *MCMCglmm* (Hadfield, 2010). Zero-inflated models are a
227 type of mixture model in which the zero class is modeled as the combined result of binomial and
228 count processes (Zuur *et al.*, 2009). In MCMCglmm, zero-inflated Poisson GLMMs are fit as
229 multi-response models with one latent variable for the binomial zero-generating process and one
230 for the Poisson count-generating process (Hadfield, 2015). We fit a model for seed number that
231 included fixed effects of rhizobia, region, the rhizobia-by-region interaction, and the reserved
232 MCMCglmm variable "trait" that indexes the binomial and Poisson latent variables. We omitted

the interaction between trait and other fixed effects in order to estimate a single effect of rhizobia, region, and the rhizobia-by-region interaction across both the binomial and Poisson processes. Block, population, family, and the block-by-treatment effect were included as random effects. Different random effect variances were fit to the binomial and Poisson processes using the "idh" variance structure in MCMCglmm (Hadfield, 2015). We fit a residual variance (R) structure using the argument rcov = ~ us(trait):units, which allows a unique residual for all predictors in the model, used the default priors for the fixed effects (mean = 0, variance = $10^{10}$) and specified parameter-expanded priors (alpha.mu = 0, alpha.v = 1000) for the random effects (Hadfield, 2010).

We ran the model for 1,300,000 iterations, discarded the first 300,000 iterations, and stored every 1,000$^{th}$ iterate. Model convergence was assessed with traceplots, running mean plots, and autocorrelation plots of the fixed and random effects using the *coda* (Plummer *et al.*, 2006) and *mcmcplots* (McKay Curtis, 2015) packages. Even though we used parameter-expanded priors on the random effects, the estimates of the population and block random effects remained close to zero, but omitting these terms from our model did not qualitatively change the results. We assessed significance of the population and family random effects in our analysis of seed number by comparing the Deviance Information Criteria (DIC; a Bayesian analog of AIC) of the full model to reduced models that omitted the random effects (Spiegelhalter *et al.*, 2002; Bolker *et al.*, 2009). We considered a random effects to be significant when the reduced model had $\Delta$DIC greater than two, indicating that it was a worse fit to the data than the full model (Spiegelhalter *et al.*, 2002).

Finally, we calculated pairwise correlations between all traits using Spearman's correlation on the family means for each trait. We obtained family means for biomass, flowering time, and number of nodules by extracting the conditional modes (also known as the best linear unbiased predictors, or BLUPs) for each level of the family random effect from the models described above. For number of seeds, we used the marginal posterior modes of the family random effect as our family mean estimates.

*Genomic data set*

A limitation of using reciprocal inoculation experiments to test for symbiont local adaptation is that the fitness benefit of a symbiosis often depends on the biotic and abiotic environmental conditions in which it is expressed (Heath & Tiffin, 2007; Heath *et al.*, 2010; Porter *et al.*, 2011; Barrett *et al.*, 2012; Simonsen & Stinchcombe, 2014a). To address this limitation, we took advantage of a pre-existing *M. lupulina* SNP dataset collected by Harrison et al. (2017) to perform genomic scans in *M. lupulina*. The goal of this analysis was to determine whether genes involved in the legume-rhizobia symbiosis are differentiated between plants associated with different *Ensifer* species in natural populations, a pattern that would be consistent with symbiont local adaptation.

Details on SNP discovery methods can be found in Supplemental Methods (Appendix 1), Harrison et al. (2017). In brief, field-collected seeds from 73 *M. lupulina* individuals were grown in the greenhouse as described in the "Reciprocal inoculation experiment" section above. We extracted DNA from leaf tissue collected from one individual per maternal line and samples were sequenced at Cornell University using genotyping-by-sequencing (GBS) in two Illumina flow

276 cell lanes (Elshire *et al.*, 2011). Genomic libraries were prepared with the restriction enzyme
277 EcoT22I, and SNPs were called using the program Stacks (Catchen *et al.*, 2011, 2013). We
278 extracted and sequenced rhizobia DNA from one nodule from each field-sampled plant, and
279 determined the species identity of each strain as described in the "Reciprocal inoculation
280 experiment" section above.

281 We searched for outlier loci between *M. lupulina* plants hosting *E. medicae* and *E.*
282 *meliloti* to assess whether there is evidence for genetic divergence between plants associated
283 with different *Ensifer* species. We used the program Bayenv2 to calculate $X^TX$ statistics for each
284 SNP in the *M. lupulina* sample (Coop *et al.*, 2010): $X^TX$ is an $F_{ST}$-like statistic that controls for
285 population variation and covariation in allele frequencies (i.e., population structure). We
286 estimated the covariance matrix using 100,000 iterations. Because we only wanted to calculate
287 $X^TX$ statistics and did not wish to calculate environmental correlations, we included an
288 environmental file of dummy values to run Bayenv2 but avoid environmental analysis. We
289 ranked SNPs from highest to lowest $X^TX$ values and identified the top 1% of SNPs to BLAST
290 against the reference genome of *M. truncatula* to identify the outlier loci involved in rhizobia
291 association in *M. truncatula* (taxonomy ID 3880) (Tang *et al.*, 2014). We aligned to the *M.*
292 *truncatula* genome to leverage its extensive genomic resources; *M. truncatula* is the closest
293 relative of *M. lupulina* with a thoroughly mapped and annotated genome. We used nucleotide
294 BLAST (blastn) to search somewhat similar sequences in the unannotated *M. truncatula* genome
295 in order to retrieve chromosome positions for our outlier loci. To identify the orthologous gene
296 associated with each outlier locus, we then looked up the chromosome position of each outlier in
297 the annotated *Medicago truncatula* genome (Mt. 4.0 http://jcvi.org/medicago/). We performed

298 the outlier loci test in three ways. First, we characterized outlier loci using the range-wide sample
299 of plants (73 plant individuals) and compared the results to outlier loci found using the southern
300 Ontario samples (49 plant individuals) to account for possible covariance between environmental
301 gradients and bacterial species composition. Second, because outlier loci are usually in linkage
302 disequilibrium (LD) with the causal genes responsible for adaptation, we searched for legume-
303 rhizobia symbiosis genes within 5 kb and 10 kb of our detected outlier loci (Branca *et al.* 2011).
304 Third, we measured the distance in base pairs between the *M. truncatula* orthologs of detected
305 outlier loci and key *M. truncatula* genes involved in the rhizobia symbiosis, assuming synteny
306 between *M. lupulina* and *M. truncatula*. Details of these analyses can be found in the
307 Supplemental Methods (Appendix 2).
308
309 **Results**
310 *Reciprocal inoculation experiment*
311 Uninoculated *Medicago lupulina* plants performed extremely poorly without rhizobia.
312 None of our uninoculated control plants flowered or set seed, and the biomass of control plants
313 was approximately 20-fold smaller than inoculated plants (least squares mean ± SE (mg):
314 controls: 21.01 ± 0.05; inoculated plants from both rhizobia treatments: 476.01 ± 0.03; $F_{1,14.808}$ =
315 610.7, $P < 0.001$). The performance of the control plants also demonstrates that cross-
316 contamination between the two rhizobia treatments was likely minimal in our experiment. Only 1
317 of 42 uninoculated control plants produced nodules, and this anomalous individual was similar in
318 size to the rest of the controls for the first several months, indicating that it probably did not
319 nodulate until late in the experiment.

In plants inoculated with *E. medicae* or *E. meliloti*, pairwise family mean correlations between all measured traits were generally low, indicating that the traits that we measured were largely independent of one another ($r \leq |0.10|$, $P \geq 0.54$). Only flowering time and aboveground biomass were significantly correlated ($r = 0.49$, $P = 0.002$); later-flowering plants had greater aboveground biomass.

Our analysis of seed number, probability of flowering, and flowering time revealed no evidence of adaptation to the local rhizobia. There was no significant rhizobia-by-region interaction for any of these reproductive traits (Figure 3, Table 1). There was a marginally significant effect of region on seed number; southern plants produced more seeds than northern plants in both rhizobia treatments (Figure 3A, Table 1). There was no significant effect of rhizobia treatment or region on either flowering trait (Figure 3C, Table 1).

The rhizobia-by-region interaction for aboveground biomass was marginally significant ($P_{\text{rhizobia-by-region interaction}} = 0.054$, Table 1). While the biomass of northern plants was unaffected by rhizobia treatment, southern plants produced more aboveground biomass when inoculated with *E. meliloti* (Figure 3B), the locally abundant rhizobia in south.

We found a highly significant rhizobia-by-region interaction for nodule number (Table 1). Northern plants produced more nodules that southern plants when inoculated with *E. medicae*, the locally abundant rhizobia in the north. The difference between northern and southern plants decreased when inoculated with *E. meliloti*, an effect that was driven by both an increase in nodulation in southern plants and a decrease in nodulation in northern plants (Figure 3D). There was also a significant effect of region, indicating that northern plants produced more nodules across both rhizobia treatments, and a significant effect of researcher (Table 1).

Aboveground biomass varied significantly among populations ($\chi^2$ = 3.405, P = 0.033) and families ($\chi^2$ = 40.490, P < 0.001). Number of nodules did not vary among populations ($\chi^2$ = 0, P = 0.500), but did vary among families ($\chi^2$ = 1305.200, P <0.001). Flowering time varied significantly among populations ($\chi^2$ = 12.058, P <0.001) and families ($\chi^2$ = 28.511, P <0.001). Number of seeds did not vary significantly among populations ($\Delta DIC_{reduced}$ = 1.54), but did vary among families ($\Delta DIC_{reduced}$ = 133.7).

*Genomic outlier analysis*

We identified three outlier loci that appeared in the top 1% of SNPs in both the range-wide *M. lupulina* sample and southern *M. lupulina* Ontario sample in our Bayenv2 analysis (Supplemental Table S2 and S3). None of these three loci mapped to a specific gene in the *M. truncatula* reference genome. Furthermore, we did not find any genes involved in the legume-rhizobia interaction within 5 kb or 10 kb of our three outlier loci (Supplemental Table S4). Finally, the base pair distances between our outlier loci and known genes involved in the *Medicago-Ensifer* mutualism were very large (minimum: 18 kb) (Supplemental Table S4). Details on summary $X^TX$ statistics, BLAST alignment scores, and gene functions are presented in the Supplemental Methods (Appendix 3) and Supplemental Table S2.

## Discussion

We performed a reciprocal inoculation experiment and genome scan to test for symbiont local adaptation in the mutualism between *M. lupulina* and nitrogen-fixing *Ensifer* bacteria across its eastern North American range. We found no evidence that plants are adapted to the

364 locally abundant rhizobia species in either analysis. Local rhizobia did not have differential
365 fitness consequences for their host plants, nor were any of the well-characterized symbiosis
366 genes differentiated between field-collected plants associated with different rhizobia. However,
367 we did find suggestive evidence for rhizobium local adaptation to the plant host; plants produced
368 more nodules with rhizobia from the same geographic region. Overall, symbiont local adaptation
369 appears to be absent or weak in this mutualism's eastern North American range despite the strong
370 cline in the relative abundances of the two rhizobia species.
371
372 *Reciprocal inoculation experiment & genomic outlier analysis*
373 Uninoculated plants performed extremely poorly without either *Ensifer* species,
374 demonstrating that *M. lupulina* is adapted to symbiosis with rhizobia. Despite differential
375 nodulation with local and foreign rhizobia ($P_{rhizobia\text{-}by\text{-}region} < 0.001$, Table 1), however, there was
376 no strong evidence for adaptation to the local rhizobia in other plant traits. One explanation for
377 this pattern is that plants modify their nodulation strategy to compensate for differences in
378 symbiotic efficiency with local and foreign rhizobia. The congeneric species *M. truncatula*
379 adjusts its nodulation strategy in response to the rhizobia nitrogen fixation efficiency (Heath &
380 Tiffin 2009), which jointly depends on plant and rhizobia genotype (Mhadhbi *et al.*, 2005). If
381 plants produce more nodules with less efficient symbionts, increased nodulation may not
382 translate to greater nitrogen uptake, masking any effects of differential nodulation on biomass
383 and seed production. The fact that seed number, a reasonable proxy for total fitness in a selfing
384 annual or short-lived perennial like *M. lupulina* (Turkington & Cavers, 1979), was unaffected by

385 the local rhizobia strongly suggests that adaptation to the local rhizobia was absent in our
386 experiment at the whole-plant level.

387 Our data suggest that symbiont local adaptation may be restricted to the rhizobia in this
388 mutualism; the rhizobia may be adapted to their local *M. lupulina* genotype even though the
389 plant does not appear to be adapted to its local rhizobium. The strongest signature of local
390 adaptation in our reciprocal inoculation experiment occurred in nodule traits, a pattern that has
391 also been documented in congeneric *Medicago* species (Porter *et al.*, 2011). Differential
392 nodulation may impact the rhizobia more than the plant, given that nodule number is correlated
393 with rhizobia fitness in *Medicago* (Heath, 2010).

394 However, even in the nodulation trait that exhibited a strong rhizobia-by-region
395 interaction—the statistical signature of local adaptation—the data are only weakly consistent
396 with the canonical pattern of local adaptation. The strongest test of local adaptation is whether
397 local genotypes outperform foreign genotypes in all environments (the "local-versus-foreign"
398 criterion) (Kawecki & Ebert, 2004). Neither trait that exhibited any rhizobia-by-region
399 interaction (number of nodules and aboveground biomass) satisfied this criterion. Instead, our
400 results were more closely aligned with a weaker test of local adaptation, which diagnoses local
401 adaptation when each genotype's fitness is greater in its native environment than in alternative
402 environments (the "home-versus-away" criterion) (Kawecki & Ebert, 2004).

403 Although reciprocal inoculation experiments are powerful because they reflect whole-
404 organism performance in native and foreign environments, genotype-by-environment
405 interactions are sensitive to experimental conditions (Kawecki & Ebert, 2004). We designed our
406 experiment to minimize this concern by mimicking key abiotic conditions relevant to the

407 *Medicago*-rhizobia mutualism. For example, we planted in sand, which reflects the nitrogen-poor
408 stony roadside soils that *M. lupulina* naturally colonizes (Turkington and Cavers 1979). Under
409 such nitrogen-limited conditions, the mutualism between legumes and nitrogen-fixing bacteria is
410 especially crucial for the plant (Heath et al. 2010). Nevertheless, the absence of local adaptation
411 in our experiment could be due to experimental conditions not adequately reflecting the typical
412 natural environment (in our case, cone-tainers, sterilized greenhouse soil, artificial day length
413 control, absence of other biotic interactors, etc.). A second caveat of our reciprocal inoculation
414 experiment is that we used only one strain for each rhizobium species. Although previous work
415 suggests that functional diversity within *E. medicae* and *E. meliloti* is extremely limited in
416 eastern North America, it is possible that we would have detected a signal of local adaptation had
417 we used different or larger sample of rhizobia strains.

418 To address these caveats, we performed a genomic outlier analysis to search for allele
419 frequency differences between plants associated with different rhizobia species. Genome scans
420 circumvent the caveats described above because they integrate across many generations of
421 selection and ancillary environmental variation. However, we also found very weak evidence of
422 symbiont local adaptation in this analysis. The loci that were highly differentiated between plants
423 hosting different *Ensifer* species (the top 1% of loci in the $X^TX$ outlier analysis) were not
424 associated with any genes involved in the legume-rhizobia symbiosis in either the range-wide or
425 Ontario samples. Moreover, none of the *M. truncatula* orthologs of our outlier loci were located
426 within the scale of linkage disequilibrium (5-10 kb in *M. truncatula*) (Branca *et al.*, 2011) from
427 known symbiosis genes. It is unlikely that the loci identified in our genome scan are novel *M.*
428 *lupulina*-specific symbiosis genes underlying adaptation to the local bacteria. The *Medicago*

genes involved in symbiotic interactions with rhizobia are well characterized and highly conserved in legumes (Rostas *et al.*, 1986; van Rhijn & Vanderleyden, 1995; De Mita *et al.*, 2006; Branca *et al.*, 2011; Gorton *et al.*, 2012; Stanton-Geddes *et al.*, 2013). *Medicago lupulina* is a close relative of *M. truncatula* (Bena, 2001; Yoder *et al.*, 2013), and both plants fix nitrogen with both *Ensifer* species tested in our experiment (Béna *et al.*, 2005). However, our results are subject to caveats common to genome scans for selection (Pavlidis *et al.*, 2012). In particular, our sample size in terms of individuals, and the number of SNPs, was low, reducing our power. Therefore, our genome scan might not have been able to detect highly differentiated loci important for symbiont local adaptation in *M. lupulina*.

*Local adaptation in the legume-rhizobia symbiosis*

Our data suggest that symbiont local adaptation may be asymmetric in this mutualism. Although plant fitness was not affected by the identity of its rhizobial symbiont, plants tended to produce more nodules with rhizobia from the same geographic region. This pattern was especially pronounced for the northern rhizobium, *E. medicae*. Stronger symbiont local adaptation in one partner commonly occurs in host-parasite systems (Hoeksema & Forde, 2008), but the phenomenon has not been systematically explored in the context of mutualism even though asymmetrical evolutionary rates in coevolving species pairs are expected in both mutualisms and antagonisms (Bergstrom & Lachmann, 2003).

Nevertheless, our phenotypic and genomic data indicate that *M. lupulina* is not adapted to the local rhizobia across its eastern North American range. The absence of symbiont local adaptation in this mutualism is surprising given that the system is characterized by several

451 features that ordinarily strongly favor its evolution. Genotype-by-genotype interactions
452 commonly occur between a congener, *M. truncatula*, and different strains of the same *Ensifer*
453 species (Heath & Tiffin, 2007; Heath, 2010; Heath *et al.*, 2012), suggesting that the genetically
454 divergent rhizobia *species* (Bailly *et al.*, 2006) we assayed would have even greater effects on
455 their plant host. Furthermore, there is a cline in the frequencies of the two rhizobia across a large
456 geographic scale that coincides with plant population genetic structure (Harrison *et al.*, 2017).
457 What might account for the lack of symbiont local adaptation in this mutualism?
458 Gene flow may overwhelm the effects of local selection, leading to a low equilibrium
459 level of genetic differentiation between plants associated with different rhizobia (McKay &
460 Latta, 2002). Although there is a strong geographic cline in the frequencies of the two *Ensifer*
461 species, Harrison et al. (2017) did detect *E. meliloti* in some northern populations and *E. medicae*
462 in some southern populations. Symbiont local adaptation within *M. lupulina* populations could
463 be swamped by gene flow from neighboring populations that encounter the alternative mutualist,
464 or by the invasion of the alternative mutualist itself. Frequent gene flow across large geographic
465 distances is consistent with the ecology of *M. lupulina* in North America. It is dispersed over
466 long distances by birds, grazing animals, and water, and is likely distributed as a component in
467 seed mixes for forage crops as well (Turkington and Cavers 1979).
468 Gene flow in the form of horizontal gene transfer between the two rhizobia species could
469 similarly homogenize any signature of local selection (Lenormand, 2002; Bailly *et al.*, 2007).
470 Bacteria that form nitrogen-fixing symbioses with legumes have been shown to horizontally
471 transfer genes involved in forming and maintaining the mutualism (Suominen *et al.*, 2001; Aoki
472 *et al.*, 2013; Lemaire *et al.*, 2015). Swapping cassettes of symbiosis genes would enable

473 genetically divergent rhizobial species to associate with the same hosts, largely eliminating
474 among-symbiont differences from the perspective of the plant. Finally, temporal variation in the
475 biotic and abiotic environment may modify the costs and benefits of the mutualism (Heath *et al.*,
476 2010; Heath & McGhee, 2012; Simonsen & Stinchcombe, 2014a), weakening selection favoring
477 local rhizobia.

478 Alternatively, symbiont local adaptation may generate relatively weak fitness tradeoffs in
479 mutualisms. The fitness tradeoffs that are the hallmark of local adaptation evolve whenever
480 adaptation to one environment results in maladaptation to another (Kawecki & Ebert, 2004). It
481 has been hypothesized that selection in coevolving mutualisms strongly favors general
482 compatibility and the reduction of fitness tradeoffs (Law & Koptur, 1986; Parker, 1999; Barrett
483 *et al.*, 2012). Selection to minimize fitness tradeoffs may be especially strong in the legume-
484 rhizobia mutualism, which is crucial for plants growing in nitrogen-poor soils (Heath *et al.*,
485 2010). Under nitrogen-limited conditions, the cost of maladaptation to a locally rare rhizobium
486 may be severe enough to outweigh the selective advantage of a marginal increase in the benefits
487 obtained from the locally abundant rhizobium (Barrett *et al.*, 2012). However, this process
488 should minimize plant-rhizobia interactions for fitness *within* rhizobia species as well,
489 inconsistent with the pervasive genotype-by-genotype interactions documented between *M.*
490 *truncatula* and *E. meliloti* (Heath *et al.*, 2012).
491
492 *Complementarity of phenotypic and genotypic approaches*
493 In the present study, we took advantage of a pre-existing genomic dataset to complement
494 and extend our test for symbiont local adaptation using a classic reciprocal inoculation

495 experiment. Our genomic outlier analysis also did not produce evidence of symbiont local
496 adaptation, possibly because of our low sample size and low SNP coverage in our data set.
497 However, we believe that combining an experimental approach and genomics is an innovative
498 and powerful way to test for local adaptation that should be applied more broadly. Although
499 genome scans and reciprocal inoculation experiments are typically treated as alternatives because
500 they draw on fundamentally different data, together the two approaches constitute a rigorous test
501 for local adaptation in environmentally sensitive symbioses such as the legume-rhizobia
502 mutualism. Combined, the two approaches integrate over the effects of all loci in the genome
503 (reciprocal inoculation experiments) and across ancillary environmental variation (genome
504 scans), producing inferences that are less vulnerable to the weaknesses of either method (Buehler
505 *et al.*, 2014; de Villemereuil *et al.*, 2015; Jensen *et al.*, 2016). Studies of (symbiont) local
506 adaptation should consider pairing phenotypic and genomic approaches to validate their results
507 with independent lines of evidence and exclude alternative interpretations of the data (de
508 Villemereuil *et al.*, 2015; Jensen *et al.*, 2016). Future directions for our research could include
509 repeating the genomic outlier test with a higher quality genomic data set to determine whether
510 the reciprocal inoculation and genome scan produce concordant results on symbiont local
511 adaptation.
512
513 **Acknowledgements**
514 Our work is supported by Discovery Grants and graduate fellowships from NSERC Canada, and
515 the EEB Postdoctoral Fellowship at the University of Toronto. We thank Bruce Hall and Andrew
516 Petrie for greenhouse assistance, and Adriana Salcedo, Michelle Afkhami, and Rebecca Batstone

517 for advice on experimental design and analysis. We are grateful for feedback from an

518 anonymous Associate Editor and two anonymous reviewers, which strengthened the manuscript.

519

520 **References**

521 Altman, N. & Krzywinski, M. 2015. Points of Significance: Split plot design. *Nat. Methods* **12**:
522 165–166.
523 Anderson, B., Johnson, S.D. & Anderson, B. 2009. Geographical covariation and local
524 convergence of flower depth in a guild of fly-pollinated plants. *New Phytol.* **182**: 533–540.
525 Ané, J.-M., Lévy, J., Thoquet, P., Kulikova, O., de Billy, F., Penmetsa, V., *et al.* 2002. Genetic
526 and cytogenetic mapping of DMI1, DMI2, and DMI3 genes of Medicago truncatula
527 involved in Nod factor transduction, nodulation, and mycorrhization. *Mol. Plant. Microbe.*
528 *Interact.* **15**: 1108–1118.
529 Antonovics, J. 1976. The Nature of Limits to Natural Selection. *Ann. Missouri Bot. Gard.* **63**:
530 224–247.
531 Aoki, S., Ito, M. & Iwasaki, W. 2013. From alpha- to beta-proteobacteria: The origin and
532 evolution of rhizobial nodulation genes nodIJ. *Mol. Biol. Evol.* **30**: 2494–2508.
533 Bailly, X., Olivieri, I., Brunel, B., Cleyet-Marel, J.C. & Béna, G. 2007. Horizontal gene transfer
534 and homologous recombination drive the evolution of the nitrogen-fixing symbionts of
535 Medicago species. *J. Bacteriol.* **189**: 5223–5236.
536 Bailly, X., Olivieri, I., De Mita, S., Cleyet-Marel, J.C. & Béna, G. 2006. Recombination and
537 selection shape the molecular diversity pattern of nitrogen-fixing Sinorhizobium sp.
538 associated to Medicago. *Mol. Ecol.* **15**: 2719–2734.
539 Barrett, L.G., Broadhurst, L.M. & Thrall, P.H. 2012. Geographic adaptation in plant-soil
540 mutualisms: Tests using Acacia spp. and rhizobial bacteria. *Funct. Ecol.* **26**: 457–468.
541 Bates, D., Mächler, M., Bolker, B. & Walker, S. 2015. Fitting Linear Mixed-Effects Models

using lme4. *J. Stat. Softw.* **67**.

Bena, G. 2001. Molecular phylogeny supports the morphologically based taxonomic transfer of the “medicagoid” Trigonella species to the genus Medicago L. *Plant Syst. Evol.* **229**: 217–236.

Béna, G., Lyet, A., Huguet, T. & Olivieri, I. 2005. Medicago-Sinorhizobium symbiotic specificity evolution and the geographic expansion of Medicago. *J. Evol. Biol.* **18**: 1547–1558.

Bergstrom, C.T. & Lachmann, M. 2003. The Red King effect: when the slowest runner wins the coevolutionary race. *Proc. Natl. Acad. Sci. U. S. A.* **100**: 593–598.

Bolker, B.M., Brooks, M.E., Clark, C.J., Geange, S.W., Poulsen, J.R., Stevens, M.H.H., *et al.* 2009. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends Ecol. Evol.* **24**: 127–35.

Bonhomme, M., Boitard, S., San Clemente, H., Dumas, B., Young, N. & Jacquet, C. 2015. Genomic Signature of Selective Sweeps Illuminates Adaptation of Medicago truncatula to Root-Associated Microorganisms. *Mol. Biol. Evol.* **32**: 2097–2110.

Branca, A., Paape, T.D., Zhou, P., Briskine, R., Farmer, A.D., Mudge, J., *et al.* 2011. Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula. *Proc. Natl. Acad. Sci. U. S. A.* **108**: E864-70.

Bridle, J.R. & Vines, T.H. 2007. Limits to evolution at range margins: when and why does adaptation fail? *Trends Ecol. Evol.* **22**: 140–147.

Brockhurst, M.A. & Koskella, B. 2013. Experimental coevolution of species interactions. *Trends Ecol. Evol.* **28**: 367–375. Elsevier Ltd.

564 Brodie, E.D.J., Ridenhour, B.J. & Brodie, E.D.I. 2002. The evolutionary response of predators to
565 dangerous prey: hotspots and coldspots in the geographic mosaic of coevolution between
566 garter snakes and newts. *Evolution (N. Y).* **56**: 2067–82.
567 Buehler, D., Holderegger, R., Brodbeck, S., Schnyder, E. & Gugerli, F. 2014. Validation of
568 outlier loci through replication in independent data sets: A test on Arabis alpina. *Ecol. Evol.*
569 **4**: 4296–4306.
570 Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. & Cresko, W.A. 2013. Stacks: An
571 analysis tool set for population genomics. *Mol. Ecol.* **22**: 3124–3140.
572 Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H. & De Koning, D.-J.
573 2011. Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3* **1**:
574 171–182.
575 Clausen, J. & Hiesey, W.M. 1958. Experimental studies on the nature of species. IV. Genetic
576 structure of ecological races. *Carnegie Inst. Washingt. 615. Washington, DC.*
577 Clausen, J., Keck, D.D. & Hiesey, W.M. 1940. Experimental studies on the nature of species. I.
578 Effect of varied environments on North American plants. *Carnegie Inst. Washingt. 520.*
579 *Washington, DC.*
580 Cook, D.R. 1999. Medicago truncatula - A model in the making! *Curr. Opin. Plant Biol.* **2**: 301–
581 304.
582 Cook, D.R., VandenBosch, K., de Bruijn, F.J. & Huguet, T. 1997. Model legumes get the nod.
583 *Plant Cell* **9**: 275–280.
584 Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. 2010. Using environmental correlations
585 to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423.

586 De Mita, S., Santoni, S., Hochu, I., Ronfort, J. & Bataillon, T. 2006. Molecular evolution and
587 positive selection of the symbiotic gene NORK in Medicago truncatula. *J. Mol. Evol.* **62**:
588 234–244.
589 De Mita, S., Santoni, S., Ronfort, J. & Bataillon, T. 2007. Adaptive evolution of the symbiotic
590 gene NORK is not correlated with shifts of rhizobial specificity in the genus Medicago.
591 *BMC Evol. Biol.* **7**: 210.
592 de Villemereuil, P., Gaggiotti, O.E., Mouterde, M. & Till-Bottraud, I. 2015. Common garden
593 experiments in the genomic era: new perspectives and opportunities. *Heredity (Edinb).* **116**:
594 249–254.
595 Ehinger, M., Mohr, T.J., Starcevich, J.B., Sachs, J.L., Porter, S.S. & Simms, E.L. 2014.
596 Specialization-generalization trade-off in a Bradyrhizobium symbiosis with wild legume
597 hosts. *BMC Ecol.* **14**: 8. BMC Ecology.
598 Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., *et al.* 2011. A
599 robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS*
600 *One* **6**: e19379.
601 Epstein, B., Branca, A., Mudge, J., Bharti, A.K., Briskine, R., Farmer, A.D., *et al.* 2012.
602 Population Genomics of the Facultatively Mutualistic Bacteria Sinorhizobium meliloti and
603 S. medicae. *PLoS Genet.* **8**: 1–10.
604 Fox, J. & Weisberg, S. 2011. *An R Companion to Applied Regression*, Second ed. Sage,
605 Thousand Oaks, CA.
606 Gómez, J.M., Abdelaziz, M., Camacho, J.P.M., Muñoz-Pajares, A.J. & Perfectti, F. 2009. Local
607 adaptation and maladaptation to pollinators in a generalist geographic mosaic. *Ecol. Lett.*

608 **12**: 672–682.

609 Gorton, A.J., Heath, K.D., Pilet-Nayel, M.-L., Baranger, A. & Stinchcombe, J.R. 2012. Mapping
610 the genetic basis of symbiotic variation in legume-rhizobium interactions in Medicago
611 truncatula. *G3* **2**: 1291–303.

612 Günther, T. & Coop, G. 2013. Robust identification of local adaptation from allele frequencies.
613 *Genetics* **195**: 205–220.

614 Hadfield, J. 2015. MCMCglmm Course Notes.

615 Hadfield, J.D. 2010. MCMC methods for multi-response generalized linear mixed models: the
616 MCMCglmm R package. *J. Stat. Softw.* **33**: 1–22.

617 Harrison, T.L., Wood, C.W., Heath, K.D. & Stinchcombe, J.R. 2017. Geographically structured
618 genetic variation in the Medicago lupulina-Ensifer mutualism. *bioRxiv*, doi:
619 http://dx.doi.org/10.1101/117192.

620 Heath, K.D. 2010. Intergenomic epistasis and coevolutionary constraint in plants and rhizobia.
621 *Evolution (N. Y).* **64**: 1446–1458.

622 Heath, K.D., Burke, P. V & Stinchcombe, J.R. 2012. Coevolutionary genetic variation in the
623 legume-rhizobium transcriptome. *Mol. Ecol.* **21**: 4735–47.

624 Heath, K.D. & McGhee, K.E. 2012. Coevolutionary Constraints? The environment alters
625 tripartite interaction traits in a legume. *PLoS One* **7**: e41567.

626 Heath, K.D. & Nuismer, S.L. 2014. Connecting functional and statistical definitions of genotype
627 by genotype interactions in coevolutionary studies. *Front. Genet.* **5**: 1–7.

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

630 Heath, K.D., Stock, A.J. & Stinchcombe, J.R. 2010. Mutualism variation in the nodulation
631 response to nitrate. *J. Evol. Biol.* **23**: 2494–2500.
632 Heath, K.D. & Tiffin, P. 2007. Context dependence in the coevolution of plant and rhizobial
633 mutualists. *Proc. R. Soc. B Biol. Sci.* **274**: 1905–1912.
634 Hereford, J. 2009. A quantitative survey of local adaptation and fitness trade-offs. *Am. Nat.* **173**:
635 579–588.
636 Hoeksema, J.D. & Forde, S.E. 2008. A Meta-Analysis of Factors Affecting Local Adaptation.
637 *Am. Nat.* **171**: 275–290.
638 Hoeksema, J.D. & Thompson, J.N. 2007. Geographic structure in a widespread plant-
639 mycorrhizal interaction: Pines and false truffles. *J. Evol. Biol.* **20**: 1148–1163.
640 Jensen, J.D., Foll, M. & Bernatchez, L. 2016. The past , present and future of genomic scans for
641 selection. *Mol. Ecol.* **25**: 1–4.
642 Johnson, N.C., Wilson, G.W.T., Bowker, M.A., Wilson, J.A. & Miller, R.M. 2010. Resource
643 limitation is a driver of local adaptation in mycorrhizal symbioses. *Proc. Natl. Acad. Sci. U.*
644 *S. A.* **107**: 2093–8.
645 Jones, K.M., Kobayashi, H., Davies, B.W., Taga, M.E. & Walker, G.C. 2007. How rhizobial
646 symbionts invade plants: the Sinorhizobium-Medicago model. *Nat. Rev. Microbiol.* **5**: 619–
647 33.
648 Kawecki, T.J. & Ebert, D. 2004. Conceptual issues in local adaptation. *Ecol. Lett.* **7**: 1225–1241.
649 Koskella, B., Lin, D.M., Buckling, A. & Thompson, J.N. 2012. The costs of evolving resistance
650 in heterogeneous parasite environments. *Proc. R. Soc. B* **279**: 1896–903.
651 Koski, M. & Ashman, T.-L. 2015. An altitudinal cline in UV floral pattern corresponds with a

behavioral change of a generalist pollinator assemblage. *Ecology* **96**: 3343–3353.

Law, R. & Koptur, S. 1986. On the evolution of non-specific mutualism. *Biol. J. Linn. Soc.* **27**: 251–267.

Lemaire, B., Van Cauwenberghe, J., Chimphango, S., Stirton, C., Honnay, O., Smets, E., *et al.* 2015. Recombination and horizontal transfer of nodulation and ACC deaminase (acdS) genes within Alpha- and Betaproteobacteria nodulating legumes of the Cape Fynbos biome. *FEMS Microbiol. Ecol.* **91**: 1–11.

Lenormand, T. 2002. Gene flow and the limits to natural selection. *Trends Ecol. Evol.* **17**: 183–189.

McKay, J.K. & Latta, R.G. 2002. Adaptive population divergence: Markers, QTL and traits. *Trends Ecol. Evol.* **17**: 285–291.

McKay Curtis, S. 2015. *mcmcplots: Create plots from MCMC output. R package version 0.4.2.*

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., *et al.* 2010. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297–1303.

Mhadhbi, H., Jebara, M., Limam, F., Huguet, T. & Aouani, M.E. 2005. Interaction between Medicago truncatula lines and Sinorhizobium meliloti strains for symbiotic efficiency and nodule antioxidant activities. *Physiol. Plant.* **124**: 4–11.

Mylona, P., Pawlowski, K. & Bisseling, T. 1995. Symbiotic Nitrogen Fixation. *Plant Cell* **7**: 869–885.

Newman, E., Manning, J. & Anderson, B. 2015. Local adaptation: Mechanical fit between floral ecotypes of Nerine humilis (Amaryllidaceae) and pollinator communities. *Evolution (N. Y).*

**69**: 2262–2275.

Oldroyd, G.E.D. 2013. Speak, friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat. Rev. Microbiol.* **11**: 252–63. Nature Publishing Group.

Oldroyd, G.E.D., Harrison, M.J. & Paszkowski, U. 2009. Reprogramming plant cells for endosymbiosis. *Science (80-. ).* **324**: 753–755.

Parker, M.A. 1999. Mutualism in metapopulations of legumes and rhizobia. *Am. Nat.* **153**: S48–S60.

Pavlidis, P., Jensen, J.D., Stephan, W. & Stamatakis, A. 2012. A critical assessment of storytelling: Gene ontology categories and the importance of validating genomic scans. *Mol. Biol. Evol.* **29**: 3237–3248.

Plummer, M., Best, N., Cowles, K. & Vinces, K. 2006. CODA: Converence Diagnostics and Output Analysis for MCMC. *R News* **6**: 7–11.

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

R Core Team. 2016. R: A language and environment for statistical computing. Vienna, Austria.

Rome, S., Cleyet-Marel, J.-C., Materon, L.A., Normand, P. & Brunel, B. 1997. Rapid identification of Medicago nodulating strains by using two oligonucleotide probes complementary to 16S rDNA sequences. *Can. J. Microbiol.* **43**: 854–861.

Rostas, K., Kondorosi, E., Horvath, B., Simoncsits, a & Kondorosi, a. 1986. Conservation of extended promoter regions of nodulation genes in Rhizobium. *Proc. Natl. Acad. Sci. U. S. A.* **83**: 1757–1761.

Savolainen, O., Lascoux, M. & Merilä, J. 2013. Ecological genomics of local adaptation. *Nat.*

696 *Rev. Genet.* **14**: 807–20.

697 Simonsen, A.K. & Stinchcombe, J.R. 2014a. Herbivory eliminates fitness costs of mutualism
698 exploiters. *New Phytol.* **202**: 651–61.

699 Simonsen, A.K. & Stinchcombe, J.R. 2014b. Standing genetic variation in host preference for
700 mutualist microbial symbionts. *Proc. R. Soc. B Biol. Sci.* **281**: 20142036–20142036.

701 Spiegelhalter, D.J., Best, N.G., Carlin, B.P. & Van Der Linde, A. 2002. Bayesian measures of
702 model complexity and fit. *J. R. Stat. Soc. Ser. B (Statistical Methodol.* **64**: 583–639.

703 Stanton-Geddes, J., Paape, T., Epstein, B., Briskine, R., Yoder, J., Mudge, J., *et al.* 2013.
704 Candidate Genes and Genetic Architecture of Symbiotic and Agronomic Traits Revealed by
705 Whole-Genome, Sequence-Based Association Genetics in Medicago truncatula. *PLoS One*
706 **8**: 1–9.

707 Suominen, L., Roos, C., Lortet, G., Paulin, L. & Lindström, K. 2001. Identification and structure
708 of the Rhizobium galegae common nodulation genes: evidence for horizontal gene transfer.
709 *Mol. Biol. Evol.* **18**: 907–916.

710 Tang, H., Krishnakumar, V., Bidwell, S., Rosen, B., Chan, A., Zhou, S., *et al.* 2014. An
711 improved genome release (version Mt4.0) for the model legume Medicago truncatula. *BMC*
712 *Genomics* **15**: 312.

713 Tiffin, P. & Ross-Ibarra, J. 2014. Advances and limits of using population genetics to understand
714 local adaptation. *Trends Ecol. Evol.* **29**: 673–680. Elsevier Ltd.

715 Turkington, R. & Cavers, P.B. 1979. The biology of Canadian weeds. 33. Medicago lupulina L.
716 *Can. J. Plant Sci.* **59**: 99–110.

717 van Rhijn, P. & Vanderleyden, J. 1995. The Rhizobium-Plant Symbiosis. *Microbiol. Rev.* **59**:

718 124–142.

719 Whitlock, M.C. 2015. Modern Approaches to Local Adaptation. *Am. Nat.* **186**: S000–S000.

720 Yan, J., Chu, H.J., Wang, H.C., Li, J.Q. & Sang, T. 2009. Population genetic structure of two

721 Medicago species shaped by distinct life form, mating system and seed dispersal. *Ann. Bot.*

722 **103**: 825–834.

723 Yoder, J.B., Briskine, R., Mudge, J., Farmer, A., Paape, T., Steele, K., *et al.* 2013. Phylogenetic

724 signal variation in the genomes of medicago (Fabaceae). *Syst. Biol.* **62**: 424–438.

725 Young, J.M. 2010. Sinorhizobium versus Ensifer: May a taxonomy subcommittee of the ICSP

726 contradict the Judicial Commission? *Int. J. Syst. Evol. Microbiol.* **60**: 1711–1713.

727 Young, N.D., Debellé, F., Oldroyd, G.E.D., Geurts, R., Cannon, S.B., Udvardi, M.K., *et al.*

728 2011. The Medicago genome provides insight into the evolution of rhizobial symbioses.

729 *Nature* **480**: 520–4.

730 Zuur, A.F., Ieno, E.N., Walker, N.J., Saveliev, A.A. & Smith, G.M. 2009. *Mixed Effects Models*

731 *and Extensions in Ecology with R*. Springer, New York, NY.

732

## Data accessibility

Sequence data will be uploaded to NCBI. VCF files and data from the reciprocal inoculation experiment will be available on Dryad. GPS coordinates of sampled plant and rhizobia populations are reported in Table S1.

737 **Tables and Figures**

738 Table 1. Results of general(ized) linear mixed models testing for local adaptation in the

739 reciprocal inoculation experiment.

| | | | | pMCMC |
|---|---|---|---|---|
| Seeds | Rhizobia | | | 0.204 |
| (*MCMC GLMM*) | Region | | | 0.070 |
| | Rhizobia × region | | | 0.350 |

| | | F | df | P |
|---|---|---|---|---|
| Biomass | Rhizobia | 1.955 | 1, 5.097 | 0.220 |
| (*LMM*) | Region | 0.131 | 1, 12.782 | 0.723 |
| | Rhizobia × region | 3.747 | 1, 248.656 | 0.054 |

| | | F | df | P |
|---|---|---|---|---|
| Flowering time | Rhizobia | 0.016 | 1, 5.436 | 0.903 |
| (*LMM*) | Region | 0.252 | 1, 12.896 | 0.624 |
| | Rhizobia × region | 1.378 | 1, 164.795 | 0.242 |

| | | Wald $\chi^2$ | df | P |
|---|---|---|---|---|
| Prob. of flowering | Rhizobia | 0.012 | 1 | 0.912 |
| (*GLMM*) | Region | 0.047 | 1 | 0.829 |
| | Rhizobia × region | 0.231 | 1 | 0.631 |

| | | Wald $\chi^2$ | df | P |
|---|---|---|---|---|
| Nodules | Rhizobia | 0.107 | 1 | 0.743 |
| (*GLMM*) | **Region** | **5.581** | **1** | **0.018** |
| | **Researcher** | **95.079** | **1** | **<0.001** |
| | **Rhizobia × region** | **34.806** | **1** | **<0.001** |

740

741 The type of model used is indicated below each trait. GLMM: generalized linear mixed model

742 (see text for error distribution). LMM: Linear mixed model (Gaussian error).

743

744 **Figure Legends**

745 Figure 1. A *Medicago lupulina* individual flowering in the greenhouse.

746

747 Figure 2. Locations of the 14 *M. lupulina* populations used in this study. The size of each circle
748 corresponds to the number of plants sampled from the population, and the color indicates the
749 rhizobia. The *E. medicae* strain used in the reciprocal inoculation experiment was obtained from
750 the northernmost population sampled ("SEG"); the *E. meliloti* strain was obtained from the
751 southernmost population ("DE"). See Table S1 for GPS coordinates.

752

753 Figure 3. Least squares means and 95% confidence intervals for northern (black) and southern
754 (white) plants grown in the two rhizobia treatments. *Ensifer medicae* is the locally abundant
755 rhizobia in the north, and *E. meliloti* is the locally abundant rhizobia in the south. (A) Number of
756 seeds; (B) aboveground biomass; (C) flowering time; (D) number of nodules.

757

# Supplementary Methods

## Appendix 1

*Bioinformatics and SNP discovery in* Ensifer

We aligned forward and reverse rhizobia reads to the reference genome of *E. meliloti* strain 1021 (Galibert *et al.*, 2001) (NCBI references chromosome AIL591688, plasmid a AE006469, plasmid b AL591985) and the *E. medicae* strain WSM419 (Reeve *et al.*, 2010) (NCBI references chromosome 150026743 plasmid b 150030273, plasmid a 150031715, accessory plasmid 150032810) using BWA (Li and Durbin, 2009) and Stampy (Lunter and Goodson, 2011) with default parameters and the bamkeepgoodreads parameter. We assigned bacterial species using a combination of the percentage of reads mapping to one reference genome, and sequences at the 16S rDNA locus (NCBI gene references 1234653 and 5324158, respectively), which differs between *E. medicae* and *E. meliloti* (Rome *et al.*, 1997). We used Integrative Genomics Viewer to visualize and check alignment quality (Robinson *et al.*, 2011). In general, 69.99 – 94.02% (median 84.71%) of reads per sample mapped to the *E. meliloti* reference genome, and 69.32 – 92.48% (median 83.49%) mapped to the *E. medicae* genome.

We used PICARD tools to format, sort, and remove duplicates in sequence alignments. We applied GATK version 3 indel realignment and GATK Unified Genotyper SNP discovery on all bacteria alignments (McKenna *et al.*, 2010) with ploidy set to haploid. We used the Select Variants parameter in GATK to select SNP variants only. We used standard hard filtering parameters and variant quality score recalibration on SNP discovery according to GATK Best Practices (DePristo *et al.*, 2011; Van der Auwera *et al.*, 2013). We filtered rhizobia SNPs for a minimum read depth (DP) of 20, a maximum DP of 226 for *E. meliloti* (230 for *E. medicae*), and

780 a genotype quality (GP) of 30 using vcftools (Danecek *et al.*, 2011). We removed indels and sites
781 with more than 10% of missing data from both *E. meliloti* and *E. medicae* data files. We
782 identified synonymous SNPs using SnpEff (Cingolani, Platts, *et al.*, 2012b) and SnpSift
783 (Cingolani, Patel, *et al.*, 2012a), using reference files GCA_000017145.1.22 and
784 GCA_000006965.1.22 (for *E. medicae* and *E. meliloti*, respectively) in the pre-built database.
785 We used the ANN annotation parameter in SnpSift to identify SNPs as synonymous variants and
786 missense variants.
787
788 *Bioinformatics and SNP discovery in* Medicago
789 We called *Medicago* SNPs in GBS samples by following the three-stage pipeline in the
790 program Stacks (Catchen *et al.*, 2011; 2013): cleaning raw data, building loci, and identifying
791 SNPs. We trimmed reads to 64 bp and filtered reads by a phred score of 33, the default value for
792 GSB reads sequenced on Illumina 2000/2500 machine. We built loci for *M. lupulina* using the *de*
793 *novo* approach in Stacks (denovo_map command), setting the –m parameter at 5, the –M
794 parameter at 1, and the -n parameter at 1. In the final stage of the pipeline, we identified SNPs
795 under the populations command by setting the –m parameter at 5. We filtered SNPs by removing
796 indels, removing sites with more than 10% of missing data, and removing sites that were less
797 than 64 bps apart with vcftools (Danecek *et al.*, 2011). We also excluded 9 SNPs with
798 heterozygosity that was higher than expected under Hardy-Weinberg.
799
800 **Appendix 2**
801 *Genomic outlier tests*

802 We first performed the BLAST test in two ways: first using the range-wide sample of
803 plants that hosted different bacterial species (73 plant individuals), and second, focusing on
804 southern Ontario samples (49 plant individuals). We performed the latter test because of the
805 possibility that many loci unrelated to bacterial specificity (e.g., climatic adaptation) could be
806 differentiated between southern Ontario and the mid-Atlantic United States due to environmental
807 gradients that covary with bacterial species composition.

808 Outlier loci detected in genotyping-by-sequencing (GBS) data are rarely the actual loci
809 responsible for adaptation; instead, they are usually in linkage disequilibrium (LD) with the
810 causal genes. To account for this possibility, we then searched for genes involved in the legume-
811 rhizobia symbiosis within either 5 or 10 kb of the *M. truncatula* orthologs of the outlier loci that
812 we detected in both the range-wide and Ontario samples. This approach assumes synteny
813 between *M. truncatula* and *M. lupulina*. We chose 5 and 10 kb based on the scale of LD in *M.*
814 *truncatula* (Branca *et al.*, 2011). While the scale of LD between even closely related species is
815 likely to differ based on mutation rates, recombination, population structure, and a host of other
816 demographic and evolutionary factors, we viewed this approach as superior to simply confining
817 our searches to the GBS loci without accounting for their potential LD with causal genes.

818 Finally, we measured the distance between the *M. truncatula* orthologs of the outlier loci
819 that we detected in both the range-wide and Ontario samples and key *M. truncatula* genes
820 involved in the rhizobia symbiosis (again assuming synteny between *M. truncatula* and *M.*
821 *lupulina*). We considered genes involved in the initial signal exchange between the legume and
822 rhizobia (NSP, IPD3, and DMI1-DMI3); genes involved in infection thread development (LIN);

and genes involved in both rhizobia signaling and infection (NFP, LYK3, and NIN) (Jones *et al.*, 2007; Oldroyd *et al.*, 2009; Young *et al.*, 2011; Oldroyd, 2013; Tang *et al.*, 2014).

## Appendix 3

*Genomic outlier results*

We identified a distribution of $X^TX$ statistics around the null expectation of $X^TX = 2$, reflecting the 2 populations assigned in Bayenv2 (*M. lupulina* plants hosting *E. medicae* and plants hosting *E. meliloti*). In the range-wide sample, 16% (354 of 2209) of SNPs had $X^TX$ scores greater than the null expectation of 2; in the Ontario sample, 29% (573 of 1977) of SNPs had $X^TX$ scores greater than 2. We detected a range of alignment scores when we used BLAST to align outlier loci with top $X^TX$ statistics from the whole sample and the Ontario sample to the *M. truncatula* reference genome (Table 2). The loci mapped to several different chromosomes in the *M. truncatula* reference genome.

Of the top 1% of SNPs detected in the range-wide sample (20 SNPs total), eight were associated with a specific *M. truncatula* gene (BLAST scores: 35.6 – 102; E value: 1.00e-19 – 0.31). Higher BLAST scores reflect higher-quality alignments; these scores indicate that our sequences generally aligned moderately well to the *M. truncatula* genome. E (expectation)-values reflect the number of hits expected by chance, so lower E-values indicate better matches. These 8 loci did not map to any genes known to be involved in the legume-rhizobia mutualism. The remaining 12 loci did not map to a specific gene in the *M. truncatula* genome (BLAST scores: 35.6 – 102; E-values: 3.00e-20 – 3.10e-1).

844 The results were qualitatively similar for the Ontario sample (Table 2). The BLAST
845 scores of the top 1% of outlier SNPs (20 SNPs total) ranged from 37.4 to 111 (E-values: 5.00e-
846 24 – 8.90e-2). Twelve of the top 1% of SNPs in the Ontario sample mapped to genes that are not
847 known to be involved in the legume-rhizobia mutualism. The remaining eight loci did not
848 associate with a specific gene in the *M. truncatula* annotated genome. The BLAST scores for
849 these loci were similar to the twelve loci that did map to specific *M. truncatula* genes (score:
850 35.6 – 95.1; E-value: 3.00e-20 – 3.10e-1).
851 There were only three outlier loci that appeared in the top 1% of SNPs in both the range-
852 wide sample and the Ontario sample (Table 3). These loci mapped to chromosomes 1, 5, and 7 in
853 the *M. truncatula* genome, but did not map to a specific gene. No genes found within 5 or 10kb
854 of the *M. truncatula* orthologs of these three outliers are known to be involved in the legume-
855 rhizobia symbiosis (assuming synteny between *M. truncatula* and *M. lupulina*). The *M.*
856 *truncatula* ortholog of the outlier on chromosome 5 had two genes within 5 kb, a phosphate
857 putative gene and a Ty3/Gypsy polyprotein/retrotransposon. The ortholog of the outlier on
858 chromosome 1 had no genes within a 5 kb window, and the ortholog of the outlier on
859 chromosome 7 had two genes within 5 kb, a DUF247 domain protein and a Gypsy-
860 likepolyprotein/retrotransposon putative gene. When we increased our window size to 10 kb we
861 found more genes, but none related to infection with rhizobia. For example, the *M. truncatula*
862 ortholog of the outlier on chromosome 5 was close to a DUF679 domain membrane protein and
863 an alpha/beta fold hydrolase putative gene. The ortholog of the outlier on chromosome 1 had a
864 reverse transcriptase zinc binding protein and a homeobox knotted-like protein in its 10 kb

865 window. The ortholog of the outlier on chromosome 7 had a phosphoenolpyruvate carboxylase
866 within its 10 kb window, along with several putative proteins.

867 Finally, we calculated the distance in base pairs between the *M. truncatula* orthologs of
868 the three outlier loci found in both the range-wide and Ontario analyses and several genes
869 involved in *Medicago*-rhizobia association. None of the symbiosis genes that we considered
870 were close to the orthologs of any of these three outliers. Most of the symbiosis genes are located
871 on chromosome 5, but none were close to the ortholog of the outlier locus on chromosome 5
872 (Table 4). The ortholog of the outlier locus on chromosome 1 was approximately 35,822 kb
873 away from the only symbiosis gene we considered that is located on chromosome 1 (LIN). The
874 remaining two symbiosis genes—DMI1 and DMI2—are located on chromosomes 2 and 8 (Ané
875 *et al.*, 2002), neither of which contained any outlier loci in our analysis.
876
877

878 **Supplemental Tables and Figures**

879

880 Table S1. Locations of *M. lupulina* and *Ensifer* populations used in population genetic analysis.

881

| Population ID | Province/State | Country | Latitude | Longitude |
|---|---|---|---|---|
| AV | Pennsylvania | United States | 40.200904 | -76.763663 |
| BR | Ontario | Canada | 43.749940 | -79.639721 |
| COB | Ontario | Canada | 44.103365 | -78.156738 |
| DE | Delaware | United States | 38.686381 | -75.07443 |
| HO | Ontario | Canada | 44.219923 | -81.049833 |
| KA | Ontario | Canada | 44.555739 | -78.838577 |
| MAP | Ontario | Canada | 43.836307 | -80.636623 |
| PA1 | Pennsylvania | United States | 39.930768 | -75.584929 |
| PAR | New York | United States | 40.659773 | -76.918652 |
| PT | Pennsylvania | United States | 40.07175 | -75.435691 |
| SEG | Ontario | Canada | 45.225519 | -79.682214 |
| SIN | Pennsylvania | United States | 41.457507 | -77.135632 |
| UBR | Delaware | United States | 39.626937 | -75.675749 |
| WA | Ontario | Canada | 43.596353 | -80.625336 |

882

883

884

885

886 Table S2. Summary statistics of Bayenv2 and BLAST results for the top 1% of SNPs in the $X^TX$
887 outlier analysis.
888

| **Ontario sample** | | | | | |
|---|---|---|---|---|---|
| **$X^TX$** | **SNP identity** | **BLAST** | **Query cover** | **E value** | **Gene** |
| 4.20 | 585117 | 37.4 | 56 | 0.089 | transmembrane protein |
| 3.83 | 1192907 | 93.3 | 100 | 1.00E-18 | indole-3-glycerol phosphate lyase IGL1 |
| 3.82 | 187811 | 93.3 | 98 | 1.00E-18 | toprim domain protein |
| 3.75 | 1110167 | 107 | 100 | 6.00E-23 | TLD-domain nuclear protein |
| 3.73 | 1610082 | 71.6 | 81 | 4.00E-12 | no result |
| 3.65 | 229813 | 35.6 | 60 | 3.10E-01 | no result |
| 3.59 | 1959186 | 107 | 100 | 6.00E-23 | TPR repeat protein |
| 3.56 | 129152 | 41 | 59 | 7.00E-03 | no result |
| 3.55 | 884266 | 89.7 | 100 | 2.00E-17 | no result |
| 3.44 | 616912 | 95.1 | 98 | 4.00E-19 | no result |
| 3.43 | 97240 | 111 | 100 | 5.00E-24 | transportin-1 protein |
| 3.39 | 666854 | 87.8 | 100 | 6.00E-17 | no result |
| 3.39 | 713735 | 96.9 | 100 | 1.00E-19 | indole-3-glycerol phosphate lyase IGL1 |
| 3.33 | 1294820 | 69.8 | 95 | 2.00E-11 | copia-like polyprotein/retrotransposon |
| 3.33 | 686219 | 69.8 | 100 | 2.00E-11 | no result |
| 3.33 | 1463455 | 91.5 | 100 | 5.00E-18 | single-stranded nucleic acid-binding protein R3H |
| 3.29 | 671122 | 37.4 | 85 | 8.90E-02 | cysteinyl-tRNA synthetase |
| 3.26 | 1071597 | 98.7 | 100 | 3.00E-20 | no result |
| 3.25 | 109965 | no result | no result | no result | no result |
| 3.25 | 734196 | 78.8 | 100 | 3.00E-14 | no result |
| **Range-wide sample** | | | | | |
| 3.25 | 825707 | 96.9 | 100 | 1.00E-19 | FAD/NAD(P)-binding oxidoreductase family protein |
| 3.06 | 1131532 | 89.7 | 100 | 2.00E-17 | novel plant SNARE-like protein |
| 3.02 | 1482582 | 35.6 | 39 | 0.31 | ASCH domain protein |
| 2.99 | 405789 | 89.7 | 93 | 2.00E-17 | CAAX amino terminal protease family protein |
| 2.99 | 511074 | 48.2 | 96 | 5.00E-05 | no result |
| 2.97 | 175058 | 55.4 | 96 | 3.00E-07 | no result |
| 2.95 | 485278 | 91.5 | 93 | 5.00E-18 | no result |
| 2.94 | 254373 | 59 | 100 | 3.00E-08 | CRS1/YhbY (CRM) domain protein |
| 2.92 | 1090953 | 55.4 | 95 | 3.00E-07 | galactose oxidase |
| 2.90 | 921907 | 93.3 | 100 | 1.00E-18 | DUF223 domain protein |
| 2.86 | 1071597 | 98.7 | 100 | 3.00E-20 | no result |
| 2.82 | 870953 | 35.6 | 51 | 3.10E-01 | |
| 2.81 | 1590352 | 102 | 92 | 3.00E-21 | no result |
| 2.75 | 686219 | 69.8 | 100 | 2.00E-11 | no result |
| 2.73 | 774291 | 69.8 | 95 | 2.00E-11 | no result |
| 2.71 | 313573 | 69.8 | 98 | 2.00E-11 | no result |
| 2.68 | 1342844 | 60.8 | 90 | 8.00E-09 | |
| 2.67 | 1217147 | 89.7 | 100 | 2.00E-17 | novel plant SNARE-like protein |

| | | | | | |
|---|---|---|---|---|---|
| 2.64 | 616912 | 84.2 | 100 | 7.00E-16 | no result |
| 2.64 | 1057909 | 66.2 | 70 | 2.00E-10 | no result |

889

Table S3. Outlier loci found in the top 1% of Bayenv2 results in both the *M. lupulina* range-wide and Ontario samples.

| $X^TX$ | SNP identity | BLAST | Query cover | E value | Gene name |
|---|---|---|---|---|---|
| 3.26 | 1071597 | 98.7 | 100 | 3.00E-20 | no result |
| 3.44 | 616912 | 95.1 | 98 | 4.00E-19 | no result |
| 3.33 | 686219 | 69.8 | 100 | 2.00E-11 | no result |

Table S4. Base pair distances between the *M. truncatula* ortholog of the outlier locus on chromosome 5 and well-characterized nodulation and rhizobial infection genes in *M. truncatula*.

| Gene name | Distance (kb) |
|---|---|
| NFP (Nod-factor receptor 5) | 12397 |
| NSP (Nodulation receptor kinase-like protein) | 6319 |
| LYK3 (LysM receptor kinase K1B) | 18 |
| NIN (Nodule inception protein) | 24 |
| IPD3 (Cyclops protein putative) | 8469 |

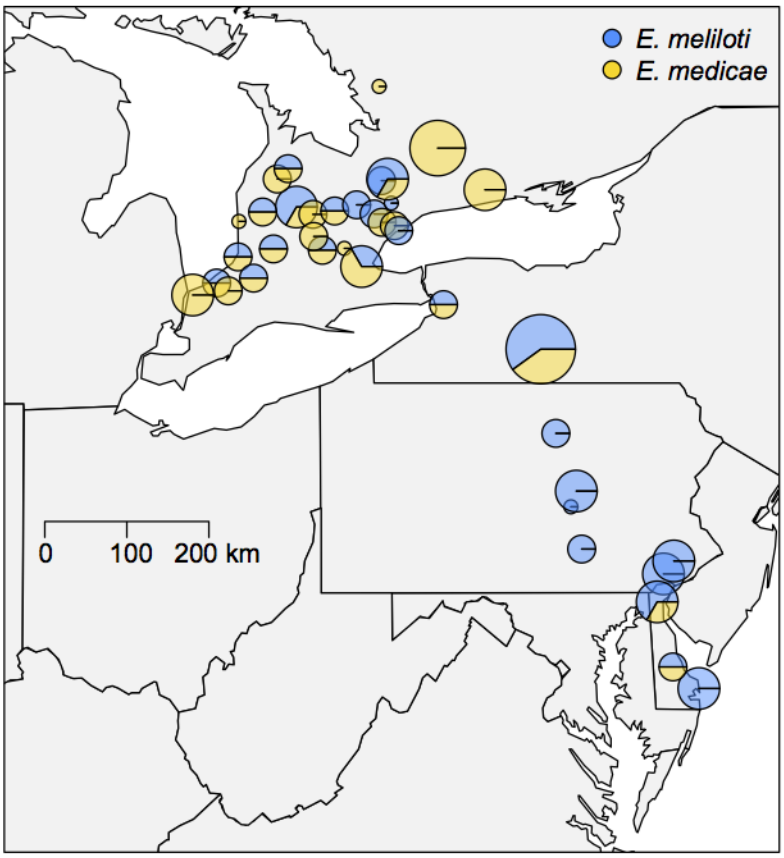
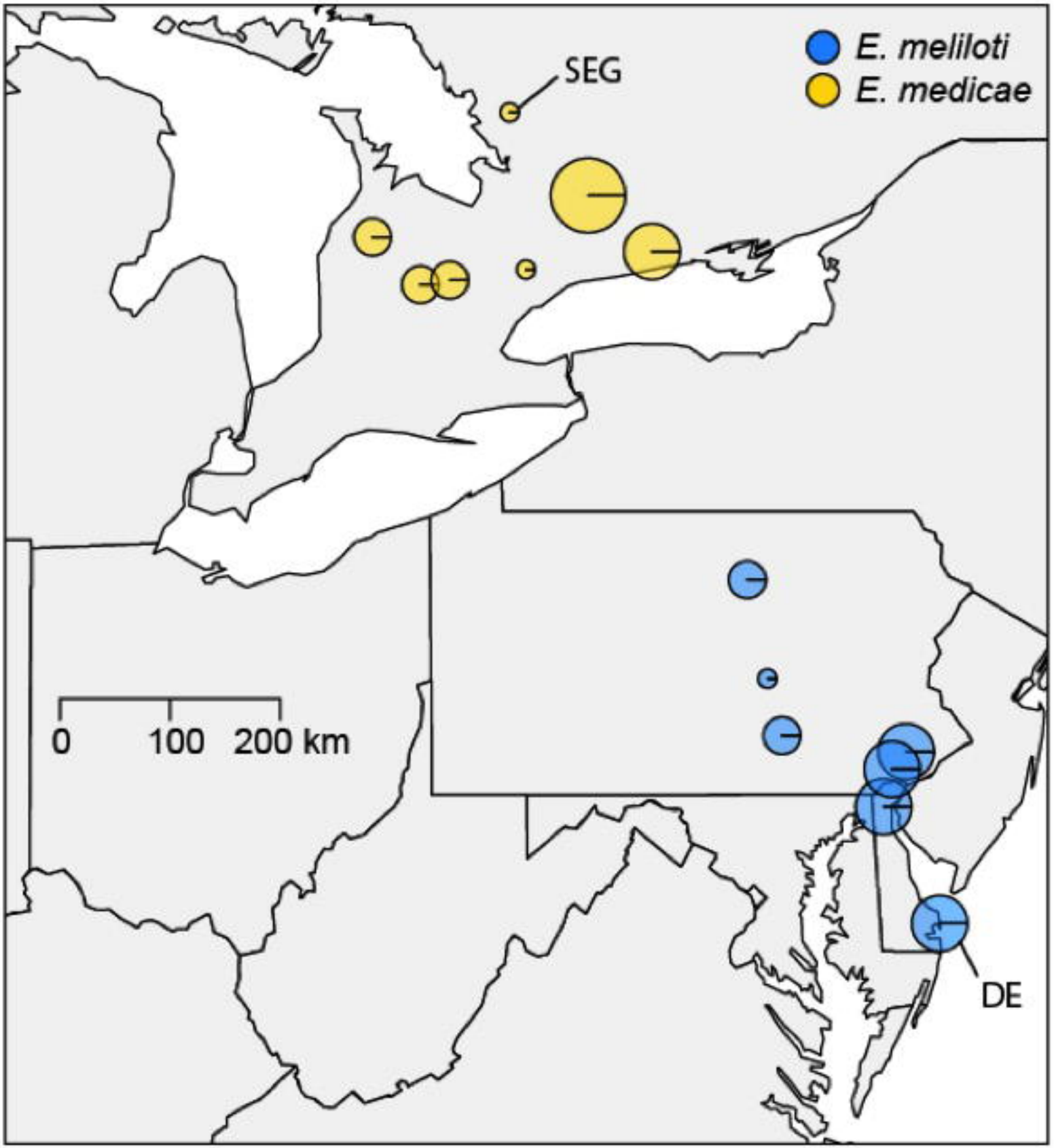



Figure S1. Locations and bacterial species present in all 39 sampled *M. lupulina* populations. The size of each circle corresponds to the number of sampled plant individuals for which their bacterial partner was identified to species. The colors correspond to the fraction of plants partnered with *E. meliloti* (blue) and *E. medicae* (yellow). The map is modified from Figure 2 in Harrison et al. (2017).


E. meliloti
E. medicae
SEG
DE
0
100
200 km