PARTNERS IN SPACE: DISCORDANT POPULATION STRUCTURE BETWEEN LEGUME HOSTS AND RHIZOBIUM SYMBIONTS IN THEIR NATIVE RANGE

BY

ALEXANDER RILEY

THESIS

Submitted in partial fulfillment of the requirements
for the degree of Master of Science in Plant Biology
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2020

Urbana, Illinois

Advisers:

Associate Professor Katy D. Heath
Associate Professor Amy Marshall-Colón

Abstract

To understand the coevolutionary dynamics of an interaction, theory suggests that we must study how genetic variation in both partners is structured in space. This is because, in coevolution, the spatial distribution of genetic variation in one species can both determine its ability to locally adapt to a partner species and also serve as an agent of selection acting on that partner species. Plant-microbe symbioses are ecologically and economically important, but broad-ranging dispersal and dynamic genome structure in bacteria present unique challenges for understanding spatial genetic processes in these systems. Here we study the model rhizobium *Ensifer meliloti* using a hierarchically structured sample of 191 strains from 21 sites in the native range and compare its population genetic structure to that of its host plant *Medicago truncatula*. We find that two of the three elements of the tripartite *Ensifer* genome lack a pattern of isolation by distance. Overall genetic variation across the symbiont genome is less spatially structured than that of its host, and variation in the two species is uncorrelated, indicating comparatively higher levels of gene flow among sampling sites in these bacterial symbionts relative to host plants. Taken together our results suggest that the spatial structure of genetic variation is impacted differently by the environment in the host and in the three symbiont genomic elements. This could lead to differing responses to selection not only between the host and symbiont, but also between the elements of the symbiont genome.

Acknowledgements

So many people have helped me through these last couple years, and I wish I had the space or memory to thank all of them, but here are a select few folks without whom this work could not have happened. First, I would like to thank my amazing advisors, Amy Marshall-Colón, and Katy Heath, who have been amazing mentors and supported me through the all the twists and turns in the last few years of my life, both professional and person. I cannot overstate how important they have both been at various times over the course of this degree, and I only hope that someday I can have the impact on folks that I teach and mentor that they've had on me.

I would also like to thank my committee members, Rachel Whitaker and Julian Catchen, who each consistently brought new perspectives to this work and always left me with a way to make it substantially stronger, and more meaningful. Moreover, I would like to thank the entire School of Integrative Biology for providing such a supportive and collaborative space to grow as a scientist. Particularly my lab mates in the Heath lab throughout this journey, who have been exceptional friends and colleagues, and who have facilitated this work as much by being a sounding board for my office ramblings, as they have through tangible contributions. Thanks everyone!

And finally I would like to thank my family for their never-ending support throughout this wild process. In particular I'd like to thank my daughter Violet, who's been a never ending source of joy and inspiration, and who's been very patient with me when she asks me a question and I don't know the answer, even though I am a scientist, thank you.

To those I've mentioned and those I haven't, thank you so much for the never-ending support, you're all appreciated more than I can ever express.

TABLE OF CONTENTS

Chapter 1: A Tale of Two Genomes

Introduction

Host-microbe symbioses are ubiquitous in nature, underlying crucial processes at every level of biological organization (Moran 2007). Our understanding of the functions of microbial symbionts in natural systems, and how the information encoded by their genomes cascades up through hosts to affect the larger communities and ecosystems in which they occur, is growing rapidly (Rosenberg and Zilber-Rosenberg 2018; Simon et al. 2019). These critical symbiotic functions have come to be through coevolutionary processes between macrobes and microbes (Thrall et al. 2007; Revillini et al. 2016; Pita et al. 2018). Coevolutionary theory has largely grown out of macrobial population biology addressing patterns of gene flow and spatially-variable selection in a metapopulation framework (Thompson 2005), while our empirical understanding of host-symbiont coevolution is built largely on microcosm experiments, which resolve mechanism but lack an explicit geographical context (Brockhurst and Koskella 2013). This disconnect points to the need for detailed studies of population-level processes in natural host-microbe symbiosis.

In coevolutionary interactions between two species, spatial genetic structure in one species can both underlie the potential for that species to locally adapt *to* its partner, and also serve as one aspect of the selective environment *for* its partner (Gandon et al. 1996; Burdon and Thrall 2000). Coevolution takes place across landscapes, with selection, drift, and migration dictating how interaction traits evolve, at what spatial scale they evolve, and even whether coevolution occurs at all (Laine 2005; Thompson 2005; Tack et al. 2014; Carlsson-Granér and Thrall 2015; Fernandes et al. 2019). Coevolutionary theory predicts that comparative population structure is important for determining the potential for coadaptation, the geographic scale at

which coadaptation should occur, and which partner (if either) has an evolutionary advantage over the other (Gandon et al. 1996; Fernandes et al. 2019). Empirical work comparing the population structure of interacting species finds a range of outcomes, from correlated spatial genetic patterns between partners (Anderson et al. 2004; Thompson et al. 2005; Caldera and Currie 2012), to largely discordant patterns, where one species exhibits substantially less population structure across the sampled geographic range, compared to its partner (Dybdahl and Lively 1996; Baums et al. 2014; Strobel et al. 2016). Understanding the population genetic structure of both hosts and symbionts provides key perspective on the potential for spatial heterogeneity to play an important role in maintaining genetic variation in mutualism and thus for interpreting spatial variation in symbiosis traits (Heath and Stinchcombe 2014; Hollowell et al. 2016a).

One challenge for understanding the (co)evolution of bacterial symbionts in nature is the difficulty of isolating populations of a single species from the environment - particularly for soil bacteria in hyper-diverse communities (Heath and Grillo 2016). To add to this complexity, bacterial genomes are often multipartite (diCenzo and Finan 2017; Harrison et al. 2010), including non-chromosomal replicons containing variable amounts of genetic information – ranging from small facultative plasmids that are often lost or transferred to large obligate segments containing core genes (often as megaplasmids or chromids) (Harrison et al. 2010). Such divided genomes might represent an adaptation allowing functional division between replicons (reviewed by diCenzo and Finan 2017). Multipartite genome organization presents a unique challenge in the study of bacterial population genetics because these different elements may move between individuals independent of each other via horizontal gene transfer, potentially resulting in distinct evolutionary histories and population genetic processes across

different elements (Galardini et al. 2013). These differences are compounded by the fact that these elements can also have widely differing rates of evolution and patterns of recombination due to their gene content ranging from facultative to obligate, leading to higher rates of evolutionary change in plasmids not required for the survival of the individual in its environment (Cooper et al. 2010).

The legume-rhizobium mutualism presents a key model for studying genetic variation in symbiosis across broad spatial scales because symbiotic bacteria are easily isolated from soil populations across the range (Barrett et al. 2012; Heath and Grillo 2016; Hollowell et al. 2016a). In this interaction, soil-borne rhizobia form symbiosis with plant roots, wherein they fix atmospheric nitrogen (N) in exchange for carbon derived from plant photosynthesis. Due to the economic and ecological importance of the mutualism, genetic resources for several legume-rhizobium partnerships are well developed (e.g., Kaneko et al. 2002; Yates et al. 2015), particularly for the model legume *Medicago truncatula* and its rhizobium symbiont *Ensifer meliloti* (Becker et al. 2009; Tang et al. 2014). In the interaction's native range around the Mediterranean genetic variation in *Medicago* exhibits a pattern of isolation by distance, and is differentiated at the population and regional levels (Bonnin et al. 1996; Ronfort et al. 2006; Siol et al. 2008; Bonhomme et al. 2015; Grillo et al. 2016). In the one previous study comparing the population structure of *Ensifer* species and a closely related host species (*Medicago lupulina*), which was focused on a 300km subset of the mutualism's introduced range in North America, Harrison et al. (2017) found that *Medicago* exhibited significant isolation by distance, while *Ensifer* symbionts did not. Understanding the coevolution of this model symbiosis requires that we address corresponding patterns of genetic variation in *Medicago* plants and *Ensifer* symbionts

in the native range, using a hierarchical sampling design to address within- and among-populations variation, and across the three elements of the multipartite genome.

Here we use whole genome sequences for 191 isolates of *E. meliloti* sampled from 21 sites in the native range around the Mediterranean to better understand the population biology of this model mutualism. We also reanalyze *Medicago* sequence data from Grillo *et al.* (2016) to directly compare population genomic patterns in the two organisms. We ask four major questions: 1) Do the three genomic elements of *E. meliloti* share the same evolutionary history and patterns of population structure at this spatial scale? 2) Do *E. meliloti* populations exhibit a pattern of isolation by distance like their plant hosts? 3) Where are major genetic clades of each species located in space? And finally 4) are the genetic distances between populations in *E. meliloti* correlated with the genetic distances between corresponding *Medicago* populations indicated corresponding population genetic structure?

Methods

Study system: M. truncatula is one of many annual legumes in its genus native to the Mediterranean basin. The species also has naturalized populations in other regions of the world with Mediterranean climates, and has been planted in limited cases as a forage crop for livestock, or as a cover crop (von Wettberg et al. 2020). Populations of *M. truncatula* (also referred to hereafter as *Medicago* or the host) are highly selfing, with rates as high as 0.99 (Bonnin et al. 1996; Siol et al. 2008).

Ensifer (formerly *Sinorhizobium*) *meliloti* is a rhizobium species in the *Alphaproteobacteria* (Young and Haukka 1996) that forms N-fixing nodules on the roots of multiple species in the genus *Medicago*, and are one of two *Ensifer* species to form root nodules

in symbiosis with *M. truncatula*, the other being *E. medicae* (Zribi et al. 2004).The genome of *Ensifer meliloti* is ~6.79Mb, divided between three genomic elements: the chromosome (3.69Mb), megaplasmid pSymA (1.41Mb), and chromid pSymB (1.69Mb) (Nelson et al. 2018). Metabolic modeling, along with molecular manipulation of the *E. meliloti* genome has shown that gene content differs functionally between the three genomic elements in *E. meliloti* (also referred to hereafter as *Ensifer* or the symbiont) (diCenzo et al. 2014). The chromosome carries primarily genes related to core metabolic function of individuals in bulk soil, while pSymA carries the majority of genes required for symbiotic N fixation, and pSymB carries primarily genes important for life in rhizosphere environments (diCenzo et al. 2014).

Sample collection: We isolated *Ensifer* strains from 21 sites in the native range of *M. truncatula* using a hierarchical sampling design with populations ranging from 1 to approximately 1400 km apart (Fig1; for collection details on each site see Supplemental Table 1). Samples were collected from the top six inches of soil directly surrounding the roots of an unearthed *Medicago* plant. To avoid cross-contamination within each site, the sampling shovel was wiped clean of excess soil in between samples and was pierced into the ground adjacent to a plant numerous times before extracting the sample. Between sampling locations, the shovel was sterilized with dilute bleach. Soil samples were stored in double wrapped plastic bags and kept in a 4°C refrigerator prior to isolating cultures. Of these 21 sites, eight corresponded to the *Medicago* sites studied in Grillo *et al*. (2016). An additional four sites from Grillo *et al*. (2016) were not sampled because they were no longer accessible or no longer harbored populations of *M. truncatula*.

Pure cultures of *E. meliloti* strains were isolated or “trapped” in the laboratory from the field samples following standard protocols (Vincent 1970). For this, *Medicago* seeds were nicked

with a razor blade, surface sterilized with 30% bleach, rinsed three times with sterile water, and imbibed in sterile water for approximately 30 minutes. These seeds were then directly sown into a given soil sample housed in a sterilized, fully self-contained Magenta box to prevent cross-contamination (see Brown et al. 2020 for details). Magenta pots were randomly placed in a temperature-controlled grow room (23 ° C) under artificial light set to 12-h days. After four weeks, plants were harvested, and the soil was washed away from the roots. Individual nodules were removed with forceps and were surface sterilized by soaking in 30% bleach for 10 minutes and then rinsed 5 times with sterilized water. Surface sterilized nodules were then crushed with sterilized forceps and were streaked on tryptone-yeast (TY) media plates. Plates were incubated at 30℃ for 48 hours, at which point sterilized glass stir rods were used to streak samples on new TY spread plates to isolate distinct colonies which were again incubated at 30℃. Individual colonies were then picked and grown in liquid TY media, and these pure cultures were stored in cryotubes in 50% glycerol and placed in -80℃ freezer. Given known variation among host genotypes in rhizobium infection rates due to G x G interactions (Heath and Tiffin 2009; Batstone et al. 2017), we used 24 different *Medicago* host genotypes to trap rhizobia. We confirmed that host genotype did not strongly impact the genetic variation of sampled strains using an AMOVA to test for among-host genotype variance in the symbiont genome, run with the R package Poppr (Kamvar et al. 2014) (Sup table 2). Because both *E. meliloti* and *E. medicae* infect the roots of *Medicago* in the native range, we used a post-PCR restriction enzyme (RsaI) digestion of the 16S to assign strains to species (following Biondi et al. 2003). Ultimately, we isolated 191 strains of *E. meliloti,* presented in the current study, 176 strains of *E. medicae,* and several additional bacterial species, which will be analyzed elsewhere.

Sequencing: We extracted DNA from cultures of *E. meliloti* grown in liquid TY media using the Qiagen DNeasy kit (Hilden, Germany), quantified DNA using the Qubit high sensitivity double stranded DNA assay (Thermo Fisher Scientific, Waltham, MA, USA), and sent samples to the DOE Joint Genome Institute (JGI) for sequencing (Berkeley, CA, USA). JGI prepared a paired end 101nt sequencing library for each strain, and sequenced samples on an Illumina HiSeq-2500 1TB platform (Illumina, Inc., San Diego, CA, USA). Of the 199 strains thought to be *E.meliloti* submitted to JGI, we received high quality whole genome sequences for 166; the remaining 33 strains were regrown from frozen cultures (as above), extracted using the Zymogen Quick-DNA kit for Fungi or Bacteria (Irvine, CA, USA), then prepped (2 X 150 or paired end 150 nt read length) and sequenced on the Novaseq 6000 platform (Illumina, Inc, San Diego, CA, USA) by the Roy J. Carver biotechnology center at the University of Illinois at Urbana-Champaign (USA). After removing sequences that did not align to the reference genome, we recovered genome sequences from another 25 isolates, for a total of 191 *E. meliloti* strains sequenced.

Genome assembly, annotation, and SNP calling: To ensure high quality SNP calling and genome assembly, we removed all nucleotides with quality scores below 30, then removed all reads with lengths less than 80nts using TrimGalore! (https://github.com/FelixKrueger/TrimGalore). Finally, we trimmed adaptor sequences, PCR replicates, and removed PhiX contamination using HT-Stream (https://github.com/s4hts/HTStream). For calling core SNPs, we first aligned reads to the *E. meliloti* reference genome USDA1106 using BWA with default settings (Li and Durbin 2009). Using Freebayes (Garrison and Marth 2012), we called haplotype variants then split variants into primitive SNPs using VCFtools (Danecek et al. 2011). We found 491,277 variable SNPs with variant qualities above 20. We then filtered SNPs to retain only those with depth

values between 20 and 230, minor allele frequencies ≥0.009, and that were present in at least 80% of individuals, after which 72,311 SNPs remained (chromosome: 34,689 SNPs, pSymA: 15,162 SNPs, pSymB: 22,460 SNPs; see Sup table 3.). For variable gene content, we also assembled genomes *de novo* using SPADES (Bankevich et al. 2012) with default parameters, followed by annotation using PROKKA (Seemann 2014) and scanned these annotations for presence absence variants using default settings in ROARY (Page et al. 2015).

Phylogenetic and Statistical Analyses: We built phylogenies based on a random subsample of 15k core genome SNPs for each of the three *Ensifer* genome elements (chromosome, pSymA, pSymB) using the neighbor joining (nj) function in the R package ape (Paradis and Schliep 2019). We used mantel tests implemented in the R package ade4 (Dray and Dufour 2007) to test whether individual-based principal component genetic distances among strains were correlated across the genome elements (e.g., between chromosome and pSymA). Based on our findings (see Results), we treated each element independently for all remaining analyses.

To explore the population structure of *Ensifer*, we first used principal components analysis (PCA) implemented in the glPca function in the adegenet (Jombart and Ahmed 2011) library in R to naively cluster individuals by genome-wide similarity. Next we used AMOVA with clone correction and 10,000 random permutations (PoppR; Kamvar, Tabima, and Gŕunwald 2014) to partition the genetic variance among individuals, sites, and regions, as well as to assess the significance of these divisions. Next we used Pearson correlations between individual-level genetic distances and geographic distances between sampling sites to test the hypothesis of isolation by distance in *Ensifer*. We calculated individual-based distance as Euclidean distance in principal component space including the number of dimensions accounting for 80% of variation

in a given element (3 for the chromosome, 26 for pSymA, and 31 for pSymB). For plotting, PC-based distances were normalized for each element separately by dividing distance values by the maximum distance value for that element. We used F_{ST}, calculated using STampp (Pembleton et al. 2013), as a metric of among-population genetic distance.

To compare the spatial genetic structure of rhizobia to that of its host plant, we reanalyzed RAD-seq data from the 192 *Medicago* genotypes studied in Grillo *et al* (2016). We first called SNPs using Stacks with default parameters (Catchen et al. 2013), then filtered the resulting variants using VCFtools (Danecek et al. 2011) to ensure that all variants were present in at least 80% of lines, had minor allele frequencies of at least 0.05, and were >5kb apart. We calculated individual and population-based genetic distances and tests for isolation by distance in *Medicago* as detailed above for *Ensifer*. We also ran an AMOVA analysis for *Medicago* as detailed above for *Ensifer*. For the subset of eight sites for which we had both hosts and symbionts, we used a Mantel test to correlate pairwise population F_{ST} values between hosts and symbionts. Finally, we used PCA on the matrix of *Ensifer* gene presence-absence variants and plotted the first three principal components to assess whether presence-absence variation was structured in space.

Results and discussion

Incongruence in the tripartite genome of the rhizobium E. meliloti*:* For each pairwise combination of strains, we computed individual-based genetic distances at each of the three genome elements, then used pairwise Mantel tests to ask whether these three matrices were correlated, i.e., whether the three replicons have distinct evolutionary histories. We found that the chromosomal distance matrix was uncorrelated with that of pSymB (Table 1). We found a

weak but significant correlation between the chromosome and pSymA, and a stronger correlation between the two symbiotic plasmids (pSymA and pSymB). Neighbor joining trees showed qualitatively different evolutionary histories between the elements. The topology of the chromosomal tree showed several tightly clustered groups of individuals, while the trees for both symbiotic plasmids are made up of broader clades with longer branches separating sister strains (Sup Figs. 1-3). Based on the lack of tight correlations in the pairwise comparisons between elements, and the differences in the structures of their phylogenies, we analyze each element independently hereafter.

Our results suggest that the three genomic elements have distinct evolutionary histories in *E. meliloti*, consistent with previous work showing independent evolution of the three elements in this species (generally using global, or regional samples of single individuals from any given population) (Galardini et al. 2013; Epstein et al. 2014; Nelson et al. 2018), as well as in other rhizobium taxa with symbiosis plasmids (Young et al. 2006; Klinger et al. 2016; Pérez Carrascal et al. 2016) or symbiosis gene regions (Hollowell et al. 2016b; Porter et al. 2019). Based on functional genetics and metabolic models (diCenzo et al. 2014; diCenzo and Finan 2017), the chromosome, pSymA, and pSymB are thought to play distinct roles in the life history of *Ensifer*. Our results suggest substantial recombination via horizontal gene transfer among the three elements and thus that evolutionary change due to selection on traits governed largely by chromosomal genes (i.e., those putatively important for life in bulk soil) might proceed independently from evolutionary changes in the pSymA and pSymB genes that are posited to govern symbiotic N fixation or life in the rhizosphere. This might imply that fitness tradeoffs between life in bulk soil and life in the nodule (symbiont quality), which have been proposed as a force maintaining partner quality variation in natural population of rhizobia (Friesen and Mathias

2010; Heath and Stinchcombe 2014), are unlikely to be common in this particular interaction; however, this hypothesis relies on many assumptions and requires careful experimental work to test rigorously.

Mismatched population structure: First we used AMOVA to partition the genome-wide genetic variation in the rhizobium to the within-population, among-population, and among-region scales, and compare it to its host plant. Interestingly our results depended on the particular element of the tripartite *Ensifer* genome. Overall, genetic variation in the rhizobium symbiont was less structured in space, i.e., varied less *among* populations and regions, compared to that of its host plant. For all three elements, we found the majority of genetic variation in the rhizobium partner at the within-site level (Table 2). This contrasts with the host plant, for which less than half of the genomic diversity was found within sites (Table 1 in Grillo et al. 2016; Table 2). Both pSymA and pSymB were less structured than the host at both the regional and population levels, with more variation partitioned at the population level than the regional (Table 2).For the chromosome, however, more genetic variation was partitioned at the regional scale, compared to the host plant (Table 2), likewise only the chromosome exhibited a significant pattern of isolation by distance (Fig 2. Sup. Fig. 4), though the relationship (r^2) was substantially stronger for the host (Fig 3 in Grillo et al. 2016; Fig 2). This finding is driven by increased variation in the genetic distance value between points from the same geographic distance for the rhizobium (i.e., increased spread above and below the line in Sup. Fig. 4), again consistent with structuring among regions but high genomic diversity within local populations.

Consistent with the relatively high degree of spatial genetic structure at the regional scale in the rhizobium chromosome, and at all scales in the host genome, analyses of genetic PCAs

indicated that rhizobium and plant individuals clustered more based on region of origin for these elements than for either rhizobium symbiotic plasmid (Fig. 3). Host individuals primarily clustered along axes that separated individuals from Spain from individuals from France and Corsica, with France and Corsica representing distinct, but overlapping, groups (Grillo et al. 2016; Fig. 3A). Rhizobium chromosomal sequences, on the other hand, most clearly clustered into two groups: mainland Europe (France and Spain) versus Corsica (Fig. 3B). This pattern was also present in the phylogeny of chromosomal sequences, where individuals from Corsica largely formed a separate clade that is more distinct than clades including Corsica individuals for the symbiotic plasmids (Sup. Figs. 1-3). For both symbiotic plasmids, on the other hand, genomic variation was continuous along the first three principal component axes, with each of the three regions overlapping significantly, consistent with less clonality and higher relative rates of recombination compared to the chromosome, with some degree of geographic structure (Fig. 3 C and D). Genetic PCA of presence-absence variants showed no clear geographic structure (Sup fig 5).

The elements of the symbiont genome differ substantially in how their variation is distributed across space. The relatively high proportion of variation present within populations for all elements of the symbiont genome, suggests relatively high levels of gene-flow between populations within a given region. The high degree of variation partitioned at the regional scale in the chromosome relative to the symbiotic plasmids, suggest that barriers to dispersal at the regional scale, such as greater spatial distances, and potential barriers to the movement of individuals, particularly the Pyrenees mountain range and Mediterranean Sea in the case of our data, have a greater impact on gene-flow in the chromosome than in either of the symbiotic plasmids. Previous work in *E. meliloti* has found that rates of horizontal gene transfer are higher

in the symbiotic plasmids than in the chromosome (Nelson et al. 2018), which could help to explain this pattern. While rhizobium individuals that migrate into established populations are likely to be lost due to drift, genetic material from individuals could remain in the population due to horizontal gene transfer, leading to genetic variation being less structured, and higher observed rates of geneflow in genomic elements with higher gene transfer rates.

Furthermore, the spatial genetic structure of all elements of the symbiont genome differ substantially from that of the host. Our analyses do not support the idea of genetic variation in pSymA or pSymB exhibiting patterns strongly separating individuals by region. In the symbiont chromosome, individuals from Corsica form a tight group, separating them from mainland individuals. The patterns observed in both symbiotic plasmids, as well as the symbiotic chromosome differ from the patterns seen in the host plant, where individuals from Spain appear to cluster, largely separately from France and Corsica. This suggests that the same environmental factors can impact the population genetics of microbes in ways that are different from their impact on interacting macrobial species, as well as having differing impacts on genomic elements found in the same microorganism. While genetic variation in the symbiotic plasmids is not as structured as the host plant at the regional level, the similar level of structuring at this scale between the chromosome and host plant, as well as the pattern of isolation by distance observed in both, suggest that gene flow in both is somehow limited across the studied landscape. The difference in regional clustering, however, between the chromosome and host plant suggest that for the symbiont chromosome the Mediterranean sea is the major dispersal barrier driving the observed pattern, while in the host plant the Pyrenees mountains appear to be the driving force. This suggests that these barriers impact the different species differently, and

that assumptions around the impact of such features of a landscape on macrobial population genetics should not be assumed to be the same as the impact on microbial populations.

Increased geneflow in the one partner compared to the other would mean that a given selection pressure (abiotic or biotic) must be stronger when acting on the symbiont than the same pressure on the host, to avoid being swamped out by gene flow from other sites. Importantly, in our study system, such effects appear to depend on the genome element in *Ensifer* as well as the geographic scale at which selection varies. While variation in the bacterial chromosome is structured similarly to the plant host at the regional scale, minimal structure at the among-population (within-region) scale might mean that the host plant is more likely to locally adapt to environmental heterogeneity at smaller spatial scales. By contrast, for the symbiotic plasmids the low degree of spatial structure observed in our data suggests that a strong selective force would need to be acting a population to become locally adapted.

The potential implications of such discordant population structure for the outcomes of mutualism coevolution depends on whether the evolution of this interaction is driven by aligned or conflicting fitness interests. The fitness alignment of interacting species in mutualism is an active area of debate (Friesen *et al.* 2012, Jones *et al.* 2015, Frederickson 2017, Sachs *et al.* 2018, Gano-Cohen *et al.* 2020) and likely varies based on environmental conditions and from mutualism to mutualism. On the one hand, the fitness of partners might be mal-aligned, or conflicting, wherein increases in each partner's fitness occur at the expense of that of the other partner's fitness (Sachs *et al.* 2018); in this case, at any given site one would expect a pattern of negative frequency dependent selection wherein novelty is favored in both partners to exploit their partner or to avoid being exploited. On the other hand, the fitness of each partner might be aligned such that improvements in one partner's fitness also benefits the interacting species

(Friesen *et al.* 2012); in this case, one would expect a pattern of positive frequency dependent selection wherein novelty is disfavored, leading to stable allele frequencies under purifying selection.

If the evolution of this interaction is driven by conflicting fitness interests between partners, we might expect patterns of coevolution similar to those documented in antagonistic interactions. In antagonistic interactions, the partner with a higher migration rate will benefit from the influx of new, potentially adaptive alleles that provide a competitive edge (Gandon *et al.* 1996**,** Carlsson-Granér and Thrall 2015). This suggests that if the evolution of the legume-rhizobium mutualism is driven by misaligned fitness interests, then the rhizobium partners should be better adapted to their hosts. While there is substantially less theory on mutualism coevolution in the presence of gene flow we can, to an extent generalize the pattern of positive frequency dependent selection described above to a metapopulation setting. Because the pattern and scale of genetic structure in the host and the symbiont are different, the potential for local co-adaptation is dependent on the relationship between selection and migration at any given site. Relatively weak selective forces acting on the rhizobia in a population may be easily swamped out by gene flow, while the effects of strong selection may overpower higher rates of migration, and lead to local coadaptation between historically high quality partners at individual sites within the symbiont genome.

Pairwise F_{ST} values for populations of the host were not significantly correlated with those of the rhizobium symbiont at the 8 aligned sites for any genetic element (Fig 4.). The lack of alignment at broad spatial scales observed here is consistent with previous work in the symbiosis between coral and symbiotic dinoflagellates (Magalon et al. 2006; Baums et al. 2014), although there is evidence in that system of correlation between the host and symbiont at smaller

distances (<200km) (Magalon et al. 2006). While some work comparing the population genetics of mutualists has found broad alignment between species' population structures (Anderson et al. 2004; Thompson et al. 2005), these studies are focused on interactions between macrobes. Factors such as environmental heterogeneity or shared barriers to dispersal can drive population structure (Wang and Bradburd 2014; Aguillon et al. 2017). While there is not a large body of work, our results, along with those from the coral-dinoflagellate symbiosis, suggest that microbial population structure may be shaped by landscapes in ways that differ fundamentally than that of their macrobial hosts, warranting further study to understand these differences and their implication for evolution.

Previous work, including work on this system, has often assumed that spatially, microbial populations are roughly aligned in scale with interacting species (Heath 2010; Carlsson-Granér and Thrall 2015). Our findings suggest that the idea of a genetically distinct population at these scales should be applied to microbial species with caution, particularly in organisms with multipartite genomes, as genomic elements can have remarkably different spatial population genomics. Future work on microbial species should work to understand the factors that determine microbial population structure, and how they differ from those that determine structure in macrobial species.

Conclusions: Our results show distinct population structure between the elements of the symbiont genome, as well as discordant population structure between the symbiont and the host, with genetic variation in the symbiont tending to be less structured among geographic regions and among site overall. While some implications for the evolution of this interaction hinge on the alignment of partners' fitness, our results have several useful takeaways. The distinct

evolutionary histories and population structure of the three elements of the rhizobium genome at this scale is further evidence for their independence in *E. meliloti*, which has not been documented within the native range at this local sampling depth prior to this work. Additionally, our results imply that the factors structuring genetic variation differ between rhizobia and their host plants. Future work should strive to understand the factors that drive this difference, and whether this observation is a specific to this interaction, or one that broadly impacts interactions between microbial and macrobial species.

Figures and Tables

Table 1. Mantel tests (r^2 values) correlating the genetic distance matrices of each of the three elements of the *E. meliloti* genome. Matrices were constructed from pairwise principal component based genetic distances among 191 strains from 21 populations in the native range.

Element 1	**Element 2**	**r^2**
Chromosome	pSymA	0.12*******
Chromosome	pSymB	0.015
pSymA	pSymB	0.53*******

***p<0.05;** p<0.01; ***p<0.001; ****p<0.0001**

Table 2. AMOVA partitioning the genome-wide genetic variation in the host plant *Medicago* and for each of the three elements of the *E. meliloti* genome for 191 strains sampled from 21 populations from three regions (Spain, France, or Corsica) in the native range of the symbiosis. For each analysis, the percent variance explained at each level of spatial division is given in the lefthand column, while phi statistics and significance for each level are given in the righthand column.

	Medicago		**Chromosome**		**pSymA**		**pSymB**	
	% variance	Phi	% variance	Phi	% variance	Phi	% variance	Phi
Among Region	19.56%	0.20***	24.52%	0.34**	7.71%	0.27***	10.06%	0.24***
Among Population within Region	36.48%	0.45***	9.68%	0.13***	19.18%	0.21***	13.82%	0.15***
Within Population	43.96%	0.56***	65.80%	0.25***	73.12%	0.08***	76.12%	0.10***

***p<0.05;** p<0.01; ***p<0.001; ****p<0.0001**

Fig 1. Map of 25 total sampling locations in Spain, France, and Corsica. Red points represent sites where both host and symbiont were sampled (n = 8), green points represent sites where only symbionts were sampled (n = 13), and blue points represent sites where only hosts were sampled (n = 4).

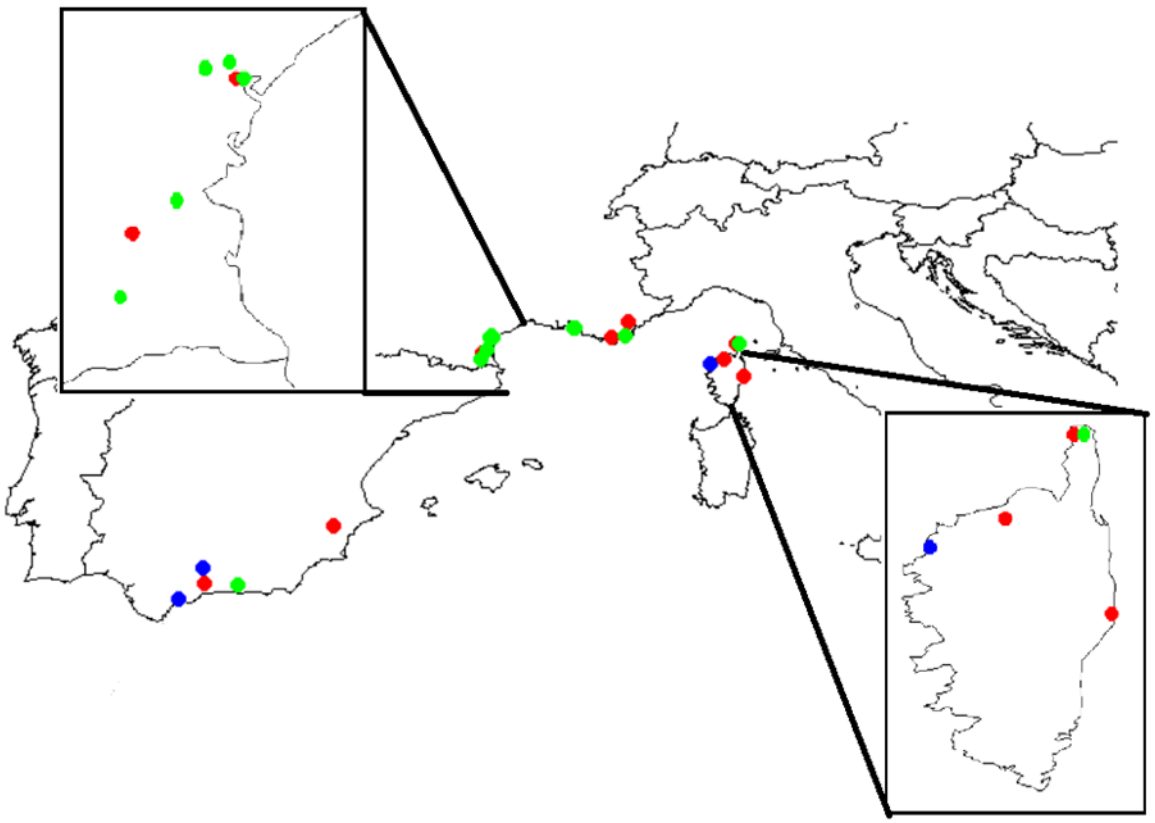


Fig 2. Correlations between geographic distance and principal component based genetic distances for *Medicago* host plants (n=192) (green), and the *E. meliloti* (n=191) chromosome (red), pSymA (blue), and pSymB (purple). Shown are trend lines with standard error and Pearson r^2 values from Mantel tests for each comparison next to the respective line ($^{***}p<0.001$)**.**

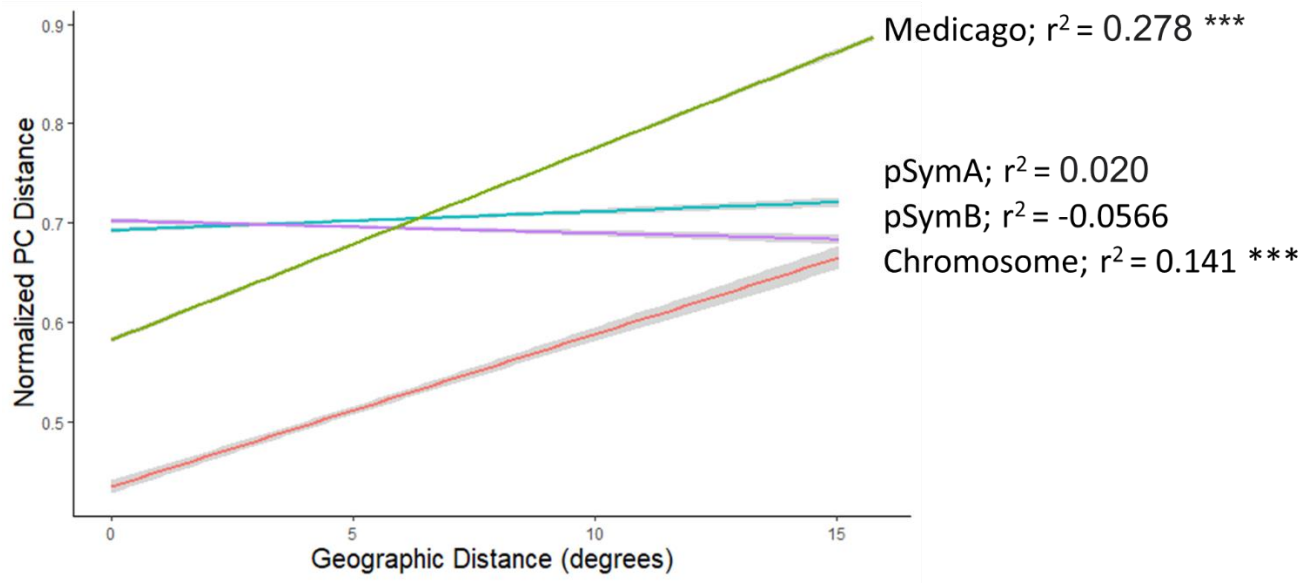


Fig 3. Principal component axis plots showing genome wide similarity on PCs 1-3 for the *Medicago* host plants (n=192) (**A**) and for symbiont (n=191) chromosome (**B**), pSymA (**C**), and pSymB (**D**). The darkest ellipse and points on each plot represent individuals from Spain, the intermediate color line represent individuals from mainland France, and the lightest color represents individuals from Corsica.

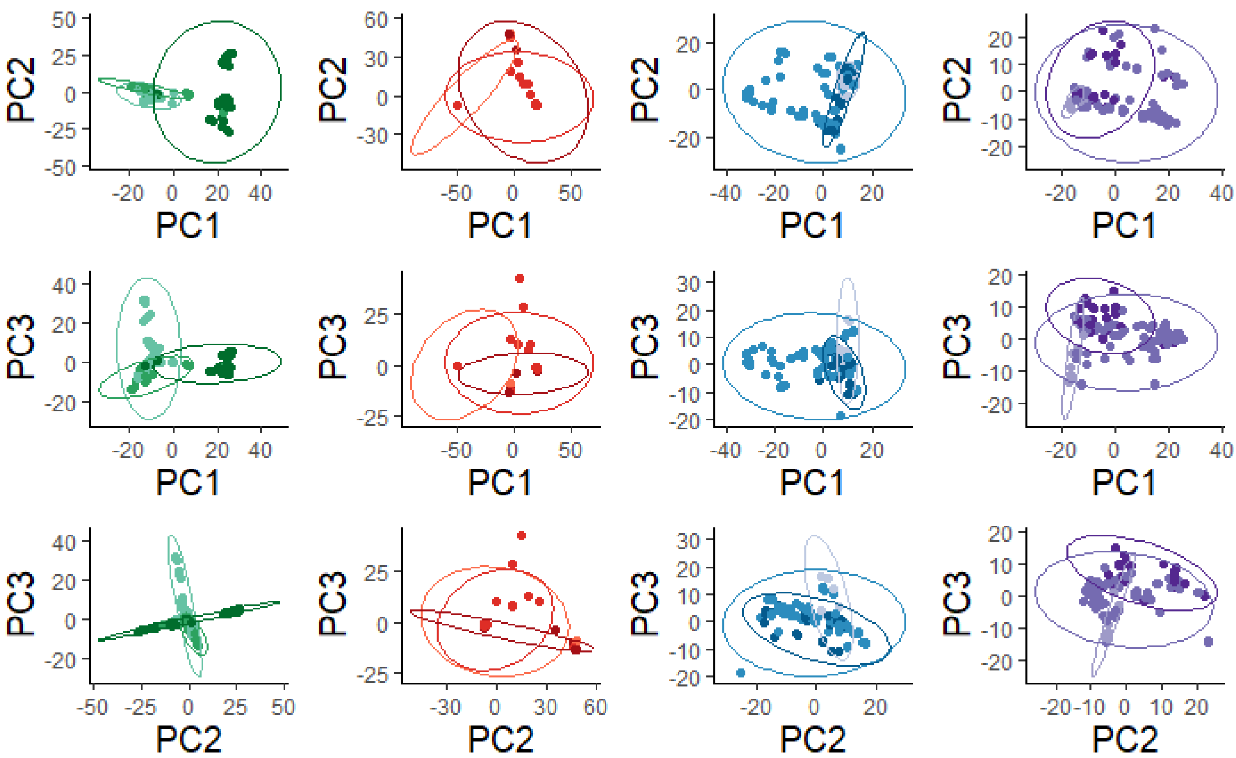
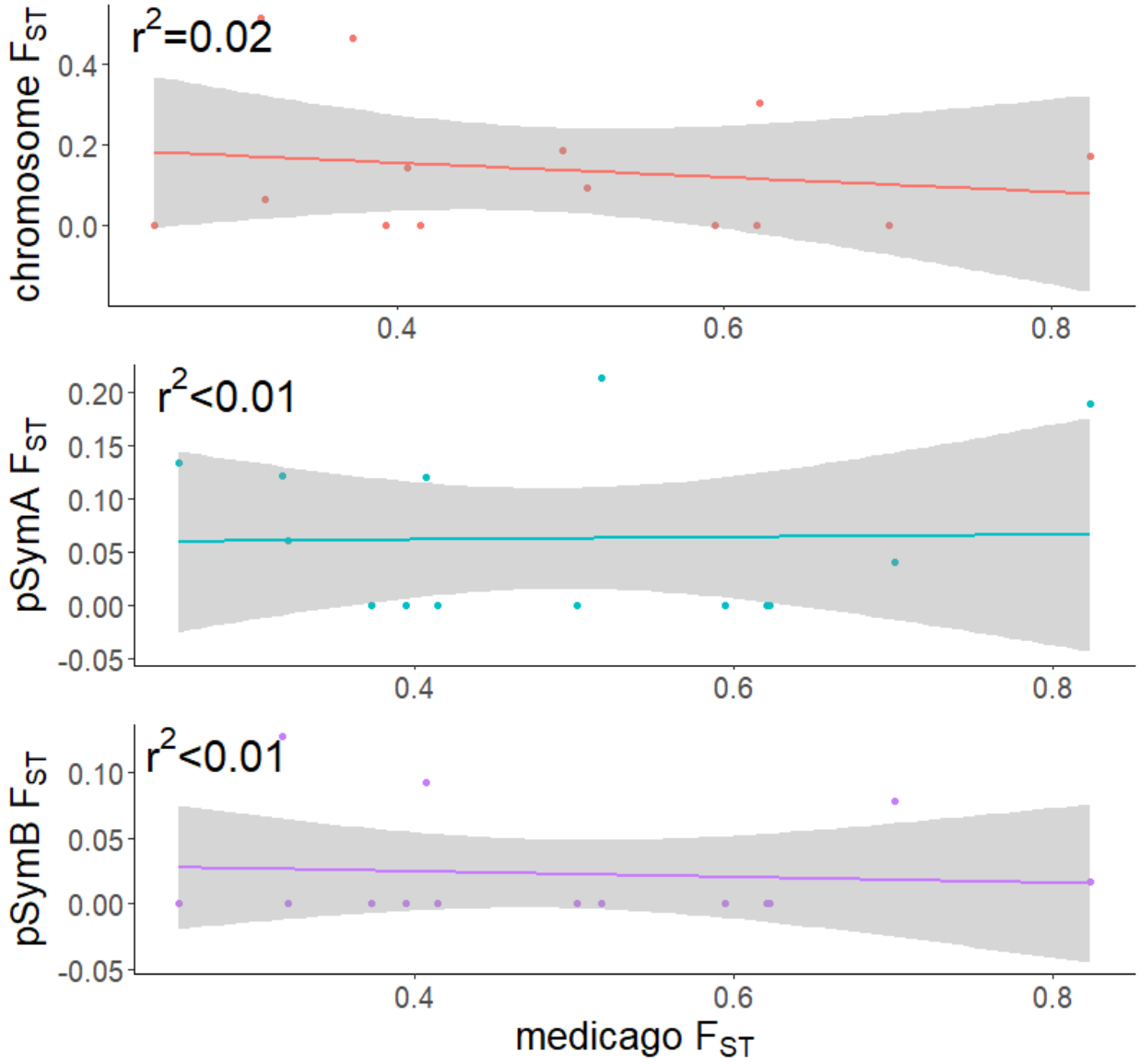


Fig 4. Tests of correlated population structure between hosts and symbionts. Shown are correlations of the between-population pairwise F_{ST} values for eight populations sampled for both *Medicago* hosts and (from top to bottom) the *E. meliloti* chromosome, pSymA, and pSymB. None of the presented correlations are significant.



References

Aguillon, S. M., J. W. Fitzpatrick, R. Bowman, S. J. Schoech, A. G. Clark, G. Coop, and N. Chen. 2017. Deconstructing isolation-by-distance: The genomic consequences of limited dispersal. PLoS Genet. 13:1–27.

Anderson, B., I. Olivieri, M. Lourmas, and B. A. Stewart. 2004. Comparative population genetic structures and local adaptation of two mutualists. Evolution (N. Y). 58:1730–1747.

Bankevich, A., S. Nurk, D. Antipov, A. A. Gurevich, M. Dvorkin, A. S. Kulikov, V. M. Lesin, S. I. Nikolenko, S. Pham, A. D. Prjibelski, A. V. Pyshkin, A. V. Sirotkin, N. Vyahhi, G. Tesler, M. A. Alekseyev, and P. A. Pevzner. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19:455–477.

Barrett, L. G., L. M. Broadhurst, and P. H. Thrall. 2012. Geographic adaptation in plant-soil mutualisms: Tests using Acacia spp. and rhizobial bacteria. Funct. Ecol. 26:457–468.

Batstone, R. T., E. M. Dutton, D. Wang, M. Yang, and M. E. Frederickson. 2017. The evolution of symbiont preference traits in the model legume Medicago truncatula. New Phytol. 213:1850–1861.

Baums, I. B., M. K. Devlin-Durante, and T. C. Lajeunesse. 2014. New insights into the dynamics between reef corals and their associated dinoflagellate endosymbionts from population genetic studies. Mol. Ecol. 23:4203–4215.

Becker, A., M. J. Barnett, D. Capela, M. Dondrup, P. B. Kamp, E. Krol, B. Linke, S. Rüberg, K. Runte, B. K. Schroeder, S. Weidner, S. N. Yurgel, J. Batut, S. R. Long, A. Pühler, and A. Goesmann. 2009. A portal for rhizobial genomes: RhizoGATE integrates a Sinorhizobium meliloti genome annotation update with postgenome data. J. Biotechnol. 140:45–50.

Biondi, E. G., E. Pilli, E. Giuntini, M. L. Roumiantseva, E. E. Andronov, O. P. Onichtchouk, O.

N. Kurchak, B. V. Simarov, N. I. Dzyubenko, A. Mengoni, and M. Bazzicalupo. 2003. Genetic relationship of Sinorhizobium meliloti and Sinorhizobium medicae strains isolated from Caucasian region. FEMS Microbiol. Lett. 220:207–213.

Bonhomme, M., S. Boitard, H. S. Clemente, B. Dumas, N. Young, and C. Jacquet. 2015. Genomic signature of selective sweeps illuminates adaptation of medicago truncatula to root-associated microorganisms. Mol. Biol. Evol. 32:2097–2110.

Bonnin, I., T. Huguet, M. Gherardi, J.-M. Prosperi, and I. Olivieri. 1996. High level of polymorphism and spatial structure in a selfing plant species, Medicago truncatula (Leguminosae), shown using RAPD markers . Am. J. Bot. 83:843–855.

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

Brown, S. P., M. A. Grillo, J. C. Podowski, and K. D. Heath. 2020. Soil origin and plant genotype structure distinct microbiome compartments in the model legume Medicago truncatula. Microbiome 8:1–17. Microbiome.

Burdon, J. J., and P. H. Thrall. 2000. Coevolution at multiple spatial scales: Linum marginale-Melampsora lini - From the individual to the species. Evol. Ecol. 14:261–281.

Caldera, E. J., and C. R. Currie. 2012. The population structure of antibiotic-producing bacterial symbionts of apterostigma dentigerum ants: Impacts of coevolution and multipartite symbiosis. Am. Nat. 180:604–617.

Carlsson-Granér, U., and P. H. Thrall. 2015. Host resistance and pathogen infectivity in host populations with varying connectivity. Evolution (N. Y). 69:926–938.

Catchen, J., P. A. Hohenlohe, S. Bassham, A. Amores, and W. A. Cresko. 2013. StackCatchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: an analysis

tool set for population genomics. Molecular Ecology, 22(11), 3124–40. http://doi.org/10.1111/mec.12354s: an analysis tool set for population genomics. Mol. Ecol. 22:3124–40.

Cooper, V. S., S. H. Vohr, S. C. Wrocklage, and P. J. Hatcher. 2010. Why genes evolve faster on secondary chromosomes in bacteria. PLoS Comput. Biol. 6.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin. 2011. The variant call format and VCFtools. Bioinformatics 27:2156–2158.

DiCenzo, G. C., and T. M. Finan. 2017. The Divided Bacterial Genome. Microbiol. Mol. Biol. Rev. 81:1–37.

diCenzo, G. C., A. M. MacLean, B. Milunovic, G. B. Golding, and T. M. Finan. 2014. Examination of Prokaryotic Multipartite Genome Evolution through Experimental Genome Reduction. PLoS Genet. 10.

Dray, S., and A. B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Dybdahl, M. F., and C. M. Lively. 1996. The geography of coevolution: Comparative population structures for a snail and its trematode parasite. Evolution (N. Y). 50:2264–2275.

Epstein, B., M. J. Sadowsky, and P. Tiffin. 2014. Selection on Horizontally Transferred and Duplicated Genes in Sinorhizobium (Ensifer), the Root-Nodule Symbionts of Medicago. Genome Biol. Evol. 6:1199–1209. Sinauer Associates, Sunderland (MA).

Fernandes, L. D., P. Lemos-Costa, P. R. Guimarães, J. N. Thompson, and M. A. M. de Aguiar. 2019. Coevolution Creates Complex Mosaics across Large Landscapes. Am. Nat. 194:000–000.

Friesen, M. L., and A. Mathias. 2010. Mixed infections may promote diversification of mutualistic symbionts: Why are there ineffective rhizobia? J. Evol. Biol. 23:323–334.

Galardini, M., F. Pini, M. Bazzicalupo, E. G. Biondi, and A. Mengoni. 2013. Replicon-dependent bacterial genome evolution: The case of Sinorhizobium meliloti. Genome Biol. Evol. 5:542–558.

Gandon, S., Y. Capowiez, Y. Dubois, Y. Michalakis, and I. Olivieri. 1996. Local adaptation and gene-for-gene coevolution in a metapopulation model. Proc. R. Soc. B Biol. Sci. 263:1003–1009.

Garrison, E., and G. Marth. 2012. Haplotype-based variant detection from short-read sequencing. 1–9.

Grillo, M. A., S. De Mita, P. V. Burke, K. L. S. Solórzano-Lowell, and K. D. Heath. 2016. Intrapopulation genomics in a model mutualist: Population structure and candidate symbiosis genes under selection in *Medicago truncatula*. Evolution (N. Y). 70:2704–2717. Wiley/Blackwell (10.1111).

Harrison, P. W., R. P. J. Lower, N. K. D. Kim, and J. P. W. Young. 2010. Introducing the bacterial 'chromid': not a chromosome, not a plasmid. Trends Microbiol. 18:141–148. Elsevier Current Trends.

Harrison, T. L., C. W. Wood, K. D. Heath, and J. R. Stinchcombe. 2017. Geographically structured genetic variation in the *Medicago lupulina* - *Ensifer* mutualism. Evolution (N. Y). 71:1787–1801. Wiley/Blackwell (10.1111).

Heath, K. D. 2010. INTERGENOMIC EPISTASIS AND COEVOLUTIONARY CONSTRAINT IN PLANTS AND RHIZOBIA. Evolution (N. Y). 64:1446–1458. Blackwell Publishing Inc.

Heath, K. D., and M. A. Grillo. 2016. Rhizobia: tractable models for bacterial evolutionary ecology. Environ. Microbiol. 18:4307–4311.

Heath, K. D., and J. R. Stinchcombe. 2014. EXPLAINING MUTUALISM VARIATION: A NEW EVOLUTIONARY PARADOX? Evolution (N. Y). 68:309–317. Wiley/Blackwell (10.1111).

Heath, K. D., and P. Tiffin. 2009. Stabilizing mechanisms in a legume-rhizobium mutualism. Evolution (N. Y). 63:652–662.

Hollowell, A. C., J. U. Regus, K. A. Gano, R. Bantay, D. Centeno, J. Pham, J. Y. Lyu, D. Moore, A. Bernardo, G. Lopez, A. Patil, S. Patel, Y. Lii, and J. L. Sachs. 2016a. Epidemic Spread of Symbiotic and Non-Symbiotic Bradyrhizobium Genotypes Across California. Microb. Ecol. 71:700–710.

Hollowell, A. C., J. U. Regus, D. Turissini, K. A. Gano-Cohen, R. Bantay, A. Bernardo, D. Moore, J. Pham, and J. L. Sachs. 2016b. Metapopulation dominance and genomicisland acquisition of Bradyrhizobium with superior catabolic capabilities. Proc. R. Soc. B Biol. Sci. 283.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071.

Kamvar, Z. N., J. F. Tabima, and N. J. Gŕunwald. 2014. Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ 2014:1–14.

Kaneko, T., Y. Nakamura, S. Sato, K. Minamisawa, T. Uchiumi, S. Sasamoto, A. Watanabe, K. Idesawa, M. Iriguchi, K. Kawashima, M. Kohara, M. Matsumoto, S. Shimpo, H. Tsuruoka, T. Wada, M. Yamada, and S. Tabata. 2002. Complete genomic sequence of nitrogen-fixing

symbiotic bacterium Bradyrhizobium japonicum USDA110. DNA Res. 9:189–197.

Klinger, C. R., J. A. Lau, and K. D. Heath. 2016. Ecological genomics of mutualism decline in nitrogen-fixing bacteria. Proc. R. Soc. B Biol. Sci. 283:20152563.

Laine, A. L. 2005. Spatial scale of local adaptation in a plant-pathogen metapopulation. J. Evol. Biol. 18:930–938.

Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Magalon, H., E. Baudry, A. Husté, M. Adjeroud, and M. Veuille. 2006. High genetic diversity of the symbiotic dinoflagellates in the coral Pocillopora meandrina from the South Pacific. Mar. Biol. 148:913–922.

Moran, N. A. 2007. Symbiosis as an adaptive process and source of phenotypic complexity. Light Evol. 1:165–181.

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

Page, A. J., C. A. Cummins, M. Hunt, V. K. Wong, S. Reuter, M. T. G. Holden, M. Fookes, D. Falush, J. A. Keane, and J. Parkhill. 2015. Roary: Rapid large-scale prokaryote pan genome analysis. Bioinformatics 31:3691–3693.

Paradis, E., and K. Schliep. 2019. Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35:526–528.

Pembleton, L. W., N. O. I. Cogan, and J. W. Forster. 2013. StAMPP: An R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol.

Ecol. Resour. 13:946–952.

Pérez Carrascal, O. M., D. VanInsberghe, S. Juárez, M. F. Polz, P. Vinuesa, and V. González. 2016. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing *Rhizobium* species associated with *Phaseolus vulgaris*. Environ. Microbiol. 18:2660–2676. Wiley/Blackwell (10.1111).

Pita, L., L. Rix, B. M. Slaby, A. Franke, and U. Hentschel. 2018. The sponge holobiont in a changing ocean: from microbes to ecosystems. Microbiome 6:46. Microbiome.

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

Revillini, D., C. A. Gehring, and N. C. Johnson. 2016. The role of locally adapted mycorrhizas and rhizobacteria in plant–soil feedback systems. Funct. Ecol. 30:1086–1098.

Ronfort, J., T. Bataillon, S. Santoni, M. Delalande, J. L. David, and J. M. Prosperi. 2006. Microsatellite diversity and broad scale geographic structure in a model legume: Building a set of nested core collection for studying naturally occurring variation in Medicago truncatula. BMC Plant Biol. 6:1–13.

Rosenberg, E., and I. Zilber-Rosenberg. 2018. The hologenome concept of evolution after 10 years. Microbiome 6:78. Microbiome.

Seemann, T. 2014. Prokka: Rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069.

Simon, J. C., J. R. Marchesi, C. Mougel, and M. A. Selosse. 2019. Host-microbiota interactions: From holobiont theory to analysis. Microbiome 7:1–5. Microbiome.

Siol, M., J. M. Prosperi, I. Bonnin, and J. Ronfort. 2008. How multilocus genotypic pattern helps

to understand the history of selfing populations: A case study in Medicago truncatula. Heredity (Edinb). 100:517–525.

Strobel, H. M., F. Alda, C. G. Sprehn, M. J. Blum, and D. C. Heins. 2016. Geographic and host-mediated population genetic structure in a cestode parasite of the three-spined stickleback. Biol. J. Linn. Soc. 119:381–396.

Tack, A. J. M., F. Horns, and A. L. Laine. 2014. The impact of spatial scale and habitat configuration on patterns of trait variation and local adaptation in a wild plant parasite. Evolution (N. Y). 68:176–189.

Tang, H., V. Krishnakumar, S. Bidwell, B. Rosen, A. Chan, S. Zhou, L. Gentzbittel, K. L. Childs, M. Yandell, H. Gundlach, K. F. X. Mayer, D. C. Schwartz, and C. D. Town. 2014. An improved genome release (version Mt4.0) for the model legume Medicago truncatula. BMC Genomics 15:1–14.

Thompson, A. R., C. E. Thacker, and E. Y. Shaw. 2005. Phylogeography of marine mutualists: Parallel patterns of genetic structure between obligate goby and shrimp partners. Mol. Ecol. 14:3557–3572.

Thompson, J. N. 2005. The Geographic Mosaic of Coevolution. 1st ed. Chicago Press.

Thrall, P. H., M. E. Hochberg, J. J. Burdon, and J. D. Bever. 2007. Coevolution of symbiotic mutualists and parasites in a community context. Trends Ecol. Evol. 22:120–126.

Vincent, J. 1970. A Manual for the Practical Study of Root-Nodule Bacteria. Blackwell Scientific, Oxford and Edinburgh.

von Wettberg, E. J. B., J. Ray-Mukherjee, K. Moriuchi, and S. S. Porter. 2020. Medicago truncatula as an ecological, evolutionary, and forage legume model: new directions forward. Model Legum. Medicago truncatula 31–40.

Wang, I. J., and G. S. Bradburd. 2014. Isolation by environment. Mol. Ecol. 23:5649–5662.

Yates, R., J. Howieson, S. E. De Meyer, R. Tian, R. Seshadri, A. Pati, T. Woyke, V. Markowitz, N. Ivanova, N. Kyrpides, A. Loi, B. Nutt, G. Garau, L. Sulas, and W. Reeve. 2015. High-quality permanent draft genome sequence of Rhizobium sullae strain WSM1592; a Hedysarum coronarium microsymbiont from Sassari, Italy. Stand. Genomic Sci. 10:1–6. Standards in Genomic Sciences.

Young, J. P. W., L. C. Crossman, A. W. B. Johnston, N. R. Thomson, Z. F. Ghazoui, K. H. Hull, M. Wexler, A. R. J. Curson, J. D. Todd, P. S. Poole, T. H. Mauchline, A. K. East, M. A. Quail, C. Churcher, C. Arrowsmith, I. Cherevach, T. Chillingworth, K. Clarke, A. Cronin, P. Davis, A. Fraser, Z. Hance, H. Hauser, K. Jagels, S. Moule, K. Mungall, H. Norbertczak, E. Rabbinowitsch, M. Sanders, M. Simmonds, S. Whitehead, and J. Parkhill. 2006. The genome of Rhizobium leguminosarum has recognizable core and accessory components. Genome Biol. 7.

Young, J. P. W., and K. E. Haukka. 1996. Diversity and phylogeny of rhizobia. New Phytol. 133:87–94.

Zribi, K., R. Mhamdi, T. Huguet, and M. E. Aouani. 2004. Distribution and genetic diversity of rhizobia nodulating natural populations of Medicago truncatula in tunisian soils. Soil Biol. Biochem. 36:903–908.

Supplementary Figures and Tables

Supplementary Table 1. Metadata on strains of *Ensifer meliloti*. For each strain the region, sampling site (soil), and specific soil samples are given, along with the genotype of the *Medicago truncatula* plant used to isolate the strain, and the latitude and longitude at which the soil sample was taken.

STRAIN	REGION	SOIL	SOIL_SAMPLE	PLANT GENOTYPE	LAT	LONG
MAG722A	CORSICA	C1	C1-3	F83026	42.16748	9.542717
MAG722B	CORSICA	C1	C1-3	F83026	42.16748	9.542717
MAG530	CORSICA	C2	C2-5	ESP161	42.9687	9.41295
MAG420	CORSICA	C3	C3-8	F20089	42.97463	9.365783
MAG421	CORSICA	C3	C3-8	F20089	42.97463	9.365783
MAG422	CORSICA	C3	C3-8	F20089	42.97463	9.365783
MAG498	CORSICA	C3	C3-8	ESP161	42.97463	9.365783
MAG508	CORSICA	C3	C3-11	F20089	42.97263	9.366483
MAG701A	CORSICA	C3	C3-8	F83026	42.97463	9.365783
MAG701B	CORSICA	C3	C3-8	F83026	42.97463	9.365783
MAG714A	CORSICA	C3	C3-11	F20093	42.97263	9.366483
MAG714B	CORSICA	C3	C3-11	F20093	42.97263	9.366483
MAG513	CORSICA	C4	C4-12	ESP161	42.59243	9.04045
MAG514	CORSICA	C4	C4-12	ESP161	42.59243	9.04045
MAG533	CORSICA	C4	C4-12	F20093	42.59243	9.04045
MAG710A	CORSICA	C4	C4-12	F20093	42.59243	9.04045
MAG710B	CORSICA	C4	C4-12	F20093	42.59243	9.04045

Supplementary Table 1. (cont.)

MAG724A	CORSICA	C4	C4-4	F20043	42.59172	9.04035
MAG724B	CORSICA	C4	C4-4	F20043	42.59172	9.04035
MAG726A	CORSICA	C4	C4-7	F20043	42.59098	9.040233
MAG726B	CORSICA	C4	C4-7	F20043	42.59098	9.040233
MAG523	FRANCE	1	1-8	ESP009	43.1474	3.000417
MAG524	FRANCE	1	1-8	ESP009	43.1474	3.000417
MAG525	FRANCE	1	1-8	ESP009	43.1474	3.000417
MAG526	FRANCE	1	1-8	ESP009	43.1474	3.000417
MAG141	FRANCE	2	2-1	F11005	43.14543	2.999367
MAG183	FRANCE	2	2-1	ESP174	43.14543	2.999367
MAG184	FRANCE	2	2-1	ESP174	43.14543	2.999367
MAG199	FRANCE	2	2-1	F11005	43.14543	2.999367
MAG220	FRANCE	2	2-9	ESP174	43.1457	2.9989
MAG32	FRANCE	2	2-1	F11005 n	43.14543	2.999367
MAG539	FRANCE	2	2-7	F11005	43.14542	2.998683
MAG540	FRANCE	2	2-7	F11005	43.14542	2.998683
MAG541	FRANCE	2	2-7	F11005	43.14542	2.998683
MAG80	FRANCE	2	2-10	ESP174	43.14572	2.998917
MAG82	FRANCE	2	2-1	ESP174	43.14543	2.999367
MAG92	FRANCE	2	2-9	F11005	43.1457	2.9989
MAG93	FRANCE	2	2-9	F11005	43.1457	2.9989

Supplementary Table 1. (cont.)

MAG123	FRANCE	3	3-3	F11005	43.12227	3.077067
MAG177	FRANCE	3	3-4	F11005	43.12235	3.075667
MAG194-mel	FRANCE	3	3-2	ESP174	43.12233	3.076717
MAG225	FRANCE	3	3-2	ESP174	43.12233	3.076717
MAG245	FRANCE	3	3-1	ESP174	43.12222	3.07635
MAG246	FRANCE	3	3-1	ESP174	43.12222	3.07635
MAG247	FRANCE	3	3-1	ESP174	43.12222	3.07635
MAG248	FRANCE	3	3-1	ESP174	43.12222	3.07635
MAG34	FRANCE	3	3-2	ESP174	43.12233	3.076717
MAG40	FRANCE	3	3-3	ESP174 no	43.12227	3.077067
MAG69	FRANCE	3	3-2	F11005	43.12233	3.076717
MAG90A	FRANCE	3	3-7	F11005	43.12232	3.075283
MAG90B	FRANCE	3	3-7	F11005	43.12232	3.075283
MAG14	FRANCE	4	4-2	ESP174	43.12198	3.095533
MAG16	FRANCE	4	4-3	F20089	43.12227	3.096033
MAG204	FRANCE	4	4-1	ESP174	43.122	3.095283
MAG205	FRANCE	4	4-1	ESP174	43.122	3.095283
MAG206	FRANCE	4	4-1	ESP174	43.122	3.095283
MAG24	FRANCE	4	4-3	F20089	43.12227	3.096033

Supplementary Table 1. (cont.)

MAG27	FRANCE	4	4-2	ESP174 n	43.12198	3.095533
MAG39	FRANCE	4	4-1	F20089	43.122	3.095283
MAG43	FRANCE	4	4-1	F20089	43.122	3.095283
MAG48	FRANCE	4	4-3	ESP174	43.12227	3.096033
MAG5	FRANCE	4	4-3	F20089	43.12227	3.096033
MAG57	FRANCE	4	4-3	ESP174	43.12227	3.096033
MAG59	FRANCE	4	4-1	ESP174	43.122	3.095283
MAG62	FRANCE	4	4-1	F20089	43.122	3.095283
MAG65	FRANCE	4	4-3	ESP174	43.12227	3.096033
MAG67	FRANCE	4	4-2	F20089	43.12198	3.095533
MAG6A	FRANCE	4	4-3	F20089	43.12227	3.096033
MAG6B	FRANCE	4	4-3	F20089	43.12227	3.096033
MAG73	FRANCE	4	4-3	ESP174	43.12227	3.096033
MAG7A	FRANCE	4	4-2	ESP174	43.12198	3.095533
MAG7B	FRANCE	4	4-2	ESP174	43.12198	3.095533
MAG17	FRANCE	5	5-3	F20089	42.82302	2.9269
MAG209	FRANCE	5	5-1	F66009	42.82278	2.927717
MAG36	FRANCE	5	5-3	F66009	42.82302	2.9269
MAG46	FRANCE	5	5-3	F66009	42.82302	2.9269
MAG71	FRANCE	5	5-2	F20089	42.82302	2.927317
MAG201	FRANCE	6	6-2	F66009	42.74198	2.81485

Supplementary Table 1. (cont.)

MAG219	FRANCE	6	6-2	F20089	42.74198	2.81485
MAG38	FRANCE	6	6-2	F20089n	42.74198	2.81485
MAG41	FRANCE	6	6-1	F20089	42.74205	2.814917
MAG42	FRANCE	6	6-1	F20089	42.74205	2.814917
MAG53	FRANCE	6	6-2	F20089	42.74198	2.81485
MAG72	FRANCE	6	6-2	F66009	42.74198	2.81485
MAG8	FRANCE	6	6-1	F66009	42.74205	2.814917
MAG85	FRANCE	6	6-1	F20089	42.74205	2.814917
MAG87	FRANCE	6	6-1	F20089	42.74205	2.814917
MAG88	FRANCE	6	6-1	F20089	42.74205	2.814917
MAG15	FRANCE	7	7-1	F66009n	42.58827	2.78495
MAG18	FRANCE	7	7-1	F66009n	42.58827	2.78495
MAG189	FRANCE	7	7-3	F20089	42.58895	2.785383
MAG191	FRANCE	7	7-2	F66009	42.5883	2.785
MAG215	FRANCE	7	7-3	F66009	42.58895	2.785383
MAG216	FRANCE	7	7-3	F66009	42.58895	2.785383
MAG26	FRANCE	7	7-2	F20089	42.5883	2.785
MAG54	FRANCE	7	7-2	F20089	42.5883	2.785
MAG78	FRANCE	7	7-4	F20089	42.58948	2.784833

Supplementary Table 1. (cont.)

MAG182	FRANCE	14	14-2	F11005	43.1612	3.05905
MAG200	FRANCE	14	14-A	ESP174	43.1612	3.059983
MAG84	FRANCE	14	14-A	F11005	43.1612	3.059983
MAG86	FRANCE	14	14-3	F11005	43.16108	3.059383
MAG89	FRANCE	14	14-1	ESP174	43.16122	3.059017
MAG117B	FRANCE	15	15-1	F83026	43.35138	5.169617
MAG145	FRANCE	15	15-1	F83026	43.35138	5.169617
MAG101	FRANCE	16	16-A	ESP174	43.37568	5.188133
MAG114	FRANCE	16	16-B	F83026	43.37568	5.1871
MAG142	FRANCE	16	16-3	ESP174	43.37693	5.188133
MAG144A	FRANCE	16	16-A	F83026	43.37568	5.188133
MAG144B	FRANCE	16	16-A	F83026	43.37568	5.188133
MAG158A	FRANCE	16	16-3	ESP174	43.37693	5.188133
MAG158B	FRANCE	16	16-3	ESP174	43.37693	5.188133
MAG169	FRANCE	16	16-3	ESP174	43.37693	5.188133
MAG97	FRANCE	16	16-B	ESP174	43.37568	5.1871
MAG303	FRANCE	17	17-5	ESP161	43.12692	6.1252
MAG372	FRANCE	17	17-12	F83026	43.12538	6.12535
MAG373	FRANCE	17	17-12	F83026	43.12538	6.12535
MAG733A	FRANCE	17	17-2	F20043	43.12693	6.125217
MAG762A	FRANCE	17	17-12	ESP174	43.12538	6.12535
MAG276	FRANCE	18	18-2	ESP161	43.16532	6.4774

Supplementary Table 1. (cont.)

MAG277	FRANCE	18	18-2	ESP161	43.16532	6.4774
MAG278	FRANCE	18	18-2	ESP161	43.16532	6.4774
MAG281	FRANCE	18	18-3	ESP161	43.1653	6.4774
MAG282	FRANCE	18	18-3	ESP161	43.1653	6.4774
MAG283	FRANCE	18	18-3	ESP161	43.1653	6.4774
MAG313	FRANCE	18	18-1	F83026	43.16528	6.477433
MAG317	FRANCE	18	18-5	F83026	43.1649	6.477233
MAG318	FRANCE	18	18-5	F83026	43.1649	6.477233
MAG319	FRANCE	18	18-5	F83026	43.1649	6.477233
MAG320	FRANCE	18	18-5	F83026	43.1649	6.477233
MAG335	FRANCE	18	18-3	F83026	43.1653	6.4774
MAG336	FRANCE	18	18-3	F83026	43.1653	6.4774
MAG382	FRANCE	18	18-2	F83026	43.16532	6.4774
MAG406	FRANCE	18	18-5	ESP161	43.1649	6.477233
MAG286	FRANCE	19	19-2	ESP161	43.16932	6.474433
MAG342	FRANCE	19	19-3	ESP161	43.16935	6.47445
MAG384	FRANCE	19	19-2	F83026	43.16932	6.474433
MAG386	FRANCE	19	19-4	F83026	43.16932	6.474483
MAG409	FRANCE	19	19-1	F83026	43.16935	6.474417
MAG258	FRANCE	20	20-2	F83026	43.53285	6.567617
MAG259	FRANCE	20	20-2	F83026	43.53285	6.567617
MAG261	FRANCE	20	20-2	F83026	43.53285	6.567617

Supplementary Table 1. (cont.)

MAG322	FRANCE	20	20-3	ESP161	43.53283	6.567467
MAG325	FRANCE	20	20-3	ESP161	43.53283	6.567467
MAG358	FRANCE	20	20-1	F83026	43.53403	6.572617
MAG417	FRANCE	20	20-3	F83026	43.53283	6.567467
MAG739A	FRANCE	20	20-2	F20093	43.53285	6.567617
MAG740A	FRANCE	20	20-1	F20093	43.53403	6.572617
MAG749A	FRANCE	20	20-2	F83026	43.53285	6.567617
MAG749B	FRANCE	20	20-2	F83026	43.53285	6.567617
MAG749C	FRANCE	20	20-2	F83026	43.53285	6.567617
MAG753A	FRANCE	20	20-3	F20043	43.53283	6.567467
MAG753B	FRANCE	20	20-3	F20043	43.53283	6.567467
MAG761A	FRANCE	20	20-4	F20043	43.53287	6.566783
MAG10	SPAIN	8	8-1	ESP174	38.41355	-1.01468
MAG21	SPAIN	8	8-1	ESP174	38.41355	-1.01468
MAG221	SPAIN	8	8-4	ESP174	38.4116	-1.01507
MAG230	SPAIN	8	8-3	ESP174	38.41155	-1.0165
MAG231	SPAIN	8	8-3	ESP174	38.41155	-1.0165
MAG238	SPAIN	8	8-2	ESP174	38.41253	-1.01568
MAG242	SPAIN	8	8-2	ESP174	38.41253	-1.01568
MAG243A	SPAIN	8	8-2	ESP174	38.41253	-1.01568
MAG243B	SPAIN	8	8-2	ESP174	38.41253	-1.01568
MAG33	SPAIN	8	8-4	ESP174	38.4116	-1.01507

Supplementary Table 1. (cont.)

MAG35	SPAIN	8	8-4	ESP174	38.4116	-1.01507
MAG66	SPAIN	8	8-3	F20050	38.41155	-1.0165
MAG79	SPAIN	8	8-3	F20050	38.41155	-1.0165
MAG83	SPAIN	8	8-3	ESP174	38.41155	-1.0165
MAG9	SPAIN	8	8-1	ESP174 nod	38.41355	-1.01468
MAG192C	SPAIN	9	8-3	F20050	36.91188	-3.47665
MAG192D	SPAIN	9	9-3	F20050	36.91188	-3.47665
MAG25A	SPAIN	9	9-3	ESP174	36.91188	-3.47665
MAG25B	SPAIN	9	9-3	ESP174	36.91188	-3.47665
MAG31	SPAIN	9	9-1	ESP174 no	36.91083	-3.47712
MAG746A	SPAIN	9	9-3	F20093	36.91188	-3.47665
MAG746B	SPAIN	9	9-3	F20093	36.91188	-3.47665
MAG121	SPAIN	11	11-3	ESP174	36.93672	-4.3563
MAG146	SPAIN	11	11-3	F83026	36.93672	-4.3563
MAG148A	SPAIN	11	11-3	F83026	36.93672	-4.3563
MAG148B	SPAIN	11	11-3	F83026	36.93672	-4.3563
MAG154	SPAIN	11	11-1	F83026	36.93678	-4.35608
MAG157	SPAIN	11	11-1	F83026	36.93678	-4.35608
MAG174	SPAIN	11	11-1	ESP174	36.93678	-4.35608
MAG176A	SPAIN	11	11-1	ESP174	36.93678	-4.35608

Supplementary Table 1. (cont.)

MAG176B	SPAIN	11	11-1	ESP174	36.93678	-4.35608
MAG748A	SPAIN	11	11-1	ESP174	36.93678	-4.35608
MAG748B	SPAIN	11	11-1	ESP174	36.93678	-4.35608
MAG755A	SPAIN	11	11-1	F83026	36.93678	-4.35608
MAG755B	SPAIN	11	11-1	F83026	36.93678	-4.35608
MAG757A	SPAIN	11	11-3	F20093	36.93672	-4.3563
MAG757C	SPAIN	11	11-3	F20093	36.93672	-4.3563
MAG758A-mel	SPAIN	11	11-2	F83026	36.93678	-4.35608

Supplementary Table 2. AMOVA results testing for the percent of genome wide genetic variation in 191 strains of E. meliloti that can be attributable to plant genotype used to isolate (" trap") the isolate from native soils. Values were computed separately for each genome element.

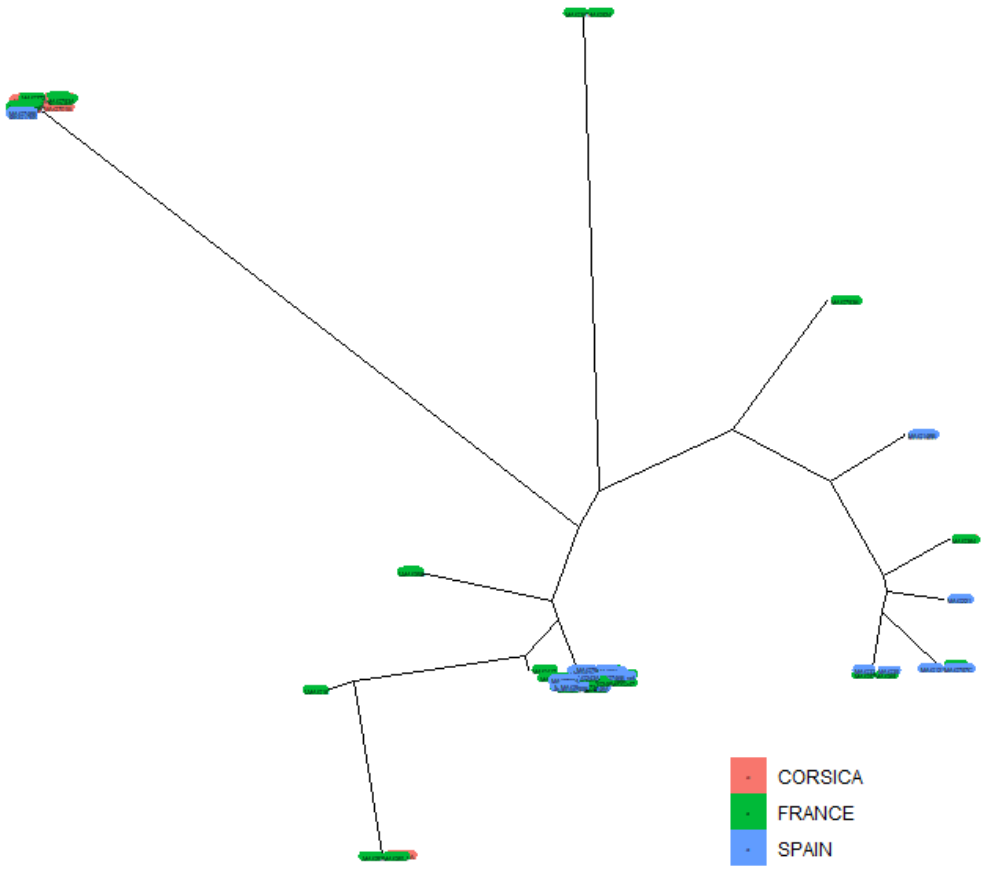
	Chromosome	pSymA	pSymB
Variation between trapping lines	14.7%***	13.1%***	12.2%***
Variation within trapping lines	85.3%***	86.9%***	87.8%***

***p<0.05;** p<0.01; ***p<0.001; ****p<0.0001**

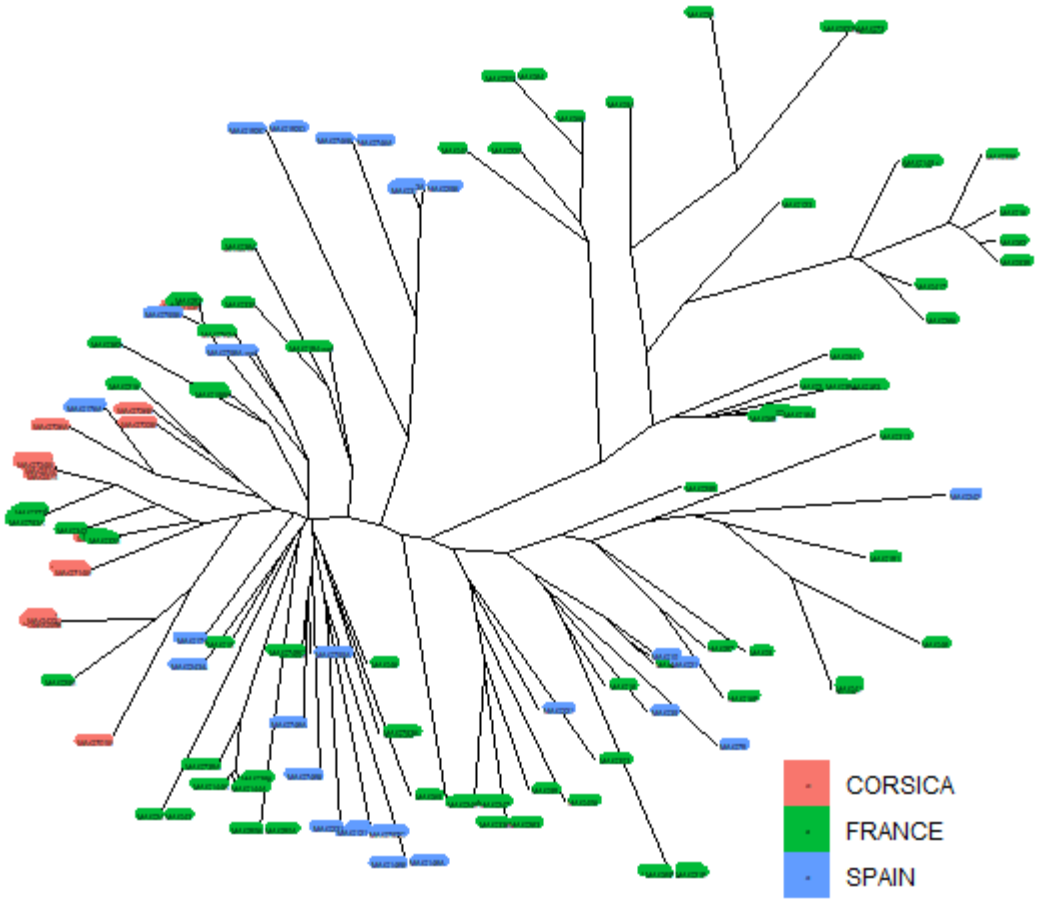
Supplementary Table 3. SNP counts for each element of the symbiont genome before any variant filtering, and after filtering for quality and depth using VCFtools.

	Chromosome	pSymA	pSymB
Pre-filtering	414,004	179,505	222,935
Quality filters	34,689	15,162	22,460

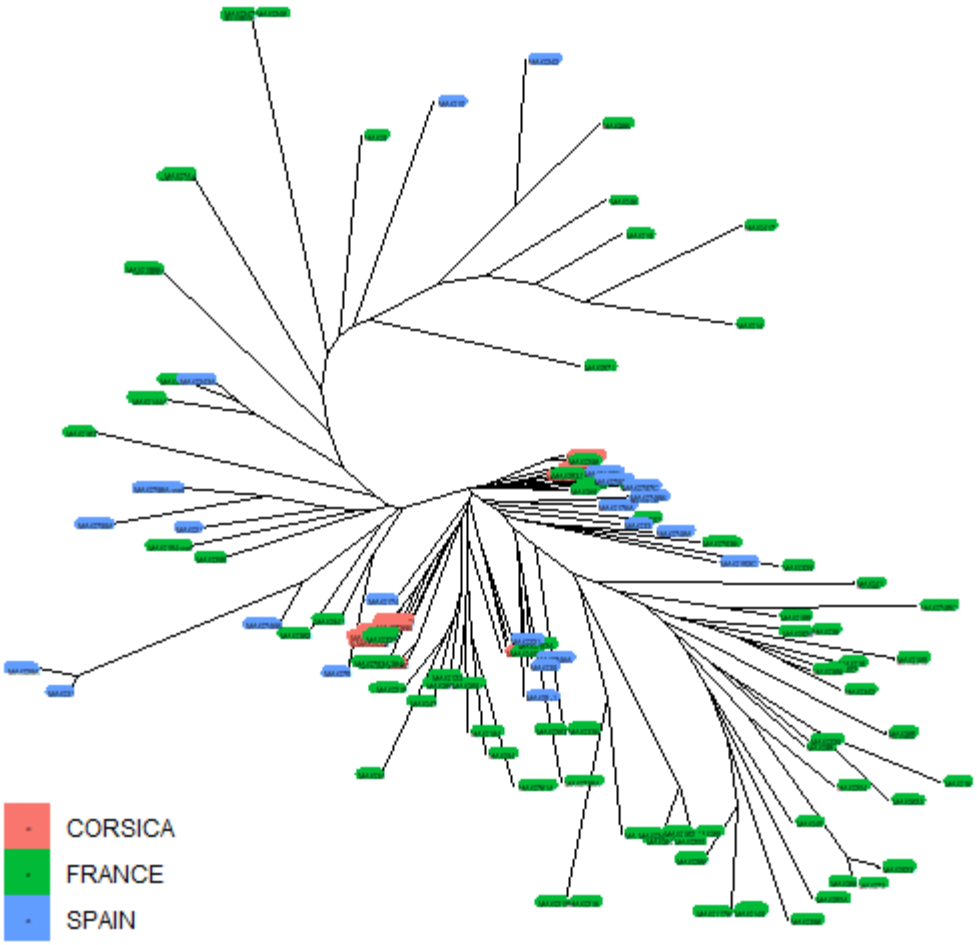
Supplementary Fig 1. Neighbor joining tree of rhizobium individuals based on chromosomal variant data. Individual tip labels are colored based on region of origin, with pink representing individuals from Corsica, green representing mainland France, and blue representing Spain.



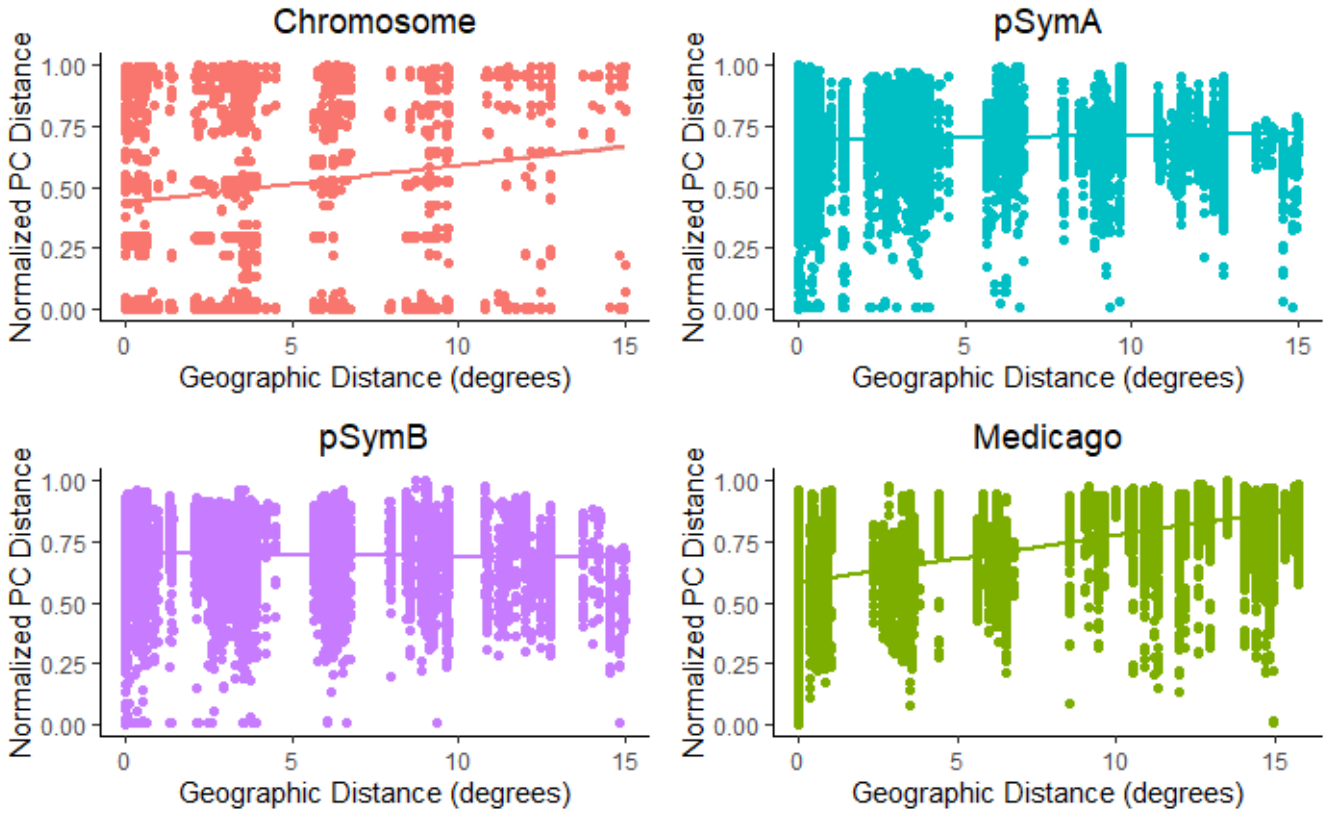
Supplementary Fig 2. Neighbor joining tree of rhizobium individuals based on pSymA variant data. Individual tip labels are colored based on region of origin, with pink representing individuals from Corsica, green representing mainland France, and blue representing Spain.



Supplementary Fig 3. Neighbor joining tree of rhizobium individuals based on pSymB variant data. Individual tip labels are colored based on region of origin, with pink representing individuals from Corsica, green representing mainland France, and blue representing Spain.



Supplementary Fig 4. Correlations between geographic distance and normalized principal component distance for the symbiont (n=191) chromosome (**A**), pSymA (**B**), pSymB (**C**), and the host (n=192) (**D**) Shown are trend lines and points representing individual comparisons.



Supplementary Fig 5. Principal component axis plot for all individuals of the symbiont (n=191) based on variable gene content. From top to bottom principal component 1 and 2, 1 and 3, and 2 and 3. The red ellipse and points on each plot represent individuals from Spain, the green dotted line and points represent individuals from mainland France, and the blue points and line represent individuals from Corsica.