Elevated Rates of Molecular Evolution Genome-wide in Mutualist Legumes and Rhizobia

Tia L. Harrison [1,2,*] John R. Stinchcombe [1,†] Megan E. Frederickson [1,†]

[1]Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3B2, Canada
[2]Department of Biology, Queen's University, Kingston, Ontario K7L 3N6, Canada

***Corresponding author:** E-mail: tia.harrison@queensu.ca.
[†]Authors contributed equally.

Associate editor: Rebekah Rogers

Abstract

Rates of molecular evolution vary greatly among even closely related species. Although theory predicts that antagonistic interactions between species increase rates of molecular evolution, predictions for how mutualism affects evolutionary rates are mixed. We compared rates of molecular evolution between (i) mutualistic and non-mutualistic legumes, (ii) an independent set of symbiotic rhizobia and their non-symbiotic close relatives, and (iii) symbiotic and non-symbiotic clades within *Ensifer*, a diverse genus of bacteria with various lifestyles. We assembled transcriptomes de novo for 12 legume species and calculated dN/dS ratios at orthologous genes in all species to determine if genes in mutualistic plants evolve faster or slower than in their non-mutualistic relatives. We also calculated dN/dS ratios in genes known to be important for symbiosis. We found that mutualists have higher rates of molecular evolution genome-wide compared to non-mutualistic legumes, but this pattern did not hold in symbiosis genes. We next calculated dN/dS ratios in 14 bacteria species across the proteobacteria phylogeny that differ in whether they associate mutualistically with plants, using published data. In most pairs, symbiotic rhizobia show higher dN/dS values compared to their non-symbiotic relatives. Within a bacterial genus with many well-characterized mutualist species (*Ensifer*), we calculated dN/dS ratios in symbiotic and non-symbiotic clades and found that symbiotic lineages have higher rates of molecular evolution genome-wide, but not at genes on the symbiotic plasmid pSymB. Our results suggest that although mutualism between legumes and rhizobia is associated with elevated rates of molecular evolution genome-wide, symbiosis genes may be evolutionarily stagnant.

Key words: **legume, rhizobia, mutualism, symbiosis, molecular evolution.**

Introduction

A fundamental question in evolutionary biology is how species interactions contribute to variation in rates of molecular evolution (Woolfit and Bromham 2003; Bromham 2009). According to the Red Queen hypothesis, species interacting antagonistically will have higher rates of molecular evolution because they are under constant pressure to evolve new defenses against their enemies, or new counter-adaptations that overcome their victims' defenses (Stahl et al. 1999; Brockhurst et al. 2014; Delaye et al. 2018). It is less clear how mutualism might impact molecular evolution. On the one hand, mutualists might have higher rates of molecular evolution than non-mutualistic species because they must adapt to both a changing environment and a changing partner (Lutzoni and Pagel 1997; Rubin and Moreau 2016). On the other hand, some theory suggests that the more slowly evolving partner in a mutualism reaps the greatest rewards (the so-called Red King hypothesis, Bergstrom and Lachmann 2003). Here, we compare rates of molecular evolution between mutualistic and non-mutualistic legumes and bacteria to determine whether the legume–rhizobium symbiosis involves rapid or slow DNA sequence evolution.

Although the literature on coevolution tends to emphasize positive selection and diversification, slower rates of evolution may be expected if mutualist partners reach evolutionary stasis (Hembry et al. 2014; Barker et al. 2017), when selection maintains interacting species near trait-matched fitness optima with little further phenotypic change (Nuismer et al. 2013). In this scenario, stabilizing selection on both hosts and rhizobia would result in fewer nucleotide substitutions in the genome, particularly at symbiosis genes (Epstein et al. 2022). In addition, if most genetic variation in mutualist quality is due to mutation-selection balance, a signature of purifying selection would be expected at symbiosis genes to maintain compatible traits in interacting partners (Heath and Stinchcombe

Received: May 21, 2024. **Revised:** November 01, 2024. **Accepted:** November 15, 2024

2014). Therefore, patterns of widespread purifying selection can be a signature of coevolution and reciprocal adaptation, even if this form of selection is not likely to lead to evolutionary diversification and elaboration of mutualistic traits. Although population genetic methods can identify neutral and selective pressures on traits (Tiffin and Ross-Ibarra 2014; O'Brien et al. 2021), it is challenging to distinguish between stabilizing and purifying selection with population genetic data (Charlesworth 2013).

The facultative legume–rhizobium mutualism, in which leguminous plants exchange carbon for fixed nitrogen provided by rhizobial partners, is an excellent system to test how mutualism influences rates of molecular evolution. Rhizobia occupy specialized root structures on legumes called nodules (van Rhijn and Vanderleyden 1995), and this symbiosis is generally mutualistic (Friesen 2012) especially in low-nitrogen environments. The legume family (Fabaceae/Leguminosae) is large, including around 19,500 species (Azani et al. 2017), but not all these species form nodules with rhizobia. Although some work suggests that nodulation in plants has evolved multiple times after a single predisposition event (Doyle 2011; Werner et al. 2014), recent phylogenomic analyses support the hypothesis that nodulation has a single evolutionary origin, followed by multiple losses of the trait across the clade (Griesmann et al. 2018; Parshuram et al. 2023).

There is also high variation in nodulation capabilities among bacteria. Rhizobia are horizontally transmitted symbionts that are taken up from the soil by new legume hosts each generation, meaning rhizobial lineages alternate between being plant-associated and free-living in soil. Bacterial strains that have nodulation genes (*nod* genes) produce Nod factors that are important for initiating nodule formation on plant roots. However, not all rhizobia with *nod* genes can form nodules on all legume species; legumes have Nod factor receptors that must recognize compatible Nod factors for a successful symbiosis to occur (Wang et al. 2018). The development of nodules and nitrogen fixation are complex processes, requiring many genes (*nif*, *fix*, etc.) that are often found together on a mobile genetic element such as a plasmid (diCenzo et al. 2016; Batstone 2022). Rhizobia can exchange these symbiosis genes or plasmids, and thus symbiotic ability, through horizontal gene transfer (Epstein and Tiffin 2021; Rahimlou et al. 2021).

Many genetic changes that accompany transitions to mutualism have been identified in endosymbiotic bacteria. For example, extremely tiny genomes are a common feature of obligately endosymbiotic bacteria that are vertically transferred to new hosts (McCutcheon and Moran 2012). These bacteria rely on their host for many functions and thus many genes are lost in their own genome (Wernegreen 2002). Endosymbiotic bacteria also experience bottlenecks each time they are passed down to a new host (Woolfit and Bromham 2003). The reduction in population size leads to a decrease in genetic variation in the new population and a greater chance that variants are fixed or lost due to this random sampling, leading to higher rates of nucleotide substitution. Although rhizobia are also endosymbiotic within plant cells, they have a free-living stage, are horizontally transmitted, and may gain nodulation abilities through horizontal gene transfer, making it less clear how mutualism will impact genome and molecular evolution. Associating with diverse plant species and spending some time in the soil without a host might weaken host selection on the rhizobia genome (Sachs et al. 2011). Many of the genomic signatures of coevolution might be observed only in symbiosis genes, if these genes are commonly horizontally transferred into the genomes of non-symbiotic lineages (Epstein et al. 2022).

In plants, we might expect that many adaptive mutations would be necessary for the evolution of nodulation, resulting in signals of positive selection in mutualist lineages. If nodulation evolved only once near the base of the legume tree (Griesmann et al. 2018), strong positive selection may have occurred in response to mutualists in the past but may no longer be detectable with population genetic methods. Previous work has shown that the evolution of polyploidy in legumes likely predated symbiosis and may have facilitated the evolution of nodulation (Parshuram et al. 2023), suggesting that nodulation is not easily gained in multiple lineages. Nonetheless, around 9% of legumes do not form nodules (Simonsen et al. 2017), with a phylogenetic distribution that suggests multiple losses of this trait. When nodulation is lost, we might expect relaxed selection on genes that were formerly important for symbiosis with rhizobia and thus higher rates of molecular evolution at symbiosis genes in non-mutualistic lineages. In addition, mutualism is expected to increase population sizes by allowing organisms to thrive despite enemies, abiotic stress, or nutrient limitation (Afkhami et al. 2014; Weber and Agrawal 2014). Consequently, purifying selection and positive selection (and therefore adaptation) may be more effective in mutualists than non-mutualists because of their larger population sizes.

We took advantage of the presence and absence of nodulation across legumes and rhizobia to test whether mutualistic species evolve more quickly or more slowly than their non-mutualistic relatives. We assessed molecular evolution in (i) six closely related pairs of mutualistic and non-mutualistic plants (i.e. those that do and do not form nodules with rhizobia) across the legume phylogeny, (ii) seven pairs of symbiotic and non-symbiotic bacteria species (strains that have *nod* genes and those that lack *nod* genes), and (iii) a widespread genus (*Ensifer*) that includes clades of legume symbionts and non-symbiotic bacteria with other lifestyles. We generated de novo transcriptomes of 12 non-model legume species to calculate ratios of nonsynonymous to synonymous substitutions (dN/dS) at orthologous genes. We also calculated dN/dS values from 14 bacteria species with sequence data deposited in NCBI and from a total of 104 strains in the *Ensifer* phylogeny. We compared dN/dS ratios between mutualistic and non-mutualistic species genome-wide and at symbiotic genes involved in nodulation.

Results

Identifying and Sequencing Closely Related Mutualist and Non-mutualist Pairs

We identified six species pairs of mutualistic legumes and non-mutualistic close relatives using available legume phylogenies (including Zanne et al. 2014 and Azani et al. 2017) and nodulation data (Werner et al. 2014). We categorized three of the species pairs as a loss of nodulation because, within these pairs, the non-mutualistic species occurred in a clade where at least 90% of the tips were mutualist species (Fig. 1). It is unclear whether nodulation has been gained or lost in the other three pairs of species in our analysis. Mutualist species in these pairs are found within a phylogenetic group where 56% of the tips represent non-mutualistic legumes, although this group also includes many tips where the mutualist status is unknown. These pairs could represent nodulation reversals (i.e. a loss followed by a regain), but without more nodulation data, it remains unclear; the six species pairs are presented in Table 1. We also identified other traits in the focal species that might influence molecular evolution, including geographic distribution, ploidy, and life history (annual or perennial) (Simonsen et al. 2017; Parshuram et al. 2023), and found that species within pairs generally shared those traits (Table 1).

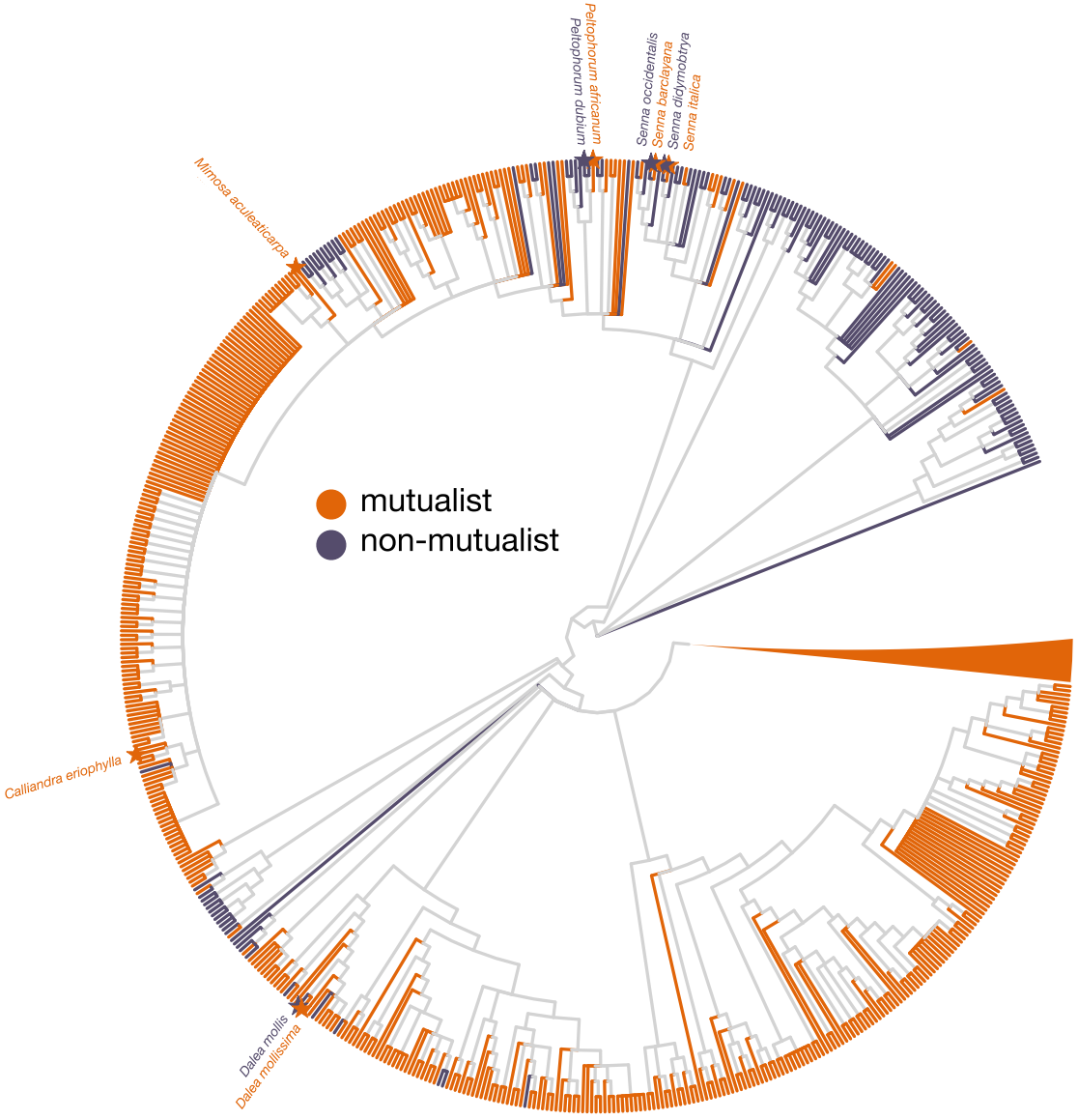

Fig. 1. Species pairs of mutualistic (orange) and non-mutualistic legumes (purple) in the study. Sampled legume species are indicated by stars at the tips of the tree and labeled with text. The legume tree developed by the Legume Phylogeny Working Group (LPWG et al. 2013) was filtered for species with symbiotic status data. The branches leading to the tips of the tree were colored based on whether the species at the tip was known to form nodules (Werner et al. 2014) and therefore is a mutualist (orange) or lacked the ability to nodulate and is therefore a non-mutualist (purple). All internal branches are colored in gray. The *Calliandra* and *Mimosa* non-symbiotic species were obtained from separate smaller phylogenies, and thus not shown here (Simon et al. 2011; de Souza et al. 2013).

All plants in our dataset likely form indeterminate nodules based on their placement in the legume phylogeny and previous records of indeterminate nodules in the Caesalpinioideae, Mimosoideae, and Papilionoideae legume subfamilies (Andrews and Andrews 2017). Because there are no available genomes for our focal legume species, we sequenced RNA for the 12 legume species in our dataset using Illumina short-read sequencing. We produced de novo assemblies of transcriptomes for all sequenced individuals.

We identified 14 bacteria genomes for analysis by searching rhizobia phylogenies for strains with *nod* genes and close relatives lacking *nod* genes (Rahimlou et al. 2021). We identified seven pairs of nodulating and non-nodulating bacteria species that spanned across both Alphaproteobacteria and Betaproteobacteria. We note that nodulating bacterial species, and their non-nodulating closest relatives, are not the rhizobia partners of the plant species used above. We then downloaded the annotated genomes from NCBI for use in our analysis (Fig. 2a; Table 2). We also downloaded genomes for 65 symbiotic members of the *Ensifer* genus (Fig. 2b; supplementary table S1, Supplementary Material online) containing *nod* genes and 39 non-symbiotic members without *nod* genes (Fagorzi et al. 2020) for separate analysis comparing symbiotic and non-symbiotic clades.

Table 1 Species of mutualistic and non-mutualistic legumes used in the study

Life style	Species	Introduced ranges	Human uses	Ploidy	Life history	Native region
Non-mutualistic	*Senna didymobotrya*	20	5	Polyploid	Perennial	Africa
Mutualist	*Senna italica*	2	1	Polyploid	Perennial	Africa, Middle East
Non-mutualistic	*Peltophorum dubium*	6	3	Polyploid	Perennial	South America
Mutualist	*Peltophorum africanum*	7	3	Polyploid	Perennial	Africa
Non-mutualistic	*Senna occidentalis*	48	6	Polyploid	Perennial	South America
Mutualist	*Senna barclayana*	0	0	NA	Annual/perennial	Australia
Non-mutualistic	*Dalea mollis*	0	1	Diploid	Annual	North America
Mutualist	*Dalea mollissima*	0	1	Diploid	Annual	North America
Non-mutualistic	*Mimosa grahamii*	0	0	NA	Perennial	North America
Mutualist	*Mimosa aculeaticarpa*	0	0	Polyploid	Perennial	North/Central America
Non-mutualistic	*Calliandra humilis*	0	0	NA	Perennial	North/Central America
Mutualist	*Calliandra eriophylla*	0	0	Diploid	Perennial	North/Central America

Senna italica, *P. africanum*, *P. dubium*, and *D. mollis* seeds were sourced from KEW Royal Botanical Gardens Millennial Seed Bank. Seeds of all other species were obtained from USDA-ARS Germplasm Resources Information Network.

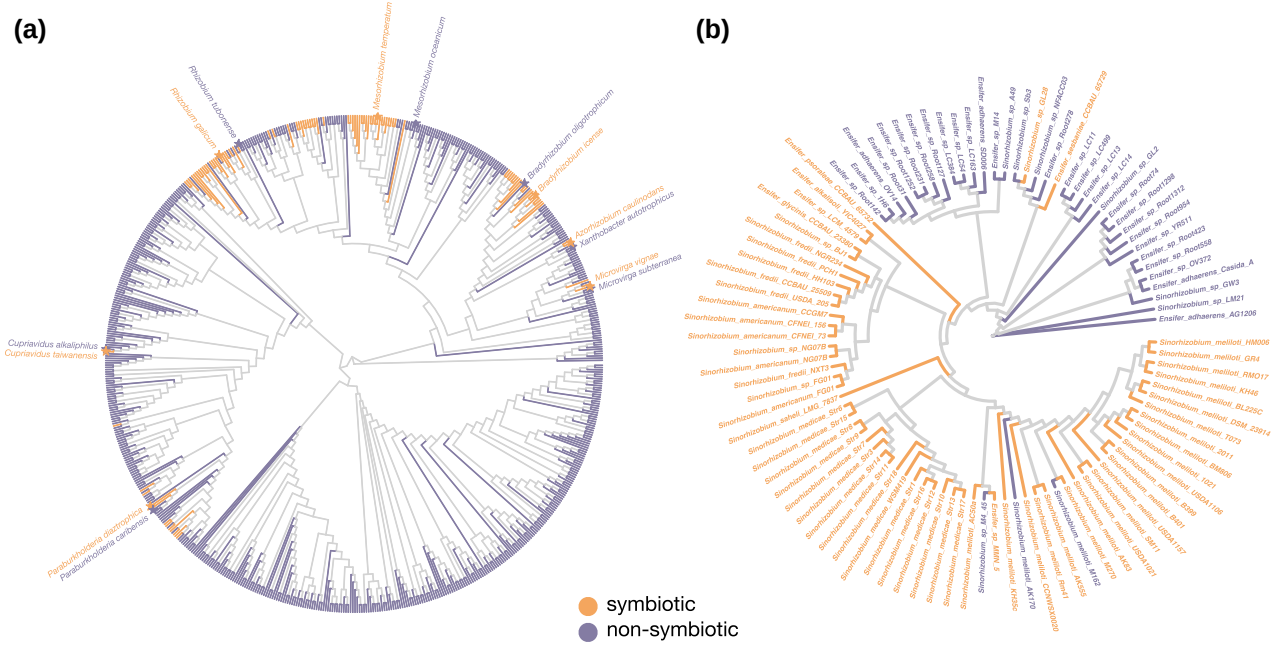

Fig. 2. Bacteria species with (orange) or without (purple) *nod* genes. a) Phylogeny from Rahimlou et al. (2021). The branches leading to the tips of the tree were colored based on whether the species at the tip is known to have *nod* genes (Fagorzi et al 2020, Rahimlou et al. 2021) and therefore is symbiotic (orange) or lacked *nod* genes and is non-symbiotic (purple). All internal branches are colored in gray. Species indicated by stars and labeled with text represent the species pairs with genomes used in the analysis. b) Phylogeny from Fagorzi et al. (2020) trimmed to only the genomes used in the analysis.

Table 2 Bacteria strains with genome data collected from NCBI used in the study

Mutualist	Species	Assembly code	Isolation source
Non-symbiotic	*Bradyrhizobium oligotrophicum*	GCA_000344805.1	Paddy field soil
Symbiotic	*Bradyrhizobium icense*	GCA_001693385.1	Root nodule
Non-symbiotic	*Xanthobacter autotrophicus*	GCA_000017645.1	Black sludge
Symbiotic	*Azorhizobium caulinodans*	GCA_000010525.1	Stem nodule
Non-symbiotic	*Microvirga subterranea*	GCA_003350535.1	Geothermal aquifer
Symbiotic	*Microvirga vignae*	GCA_001017175.1	Root nodule
Non-symbiotic	*Mesorhizobium oceanicum*	GCA_001889605.1	Sea water
Symbiotic	*Mesorhizobium temperatum*	GCA_002284575.1	Root nodule
Non-symbiotic	*Rhizobium tubonense*	GCA_003240585.1	Plant
Symbiotic	*Rhizobium gallicum*	GCA_001908615.1	Root nodule
Non-symbiotic	*Cupriavidus alkaliphilus*	GCA_003254285.1	Alkaline soils
Symbiotic	*Cupriavidus taiwanensis*	GCA_900250065.1	Root nodule
Non-symbiotic	*Paraburkholderia caribensis*	GCA_001449005.1	Soil
Symbiotic	*Paraburkholderia diazotrophica*	GCA_900108945.1	Root nodule

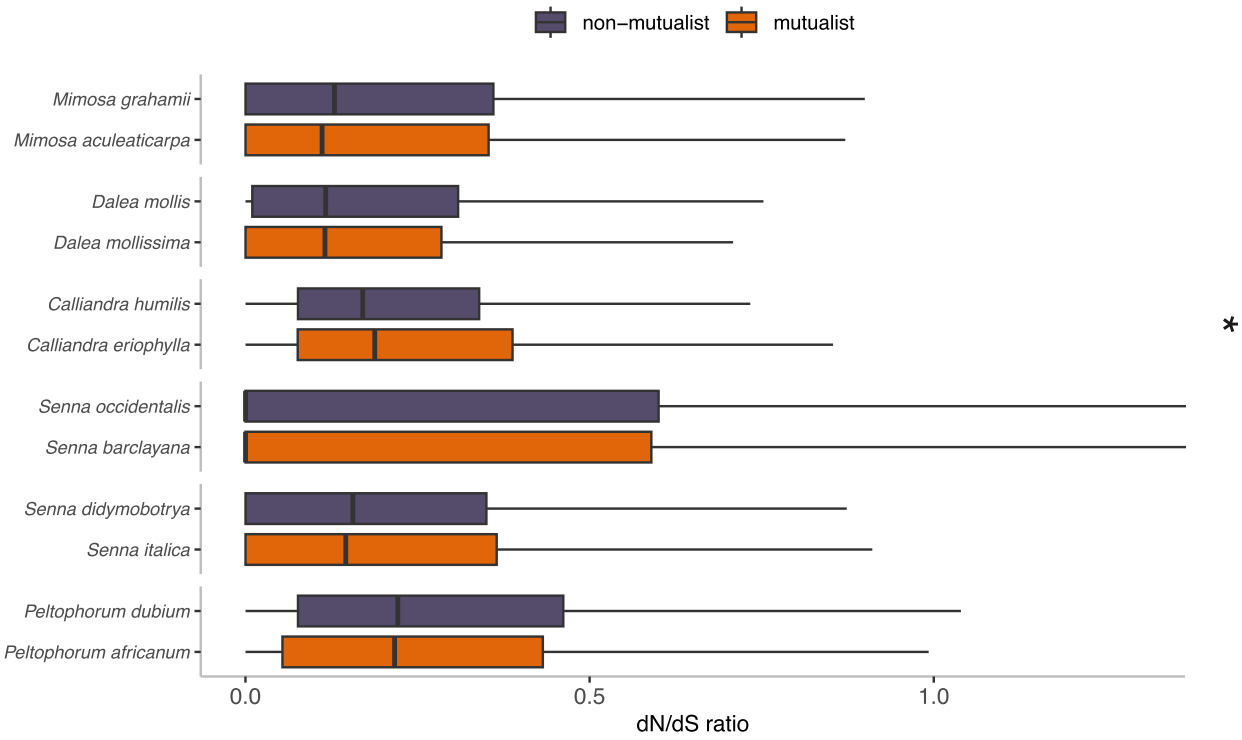

Fig. 3. Genome-wide dN/dS ratios estimated from free-ratio models in PAML for mutualistic (orange; bottom boxplot in species pair) and non-mutualistic (purple; top boxplot in species pair) legumes. A * indicates species pairs that showed significance at $P < 0.05$ in paired Wilcoxon tests. Outliers (1.5× inter quartile range) have been removed from the plot for improved visualization but were included in the paired Wilcoxon signed-rank tests.

Molecular Evolution in Legumes

To investigate whether the evolution of mutualism in legumes is associated with higher rates of molecular evolution, we first identified 761 to 1339 matching orthologous genes in the paired legume species. We aligned orthologs, created gene trees for each ortholog, and estimated the ratio of nonsynonymous to synonymous substitutions between species (dN/dS). We used a "free-ratio" model in the phylogenetic analysis by maximum likelihood program (PAML) to allow for an independent dN/dS estimation for each branch in every gene tree. We compared dN/dS ratios across all orthologs between each mutualist and non-mutualist species pair. Most pairs showed very similar rates of molecular evolution when we considered genome-wide dN/dS values (Fig. 3). Only one of these comparisons (*Calliandra humilis* and *Calliandra eriophylla*) showed a significant increase in rates of molecular evolution in the mutualist in paired Wilcoxon signed-rank statistical tests (Table 3). The other species comparisons showed non-significant differences between mutualists and their non-mutualist relatives.

We also subset our dataset to the three legume pairs that represent a loss of mutualism. The species in these pairs are also non-invasive, found in the same habitat (desert habitat in southern United States of America), and do not play a large role in human agriculture, medicine, or

Table 3 Results of paired Wilcoxon signed-rank tests comparing dN/dS ratios at matching genes in mutualistic legumes and non-mutualistic relatives

Mutualist	Non-mutualist	Gene no.	*V*	*U*	*P* value
M. aculeaticarpa	*M. grahamii*	1077	156,945	164,256	0.5769
D. mollissima	*D. mollis*	761	109,734	95,386	0.1253
C. humilis	*C. eriophylla*	1191	274,018	329,333	**0.0085**
S. occidentalis	*S. barclayana*	1339	95,093	84,607	0.2160
S. italica	*S. didymobotrya*	1003	185,751	174,225	0.4193
P. africanum	*P. dubium*	964	198,736	183,639	0.3120

The *V* value is the total sum of ranked genes where the non-mutualistic species had positive values. The *U* value is the total sum of ranked genes where the mutualist had positive values. The *P* value is reported for paired Wilcoxon tests where the null hypothesis was that the shift in rank is 0. Significant tests at $P < 0.05$ are bolded.

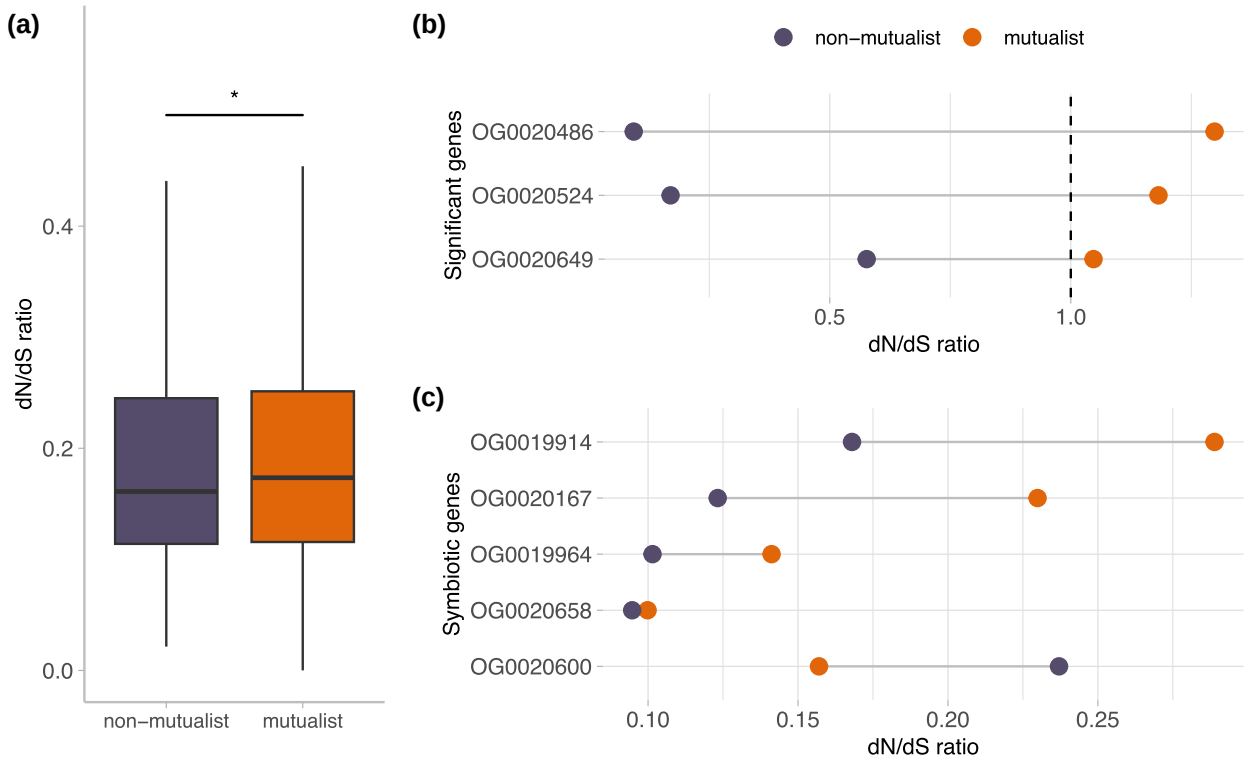

Fig. 4. Results from two-ratio models performed in PAML on legume species representing a loss of mutualism in the phylogeny. a) Genome-wide average dN/dS ratios for mutualistic (orange, right boxplot) and non-mutualistic (purple; left boxplot) legumes. A * indicates species pairs that showed significance at $P < 0.05$ in paired Wilcoxon tests. In this test, a total of 277 genes with significant differences in dN/dS ratios between mutualists and non-symbiotic species (estimated from two-rate PAML models) were included in the analyses. Outliers (>1.5× inter quartile range) have been removed from the plot for improved visualization but were included in the paired Wilcoxon signed-rank tests. b) Differences in dN/dS ratios for genes under positive selection. c) Differences in dN/dS ratios for symbiotic genes involved in symbiosis with rhizobia.

industry (genera *Dalea*, *Mimosa*, and *Calliandra*). For these plant genomes, we implemented a two-ratio model in PAML, where we allowed all mutualists to evolve at one dN/dS and all non-mutualist species to evolve at a separate dN/dS. With this method, we identified 277 genes out of 308 that showed significant differences in dN/dS ratios between mutualist and non-mutualist species. In paired Wilcoxon signed-rank tests, we found that mutualist species exhibit increased rates of molecular evolution overall (Fig. 4a).

Molecular Evolution in Rhizobia Genomes

We identified 836 to 1288 orthologous genes in the symbiotic and non-symbiotic bacteria pairs and estimated dN/dS values for each of these genes. We used these dN/dS values to compare rates of molecular evolution between symbiotic and non-symbiotic bacteria across the genome. When there were significant differences in evolutionary rates between bacteria species, symbiotic species always had higher dN/dS ratios than non-symbiotic species. Four pairs showed significantly higher rates of molecular evolution in the symbiotic species (Fig. 5). An additional pair, *C. alkaliphilus* and *C. taiwanesis*, also showed higher dN/dS ratios in the symbiotic rhizobia species (P value = 0.0603, Table 4).

To evaluate rates of molecular evolution in the *Ensifer* genus, we performed a two-ratio model in PAML where we allowed the nodulating rhizobia (largely the *Sinorhizobium* clade within the *Ensifer* phylogeny) to evolve at one dN/dS rate and the non-nodulating bacteria to evolve at a separate rate. We identified 405 orthologous genes that showed significant differences between

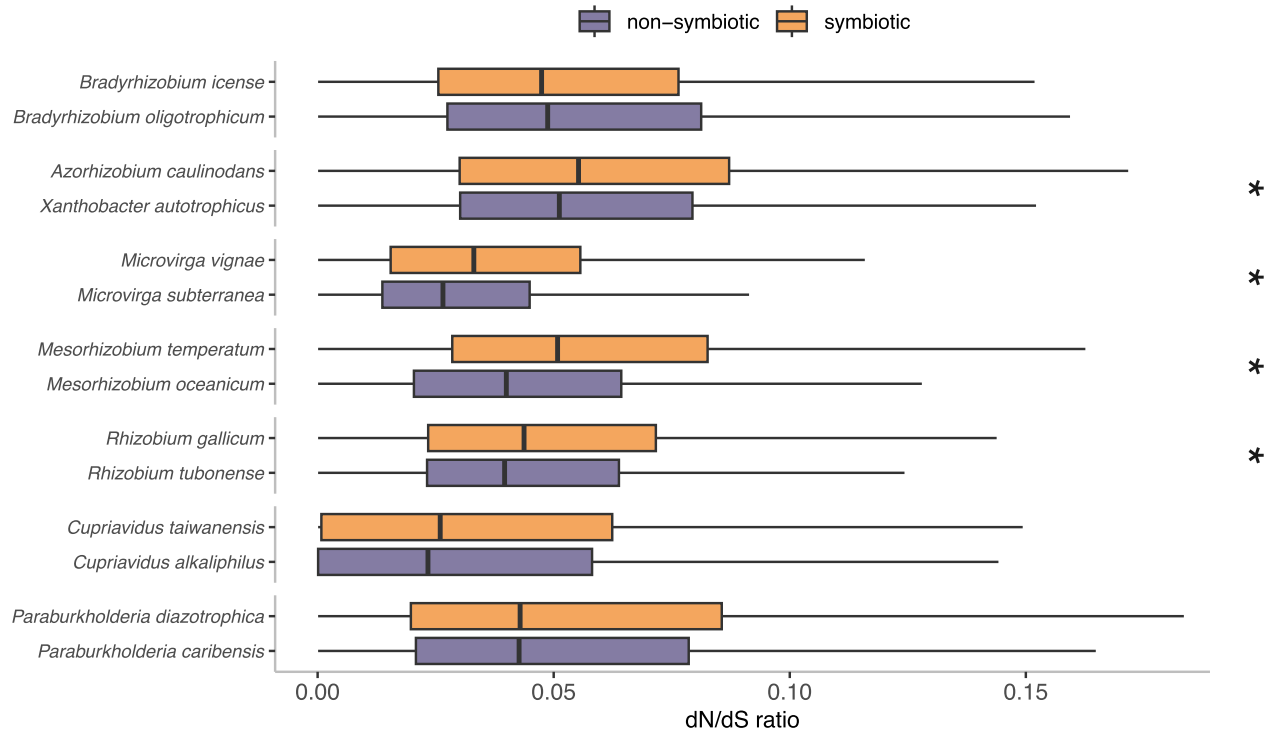

Fig. 5. Genome-wide dN/dS ratios estimated from free-ratio models in PAML for symbiotic (orange; top boxplot in species pair) and non-symbiotic (purple; bottom boxplot in species pair) bacteria strains. A * indicates species pairs that showed significance at $P < 0.05$ in paired Wilcoxon tests. Outliers (>1.5× inter quartile range) were removed from the plot for improved visualization but were included in the paired Wilcoxon signed-rank tests.

Table 4 Results of paired Wilcoxon signed-rank tests comparing dN/dS ratios at matching genes in symbiotic rhizobia and non-symbiotic relatives

Symbiotic	Non-symbiotic	Gene no.	*V*	*U*	*P* value
B. icense	***B. oligotrophicum***	**1154**	**335,686**	**326,140**	**0.6718**
A. caulinodans	***X. autotrophicus***	**988**	**200,434**	**285,171**	**<0.0001**
M. vignae	***M. subterranea***	**1052**	**216,521**	**331,060**	**<0.0001**
M. temperatum	***M. oceanicum***	**1090**	**204,387**	**390,208**	**<0.0001**
R. gallicum	***R. tubonense***	**1288**	**373,359**	**451,611**	**0.0032**
C. taiwanensis	***C. alkaliphilus***	**836**	**116,267**	**136,849**	**0.0603**
P. diazotrophica	***P. caribensis***	**1074**	**262,824**	**281,622**	**0.3341**

The *V* value is the total sum of ranked genes where the non-symbiotic species had positive values. The *U* value is the total sum of ranked genes where the symbiotic species had positive values. The *P* value is reported for paired Wilcoxon tests where the null hypothesis was that the shift in rank is 0. Significant tests at $P < 0.05$ are bolded.

nodulating and non-nodulating *Ensifer* strains. Paired Wilcoxon signed-rank tests showed that genome-wide, the symbiotic strains had higher dN/dS ratios (Fig. 6a, $P = 0.0003$). In the *Ensifer* genus, many genes that are important for symbiosis with plants are located within plasmids and symbiotic islands (Geddes et al. 2020). The plasmid pSymB is common among all 104 species in our analysis (Fagorzi et al. 2020), while the presence of pSymA is more variable across strains. Although many *nif* and *nod* genes are located on pSymA (Barnett et al. 2001), there are also symbiosis genes involved in nitrate/nitrite reduction on pSymB (Finan et al. 2001). When we analyzed genes on pSymB separately from the rest of the genome, there was no significant difference between nodulating and non-nodulating strains (Fig. 6b, $P = 0.2126$). Wilcoxon paired tests performed on the chromosome (plus other various accessory plasmids) showed higher dN/dS ratios in the nodulating strains (Fig. 6b, $P = 0.0007$).

Symbiosis Genes

We also compared dN/dS ratios between mutualist and non-mutualist species at genes expected to be involved in the legume–rhizobium symbiosis. As noted above, there are no annotated genomes available for the 12 legume species in our analysis. We identified 17 unique symbiosis genes in our legume ortholog dataset by mapping our sequences to a list of previously identified symbiosis gene sequences in *Medicago* (Epstein et al. 2022). There was no consistent pattern as to whether these symbiosis genes had higher dN/dS values in the mutualist or non-mutualist plant (supplementary fig. S1 and table S2,

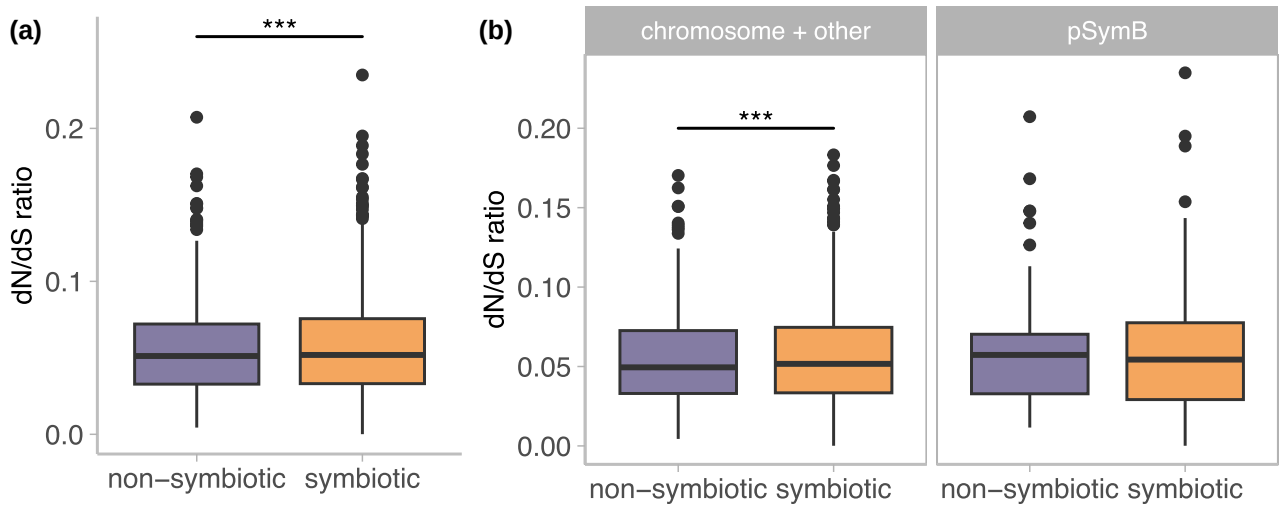

Fig. 6. dN/dS ratios estimated from two-rate models in PAML for symbiotic (orange; right boxplot) and non-symbiotic (purple; left boxplot) strains from the *Ensifer* genus. A *** indicates comparisons that showed significance at $P < 0.0005$ in paired Wilcoxon tests. a) dN/dS ratios estimated in all genes across the whole genome. b) dN/dS ratios estimated on genes separated by their location in the genome. The right-hand panel represents dN/dS ratios calculated from genes found on the pSymB plasmid and the left-hand panel shows dN/dS ratios calculated on genes from the rest of the genome including the chromosome and other plasmids (excluding pSymB).

Supplementary Material online). Differences in dN/dS ratios at symbiosis genes between mutualist and non-mutualistic species were also generally low across the species pairs we tested. One exception was the *Mimosa* pair, in which dN/dS ratios in the non-mutualist *M. grahamii* were much larger than in the mutualist relative *M. aculeaticarpa* (supplementary fig. S1, Supplementary Material online).

To determine if symbiosis genes generally show different patterns of molecular evolution compared to the rest of the genome, we compared dN/dS values for all symbiosis genes to the genome-wide distribution of dN/dS values within each legume species (supplementary fig. S2, Supplementary Material online). Overall, dN/dS values at symbiosis genes were similar to the genome-wide data (supplementary table S3, Supplementary Material online). We also compared dN/dS values across symbiosis genes that varied in degree of conservation. Rates of molecular evolution at symbiosis genes that were found in all legume species in our dataset (conserved) were not significantly different from other symbiosis genes (genes found in at least one species but not in all 12) in unpaired Wilcoxon–Mann–Whitney tests ($P = 0.2173$, supplementary fig. S3, Supplementary Material online).

We identified only five unique symbiosis genes in our filtered dataset for legumes that represent a loss of mutualism when we performed a two-ratio model on these genes in PAML. All but one of these genes had a higher dN/dS value in the mutualist species (Fig. 4c). Only one gene showed an increased rate of molecular evolution in the non-mutualist species. The top hits from the blastn search predicted this gene encodes for a leucine-rich repeat receptor-like serine/threonine-protein kinase.

In the rhizobia genomes, we identified 33 distinct symbiosis genes by searching annotated mutualist genomes for key words that indicate a role in symbiosis (e.g. *nod*, *nfe*, and *nif*). For most pairs, there was a nearly equal number of symbiotic and non-symbiotic species with higher dN/dS values at these genes (supplementary fig. S4, and table S4, Supplementary Material online). One exception was the non-symbiotic *X. autotrophicus*, which had many more genes with higher rates of molecular evolution (14 genes) compared to its symbiotic relative *A. caulinodans* (5 genes). Average dN/dS for symbiosis genes were also evaluated relative to genome-wide dN/dS distributions within each species. Patterns of molecular evolution at symbiosis genes were similar to genome-wide genes, although there were two exceptions (supplementary fig. S5, Supplementary Material online). The symbiotic species *M. vignae* showed higher rates of molecular evolution at symbiosis genes (supplementary fig. S5c and table S3, Supplementary Material online) while the symbiotic bacteria *A. caulinodans* showed the opposite pattern, an increase in dN/dS ratios genome-wide (supplementary fig. S5b and table S3, Supplementary Material online). When we compared dN/dS ratios at symbiosis genes that are shared among all 14 bacteria species to genes that are unique to certain species or shared among some of the species in our dataset, we observed significantly lower dN/dS values at conserved genes ($P = 0.0094$, supplementary fig. S6, Supplementary Material online).

Genes Under Positive Selection

Using the free-ratio dataset, we identified a total of 797 unique genes that were under positive selection in plant species. The number of genes under positive selection varied across species pairs (supplementary table S5, Supplementary

Material online). The dN/dS ratios at these genes were not consistently higher in one species over the other. When we subset our dataset to genes found in all 12 species, we found 78 genes under positive selection. Few of these genes were under positive selection in more than one species in our dataset (supplementary fig. S7, Supplementary Material online). We found only three genes under positive selection when we considered non-invasive legume dN/dS ratios calculated from two-ratio models in PAML (Fig. 4b). All three genes had higher dN/dS ratios in the mutualist species and a dN/dS ratio under 1 in the non-mutualistic relatives. We identified these genes as a telomere repeat-binding factor, very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase, and 50S ribosomal chloroplastic protein using blastn searches.

In rhizobia, we found 110 unique genes across all pairs that were under positive selection. There was no consistent pattern as to whether symbiotic species or non-symbiotic species contained more genes under positive selection (supplementary table S6, Supplementary Material online). When we considered genes that were common to all 14 rhizobia genomes, only 13 were under positive selection. None of these genes had dN/dS ratios greater than one in multiple species (supplementary fig. S8, Supplementary Material online).

Discussion

In this study, we investigated shifts in rates of molecular evolution genome-wide and at symbiotic genes in (i) mutualistic versus non-mutualistic legumes, (ii) symbiotic versus non-symbiotic rhizobia, and (iii) symbiotic and non-symbiotic clades in the *Ensifer* phylogeny. We sequenced and assembled transcriptomes for 12 non-model plant species from across the legume phylogeny. In bacteria, we collected rhizobia genomes from across the Alphaproteobacteria and Betaproteobacteria clades and analyzed 104 genomes across the *Ensifer* phylogeny. When there were significant differences in rates of molecular evolution between mutualists and non-mutualistic species, mutualists always showed faster evolutionary rates genome-wide. When we examined symbiosis genes in both legumes and rhizobia, mutualists did not consistently show higher dN/dS values compared to non-mutualist species. We consider in turn several possible explanations for faster evolutionary rates genome-wide, but not at symbiosis genes, in mutualist legumes and rhizobia.

The first potential explanation we consider is coevolution between legumes and rhizobia. If legumes and rhizobia are engaged in ongoing coevolution (i.e. reciprocal adaptation), we might predict that increased positive selection would elevate rates of molecular evolution genome-wide and especially at symbiosis genes, as reported in parasitic systems (Paterson et al. 2010; Bromham et al. 2013). However, we could also expect to see strong purifying selection at symbiosis genes in mutualist lineages if there is pressure to maintain symbiosis. All the symbiosis genes in our dataset were under purifying selection in both legumes and rhizobia (except for one gene in the *Mimosa* genus), and whether rates were relatively higher or lower in the mutualist partner varied among genes and taxa. Only our analyses on legumes that have apparently lost nodulation showed that mutualists consistently had higher dN/dS ratios than non-mutualistic legumes at symbiotic genes (though still less than one). Our results are consistent with previous work that has also failed to find strong evidence for population genetic signatures of balancing or positive selection driving evolution of symbiosis genes in legumes (Yoder 2016; Epstein et al. 2022) but have produced evidence that symbiosis genes are experiencing purifying selection (Epstein et al. 2022). In our data, strong purifying selection could be maintaining symbiosis genes in mutualists or we have not yet identified the most important genes for symbiosis in our study species. Comparing dN/dS ratios at symbiosis genes and across whole genomes within species, we saw no obvious differences, suggesting that selection is not acting differently on genes involved in symbiosis and genes found across the genome.

It is possible that different legumes may not have evolved the same symbiosis genes to associate with rhizobia and by mapping known symbiosis genes in *Medicago* to our transcriptomes, we may be missing some key symbiosis genes in our non-model legume species. We found little overlap in genes under positive selection across the dataset, also suggesting that different species are experiencing different selective pressures targeting different genes. However, some symbiosis genes in legumes are highly conserved across plant taxa (Schnabel et al. 2011; Roy et al. 2020) and might be involved in processes other than symbiosis with rhizobia. Therefore, conserved symbiosis genes are unlikely to show signs of positive selection expected from coevolution. In our plant dataset, symbiosis genes that were found in all species (i.e., highly conserved genes) showed similar dN/dS ratios to genes that were lineage-specific. However, less conserved genes did show more variation in dN/dS values suggesting that these symbiosis genes are less constrained. Overall, differences between mutualist and non-mutualist lineages appear to be relatively modest and therefore may only be detectable with increased power at the genome-wide level.

Symbiosis genes in rhizobia also show evolutionary conservation (Laranjo et al. 2008). We saw some evidence of stronger purifying selection acting on conserved genes in bacteria. Symbiosis genes that are unique to specific bacterial lineages should be less constrained. Indeed, we see that less conserved genes in symbiotic bacteria are potentially experiencing more positive selection compared to less conserved genes in non-symbiotic bacteria. However, in many cases, known symbiosis genes may not be where the most striking signals of coevolution occur in the genome. Transitioning to a mutualistic lifestyle could lead to drastic changes in the ecology and demography of interacting species, ultimately resulting in more changes across the whole genome rather than at a few key symbiosis genes. In rhizobia, the ability to nodulate plants is largely acquired through the horizontal transfer of symbiotic plasmids (Wernegreen and Riley 1999) or symbiosis islands

(Sullivan et al. 1995). Therefore, after accepting a symbiotic plasmid or island, a bacterium may undergo many mutations in the rest of the genome to accommodate the new genetic material that allows it access to a plant host. Previous research suggests that the initial introduction of plasmids with symbiotic genes is not enough to maintain cooperation in bacteria long term (Dewar et al. 2021). Therefore, a large-scale change to the genome plus living in a new habitat (the root nodule) may drive up dN/dS ratios across the chromosome, but not at genes on the symbiotic plasmid itself.

There could be other mechanisms driving differences in substitution rates between species other than mutualism. Differences in life history strategies in plants have been previously shown to influence rates of molecular evolution. For instance, the generation time hypothesis predicts that long-lived species (i.e. perennials) evolve more slowly compared to annuals (Smith and Donoghue 2008). Additionally, asexual plant species show increased accumulation of substitutions compared to sexually reproducing species (Hollister et al. 2015). Duplicated genes are likely to experience higher rates of molecular evolution (Kimura and Ohta 1974), indicating that organisms with higher ploidy levels might show elevated dN/dS values. Mutualists and non-mutualists generally had similar life history traits (ploidy level, generation time, and reproduction) within species pairs in our datasets. Therefore, whether plants participate in mutualism with rhizobia should be the main lifestyle difference between species in our dataset. Across pairs in our dataset, species differed greatly in their invasion history. High rates of molecular evolution may make plants better at invading new habitats as fast evolving organisms could have greater niche breadth and environmental tolerances (Whitney and Gabler 2008). Alternatively, elevated rates of molecular evolution may be a consequence of invasion because once established in a new environment, plants may have to adapt quickly (Young et al. 2018). While we cannot account for all possible differences in life history strategies (or other traits) between species, we were able to match most life history traits within species pairs when we could find data on these key traits in the literature. Therefore, differences in molecular evolution between mutualists and non-mutualists in our analyses are unlikely to be fully explained by differences in life history traits.

Another potential class of explanations for our results is the efficacy of selection in mutualist populations. Population size is predicted to have drastic effects on genome and molecular evolution. Species with small population sizes are expected to accumulate more deleterious mutations due to genetic drift (Charlesworth 2009). Previous work comparing island to mainland species (Woolfit and Bromham 2005) and small mammals to large mammals (Popadin et al. 2007) has shown that species with small population sizes experience faster rates of molecular evolution. In contrast, genetic drift is less pronounced in large populations and selection is more efficient (Charlesworth 2009). Species engaged in mutualism are expected to grow to larger population sizes because having a beneficial partner can help organisms occupy novel habitats, access nutrients when resources are scarce, and resist natural enemies (Afkhami et al. 2014; Weber and Agrawal 2014; Hayward et al. 2015). Therefore, the high dN/dS ratios in mutualists might be a result of large mutualist populations experiencing (more) positive selection than non-mutualists. Alternatively, higher dN/dS ratios across mutualist genomes could be a result of relaxed negative selection (Bromham et al. 2013). If mutualist legumes always rely on a rhizobia partner for access to nitrogen, there may be less selective pressure to maintain other genes that are important for accessing nutrients in the absence of a rhizobia partner. For instance, genes responsible for root proliferation might be under relaxed selection if it is less important for plants to "forage" for soil nutrients when rhizobia are present. Previous work has shown that mutualist traits and root foraging show a weak (quantitative) genetic correlation (Batstone et al. 2017), suggesting that traits could be evolving largely independently (i.e. relaxed selection on root foraging genes while purifying selection acts on symbiotic genes). Given that symbiotic genes in rhizobia are clustered on plasmids or in genomic islands (Wernegreen and Riley 1999), there is also opportunity for the chromosome to experience relaxed selection while purifying selection simultaneously acts on symbiotic plasmids or islands in bacteria. In addition, if rhizobia undergo more replication events while inside nodules than in the soil, then there could be relaxed selection on genes encoded on the chromosome for traits that only matter in the soil environment (e.g. competition with other microbes) and not inside the intracellular nodule environment.

Conclusion

A limited number of studies use both a comparative approach and a population genetics approach to investigating patterns of coevolution in mutualist species. In our study, we use a comparative approach to directly contrast related mutualist and non-mutualist species at a broad phylogenetic scale. We find that genome-wide elevated rates of molecular evolution are a common feature of both mutualist partners in the legume–rhizobium symbiosis. Genetic analyses of other positive species interactions also show accelerated rates of molecular evolution in mutualist species, suggesting that our findings are a general characteristic of mutualism overall (Lutzoni and Pagel 1997; Rubin and Moreau 2016). When we analyzed species pairs that represent a loss of the nodulation trait separately from the full legume dataset, slower evolution was particularly evident in three non-mutualistic legumes compared to mutualist legumes. In plants, there was some evidence that symbiotic genes are experiencing stronger purifying selection in mutualists compared to non-mutualist legumes. Our approach is complementary to the population genetics analyses performed in Epstein et al. (2022) where they found that symbiosis genes were also associated with purifying selection. Overall, a

combination of relaxed selection and more effective positive selection in large mutualist populations may be responsible for the high rates of molecular evolution we observed genome-wide in mutualist legumes and rhizobia.

Methods

Plant Materials and RNA Sequencing

The six legume species pairs in our analysis are: *Senna italica* and *Senna didymobotrya* (Azani et al. 2017), *Peltophorum africanum* and *Peltophorum dubium* (Haston et al. 2005), *Senna barclayana* and *Senna occidentalis* (LPWG et al. 2013; Azani et al. 2017), *Dalea mollissima* and *Dalea mollis* (McMahon and Hufford 2004; Zanne et al. 2014), *Calliandra eriophylla* and *C. humilis* (de Souza et al. 2013), and *Mimosa aculeaticarpa* and *Mimosa grahamii* (Simon et al. 2011). We obtained seed for each of these legume species from either the USDA-ARS Germplasm Resources Information Network or the KEW Royal Botanical Gardens Millennial Seed Bank. We grew one plant of each species in a growth chamber with daytime temperature set to 28 °C, nighttime temperature set to 19 °C, and a light period of 15 h. We prepared all seeds for germination by nicking the seed coat with a razor blade and incubating the scarified seed at 30 °C overnight on wet filter paper in a petri dish. *Senna occidentalis* and *S. barclayana* seeds were placed in boiling water for 10 min prior to scarification. *Peltophorum, S. didymobotrya*, and *S. italica* seeds were treated with sulfuric acid for 10 min (Alves et al. 2011) and rinsed with distilled water before scarification. All plants were grown in sterile sand in Magenta boxes. Once a week, the bottom compartment of each box was filled with a high-nitrogen fertilizer diluted to one-quarter strength (recipe in Zhang et al. 2020). We did not inoculate plants with rhizobia so that we could collect and compare RNA from roots without nodules from both the mutualistic and non-mutualistic plant species. Previous work has shown that association with rhizobia causes differential expression of many genes with diverse functions, but symbiosis genes are still generally expressed in legumes even in the non-symbiotic state (Afkhami and Stinchcombe 2016). Therefore, symbiosis genes are still captured in transcriptomes from legumes without rhizobia (see Results, above). Plants were harvested for root tissue after 5 weeks of growth or when the plant had 10 true leaves. Roots were rinsed with water and a small amount of fresh root tissue was cut and stored in Eppendorf tubes. We collected an average of 86 mg of tissue for all species except for the *Peltophorum* species for which we collected 30 mg each. Tubes were flash frozen in liquid nitrogen and stored in a −80 °C freezer. We followed the Sigma Aldrich Plant Total RNA Kit instructions to isolate RNA and obtained between 44.1 and 333 ng/µl of RNA per sample. Samples were submitted to Genome Quebec for sequencing on the NovaSeq 6000 Sequencing System (PE100). We received 67,812,540 to 98,924,616 paired end reads per sample with an average quality of 36.

De Novo Transcriptome Assembly

We checked the quality of the sequences with FastQC v0.11.9 (Andrews 2010). We cut adapters, removed leading and trailing low-quality bases (below quality 3), and trimmed sequences to a minimum length of 30 using Trimmomatic v0.39 (Bolger et al. 2014). We assembled de novo transcriptomes for each species from the cleaned reads using RNASpades v3.15.2 (Bushmanova et al. 2019) and Trinity v2.8.1 (Haas et al. 2013) with default parameters. We ran CD-HIT v4.8.1 (Fu et al. 2012) to remove redundant transcripts from the assemblies and checked assembly quality with rnaQUEST v2.2.1 (Bushmanova et al. 2019). Transcriptomes produced from RNASpades had fewer but longer contigs, therefore the rest of the analysis was performed on the RNASpades assemblies (supplementary table S6, Supplementary Material online). We predicted coding regions using TransDecoder v5.5.0 and we removed any contigs with no predicted peptide.

Ortholog Identification

We identified a total of 308 single-copy orthologous genes shared in the proteomes of all 12 legume species using OrthoFinder v2.4 (Emms and Kelly 2019). We expanded this set to also include orthologous genes found in at least four species, resulting in a total of 3,548 genes for analysis. We found 438 single-copy orthologous genes in all 14 rhizobia species using default settings in OrthoFinder v2.4 (Emms and Kelly 2019). When we included orthologs present in at least four species, we identified 2,812 genes shared among the bacteria strains in our dataset. We identified 456 single copy orthologous genes shared among all 104 *Ensifer* genomes.

Estimating Rates of Molecular Evolution

To calculate dN/dS values in each species for each of the 3,548 plant orthologous genes and 2,812 bacteria genes, we first compiled orthologous nucleotide sequences (cds files) into single fasta files. For each gene, we executed alignments in PRANK v.170427 with the "-codon" option (Markova-Raina and Petrov 2011). We constructed maximum-likelihood gene trees for each orthologous gene using default settings in RAxML with the substitution model set to GTRCATX (Stamatakis 2014). Gene trees and sequence alignments were used as input for dN/dS analysis in PAML v.4.9j (Yang 2007). We implemented a "free-ratios" model in PAML to calculate a separate dN/dS value for each branch in the gene tree. Including six closely related pairs of legumes in the gene trees allows for the pairs to serve as outgroups for the different ingroup tests. We extracted dN/dS ratios for each tip of the tree to obtain a unique dN/dS value for each species and performed the remaining analyses in R (R Core Team 2024). We removed all dN/dS ratios >10 as values this high are likely a result of either an error in assembly or overparameterization in the PAML model for complicated gene trees. After filtering abnormally high dN/dS ratios, our sample size included 210 orthologs shared among all 12 plant species

and 227 orthologs shared among all 14 bacteria species in our paired analysis. To compare dN/dS ratios between mutualists and non-mutualistic taxa genome-wide, we performed paired Wilcoxon signed-rank tests on dN/dS values for all orthologous genes between mutualist species and their non-mutualistic relative in R (Danneels et al. 2021; R Core Team 2024).

To compare molecular evolution in the legume species pairs that represent a loss of mutualism, we labeled all species in the gene trees as a mutualist (test) or non-mutualist (reference) and allowed PAML to model separate dN/dS ratios for test and reference branches. We compared model fit between the two-rate model and a model where all species were constrained to evolve at the same rate. We identified the number of genes that showed significantly different dN/dS values in mutualists and non-mutualistic relatives. We then performed paired Wilcoxon signed-rank tests on dN/dS values calculated at these significant genes.

To evaluate rates of molecular evolution within the *Ensifer* genus, we performed a two-ratio model in PAML where we labeled symbiotic species (largely the *Sinorhizobium* clade within the *Ensifer* phylogeny) as the test branches and non-symbiotic species as reference branches. Separate dN/dS ratios were calculated in PAML for test and reference lineages in the *Ensifer* phylogeny. Out of 456 single copy orthologs tested in the two-ratio model, 405 genes showed significant differences between the symbiotic and non-symbiotic clades. We performed paired Wilcoxon signed-rank tests using dN/dS ratios calculated from these 405 significant genes. We also split up the dN/dS ratios across the genome into two categories: genes found on the pSymB plasmid and genes found elsewhere in the genome. We used the annotated genome assembly for *Sinorhizobium meliloti* USDA1021 (GCA_002197445.1) to identify the location of our orthologs in the genome. We found 70 genes in our dataset and performed paired Wilcoxon signed-rank tests to compare rates of molecular evolution in symbiotic and non-symbiotic bacteria strains.

Identifying Symbiosis Genes

To identify symbiosis genes in our transcriptomes, we first obtained a list of sequences (Epstein et al. 2022) for 224 genes that Roy et al. (2020) identified as important for symbiosis in *Medicago*. We mapped these sequences to all 12 legume transcriptomes using bwa-mem. We extracted the positions in our transcriptomes where the symbiosis sequences mapped and filtered our full dataset of dN/dS ratios for these genes. We then compared the dN/dS ratios for these symbiosis genes in the mutualist legume and their matching non-symbiotic relative. We also searched for symbiosis genes among the genes that were significantly different in mutualist and non-mutualist species (calculated using two-ratio models in PAML). We performed a blastn search on sequences for any significant symbiosis genes against the flowering plant database (taxid: 3398). We performed unpaired Wilcoxon–Mann–Whitney tests to test for significant differences in patterns of molecular evolution at symbiosis genes in our dataset and genes found across the rest of the genome within species. To determine if symbiosis genes that show differing levels of conservation also have differing rates of molecular evolution, we first calculated how many species each symbiosis gene was found in. We then calculated average dN/dS values for each symbiosis gene and compared dN/dS values in genes shared among all species (i.e. highly conserved genes) to genes found in only one or less than all 12 species.

To identify symbiosis genes in bacteria, we searched the annotated mutualist genomes for gene descriptions including the following key words: nod, noe, nfe, nodul, nif, fix, fixation, and nitrogenase. Within our set of orthologs, we found 33 genes that contained key words related to nitrogen fixation and nodulation in our dataset of 14 bacteria species. We then compared dN/dS ratios between symbiotic and free-living strains at these symbiosis genes across all pairs in the dataset. We also compared dN/dS values at symbiosis genes to the rates across the rest of the genome within species using the same methods described above in the legume symbiosis analysis. Bacteria symbiosis genes also varied in how many species they occurred in. Therefore, we also compared average dN/dS values at highly conserved symbiosis genes and less conserved symbiosis genes using unpaired Wilcoxon–Mann–Whitney tests.

Supplementary Material

Supplementary material is available at *Molecular Biology and Evolution* online.

Acknowledgments

We thank Stephen Wright, Mark Hibbins, and Tyler Kent for advice on running PAML models and Matt Pennell for suggestions on analyzing dN/dS ratios. We also thank Maria Tocora for advice on running transcriptome assembly programs and George diCenzo for feedback on analyzing the *Ensifer* samples. T.L.H. received a Queen Elizabeth II Graduate Scholarship and a Natural Sciences and Engineering Research Council of Canada (NSERC) Graduate Scholarship (PGSD3-502550-207) during her PhD when this research was performed. T.L.H. also secured an internal EEB PhD Student Research Grant from the University of Toronto, which was used for sequencing costs of the transcriptomes. M.E.F. (RGPIN-2021-03711) and J.R.S. (RGPIN-2022-04366) both supported the research with NSERC Discovery Grants.

Data Availability

Sequence data has been deposited on Sequence Read Archive (BioProject PRJNA1188642). Assembly codes for the bacteria and rhizobia genomes used in the analysis are listed in Table 2 and supplementary table S2, Supplementary Material online. All code for reproducing the analysis is available on the github repository named mutualism-molecular-evolution (https://github.com/harri318/mutualism-molecular-evolution.git).

References

Afkhami ME, McIntyre PJ, Strauss SY. Mutualist-mediated effects on species' range limits across large geographic scales. Ecol Lett. 2014:**17**(10):1265–1273. https://doi.org/10.1111/ele.12332.

Afkhami ME, Stinchcombe JR. Multiple mutualist effects on genome-wide expression in the tripartite association between *Medicago truncatula*, nitrogen-fixing bacteria and mycorrhizal fungi. Mol Ecol. 2016:**25**(19):4946–4962. https://doi.org/10.1111/mec.13809.

Alves EU, Guedes RS, Gonçalves EP, Viana JS, Santos SDS, de Moura MF. Effect of temperature and substrate on germination of *Peltophorum dubium* (Sprengel) Taubert seeds. Acta Sci Biol Sci. 2011:**33**(1):113–118. https://doi.org/10.4025/actascibiolsci.v33i1.7057.

Andrews M, Andrews ME. Specificity in legume-rhizobia symbioses. Int J Mol Sci. 2017:**18**(4):705. https://doi.org/10.3390/ijms18040705.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. [accessed 2022 Sep]. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Azani N, Babineau M, Bailey CD, Banks H, Barbosa AR, Pinto RB, Boatwright JS, Borges LM, Brown GK, Bruneau A, et al. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny: The Legume Phylogeny Working Group (LPWG). Taxon. 2017:**66**(1):44–77. https://doi.org/10.12705/661.3.

Barker JL, Bronstein JL, Friesen ML, Jones EI, Reeve HK, Zink AG, Frederickson ME. Synthesizing perspectives on the evolution of cooperation within and between species. Evolution. 2017:**71**(4):814–825. https://doi.org/10.1111/evo.13174.

Barnett MJ, Fisher RF, Jones T, Komp C, Abola AP, Barloy-Hubler F, Bowser L, Capela D, Galibert F, Gouzy J, et al. Nucleotide sequence and predicted functions of the entire *Sinorhizobium meliloti* pSymA megaplasmid. Proc Natl Acad Sci U S A. 2001:**98**(17):9883–9888. https://doi.org/10.1073/pnas.161294798.

Batstone RT. Genomes within genomes: nested symbiosis and its implications for plant evolution. New Phytol. 2022:**234**(1):28–34. https://doi.org/10.1111/nph.17847.

Batstone RT, Dutton EM, Wang D, Yang M, Frederickson ME. The evolution of symbiont preference traits in the model legume *Medicago truncatula*. New Phytol. 2017:**213**(4):1850–1861. https://doi.org/10.1111/nph.14308.

Bergstrom CT, Lachmann M. The Red King effect: when the slowest runner wins the coevolutionary race. Proc Natl Acad Sci U S A. 2003:**100**(2):593–598. https://doi.org/10.1073/pnas.0134966100.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014:**30**(15):2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Brockhurst MA, Chapman T, King KC, Mank JE, Paterson S, Hurst GDD. Running with the Red Queen: the role of biotic conflicts in evolution. Proc Biol Sci. 2014:**281**(1797):20141382. https://doi.org/10.1098/rspb.2014.1382.

Bromham L. Why do species vary in their rate of molecular evolution? Biol Lett. 2009:**5**(3):401–404. https://doi.org/10.1098/rsbl.2009.0136.

Bromham L, Cowman PF, Lanfear R. Parasitic plants have increased rates of molecular evolution across all three genomes. BMC Evol Biol. 2013:**13**(1):126. https://doi.org/10.1186/1471-2148-13-126.

LPWG; Bruneau A, Doyle JJ, Herendeen P, Hughes C, Kenicer G, Lewis G, Mackinder B, Pennington RT, Sanderson MJ, et al. Legume phylogeny and classification in the 21st century: progress, prospects and lessons for other species-rich clades. Taxon. 2013:**62**(2):217–248. https://doi.org/10.12705/622.8.

Bushmanova E, Antipov D, Lapidus A, Prjibelski AD. rnaSPAdes: a de novo transcriptome assembler and its application to RNA-Seq data. GigaScience. 2019:**8**(9):giz100. https://doi.org/10.1093/gigascience/giz100.

Charlesworth B. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 2009:**10**(3):195–205. https://doi.org/10.1038/nrg2526.

Charlesworth B. Stabilizing selection, purifying selection, and mutational bias in finite populations. Genetics. 2013:**194**(4):955–971. https://doi.org/10.1534/genetics.113.151555.

Danneels B, Viruel J, Mcgrath K, Janssens SB, Wales N, Wilkin P, Carlier A. Patterns of transmission and horizontal gene transfer in the *Dioscorea sansibarensis* leaf symbiosis revealed by whole-genome sequencing. Curr Biol. 2021:**31**(12):2666–2673.e4. https://doi.org/10.1016/j.cub.2021.03.049.

Delaye L, Ruiz-Ruiz S, Calderon E, Tarazona S, Conesa A, Moya A. Evidence of the red-queen hypothesis from accelerated rates of evolution of genes involved in biotic interactions in pneumocystis. Genome Biol Evol. 2018:**10**(6):1596–1606. https://doi.org/10.1093/gbe/evy116.

de Souza ÉR, Lewis GP, Forest F, Schnadelbach AS, van den Berg C, de Queiroz LP. Phylogeny of Calliandra (Leguminosae: Mimosoideae) based on nuclear and plastid molecular markers. Taxon. 2013:**62**(6):1200–1219. https://doi.org/10.12705/626.2.

Dewar AE, Thomas JL, Scot TW, Wild G, Griffin AS, West SA, Ghoul M. Plasmids do not consistently stabilize cooperation across bacteria but may promote broad pathogen host-range. Nat Ecol Evol. 2021:**5**(12):1624–1636. https://doi.org/10.1038/s41559-021-01573-2.

diCenzo G, Checcucci A, Bazzicalupo M, Mengoni A, Viti C, Dziewit L, Finan TM, Galardini M, Fondi M. Metabolic modelling reveals the specialization of secondary replicons for niche adaptation in *Sinorhizobium meliloti*. Nat Commun. 2016:**7**(1):12219. https://doi.org/10.1038/ncomms12219.

Doyle JJ. Phylogenetic perspectives on the origins of nodulation. Mol Plant Microbe Interact. 2011:**24**(11):1289–1295. https://doi.org/10.1094/MPMI-05-11-0114.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:**20**(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Epstein B, Burghardt LT, Heath KD, Grillo MA, Kostanecki A, Hämälä T, Young ND, Tiffin P. Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis. Mol Ecol. 2022:**32**(14):3798–3811. https://doi.org/10.1111/mec.16602.

Epstein B, Tiffin P. Comparative genomics reveals high rates of horizontal transfer and strong purifying selection on rhizobial symbiosis genes. Proc R Soc Lond B Biol Sci. 2021:**288**(1942):20201804. https://doi.org/10.1098/rspb.2020.1804.

Fagorzi C, Ilie A, Decorosi F, Cangioli L, Viti C, Mengoni A, diCenzo GC. Symbiotic and nonsymbiotic members of the genus ensifer (syn. sinorhizobium) are separated into two clades based on comparative genomics and high-throughput phenotyping. Genome Biol Evol. 2020:**12**(12):2521–2534. https://doi.org/10.1093/gbe/evaa221.

Finan TM, Weidner S, Wong K, Buhrmester J, Chain P, Vorhölter FJ, Hernandez-Lucas I, Becker A, Cowie A, Gouzy J, et al. The complete sequence of the 1,683-kb pSymB megaplasmid from the N_2-fixing endosymbiont *Sinorhizobium meliloti*. Proc Natl Acad Sci U S A. 2001:**98**(17):9889–9894. https://doi.org/10.1073/pnas.161294698.

Friesen ML. Widespread fitness alignment in the legume–rhizobium symbiosis. New Phytol. 2012:**194**(4):1096–1111. https://doi.org/10.1111/j.1469-8137.2012.04099.x.

Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012:**28**(23):3150–3152. https://doi.org/10.1093/bioinformatics/bts565.

Geddes BA, Kearsley J, Morton R, diCenzo GC, Finan TM. The genomes of rhizobia. Advances in botanical research. Vol. 94. Amsterdam, The Netherlands: Elsevier; 2020. p. 213–249. https://linkinghub.elsevier.com/retrieve/pii/S0065229619300916.

Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, et al.

Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. Science. 2018:**361**(6398):eaat1743. https://doi.org/10.1126/science.aat1743.
Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. De novo transcript sequence reconstruction from RNA-Seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013:**8**(8):1494–1512. https://doi.org/10.1038/nprot.2013.084.
Haston EM, Lewis GP, Hawkins JA. A phylogenetic reappraisal of the Peltophorum group (Caesalpinieae: Leguminosae) based on the chloroplast trnL-F, rbcL and rps16 sequence data. Am J Bot. 2005:**92**(8):1359–1371. https://doi.org/10.3732/ajb.92.8.1359.
Hayward J, Horton TR, Pauchard A, Nuñez MA. A single ectomycorrhizal fungal species can enable a Pinus invasion. Ecology. 2015:**96**(5):1438–1444. https://doi.org/10.1890/14-1100.1.
Heath KD, Stinchcombe JR. Explaining mutualism variation: a new evolutionary paradox? Evolution. 2014:**68**(2):309–317. https://doi.org/10.1111/evo.12292.
Hembry DH, Yoder JB, Goodman KR. Coevolution and the diversification of life. Am Nat. 2014:**184**(4):425–438. https://doi.org/10.1086/677928.
Hollister JD, Greiner S, Wang W, Wang J, Zhang Y, Wong GK-S, Wright SI, Johnson MTJ. Recurrent loss of sex is associated with accumulation of deleterious mutations in oenothera. Mol Biol Evol. 2015:**32**(4):896–905. https://doi.org/10.1093/molbev/msu345.
Kimura M, Ohta T. On some principles governing molecular evolution. Proc Natl Acad Sci U S A. 1974:**71**(7):2848–2852. https://doi.org/10.1073/pnas.71.7.2848.
Laranjo M, Alexandre A, Rivas R, Velázquez E, Young JPW, Oliveira S. Chickpea rhizobia symbiosis genes are highly conserved across multiple Mesorhizobium species. FEMS Microbiol Ecol. 2008:**66**(2):391–400. https://doi.org/10.1111/j.1574-6941.2008.00584.x.
Lutzoni F, Pagel M. Accelerated evolution as a consequence of transitions to mutualism. Proc Natl Acad Sci U S A. 1997:**94**(21):11422–11427. https://doi.org/10.1073/pnas.94.21.11422.
Markova-Raina P, Petrov D. High sensitivity to aligner and high rate of false positives in the estimates of positive selection in the 12 Drosophila genomes. Genome Res. 2011:**21**(6):863–874. https://doi.org/10.1101/gr.115949.110.
McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012:**10**(1):13–26. https://doi.org/10.1038/nrmicro2670.
McMahon M, Hufford L. Phylogeny of amorpheae (Fabaceae: papilionoideae). Am J Bot. 2004:**91**(8):1219–1230. https://doi.org/10.3732/ajb.91.8.1219.
Nuismer SL, Jordano P, Bascompte J. Coevolution and the architecture of mutualistic networks. Evolution. 2013:**67**(2):338–354. https://doi.org/10.1111/j.1558-5646.2012.01801.x.
O'Brien AM, Jack CN, Friesen ML, Frederickson ME. Whose trait is it anyways? Coevolution of joint phenotypes and genetic architecture in mutualisms. Proc R Soc Lond B Biol Sci. 2021:**288**(1942):20202483. https://doi.org/10.1098/rspb.2020.2483.
Parshuram ZA, Harrison TL, Simonsen AK, Stinchcombe JR, Frederickson ME. Nonsymbiotic legumes are more invasive, but only if polyploid. New Phytol. 2023:**237**(3):758–765. https://doi.org/10.1111/nph.18579.
Paterson S, Vogwill T, Buckling A, Benmayor R, Spiers AJ, Thomson NR, Quail M, Smith F, Walker D, Libberton B, et al. Antagonistic coevolution accelerates molecular evolution. Nature. 2010:**464**(7286):275–278. https://doi.org/10.1038/nature08798.
Popadin K, Polishchuk LV, Mamirova L, Knorre D, Gunbin K. Accumulation of slightly deleterious mutations in mitochondrial protein-coding genes of large versus small mammals. Proc Natl Acad Sci U S A. 2007:**104**(33):13390–13395. https://doi.org/10.1073/pnas.0701256104.
Rahimlou S, Bahram M, Tedersoo L. Phylogenomics reveals the evolution of root nodulating alpha- and beta-proteobacteria (rhizobia). Microbiol Res. 2021:**250**:126788. https://doi.org/10.1016/j.micres.2021.126788.
R Core Team. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2024. https://www.R-project.org/.
Roy S, Liu W, Nandety RS, Crook A, Mysore KS, Pislariu CI, Frugoli J, Dickstein R, Udvardi MK. Celebrating 20 years of genetic discoveries in legume nodulation and symbiotic nitrogen fixation. Plant Cell. 2020:**32**(1):15–41. https://doi.org/10.1105/tpc.19.00279.
Rubin BER, Moreau CS. Comparative genomics reveals convergent rates of evolution in ant–plant mutualisms. Nat Commun. 2016:**7**(1):12679. https://doi.org/10.1038/ncomms12679.
Sachs JL, Russell JE, Hollowell AC. Evolutionary instability of symbiotic function in *Bradyrhizobium japonicum*. PLoS One. 2011:**6**(11):e26370. https://doi.org/10.1371/journal.pone.0026370.
Schnabel EL, Kassaw TK, Smith LS, Marsh JF, Oldroyd GE, Long SR, Frugoli JA. The ROOT DETERMINED NODULATION1 gene regulates nodule number in roots of *Medicago truncatula* and defines a highly conserved, uncharacterized plant gene family. Plant Physiol. 2011:**157**(1):328–340. https://doi.org/10.1104/pp.111.178756.
Simon MF, Grether R, de Queiroz LP, Särkinen TE, Dutra VF, Hughes CE. The evolutionary history of Mimosa (Leguminosae): toward a phylogeny of the sensitive plants. Am J Bot. 2011:**98**(7):1201–1221. https://doi.org/10.3732/ajb.1000520.
Simonsen AK, Dinnage R, Barrett LG, Prober SM, Thrall PH. Symbiosis limits establishment of legumes outside their native range at a global scale. Nat Commun. 2017:**8**(1):14790. https://doi.org/10.1038/ncomms14790.
Smith SA, Donoghue MJ. Rates of molecular evolution are linked to life history in flowering plants. Science. 2008:**322**(5898):86–89. https://doi.org/10.1126/science.1163197.
Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. Dynamics of disease resistance polymorphism at the Rpm1 locus of Arabidopsis. Nature. 1999:**400**(6745):667–671. https://doi.org/10.1038/23260.
Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014:**30**(9):1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of Rhizobium loti arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci U S A. 1995:**92**(19):8985–8989. https://doi.org/10.1073/pnas.92.19.8985.
Tiffin P, Ross-Ibarra J. Advances and limits of using population genetics to understand local adaptation. Trends Ecol Evol. 2014:**29**(12):673–680. https://doi.org/10.1016/j.tree.2014.10.004.
van Rhijn P, Vanderleyden J. The Rhizobium-plant symbiosis. Microbiol Rev. 1995:**59**(1):124–142. https://doi.org/10.1128/mr.59.1.124-142.1995.
Wang Q, Liu J, Zhu H. Genetic and molecular mechanisms underlying symbiotic specificity in legume-rhizobium interactions. Front Plant Sci. 2018:**9**:313. https://doi.org/10.3389/fpls.2018.00313.
Weber MG, Agrawal AA. Defense mutualisms enhance plant diversification. Proc Natl Acad Sci U S A. 2014:**111**(46):16442–16447. https://doi.org/10.1073/pnas.1413253111.
Wernegreen JJ. Genome evolution in bacterial endosymbionts of insects. Nat Rev Genet. 2002:**3**(11):850–861. https://doi.org/10.1038/nrg931.
Wernegreen JJ, Riley MA. Comparison of the evolutionary dynamics of symbiotic and housekeeping loci: a case for the genetic coherence of rhizobial lineages. Mol Biol Evol. 1999:**16**(1):98–113. https://doi.org/10.1093/oxfordjournals.molbev.a026041.
Werner GDA, Cornwell WK, Sprent JI, Kattge J, Kiers ET. A single evolutionary innovation drives the deep evolution of symbiotic N2-fixation in angiosperms. Nat Commun. 2014:**5**(1):4087. https://doi.org/10.1038/ncomms5087.
Whitney KD, Gabler CA. Rapid evolution in introduced species, 'invasive traits' and recipient communities: challenges for predicting invasive potential. Divers Distrib. 2008:**14**(4):569–580. https://doi.org/10.1111/j.1472-4642.2008.00473.x.

Woolfit M, Bromham L. Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol Biol Evol. 2003:**20**(9):1545–1555. https://doi.org/10.1093/molbev/msg167.

Woolfit M, Bromham L. Population size and molecular evolution on islands. Proc Biol Sci. 2005:**272**(1578):2277–2282. https://doi.org/10.1098/rspb.2005.3217.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007:**24**(8):1586–1591. https://doi.org/10.1093/molbev/msm088.

Yoder JB. Understanding the coevolutionary dynamics of mutualism with population genomics. Am J Bot. 2016:**103**(10):1742–1752. https://doi.org/10.3732/ajb.1600154.

Young RG, Mitterboeck TF, Loeza-Quintana T, Adamowicz SJ. Rates of molecular evolution and genetic diversity in European vs. North American populations of invasive insect species. EJE. 2018:**115**:718–728. https://doi.org/10.14411/eje.2018.071.

Zanne AE, Tank DC, Cornwell WK, Eastman JM, Smith SA, FitzJohn RG, McGlinn DJ, O'Meara BC, Moles AT, Reich PB, et al. Three keys to the radiation of angiosperms into freezing environments. Nature. 2014:**506**(7486):89–92. https://doi.org/10.1038/nature12872.

Zhang X, Wang L, Li J, Batstone RT, Frederickson ME. Medicago truncatula adjusts root proliferation, nodule formation, and partner choice in response to local N heterogeneity. Plant Soil. 2020:**450**(1-2):417–428. https://doi.org/10.1007/s11104-020-04433-3.