



Pangenome Evolution Reconciles Robustness and Instability of Rhizobial Symbiosis

 Alexandra J. Weisberg,^a Arafat Rahman,^b Dakota Backus,^a Parinita Tyavanagimatt,^a  Jeff H. Chang,^a  Joel L. Sachs^{b,c,d}

^aDepartment of Botany and Plant Pathology, Oregon State University, Corvallis, Oregon, USA

^bDepartment of Evolution, Ecology and Organismal Biology, University of California Riverside, Riverside, California, USA

^cDepartment of Microbiology and Plant Pathology, University of California Riverside, Riverside, California, USA

^dInstitute for Integrative Genome Biology, University of California Riverside, Riverside, California, USA

ABSTRACT Root nodulating rhizobia are nearly ubiquitous in soils and provide the critical service of nitrogen fixation to thousands of legume species, including staple crops. However, the magnitude of fixed nitrogen provided to hosts varies markedly among rhizobia strains, despite host legumes having mechanisms to selectively reward beneficial strains and to punish ones that do not fix sufficient nitrogen. Variation in the services of microbial mutualists is considered paradoxical given host mechanisms to select beneficial genotypes. Moreover, the recurrent evolution of non-fixing symbiont genotypes is predicted to destabilize symbiosis, but breakdown has rarely been observed. Here, we deconstructed hundreds of genome sequences from genotypically and phenotypically diverse *Bradyrhizobium* strains and revealed mechanisms that generate variation in symbiotic nitrogen fixation. We show that this trait is conferred by a modular system consisting of many extremely large integrative conjugative elements and few conjugative plasmids. Their transmissibility and propensity to reshuffle genes generate new combinations that lead to uncooperative genotypes and make individual partnerships unstable. We also demonstrate that these same properties extend beneficial associations to diverse host species and transfer symbiotic capacity among diverse strains. Hence, symbiotic nitrogen fixation is underpinned by modularity, which engenders flexibility, a feature that reconciles evolutionary robustness and instability. These results provide new insights into mechanisms driving the evolution of mobile genetic elements. Moreover, they yield a new predictive model on the evolution of rhizobial symbioses, one that informs on the health of organisms and ecosystems that are hosts to symbionts and that helps resolve the long-standing paradox.

IMPORTANCE Genetic variation is fundamental to evolution yet is paradoxical in symbiosis. Symbionts exhibit extensive variation in the magnitude of services they provide despite hosts having mechanisms to select and increase the abundance of beneficial genotypes. Additionally, evolution of uncooperative symbiont genotypes is predicted to destabilize symbiosis, but breakdown has rarely been observed. We analyzed genome sequences of *Bradyrhizobium*, bacteria that in symbioses with legume hosts, fix nitrogen, a nutrient essential for ecosystems. We show that genes for symbiotic nitrogen fixation are within elements that can move between bacteria and reshuffle gene combinations that change host range and quality of symbiosis services. Consequently, nitrogen fixation is evolutionarily unstable for individual partnerships, but is evolutionarily stable for legume-*Bradyrhizobium* symbioses in general. We developed a holistic model of symbiosis evolution that reconciles robustness and instability of symbiosis and informs on applications of rhizobia in agricultural settings.

KEYWORDS evolution, symbiosis, robustness, nitrogen fixation, mobile genetic elements, rhizobia

Editor Vaughn S. Cooper, University of Pittsburgh

Copyright © 2022 Weisberg et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Joel L. Sachs, joels@ucr.edu.

The authors declare no conflict of interest.

Received 20 January 2022

Accepted 14 March 2022

A predictive understanding of symbiosis evolution is critical to inform on the health of hosts and ecosystems in which microbial symbionts reside. A central feature of symbiosis is the variation in the magnitude of services that symbionts provide (1). At one extreme are uncooperative strains, those that abandon intimate association with hosts or that are ineffective in providing benefits (2). Their recurrent origins are predicted to destabilize symbiosis, but breakdown has rarely been observed, presenting a paradox of instability and robustness (2, 3). Discovering mechanisms that generate and maintain variation in microbial symbionts is foundational for building a unifying framework for symbiosis evolution and resolving the paradox (1).

Genetic variation occurs through mutation, recombination, and gene flow, which together underlie the concept of the pangenome, a nonredundant set of genes in organisms related through ancestry and divisible into core and accessory genomes (4). The core represents genes predicted to encode functions common and essential among related strains. The accessory genome consists of genes polymorphic in presence/absence and confers upon related individuals the ability to adopt diverse lifestyles. In bacteria, mobile genetic elements (MGEs) are important molecules of accessory genomes. MGEs tend to be arranged into functional units, an organization that promotes modularity, a property that preserves functionality while allowing components to separate and recombine, thereby conferring flexibility and robustness to adapt to different conditions (5). Thus, MGEs are important to bacterial evolution because their exchangeability increases opportunities to recombine, reassort accessory cargo genes, and diversify. Resolving relationships of MGEs is essential for understanding the evolution of traits they encode.

Integrative conjugative elements (ICEs) and plasmids are two major classes of MGEs and both carry cargo genes that encode traits, such as virulence and antibiotic resistance, associated with transitions in the evolution of bacteria (6, 7). ICEs typically recombine into chromosomes and replicate passively, while plasmids typically replicate independently from the chromosome. ICEs encode integrases that can mediate site-specific recombination between homologous attachment (*att*) sequences located on the ICE and chromosome, which are often in a conserved gene, such as a tRNA gene. ICEs can also excise, circularize, mobilize one strand through a type IV secretion system (T4SS), and recombine into the genome of recipient strains and back into that of the donor strain. The identification of ICEs, and their distinction from nonmobile genomic islands, is confounded by challenges in determining *att* sites as well as by compositional variation and presence of many repeat sequences that fragment these elements (8). Consequently, fundamental aspects of the mechanisms that generate diversity and the extent of variation within related ICEs are poorly understood.

MGEs are crucial for the ability of many taxa of rhizobia to carry out symbiotic nitrogen fixation (SNF), a service essential to ecosystems (9). Beneficial rhizobia are defined by two sets of functions, the capacity to nodulate hosts and the ability to fix nitrogen, that are often encoded on MGEs. Symbiosis ICEs and plasmids have clusters of *nod*, *nol*, and *noe* genes (collectively *nod* genes here) for the synthesis of Nod factors, signaling molecules that initiate interactions and influence host specificity, as well as clusters of *nif/fix* genes for the catalysis of nitrogen fixation (10). Common NodABC proteins synthesize the core signaling structure while others, encoded by genes polymorphic in presence/absence, decorate the core with diverse substitutions (9). Type III secretion system (T3SS)- and effector-encoding genes are frequently linked to symbiosis genes (11). Effector genes influence host specificity because of their dichotomous abilities in dampening and inciting plant immune responses (12).

Symbiosis ICEs (symICEs) were first characterized in *Mesorhizobium* (10, 13, 14). *Mesorhizobium* symICEs adopt monopartite or polypartite structures, with the latter in *Mesorhizobium* consisting of three elements, each encoding their own integrase, that interact to circularize and mediate genomic rearrangements during integration or excision from the chromosome (13). In agronomic landscapes – where symICEs have been extensively studied – transfer of entire symICEs promotes diversification of *Mesorhizobium*

strains that nodulate crop hosts, but with mixed results for effective symbiosis. For instance, in settings where legume crops and compatible rhizobia were introduced by growers, transfer of an entire symICE from highly effective inoculant strains to native rhizobia occurred and generated a diversity of novel nodulating strains; however, many of them were ineffective in fixing nitrogen and the bases for loss of SNF on target crops remain unknown (15–17). Moreover, processes that drive symICE variation, e.g., monopartite and polypartite and diverse integration sites, as well as the selective advantages for such variation in nitrogen fixing bacteria are also unknown (18, 19).

In contrast to other rhizobia, most members of *Bradyrhizobium* are traditionally thought to have genes necessary for SNF clustered in a genomic island referred to as a symbiosis island (SI) (20). *Bradyrhizobium* is cosmopolitan and its members can fix nitrogen in facultative associations with diverse members of the legume family, Fabaceae. (21). Host species include at least 24 of the 33 legume tribes that can form nodules, spread across the three legume subfamilies, Caesalpinioideae, Mimosoideae, and Papilionoideae. *Bradyrhizobium* populations have been extensively characterized in native host communities, and shown to exhibit broad variation in symbiotic capacity, providing natural tests to investigate genomic drivers of this variation (22). We compiled and analyzed a data set of genome sequences of genetically and phenotypically diverse strains of *Bradyrhizobium* (Data set S1; Extended Data sets S1–S3 available at https://github.com/osuchanglab/BradyrhizobiumPangenomeManuscript/tree/main/Extended_Supplementary_Materials). Critical to resolving the drivers of symbiosis variation in natural settings, we sequenced genomes of 85 strains (here metapopulation strains) isolated from across an 800 km transect of wild *Acmispon strigosus* populations in California and phenotyped the strains as beneficial (Nod⁺/Fix⁺), ineffective (Nod⁺/Fix⁻), or non-nodulating (Nod⁻/Fix⁻) on this host species (22, 23). We additionally included 167 publicly available genome sequences of strains beneficial to plants of many legume tribes as well as strains considered non-nodulating, photosynthetic, or not isolated from plants. Here, findings suggested that the SI of *Bradyrhizobium* represents a diverse set of symICEs. We additionally demonstrate that recombination among symICEs and with nonsymbiosis ICEs as well as nonsymbiosis plasmids generates tremendous structural and functional diversity. Modularity of genes that contribute to SNF and their presence on mobile genetic elements are key to generating variation and conferring robustness to this ecologically important trait.

RESULTS

Symbiosis genes of *Bradyrhizobium* cluster within a strikingly diverse set of mobile genetic elements. We first identified clusters of symbiosis genes in *Bradyrhizobium* genome sequences and searched for hallmark features indicative of being within mobile genetic elements. A total of 179 strains have symbiosis genes clustered within regions with features of ICEs, here called symbiosis ICEs (symICEs), and two have them clustered within megaplasmids (Fig. 1). Both classes of genetic elements are associated with a T4SS-encoding locus that mediates interbacterial conjugation. In addition, the symICEs have integrase-associated genes, while the symbiosis megaplasmids (Sym plasmids) have a *repABC* origin of replication. The symICEs are also associated with tRNA genes and represent five types based on these associations and multiple subtypes based on differences in sequence signatures and gene composition (Fig. S1, S2A, S3; Data set S1). Analysis of replication genes suggest that the two Sym plasmids have independent origins (Fig. 1; Ext. Fig. S1 available at https://github.com/osuchanglab/BradyrhizobiumPangenomeManuscript/tree/main/Extended_Supplementary_Materials) (24). Twenty additional strains have *nif/fix* islands that lack features of ICEs or plasmids and two of these strains also have *nod* genes located in different regions originally classified as islands (25, 26). The remaining strains lack symbiosis genes and are incapable of SNF (Data set S1) (23).

Previous work used genomic comparisons, *att* site identification, and experimental transfer to characterize tripartite symICEs in *Mesorhizobium*, wherein sequential steps assemble the parts into a single circular element during recombination (13). Molecular

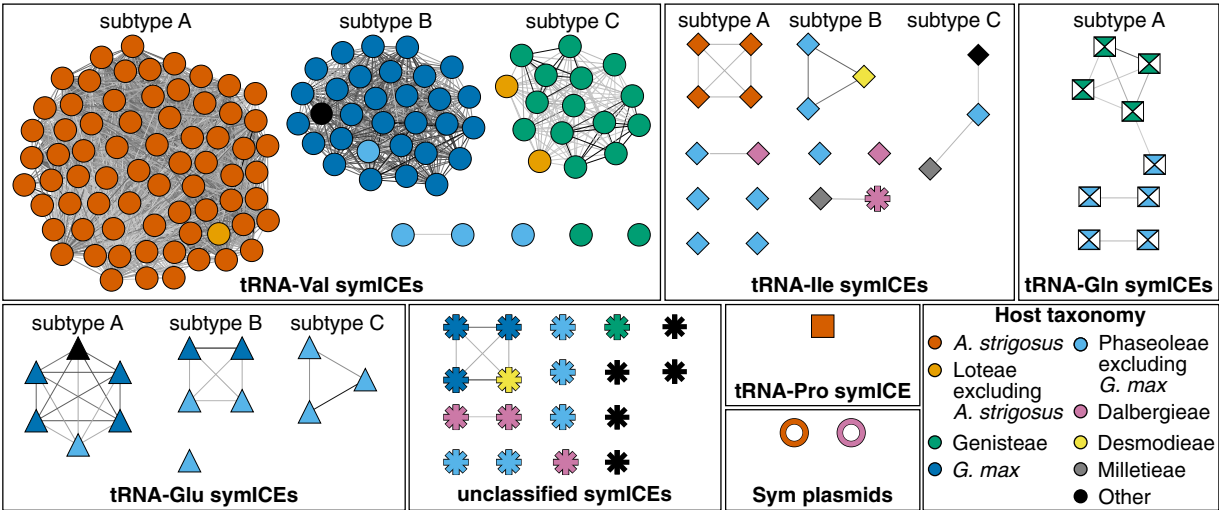


FIG 1 Symbiosis MGEs of *Bradyrhizobium* are diverse. (A) Weighted undirected network of symbiosis MGEs. Graphs are arranged according to type or class (shape) of symbiosis MGE. Graphs with more than two nodes were further classified into subtypes. Nodes represent individual symbiosis MGEs. Colors indicate taxonomic classification of associated host. Darker edges indicate greater Jaccard similarity of *k*-mer signatures.

evidence herein is consistent with symICEs of *Bradyrhizobium* having the capacity to mobilize and for some to induce large-scale rearrangements during recombination with chromosomes. The tRNA-Val (subtype A) symICE is the most common subtype. However, unlike those in *Mesorhizobium*, most of these adopt a bipartite configuration with a symICE element integrated near a tRNA-Val gene and a “B” element near the *ybgC* gene (Fig. 2). Each of the two elements has an integrase gene bordering a predicted *att* site and the other *att* site at the other boundary of the element; and recombination is predicted to occur in two steps, causing a chromosomal inversion around cognate *att* sites in each element (Fig. 2A; Fig. S1 and S4; Data set S1). This proposed mechanism of recombination is supported by alterations seen in four variants of tRNA-Val (subtype A). The tRNA-Val (subtype A) symICE of strain #195 is organized in a manner consistent with being in an intermediate configuration, as its two elements are adjacent to each other at one of the predicted *att* sites. Those of strain USDA6 and closely related E109 are predicted to be locked in the chromosome due to them having an *att* site transposed proximal to its partner *att* site, precluding its integrases from mediating recombination (Fig. 2B and C). Last is the symICE of strain number 186, which adopts a tripartite configuration (Fig. 2D). This tRNA-Val (subtype A) symICE acquired a third integrase and *att* site that resulted in *nif/fix* genes splitting into a third region adjacent to a tRNA-His gene. Similar to multipartite *Mesorhizobium* symICEs, this variant is predicted to require three steps to integrate and excise and cause two genome inversions (13). Four other types of symICEs are monopartite and likely integrate in one step at different tRNA genes.

Phylogenetic patterns of symbiosis MGEs are consistent with them having been mobilized horizontally across the genus (9, 10). Diverse species-level groups of *Bradyrhizobium* have similar symbiosis MGEs and are associated with the same host taxon while clades of closely related *Bradyrhizobium* vary in symbiosis MGEs and include ineffective and non-nodulating strains (Fig. 3; Fig. S3; Extended Data sets S1–S3). We examined conserved single copy genes in symbiosis MGEs to identify evidence for transmission events. Genes formed 14 groups based on similar evolutionary histories, reflecting substantial recombination and reshuffling (Fig. S5; Extended Fig. S2; Extended Data set S4). Nonetheless, symICE types and subtypes tended to cluster similarly across the 14 trees, each of which has a topology that differs from that of the genus tree, suggesting that symbiosis MGEs are largely inherited vertically and horizontally, as single units.

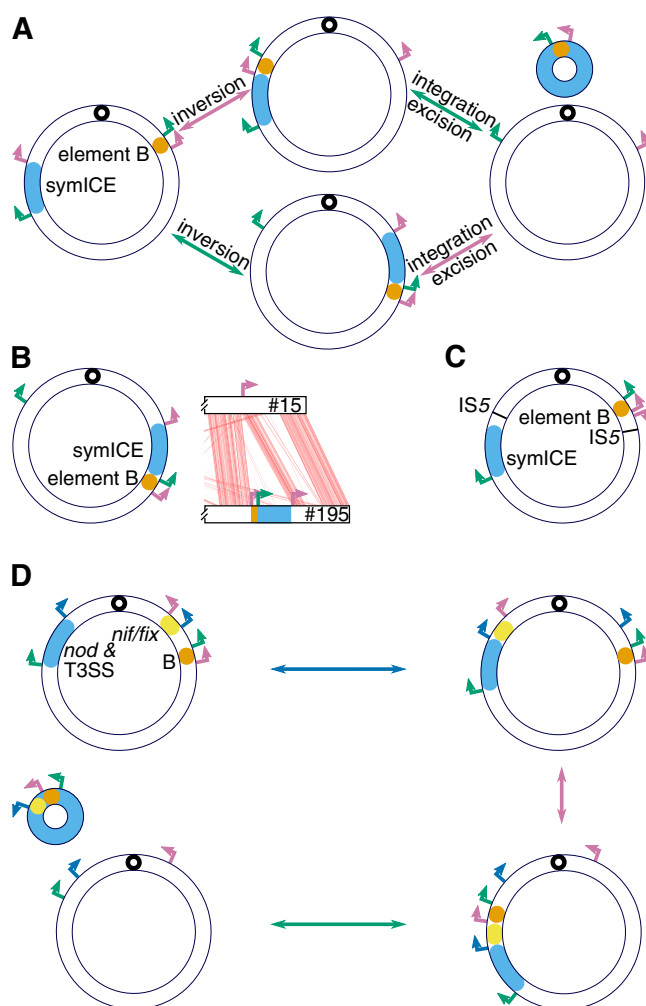


FIG 2 Recombination and organization of symICEs in chromosomes of *Bradyrhizobium* strains. (A) Model for recombination of the bipartite tRNA-Val symICE. Two recombination events are predicted to occur between *att* sites and result in the excision (left to right) or integration (right to left) of a symICE into a chromosome. Some recombination events are predicted to be associated with inversions to the genome. The order in which pair of *att* sites recombine determines whether recombination between the symICE and chromosome proceeds along the top or bottom path. (B) Organization of the bipartite tRNA-Val symICE in chromosome of strain number 195 (left is a model; right is an alignment of a portion of the chromosome to that of non-nodulating strain number 15; red lines link homologous regions). (C) Organization of the bipartite tRNA-Val symICE in the chromosome of strain USDA6. (D) One model for recombination (excision follows the clockwise path while integration follows the counterclockwise path) of the tripartite tRNA-Val symICE variant in strain number 186. Its *nif/fix* genes are split off to a region near a tRNA-His and are linked to an *intH* gene. In all panels, small black circles represent origins of replication. Single headed arrows are colored according to sequence and indicate locations and orientations of *att* sites. Double headed arrows indicate excision or integration, and are colored like the color of the recombining *att* sites.

Modularity of symbiosis MGEs enables diversification while preserving SNF.

Like other rhizobia, most genes implicated in SNF of *Bradyrhizobium* are encoded within MGEs (9, 10, 13, 14, 16). Thus, transfer of MGEs and recombination of symbiosis genes could be important to inform on mechanisms that diversify and impact SNF functions. We used principal-component analysis (PCA) to visualize the compositional variation of the symbiosis MGEs and showed that a large proportion is associated with differences in host plant species (Fig. 4A). Two additional observations of metapopulation strains provided critical insights. First, symbiosis MGEs of ineffective strains are each more similar in composition to that of a beneficial strain than to each other, suggesting that symbiosis MGEs of ineffective strains are derived from those of beneficial strains. Second, both tRNA-Val (subtype A) and tRNA-Ile (subtype A) symICEs can confer benefit to *A. strigosus*,

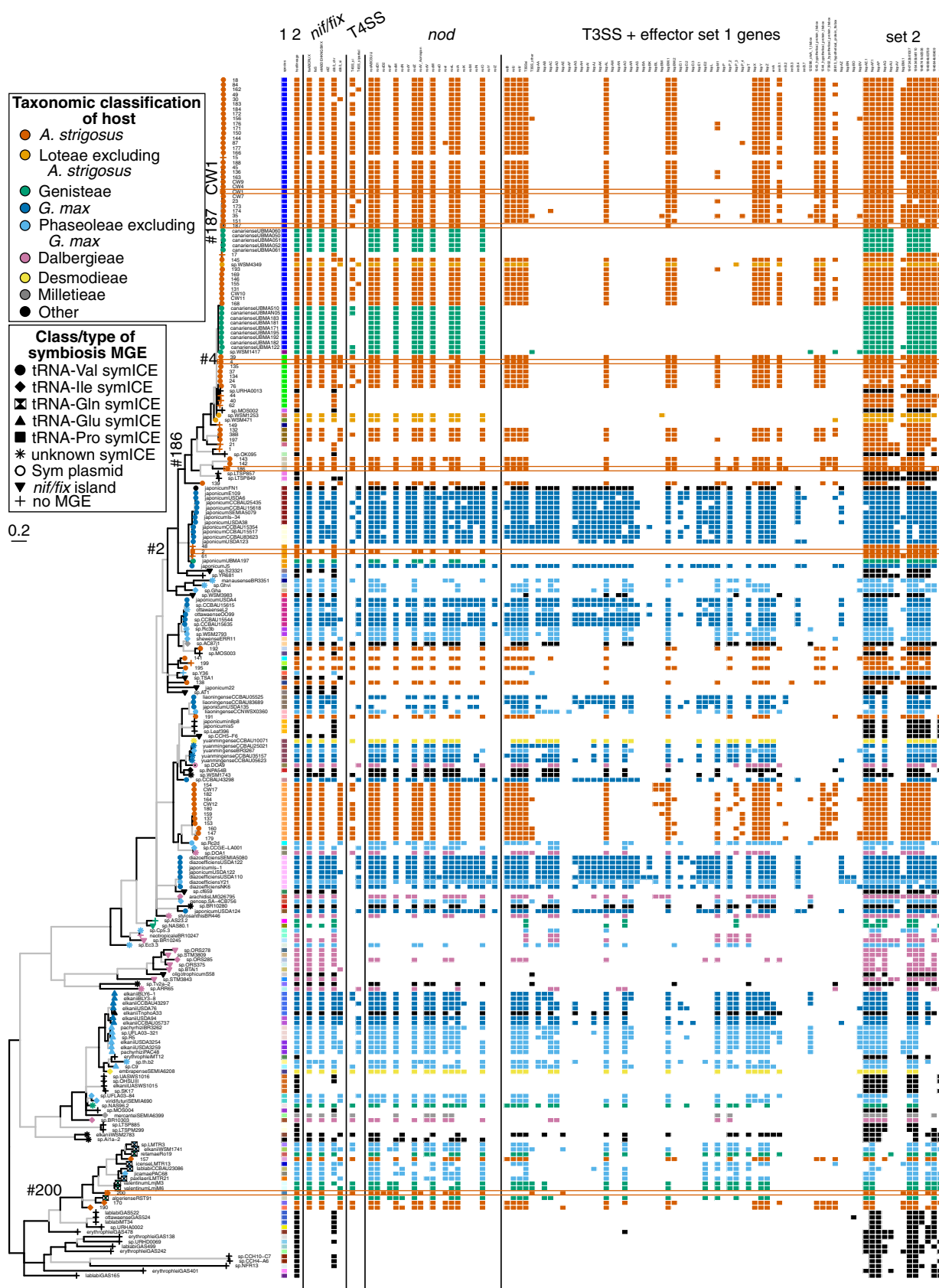


FIG 3 Diversity in combinations of *Bradyrhizobium* species and symbiosis GMEs. A multilocus sequence analysis (MLSA) maximum likelihood (ML) tree of *Bradyrhizobium* strains is shown on the left. The tree is midpoint rooted. Black colored branches exceed 70% bootstrap support. (Continued on next page)

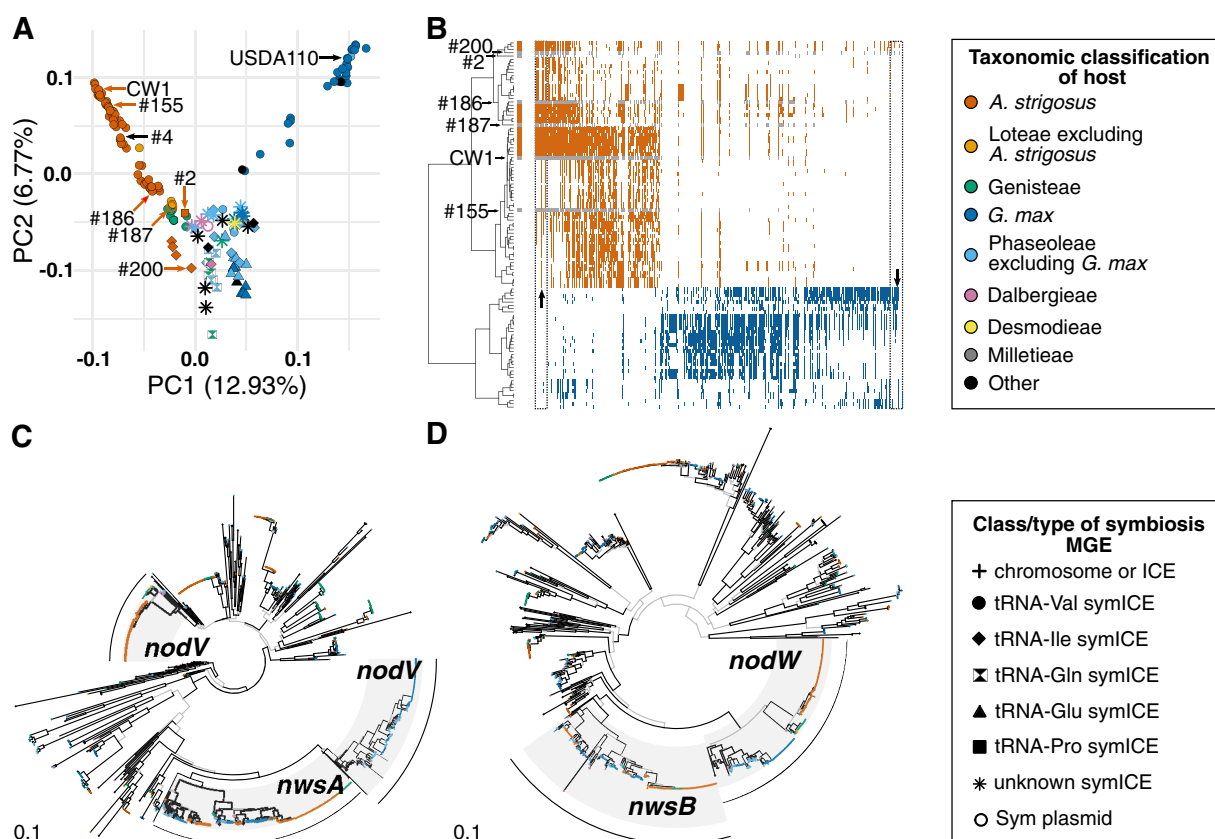


FIG 4 Symbiosis MGEs differ substantially in gene composition. (A) Principal-component analysis (PCA) of gene presence/absence of symbiosis MGEs. Each point represents a symbiosis MGE. Metapopulation strains that are ineffective on *A. strigosus* are labeled. Strains number 4 and USDA110 are shown as references. (B) A genome-wide association study was used to identify genes (columns) that are enriched in symbiosis MGEs (rows) of strains associated with *A. strigosus* or *G. max* (Data set S3). Shown are significant genes (Bonferroni corrected) and sorted based on enrichment in strains associated with *A. strigosus*. The two boxes outline regions with signature genes which were highly ranked, in $\geq 95\%$ of the strains of the group, and in $\leq 5\%$ of the comparator group. Gray colored rows correspond to ineffective strains. Ward's distance of gene presence/absence patterns was used to hierarchically cluster symbiosis MGEs. Ineffective metapopulation strains are indicated. (C) and (D) Deep maximum likelihood phylogenetic gene trees for homologs of *nodV* and *nodW*, respectively. Clades with *nodV*, *nwsA*, *nodW*, and *nwsB* are labeled. Branches colored in black exceed 95% UFBoot and 80% SH-aLRT support. Scale bars indicate average number of substitutions per site. Data and tip points are colored according to the taxonomic classification of plant hosts and shaped according to the location of the gene or class/type of symbiosis MGE.

suggesting that symbiosis genes have been shuffled across symICE types. Collectively, these observations indicate that modularity impacts both loss and gain of plant species as hosts.

To identify reshuffling events that are associated with ineffectiveness, we first grouped genes of the symbiosis MGEs into three categories, those representing the minimal set necessary for symbiotic nitrogen fixation, e.g., *nif/fix*, and *nodABC*, those core to strains beneficial to *A. strigosus*, and those identified in a genome-wide association study defined as signatures of being associated with *A. strigosus* relative to *Glycine max* (Data set S2). The genes within these categories were then used to query for differences in ineffective strains. Based on comparisons of genes essential for SNF present in beneficial *Bradyrhizobium*, the symbiosis MGEs of the six ineffective strains number

FIG 3 Legend (Continued)

Gray colored branches have 51% to 70% bootstrap support. Branches with $\leq 50\%$ bootstrap support were collapsed. The scale bar indicates average number of substitutions per site. Tip points are colored according to the taxonomic classification of plant hosts and shaped according to the class or type of symbiosis MGE. The next two columns show the species-level classification for strains (Extended Data set S2) and the associated class or type of symbiosis element. Remaining columns represent genes or functional gene clusters. A colored box indicates the presence in the corresponding symbiosis element. A white box indicates an absence. T3SS-associated effector genes are separated into two sets. Set 2 consists of predicted effector genes located in the chromosome, distal to symICEs.

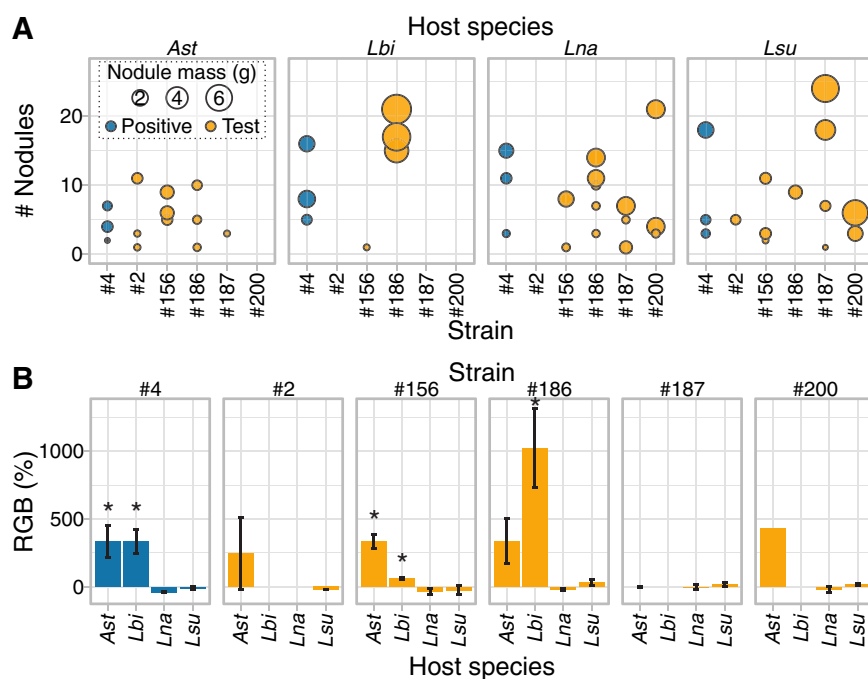


FIG 5 Strains ineffective on *A. strigosus* can nodulate and benefit other plant species. (A) Balloon plots relating number and total dry mass of nodules in four plant species infected with strains of *Bradyrhizobium*. Only plants with a minimum of two nodules were included in analyses. *A. wrangelianus* was tested as a partner species but did not form nodules and is not shown. (B) Relative growth benefit (RGB) provided by six strains of *Bradyrhizobium* on four species of plants. Relative growth is expressed as percent change in biomass compared to the biomass of plants from the corresponding host species treated with only water. A "*" indicates a significant relative change (P -value ≤ 0.05). Abbreviations used are: Ast = *A. strigosus*, Lbi = *L. bicolor*, Lsu = *L. succulentus*, and Lna = *L. nanus*. Measurements are provided (Extended Data set S6).

2, number 155, number 186, number 187, number 200, and CW1 are predicted to have genes necessary and sufficient for SNF but relative to those of beneficial metapopulation strains, have gene polymorphisms predicted to affect specificity of *A. strigosus* as a host species (Fig. 3; Fig. 4B; Fig. S2B; Fig. S6; Data sets S2–S3) (22). For example, the symbiosis MGEs of strains number 2, number 187, and number 200 lack symbiosis-associated T3SS and effector genes and several signature genes associated with being a symbiont of *A. strigosus*. Other differences in sequences or composition of *nod* genes and other genes in symbiosis MGEs were also identified (Fig. S6A; Fig. S7A–C, Data sets S2 and S3; Extended Data set S5) (27–29).

To test predictions that ineffective strains maintain SNF functions but benefit different plant species, we measured symbiosis outcomes on five sympatric legume species (Extended Data set S6). Four ineffective strains, marginally beneficial strain number 156, and beneficial strain number 4 can all nodulate multiple host species that they overlap in their native range (Fig. 5A). On *A. strigosus*, benefits derived from strain number 186 were not significant, consistent with its classification as ineffective (22). However, strain number 186 exhibited a qualitative shift in host range by inducing nodules on *Lupinus bicolor*. This strain caused significant growth benefits that were strikingly greater than those caused by strains number 4 and number 156, which also caused significant beneficial effects to *L. bicolor* (Fig. 5B). These data suggest that shuffling of symbiosis genes preserve SNF functions while changing host specificity, thus providing an explanation for the maintenance of variation in effectiveness in metapopulation strains.

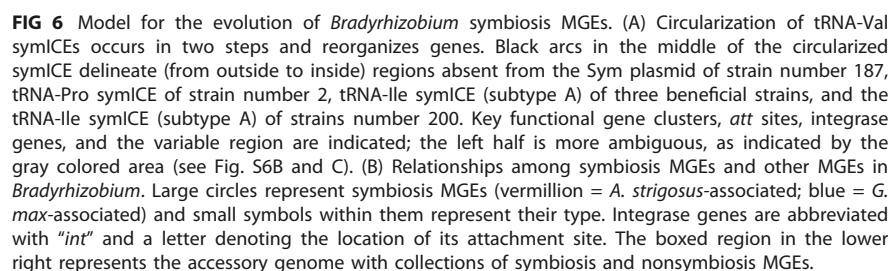
We also uncovered processes in the evolution of symbiosis MGEs predicted to have allowed metapopulation strains to gain *A. strigosus* as a host species and provide benefits. First, *nodV* was identified as a signature gene in the GWAS analysis. NodVW is a

two-component system that induces *nod* gene expression upon sensing specific plant isoflavones (30). In all sequenced strains, the *nodV* and *nodW* genes are adjacent to each other in the symICE. In addition, *nodVW* are close paralogs of *nwsAB* in strain USDA110, a symbiont of *G. max* and other hosts in the Phaseoleae tribe (30). The phylogeny of *nodV* is polyphyletic and incongruent with the phylogeny of *nodW*, supporting a scenario where a distantly related *nodV* homolog was acquired by an ancestral tRNA-Val symICE (Fig. 4C and D). We predict that it displaced the original and maintained the ability to function with the original *nodW*, thereby gaining a new capacity to perceive *A. strigosus* as a host species while retaining the ability to regulate *nod* gene expression.

Second, we predict that reshuffling of genes among symICEs led to strains gaining *A. strigosus* as a host species. Our analysis suggested a tRNA-Ile symICE of a symbiont associated with a host outside the Loteae tribe (i.e., *Acmispon*, *Lotus*, related genera) acquired a region from a tRNA-Val (subtype A) symICE, yielding the tRNA-Ile (subtype A) symICE with two *nod* clusters (Fig. S2B, Fig. S6A, S7). The acquired region includes most of the *nod* genes common to beneficial metapopulation strains, including *nodVW* with the *A. strigosus*-associated *nodV* allele. In beneficial strains, the region includes symbiosis-associated T3SS- and effector-encoding genes but in ineffective strain number 200, they are absent. The nonfunctional *nodD1* allele of strain number 200 also resides in this region. Across tRNA-Ile (subtype A) symICEs, the acquired region lacks a complete set of *nod* genes necessary for host-specific nodulation and normal nodule development (30). However, it is complemented by the other *nod* region, and three of four strains with a tRNA-Ile (subtype A) symICE can benefit *A. strigosus*. Thus, modularity allows genes to reconfigure at different levels of organization and diversify *Bradyrhizobium* strains without compromising essential functions of SNF.

Reorganization of symICE modules promotes diversification. The process by which symICEs excise causes important changes to gene organization that inform on mechanisms that led to the diversity of SNF uncovered herein (Fig. 6). Upon circularization, *Bradyrhizobium* symICEs reorganize SNF-associated genes into a small “Sym region,” including *nif/fix*, T3SS-encoding, and *nod* genes (13) (Fig. 6A; Fig. S6B–C). We predict that the reconfiguration, with SNF-associated genes more closely clustered and separated from a large variable region, is crucial for reshuffling across elements and generating the diversity observed in symbiosis MGEs among metapopulation strains. We suggest that the Sym region recombines with other elements, as the *Bradyrhizobium* genus is predicted to have an extremely large and diverse collection of nonsymbiosis associated MGEs (Fig. 1 and 4; Extended Fig. S3). For example, the Sym region from a tRNA-Val (subtype A) symICE likely recombined with a tRNA-Pro ICE and plasmid to yield the symbiosis MGEs in strains number 2 and number 187, respectively (Fig. 6B; Fig. S1, S2A, and S6A). Likewise, evidence strongly supports a scenario where the Sym region recombined with a tRNA-Ile symICE to yield the A subtype (Fig. 6B; Fig. S6A). Recombination of Sym regions also occurred among tRNA-Val symICEs and ICEs, generating members with highly similar symbiosis genes and diverged variable regions (Fig. S5 and S6B–C). Consistent with this, the variable regions of examined tRNA-Val symICEs have closely related T4SS-encoding loci flanked by large polymorphic regions, nested within closely related Sym regions (Fig. S2A and S6B–C). However, evidence also suggests that in other symbiosis MGEs, recombination can occur across more distantly related molecules and lead to the acquisition of more divergent T4SS-encoding loci and potentially, its flanking variable regions (Fig. S2A).

The model of symICE excision predicts that joining and circularization of bipartite tRNA-Val symICEs will link integrase genes together with *att* sites and the small B element. This subsequent predicted reorganization is critical in facilitating losses of this region, whose absence is necessary to transition one type of symICE to another that integrates at a different site, or to a Sym plasmid that does not integrate into the chromosome. The Sym plasmid of strain number 187 has a scar, one *att* sequence of a tRNA-Val symICE, supporting the proposed loss. Furthermore, loss of one or both



However, organizational differences can also restrict recombination between symICEs. Subtypes A and B of the tRNA-Val symICEs, associated with symbiosis on Loteae and Phaseoleae hosts, respectively, are derived from a common ancestor, but there is little evidence for them recombining (Fig. 4A; Fig. S1A–B and S5). When tRNA-Val (subtype B) symICEs circularize, *nod* and T35S-encoding genes will separate to opposite regions flanking the B region, as opposed to the same region predicted for subtype A. We suggest this difference reduces productive recombination events between these two subtypes (Fig. S6A).

DISCUSSION

Here, we showed that modularity and reshuffling of genes by mobile genetic elements generate uncooperative genotypes of rhizobia and make individual partnerships unstable, but these same properties are fundamental for robustness and extending beneficial associations to diverse host species, as well as transferring symbiotic capacity among diverse rhizobia. We sampled native bacteria from an 800 km transect of wild *Acmispon strigosus* populations and included strains various in symbiosis phenotypes. We applied a strategy, developed for virulence plasmids, to study symbiosis ICEs and plasmids in *Bradyrhizobium*, and in doing so we were able to infer relationships among many symbiosis MGEs and model interactions among MGEs that led to their diversification (31). Findings suggested that symICE circularization reorganizes genes and is predicted to promote the shuffling of blocks of SNF genes into different ICE backbones, the generation of new combinations of modules, the acquisition of genes from the chromosome onto the symICE, and transfer of the SNF trait to plasmids. In this regard, despite being integrated in chromosomes for most of their life cycle, symICEs are like plasmids in promoting more rapid evolution of their cargo genes (32). The symICEs described here are incredibly diverse, and even metapopulation strains isolated from a single plant host species and sampled in one US state can have different symICE types, subtypes, or symbiosis plasmids. No two symICEs isolated from metapopulation strains or in this data set have identical gene content or sequence. This reflects multiple scales of modularity. Plants can be host to diverse species of *Bradyrhizobium*, strains can host diverse and exchangeable symbiosis MGEs, which themselves exchange and acquire genes, and individual genes within functional units can vary.

The patterns that we uncovered in the metapopulation strains, including the role of MGEs in gene reshuffling and diversification, extend broadly across the *Bradyrhizobium* genus. For example, most analyzed photosynthetic nitrogen fixing *Bradyrhizobium* strains have only *nif/fix* islands and are restricted in host range (25, 33). But two strains acquired symICEs that expanded their host ranges. Notably, strain ORS285 has an island that we suggest is a remnant of a symICE located at tRNA-Ile that includes *nod*, T3SS-associated genes, and an integrase gene but no *nif/fix* genes or conjugation machinery-encoding locus (26) (Fig. S1B and S2B). Moreover, several variants of symICEs have also recurrently gained *fixNOQP* and *fixGHIS*, genes necessary for respiration in microoxic root nodules and typically in *fix* gene cluster III located in the chromosome (34) (Extended Fig. S4). Acquisition of these symICE variants extended SNF to strains that lack *fix* gene cluster III and would not otherwise be capable of supporting SNF. Thus, acquisition of MGEs, and the traits that they encode, appear to play a major role in the adaptation of *Bradyrhizobium* to novel lifestyles.

We propose that diversification in the *Mesorhizobium* genus, which exhibits many parallel patterns, is also driven by acquisition and reshuffling of MGEs (19). Importantly, key aspects of our study differ from those of *Mesorhizobium*, where diversification has been primarily characterized in managed systems. In *Mesorhizobium*, entire symICEs originating from inoculum strains were predicted to be transferred into indigenous nonsymbiotic rhizobia, or among strains already naturalized under monoculture crops, a scenario that imposes intense selection for host crop compatibility (14, 16, 17, 19). Conversely, our findings are based on investigation of phenotypically variable strains of *Bradyrhizobium* isolated from diverse native plant communities, where multiple legume species overlap and select for differential subsets of rhizobia (35–37). Our study suggested that symICE transfer has recurrently promoted novel host acquisition, and that loss of effectiveness on one host is associated with gains of other hosts, processes that likely require a diverse array of potential hosts.

Reconceiving symbiosis as a dynamical system with links that can form and dissolve among symbionts and between symbionts and hosts is essential for revealing emergent properties. Modularity and flexibility of genetic elements, coupled to their mobility, drive diversification, giving rise to variation in symbiosis, such as that revealed in *Bradyrhizobium*. Modularity and flexibility are central to robustness (3). A fundamental

principle of robustness is that it maintains the function of a system, not a state (38). Therein lies the source of the paradox where symbiosis functions are maintained at the system level but lost from individual states, such as a symbiont, partnership, or host (2). Models that reduce symbiosis to bipartite partnerships and ignore symbiont-symbiont interactions unknowingly neglect major sources of variation and overlook robustness (2, 39). Additionally, by separating symbiosis into categorical partnerships, these models fail to recognize the effects of multiple and various symbiont-host interactions within the system. Host species select for different combinations of symbiosis genes in their bacterial partners. Pangenome evolution, shaped by individual and gene-level selection, reassorts genes into new combinations that can extend symbiosis to new host species. Hence, alignment of fitness interests between host and symbiont is necessary for persistence of a partnership while interactions diverse in partners are necessary for robustness and evolutionary stability of symbiosis (3).

This alternative framework provides a predictive understanding of symbiosis functions that are encoded on MGEs. SNF is evolutionarily stable despite repeated abandonment by both symbiont and host species (22, 40). Conversely, symbioses involving vertically transmitted endosymbiotic bacteria with closed pangenomes are not as robust and are at higher risks of extinction (41). In agriculture, elite rhizobia strains are often added to monocultures in attempt to establish a highly specific and optimal partnership. Success is difficult to achieve because the system is flexible, and plants can partner with different genotypes of rhizobia (9). Even if the optimal partnership is attained, the likelihood for it to persist is low because of potential trade-offs between state optimality and system robustness (42). Strategies that promote interactions between multiple lineages of beneficial nitrogen-fixing rhizobia and diverse crops will have greater success for long-term sustainability.

MATERIALS AND METHODS

Genome sequencing, assembly, and annotation. Strains were selected from a metapopulation that was previously generated and phenotyped (22, 23) (Data set S1). Bacteria were grown overnight in a modified arabinose-gluconate medium at 29°C with shaking (43). The Qiagen DNeasy blood and tissue kit (Qiagen, Valencia, CA, USA) was used to extract total genomic DNA, and according to methods previously described, prepared, multiplexed, and sequenced on one channel of an Illumina HiSeq 3000 (Illumina Inc., San Diego, CA, USA) by the Center for Genome Research and Biocomputing (CGRB; Oregon State University, Corvallis, OR, USA) to generate 150mer paired end sequencing reads (31).

For long read sequencing, strains were either sequenced on two flow cells of an Oxford Nanopore GridION X5 by De Novo Genomics Corporation (Kansas City, KS, USA) or on a MinION flowcell (1D chemistry, LSK109) on a Mk1b MinION sequencer controlled by a MinIT coprocessor. In both cases, samples were multiplexed and prepared using a pooled 1D Native genomic DNA library prep (SQK-LSK109) prior to sequencing (44). Previously described methods were used to process and assemble sequencing reads and annotate genome assemblies (31).

BLASTN v. 2.6.0 searches were used, with “TTACACGGGAGAGGTCCAAGGTTTCGATCCCTTGTGCGCCC ACCATTCACCT” and “TCCGTATCTTGAATAGACGCGGACCTGCATATGATGGTGACCGCCGCAATTCGCCA TCAAGAGAAGCTGTGACG” as queries and the parameters “-word_size 7 -evalue 100” to identify attachment sequences of *Int_v* and *Int_h*, respectively (45). HMMER v 3.3 *hmmsearch* and custom *hmm* profiles were used to annotate putative *nod*-box and *tts*-box sequences (46–48). Macsfinder 1.0.5 with the default options and the *hmm* profiles TXSS and CONJ were used to annotate secretion system-encoding loci (49). BLASTP and translated sequences of type III effector genes from rhizobia were used as queries to identify homologs of effector genes (48, 50).

Analyses of genome sequences. Previously described methods were used to calculate percentage of conserved proteins (POCP; 50% threshold) and pairwise average nucleotide identity (ANI; threshold ≥ 95%) to operationally classify strains into genus- and species-level groups, respectively (31). The software package *get_homologues* v. 20170418 with the options “-M -t 0” was used to cluster genes into orthologous groups (51). Gene presence/absence heatmaps were generated using the R package *heatmap.plus* with “complete” clustering of binary distances, or with the *ggtree* function *gheatmap* (52, 53). Publicly available genome sequences were downloaded from NCBI on May 7, 2018.

BWA-MEM, Picard tools, and GATK HaplotypeCaller, following previously described methods, were used to call single nucleotide polymorphisms (SNPs) (31, 54, 55).

Panaroo v. 1.2.3 with the option “—clean-mode sensitive” was used to generate a pangenome graph for select genomes, which was visualized using Cytoscape v. 3.8.0 (56, 57).

Islander v. 1.2 with the options “-translate -trna -annotate -reisland -table 11 -nocheck” was used to identify tRNA-associated ICEs in finished or hybrid assembled genome sequences (58). Islander v. 1.2 was modified by increasing maximum island size thresholds to 2 Mb to search for large symICEs.

Predicted regions, conserved and syntenic across multiple strains and spanning large clusters of tRNA loci, were identified manually and filtered out.

Plasmids were identified from finished or hybrid assembled genome sequences as separate replicons or from draft genome sequences as contigs with a *repABC* locus.

Analysis of symICE and symbiosis genes. The symICEs in finished or hybrid assembled genome sequences were identified based on presence of *nod*, T3SS-associated, and *nif/fix* genes. The program progressiveMauve was used to identify symICE boundaries by comparing genome sequences with those from closely related strains that lack the same type of symICE types (59). The symICEs were classified into types based on concordance with identified ICE regions, location of tRNA genes, and the presence of integrase genes next to or overlapping border sequences. Repetitive sequences at border regions, including putative *att* sites were also identified. Sequences corresponding to symICEs of draft genome sequences were identified by using CONTIGuator v. 2.7, with default parameters, to map contigs to complete genome sequences with most similar *nif/fix* genes (60). Contigs and regions mapping within the symICE region of the reference genome were extracted.

A symICE gene presence/absence matrix, representing the diversity of gene content of symICEs, was inferred from the ortholog clusters of the full genome get_homologues analysis. The ortholog group for each symICE-associated gene was identified from the larger analysis and using the “spread” function of the R package tidyR, a new presence/absence matrix containing only symICE genes and paralogs was generated (61). These new gene cluster sequences were subset in fasta format and used in phylogenetic and topology clustering analysis. The base R function “prcomp” was used to perform a principal-component analysis (PCA) of symICE gene content (62). Scoary v. 1.6.16 with the options “-s 4 -collapse -e 100” was used to perform a genome wide association study (GWAS) analysis of symICE gene content, comparing symICEs of strains isolated from *A. strigosus* and *G. max* (63). LAST v. 1066 lastal with the option “-f BlastTab+” was used to compare symICEs, and homologous regions were visualized using the BioPython package GenomeDiagram v. 1.72 (64, 65). A previously described method was used to generate and visualize gene synteny networks for select symICEs (31).

Sourmash v. 2.0.0a11 with the option “compute -scaled 100” and “compare -k 21” was used to estimate a Jaccard Index between symbiosis MGEs based on their *k*-mer signatures (66). These values were used to build graphs where symbiosis MGEs are nodes, and edges connect them in which at least one symbiosis MGE has a Jaccard Index ≥ 0.1 to the other. Cytoscape v. 3.8.0 was used to visualize graphs (56).

Construction of phylogenetic trees. The translated sequences of *dnaG*, *frr*, *infC*, *nusA*, *pgk*, *pyrG*, *rplA*, *rplB*, *rplC*, *rplE*, *rplF*, *rplK*, *rplL*, *rplM*, *rplN*, *rplP*, *rplS*, *rplT*, *rpmA*, *rpoB*, *rpsB*, *rpsC*, *rpsE*, *rpsI*, *rpsJ*, *rpsK*, *rpsM*, *rpsS*, *smpB*, and *tsf* were used to construct a multilocus sequence analysis (MLSA) phylogeny (67, 68). Previously described methods were followed to construct phylogenetic trees and for phylogenetic topology clustering analysis (31). IQ-TREE v. 1.6.12 with the options “-bb 1000 -alrt 1000” was used to generate phylogenies for some data sets (69). Cophylo plots were generated using the R package phytools (70).

The top 100 hits of a BLASTP search, using SctU as a query against the NCBI nr database, was retrieved on March 4, 2019. BLASTP with the default parameters was used to identify homologs of *sctU*, *nodV*, *nodW*, *repC*, and *trbE* in the genomes of analyzed strains. Data sets for integrases and other gene/protein phylogenies were comprised of ortholog groups from the get_homologues analysis.

Plant inoculations. *Bradyrhizobium* strains number 2, number 4, number 156, number 186, number 187, and number 200 were each inoculated on to five sympatric host species, including *A. strigosus* AcS049, *A. wrangelianus* AcW10-R5, and *L. bicolor*, *L. succulentus*, *L. nanus*. Inbred lines were used for *Acmispon* whereas mixed seed sets of *Lupinus* were used (S & S Seeds, Carpinteria, CA). Seeds were surface sterilized in 5.0% sodium hypochlorite, rinsed in sterile water, and scarified. For *A. wrangelianus*, seeds were vernalized for a week at 4°C before planting (71). For others, immediately after scarification, seeds were planted in sterilized containers (Steuwe and Sons, Corvallis, OR), each filled with autoclaved Turface Pro League soil mixed 1:1 with small- and coarse-grain sand. The seedlings were germinated in a growth facility and when true leaves appeared, moved to a greenhouse. Plants were hardened to greenhouse conditions for 1 week before inoculation.

Previously published protocols were followed to inoculate plants with *Bradyrhizobium* strains (43). A minimum of six plant replicates was used for each host-strain combination. Negative-control plants received sterilized reverse-osmosis H₂O only. Immediately before inoculation, plants were arranged in a randomized design blocking by plant size, determined based on the number of true leaves. Plants were fertilized weekly with 5.0 mL nitrogen-free Jenson’s fertilizer. After 7 weeks postinoculation, plants were depotted, rinsed of remaining soil, and photographed. We counted and weighed nodules and measured separately dry biomasses of shoots and roots. Tissues were separated and dried at 60°C for at least 72 h prior to weighing.

The effect of strain inoculation on plants was calculated as percent relative growth benefit (RGB), the mean percent biomass of an inoculated plant relative to its corresponding control plants:

$$\text{RGB} = 100 * (\text{Inoculated host biomass} - \text{control host biomass}) / \text{control host biomass}.$$

Linear models were used to investigate variation in nodulation (total nodule number, dry nodule biomass) and RGB for effects of strain, host, and interaction effects. We used the same model for all tests to improve normality of the residuals and homoscedasticity. Significant differences among strains or hosts were assessed using Tukey’s HSD test. To test if RGB was significantly larger than zero, we performed a series of one sample unpaired t-tests. All statistical analyses were carried out using R version 3.6.0 (62).

Data and materials availability. Short reads and assemblies have been deposited in NCBI as BioProject PRJNA671608 and accession numbers are listed in Data set S1. Network graphs in nexus or sif format,

phylogenetic trees in Newick format, genome annotations, and scripts can be downloaded from <https://github.com/osuchanglab/BradyrhizobiumPangenomeManuscript>. Strains sequenced in this study are available from JLS upon request. Extended supplementary figures and data sets can be downloaded from https://github.com/osuchanglab/BradyrhizobiumPangenomeManuscript/tree/main/Extended_Supplementary_Materials.

SUPPLEMENTAL MATERIAL

Supplemental material is available online only.

DATA SET S1, XLSX file, 0.03 MB.

DATA SET S2, XLSX file, 0.02 MB.

DATA SET S3, XLSX file, 0.1 MB.

FIG S1, PDF file, 0.1 MB.

FIG S2, PDF file, 0.1 MB.

FIG S3, PDF file, 0.4 MB.

FIG S4, PDF file, 0.1 MB.

FIG S5, PDF file, 0.3 MB.

FIG S6, PDF file, 0.3 MB.

FIG S7, PDF file, 0.1 MB.

ACKNOWLEDGMENTS

We thank those that made whole-genome sequences of *Bradyrhizobium* publicly available. We thank Drs. Lorena Torres-Martínez and David Fronk as well as other members of the Chang and Sachs labs for their assistance, insightful comments, and critical reading of the manuscript. We acknowledge the staff of the CGRB for their services and thank the Department of Botany and Plant Pathology at OSU for its support of the computing infrastructure. This work was supported by the National Science Foundation under grant numbers DEB-1738028 to JHC and DEB-1150278 and DEB-1738009 to JLS, respectively. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

A.J.W., A.R., J.H.C., and J.L.S. conceptualized and designed the experiments. A.J.W., A.R., D.B., and P.T. performed the research. A.J.W., A.R., D.B., P.T., J.H.C., and J.L.S. analyzed the data. A.J.W., A.R., J.H.C., and J.L.S. wrote the paper. J.H.C. and J.L.S. acquired funding.

We have no conflicts of interest to declare.

REFERENCES

1. Heath KD, Stinchcombe JR. 2014. Explaining mutualism variation: a new evolutionary paradox? *Evolution* 68:309–317. <https://doi.org/10.1111/evo.12292>.
2. Sachs JL, Skophammer RG, Regus JU. 2011. Evolutionary transitions in bacterial symbiosis. *Proc Natl Acad Sci U S A* 108:10800–10807. <https://doi.org/10.1073/pnas.1100304108>.
3. Kitano H. 2004. Biological robustness. *Nat Rev Genet* 5:826–837. <https://doi.org/10.1038/nrg1471>.
4. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS, DeBoy RT, Davidsen TM, Mora M, Scarselli M, Ros IMY, Peterson JD, Hauser CR, Sundaram JP, Nelson WC, Madupu R, Brinkac LM, Dodson RJ, Rosovitz MJ, Sullivan SA, Daugherty SC, Haft DH, Selengut J, Gwinn ML, Zhou LW, Zafar N, Khouri H, Radune D, Dimitrov G, Watkins K, O'Connor KJB, Smith S, Utterback TR, White O, Rubens CE, Grandi G, Madoff LC, Kasper DL, Telford JL, Wessels MR, Rappuoli R, Fraser CM. 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial “pan-genome”. *Proc Natl Acad Sci U S A* 102:13950–13955. <https://doi.org/10.1073/pnas.0506758102>.
5. Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. *Nat Rev Genet* 8:921–931. <https://doi.org/10.1038/nrg2267>.
6. Mutreja A, Kim DW, Thomson NR, Connor TR, Lee JH, Kariuki S, Croucher NJ, Choi SY, Harris SR, Lebens M, Niyogi SK, Kim EJ, Ramamurthy T, Chun J, Wood JLN, Clemens JD, Czerkinsky C, Nair GB, Holmgren J, Parkhill J, Dougan G. 2011. Evidence for several waves of global transmission in the seventh cholera pandemic. *Nature* 477:462–465. <https://doi.org/10.1038/nature10392>.
7. The HC, Thanh DP, Holt KE, Thomson NR, Baker S. 2016. The genomic signatures of *Shigella* evolution, adaptation and geographical spread. *Nat Rev Microbiol* 14:235–250. <https://doi.org/10.1038/nrmicro.2016.10>.
8. Cury J, Touchon M, Rocha Eduardo PC. 2017. Integrative and conjugative elements and their hosts: composition, distribution and organization. *Nucleic Acids Res* 45:8943–8956. <https://doi.org/10.1093/nar/gkx607>.
9. Masson-Boivin C, Sachs JL. 2018. Symbiotic nitrogen fixation by rhizobia - the roots of a success story. *Curr Opin Plant Biol* 44:7–15. <https://doi.org/10.1016/j.pbi.2017.12.001>.
10. Poole P, Ramachandran V, Terpolilli J. 2018. Rhizobia: from saprophytes to endosymbionts. *Nat Rev Microbiol* 16:291–303. <https://doi.org/10.1038/nrmicro.2017.171>.
11. Deakin WJ, Broughton WJ. 2009. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat Rev Microbiol* 7:312–320. <https://doi.org/10.1038/nrmicro2091>.
12. Jones JDG, Dangl JL. 2006. The plant immune system. *Nature* 444:323–329. <https://doi.org/10.1038/nature05286>.
13. Haskett TL, Terpolilli JJ, Bekuma A, O'Hara GW, Sullivan JT, Wang P, Ronson CW, Ramsay JP. 2016. Assembly and transfer of tripartite integrative and conjugative genetic elements. *Proc Natl Acad Sci U S A* 113:12268–12273. <https://doi.org/10.1073/pnas.1613358113>.
14. Sullivan JT, Ronson CW. 1998. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. *Proc Natl Acad Sci U S A* 95:5145–5149. <https://doi.org/10.1073/pnas.95.9.5145>.
15. Elias NV, Herridge DF. 2015. Naturalised populations of mesorhizobia in chickpea (*Cicer arietinum* L.) cropping soils: effects on nodule occupancy

- and productivity of commercial chickpea. *Plant Soil* 387:233–249. <https://doi.org/10.1007/s11104-014-2298-z>.
16. Hill Y, Colombi E, Bonello E, Haskett T, Ramsay J, O'Hara G, Terpolilli J, Semrau JD. 2021. Evolution of diverse effective N₂-fixing microsymbionts of *Cicer arietinum* following horizontal transfer of the *Mesorhizobium ciceri* CC1192 symbiosis integrative and conjugative element. *Appl Environ Microbiol* 87:e02558-20. <https://doi.org/10.1128/AEM.02558-20>.
 17. Nandasena KG, O'Hara GW, Tiwari RP, Howieson JG. 2006. Rapid in situ evolution of nodulating strains for *Biserrula pelecinus* L. through lateral transfer of a symbiosis island from the original mesorhizobial inoculant. *Appl Environ Microbiol* 72:7365–7367. <https://doi.org/10.1128/AEM.00889-06>.
 18. Colombi E, Perry BJ, Sullivan JT, Bekuma AA, Terpolilli JJ, Ronson CW, Ramsay JP. 2021. Comparative analysis of integrative and conjugative mobile genetic elements in the genus *Mesorhizobium*. *Microb Genom* 7.
 19. Greenlon A, Chang PL, Damte ZM, Muleta A, Carrasquilla-Garcia N, Kim D, Nguyen HP, Suryawanshi V, Krieg CP, Yadav SK, Patel JS, Mukherjee A, Udupa S, Benjelloun I, Thami-Alami I, Yasin M, Patil B, Singh S, Sarma BK, von Wettberg EJB, Kahraman A, Bukun B, Assefa F, Tesfaye K, Fikre A, Cook DR. 2019. Global-level population genomics reveals differential effects of geography and phylogeny on horizontal gene transfer in soil bacteria. *Proc Natl Acad Sci U S A* 116:15200–15209. <https://doi.org/10.1073/pnas.1900056116>.
 20. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, Watanabe A, Idesawa K, Iriguchi M, Kawashima K, Kohara M, Matsumoto M, Shimpo S, Tsuruoka H, Wada T, Yamada M, Tabata S. 2002. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 9:189–197. <https://doi.org/10.1093/dnares/9.6.189>.
 21. Parker MA. 2015. The spread of *Bradyrhizobium* lineages across host legume clades: from *Abarema* to *Zygia*. *Microb Ecol* 69:630–640. <https://doi.org/10.1007/s00248-014-0503-5>.
 22. Gano-Cohen KA, Wendlandt CE, Al Moussawi K, Stokes PJ, Quides KW, Weisberg AJ, Chang JH, Sachs JL. 2020. Recurrent mutualism breakdown events in a legume rhizobia metapopulation. *Proceedings of the Royal Society B-Biological Sciences* 287. <https://doi.org/10.1098/rspb.2019.2549>.
 23. Hollowell AC, Regus JU, Turissini D, Gano-Cohen KA, Bantay R, Bernardo A, Moore D, Pham J, Sachs JL. 2016. Metapopulation dominance and genomic-island acquisition of *Bradyrhizobium* with superior catabolic capabilities. *Proceedings of the Royal Society B-Biological Sciences* 283. <https://doi.org/10.1098/rspb.2016.0496>.
 24. Okazaki S, Noisangiam R, Okubo T, Kaneko T, Oshima K, Hattori M, Teamtisong K, Songwattana P, Tittabutr P, Boonkerd N, Saeki K, Sato S, Uchiumi T, Minamisawa K, Teamroong N. 2015. Genome analysis of a novel *Bradyrhizobium* sp DOA9 carrying a symbiotic plasmid. *PLoS One* 10:e0117392. <https://doi.org/10.1371/journal.pone.0117392>.
 25. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre J-C, Jaubert M, Simon D, Cartieaux F, Prin Y, Bena G, Hannibal L, Fardoux J, Kojadinovic M, Vuillet L, Lajus A, Cruveiller S, Rouy Z, Mangenot S, Segurens B, Dossat C, Franck WL, Chang W-S, Saunders E, Bruce D, Richardson P, Normand P, Dreyfus B, Pignol D, Stacey G, Emerich D, Vermiglio A, Medigue C, Sadowsky M. 2007. Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. *Science* 316:1307–1312. <https://doi.org/10.1126/science.1139548>.
 26. Gully D, Teulet A, Busset N, Nouwen N, Fardoux J, Rouy Z, Vallenet D, Cruveiller S, Giraud E. 2017. Complete genome sequence of *Bradyrhizobium* sp. ORS285, a photosynthetic strain able to establish nod factor-dependent or nod factor-independent symbiosis with *Aeschynomene* legumes. *Microbiology Resource Announcements* 5.
 27. Demont N, Debelle F, Aurelle H, Denarie J, Prome JC. 1993. Role of the *Rhizobium meliloti* *nodF* and *nodE* genes in the biosynthesis of lipo-oligosaccharidic nodulation factors. *J Biological Chemistry* 268:20134–20142. [https://doi.org/10.1016/S0021-9258\(20\)80704-2](https://doi.org/10.1016/S0021-9258(20)80704-2).
 28. Kossak RM, Bookland R, Barkei J, Paaren HE, Appelbaum ER. 1987. Induction of *Bradyrhizobium japonicum* common nod genes by isoflavones isolated from *Glycine max*. *Proc Natl Acad Sci U S A* 84:7428–7432. <https://doi.org/10.1073/pnas.84.21.7428>.
 29. Lamrabet Y, Bellogin RA, Cubo T, Espuny R, Gil A, Krishnan HB, Megias M, Ollero FJ, Pueppke SG, Ruiz-Sainz JE, Spaink HP, Tejero-Mateo P, Thomas-Oates J, Vinardell JM. 1999. Mutation in GDP-fucose synthesis genes of *Sinorhizobium fredii* alters nod factors and significantly decreases competitiveness to nodulate soybeans. *Mol Plant Microbe Interact* 12:207–217. <https://doi.org/10.1094/MPMI.1999.12.3.207>.
 30. Loh J, Stacey G. 2003. Nodulation gene regulation in *Bradyrhizobium japonicum*: a unique integration of global regulatory circuits. *Appl Environ Microbiol* 69:10–17. <https://doi.org/10.1128/AEM.69.1.10-17.2003>.
 31. Weisberg AJ, Davis EW, Tabima J, Belcher MS, Miller M, Kuo C-H, Loper JE, Grünwald NJ, Putnam ML, Chang JH. 2020. Unexpected conservation and global transmission of agrobacterial virulence plasmids. *Science* 368:eaba5256. <https://doi.org/10.1126/science.aba5256>.
 32. Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. 2021. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. *Nat Rev Microbiol* 19:347–359. <https://doi.org/10.1038/s41579-020-00497-1>.
 33. Molouba F, Lorquin J, Willems A, Hoste B, Giraud E, Dreyfus B, Gillis M, de Lajudie P, Masson-Boivin C. 1999. Photosynthetic bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group. *Appl Environ Microbiol* 65:3084–3094. <https://doi.org/10.1128/AEM.65.7.3084-3094.1999>.
 34. Preisig O, Anthamatten D, Hennecke H. 1993. Genes for a microaerobically induced oxidase complex in *Bradyrhizobium japonicum* are essential for a nitrogen-fixing endosymbiosis. *Proc Natl Acad Sci U S A* 90:3309–3313. <https://doi.org/10.1073/pnas.90.8.3309>.
 35. Torres-Martínez L, Porter SS, Wendlandt C, Purcell J, Ortiz-Barbosa G, Rothschild J, Lampe M, Warisha F, Le T, Weisberg AJ, Chang JH, Sachs JL. 2021. Evolution of specialization in a plant-microbial mutualism is explained by the oscillation theory of speciation. *Evolution* 75:1070–1086. <https://doi.org/10.1111/evo.14222>.
 36. Ehinger M, Mohr TJ, Starcevich JB, Sachs JL, Porter SS, Simms EL. 2014. Specialization-generalization trade-off in a *Bradyrhizobium* symbiosis with wild legume hosts. *BMC Ecol* 14:8. <https://doi.org/10.1186/1472-6785-14-8>.
 37. Pahua VJ, Stokes PJN, Hollowell AC, Regus JU, Gano-Cohen KA, Wendlandt CE, Quides KW, Lyu JY, Sachs JL. 2018. Fitness variation among host species and the paradox of ineffective rhizobia. *J Evol Biol* 31:599–610. <https://doi.org/10.1111/jeb.13249>.
 38. Kitano H. 2007. Towards a theory of biological robustness. *Mol Syst Biol* 3:137. <https://doi.org/10.1038/msb4100179>.
 39. Sachs JL, Simms EL. 2006. Pathways to mutualism breakdown. *Trends Ecol Evol* 21:585–592. <https://doi.org/10.1016/j.tree.2006.06.018>.
 40. Griesmann M, Chang Y, Liu X, Song Y, Haberer G, Crook MB, Billault-Penneteau B, Lauressergues D, Keller J, Imanishi L, Roswanjaya YP, Kohlen W, Pujic P, Battenberg K, Alloisio N, Liang Y, Hillhorst H, Salgado MG, Hoche V, Gherbi H, Svistoonoff S, Doyle JJ, He S, Xu Y, Xu S, Qu J, Gao Q, Fang X, Fu Y, Normand P, Berry AM, Wall LG, Ane J-M, Pawlowski K, Xu X, Yang H, Spannagl M, Mayer KFX, Wong GK-S, Parniske M, Delaux P-M, Cheng S. 2018. Phylogenomics reveals multiple losses of nitrogen-fixing root nodule symbiosis. *Science* 361. <https://doi.org/10.1126/science.aat1743>.
 41. Bennett GM, Moran NA. 2015. Heritable symbiosis: the advantages and perils of an evolutionary rabbit hole. *Proc Natl Acad Sci U S A* 112:10169–10176. <https://doi.org/10.1073/pnas.1421388112>.
 42. Kitano H. 2010. Violations of robustness trade-offs. *Mol Syst Biol* 6:384. <https://doi.org/10.1038/msb.2010.40>.
 43. Sachs JL, Kembel SW, Lau AH, Simms EL. 2009. In situ phylogenetic structure and diversity of wild *Bradyrhizobium* communities. *Appl Environ Microbiol* 75:4727–4735. <https://doi.org/10.1128/AEM.00667-09>.
 44. Savory EA, Weisberg AJ, Stevens DM, Creason AL, Fuller SL, Pearce EM, Chang JH. 2020. Phytopathogenic *Rhodococcus* have diverse plasmids with few conserved virulence functions. *Front Microbiol* 11:1022. <https://doi.org/10.3389/fmicb.2020.01022>.
 45. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
 46. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
 47. Fisher RF, Long SR. 1989. DNA footprint analysis of the transcriptional activator proteins NodD1 and NodD3 on inducible nod gene promoters. *J Bacteriol* 171:5492–5502. <https://doi.org/10.1128/jb.171.10.5492-5502.1989>.
 48. Kimbrel JA, Thomas WJ, Jiang Y, Creason AL, Thireault CA, Sachs JL, Chang JH. 2013. Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLoS Pathog* 9:e1003204. <https://doi.org/10.1371/journal.ppat.1003204>.
 49. abby ss, rocha epc. 2017. identification of protein secretion systems in bacterial genomes using MacSyFinder. *Methods Mol Biol* (Clifton, NJ) 1615:1–21. https://doi.org/10.1007/978-1-4939-7033-9_1.

50. Staehelin C, Krishnan HB. 2015. Nodulation outer proteins: double-edged swords of symbiotic rhizobia. *Biochem J* 470:263–274. <https://doi.org/10.1042/BJ20150518>.
51. Contreras-Moreira B, Vinuesa P. 2013. GET_HOMOLOGUES, a versatile software package for scalable and robust microbial pangenome analysis. *Appl Environ Microbiol* 79:7696–7701. <https://doi.org/10.1128/AEM.02411-13>.
52. Day A. 2012. heatmap.plus: Heatmap with more sensible behavior <https://CRAN.R-project.org/package=heatmap.plus>. Accessed.
53. Yu G, Smith DK, Zhu H, Guan Y, Lam TT-Y. 2017. GGTREE: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods Ecol Evol* 8:28–36. <https://doi.org/10.1111/2041-210X.12628>.
54. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>.
55. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303. <https://doi.org/10.1101/gr.107524.110>.
56. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 13:2498–2504. <https://doi.org/10.1101/gr.1239303>.
57. Tonkin-Hill G, MacAlasdair N, Ruis C, Weimann A, Horesh G, Lees JA, Gladstone RA, Lo S, Beaudoin C, Floto RA, Frost SDW, Corander J, Bentley SD, Parkhill J. 2020. Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* 21. <https://doi.org/10.1186/s13059-020-02090-4>.
58. Hudson CM, Lau BY, Williams KP. 2015. Islander: a database of precisely mapped genomic islands in tRNA and tmRNA genes. *Nucleic Acids Res* 43:D48–D53. <https://doi.org/10.1093/nar/gku1072>.
59. Darling AE, Mau B, Perna NT. 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS One* 5:e11147. <https://doi.org/10.1371/journal.pone.0011147>.
60. Galardini M, Biondi EG, Bazzicalupo M, Mengoni A. 2011. CONTIGuator: a bacterial genomes finishing tool for structural insights on draft genomes. *Source Code Biol Med* 6:11. <https://doi.org/10.1186/1751-0473-6-11>.
61. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H. 2019. Welcome to the tidyverse. *JOSS* 4:1686. <https://doi.org/10.21105/joss.01686>.
62. Anonymous. 2017. R Core Team, R Foundation for Statistical Computing R: a language and environment for statistical computing. <http://www.R-project.org>.
63. Brynildsrud O, Bohlin J, Scheffer L, Eldholm V. 2016. Rapid scoring of genes in microbial pan-genome-wide association studies with Scoary. *Genome Biol* 17. <https://doi.org/10.1186/s13059-016-1108-8>.
64. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res* 21:487–493. <https://doi.org/10.1101/gr.113985.110>.
65. Pritchard L, White JA, Birch PRJ, Toth IK. 2006. GenomeDiagram: a python package for the visualization of large-scale genomic data. *Bioinformatics* 22:616–617. <https://doi.org/10.1093/bioinformatics/btk021>.
66. Brown CT, Irber L. 2016. sourmash: a library for MinHash sketching of DNA. *J Open Source Softw*.
67. Davis EW, Weisberg AJ, Tabima JF, Grunwald NJ, Chang JH. 2016. Gall-ID: tools for genotyping gall-causing phytopathogenic bacteria. *PeerJ* 4:e2222. <https://doi.org/10.7717/peerj.2222>.
68. Wu M, Scott AJ. 2012. Phylogenomic analysis of bacterial and archaeal sequences with AMPHORA2. *Bioinformatics* 28:1033–1034. <https://doi.org/10.1093/bioinformatics/bts079>.
69. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.
70. Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* 3:217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>.
71. Porter SS, Simms EL. 2014. Selection for cheating across disparate environments in the legume-rhizobium mutualism. *Ecol Lett* 17:1121–1129. <https://doi.org/10.1111/ele.12318>.