Plasmid transmission dynamics and evolution of partner quality in a natural population of *Rhizobium leguminosarum*

David Vereau Gorbitz[a,b,#], Chase P. Schwarz[a,c], John G. McMullen[d,*], Mario Ceron-Romero[e], Rebecca T. Doyle[f], Jennifer A. Lau[d], Rachel J. Whitaker[a,b], Carin K. Vanderpool[a,b], Katy D. Heath[a,b,c,#]

[a]Carl R. Woese Institute for Genomic Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA

[b]Department of Microbiology, University of Illinois Urbana-Champaign, Urbana, IL, USA

[c]Department of Plant Biology, University of Illinois Urbana-Champaign, Urbana, IL, USA

[d]Department of Biology, Indiana University, Bloomington, IN, USA

[e]Department of Biology, Case Western Reserve University, Cleveland, OH, USA

[f]Department of Biology, McMaster University, Hamilton, Ontario, Canada

Running Head: Genomic diversity found in a *R. leguminosarum* natural population

#Address Correspondence to David Vereau Gorbitz, davidv3@illinois.edu; or Katy D. Heath, kheath@illinois.edu

*Present address: Biologics, Bayer Crop Sciences, Chesterfield, MO, USA

Abstract word count: 239

Text Word count: 5111

ABSTRACT

Many bacterial traits important to host-microbe symbiosis are determined by genes carried on extrachromosomal replicons such as plasmids, chromids, and integrative and conjugative elements. Multiple such replicons often coexist within a single cell and, due to horizontal mobility, have patterns of variation and evolutionary histories that are distinct from each other and from the bacterial chromosome. In nitrogen-fixing *Rhizobium*, genes carried on multiple plasmids make up almost 50% of the genome, are necessary for the formation of symbiosis, and underlie bacterial traits including host plant benefits. Thus the genomics and transmission of plasmids in *Rhizobium* underlie the ecology and evolution of this important model symbiont. Here we leverage a natural population of clover-associated *Rhizobium* in which partner quality has declined in response to long-term nitrogen fertilization. We use 62 novel, reference-quality genomes to characterize 257 replicons in the plasmidome and study their genomics and transmission patterns. We find that, of the four most frequent plasmid types, two (types II & III) have more stable size, larger core genomes, and track the chromosomal phylogeny (display more vertical transmission), while others (types I & IV – the symbiosis plasmid, or pSym) vary substantially in size, shared gene content, and have phylogenies consistent with frequent horizontal transmission. We also find differentiation in pSym subtypes driven by long-term nitrogen fertilization. Our results highlight the variation in plasmid transmission dynamics within a single symbiont and implicate plasmid horizontal transmission in the evolution of partner quality.

IMPORTANCE

Understanding how bacterial genes move through natural populations is critical for understanding how bacterial traits evolve. The nitrogen-fixing bacterium *Rhizobium*

leguminosarum lives in symbiosis with plants and is a model for studying how gene transmission from one cell to another on mobile genetic elements called plasmids impacts the evolution of bacteria and plants. Here we characterize the genomes of a natural bacterial population, then use novel approaches to show that mechanisms of plasmid gene transmission varies across multiple plasmid types possessed by *R. leguminosarum.* We find that changes in plasmid genes are associated with the decline of symbiotic partner quality in strains isolated from environments undergoing long-term fertilization. Together, these results underscore the importance of plasmid evolution in shaping ecosystem processes like nitrogen cycling. Our study provides a framework for probing the plasmid dynamics within natural bacterial populations and how plasmid transmission affects genetic diversity and ecological interactions in bacteria.

INTRODUCTION

Predictive models of bacterial trait evolution require a comprehensive understanding of how bacterial genes are inherited in natural populations. Bacterial traits arise and evolve via point mutations, gene duplication, homologous and non-homologous recombination, structural variants like transposons, and rare events like functionalization of non-coding regions (1). In addition, bacterial genomes can also undergo size-reduction in nutrient-limited environments or through endosymbiosis (2,3). At the core of all these processes are the vertical and horizontal transfer of genes and extrachromosomal elements (ECEs) such as plasmids, megaplasmids, and chromids (4,5) between strains. Diverse ECEs are characterized by distinct patterns of mutation, recombination, gene flow, and co-inheritance, which in turn influence how traits on these elements evolve through time (6–8). Bacterial genomes having two or more independent replicons – multipartite genomes – allow us to test how patterns of gene content, sequence similarity, size variation, and horizontal transmission rates vary across multiple extra-chromosomal elements within a single lineage of bacteria and thereby impact trait evolution.

Species with multipartite genomes are prevalent in nature and fill important ecological niches, including as opportunistic pathogens/mutualists and obligate endosymbionts. Multipartite genomes are particularly enriched in Pseudomonadota, but have been found in distant phyla including Cyanothece, Leptospira, and Deinococcus (9). The ECEs of multipartite genomes carry genes in symbionts that expand their hosts' metabolism, as well as genes that confer resistance to antibiotics and heavy metal toxicity (10–15). ECEs can replicate independently of the chromosome, be transmitted in whole or in part between bacteria via horizontal gene transfer (HGT), and increase the size and diversity of a pangenome (the repertoire of core and variable genes found in either all, or some, members of the species, respectively) (16–19). Moreover this

pangenomic variation is due in part to gene gain and loss via the action of mobile genetic elements within the pangenome, often found on ECEs (20). Due to HGT, ECEs can have evolutionary histories that are quite distinct from those of their bacterial chromosomes (21), effectively decoupling the fitness interests of ECEs and chromosomes and even driving coevolutionary dynamics among the elements within a single cell (22,23). Given the importance of ECEs in symbionts and their ability to modify their hosts' ecology, there is a critical need for an ECE-centric approach in natural bacterial populations.

Genomic variation in bacterial pangenomes including single nucleotide polymorphisms (SNPs), indels, translocations, inversions, and duplications (24) are often concentrated on ECEs (25). Yet approaches to studying genetic variation using high-throughput shotgun sequencing tend to exclude or minimally investigate ECEs (26). For instance, common markers of phylogenetic distance such as average nucleotide identity (ANI), percentage of shared single nucleotide polymorphisms (SNPs), and shared genes are usually calculated on chromosomal genes. Reference-based assemblies fail to fully capture presence-absence diversity, and can miss major genome rearrangements, meaning that much of the genetic variation driving genome evolution can be missed or misassembled (27,28). Non-reference based, *de novo* assembly of closed, circular replicons is made difficult by repetitive elements in bacterial genomes and particularly ECEs, resulting in genome fragmentation (29) and hindering the assembly and analysis of ECEs and pan-genomes (30,31). Harnessing long-read technology to assemble populations of closed, reference-quality genomes enables pan-genome analysis of genetic diversity in natural populations of ECEs – the diversity on which selection acts.

Nitrogen-fixing rhizobia, bacteria that fix nitrogen in symbiosis with leguminous plants (32), are good models for the evolutionary ecology of ECEs since they contain multiple and

diverse plasmids in their multi-partite genomes and because the genes required for associating intracellularly with hosts are found on plasmids or Integrative and Conjugative Elements (ICEs) (7,33,34). The *Rhizobium leguminosarum* species complex, formerly all one polyphyletic species, was recently divided into at least 5 genospecies (35); most recently, certain genospecies (*gsA, gsB, gsD*) have been classified as new species (*Rhizobium brockwellii*, *johnstonii,* and *beringeri,* respectively) (36). *Rhizobium* isolates are well-known to carry many plasmids, which can coexist in the same cells due to variation at the *repABC* operon (35,37); *repA* in particular determines incompatibility group, and the same *repA* "Rh group" has been found across multiple genospecies (35). These diverse *Rhizobium* plasmids tend to contain non-essential genes and are mobile, based on studies showing that the phylogenetic histories of plasmid genes are distinct from those of the chromosome (38). Fully-resolved genomes for many natural isolates are required to address variation in phylogenetic history within and among plasmids and thus understand plasmid transmission patterns on timescales relevant to bacterial trait evolution.

Clover-associated *Rhizobium* are a particularly powerful tool for understanding the contribution of plasmid diversity and plasmid transmission to bacterial evolution in nature. The canonical symbiosis genes that enable *Rhizobium* to interact with clover, vetch, and other hosts are not only found on plasmids, but have been found on *different* plasmids among natural isolates, indicating HGT of this critical gene region across replicons (32,35,39,40). In our previous work using a population of *Rhizobium* from the Long Term Ecological Research (LTER) site at Kellogg Biological Station (KBS) in Michigan, we found that strains from old field communities exposed to long-term nitrogen (N) since 1998 are less-beneficial on average for clover (*Trifolium spp*.) host plants, compared to control (41). Moreover we used reference-based SNP analysis to associate this decline in partner quality, *i.e.,* the benefits that symbionts

125 provide to their host, with differentiation at the canonical symbiosis gene region (42). Beyond
126 contributing to fundamental knowledge on plasmids, a better understanding of plasmid
127 inheritance in these multi-partite genomes is critical for understanding symbiosis gene
128 transmission and thus the role of HGT in symbiosis trait evolution.

129 In this study, we generate high-quality reference genomes to study ECE dynamics in a
130 natural bacterial population, leveraging 62 previously-studied strains of *Rhizobium* (41–43). We
131 use this population to understand how sequence diversity, gene function, and size differ among
132 plasmids within a pan-genome (the plasmidome). We then use phylogenomic approaches to infer
133 how the propensity for horizontal and vertical inheritance differs across plasmids in the pan-
134 genome. Finally we study the pSym to examine the role of plasmid HGT in partner quality
135 decline.

136

137 **RESULTS**

138 *Delineating the plasmidome:* To assess the ECE diversity in natural populations of *R.*
139 *leguminosarum*, we used long-read technology to sequence and assemble complete genomes *de*
140 *novo* for all 62 nodule isolates from Weese et al. (41). Each strain's assembled genome contained
141 1-6 extrachromosomal elements. In total, we identified 257 extrachromosomal elements in our
142 bacterial population. Of those, 256 had at least one *repABC* operon, which has been used as a
143 basis for plasmid categorization (35,38,44,45), and 17 replicons had two distinct *repABC*
144 operons. One 1.4 Mbp replicon in strain 773_N did not contain identifiable plasmid replication
145 genes.

146 We first reconstructed a phylogenetic tree from a core chromosome alignment (Fig. S1),
147 which showed that 56 of our 62 strains formed a single clade identified as *R. leguminosarum*

148 sensu stricto, or *genospecies E* (*gsE*) (35). Four of the 62 strains (061_N, 173_C, 209_N, 231_N)
149 were identified as *gsB*/*Rhizobium johnstonii* (36), and were characterized within our population
150 by low diversity in the chromosomes shared between these four strains. The remaining two
151 strains, 717_N and 773_N, fell outside both *gsE* and *gsB* clades; strain 717_N is sister to the *gsB*
152 clade, and strain 773_N is the most basal strain. To further investigate the identity of strain
153 773_N, we queried NCBI non-redundant database with its 16S rRNA and *dnaA* sequences and
154 found that the closest sequence identity match with a complete genome for both sequences was
155 *Rhizobium sp.* WYJ-E13 (accession ID: PRJNA738292), 99.12% and 92.86% respectively.
156 To categorize all plasmids into discrete groups, we used a k-mer approach that considers
157 core and non-core regions of all plasmids. We calculated the k-mer signatures for all 257
158 replicons and generated a weighted undirected network component graph based on the Jaccard
159 Index for all-vs-all combinations (Fig 1). This approach benefits from using all genetic
160 information in both the core and non-core regions of the plasmid, as opposed to using only core
161 gene-coding regions present in all plasmids. This more inclusive approach is especially
162 important since there is little overlap in genes across all replicons. Of 257 plasmids, most (226)
163 fell into one of four plasmid clusters (hereafter type I, II, III, and IV; Fig. 1), while the few
164 remaining plasmids fell into an additional four types (hereafter type V, VI, VII, and VIII).
165 Plasmids were numbered based on sample number, and then size within the population. Both
166 Type I and Type IV plasmids clustered into 3 and 4 smaller sub-groups, respectively (labeled a,
167 b, c, d; Fig. 1)
168 The distribution of these plasmid types in our study correspond to the bifurcations in the
169 chromosome phylogeny (Fig. 2), with different genospecies possessing distinct plasmidomes.
170 Type IV plasmids were found in 58/62 strains across both genospecies (*i.e., gsB* and *gsE*); type I,

II, and III plasmids were present in all 56 *gsE* strains (Fig. 2), while the plasmid type groups V, VI, VII, and VIII were present in all four *gsB* strains, indicating that plasmid types delineated by k-mer analysis track with chromosomal genospecies. Like their associated *gsB* chromosomes, Type V-VIII plasmids contain little genomic diversity (table 1).

We next examined the presence of the canonical symbiosis genes (*i.e. nif, fix,* and *nod*) in our annotated genomes, and found they were limited to type IV plasmids (pSym hereafter). Therefore, the presence of type IV plasmids appears to be required for symbiosis with clovers (*Trifolium* spp.) in this population of *Rhizobium*. Interestingly, we found that four of the 62 strains lacked the pSym, further corroborating that this non-essential element could be lost in strains present within a naturally occurring population (42). Given that all 62 strains were isolated from nodules, it is not known whether this loss occurred in culture or whether these isolates co-infected nodules with symbiotically-capable strains (46,47).

Of the remaining smaller clusters of plasmids, fourteen unique plasmids did not group with any other plasmid within the population (Fig. 1). Of these singleton replicons, six were found in strains outside the main *gsE/gsB* clades; five of the singleton replicons belonged to strain 717_N (type IX, X, XI, XII, and XIII), and one belonged to strain 773_N (Fig. 2). The other eight singleton plasmids (hereafter, accessory plasmids) were in strains scattered across the *gsE* phylogeny (Fig. 2).

Genomic content of the plasmidome: To visualize the difference in gene composition between plasmids, we annotated the genomes using NCBI's PGAP (48), assigned genes to clusters of orthologous genes (COG) groups (49), and performed a Principal Coordinate Analysis (PCoA) with the Jaccard distance calculated from the presence-absence of orthologous gene clusters of

194 the plasmidome. We found that, using PCoA based on orthologous gene content, all 257
195 plasmids formed clear clusters that correlated with plasmid types from k-mer clustering (Fig. 3),
196 further indicating that these plasmid types differ in gene content and have distinct functional
197 roles. In the PCoA, the pSyms cluster together in the center between plasmids I-III, and
198 alongside the accessory plasmids, suggesting some shared gene content among these elements. In
199 particular, many (209) genes were shared between at least one accessory plasmid and one Type
200 IV plasmid.

201 Although the *gsB* plasmid types V-VIII formed k-mer clusters distinct from *gsE* plasmids
202 (Fig. 1), these four plasmids were similar to (had low Jaccard distances with) *gsE* plasmid types I
203 (V), II (VI), and III (VII and VIII) (Fig S2A), suggesting that these sets of elements serve
204 somewhat analogous functions in *gsB* and *gsE*. Indeed pairwise comparison of type I with V and
205 type II with VI indicated 56% and 74% core gene content overlap, respectively (Fig. S2B).
206 Interestingly, while pairwise distances between *gsE* plasmid type III and either *gsB* type VII or
207 VIII were low (Fig. S2A), pairwise distance between *gsB* types VII and VIII was very high (Fig.
208 S2B); these patterns suggest that the core gene content of type III in *gsE* is comprised of a
209 combination of the cores of Types VII and VIII in *gsB*. This is supported by an alignment of
210 types VII and VIII to the type III plasmid, showing large syntenic aligned blocks consistent with
211 shared plasmid history (Fig. S2C). Interestingly, while plasmid types I and V were both in Rh
212 group 1, and II and VI were both in Rh group 2, plasmid types III, VII, and VIII belonged to
213 three different Rh groups (4,5, and 3, respectively) (Table 1) – indicating that plasmid gene
214 content often, but not always, tracks *repABC*-based Rh group.

215 Overall, COG analysis suggested that high-level distribution of functional gene content
216 was similar across all plasmids (Fig. S3), despite their unique patterns of gene presence-absence

variation (PAV). Across all plasmid types in the population, four COG categories were completely absent in the functional prediction (A, B, W, Y). The absence of these functions was not surprising as they relate to RNA processing (A), chromatin (B), extracellular (W), and nuclear (Y) structures, which are more commonly associated with eukaryotic or chromosomal processes. The one outlier was the pSym, which features more genes for intracellular trafficking and secretion (U), and recombination and replication (L) (Fig. S3).

Variable size distributions in gsE *Rhizobium plasmids*: Because we have the largest sample of plasmids from *gsE* strains (I-III, and pSym), hereafter we primarily focus on the plasmidome of the *gsE* subgroup in order to examine within-type variation and transmission. Plasmid types varied in size and in the shape of their size distribution (Fig. 4). Type I plasmids were the largest in our population, had the largest size variation, and had a bimodal distribution centered at either ~0.9 Mbp or ~1.2 Mbp. Type II and type III plasmids were similar in size, with intermediate lengths ranging from ~0.5-0.6 Mbp. Of the *gsE* plasmidome (types I-IV), pSyms were the smallest (~0.2-0.4 Mbp). Notably, this symbiosis gene location is distinct from the reference genome WSM1325 (*gsA*), in which the symbiosis genes are found on the largest plasmid (32).

Comparing the size variation within the type I plasmid with its core length, content, and phylogeny (Fig. 4, Table 1) allows us to reconstruct the history of how dramatic changes in plasmid size have evolved in this element. The three type I k-mer subgroups (I-a, I-b, I-c; Fig. 1) map onto the three distinct groups in the type I core phylogeny (Fig. S4a). However although subgroup I-b and I-c have similar size distributions, they contain distinct large insertions (Fig. S4b). All small type I plasmids (subgroup I-a) form a clade that diverged from I-b after a single loss of the ~0.3 Mb insertion (Fig. S4b, c). Thus independent gain and loss of large insertions is

responsible for the bimodal size distribution; in addition, these large insertions occur in similar locations along the plasmid (Fig. S4c), suggesting that type I plasmids might have specific hotspots where insertions are more likely.

Mode of inheritance varies in the Rhizobium *plasmidome:* To study plasmid transmission modes, we used a phylogenetic approach with random resampling of orthologous genes to quantify patterns of gene tree heterogeneity within each plasmid type. We then calculated the Generalized Robinson-Foulds (GRF) distance between two trees as a measure (0-100) of phylogenetic congruence between two trees. Mean GRF distance between gene trees of the chromosome was 72.46, with a standard deviation of 5.40 (Fig. 5). The type I and type II plasmids had similar distributions to each other and to the chromosome (mean = 72.84 $\pm$ sd 5.08 and 68.00 $\pm$ sd 6.31, respectively) – suggesting similar levels of within-element vertical versus horizontal transmission on these three elements. By contrast, the distribution of type III plasmid GRF distances (mean = 56.41 $\pm$ sd 5.09) indicates that gene trees within this element were more similar – suggesting more internal consistency and thus less horizontal transmission of genes compared even to the chromosome. Finally, the pSym GRF distribution was wider and centered at mean = 86.79 $\pm$ sd 9.46, indicating higher divergence of gene trees among the loci on this element, compared to the other four – consistent with abundant HGT of genes on the pSym.

Next, to assess the degree to which plasmids are vertically inherited (together with the chromosome), versus horizontally (separately from the chromosome), we compared the distributions of GRF distances when gene trees from each plasmid were compared to those from the chromosome. The distributions of GRF values for plasmid types II and III fell inside the 95th percentile of the chromosome GRF distance distribution (Fig. 5), indicating that the likelihoods

of co-inheritance of chromosomal genes with genes on these plasmids were not different from the co-inheritance of chromosomal genes with each other – supporting abundant vertical transmission alongside the chromosome in this population. By contrast, the chromosome-type I gene distribution (mean = 83.24 ± sd 2.97) fell above the 95th percentile of the chromosomal gene distribution (Fig. 5), indicating that the genes on these two elements tend to have distinct evolutionary histories, suggesting HGT. Finally, GRF distances between genes on the chromosome and the pSym were particularly high (mean = 99.56 ± sd = 0.83, with 881/1000 resamplings resulting in a maximum GRF = 100; Fig. 5), also suggesting high levels of horizontal transmission of pSym genes relative to the chromosome – likely due to the HGT of entire pSym plasmids across chromosomal lineages. Due to the small core, robust clade patterns, and high levels of within-subclade core genes of the pSym (see below Fig. 6A,6B), we separately ran the resampling analysis of gene tree heterogeneity by pSym subclade to test whether these patterns of HGT result from across-clade differences rather than individual plasmids moving independently of the chromosome. Despite high sequence similarity within Type IV-a, b, c, these subclades show elevated GRF distances compared to the chromosome (Fig. S5A), suggesting that the entire pSym is moving horizontally.

pSym (Type IV) plasmid movement drives differences in partner quality: Given the evidence for abundant HGT of the pSym plasmid, and previous population genetic analyses focusing on symbiosis-related loci (42), we next studied patterns of differentiation in the pSym in order to relate pSym variation to symbiotic partner quality. First, we inferred a phylogenetic tree with concatenated core sequences from the type IV plasmid and found four clades (Fig. 6A); these clades mirror the four major k-mer clusters of type IV plasmids (Fig. 1).

286 The pSym is characterized by a particularly small core (52.1 kb, or 49 genes) relative to
287 other types (Table 1). Although a gene presence-absence plot shows modularity and unique gene
288 content unique in subclades IVa-c (Fig. S5B,C), groups of shared genes are often found in
289 unrelated clades (Fig. S5C), and the core gene set of any two pSym subgroups (regardless of
290 phylogenetic distance) is noticeably larger than the universal pSym core of 49 genes – indicating
291 abundant gene-sharing across groups on the pSym phylogeny (Fig. S5C, D). For example,
292 despite being more phylogenetically distinct and most often found in the *gsB* minor clade, nearly
293 all IV-d genes are shared with other type IV sub-clades (Fig. S5C).
294 We found clear cases of pSym HGT across distinct chromosomal lineages in our
295 population. First, although strain 717_N is more closely related to *gsB* than *gsE* at the
296 chromosome (Fig. S1) and shares no other plasmids with either group (Fig. 2), the 717_N pSym
297 falls alongside a group of high-quality *gsE* strains in the pSym tree (clade IV-b, Fig. 6A).
298 Similarly, although the *gsB* and *gsE* chromosomal clades generally have distinct plasmidomes
299 including the pSyms (Fig. 2, Fig 6A), two *gsE* strains (308_C and 859_N, Fig. S1) were found to
300 carry *gsB*-like pSyms (clade IV-d; Fig. 6A) indicating cross-genospecies HGT.
301 Finally we found striking correspondence between the core pSym phylogeny and the
302 benefits of symbiosis for plant hosts, as measured in a previous common garden experiment (41).
303 Within the major pSym groups, clades IV-a and IV-b contain mostly higher-quality strains that
304 originated from unfertilized control plots (Fig. 6A), whereas clade IV-c contains an abundance of
305 lower-quality strains isolated from N-fertilized plots and features longer branch lengths due to
306 SNPs in core genes (Fig. 6A). A Fisher's exact test supported a significant difference in
307 frequency of Type IV sub-clades between control and N treatments ($p = 0.042$; Fig. 6B).
308

DISCUSSION

Plasmid inheritance is a key process underlying bacterial trait evolution in natural and managed ecosystems. Studies of plasmid variation and transmission within local-scale, recombining populations are needed to quantify patterns of gene co-inheritance as well as their influence on bacterial phenotypes. Here we delineate the major types, size, gene content variation, and transmission patterns of coexisting plasmids from a single population of clover-associated *Rhizobium*. We find that *Rhizobium* plasmids vary considerably in their size structures and modes of transmission; some plasmids (type II, type III) appear to be primarily vertically transmitted, while others (type I, pSym) are more likely to be horizontally transmitted. Concomitant with these findings from within our best-sampled genospecies (*gsE*), we find that most of the plasmidome is delimited by chromosomal lineages. Nevertheless, the extent of this limitation varies across plasmids; for example, we find clear examples of cross-genospecies HGT of the pSym. Finally, our analysis of pSym subclades indicates a role for pSym HGT in the decline of clover-associated partner quality in N-fertilized environments. Below we discuss the important lessons learned from each of the major element in our *Rhizobium* plasmidome, then finish with a holistic discussion of the potential importance of transmission variation to the evolution of bacterial traits.

Type I plasmids: The largest plasmid in *Rhizobium* frequently carries the symbiosis genes (42,43). Indeed in previous work (42) using referenced-based assembly to WSM1325, we assumed that the Type I plasmid was the symbiosis plasmid (see discussion of the pSym below). This highlights the value of using non-reference-based genome assembly facilitated by long read sequencing, as well as naïve k-mer based clustering, for studying plasmid variation, plasmid

transmission, and the evolution of plasmid-borne traits. Plasmid size variation is usually caused by differing patterns of presence and absence of genes, which is caused by homologous recombination or horizontal gene transfer (50,51). Studies of plasmid evolution usually limit analyses to a handful of genes involved in plasmid replication, maintenance and transfer (52,53). The ability to interrogate fully closed plasmid genomes allowed the separation of core and variable content within this single plasmid type, revealing the gain and loss of large insertions that dramatically alter plasmid size and track the core phylogeny. The internal consistency of the type I plasmid (based on the comparison of core and variable content, and similar *within*-plasmid gene tree distances to those of the chromosome) at first appears at odds with evidence for HGT *between* type I plasmids and the chromosome. These results might hint at distinct mechanisms governing within-element stability versus whole-element transfer in this plasmid (*e.g.,* homologous recombination versus conjugation), though more functional studies are required to make any generalizations.

Type II and III plasmids: We found that the type II and type III plasmids are more frequently vertically transmitted alongside the chromosome, compared to the other plasmids, and accordingly, have a large fraction of core genes. In contrast with type I and IV plasmids, the gene PAV in type II and III plasmids is comprised of singleton genes being present at low frequencies in the plasmids as opposed to large gene-clusters being present in multiple closely related plasmids. This suggests that the mechanisms by which type II & III, and type I & IV are acquiring and losing genes are different. In fact, the type III plasmid shows particularly low *within*-plasmid GRF distances, suggesting particularly low rates of recombination, though the potential mechanisms remain unclear.

355 Our observation that the *gsE* type III plasmid contains the gene content of *two* other
356 plasmids from the *gsB* group (type VII and VIII) might suggest a key evolutionary event in the
357 history of these *Rhizobium* plasmids – either a subdivision of one (fission) or a merging of two
358 (fusion) – and denotes a key change in how *gsE* and *gsB* strains subdivide their respective
359 genomes. Importantly, however, the type III does not share its Rh incompatibility group (35)
360 with either of these plasmids, making historical reconstruction difficult. Chromosomal fusion is
361 known in closely-related *Agrobacterium tumefaciens (54)*. The "schism hypothesis" of
362 chromosome fission has been developed as an explanation for the evolution of multi-partite
363 genomes (51); such processes might generate plasmid diversity as well. Fusion and fission of
364 eukaryotic chromosomes is well-known to drive reproductive isolation and thus speciation in
365 plants and animals (54). Understanding the impact of these processes in the diverse plasmids
366 of *Rhizobium* and other species is paramount to further understanding bacterial genome evolution
367 and the diversification of bacterial species.
368
369 *Type IV plasmids*: The pSym is usually defined as the replicon that harbors genes necessary for
370 symbiosis with a host plant (23). In our population, the pSym is the type IV plasmid, the smallest
371 of the main plasmids and the only one present in both *gsE* and *gsB* chromosomal genospecies. It
372 is also the most variable, having a very small pool of core genes and deeply diverged lineages.
373 Nevertheless, we treat the pSyms as one "type" of plasmid for multiple reasons. First is their
374 clustering based on the k-mer approach we used to categorize plasmids. Though previous work
375 has used *repABC* Rh types to categorize *Rhizobium* plasmids (35), we found that 17 pSyms in
376 our population contain two distinct full *repABC* operons, suggesting the need for an additional
377 approach that reflects gene content similarity. Because plasmids with the same *repABC* groups

generally cannot be maintained in the same cell (55), it is unclear what the evolutionary advantage of having two distinct copies of the operon in a single plasmid might be. Second, while some pSym lineages possessed subclade-specific gene clusters, there was abundant shared gene content in pairwise comparisons of the pSym subclades regardless of relatedness at the core – suggesting mobility of pSym genes across the pSym phylogeny. Finally, examining all pSyms together based on gene content and function allowed us to relate pSym subclade to symbiotic partner quality and detect shifts in subclade frequency between N-fertilized and control plots (see below).

The symbiosis plasmids in the *R. leguminosarum* species complex are well-known to exhibit high levels of variable gene content and horizontal mobility (23,56,57). Nevertheless we found that the pSym subclades tended to be associated with different chromosomal genospecies (*gsE* and *gsB*), and the few notable exceptions allow us to pinpoint clear cases of pSym HGT across the genospecies boundary. Previous approaches designed to detect introgression events across genospecies support shared alleles at symbiosis genes *nifB*, *nodC*, and *fixT* across *gsE* and *gsB* (35); our results suggest that the symbiosis plasmids move across these genospecies boundaries as well. Although a much more thorough functional analysis would be required to pinpoint the underlying drivers of transmission in our plasmids, extra recombination genes (L) in the pSym might explain higher levels of recombination and PAV in this plasmid. Nevertheless it is interesting to consider the genetic drivers and selective forces that might reinforce, versus break up, this type of structure in chromosome-plasmid relationships through time and across environmental conditions (*e.g.,* the presence or absence of hosts, changing abiotic conditions (51,58)).

Previously, we had reported differentiation of the symbiosis gene region between high-quality partners in the control plots and low-quality partners in the N-fertilized plots in this *Rhizobium* population, relative to the rest of the genome (42). Here, we add evidence that it is the entire pSym, and not just the symbiosis gene region, that is differentiated. Together with evidence for HGT of the pSym – both between chromosomal genospecies and within a single, well-sampled genospecies (*gsE*) – we infer a shift in pSym subclade frequencies with a change in the environment, rather than a gene-specific sweep at the symbiosis gene region. The shorter branch lengths in higher partner quality pSym subclades might indicate purifying selection in control plots, or relaxed selection in N-fertilized plots. Our new, plasmid-centric interpretation of the genetic underpinnings of partner quality decline stems from both data type and analytical methods; our long read-enabled full genome assemblies reveal diversity in symbiosis gene location, duplicate pSym *repABC* types, and pSym gene content that was not previously visible. What's more, these fully-resolved plasmidome sequences, combined with novel phylogenomic analyses, allow us to quantify patterns of gene tree heterogeneity not only at the pSym, but across all plasmid types and thus make inferences about the variation in transmission modes among elements within a single genome.

Transmission in the Rhizobium *plasmidome*: The degree to which plasmid vertical versus horizontal transmission modes are determined by chromosomal mechanisms, plasmid-specific mechanisms, and/or the interaction is still being worked out (45,59). It has long been recognized that plasmid transmission mode can evolve as the costs and benefits of conjugation-growth tradeoffs shift (60,61), though the selective drivers in nature are not well-known. Most plasmid transmission studies take place in the lab under strongly selective conditions in one or a few

laboratory or clinical strains (62,63), whereas population genomic studies on natural diversity have historically not focused on ECEs given sequencing limitations (64,65). Studies of plasmid transmission in natural populations are rare, though others have compared gene content and plasmid diversity among non-coexisting plasmids within a single species (60,66). Here we establish a novel framework for studying plasmid transmission in nature and find quite different patterns of inheritance among the multiple plasmids that coexist within *Rhizobium* host cells. Some plasmids appear to move more vertically alongside the chromosome, while others move horizontally. Given that these plasmids co-occur in the same cells, plasmid-specific mechanisms must explain this variation, at least in part. Our findings support evolutionary conceptions of plasmids as having their own agency and fitness interests (61,67,68).

Rhizobia also provide a rare opportunity to study how plasmids and HGT facilitate quantitative trait evolution in natural bacterial populations. Rhizobial symbiosis genes are known to be both horizontally-transmitted and selected in nature (34,35). In our *Rhizobium* population, we find frequent HGT of the pSym relative to the other plasmids paired with differentiation of the pSym between N environments, suggesting environmentally-dependent selection on this plasmid. The symbiosis genes in *Rhizobium* can move between plasmids (32,35,38,39), and as we show here, those plasmids can vary in rates of HGT – potentially suggesting that the traits governed by these mobile loci will evolve more or less rapidly depending on their genomic location. Rapid evolution of symbiotic traits via gene-specific sweeps, whole plasmid sweeps, and even plasmid loss (69) might be advantageous in context dependent mutualisms where the costs and benefits of symbiosis shift with the biotic and abiotic context in which partners interact (70,71).

By integrating genomics, plasmid biology, phylogenetics, and plant-microbe interactions in wild bacteria, our study provides a framework for quantifying the relative rates of vertical and horizontal transmission among the ECEs that coexist within a single species and elucidates the role of plasmid HGT in an ecologically-important symbiosis that plays a critical role in global nitrogen cycling.

MATERIALS AND METHODS

Strain isolation and growth: Here we generate novel genomes for a population of 62 *Rhizobium* strains originally isolated from old field successional plots at the KBS LTER. Full methods detailing the long-term N fertilization experiment, strain isolations, and phenotypic experiments to characterize partner quality symbiosis with three clover host species are described elsewhere (41). Briefly, rhizobium strains were isolated from both N-fertilized and control plots. Fertilized plots had been supplemented with 12.3 g N m^{-2} per year granular ammonium nitrate for 22 years prior to sampling, whereas control plots remained unfertilized.

DNA and sequencing: We grew the strain isolates in solid tryptone yeast (TY) media (5 g L^{-1} tryptone, 3 g L^{-1} yeast extract, 6 mM CaCl2, and 16 g L^{-1} agar) plates at 30 C for 2 days. After growth on solid media, single colonies were selected to inoculate 5 mL liquid TY media for 1 day at 30 C in a roller drum. The PacBio Nanobind CBB kit (Pacbio, San Diego CA) was used to extract high molecular weight (50–300+ kbp) DNA from 1 mL of bacterial culture for all strain isolates. The DNA was sent for PacBio hifi long-read sequencing (72) at the W. M. Keck center at the University of Illinois, where gDNAs were sheared with a Megaruptor 3 to an average fragment length of 10kb then converted to barcoded libraries with the SMRTBell Express Template Prep kit 3.0 and pooled in equimolar concentration. The pooled libraries were

sequenced on 2 SMRTcell 8M on a PacBio Sequel IIe using the CCS sequencing mode and a 30hs movie time. Circular consensus sequence (CCS) analysis was done using SMRTLink V11.0 using the following parameters: ccs --min-passes 3 --min-rq 0.99 lima --hifi-preset SYMMETRIC --split-bam-named --peek-guess.

Genome assembly and annotation: All genomes were assembled using recommended workflows in Trycycler (73). Briefly, raw reads were filtered for quality using Filtlong, assessing both length and quality of the reads. The raw reads where then divided into a 12 maximally-independent subsets using the subsample function in Tryclycler. Next Trycycler uses three assembly methods (Flye (74), Hifiasm (75), and Raven (76)) to generate independent whole-genome assemblies for four read subsets each. Finally we used Trycycler to generates a consensus genome based on these 12 assemblies, followed by manual curation to ensure consistent genome structure across assemblies. Genomes were then annotated using NCBI's Prokaryotic Genome Annotation Pipeline (PGAP).

Classification of plasmids: We generated individual fasta files for every replicon in every strain then used sourmash (77) to generate and compare k-mer signatures and calculate pairwise Jaccard Index (JI) values using the parameters: sketch dna -p scaled=10000, k=31, compare -p 8. Signature tables were then imported into Cytoscape, and component graphs were created using a minimum JI value of 0.1 to delineate clusters; this cutoff was chosen based on similar analyses of *Rhizobium*, *Agrobacterium*, *Bradyrhizobium*, as well as a global plasmid analysis (7,37,78). We classified the resulting clusters as plasmid "types", then layered this plasmid presence onto phylogenetic trees using "ggtree" and "ggplot" libraries in R (version 4.3.2 "eye holes"). We used the package popgenome to calculate pi and Average Nucleotide Identity (ANI) for aligned core regions of each plasmid type.

Phylogenetic trees: We used a custom SPINE-Nucmer pipeline (https://github.com/Alan-Collins/Spine-Nucmer-SNPs) to generate core genome alignments for all 62 genomes (core genes were all chromosomal), all chromosomes (resulting in the same core), and subsequently for each plasmid type. For each, core components were concatenated, aligned using MAFFT, and used to generate phylogenetic trees in IQTree2 with parameters: -bb 10000 -st DNA.

Presence-absence data and Principal Coordinate Analysis: We used PIRATE (79) on PGAP-annotated genomes to define orthologous gene presence-absence using PIRATE (79)., followed by Principal Coordinate Analysis (PCoA) of this output using the "dplyr", "vegan" packages in Rstudio to calculate Jaccard distances between all samples based on shared gene content. Pairwise distances were then transformed using the "cmdscale" function from the "stats" package and sketched using the "ggplot" package. Gene presence-absence plots were made using "pheatmap" package.

Gene-tree simulations: To assess patterns of plasmid inheritance, we created custom scripts to generate distributions of gene tree distances among genes randomly subsampled from within (or across) replicons in our population – with larger gene tree distances indicating more gene tree heterogeneity and thus increased horizontal (versus vertical) transmission. First we generated a list of the core genes in each plasmid using PIRATE. For each replicon, we used custom R scripts to randomly subsample two sets of 100 genes (10 from the small core of the type IV/pSym) from each strain, align them, create two phylogenetic trees, and calculate Generalized Robinson-Foulds (GRF) distance between trees, then repeated this process over 1000 random resamplings. We plotted these GRF distributions using geom_density function in the ggplot2 package. Next we repeated this process comparing samples from each plasmid to the chromosome. Because the chromosome is necessarily vertically transmitted each generation, the

514 distribution of chromosome-to-chromosome comparisons serves as a null expectation for the
515 GRF distribution of a vertically inherited element in the presence of horizontal gene transfer,
516 recombination, and gene tree uncertainty.

517 **Availability of data and materials**

518 Phenotypic data, and scripts used for generating analyses will be available upon acceptance.
519 Genomes will be uploaded to NCBI genome database upon acceptance.

520

521 **ACKNOWLEDGMENTS**

522 D.V.G., K.D.H., C.K.V., and R.J.W. conceived the project. D.V.G. extracted and submitted
523 DNA for sequencing. D.V.G. and C.P.S. generated the genomes and performed bioinformatic
524 analyses. D.V.G. and K.D.H. drafted the article, and all authors participated in critical revisions
525 and approved the final version for submission. K.D.H., C.K.V., R.J.W., J.A.L. acquired funding
526 for the project.

527

528 We would like to acknowledge Dr. Alvaro Hernandez and Chris Wright from the Roy J. Carver
529 Biotechnology Center at the University of Illinois for library preparation and sequencing. We
530 thank Dr. Jaya Chandrashekhar, Dr. Susan Thomas, and the rest of GEMS for their feedback,
531 help, and coordination of lab protocols. We thank Dr. Ilan Shomorony, Dr. Pamela Martinez, and
532 Ivan Sosa-Marquez for their invaluable help, motivation, and insightful advice during the
533 manuscript writing process. The results and interpretation of this manuscript accomplished by
534 J.G.M. were in their personal capacity and are the author's own views and do not reflect the view
535 of their employer.

536

Competing interests

The authors declare that they have no competing interests

Funding

This research is a contribution of the GEMS Biology Integration Institute, funded by the National Science Foundation DBI Biology Integration Institutes Program, Award # 2022049, as well as NSF Award #1257938.

REFERENCES

1. Kirchberger PC, Schmidt ML, Ochman H. The Ingenuity of Bacterial Genomes. Annu Rev Microbiol. 2020 Sep 8;74(Volume 74, 2020):815–34.

2. Giovannoni SJ, Tripp HJ, Givan S, Podar M, Vergin KL, Baptista D, et al. Genome Streamlining in a Cosmopolitan Oceanic Bacterium. Science. 2005 Aug 19;309(5738):1242–5.

3. McCutcheon JP, Moran NA. Extreme genome reduction in symbiotic bacteria. Nat Rev Microbiol. 2012 Jan;10(1):13–26.

4. Finan TM. Evolving Insights: Symbiosis Islands and Horizontal Gene Transfer. J Bacteriol. 2002 Jun;184(11):2855–6.

5. Karcagi I, Draskovits G, Umenhoffer K, Fekete G, Kovács K, Méhi O, et al. Indispensability of Horizontally Transferred Genes and Its Impact on Bacterial Genome Streamlining. Mol Biol Evol. 2016 May 1;33(5):1257–69.

6. Riley AB, Grillo MA, Epstein B, Tiffin P, Heath KD. Discordant population structure among rhizobium divided genomes and their legume hosts. Mol Ecol. 2023;32(10):2646–59.

7. Weisberg AJ, Rahman A, Backus D, Tyavanagimatt P, Chang JH, Sachs JL. Pangenome Evolution Reconciles Robustness and Instability of Rhizobial Symbiosis. mBio. 2022 Apr 13;13(3):e00074-22.

8. Epstein B, Burghardt LT, Heath KD, Grillo MA, Kostanecki A, Hämälä T, et al. Combining GWAS and population genomic analyses to characterize coevolution in a legume-rhizobia symbiosis. Mol Ecol. 2023;32(14):3798–811.

9. Landeta C, Dávalos A, Cevallos MÁ, Geiger O, Brom S, Romero D. Plasmids with a Chromosome-Like Role in Rhizobia. J Bacteriol. 2011 Feb 24;193(6):1317–26.

10. Baker-Austin C, Wright MS, Stepanauskas R, McArthur JV. Co-selection of antibiotic and metal resistance. Trends Microbiol. 2006 Apr 1;14(4):176–82.

11. Foster TJ. Plasmid-determined resistance to antimicrobial drugs and toxic metal ions in bacteria. Microbiol Rev. 1983 Sep;47(3):361–409.

12. Ghosh A, Singh A, Ramteke PW, Singh VP. Characterization of Large Plasmids Encoding Resistance to Toxic Heavy Metals in *Salmonella abortus equi*. Biochem Biophys Res Commun. 2000 May 27;272(1):6–11.

13. Carattoli A. Resistance Plasmid Families in Enterobacteriaceae. Antimicrob Agents Chemother. 2009 Jun;53(6):2227–38.

14. Jackson CR, Davis JA, Frye JG, Barrett JB, Hiott LM. Diversity of Plasmids and Antimicrobial Resistance Genes in Multidrug-Resistant Escherichia coli Isolated from Healthy Companion Animals. Zoonoses Public Health. 2015;62(6):479–88.

15. McMillan EA, Gupta SK, Williams LE, Jové T, Hiott LM, Woodley TA, et al. Antimicrobial Resistance Genes, Cassettes, and Plasmids Present in Salmonella enterica Associated With United States Food Animals. Front Microbiol [Internet]. 2019 Apr 17 [cited 2024 Mar 30];10. Available from: https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2019.00832/full

16. Brockhurst MA, Harrison E, Hall JPJ, Richards T, McNally A, MacLean C. The Ecology and Evolution of Pangenomes. Curr Biol. 2019 Oct;29(20):R1094–103.

17. del Solar G, Giraldo R, Ruiz-Echevarría MJ, Espinosa M, Díaz-Orejas R. Replication and Control of Circular Bacterial Plasmids. Microbiol Mol Biol Rev. 1998 Jun;62(2):434–64.

18. Domingues CPF, Rebelo JS, Monteiro F, Nogueira T, Dionisio F. Harmful behaviour through plasmid transfer: a successful evolutionary strategy of bacteria harbouring conjugative plasmids. Philos Trans R Soc B Biol Sci. 2021 Nov 29;377(1842):20200473.

19. Virolle C, Goldlust K, Djermoun S, Bigot S, Lesterlin C. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. Genes. 2020 Nov;11(11):1239.

20. Mira A. The bacterial pan-genome: a new paradigm in microbiology. Int Microbiol. 2010;(13):45–57.

21. van Passel MW, Bart A, Luyf AC, van Kampen AH, van der Ende A. Compositional discordance between prokaryotic plasmids and host chromosomes. BMC Genomics. 2006 Feb 15;7(1):26.

22. Harrison E, Brockhurst MA. Plasmid-mediated horizontal gene transfer is a coevolutionary process. Trends Microbiol. 2012 Jun 1;20(6):262–7.

23. Wardell GE, Hynes MF, Young PJ, Harrison E. Why are rhizobial symbiosis genes mobile? Philos Trans R Soc B Biol Sci. 2021 Nov 29;377(1842):20200471.

24. Mira A, Klasson L, Andersson SGE. Microbial genome evolution: sources of variability. Curr Opin Microbiol. 2002 Oct 1;5(5):506–12.

25. Iwasaki W, Takagi T. Rapid Pathway Evolution Facilitated by Horizontal Gene Transfers across Prokaryotic Lineages. PLOS Genet. 2009 Mar 6;5(3):e1000402.

26. Valiente-Mullor C, Beamud B, Ansari I, Francés-Cuesta C, García-González N, Mejía L, et al. One is not enough: On the effects of reference genome for the mapping and subsequent analyses of short-reads. PLOS Comput Biol. 2021 Jan 27;17(1):e1008678.

610 27. Hurgobin B, Edwards D. SNP Discovery Using a Pangenome: Has the Single Reference
611 Approach Become Obsolete? Biology. 2017 Mar;6(1):21.

612 28. Onishi-Seebacher M, Korbel JO. Challenges in studying genomic structural variant formation
613 mechanisms: The short-read dilemma and beyond. BioEssays. 2011;33(11):840–50.

614 29. Alkan C, Sajjadian S, Eichler EE. Limitations of next-generation genome sequence assembly.
615 Nat Methods. 2011 Jan;8(1):61–5.

616 30. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A Bioinformatic Analysis of
617 Integrative Mobile Genetic Elements Highlights Their Role in Bacterial Adaptation. Cell Host
618 Microbe. 2020 Nov 11;28(5):767.

619 31. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile Genetic Elements Associated with
620 Antimicrobial Resistance. Clin Microbiol Rev. 2018 Aug;31(4):10.1128/cmr.00088-17.

621 32. Reeve W, O'Hara G, Chain P, Ardley J, Bräu L, Nandesena K, et al. Complete genome
622 sequence of Rhizobium leguminosarum bv. trifolii strain WSM1325, an effective
623 microsymbiont of annual Mediterranean clovers. Stand Genomic Sci. 2010 Jun 30;2(3):347.

624 33. Burghardt LT, diCenzo GC. The evolutionary ecology of rhizobia: multiple facets of
625 competition before, during, and after symbiosis with legumes. Curr Opin Microbiol. 2023
626 Apr 1;72:102281.

627 34. Epstein B, Tiffin P. Comparative genomics reveals high rates of horizontal transfer and
628 strong purifying selection on rhizobial symbiosis genes. Proc R Soc B Biol Sci. 2021 Jan
629 6;288(1942):20201804.

630 35. Cavassim MIA, Moeskjær S, Moslemi C, Fields B, Bachmann A, Vilhjálmsson BJ, et al.
631 Symbiosis genes show a unique pattern of introgression and selection within a Rhizobium
632 leguminosarum species complex. Microb Genomics [Internet]. 2020 Mar 16 [cited 2021 Apr
633 21];6(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7276703/

634 36. Young JPW, Jorrin B, Moeskjær S, James EK. Rhizobium brockwellii sp. nov., Rhizobium
635 johnstonii sp. nov. and Rhizobium beringeri sp. nov., three genospecies within the
636 Rhizobium leguminosarum species complex. Int J Syst Evol Microbiol. 2023;73(7):005979.

637 37. Weisberg AJ, Miller M, Ream W, Grünwald NJ, Chang JH. Diversification of plasmids in a
638 genus of pathogenic and nitrogen-fixing bacteria. Philos Trans R Soc B Biol Sci. 2021 Nov
639 29;377(1842):20200466.

640 38. Young JPW, Moeskjær S, Afonin A, Rahi P, Maluk M, James EK, et al. Defining the Rhizobium
641 leguminosarum Species Complex. Genes. 2021 Jan;12(1):111.

642 39. Reeve W, O'Hara G, Chain P, Ardley J, Bräu L, Nandesena K, et al. Complete genome
643 sequence of Rhizobium leguminosarum bv trifolii strain WSM2304, an effective

644 microsymbiont of the South American clover Trifolium polymorphum. Stand Genomic Sci.
645 2010 Feb 28;2(1):66–76.

646 40. Terpolilli J, Rui T, Yates R, Howieson J, Poole P, Munk C, et al. Genome sequence of
647 Rhizobium leguminosarum bv trifolii strain WSM1689, the microsymbiont of the one
648 flowered clover Trifolium uniflorum. Stand Genomic Sci. 2014 Jun 15;9(3):527–39.

649 41. Weese DJ, Heath KD, Dentinger BTM, Lau JA. Long-term nitrogen addition causes the
650 evolution of less-cooperative mutualists. Evolution. 2015;69(3):631–42.

651 42. Klinger CR, Lau JA, Heath KD. Ecological genomics of mutualism decline in nitrogen-fixing
652 bacteria. Proc R Soc B Biol Sci. 2016 Mar 16;283(1826):20152563.

653 43. Gordon BR, Klinger CR, Weese DJ, Lau JA, Burke PV, Dentinger BTM, et al. Decoupled
654 genomic elements and the evolution of partner quality in nitrogen-fixing rhizobia. Ecol Evol.
655 2016;6(5):1317–27.

656 44. Mazur A, Majewska B, Stasiak G, Wielbo J, Skorupska A. repABC-based replication systems
657 of Rhizobium leguminosarum bv. trifolii TA1 plasmids: Incompatibility and evolutionary
658 analyses. Plasmid. 2011 Jul 1;66(2):53–66.

659 45. Pinto UM, Pappas KM, Winans SC. The ABCs of plasmid replication and segregation. Nat Rev
660 Microbiol. 2012 Nov;10(11):755–65.

661 46. Dunny GM, Brickman TJ, Dworkin M. Multicellular behavior in bacteria: communication,
662 cooperation, competition and cheating. BioEssays. 2008;30(4):296–8.

663 47. Sachs JL, Ehinger MO, Simms EL. Origins of cheating and loss of symbiosis in wild
664 Bradyrhizobium. J Evol Biol. 2010;23(5):1075–89.

665 48. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et al. NCBI
666 prokaryotic genome annotation pipeline. Nucleic Acids Res. 2016 Aug 19;44(14):6614–24.

667 49. Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2:
668 Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic
669 Scale. Mol Biol Evol. 2021 Dec 1;38(12):5825–9.

670 50. Dewan I, Uecker H. Is the distribution of plasmid lengths bimodal? Plasmid. 2024 Jan 1;129–
671 130:102721.

672 51. diCenzo GC, Finan TM. The Divided Bacterial Genome: Structure, Function, and Evolution.
673 Microbiol Mol Biol Rev. 2017 Aug 9;81(3):e00019-17.

674 52. Garcillán-Barcia MP, Alvarado A, de la Cruz F. Identification of bacterial plasmids based on
675 mobility and plasmid population biology. FEMS Microbiol Rev. 2011 Sep 1;35(5):936–56.

53. Thomas CM. Paradigms of plasmid organization. Mol Microbiol. 2000;37(3):485–91.

54. Liao Q, Ren Z, Wiesler EE, Fuqua C, Wang X. A dicentric bacterial chromosome requires XerC/D site-specific recombinases for resolution. Curr Biol. 2022 Aug 22;32(16):3609-3618.e7.

55. Pérez-Oseguera Á, Cevallos MA. RepA and RepB exert plasmid incompatibility repressing the transcription of the *repABC* operon. Plasmid. 2013 Nov 1;70(3):362–76.

56. González V, Bustos P, Ramírez-Romero MA, Medrano-Soto A, Salgado H, Hernández-González I, et al. The mosaic structure of the symbiotic plasmid of Rhizobium etliCFN42 and its relation to other symbiotic genome compartments. Genome Biol. 2003 May 13;4(6):R36.

57. Pérez Carrascal OM, VanInsberghe D, Juárez S, Polz MF, Vinuesa P, González V. Population genomics of the symbiotic plasmids of sympatric nitrogen-fixing Rhizobium species associated with Phaseolus vulgaris. Environ Microbiol. 2016;18(8):2660–76.

58. Ling J, Wang H, Wu P, Li T, Tang Y, Naseer N, et al. Plant nodulation inducers enhance horizontal gene transfer of Azorhizobium caulinodans symbiosis island. Proc Natl Acad Sci. 2016 Nov 29;113(48):13875–80.

59. Funnell BE. Coordinating plasmid partition with bacterial chromosome segregation. Proc Natl Acad Sci. 2024 May 21;121(21):e2407081121.

60. Bethke JH, Ma HR, Tsoi R, Cheng L, Xiao M, You L. Vertical and horizontal gene transfer tradeoffs direct plasmid fitness. Mol Syst Biol. 2023 Feb 10;19(2):e11300.

61. Turner PE, Cooper VS, Lenski RE. Tradeoff Between Horizontal and Vertical Modes of Transmission in Bacterial Plasmids. Evolution. 1998 Apr;52(2):315–29.

62. Humphrey S, San Millán Á, Toll-Riera M, Connolly J, Flor-Duro A, Chen J, et al. Staphylococcal phages and pathogenicity islands drive plasmid evolution. Nat Commun. 2021 Oct 6;12(1):5845.

63. Peter S, Bosio M, Gross C, Bezdan D, Gutierrez J, Oberhettinger P, et al. Tracking of Antibiotic Resistance Transfer and Rapid Plasmid Evolution in a Hospital Setting by Nanopore Sequencing. mSphere. 2020 Aug 19;5(4):10.1128/msphere.00525-20.

64. Denamur E, Clermont O, Bonacorsi S, Gordon D. The population genetics of pathogenic Escherichia coli. Nat Rev Microbiol. 2021 Jan;19(1):37–54.

65. Wyres KL, Lam MMC, Holt KE. Population genomics of Klebsiella pneumoniae. Nat Rev Microbiol. 2020 Jun;18(6):344–59.

66. Matamoros S, van Hattem JM, Arcilla MS, Willemse N, Melles DC, Penders J, et al. Global phylogenetic analysis of Escherichia coli and plasmids carrying the mcr-1 gene indicates bacterial diversity but plasmid restriction. Sci Rep. 2017 Nov 10;7(1):15364.

67. Carroll AC, Wong A. Plasmid persistence: costs, benefits, and the plasmid paradox. Can J Microbiol. 2018 May;64(5):293–304.

68. Alonso-del Valle A, León-Sampedro R, Rodríguez-Beltrán J, DelaFuente J, Hernández-García M, Ruiz-Garbajosa P, et al. Variability of plasmid fitness effects contributes to plasmid persistence in bacterial communities. Nat Commun. 2021 May 11;12(1):2653.

69. Gano-Cohen KA, Wendlandt CE, Al Moussawi K, Stokes PJ, Quides KW, Weisberg AJ, et al. Recurrent mutualism breakdown events in a legume rhizobia metapopulation. Proc R Soc B Biol Sci. 2020 Jan 29;287(1919):20192549.

70. Heath KD, Stinchcombe JR. Explaining Mutualism Variation: A New Evolutionary Paradox? Evolution. 2014;68(2):309–17.

71. Negotiation, Sanctions, and Context Dependency in the Legume-Rhizobium Mutualism. [Internet]. [cited 2024 Jul 29]. Available from: https://www-journals-uchicago-edu.proxy2.library.illinois.edu/doi/epdf/10.1086/659997

72. Wenger AM, Peluso P, Rowell WJ, Chang PC, Hall RJ, Concepcion GT, et al. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat Biotechnol. 2019 Oct;37(10):1155–62.

73. Wick RR, Judd LM, Cerdeira LT, Hawkey J, Méric G, Vezina B, et al. Trycycler: consensus long-read assemblies for bacterial genomes. Genome Biol. 2021 Sep 14;22(1):266.

74. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019 May;37(5):540–6.

75. Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021 Feb;18(2):170–5.

76. Vaser R, Šikić M. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci. 2021 May;1(5):332–6.

77. Pierce NT, Irber L, Reiter T, Brooks P, Brown CT. Large-scale sequence comparisons with *sourmash* [Internet]. F1000Research; 2019 [cited 2024 Jul 1]. Available from: https://f1000research.com/articles/8-1006

78. Acman M, van Dorp L, Santini JM, Balloux F. Large-scale network analysis captures biological features of bacterial plasmids. Nat Commun. 2020 May 15;11(1):2452.

79. Bayliss SC, Thorpe HA, Coyle NM, Sheppard SK, Feil EJ. PIRATE: A fast and scalable pangenomics toolbox for clustering diverged orthologues in bacteria. GigaScience. 2019 Oct 1;8(10):giz119.

FIGURES AND TABLES

Table 1: Summary of *Rhizobium* plasmids featured in this study. For plasmids with more than one representative, we report the range of plasmid size, core length, number of core genes, nucleotide diversity (π), and average nucleotide identity (ANI). Rh type is based on *repABC* sequence as in previous studies (35).

Plasmid Type	N	present in	Size (Mbp)	Core length (Mbp)	Core genes	π	ANI	Rh Type
I	56	gsE	0.92-1.43	0.62	510	0.015	97.94	1
II	56	gsE	0.51-0.63	0.39	335	0.02	97.32	2
III	56	gsE	0.56-0.73	0.50	421	0.017	97.89	4
IV	58	gsE, gsB, Rht_717_N	0.26-0.43	0.05	49	0.021	95.25	4, 6, 7, 8
V	4	gsB	0.93	0.93	587	0	100	1
VI	4	gsB	0.67	0.67	419	0	100	2
VII	4	gsB	0.39	0.39	219	0	100	5
VIII	4	gsB	0.35	0.35	246	0	100	3
IX	1	Rht_717_N	1.10	-	-	-	-	1
X	1	Rht_717_N	0.61	-	-	-	-	2
XI	1	Rht_717_N	0.31	-	-	-	-	3
XII	1	Rht_717_N	0.29	-	-	-	-	9
XIII	1	Rht_717_N	0.21	-	-	-	-	4

Figure 1: Weighted, undirected network of all 257 plasmids from a population of 62 clover-associated natural isolates of *Rhizobium*. Plasmids (nodes) were grouped by Jaccard-Index similarity > 0.1 into types. Darker edges indicate more pairwise similarity between nodes. Nodes were colored blue (type I), red (type II), orange (Type III), and green (Type IV) and shaded by subgroup for type I and IV plasmids. Plasmid types with few representatives (V-VIII and singletons) were left unshaded.

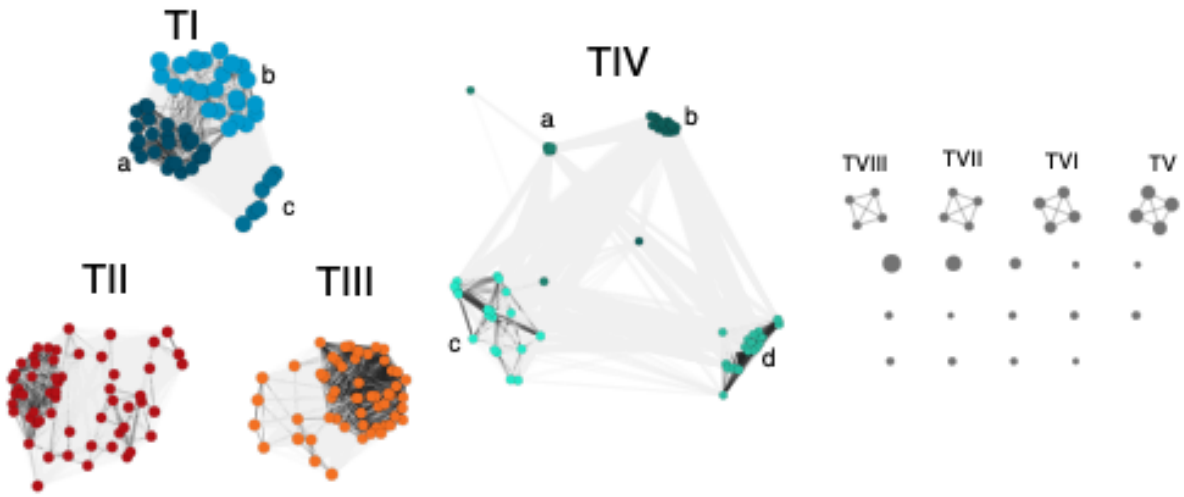
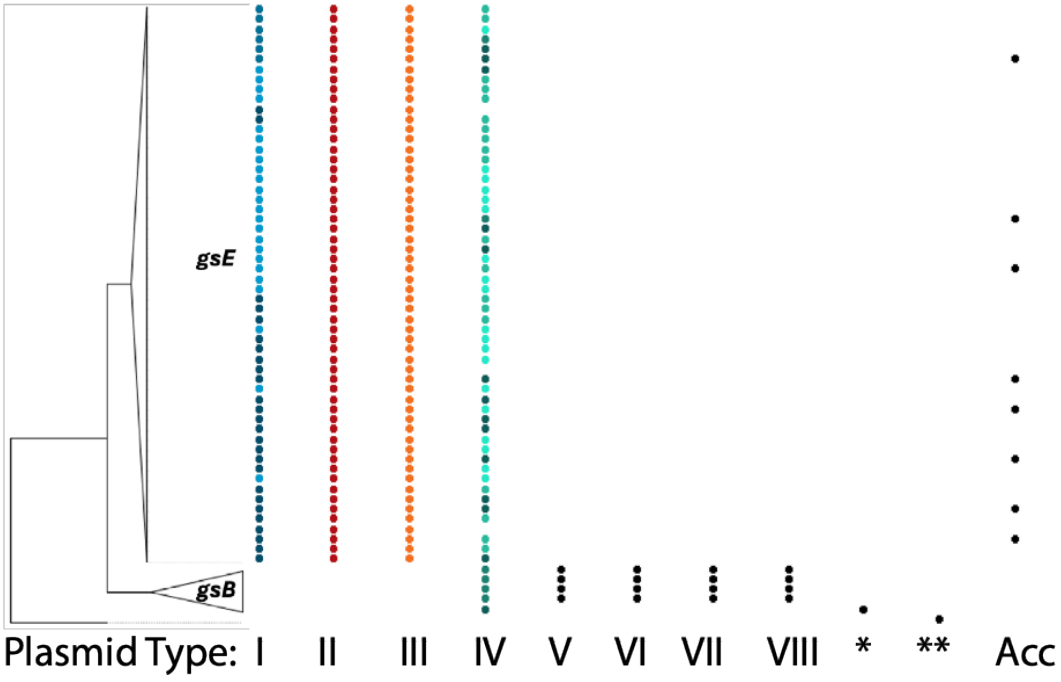
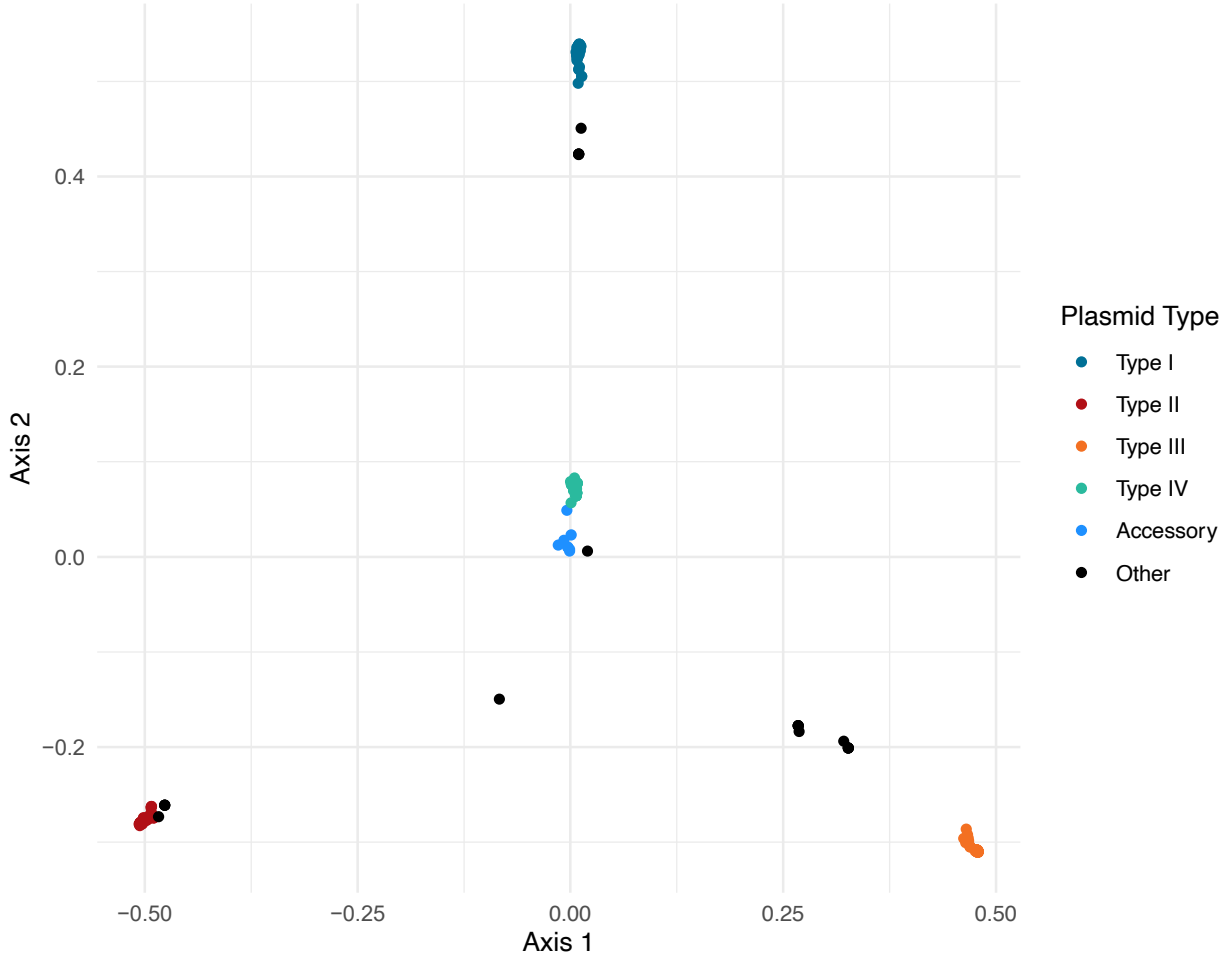


Figure 2: Distribution of plasmid types across the concatenated core chromosomal tree from a natural population of 62 clover-associated strains of *Rhizobium*. Tree tips were collapsed to highlight differences in plasmid composition between genospecies (*gsE* and *gsB*). Subclades within types are represented by shape for Type I: circles (Type I-a), squares (Type I-b), and triangles (Type I-c) and Type IV: circles (Type IV-a), squares (Type IV-b), triangles (Type IV-c), and diamonds (type IV-d). Six replicons from strain 717_N (*) and a single 1.4 Mbp replicon from strain 773_N (**) are each represented by a single dot.

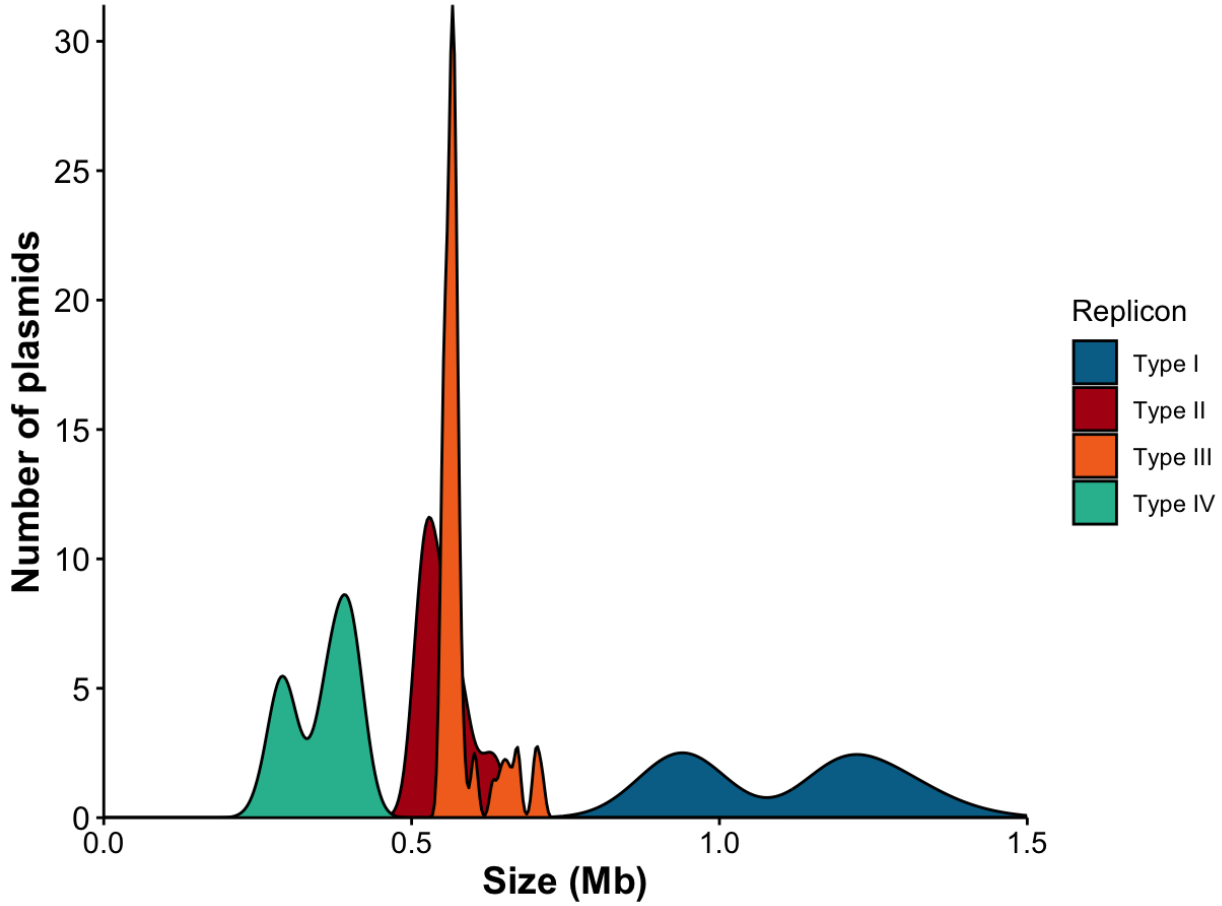


768 Figure 3: Principal Coordinate Analysis (PCoA) based on the presence-absence matrix of
769 orthologous gene clusters in the *Rhizobium* plasmidome of 62 natural isolates. Each point
770 represents a single plasmid, with points closer together indicating more shared gene content.



771

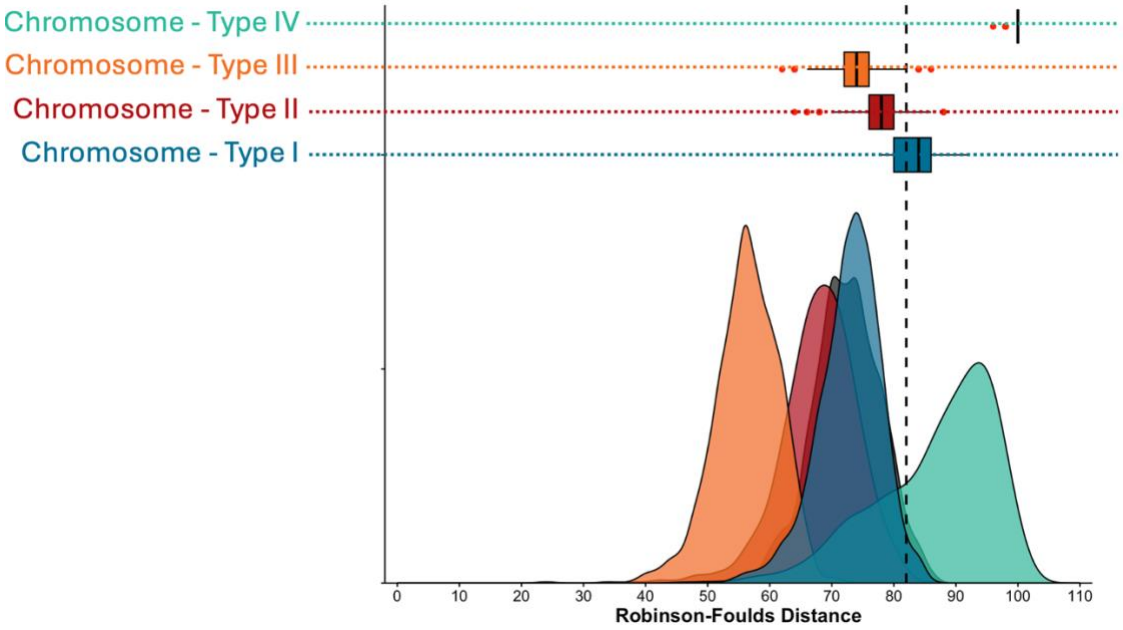
772 Figure 4: Distribution of plasmid size for each of the four major plasmid types (I-IV) commonly
773 found in *Rhizobium* genospecies (*gsE*). Distributions of types I-III include plasmids from all 56
774 *gsE* strains, whereas the type IV distribution includes 58 type IV plasmids from *gsE* and *gsB*
775 strains, as well as 717_N and 773_N.



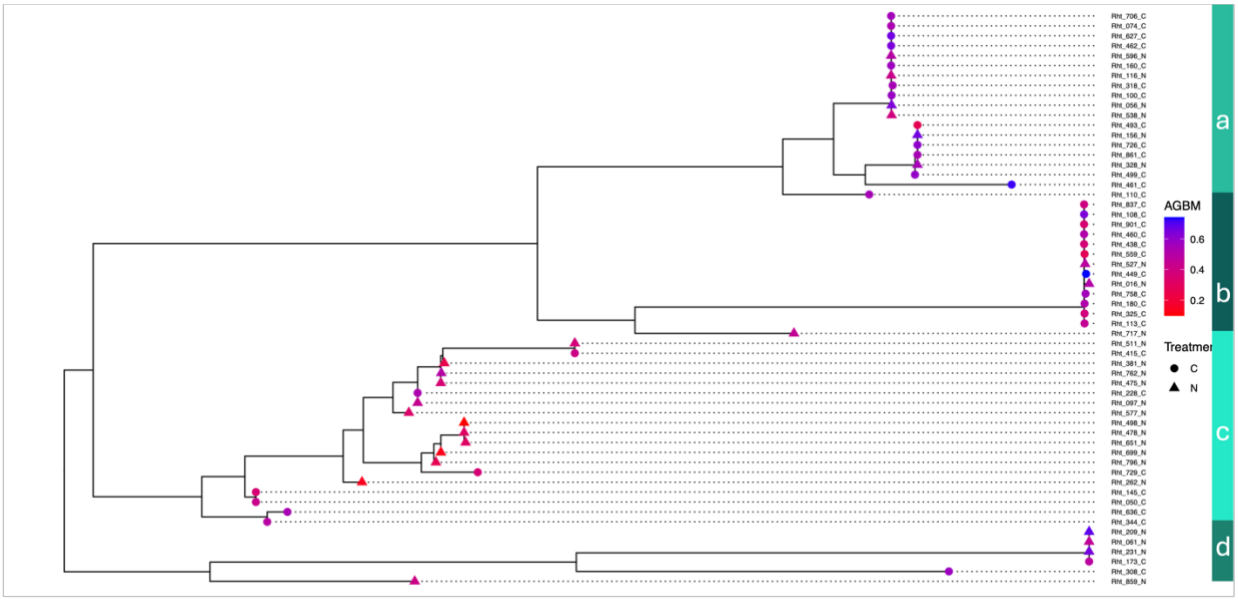
776

777

Figure 5: GRF distance distributions of within-element and across-element gene trees. Values closer to 0 represent higher levels of across-tree congruence, while values closer to 100 represent higher levels of incongruence among trees. Within-element distances for the chromosome (gray), Type I plasmid (blue), Type II plasmid (red), Type III plasmid (orange), and Type IV plasmid (green) are shown as geometric smoothed distributions along the bottom. GRF distance distributions between each plasmid and the chromosome are represented as horizontal box and whisker plots. The vertical dotted gray line represents the 95th percentile of the within-chromosome distribution.

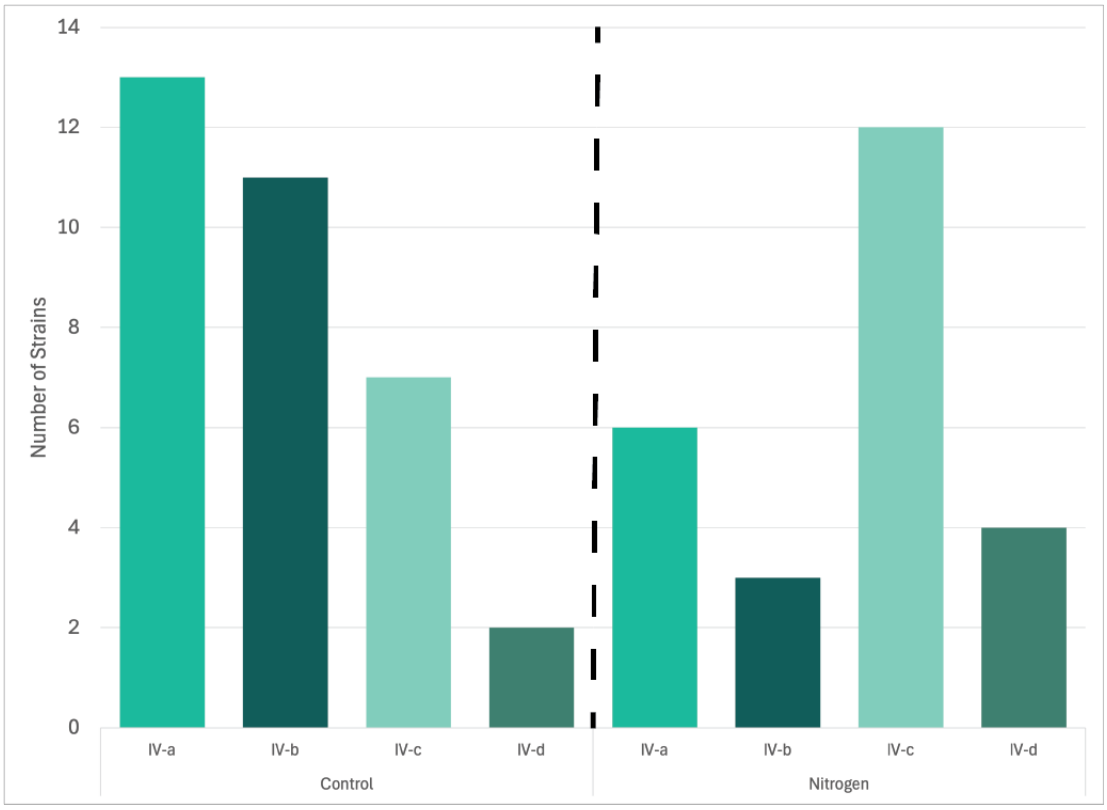


790 Figure 6: (A) Phylogenetic tree of the type IV plasmid (pSym) based on concatenated core gene
791 alignment, with nodes labelled with aboveground biomass (AGBM as color gradient) and plot
792 treatment of origin (C for control or N for N-fertilized). Type IV subclades (a-d) as in Figure 1
793 are indicated in shades of green on the right. (B) Number of strains from each Type IV sub-clade
794 isolated from either control (left) or N-fertilized (right) plots from the KBS LTER. Any nodes
795 with < 85% bootstrap support were collapsed.
796 A.

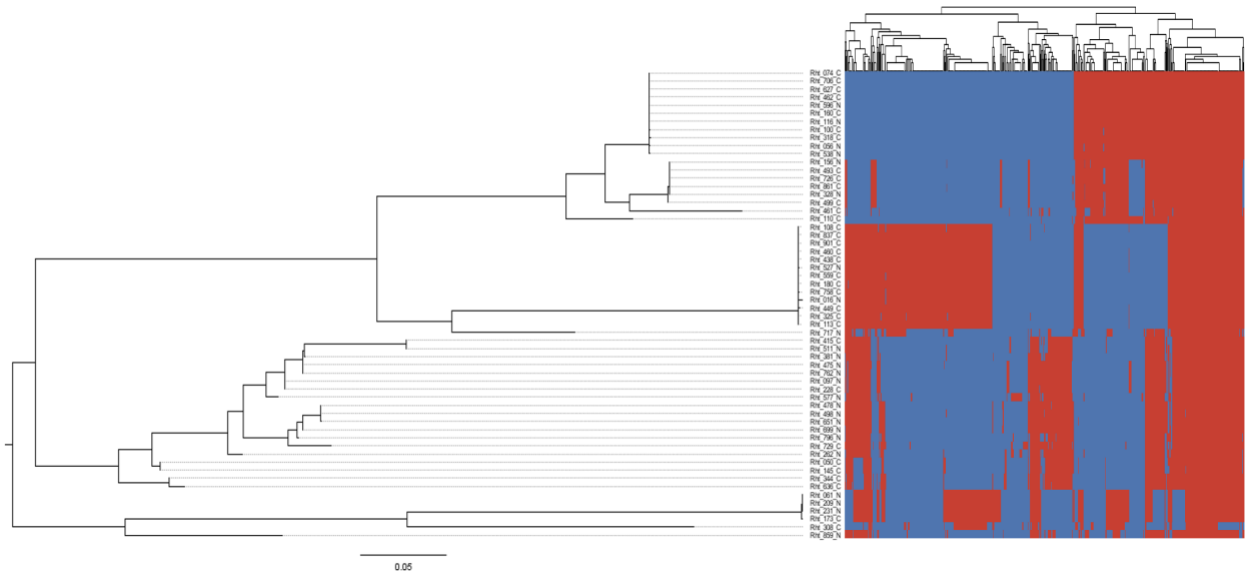


797
798

B.



C.



803 **SUPPLEMENTARY FIGURES AND TABLES**

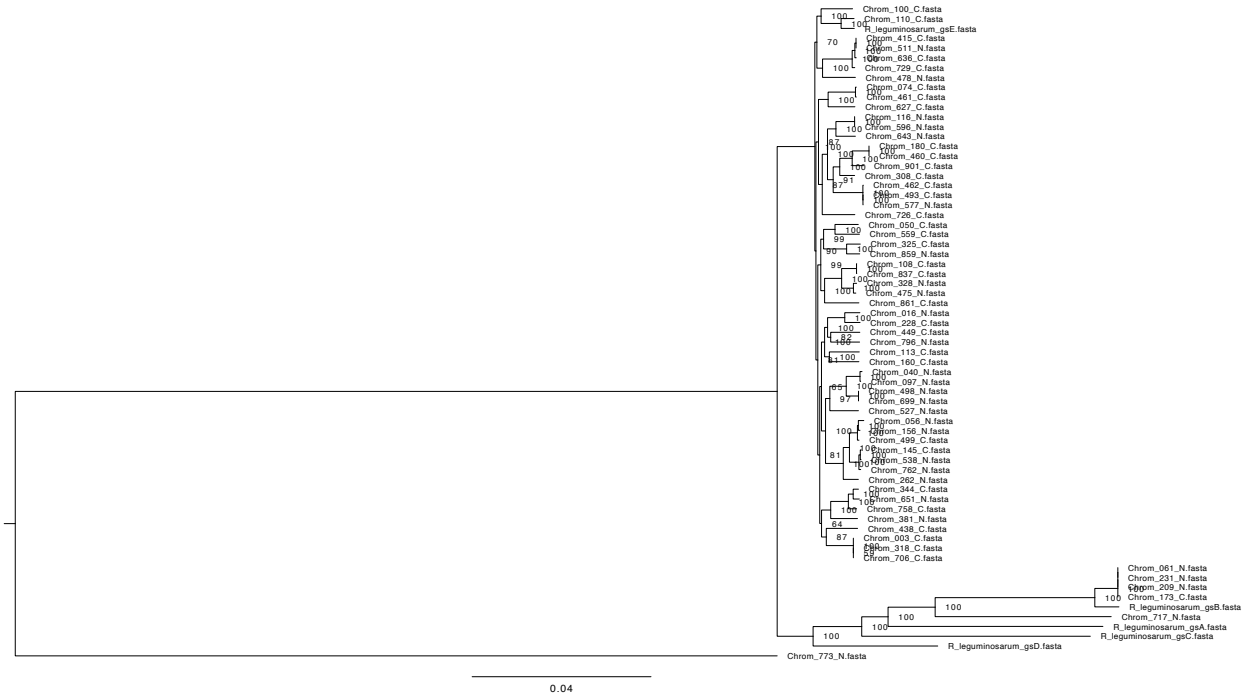
804 Table S1: Size and strain of origin of accessory plasmids in the pangenome.

Accessory plasmid's strain	Size (bp)
Rht_108_C	72609
Rht_180_C	122968
Rht_228_C	73852
Rht_318_C	188843
Rht_527_N	277757
Rht_538_N	245492
Rht_758_C	22509
Rht_859_N	208987

805

806

807 Figure S1: Chromosomal phylogeny of clover-associated *Rhizobium* strains from the KBS LTER
808 population alongside one representative each from five genospecies (A-E) of the *Rhizobium*
809 species complex, as described in Cavassim et al. 2020. Tree was manually rooted on strain
810 717_N. Any nodes with < 85% bootstrap support were collapsed.



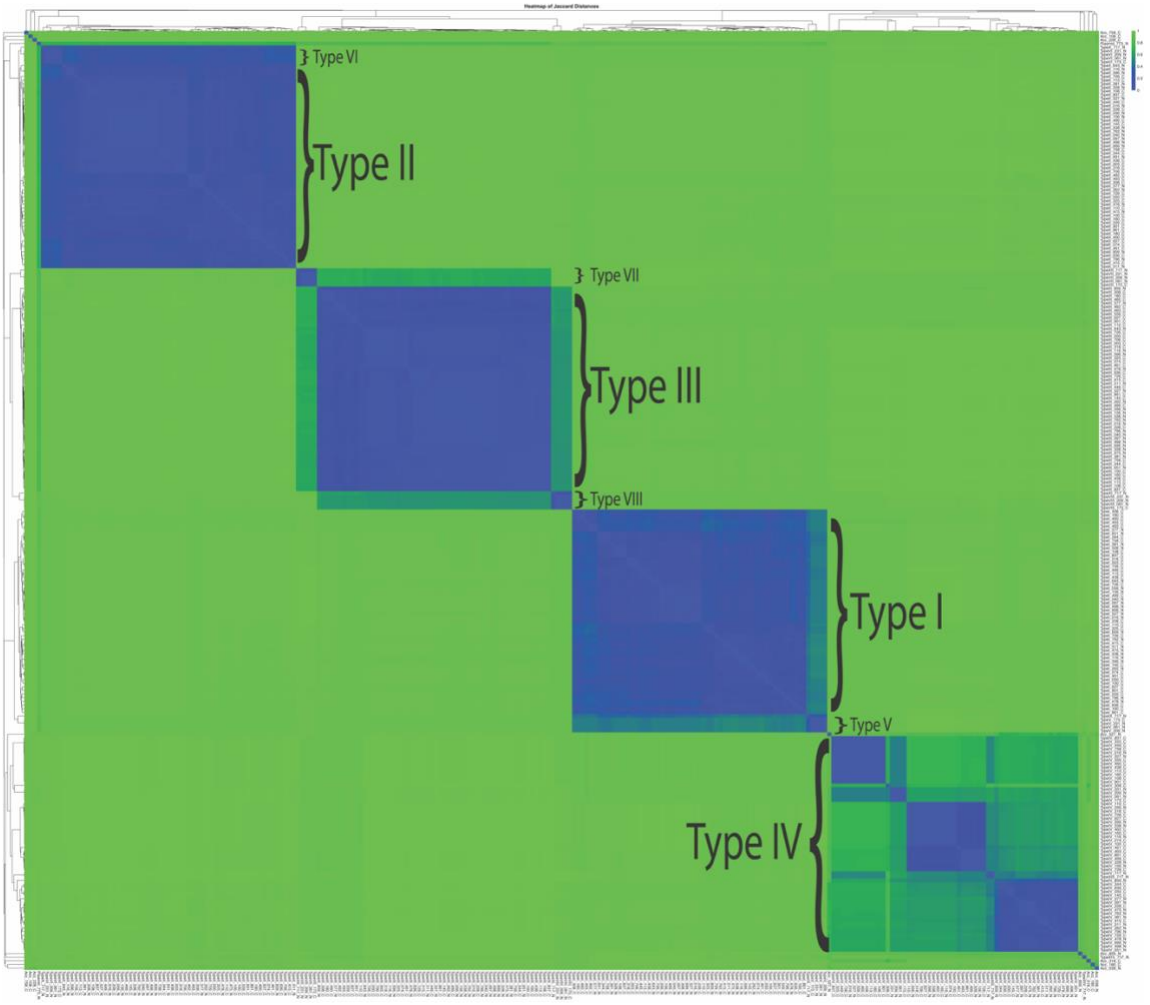
811

812

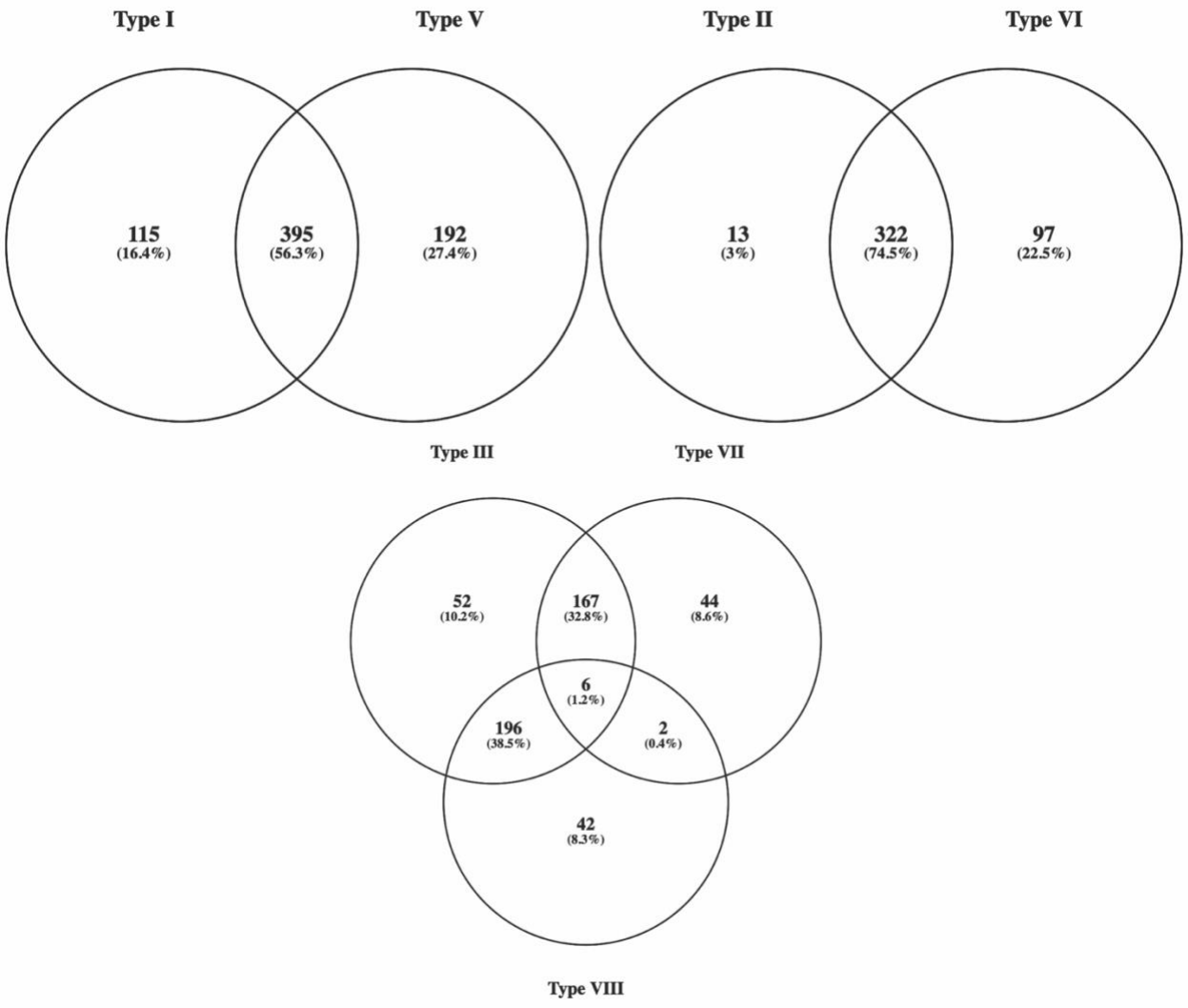
813

Figure S2: (A) Heatmap of pairwise Jaccard distances based on gene presence-absence between all plasmids from our study, highlighting similarity in gene content between plasmid types from *gsE* and *gsB*: I with V, II with VI, and III with VII/VIII. (B) Venn diagram of genes shared across plasmid types III, VII, and VIII. (C) local alignment of a representative of type III (strain 110_C) with representatives of types VII and VIII (both from 231_N).

A.

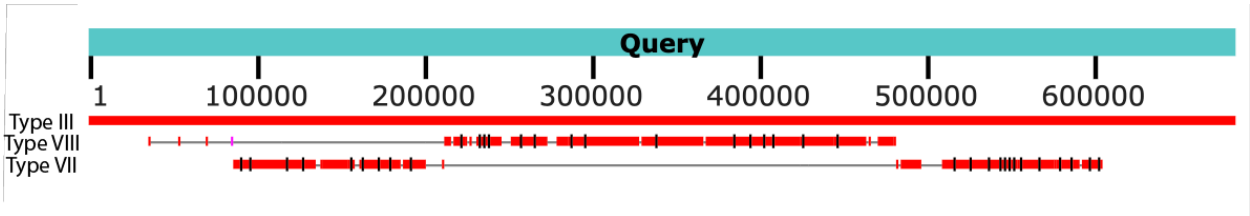


821 B.



822

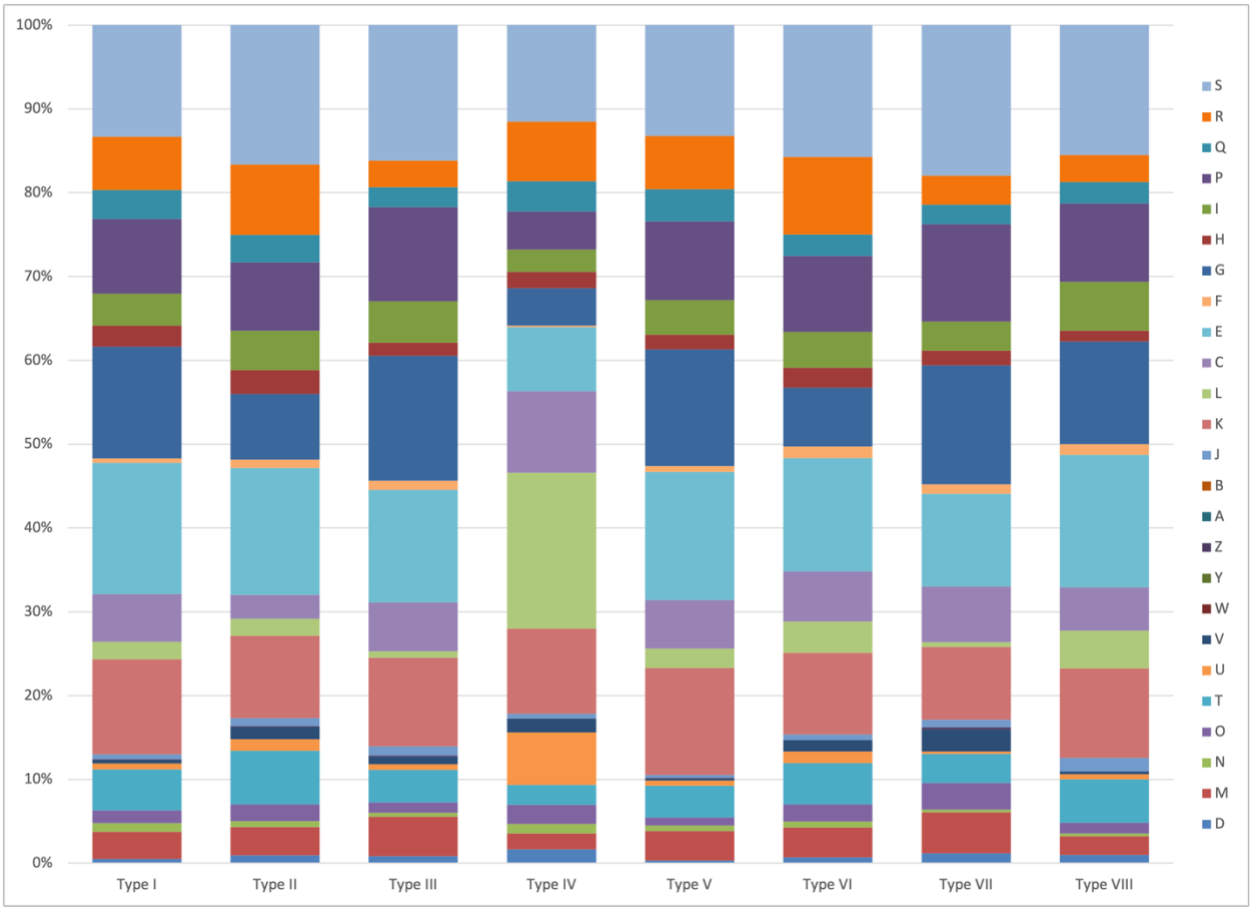
823 C.



824

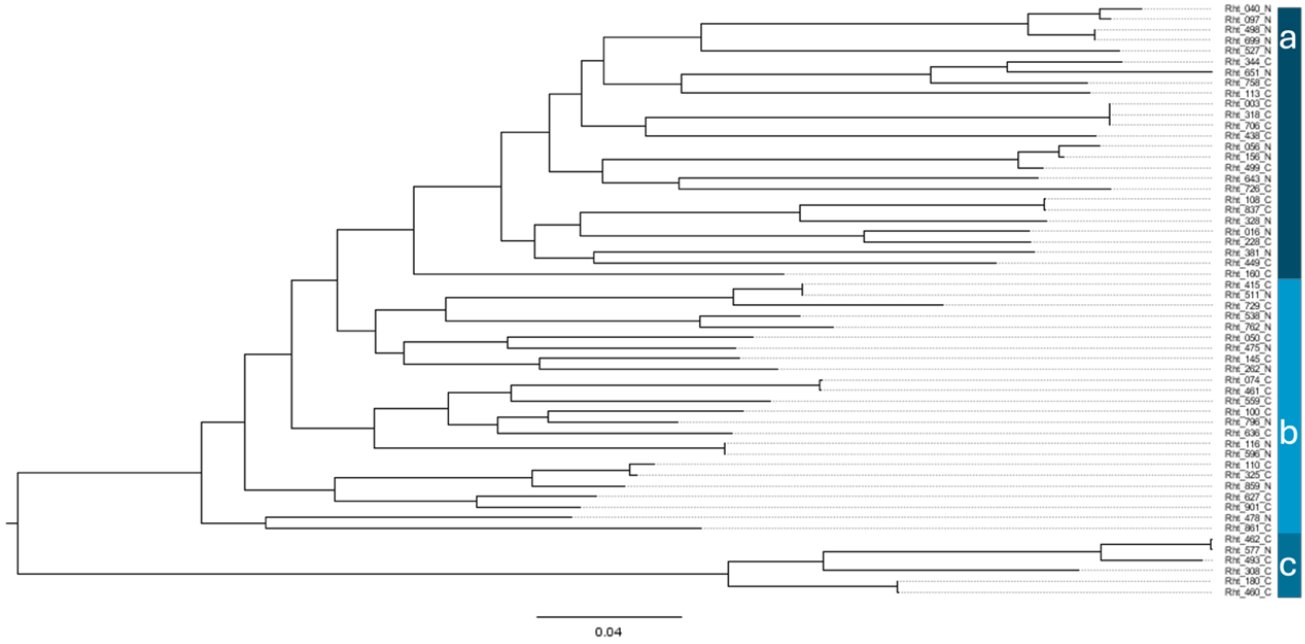
825

Figure S3: Distribution of COG functions across the plasmids of clover-associated *Rhizobium* from a natural population: (S) Function unknown, (R) General function prediction only, (Q) Secondary metabolites biosynthesis, transport and catabolism, (P) Inorganic ion transport and metabolism, (I) Lipid transport and metabolism, (H) Coenzyme transport and metabolism, (G) Carbohydrate transport and metabolism, (F) Nucleotide transport and metabolism, (E) Amino acid transport and metabolism, (C) Energy production and conversion, (L) Replication, recombination and repair, (K) Transcription, (J) Translation, ribosomal structure and biogenesis, (B) Chromatin structure and dynamics, (A) RNA processing and modification, (Z) Cytoskeleton, (Y) Nuclear structure, (W) Extracellular structures, (U) Intracellular trafficking, secretion, and vesicular transport, (T) Signal transduction mechanisms, (O) Posttranslational modification, protein turnover, chaperones, (N) Cell motility.



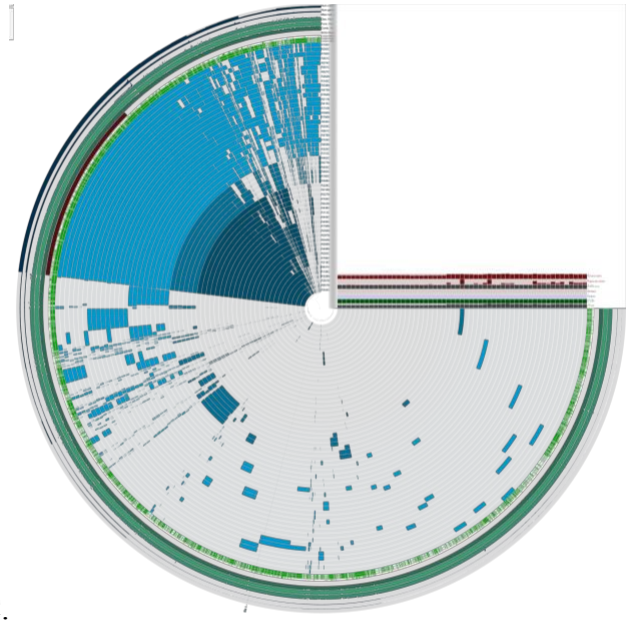
838 Figure S4: (A) Phylogenetic tree of the concatenated core genes of the type I plasmid, showing
839 the relationships among the three subclades clades as in Figure 1. Any nodes with < 85%
840 bootstrap support were collapsed. (B) Gene presence-absence anv'io plot of Type I plasmids,
841 colored by clade and showing that larger subclades I-b and I-c possess distinct large insertions.
842 (C) Pangenome graph view of the Type I plasmid population, showing distinct insertions in
843 larger subclades I-b and I-c occur in the same genomic location. Aligned sequences are in black,
844 and inverted regions are in red.

845 A.

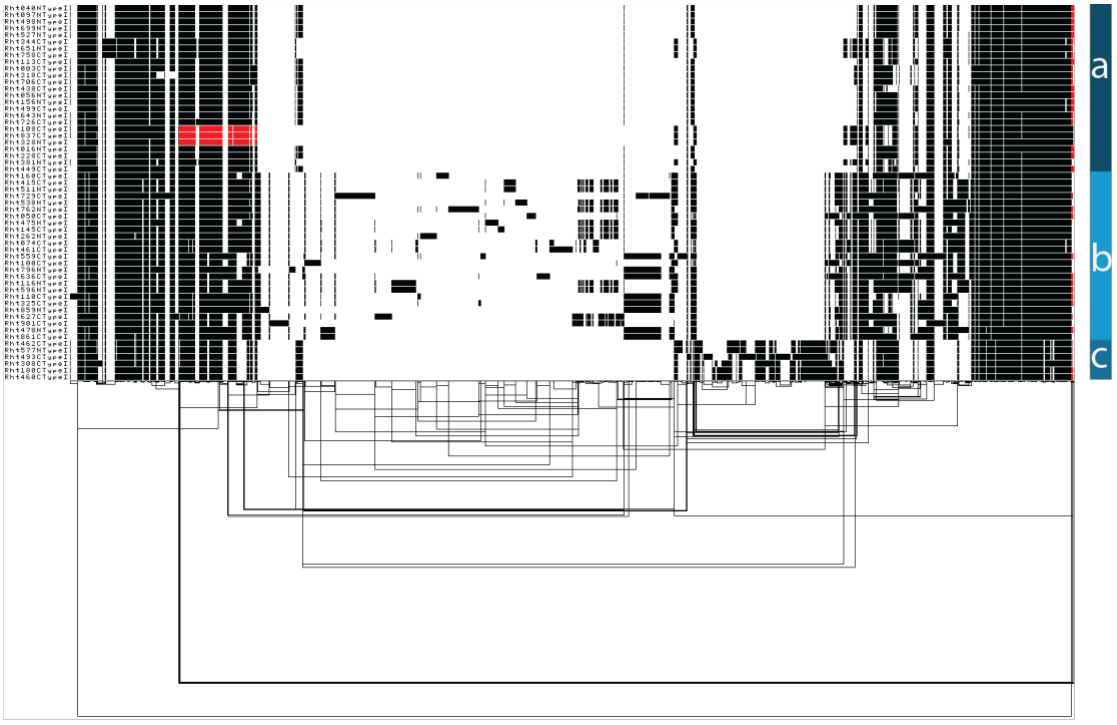


846

847 B.



848 C.

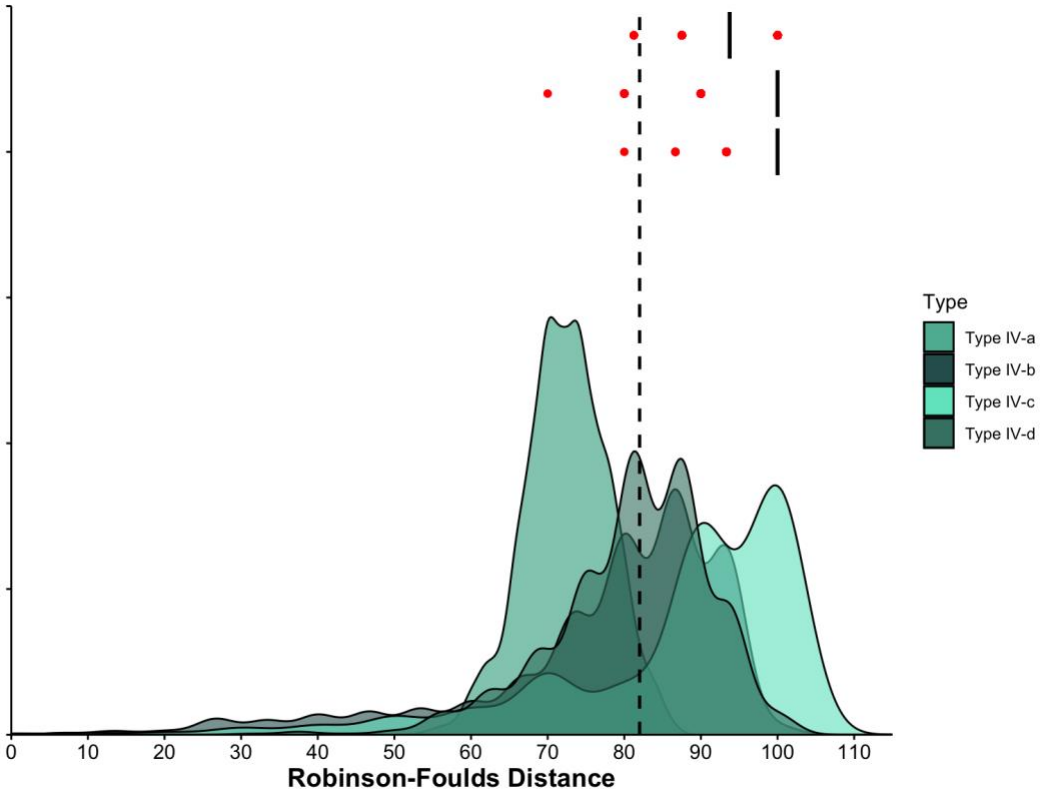


849

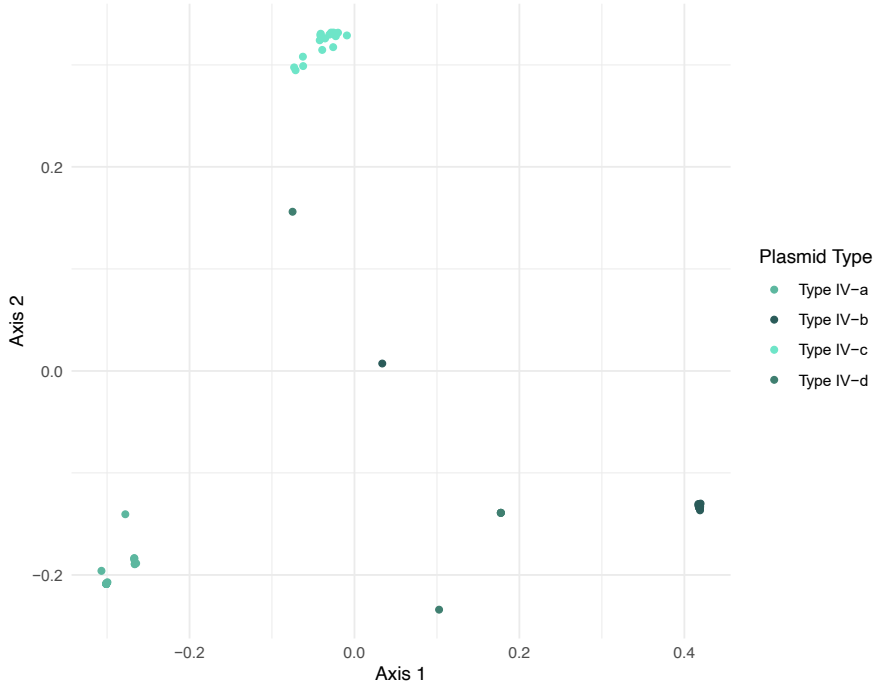
850

Figure S5: (A) GRF distances within Type IV subclades (smoothed distributions) and between subclades and the chromosome (box and whisker plots). (B) Principal Coordinate Analysis (PCoA) of orthologous gene clusters in the type IV plasmids in *Rhizobium*. (C) Tree of type IV pSym showing orthologous genes that are present (red) or absent (blue) across strains. (D) Venn diagram showing the number of core genes in different combinations of type IV subclades.

A.

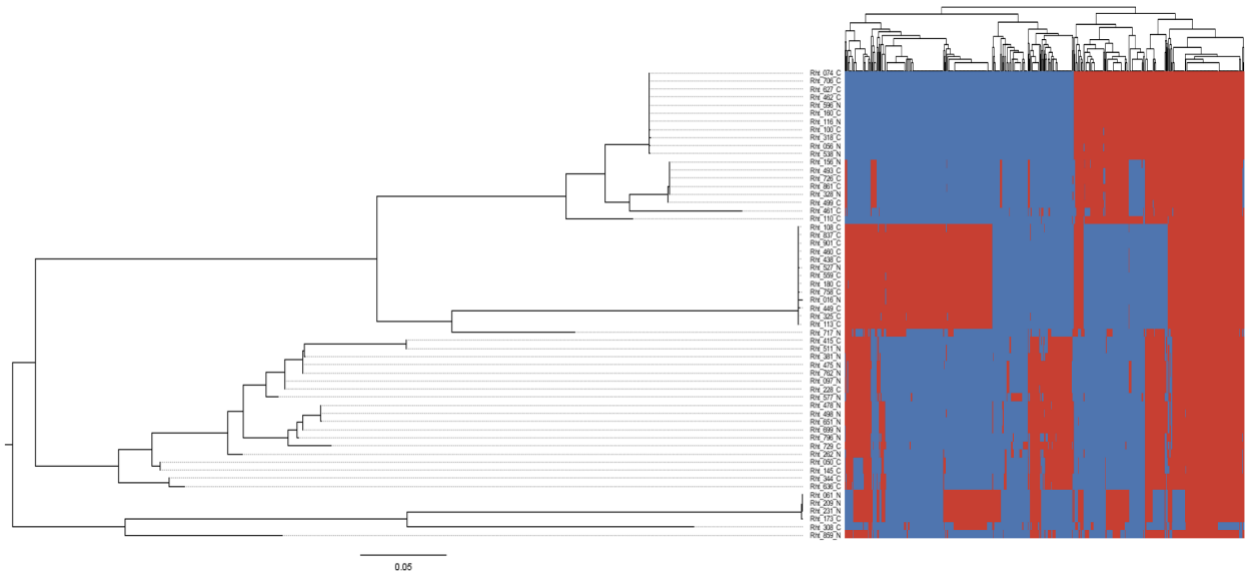


B.



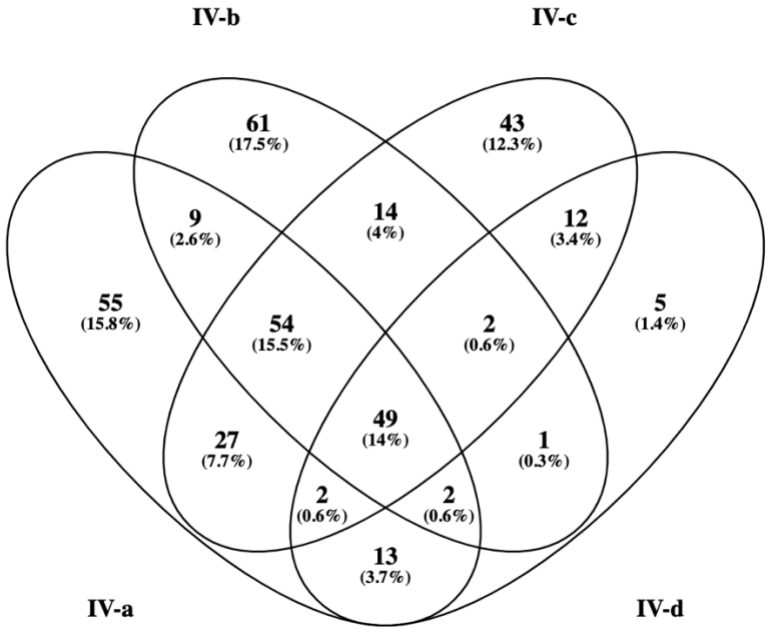
859

860 C.



861

862 D.



863