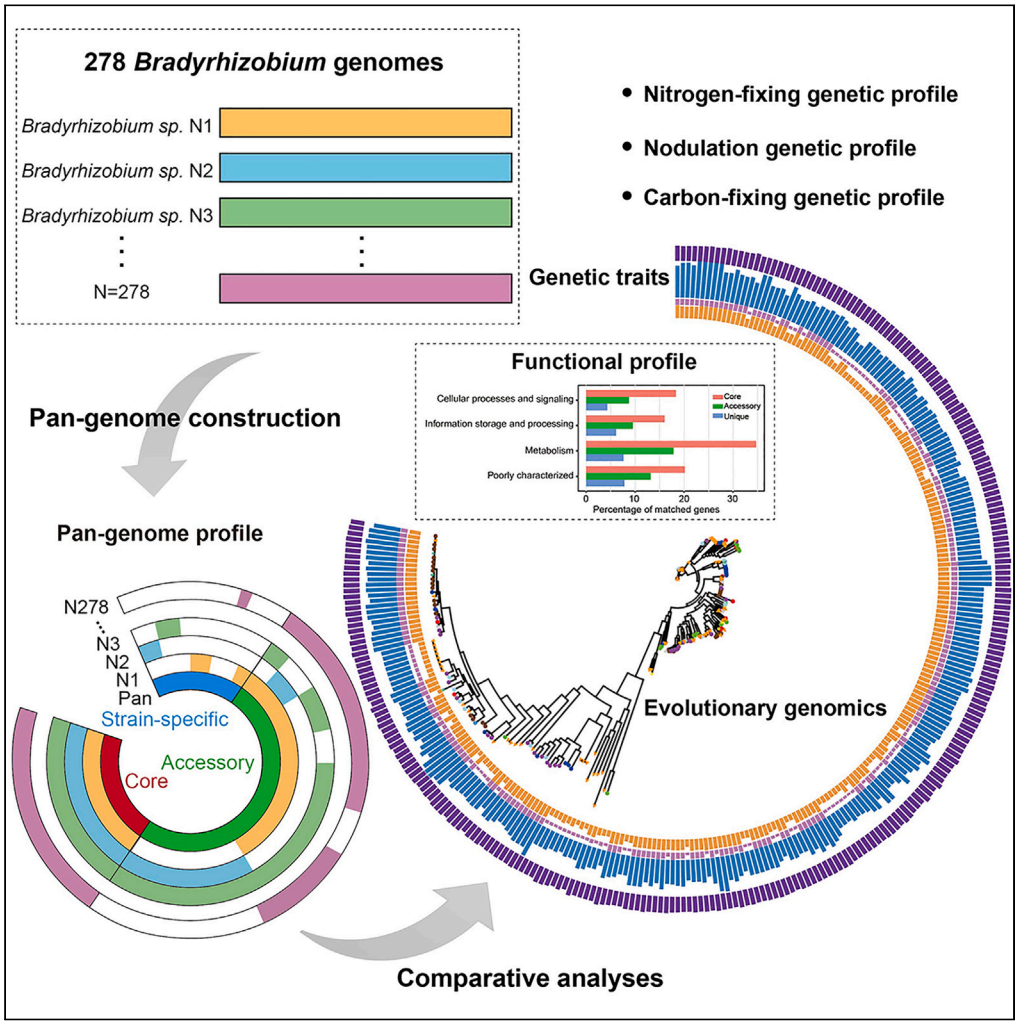
iScience

CellPress
OPEN ACCESS

Article

Comparative genomics analysis reveals genetic characteristics and nitrogen fixation profile of *Bradyrhizobium*



Chaofang Zhong, Gang Hu, Cong Hu, Chaohao Xu, Zhonghua Zhang, Kang Ning

gxtczzh@126.com (Z.Z.)
ningkang@hust.edu.cn (K.N.)

Highlights

Comparative genomics on a collection of 278 *Bradyrhizobium* strains was performed

The genomic diversity of *Bradyrhizobium* represents an "open" pan-genome model

Bradyrhizobium strains have dramatically different symbiotic nitrogen fixation genes

Most *Bradyrhizobium* strains from soil lack the *nif* and *nod* gene clusters

Zhong et al., iScience 27, 108948
February 16, 2024
https://doi.org/10.1016/j.isci.2024.108948

iScience

CellPress
OPEN ACCESS

Article

Comparative genomics analysis reveals genetic characteristics and nitrogen fixation profile of *Bradyrhizobium*

Chaofang Zhong,[1] Gang Hu,[1] Cong Hu,[1] Chaohao Xu,[1] Zhonghua Zhang,[1,3,*] and Kang Ning[2,*]

SUMMARY

***Bradyrhizobium* is a genus of nitrogen-fixing bacteria, with some species producing nodules in leguminous plants. Investigations into *Bradyrhizobium* have recently revealed its substantial genetic resources and agricultural benefits, but a comprehensive survey of its genetic diversity and functional properties is lacking. Using a panel of various strains (N = 278), this study performed a comparative genomics analysis to anticipate genes linked with symbiotic nitrogen fixation. *Bradyrhizobium*'s pan-genome consisted of 84,078 gene families, containing 824 core genes and 42,409 accessory genes. Core genes were mainly involved in crucial cell processes, while accessory genes served diverse functions, including nitrogen fixation and nodulation. Three distinct genetic profiles were identified based on the presence/absence of gene clusters related to nodulation, nitrogen fixation, and secretion systems. Most *Bradyrhizobium* strains from soil and non-leguminous plants lacked major *nif/nod* genes and were evolutionarily more closely related. These findings shed light on *Bradyrhizobium*'s genetic features for symbiotic nitrogen fixation.**

INTRODUCTION

As important members of the Rhizobia, bacterial species from the genus *Bradyrhizobium* are characterized by their diverse biochemical functions, including biological nitrogen fixation and carbon fixation.[1,2] They are one of the most prevalent groups in the soil, and many representatives have been identified as being involved in the development and adaptation of host plants.[3] This genus belongs to the rhizobia family, which comprises commercially important soil bacteria that interact with the roots of leguminous plants to fix atmospheric nitrogen.[4,5] The nitrogen-fixing *Bradyrhizobium* symbiosis leads to the formation of plant nodules, which boosts nitrogen availability to plants and has agricultural applications.[6] The genus *Bradyrhizobium* is currently encompassing more than 70 species that have been isolated worldwide. Some are symbiotic, able to fix nitrogen and form nodules on legumes, while others are free-living and unable to form nodules.[7,8] Variation in the genetic composition of *Bradyrhizobium* strains has been documented with respect to symbiotic associations and nitrogen fixation.[9] The soybean symbionts *Bradyrhizoium elkanii, B radyrhizoium Japonicum*, and *B radyrhizoium diazoefficiens* can nodulate soybean and have a remarkable genetic diversity.[10] *Bradyrhizobium* sp. S23321 has been described as free-living and unable to form nodules.[5]

Bradyrhizobium members have excellent prospects for agricultural research, emphasizing the importance of studying their genomes and identifying the variety embedded in this genus. There is a significant level of diversity knowledge within the *Bradyrhizobium* genus, as shown by studies on well-adapted *Bradyrhizobium* species.[11,12] Numerous *Bradyrhizobium* nod and nif factors have recently been characterized, which are responsible for nod-independent nodulation formation and nitrogen fixation, respectively.[13] These nodulation or nitrogen fixation genes are also highly variable among different bradyrhizobial species.[14] For example, *Bradyrhizobium* sp. G22 lacked both the nodulation gene *nodABC* and the nitrogen fixation gene *nifDKH* that were present in *B. diazoefficiens* USDA 110.[15] According to comparative genomic analysis research, nod genes and nif factors are species and host specific.[16,17] Although legumes of the genus *Bradyrhizobium* are known to be capable of forming symbiotic nodules, the genetic repertoire for nitrogen/carbon fixation is still under investigation. Additionally, studies have shown that nod and *nif* factors are typically arranged as gene clusters that govern nodulation and nitrogen fixation. Recent studies have shown that while some nonsymbiotic *Bradyrhizobium* isolates lacking the nod gene cluster are unable to nodulate legumes but are nevertheless common in soils.[7,18] These soil isolates have been demonstrated to be deficient in both nodulation and nitrogen fixation, and their genomes lack functional genes related to both processes.[7] According to a recent study, some *Bradyrhizobium* can participate in nod-independent nodulation via the T3SS pathway and do not require nodulation components for nodulation.[15] Even though some

[1]Key Laboratory of Wildlife Evolution and Conservation in Mountain Ecosystem of Guangxi, College of Environmental and Life Sciences, Nanning Normal University, Nanning 530001, China
[2]Key Laboratory of Molecular Biophysics of the Ministry of Education, Hubei Key Laboratory of Bioinformatics and Molecular-imaging, Department of Bioinformatics and Systems Biology, College of Life Science and Technology, Huazhong University of Science and Technology, Wuhan, Hubei 430074, China
[3]Lead contact
*Correspondence: gxtczzh@126.com (Z.Z.), ningkang@hust.edu.cn (K.N.)
https://doi.org/10.1016/j.isci.2024.108948

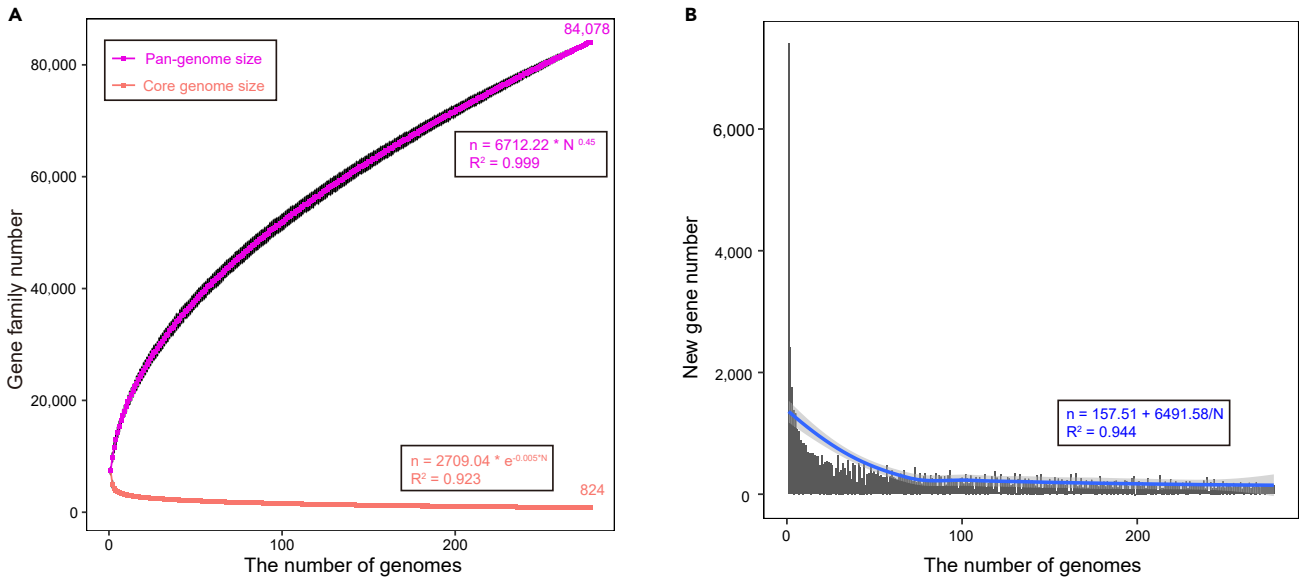


Figure 1. Prediction of *Bradyrhizobium* pan- and core genome size

(A) Genome size evolution curves of the pan-genome (purple) and core-genome (red).

(B) Curve (blue) for the number of new genes with an increase in the number of *Bradyrhizobium* genomes.

nonsymbiotic photosynthetic *Bradyrhizobium* from legumes lack the nod gene, they can fix nitrogen.[19] The genetic properties of nod-dependent or nod-independent nitrogen also vary on the host type, which requires the determination of their genomic variation.

The *Bradyrhizobium* genomes' accessibility has made it possible to characterize genetic diversity in great detail. Recent research has identified the processes that differ in symbiotic nitrogen fixation in *Bradyrhizobium* strains with varied genotypes and phenotypes.[20] *Bradyrhizobium arachidis* from *Arachis hypogaea* and *Sophora flavescens* have been compared for symbiotic nitrogen fixation, but the genetic diversity within the genus is still under-represented.[21] With the availability of a large number of *Bradyrhizobium* genomes, a pan-genome investigation would certainly be beneficial in uncovering intra-species diversity, such as nitrogen fixation capacity. As the primary determinants of nodulate legumes are frequently encoded by genes that are not shared by all strains, capturing functional gene differences is particularly relevant for *Bradyrhizobium*.

In this study, we performed a detailed comparative genomics analysis of 278 *Bradyrhizobium* strains. We constructed the pan-genome of *Bradyrhizobium*, allowing to characterize different functional gene sets within the genus. We explored the role of core, accessory, and unique genomes to obtain in-depth comparisons of the functional potential of representative strains. We developed a novel phylogenetic understanding of the genus, and also elucidated the genomic variation of symbiosis and nitrogen fixation in *Bradyrhizobium*, and further explained the genetic diversity of *Bradyrhizobium* assimilating carbon by fixation. The results of this study will shed new light on the nitrogen and carbon-fixing lifestyle of *Bradyrhizobium*.

RESULTS AND DISCUSSION

Bradyrhizobium open pan-genome

To study the genomic diversity among the members of the genus *Bradyrhizobium*, 278 strains with complete genomes or draft assemblies (Table S1) from 70 species were downloaded from NCBI for comparative genomics analysis. The OrthoMCL algorithm was applied to protein sequences from all 278 genomes to generate the *Bradyrhizobium* pan-genome, which contained a total repertoire of 84,078 gene families (non-redundant genes). The pan-genome of *Bradyrhizobium* consisted of total 824 core gene families (genes shared by all strains), 42,409 accessory gene families (genes shared by two or more strains), and 40,845 unique gene families (genes unique to individual strains). *Bradyrhizobium* genomes were found to be remarkably diverse, with only a minimal core genome component (0.98% of the pan-genome shared by the genus, and the remaining 99.02% being variable genes, of which up to 48.58% of which were strain-specific genes).

Bradyrhizobium genomes were highly heterogeneous, with an open pan-genome. We constructed accumulation curves for pan- and core genes against the sampled genomes. The power-law regression based on Heap's law was used to determine the fitted curve of pan-genome accumulation. The developed pan-genome fitting model was represented as $n = 6712.22 \times N^{0.45}$, where n is the predicted number of genes (pan-genome size) for a given number of genomes and N is the number of genomes. The *Bradyrhizobium* pan-genome could be deemed 'open' with a power-law coefficient α of 0.45, and demonstrated a progressive expansion by adding new genomes, with no obviously sharp plateau when adding up to 278 strain genomes (Figure 1A). The core genome analysis of 278 genomes converged to a core gene subset of 824 genes, and the number of shared genes decreased when more genomes were included (Figure 1A). The core genome fitting curve

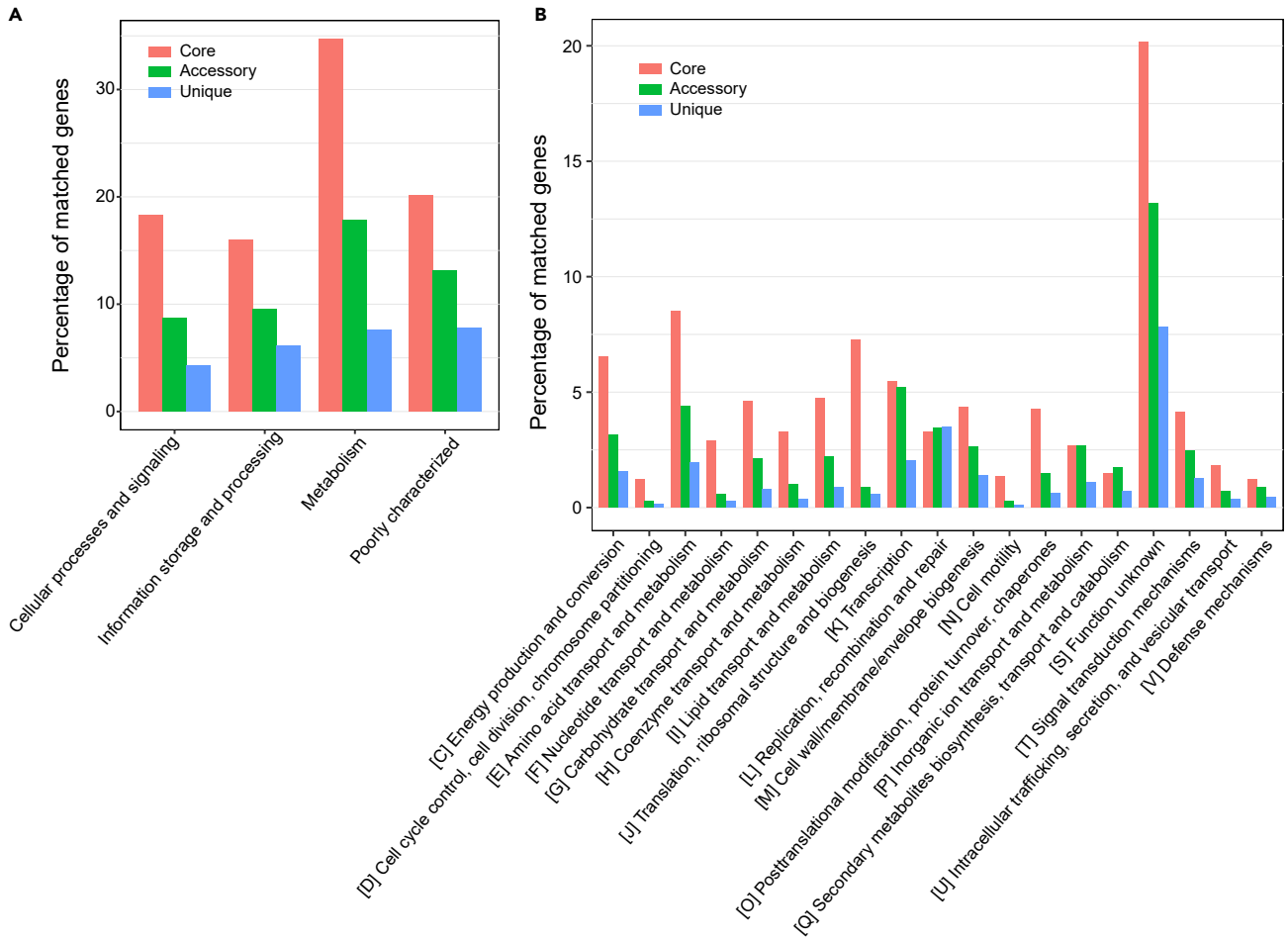


Figure 2. Classifications of core, accessory and unique genes based on COG categories

(A) Distribution in major COG categories.

(B) Type of gene groups by detailed COG categories. The horizontal coordinate was the content of each COG classification and the vertical coordinate was the percentage of matched genes. Only the proportion of genes that could be annotated to COG was shown.

followed the exponential regression decay $n = 2709.04e^{-0.005 \times N}$. Furthermore, each genome acquired an average of 302 novel genes (Figure 1B). The number of novel genes and core genes gradually decreased as more strains were added, approaching saturation with the addition of 278 strains.

Functional characterization of the core, accessory and unique genes

To assess the functional diversity within *Bradyrhizobium* pan-genome, the COG analysis was employed. The core, accessory, and unique genes were mostly found in the COG functional categories of metabolism (7.63–34.71%), cellular processes and signaling (4.31–18.33%), information storage and processing (6.10–16.02%), and poorly characterized (7.81–20.15%). Approximately 7.8–20.15% of the core, accessory, and unique genes with poorly characterized function indicate that *Bradyrhizobium* strains have potential pathways or capabilities that have not been determined by the existing COG classification. The percentage of genes related to metabolism was higher in the core, accessory and unique genomes (Figure 2A). The detailed categories of the pan-genome with the highest representation were 'transcription' (n = 3,075), 'replication, recombination and repair' (n = 2,912), 'amino acid transport and metabolism' (n = 2,718) and 'function unknown' (n = 8,936). The most common functions in the core genomes are associated with 'amino acid transport and metabolism' (8.5%), 'translation, ribosomal structure and biogenesis' (7.28%), and 'energy production and conversion' (6.55%) (Figure 2B). COG analysis revealed that core genes are functionally diverse and perform basic housekeeping functions regardless of their isolated host. The most abundant categories in the accessory genome include 'transcription' (5.19%), 'amino acid transport and metabolism' (4.38%) and 'replication, recombination, and repair' (3.45%). The unique genes were expected to encode a large number of putative proteins with unknown functions. And 'replication, recombination, and repair' accounted for around 3.48% of the unique genome content. Given the open state of the *Bradyrhizobium* pan-genome, the results of COG enrichment analysis for accessory and unique genomes, in particular 'replication, recombination, and repair', support the perspective

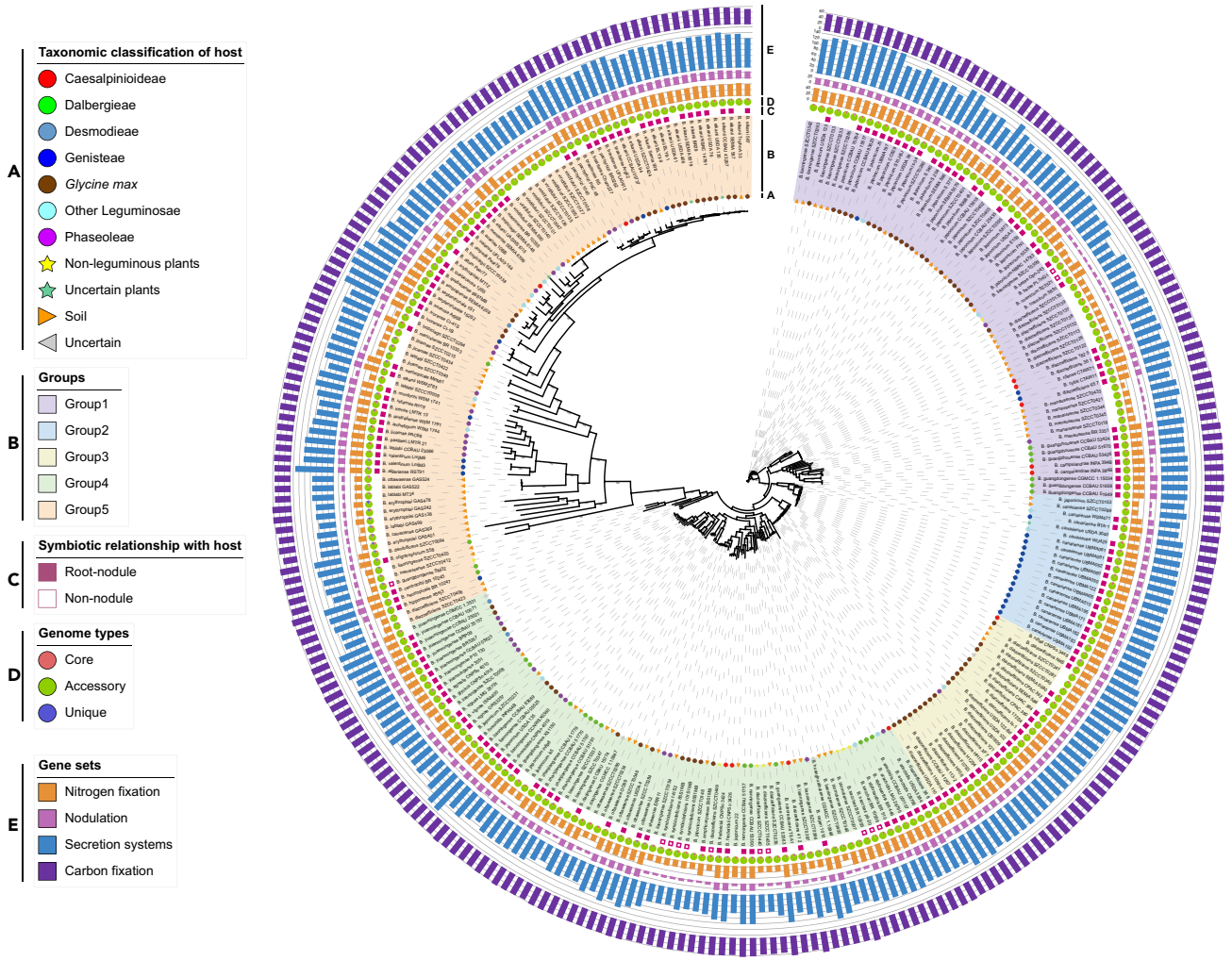


Figure 3. Phylogenetic tree and genetic profile of the 278 *Bradyrhizobium* strains

Phylogenetic tree was based on 625 single-copy core genes, and bootstrap support values were calculated from 100 replicates; values above 75 are indicated. Representation of the data (inner to outer layers): (A). Taxonomic classification of isolate host; (B). Phylogenetic clades of strains; (C). Nod-dependent (solid box) and nod-independent (hollow box) relationships with plant hosts, the strains isolated uncertain or from soil were not shown.

(D) Proportion of core, accessory, and strain-specific genes in each strain; (E) The number of functional genes of nitrogen fixation, nodulation, secretion systems, and carbon fixation.

that the *Bradyrhizobium* genome was highly plastic. However, a high proportion (10.80%, 50.73%, and 74.14%) of the core, accessory, and unique genome did not have homologs outside the genus. This suggests that deeper gene function analyses would provide more detailed and novel functions for the genus *Bradyrhizobium*. In addition, the Gene Ontology (GO) enrichment analysis of the pan-genome showed that core genes were enriched in 'structural constituent of ribosome', 'cytosolic large ribosomal subunit', 'large ribosomal subunit', etc. The accessory genes were enriched in 'cytosol', 'cytoplasm', 'ATPase activity', '4 iron, 4 sulfur cluster binding', and 'integral component of plasma membrane'. Unique genes were enriched in 'cytosolic large ribosomal subunit', 'integral component of plasma membrane', 'cytoplasm', '4 iron, 4 sulfur cluster binding', and 'large ribosomal subunit' (Figure S1).

Phylogenetic relationship of *Bradyrhizobium* strains

Using concatenated single-copy orthologous genes to resolve phylogenetic relationships could avoid rearrangement-biased phylogenetic tree reconstructions, although it may miss many differences caused by single genes.[22,23] The phylogenetic analysis based on the single-copy core genes of the pan-genome provided a complementary perspective on the evolutionary relationships among *Bradyrhizobium* strains. The phylogenetic tree showed the existence of several clades but revealed a more complex structure within *Bradyrhizobium* with additional groups (Figure 3). The genetic similarity of most strains varies with different groups of type species of this genus, but not all. Interestingly, the majority of *B. japonicum* species were assigned to group 1, whereas *B. japonicum* SZCCT0153 was assigned to group 2 and *B. japonicum*

SZCCT0231, *B. japonicum* USDA 135, *B. japonicum* in8p8, *B. japonicum* is5, *B. japonicum* SZCCT0148, *B. japonicum* 22 were assigned to group 4. We found that the strains of *B. canariense* species clustered in group 2, with the exception of *B. canariense* GAS369, which was in group 5. In addition, the strain *Bradyrhizobium manausense* SZCCT0412 also diverged from other *B. manausense* strains. The genomic profile of the number of accessory genes in these strains was identical, with an exceptionally high proportion of accessory genes in the genome of each strain.

The phylogeny reconstructed using these single-copy genes was largely congruent with previous ones inferred mainly from marker genes by multilocus sequence analysis (MLSA).[20] Our results were based on the study of phenotypically variable strains of *Bradyrhizobium* isolated from a wide range of plants and soils, in which many legume species overlapped and clustered in different subgroups. Despite having partially overlapping ecological niches, there was significant genetic differentiation between *B. diazoefficiens*, *B. japonicum*, and *B. elkanii*. Previous studies have shown that recombination and migration are important evolutionary forces that make species cohesive.[24] Several new positions were discovered for some species. For example, *B. japonicum*, *B. diazoefficiens*, etc. showed significant intraspecific genetic differentiation. Strains of the same species appeared to be clearly divergent and appeared in different monophyletic clusters, which could be due to a tendency for interspecific recombination or, alternatively, that previous classifications were inaccurate, which requires further investigation and research to confirm. Host-specific clustering was identified in subclades based on the single-copy core genome phylogenetic tree. There appeared to be a relationship between the isolation source of the strains and their evolutionary relationship with other strains. For example, in group 3, the evolutionary relationships of *B. diazoefficiens* strains isolated from plants tend to cluster more closely together, as do strains from soil, which are also evolutionarily closer to each other. Similar clustering was observed for *B. diazoefficiens* in group 1.

Repertoire of nitrogen fixation genetic traits

Bradyrhizobium includes both symbiotic nitrogen-fixing bacteria that can form nodules on soybeans and non-symbiotic members that cannot.[7,8] The superior symbiotic nitrogen-fixing ability of *Bradyrhizobium* members is often generated by *nif* genes and *nod* genes.[13] Characterizing of the genomic features of strains with different symbiotic phenotypes can provide a better understanding of *Bradyrhizobium* nitrogen fixation and distinguish symbiotic from non-symbiotic strains in *Bradyrhizobium*. To identify possible genomic signatures associated with nitrogen fixation phenotypes, we examined the distribution of nodulation and nitrogen fixation genes in the pan-genome of *Bradyrhizobium*, and found 190 genes (Table S3) involved in nodulation and nitrogen fixation, of which 5 were core genes, 141 were accessory genes, and 44 were strain-specific genes. The number of nitrogen fixation and nodulation genes within individual genomes ranged from 10 to 51 and 3 to 29, respectively (Figure 3). The distribution of nitrogen fixation and nodulation genes of these *Bradyrhizobium* strains appeared to be host specific, with 85.05% of strains from soil harboring fewer nodulation genes (less than 10), whereas strains from legumes had more nodulation genes (Figure 3). For example, *B. japonicum* SZCCT0280, *B. japonicum* SZCCT0401, *B. japonicum* SZCCT0402, *B. japonicum* SZCCT0403, and *B. japonicum* SZCCT0395, which were isolated from soil, had fewer nitrogen fixation genes and nodulation genes than those *B. japonicum* isolated from *Glycine max* nodules, despite their closer evolutionary relationship in the phylogenetic tree. A similar genetic profile was also observed in species such as *B. diazoefficiens* and *B. canariense*. We carried out a clustering analysis and comparisons based on the presence or absence of genes involved in nitrogen fixation and nodulation. It is noteworthy that the clustering presence/absence profiles showed that 278 *Bradyrhizobium* strains could be classified into two major functionally similar groups (Figure S2), and that the diversity of gene combinations in different nodulation and nitrogen fixation was responsible for the diverse spectrum of host-symbiont interactions.

Specific T3SS components in rhizobia have been found to mediate nodule development and have either positive or negative effects in various host nodules.[15,25] To further reveal the potential for symbiotic nitrogen fixation, we also detected 401 genes involved in secretion systems, including type I (T1SS), type II (T2SS), type III (T3SS), type IV (T4SS), type V (T5SS) and type VI (T6SS), which have been reported to be important in symbiosis.[26] The genes required for nodulation or nitrogen fixation by *Bradyrhizobium* have been reported typically occur together in a gene cluster.[15,20] *FixLJK*, which responds to hypoxic conditions in soil and nodules and controls the expression of nitrogen fixation and denitrification genes,[27] was common among these strains. The critical symbiosis gene clusters for nodulation (*nodABCSUIJ*) and nitrogen fixation (*nifABDEHKNQTWXZ*) are functional determinants of symbiotic nitrogen fixation, and *Bradyrhizobium* lacking these genes is likely to be incapable of nodule formation and nitrogen fixation.[20] These gene clusters were found to be associated with host type. We found that the *Bradyrhizobium* strains isolated from root nodules of *Glycine max*, Desmodieae, Dalbergieae and other Leguminosae plant hosts had both the *nif* and *nod* gene clusters, as well as the *ysc* gene cluster of the T3SS system (Figure 4), indicating that they could be used as rhizobial inoculants. Although isolated from non-legumes, strains *B. sacchari* BR 10555, *B. sacchari* BR 10556, and *B. sacchari* p9-20 were symbiotic strains carrying *nif*, *nod*, and *ysc* gene clusters. Most of the strains from Genisteae nodules carried the *nif* and *nod* gene clusters, but not the *ysc* gene cluster. The majority of the soil-isolated strains lacked the *nif*, *nod* and *ysc* gene clusters, suggesting that these isolates may not exhibit nod-dependent symbiotic nitrogen fixation, but do contain additional nitrogen-fixing genes such as *fixLJK*, *nifS*, *nodM* (Figures 4 and S2). The strains *Bradyrhizobium symbiodeficiens* 141S2, *B. symbiodeficiens* 101S1MB, *B. symbiodeficiens* 85S1MB, *B. symbiodeficiens* 65S1MB isolated from *Glycine max* were free-living members, which could be due to the absence or inactivation of efficient nodulation gene clusters, consistent with their inability to form symbiotic nodules with legumes.[28]

Some strains from the same species, despite their close phylogenetic relationship, they exhibited differences in symbiosis gene cluster, for example, the content of the symbiotic gene content of *B. viridifuturi* SEMIA 690 differed from that *B. viridifuturi* species of other strains, which could reflect host characteristics. Previous studies have shown that horizontal gene transfer likely facilitates *nif* expansion among the free-living *Bradyrhizobium*.[29] Free-living *Bradyrhizobium* containing *nif* gene clusters are of particular interest as they may play as yet unidentified roles in the nitrogen cycle in soil.[29] *Bradyrhizobium* members without nitrogen-fixing gene clusters do not fix nitrogen in the soil, but may be

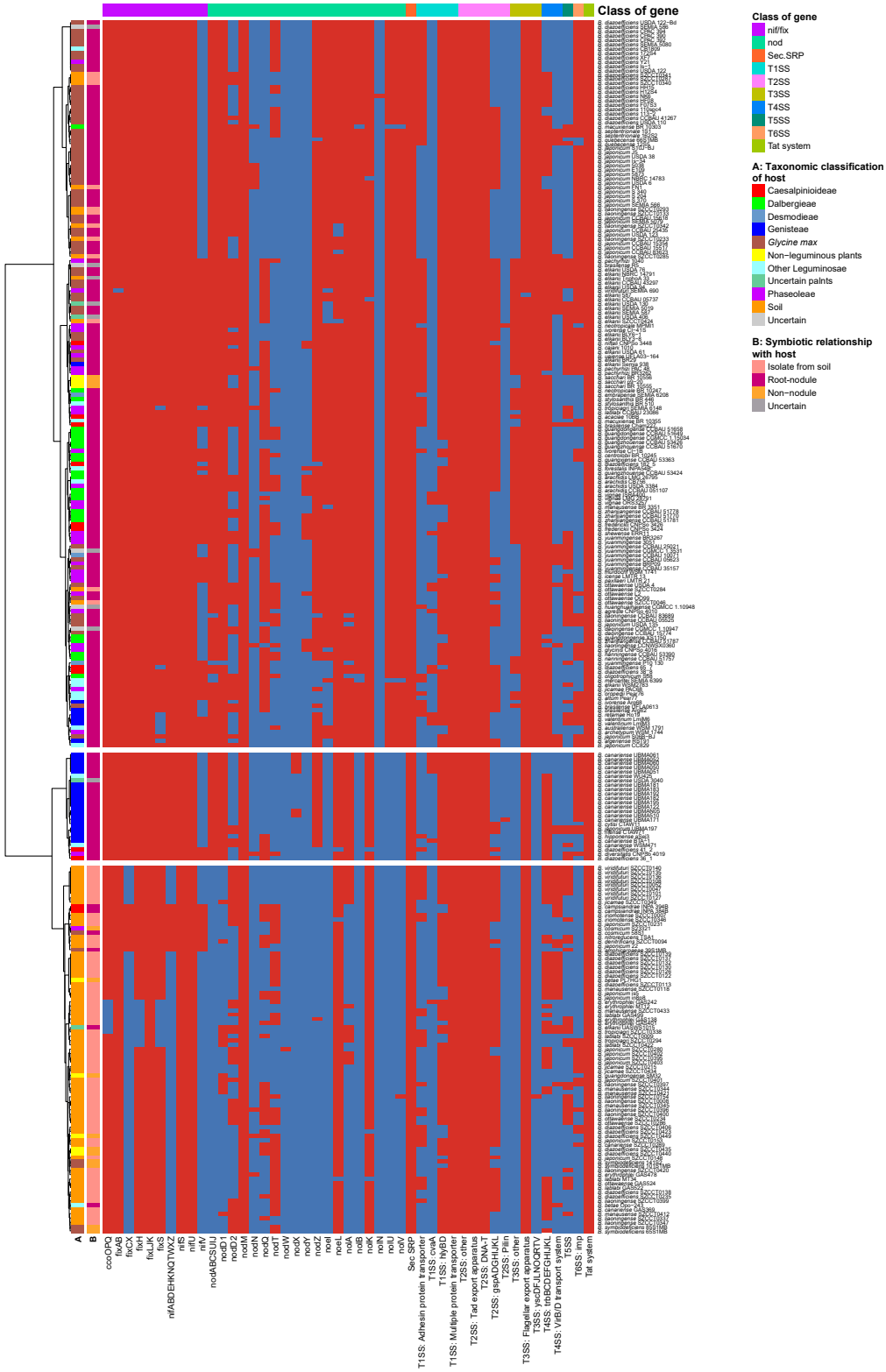


Figure 4. Presence (red)/absence (blue) pattern of functional module gene clusters involved in nodules, nitrogen fixation, and secretion system in each *Bradyrhizobium* strain

Each column represents a cluster of homologous genes. A red box indicates presence of the gene in the corresponding strain while a blue box indicates an absence. The dendrogram was generated based on the ward D method for clustering of Euclidean distances. A strain is considered to contain a gene cluster if 80% of the genes in that cluster are present.

adapted to use aromatic soil components.[7] The *B. cosmicum* 58S1 and *B. amphicarpaeae* 39S1MB strains, which were isolated from *Glycine max*, did not have the *nod* and *ysc* gene clusters, but did have *nif*. They may use a nod-independent mechanism or they may be nodule-inhabiting strains that have the potential to interact with plants while not nodulating legumes. *B. cosmicum* S23321 isolated from Phaseoleae is non-nodulating, it contained the *nif* cluster, and the *fixAB*, *fixCX*. The vast majority of the soil group are non-symbiotic lineages. Strains isolated from soil, such as *B. japonicum* 22, *B. nitroreducens* TSA1, *B. iriomotense* SZCCT0007, *B. denitrificans* SZCCT0094, *B. japonicum* SZCCT0231, and *B. iriomotense* SZCCT0346, were free-living *nif*-carrying *Bradyrhizobium*. In addition, in combination with the phylogeny, the vast majority of free-living members of the soil *B. diazoefficiens* formed an independent cluster (Figure 3), and they lacked the *nif*, *nod* and *ysc* gene clusters. We could predict that these strains cannot fix atmospheric nitrogen. Overall, each *Bradyrhizobium* strain had its own set of genes for nodule formation, nitrogen fixation, and secretion systems, which were expected to collaborate to establish the molecular basis for symbiotic association and nitrogen fixation. We have accessed a previously unrecognized part of the functional diversity of this taxonomic group in soils. However, many gaps remain, including the identification of the correlations between genome size variance and the phylogeny of strains, and symbiotic relationship with the host. This could be overcome by developing a holistic model of symbiosis evolution.[20] It is important to continue this research on genes with unknown functions to better estimate mechanisms driving the evolution of mobile genetic elements.

Diversities of genes involved in carbon fixation

The genus *Bradyrhizobium* is characterized by the presence of a number of genes for carbon fixation. The pan-genome of the 278 strains contained a collection of genes (139, Table S4) for carbon fixation, including 18 core genes, 95 accessory genes, and 26 strain-specific genes (Figure S3). Each strain contained carbon fixation genes ranging from 45 to 67 (Figure 3). Previous studies have shown that not all *Bradyrhizobium* are capable of photosynthesis,[30] and in our study, the genes for photosynthetic carbon fixation, such as *ACAT*, *FBA*, *GAPDH*, *glpX-SEBP*, *maeB*, *mdh*, *pgk*, *rpiA*, *tktA*, and *TPI*, were present in all strains, suggesting that despite the presence of carbon fixation genes many strains do not possess photosynthetic ability. The strain *B. elkanii* USDA61 has been noted to be non-photosynthetic,[31] but carbon fixation genes were present. This was consistent with the nitrogen-fixing ability of the strain *Bradyrhizobium* sp. S23321 isolated from paddy soil in Japan, which was found to lack nodulation ability despite the presence of nitrogen fixation (nif) genes.[5] These carbon fixation genes did not exhibit host-specific grouping, in contrast to the nodule genes. The *accBD*, *fumAB*, and *ACAT* genes were present in all strains. The succinate dehydrogenase, represented by the *sdhABD* genes and engaged in carbon oxidation, was detected in all members of the genus. Many *Bradyrhizobium* strains can fix carbon dioxide via the Calvin cycle, and RuBisCo is the key enzyme at the first step in the Calvin cycle.[32] The *rbcL* gene of RuBisCO was found in 272 strains of the genus *Bradyrhizobium*, but the remaining six strains all lacked it: *B. elkanii* 587, *B. manausense* BR 3351, *Bradyrhizobium lablabi* SZCCT0009, *Bradyrhizobium erythrophlei* GAS138, *B. lablabi* GAS499, and *B. erythrophlei* GAS401. Our results highlight an unsuspected distribution and diversity of carbon fixation genes within the genus *Bradyrhizobium*, which may contribute to the genus's agricultural and environmental applications.

Conclusions

This study presented a comprehensive analysis of *Bradyrhizobium*, including the largest number (N = 278) of whole *Bradyrhizobium* genomes retrieved from publicly available repositories. We outlined the genomic variability of this *Bradyrhizobium* genus, demonstrating that approximately 0.98% (824/84,078 genes) of the pan-genome was found to constitute the core genome and was present in 100% of the strains studied, accomplishing general molecular functions for cell maintenance such as amino acid transport and metabolism, translation, ribosomal structure and biogenesis, energy production and conversion. The accessory genome, on the other hand, demonstrated a wide range of functions, from broad-spectrum nitrogen fixation to nodulation factors potentially establishing effective symbiosis with host cells. However, nitrogen fixation genetic traits were also observed in the core genome.

This study aimed to uncover the evolutionary processes occurring in *Bradyrhizobium* that may help to understand the molecular mechanisms that drive nitrogen fixation. The phylogenetic relationships of these strains have a tendency to host-specific clustering, with strains from the same host forming many monophyletic clusters, especially those from soil. These strains differed in the gene clusters of functional modules involved in symbiotic nitrogen fixation, such as nodulation, nitrogen fixation and secretion systems, primarily the *nod*, *nif*, and *ysc* gene clusters. Most strains from legumes carried both the *nif*, *nod*, and *ysc* gene clusters, suggesting that they could be used as rhizobial inoculants. Some strains from Genisteae primarily lacked the *ysc* gene cluster, and most strains from soil were *nif*-carrying and non-*nif*-carrying free-living (nod-free) members. This study provided a relatively complete pan-genomic description of the genus *Bradyrhizobium* as well as a comprehensive understanding of *Bradyrhizobium*'s nitrogen fixation and carbon fixation genes, which would provide useful understanding in the agricultural applications of *Bradyrhizobium*.

Limitations of the study

A limitation of this work is that 278 draft genomes were examined, while some genes may be missing from the draft genome assembly but allow for a more comprehensive analysis of *Bradyrhizobium*. Also, a role for the *nif* gene cluster in the free-living lifestyle has not yet been recognized, although we have accessed a previously unrecognized part of the functional diversity in symbiotic and nonsymbiotic strains. It remains unclear whether *nif*-carrying *Bradyrhizobium* can use a nod-independent strategy to fix nitrogen, and if so, how this mechanism functions needs to be further investigated.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
- METHOD DETAILS
  - *Bradyrhizobium* genome sequences
  - Pan- and core genome construction
  - Phylogenetic analysis
  - Functional analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2024.108948.

ACKNOWLEDGMENTS

This research was supported by the Guangxi Science and Technology Base and Talent Special Project (AD22080023), the Guangxi Natural Science Foundation (2021GXNSFFA196005), and the National Natural Science Foundation of China (42301073).

AUTHOR CONTRIBUTIONS

Z.H.Z., K.N., and C.F.Z. designed and managed the whole project; Z.H.Z. and K.N. led the analyses and manuscript preparation; C.F.Z. performed the analyses and wrote the initial manuscript, and all authors commented on the draft and revised the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: June 15, 2023
Revised: September 12, 2023
Accepted: January 15, 2024
Published: January 18, 2024

REFERENCES

1. Rivas, R., Martens, M., de Lajudie, P., and Willems, A. (2009). Multilocus sequence analysis of the genus Bradyrhizobium. Syst. Appl. Microbiol. *32*, 101–110.
2. Gupta, R.S., and Mok, A. (2007). Phylogenomics and signature proteins for the alpha proteobacteria and its main groups. BMC Microbiol. *7*, 106.
3. Uroz, S., Buée, M., Murat, C., Frey-Klett, P., and Martin, F. (2010). Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. Environ. Microbiol. Rep. *2*, 281–288.
4. Ormeño-Orrillo, E., and Martínez-Romero, E. (2019). A Genomotaxonomy View of the Bradyrhizobium Genus. Front. Microbiol. *10*, 1334.
5. Okubo, T., Tsukui, T., Maita, H., Okamoto, S., Oshima, K., Fujisawa, T., Saito, A., Futamata, H., Hattori, R., Shimomura, Y., et al. (2012). Complete Genome Sequence of Bradyrhizobium sp. S23321: Insights into Symbiosis Evolution in Soil Oligotrophs. Microbes Environ. *27*, 306–315.
6. Parker, M.A. (2015). The spread of Bradyrhizobium lineages across host legume clades: from Abarema to Zygia. Microb. Ecol. *69*, 630–640.
7. VanInsberghe, D., Maas, K.R., Cardenas, E., Strachan, C.R., Hallam, S.J., and Mohn, W.W. (2015). Non-symbiotic Bradyrhizobium ecotypes dominate North American forest soils. ISME J. *9*, 2435–2441.
8. Rowan-Nash, A.D., Korry, B.J., Mylonakis, E., and Belenky, P. (2019). Cross-Domain and Viral Interactions in the Microbiome. Microbiol. Mol. Biol. Rev. *83*, e00044-18.
9. Itakura, M., Saeki, K., Omori, H., Yokoyama, T., Kaneko, T., Tabata, S., Ohwada, T., Tajima, S., Uchiumi, T., Honnma, K., et al. (2009). Genomic comparison of Bradyrhizobium japonicum strains with different symbiotic nitrogen-fixing capabilities and other Bradyrhizobiaceae members. ISME J. *3*, 326–339.
10. Kaneko, T., Maita, H., Hirakawa, H., Uchiike, N., Minamisawa, K., Watanabe, A., and Sato, S. (2011). Complete Genome Sequence of the Soybean Symbiont Bradyrhizobium japonicum Strain USDA6T. Genes *2*, 763–787.
11. Jaiswal, S.K., and Dakora, F.D. (2019). Widespread Distribution of Highly Adapted Bradyrhizobium Species Nodulating Diverse Legumes in Africa. Front. Microbiol. *10*, 310.
12. Siqueira, A.F., Ormeño-Orrillo, E., Souza, R.C., Rodrigues, E.P., Almeida, L.G.P., Barcellos, F.G., Batista, J.S.S., Nakatani, A.S., Martínez-Romero, E., Vasconcelos, A.T.R., and Hungria, M. (2014). Comparative genomics of Bradyrhizobium japonicum CPAC 15 and Bradyrhizobium diazoefficiens CPAC 7: elite model strains for understanding symbiotic performance with soybean. BMC Genom. *15*, 420.
13. Stepkowski, T., Hughes, C.E., Law, I.J., Markiewicz, Ł., Gurda, D., Chlebicka, A., and Moulin, L. (2007). Diversification of lupine Bradyrhizobium strains: evidence from nodulation gene trees. Appl. Environ. Microbiol. *73*, 3254–3264.
14. Grönemeyer, J.L., Kulkarni, A., Berkelmann, D., Hurek, T., and Reinhold-Hurek, B. (2014). Rhizobia Indigenous to the Okavango Region in Sub-Saharan Africa: Diversity, Adaptations, and Host Specificity. Appl. Environ. Microbiol. *80*, 7244–7257.
15. Patra, D., and Mandal, S. (2022). Nod–factors are dispensable for nodulation: A twist in

bradyrhizobia-legume symbiosis. Symbiosis *86*, 1–15.

16. Lozano, M.J., Redondo-Nieto, M., Garrido-Sanz, D., Mongiardini, E., Quelas, J.I., Mengucci, F., Dardis, C., Lodeiro, A., and Althabegoiti, M.J. (2021). Comparative Analysis of Three Bradyrhizobium diazoefficiens Genomes Show Specific Mutations Acquired during Selection for a Higher Motility Phenotype and Adaption to Laboratory Conditions. Microbiol. Spectr. *9*, e0056921.
17. Li, R., Feng, Y., Chen, H., Zhang, C., Huang, Y., Chen, L., Hao, Q., Cao, D., Yuan, S., and Zhou, X. (2020). Whole-Genome Sequencing of Bradyrhizobium diazoefficiens 113-2 and Comparative Genomic Analysis Provide Molecular Insights Into Species Specificity and Host Specificity. Front. Microbiol. *11*, 576800.
18. Hollowell, A.C., Regus, J.U., Gano, K.A., Bantay, R., Centeno, D., Pham, J., Lyu, J.Y., Moore, D., Bernardo, A., Lopez, G., et al. (2016). Epidemic Spread of Symbiotic and Non-Symbiotic Bradyrhizobium Genotypes Across California. Microb. Ecol. *71*, 700–710.
19. Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., et al. (2007). Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia. Science *316*, 1307–1312.
20. Weisberg, A.J., Rahman, A., Backus, D., Tyavanagimatt, P., Chang, J.H., and Sachs, J.L. (2022). Pangenome Evolution Reconciles Robustness and Instability of Rhizobial Symbiosis. mBio *13*, e0007422.
21. Chen, W.F., Meng, X.F., Jiao, Y.S., Tian, C.F., Sui, X.H., Jiao, J., Wang, E.T., and Ma, S.J. (2023). Bacteroid Development, Transcriptome, and Symbiotic Nitrogen-Fixing Comparison of Bradyrhizobium arachidis in Nodules of Peanut (Arachis hypogaea) and Medicinal Legume Sophora flavescens. Microbiol. Spectr. *11*, e0107922.
22. Salichos, L., and Rokas, A. (2013). Inferring ancient divergences requires genes with strong phylogenetic signals. Nature *497*, 327–331.
23. Zhang, N., Zeng, L., Shan, H., and Ma, H. (2012). Highly conserved low-copy nuclear genes as effective markers for phylogenetic analyses in angiosperms. New Phytol. *195*, 923–937.
24. Vinuesa, P., Silva, C., Werner, D., and Martínez-Romero, E. (2005). Population genetics and phylogenetic inference in bacterial molecular systematics: the roles of migration and recombination in Bradyrhizobium species cohesion and delineation. Mol. Phylogenet. Evol. *34*, 29–54.
25. Okazaki, S., Tittabutr, P., Teulet, A., Thouin, J., Fardoux, J., Chaintreuil, C., Gully, D., Arrighi, J.-F., Furuta, N., Miwa, H., et al. (2016). Rhizobium–legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. ISME J. *10*, 64–74.
26. Tampakaki, A.P. (2014). Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria. Front. Plant Sci. *5*, 114.
27. Ferrières, L., Francez-Charlot, A., Gouzy, J., Rouillé, S., and Kahn, D. (2004). FixJ-regulated genes evolved through promoter duplication in Sinorhizobium meliloti. Microbiology *150*, 2335–2345.
28. Bromfield, E.S.P., Cloutier, S., and Nguyen, H.D.T. (2020). Description and complete genome sequences of Bradyrhizobium symbiodeficiens sp. nov., a non-symbiotic bacterium associated with legumes native to Canada. Int. J. Syst. Evol. Microbiol. *70*, 442–449.
29. Tao, J., Wang, S., Liao, T., and Luo, H. (2021). Evolutionary origin and ecological implication of a unique nif island in free-living Bradyrhizobium lineages. ISME J. *15*, 3195–3206.
30. Jin, C.-Z., Wu, X.-W., Zhuo, Y., Yang, Y., Li, T., Jin, F.-J., Lee, H.-G., and Jin, L. (2022). Genomic insights into a free-living, nitrogen-fixing but non nodulating novel species of Bradyrhizobium sediminis from freshwater sediment: Three isolates with the smallest genome within the genus Bradyrhizobium. Syst. Appl. Microbiol. *45*, 126353.
31. Okazaki, S., Kaneko, T., Sato, S., and Saeki, K. (2013). Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. Proc. Natl. Acad. Sci. USA *110*, 17131–17136.
32. Hügler, M., and Sievert, S.M. (2011). Beyond the Calvin Cycle: Autotrophic Carbon Fixation in the Ocean. Ann. Rev. Mar. Sci *3*, 261–289.
33. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.
34. Edgar, R.C. (2021). MUSCLE v5 enables improved estimates of phylogenetic tree confidence by ensemble bootstrapping. bioRxiv. https://doi.org/10.1101/2021.06.20.449169.
35. Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577.
36. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
37. Huerta-Cepas, J., Forslund, K., Coelho, L.P., Szklarczyk, D., Jensen, L.J., von Mering, C., and Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. Mol. Biol. Evol. *34*, 2115–2122.
38. Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., and Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. Bioinformatics *36*, 2251–2252.
39. Kanehisa, M., Sato, Y., and Kawashima, M. (2022). KEGG mapping tools for uncovering hidden features in biological data. Protein Sci. *31*, 47–53.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
Genome data	NCBI	Table S1
Software and algorithms		
OrthoMCL	Li et al.[33]	http://orthomcl.org/orthomcl/
MUSCLE v5.1.0	Edgar et al.[34]	https://www.drive5.com/muscle/
Gblocks v0.91b	Talavera et al.[35]	http://molevol.cmima.csic.es/castresana/Gblocks/
RAxML v8.2.12	Stamatakis et al.[36]	https://github.com/stamatak/standard-RAxML
eggNOG-mapper	Huerta-Cepas et al.[37]	http://eggnog-mapper.embl.de/
KofamKOALA	Aramaki et al.[38]	https://www.genome.jp/tools/kofamkoala/
KEGG mapping	Kanehisa et al.[39]	https://www.kegg.jp

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Prof. Zhonghua Zhang (gxtczzh@126.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The published article contains all data sets generated or analyzed during this study.
- The codes and scripts used for the phylogenomic and comparative genomic analyses in this study are deposited in the online repository at https://github.com/zhongchaofang/Bradyrhizobium_pangenome.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

This study does not include experimental model or study participant.

METHOD DETAILS

Bradyrhizobium genome sequences

For pan-genome analysis, whole-genome sequences for *Bradyrhizobium* strains were available from the NCBI Reference Sequence (RefSeq) database. If a strain had multiple genomes, we selected the one with the highest level of genomic integrity for comparison study. In addition, if a strain had multiple replicons (chromosomes and plasmids), a primary chromosome was selected. As a result, 278 *Bradyrhizobium* genomes from 66 species were carefully included in this study, consisting 52 complete genomes and 226 draft genomes, and 62 genome sequences were from type strains. Their genome information is given in Table S1. The quality and Average Nucleotide Identity (ANI) of the genomes were assessed using CheckM and dRep, and the results were shown in Table S2 and Figure S4.

Pan- and core genome construction

Orthologous pairs were determined for all 278 *Bradyrhizobium* genomes using OrthoMCL version 2.0.9.[33] The BLASTP searches were conducted among mixed protein sequences with an E value of 1.0×10^{-5}, and the minimum score value used in BLASTP results was set to 50%. The MCL with an expansion parameter of 1.5 was used to cluster the filtered BLASTP results, which has been widely utilized in earlier studies of microbial genomes to uncover orthologs between various strains.

Phylogenetic analysis

For phylogenetic analysis, the protein sequences of 625 single-copy core genes were first subjected to the multiple sequence alignment using MUSCLE v5.1.0[34] with default parameters, and then Gblocks v0.91b[35] were used to make the selection of blocks from the multiple sequence alignment results (parameters: -t=p), the Gblocks results were then concatenated sequentially, followed by phylogenetic analysis using RAxML v8.2.12[36] to construct a maximum likelihood tree with 100 bootstrap iterations (parameters: -f a -x 12345 -N 100 -m PROTGAMMALGX -n ex -T 30 -p 12345).

Functional analysis

Clusters of orthologous groups (COG) of pan-genome of strains were determined using eggNOG-mapper software.[37] Functional annotations of the pan-genome were performed using Kyoto encyclopedia of genes and genomes (KEGG) database. KofamKOALA was used to annotate the functions of pan-genome proteins, which assigns KEGG orthology identifier to proteins via HMMER/HMMSEARCH.[38] The results of KofamKOALA were processed by the KEGG Mapper-Reconstruct Pathway utility with manual curation.[39] The gene function information associated with nitrogen fixation, nodulation, secretion system, and carbon fixation were then analyzed using R statistical analysis.

The 'pheatmap' function in the 'R' software was used to analyze and visualize the presence/absence pattern of the functional module gene clusters. We clustered the presence (1)/absence (0) matrix using a hierarchical clustering approach, where the Euclidean distance was chosen as the distance metric for the clustering process.

QUANTIFICATION AND STATISTICAL ANALYSIS

The genome quality assessment was performed in dRep with ANI clustering threshold of 0.95. The statistical analyses of homologous sequences were performed in BLASTP with the E-value < 1e-5. COG, GO and KEGG functions were statistically analyzed with an E-value < 1e-5. Significant enrichment of GO terms was measured by the Fisher's exact test with a p-value < 0.05.