The ISME Journal (2021) 15:3195–3206
https://doi.org/10.1038/s41396-021-01002-z

ARTICLE

Evolutionary origin and ecological implication of a unique *nif* island in free-living *Bradyrhizobium* lineages

Jinjin Tao[1,2] · Sishuo Wang[1] · Tianhua Liao[1] · Haiwei Luo[1,2]

Received: 3 February 2021 / Revised: 21 April 2021 / Accepted: 28 April 2021 / Published online: 14 May 2021

Abstract

The alphaproteobacterial genus *Bradyrhizobium* has been best known as N_2-fixing members that nodulate legumes, supported by the *nif* and *nod* gene clusters. Recent environmental surveys show that *Bradyrhizobium* represents one of the most abundant free-living bacterial lineages in the world's soils. However, our understanding of *Bradyrhizobium* comes largely from symbiotic members, biasing the current knowledge of their ecology and evolution. Here, we report the genomes of 88 *Bradyrhizobium* strains derived from diverse soil samples, including both *nif*-carrying and non-*nif*-carrying free-living (*nod* free) members. Phylogenomic analyses of these and 252 publicly available *Bradyrhizobium* genomes indicate that *nif*-carrying free-living members independently evolved from symbiotic ancestors (carrying both *nif* and *nod*) multiple times. Intriguingly, the *nif* phylogeny shows that the vast majority of *nif*-carrying free-living members comprise an independent cluster, indicating that horizontal gene transfer promotes *nif* expansion among the free-living *Bradyrhizobium*. Comparative genomics analysis identifies that the *nif* genes found in free-living *Bradyrhizobium* are located on a unique genomic island of ~50 kb equipped with genes potentially involved in coping with oxygen tension. We further analyze amplicon sequencing data to show that *Bradyrhizobium* members presumably carrying this *nif* island are widespread in a variety of environments. Given the dominance of *Bradyrhizobium* in world's soils, our findings have implications for global nitrogen cycles and agricultural research.

Introduction

Members of the slow-growing genus *Bradyrhizobium* constitute an important group of rhizobia [1, 2]. They symbiotically fix N_2 with diverse legume tribes and are the predominant symbionts of a wide range of nodulating legumes [3, 4]. Recent studies, however, have provided accumulating evidence for the existence of close relatives of rhizobia including *Bradyrhizobium* that do not carry *nod* genes which are essential for the establishment of nodulation, and thus are not able to nodulate legumes [5]. Increasingly, *Bradyrhizobium* strains without *nod* have been found to be particularly abundant in soils [6–8]. VanInsberghe et al. [8] showed that non-symbiotic *Bradyrhizobium* was the dominant bacterial lineage in North America coniferous soils. A global atlas of the dominant soil-dwelling bacteria based upon 16S rRNA gene amplicon sequencing indicated that the genus *Bradyrhizobium* was the most abundant bacterial lineage in soils across the world [6]. A recent study suggested that ancestral lineages of *Bradyrhizobium* might adapt to a free-living lifestyle, highlighting the importance of the free-living lifestyle in the evolution of *Bradyrhizobium* [9].

Although the majority of N_2 fixation in terrestrial ecosystems is generally thought to be performed by symbiotic rhizobia [10–12], free-living N_2-fixing bacteria may be important contributors to the nitrogen budgets in a number of environments, for example soil ecosystems that lack leguminous plants [13]. Free-living N_2 fixation also occurs in understudied ecosystems such as deep soil and canopy soil [14, 15], which may lead to the underestimation of its global

These authors contributed equally: Jinjin Tao, Sishuo Wang

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41396-021-01002-z.

✉ Haiwei Luo
hluo2006@gmail.com

[1] Simon F. S. Li Marine Science Laboratory, School of Life Sciences and State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Shatin, Hong Kong SAR

[2] Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China

rates. However, rhizobia are generally thought to be capable of N_2 fixation only in nodules [16, 17]. The exceptions are found in some members of *Bradyrhizobium* and *Azorhizobium* which were shown to be able to fix N_2 in both the symbiotic and free-living states [18, 19]. Such cases in *Bradyrhizobium* were mostly explored in photosynthetic members [18, 19], but have recently been found in other strains. For instance, *Bradyrhizobium* sp. AT1, a non-symbiotic strain isolated from the root of sweet potato was reported to fix N_2 with a rate comparable to the photosynthetic strain *Bradyrhizobium* sp. ORS278 under the molecular oxygen (O_2) concentrations at 1–5% (for both) in the free-living state [20]. Another example is *Bradyrhizobium* sp. DOA9, which harbors a *nif* cluster on its chromosome and another on a symbiotic plasmid: while both could participate in N_2 fixation during symbiosis, only the chromosomal one participated in N_2 fixation in the free-living state [21].

Compared with the nodules, the condition of soils is not favorable for N_2 fixation for several reasons. First, in contrast to being at nanomolar levels in nodules, the O_2 concentration in soils could be up to atmospheric levels [22], which is highly detrimental to N_2 fixation, a process that requires an anaerobic microenvironment [23]. Second, unlike symbiotic rhizobia, free-living rhizobia lack readily available organic matter provided by the host plants in exchange for fixed nitrogen. Last, free-living soil bacteria are frequently exposed to stresses like droughts, high osmolarity, and high temperature. It is therefore imperative to investigate the strategies that free-living N_2-fixing *Bradyrhizobium* may use to deal with the harsh condition in soils.

Despite the potential ecological significance, free-living *Bradyrhizobium*, particularly those able to fix N_2, have been poorly characterized. To this end, we isolated 88 *Bradyrhizobium* and five related strains from soils of soybean cropland, artificial park, forest, and grassland at several geographic locations in China, and had their genomes sequenced. By building a phylogenomic tree and reconstructing the ancestral lifestyles with additional 252 genomes where 42 are free-living *Bradyrhizobium* strains deposited in public databases, we revealed complex lifestyle shifts along the evolutionary history of *Bradyrhizobium*. Notably, we showed that horizontal gene transfer (HGT) of a unique *nif* island may have facilitated their transitions from symbiotic to free-living lifestyle and adaptation to soil habitats.

Methods and materials

Soil and plant tissue sampling and processing

Samples were collected from five different sites in China (Supplementary Fig. S1A) that cover several ecosystem types: soybean cropland (Heihe and Lvliang), artificial park (Shenzhen), undeveloped forest (Lanzhou), and grassland (Hefei). In Heihe, Lvliang and Hefei sites, intact *Glycine max* (soybean) and *Erigeron annuus* (annual fleabane) plants were excavated from soil, and shaken mildly to remove soil loosely adhering to the roots. The remaining adherent soil was separated from the roots as rhizosphere soil. For Shenzhen and Lanzhou sites, 0–5 cm and 5–10 cm depth bulk soil was collected from the vicinity of plant root. At the Shenzhen sampling site, the soil was collected near *Acacia confusa* (legume), *Calliandra haematocephala* (legume) and *Bambusoideae* (non-legume). At the Lanzhou site, the soil was collected near a non-leguminous plant *Picea asperata*. Detailed information about sampling sites was provided in Supplementary Dataset S1. Soil samples were placed in sterile bags, kept on ice, and immediately transported to the laboratory for further processing.

Bradyrhizobium isolation and identification

To prepare soil inoculum, 5.0 g of fresh soil was put in a 50 mL conical tube with 45 mL of sterile deionized water. After mixing with a vortex mixer, 1 mL of soil suspension was serially diluted, and the dilutions were used for inoculation. Roots of *Erigeron annuus* were processed according to Coombs and Franco [24]. One gram of surface-sterilized roots was grinded in a mortar and diluted with PBS buffer, and the dilutions of the root slurry were used for inoculation. Five different media were applied to retrieve target bacteria from the samples, including modified arabinose-gluconate (MAG) media adapted from the study Sachs et al. [25], 10- and 100-fold diluted MAG, ρMAG (MAG supplemented with ρ-coumaric acid) and vanillic acid media (detailed in Supplementary Text S1.2). After the sterilized medium was cooled to 60 °C, 55 mg/L cycloheximide was added to inhibit fungal growth. All agar plates were incubated at 28 °C. Colonies with small, white and raised morphology formed after 7 days were picked for species identification.

Colonies were identified by a 1465 bp fragment PCR product using universal bacteria 16S rRNA primers 27F and 1492R. Chelex 100 resin [Bio-rad, USA] was used to extract DNA from bacterial colonies for PCR reaction, and the recipe of PCR was prepared using Premix Taq [Takara Bio, USA]. The PCR conditions were as follows: denaturation at 95 °C for 5 min, followed by 32 cycles (95 °C for 45 s, 55 °C for 45 s and 72 °C for 90 s), final extension at 72 °C for 10 min. The taxonomic information of the isolates was obtained by comparing the 16S rRNA gene sequences using EzBioCloud [26], which is a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies.

Phylogenomics analysis and reconstruction of ancestral lifestyles

We obtained 93 isolates that likely resembled *Bradyrhizobium*, and had their genomes sequenced with the HiSeq X platform (Illumina), followed by genome assembly and gene identification using SPAdes v3.10.1 [27] and Prokka v1.12 [28], respectively. Completeness of these genomes was calculated using CheckM v1.0.7 [29] with default parameters.

The phylogenomic tree of *Bradyrhizobium* was constructed using IQ-Tree v1.6.2 [30] with the 93 strains isolated in the current study and 252 genomes deposited in the NCBI Genbank database based on amino acid sequences of 123 shared single-copy genes identified by OrthoFinder v2.2.7 [31]. *Bradyrhizobium* members were divided into three lifestyle types based on the presence or absence of *nod* and *nif* genes (Supplementary Datasets S1 and S2). Strains that possess *nifBDEHKN* and *nodABCIJ* genes (at most one missing gene allowed; the same thereafter) were classified as symbiotic (Sym), strains possessing only *nifBDEHKN* genes were classified as free-living *nif*-carrying (FL_{nif}), and those lacking both *nod* and *nif* genes were classified as free-living non-*nif*-carrying (FL_{nonnif}). The lifestyle of ancestral nodes in the phylogenomic tree was inferred using Mesquite v3.61 [32] by the maximum parsimony reconstruction method, which infers the ancestral lifestyles by minimizing the number of steps of lifestyle change along the phylogenomic tree. Similar to a recent study [9], we did not apply the methodologically complex maximum likelihood method because the frequent lifestyle transitions across short branches in the *Bradyrhizobium* phylogeny likely leads to overestimates of the transition rate and hence inaccurate reconstruction of ancestral lifestyles.

Amplicon sequence analysis

Metadata of 4958 sequence files (runs) were retrieved from the NCBI Sequence Read Archive (SRA) using the term "*nifH* metagenome" (last accessed in December 2020), which actually means *nifH* amplicon datasets. Raw data were downloaded using the SRA Toolkit (https://github.com/ncbi/sra-tools) and were further processed with QIIME2 [33]. We applied evolutionary placement methods to assign the short sequence reads to the four phylogenetically distinguishable types of *nif* clusters (i.e., FL, PB, Sym and SymBasal) using PaPaRa v2.5 [34] and EPA-ng v0.3.8 [35]. The normalized abundance of each type of *Bradyrhizobium nifH* genes was calculated as the number of reads assigned to the corresponding *nifH* type divided by the total number of reads in the amplicon sequencing dataset assigned to *Bradyrhizobium* (Supplementary Dataset S3).

Identification of potential transposons, prophages and direct sequence repeats of the *nif* island

We searched for potential transposable elements and prophages on the *nif* island using MGEfinder v2.3.0 [36] and phigaro v.1.0.6 [37] with default parameters, respectively. Potential direct sequence repeats surrounding the island were searched using BLASTN with 10 kb sequences at the left and right boundaries of the island as query and subject, respectively ("blastn –query left_10kb.fasta –subject right_10kb.fasta –word_size 7 –evalue 10").

Results and discussion

Newly isolated free-living strains expand ecological diversity of *Bradyrhizobium*

We isolated and sequenced the genomes of 93 *Bradyrhizobium* strains initially identified by 16S rRNA gene (five later shown to belong to *Afipia* based on phylogenomic construction) isolated from diverse types of soils, i.e., lateritic red earths, black soils, brown coniferous forest soils, cinnamon soils, and yellow-brown earths in different sites of China (Supplementary Fig. S1A). Among these isolates, 81 are non-symbiotic strains, as evidenced by the lack of the nodulation-determining *nod* genes, four of which carry *nif* genes. The remaining twelve are potential symbiotic strains: 11 of them carry *nod* genes and the other one (*B. denitrificans* SZCCT0094) is phylogenetically clustered with photosynthetic members which use a nod-independent strategy for nodulation [38] (Fig. 1, Supplementary Fig. S2); all of the twelve were isolated from the rhizosphere or bulk soil of leguminous plants (Supplementary Dataset S1). Together with the 215 publicly available *Bradyrhizobium* genomes, our dataset included 168 symbiotic (Sym), 21 free-living *nif*-carrying (FL_{nif}), and 119 free-living non-*nif*-carrying (FL_{nonnif}) strains (Supplementary Datasets S1 and S2).

Despite a significantly expanded dataset, all the *Bradyrhizobium* isolates fall into seven phylogenetic supergroups (Fig. 1) named in a recent study [2]. These are the Soil 1, Soil 2, *B. jicamae*, *B. elkanii*, Kakadu, Photosynthetic, and *B. japonicum* supergroups (Fig. 1). Among the 93 newly isolated strains, 88 contribute to all these supergroups except for the Soil 2 and Kakadu supergroups, and the remaining five strains are related to the genus *Afipia* in the outgroup. The newly isolated strains are phylogenetically nested within those that have publicly available genomes, except for the 13 strains in the *B. jicamae* supergroup which form a sister clade to the previously sequenced strains of that supergroup (Fig. 1).

Our ancestral lifestyle reconstruction analysis inferred that the last common ancestor (LCA) of *Bradyrhizobium*

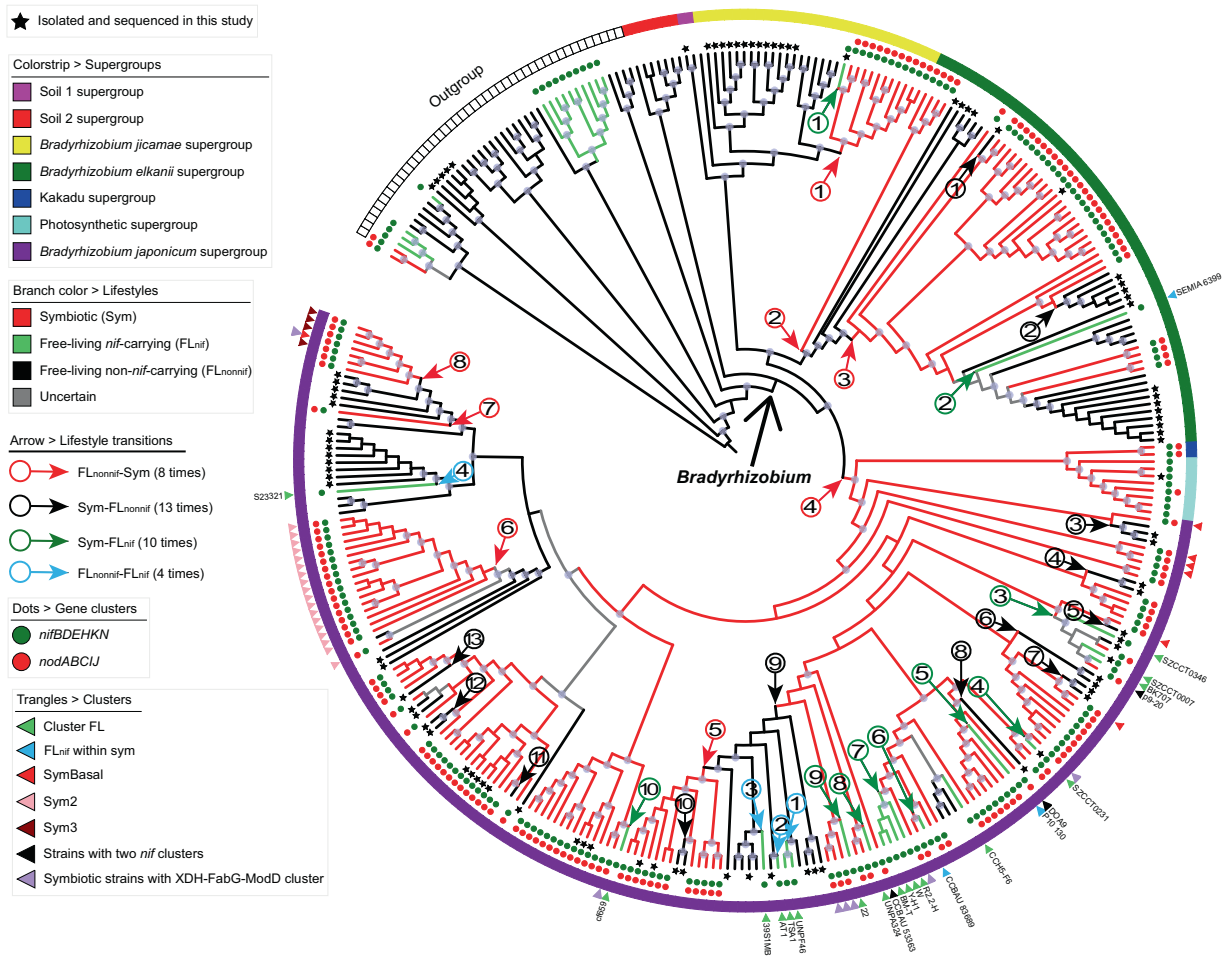

Fig. 1 The maximum-likelihood phylogenomic tree of *Bradyrhizobium* and inferred lifestyle evolutionary history. Strains from Xanthobacteraceae were used as an outgroup. Ancestral lifestyles were inferred using the parsimony method in Mesquite. Purple circles on the nodes indicate ultrafast bootstrap values higher than or equal to 95% calculated by IQ-Tree. The red and green triangles on the outer layer indicate symbiotic members of the basal lineages of the *B. japonicum* supergroup and free-living *nif*-carrying strains, respectively.

adapted to a free-living lifestyle, consistent with our prior study based on a very limited taxon sampling [9] (Fig. 1). This is evident by the pattern that a vast majority of the outgroup lineages as well as all members of the basal *Bradyrhizobium* clade (Soil 1 and Soil 2 supergroups) are non-symbiotic lineages. The symbiotic lifestyle independently originated eight times based on the sampled lineages, including three major origins: one in the *B. jicamae* supergroup (FL_{nonnif}-Sym1), one in the *B. elkanii* supergroup (FL_{nonnif}-Sym3), and another one at the LCA of the Kakadu, Photosynthetic, and *B. japonicum* supergroups (FL_{nonnif}-Sym4). These transitions occurred in relatively deep phylogenetic positions, suggesting that they represent evolutionarily ancient events. Because the ancestors of both the *B. jicamae* and *B. elkanii* supergroups were inferred to be free-living non-*nif*-carrying bacteria (Fig. 1), it is no surprise to observe that nearly half of the analyzed strains from these two supergroups are free-living non-*nif*-carrying strains. Free-living lifestyles originated from symbiotic strains 23 times: 13 of these events gave rise to non-*nif*-carrying strains and 10 gave rise to *nif*-carrying members (Fig. 1). Transitions from symbiotic to free-living lifestyle (both Sym-FL_{nif} and Sym-FL_{nonnif}) likely occurred recently as indicated by their shallow phylogenetic positions (Fig. 1). Hence, the loss of the ability to nodulate legume plants occurred more frequently than the reverse process in *Bradyrhizobium*, consistent with a prior study based on the analysis of phenotypic markers [7]. This pattern remains when we combined FL_{nif} and FL_{nonnif} as a single free-living lifestyle (Supplementary Fig. S3). Furthermore, the free-living *nif*-carrying lifestyle originated from free-living non-*nif*-carrying ancestors four times (Fig. 1). The infrequency of this type of lifestyle transition does not necessarily mean that this type of transition is rare, as it may result from a potential bacterial cultivation bias against those carrying the *nif* cluster under aerobic conditions with readily available N sources (see also Caveats and concluding remarks). The above patterns held true based on a condensed phylogeny with 90% ultrafast

bootstrap as the cutoff (Supplementary Fig. S4, branches that split in <90% of the sampled trees were condensed).

HGT of the *nif* island drives the expansion of the free-living *nif*-carrying lifestyle

Free-living *nif*-carrying *Bradyrhizobium* is of much interest, because these *Bradyrhizobium* members may play a previously unrecognized role in the N cycle in soils, especially given the reportedly high abundance of *Bradyrhizobium* in soils [6]. We asked where the N_2-fixing (*nif*) genes in free-living *Bradyrhizobium* came from and whether the *nif* genes differ between free-living and symbiotic members. It is apparent from Fig. 1 that most free-living strains evolved from their symbiotic ancestors, hence one possibility is that free-living *nif*-carrying strains inherited the same version of *nif* genes from their nodulating ancestors. Alternatively, free-living *nif*-carrying *Bradyrhizobium* might have lost the entire symbiosis island which includes both *nod* and *nif* clusters among other genes, and instead recruited a new set of *nif* genes from an external donor by HGT. To test these competing hypotheses, we constructed the phylogenetic tree of the *nif* cluster based on the concatenated alignment of the genes *nifABDEHKNX* for those from the *Bradyrhizobium* and several other rhizobia including *Azorhizobium*, *Rhizobium*, *Sinorhizobium* and *Mesorhizobium* from Alphaproteobacteria, and *Burkholderia* and *Cupriavidus* from Betaproteobacteria (see Materials and methods).

The *nif* genes of all *Bradyrhizobium* and the two strains from *Azorhizobium* form a monophyletic group (Fig. 2, Supplementary Fig. S5). The topology of *nif* tree suggests a single origin of *nif* genes in *Bradyrhizobium*, and the *nif* genes likely transmitted to *Azorhizobium* by HGT (Fig. 2). Within the *Bradyrhizobium* clade, the *nif* tree (Fig. 2) displays substantial topological incongruence with the species tree (Fig. 1), suggesting different evolutionary histories between *nif* genes and the core genome. Following its origin, the *nif* genes diverged into two major clades: the Sym1 clade consisting of mainly symbiotic strains and the other clade composed of photosynthetic and free-living strains as well as some symbiotic strains. For the latter clade, the Photosynthetic *Bradyrhizobium* supergroup together with a strain from the *B. jicamae* supergroup (SZCCT0283) and the two *Azorhizobium* sequences branched off first, followed by a group of strains from the *B. japonicum* supergroup (Fig. 2) with mixed evolutionary relatedness in the species tree (Fig. 1). These mixed *B. japonicum* supergroup members further diverged into four clades in the *nif* tree, three of which are constituted by symbiotic strains: SymBasal is largely comprised by several early-split lineages of the *B. japonicum* supergroup and spans over the species tree, whereas Sym2 and Sym3 are each composed of strains at shallow phylogenetic positions clustered in the species tree (Fig. 1). Apparently, HGT occurred within SymBasal and between these three clusters. The other clade is mostly composed of free-living *nif*-carrying members (Cluster FL) (Fig. 2) traversing the phylogenomic tree but restricted to the *B. japonicum* supergroup (Fig. 1), which is strong evidence for HGT of the *nif* genes among these free-living *Bradyrhizobium* members.

Strikingly, despite a scattered distribution in the species phylogeny (Fig. 1), 17 free-living *nif*-carrying strains group on the *nif* phylogeny (Cluster FL in Fig. 2). Also grouping in Cluster FL are the *nif* genes from three symbiotic strains (DOA9, CCBAU 53363 and p9-20). These three strains encode another *nif* cluster located in the symbiotic plasmid (DOA9) or symbiosis island on the chromosome (CCBAU 53363 and p9-20), which group with other symbiotic strains in the *nif* phylogeny (Fig. 2). Remarkably, as introduced above, two strains (*Bradyrhizobium* sp. AT1 and *Bradyrhizobium* sp. DOA9) whose *nif* genes from Cluster FL in the *nif* phylogeny were previously shown to perform free-living N_2 fixation under micro-aerobic conditions (marked by a box around the strain name in Fig. 2) [20, 21]. The above evidence together indicates that *nif* genes from Cluster FL, which is mainly comprised by free-living *Bradyrhizobium*, are likely to be a special group of *nif* genes that specifically take part in free-living N_2 fixation.

A unique *nif* island likely contributes to the oxygen tolerance of free-living *nif*-carrying *Bradyrhizobium*

We sought more evidence for HGT of the *nif* cluster by analyzing genes surrounding it. This led to the identification of a ~50 kb genomic region containing *nif* genes (*nif* island) conserved among all members of Cluster FL (Fig. 3A). The regions flanking the *nif* island are not conserved across most free-living strains (Supplementary Fig. S6), providing further evidence for HGT of the *nif* island. Note that four strains in Cluster FL (BM-T, Y-H1, R2.2-H and W) share highly conserved genomic contexts (Supplementary Fig. S6). These strains are closely related and comprise a monophyletic group in the phylogenomic tree (marked by a box around the strain names in Supplementary Fig. S2), indicating that the shared genomic context of their *nif* islands was inherited from their LCA. Likewise, the conserved genomic context of the *nif* island among strains from the Photosynthetic *Bradyrhizobium* supergroup (Supplementary Fig. S6) is consistent with their vertical descent shown in the species phylogeny (Fig. 1).

The upstream boundary of the *nif* island in Cluster FL *Bradyrhizobium* (Fig. 3A) is marked by *nifA*, which codes for a transcriptional activator required for the expression of *nif* operons [39]. Adjacent to *nifA* is a *suf* cluster, which is responsible for synthesizing and inserting Fe-S clusters into nitrogenase [40]. The core genes encoding the nitrogenase

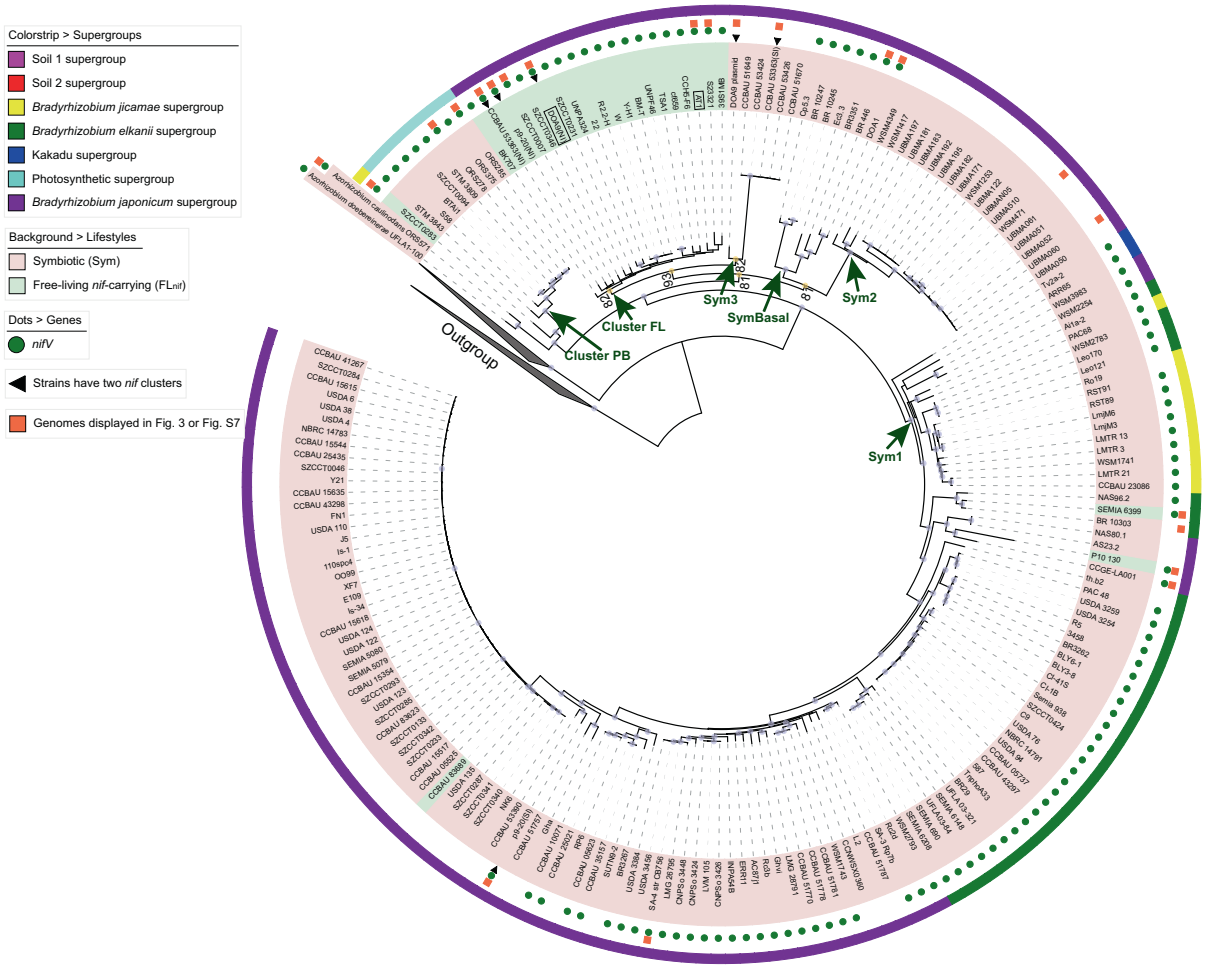

Fig. 2 *Nif* gene phylogeny of *Bradyrhizobium*. The phylogeny was constructed using the concatenated alignments of *nifABDEHKNX*. *Nif* genes from *Rhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Azorhizobium* and Beta-rhizobia (rhizobia from Betaproteobacteria) are used as the outgroup. The purple circles on the nodes indicate ultrafast bootstrap values higher than or equal to 95% calculated by IQ-Tree. The ultrafast bootstrap of some key nodes related to Cluster FL is also indicated. The tips with a black box around within Cluster FL denote those with experimental evidence for their N_2 fixation in the free-living state. Note that three symbiotic strains, namely DOA9, p9-20 and CCBAU 53363, carried a *nif* cluster similar to the one on the free-living *nif* island (NI) in addition to another one on the symbiosis island (SI) or plasmid. Members of Cluster FL, SymBasal, Sym2 and Sym3 are also marked on the phylogenomic tree (Fig. 1). The geographic locations of *nif*-carrying strains in Cluster FL are displayed in Supplementary Fig. S1B.

are located downstream of the *suf* cluster, and are mainly separated into two operons, *nifDKENX* and *nifHQ*. *FixABCX* which encode a membrane complex involved in electron transfer to nitrogenase are located downstream of the *nif* genes [41]. Extensively studied in rhizobia, the role of *fixABCX* in N_2 fixation by providing electrons for nitrogenase has been suggested for free-living diazotrophs [42]. At the downstream boundary of the *nif* island is a *mod* cluster encoding molybdate transporter. In general, the *nif* island conserved in free-living *Bradyrhizobium* is similar to that in photosynthetic strains (Fig. 3A), as also noted in previous studies based on genomic analysis of only a few photosynthetic and free-living strains [43, 44], and that in a newly sequenced soil-dwelling strain (SZCCT0283) from the *B. jicamae* supergroup (Supplementary Fig. S7A). A few differences are that those from Cluster PB carry an additional copy of *nifH* but fewer genes in the *mod* operon compared with those from Cluster FL (Fig. 3).

The *nif* clusters of the *nif* island conserved in Cluster PB and FL (Fig. 3A) share many genes with those on the symbiosis island (Fig. 3B), such as *nifDKENBH* and *fixBCX*, agreeing with their essential roles in N_2 fixation. However, they also display notable differences. For most symbiotic *Bradyrhizobium*, *nifH* gene is located together with *nifQ* on their genomes (Fig. 3B), but in SymBasal *nifH* is located in the upstream of the *nifDKENX* cluster (Fig. 3B, Supplementary Fig. S7B). A notable feature of Sym1 is that a gene cluster encoding *hya* and *hyp*, which is responsible for hydrogenase synthesis (*hyaABCDF*) and maturation (*hypVABFCDE*), is inserted between *fixJL* and *nifDKENX* gene clusters (Fig. 3B, Supplementary Fig. S7B). Although *nod* genes are usually separated far away from *nif* genes on

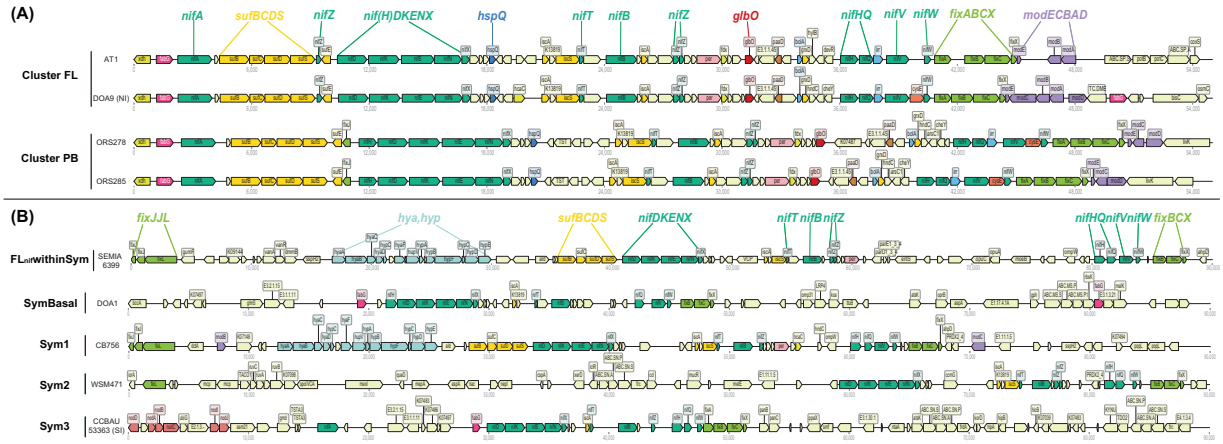

Fig. 3 The comparison of the gene arrangement of the *nif* gene cluster located in the *nif* island and in the symbiosis island. Gene functions are distinguished by different colors. The visualization of gene arrangement is performed with dna-features-viewer v3.0.3 [74]. The structures of the gene arrangement classified the *nif* clusters into three types: those belonging to the Cluster FL or Cluster PB (**A**), those found in the symbiosis island in most symbiotic strains or in FL_{nif} (free-living *nif*-carrying) strains that cluster with symbiotic strains (**B**). In panel **B**, the free-living strain SEMIA 6399 whose phylogenetic position is nested within symbiotic strains, labeled as "FL_{nif}within-Sym". "SymBasal" indicate basal lineages of the *B. japonicum* supergroup (referred to Fig. 2). "Sym1, 2, 3" stands for different cluster of symbiotic *Bradyrhizobium* denoted in Fig. 2. Note that on the symbiosis island of symbiotic *Bradyrhizobium*, *nod* genes are often separated far away from *nif* genes [43] so they are usually not shown in the figure.

the symbiosis island, these genes are located relatively close to *nif* genes in Sym3 (Fig. 3B, Supplementary Figs. S6, S7B).

Apparently, the gene arrangement of the *nif* island is more compact and more conserved across different free-living and photosynthetic strains than across symbiotic strains (Fig. 3). The *nif* island additionally carries a *nifV* gene, which encodes homocitrate synthase to synthesize homocitrate, a component of the Fe-Mo cofactor of nitrogenase [45]. A previous study reported that the vast majority of symbiotic rhizobia (not limited to *Bradyrhizobium*) do not harbor *nifV* [45], which is known to be compensated for by the homocitrate provided by their legume host during the symbiotic stage [46]. Here, we found that *nifV* is present in around half of the *Bradyrhizobium* genomes, not restricted to the free-living and photosynthetic strains (Fig. 2). Phylogenetic analysis suggests that *nifV* was presumably present at the LCA of *Bradyrhizobium*, but lost within the *B. japonicum* supergroup (Supplementary Fig. S8). This is consistent with the hypothesis that *nifV* was likely required in early *Bradyrhizobium* ancestors which took a free-living lifestyle (Fig. 1), and became dispensable later when the descendant *Bradyrhizobium* switched to a symbiotic environment in which homocitrate is readily available.

We further identified that several genes involved in O_2 tolerance and stress response are also specific to the *nif* island of free-living and photosynthetic members (Fig. 3), including *glbO* and *hspQ*. Specifically, *glbO* encodes a 17-kDa group-II truncated hemoglobin that binds O_2 with high affinity [47, 48]. The expression of truncated hemoglobin *glbO* is induced upon exposure to oxidative stress [49]. The product of *glbO* might therefore play an important role in protecting nitrogenase from O_2 inactivation, endowing the free-living *Bradyrhizobium* with higher tolerance to O_2. *HspQ* encodes a chaperone protein, which combats the detrimental effects on proteins caused by stressors such as increased temperature, oxidative stress, and heavy metals [50]. We also observed copy number differences between the free-living and symbiotic members, such as the *nifZ* which has a function in the maturation of the Fe-Mo protein nitrogenase [51], but whether the additional copies contribute to adaptation to the free-living lifestyle is unknown. In summary, we predict that in addition to *nifV*, the genes involved in oxygen tolerance and stress response in the *nif* island (e.g., *glbO*, *hspQ* and *mod*) may play a role in the adaptation to free-living lifestyle. Intriguingly, the *nif* island from *Azorhizobium*, which is embedded in the *Bradyrhizobium nif* tree (Fig. 2) and is able to perform free-living N_2 fixation [16], shares some of the above genes like *nifV*, *glbO* and *mod*, and a generally similar gene arrangement with the free-living *nif* island in *Bradyrhizobium* (Supplementary Fig. S7C). It would be interesting to conduct experiments to assess the detailed functions of the genes on the *nif* island under O_2 concentrations resembling those of soil environments.

An interesting finding is that despite the high similarity in gene arrangement of the *nif* island between Cluster PB and Cluster FL, these two groups cluster in different clades in the *nif* phylogeny (Fig. 2). Given the high similarity of the *nif* island between Cluster PB and Cluster FL strains (Fig. 3), it seems unlikely that the *nif* island is the result of convergent evolution between photosynthetic and free-

living *Bradyrhizobium* lineages. One possibility is that early in evolution Cluster FL acquired the *nif* island from Cluster PB or an even unsampled lineage related to Cluster PB that carries the *nif* island, and that the *nif* genes on the *nif* island might be later replaced by those derived from symbiotic lineages related to Sym3, Sym2 and SymBasal through recombination. To gain insights into this evolutionary process, it would be useful to sequence more genomes from related lineages in *Bradyrhizobium*.

Note that, for simplicity, we defined free-living strains based on the absence of *nod*. However, four strains claimed as free-living *nif*-carrying *Bradyrhizobium* in the current study, namely *B. mercantei* SEMIA 6399, *B. yuanmingense* P10 130, *B. liaoningense* CCBAU 83689 and *B. amphicarpaeae* 39S1MB, were originally isolated from nodules (Supplementary Dataset S2). Interestingly, except the last isolate whose *nif* genes were found in Cluster FL, the *nif* genes of the other three are embedded in symbiotic lineages in the *nif* phylogeny (Fig. 2), indicating that the *nif* genes were inherited from nodulating ancestors. Consistent with this idea, genes encoding hydrogenase are adjacent to *nif*, which is similar to Sym1 (Fig. 3, Supplementary Fig. S7B). These three strains also carry genes encoding components of T3SS and the effectors it injects (Supplementary Dataset S4), suggesting that they are likely nodule-inhabiting strains that may not nodulate legumes but with the potential to interact with plants. This is in contrast to other free-living *nif*-carrying strains where symbiosis genes were commonly lost during lifestyle transitions Sym-FL_{nonnif} 1, 4, 7 and 10 (Supplementary Fig. S9C; Supplementary Text S2; see also Supplementary Datasets S5–S8 for gene gains/losses during different lifestyle transitions). In general, the arrangement of *nif* genes and their surrounding regions of these strains are similar to that of symbiotic strains, particularly when compared with their closest relatives: P10 130 shows a very high similarity to its closest *nod*-carrying relative CCGE-LA001 in the genomic context of the *nif* genes, SEMIA 6399 is also highly similar to its closest *nod*-carrying relative BR 10303, in addition to the absence of *nifV* in BR 10303 (Supplementary Fig. S7B). We did not analyze CCBAU 83689 as the contig carrying *nif* genes is too short. The results indicate the different origins of the free-living *nif*-carrying *Bradyrhizobium*: while some were derived from their symbiotic ancestors, most analyzed strains of this lifestyle originated via HGT of the free-living *nif* island (Fig. 3).

Putative mechanisms of the horizontal transfer of the *nif* island

Genomic islands are defined as large genomic segments that have probably been horizontally acquired by prokaryotes, which are usually 10–200 kb in length [52, 53]. Clearly, the 50-kb region containing the genes necessary for N_2 fixation conserved among free-living *nif*-carrying *Bradyrhizobium* (*nif* island) well meets this criterion, as the genes it contains, not only *nif* genes (Fig. 2) but *fix* genes at the other end of the island (Supplementary Fig. S10), show a similar phylogeny that is very different from the species tree. Other features of mobile islands include different G + C content from the surrounding regions, the flanking of direct repeats or tRNA genes which serve as the sites for site-specific recombination [54, 55], and the presence of genes coding for transposable elements, integrases, or prophages (reviewed in [53]). However, we found that the *nif* island is rarely associated with any of the above features (Supplementary Fig. S11; see also Methods). This hints that the free-living *nif*-carrying *Bradyrhizobium*-specific *nif* island likely uses a mechanism distinct from other genomic islands for its movement across bacteria.

Intriguingly, we noticed that the flanking regions of the *nif* island (Fig. 3) are conserved in the symbiotic relatives of some free-living *nif*-carrying members (denoted by purple triangles in Fig. 1). As shown in Supplementary Fig. S12, this conserved flanking region consists of a gene encoding xanthine dehydrogenase (XDH), *fabG* which participates in fatty acid biosynthesis, a short gene of unknown function, and *modD* presumably involved in molybdate transport. In the strains carrying the *nif* island, these genes are separated by the *nif* island, with *modD* located on one side, and *fabG* and XDH located on the other side (Fig. 3). This opens up the possibility that homologous recombination is responsible for the spread of *nif* island among free-living *Bradyrhizobium*: recombination at both ends of the XDH-fabG-modD cluster can well lead to the insertion of the *nif* island. This mechanism, if true, suggests that the *nif* island can be horizontally transferred only between strains with the XDH-fabG-modD gene cluster. Such strains in the genomes analyzed in the current study were only found in the *B. japonicum* supergroup (Fig. 1), consistent with the limited distribution of free-living *nif*-carrying strains in this supergroup. This mechanism has been shown to be responsible for the gains and losses of genomic islands associated with pathogenic and commensal lifestyles among *Pseudomonas* lineages [56]. Further, other scenarios, such as illegitimate recombination, which could be mediated by non-homologous end joining as shown in a recent study [57], or hijacking the machinery produced by other genomic islands (e.g., by using the integrase and conjugation system encoded by other genomic island) [58], might also be possible.

Widespread distribution of the unique free-living *nif* island in diazotrophic communities

We further asked how prevalent the *nif* island specific to free-living *Bradyrhizobium* is in the diazotrophic communities of different environments and how it compares to the *nif* genes of

symbiotic members. We therefore assessed the normalized abundances of *nifH*, the most commonly used marker gene to identify N_2-fixing microbes [59], that potentially correspond to free-living (Cluster FL in Fig. 2), photosynthetic (Cluster PB in Fig. 2), and symbiotic *Bradyrhizobium* (Sym and SymBasal in Fig. 2) in 4958 amplicon sequencing datasets (see Materials and methods). We divided these datasets into four categories according to the environments where the samples were collected: soils (e.g., bulk soil, rhizosphere and freshwater lake sediment), plants (root and phyllosphere samples), marine (e.g., seawater, marine sediment, coral and mangrove samples), and others (e.g., bioreactor and unidentified samples) (Supplementary Dataset S3). Amplicon reads were assigned to different types of *Bradyrhizobium nifH* according to their phylogenetic placements (see Materials and methods).

Amplicon analysis showed that *Bradyrhizobium* constitute an important group of diazotrophic communities in different habitats, and the relative abundance of *Bradyrhizobium* was 7.52 ± 18.57%, 6.67 ± 12.75%, 17.39 ± 18.32%, 14.54 ± 18.85% (mean ± standard deviation) in marine, soil, plant, and other environmental types, respectively (Supplementary Dataset S3). The *nifH* that resembles those belonging to Cluster FL displayed the highest normalized abundance in samples from marine, soil, and habitats classified as "others" (Fig. 4; $p < 0.001$ for all comparisons, paired Wilcoxon–Mann–Whitney test). Specifically, Cluster FL accounts for 56 ± 47%, 46 ± 38%, and 51 ± 39% of reads that were assigned to *Bradyrhizobium* in these three types of habitats (marine, soil, and others), respectively. In particular, for marine samples, nearly all *Bradyrhizobium nifH* were assigned to Cluster FL (Fig. 4; Supplementary Dataset S3). For the datasets of plant samples, while symbiotic *nifH* displays the highest abundance (Fig. 4), *nifH* assigned to Cluster FL also shows a considerably high normalized abundance in plants (29 ± 39%), and is even the dominant ecotype in diazotrophic communities associated with some Asteraceae species and the roots of several perennial grass species (Supplementary Dataset S3). Hence, though the presence of free-living *Bradyrhizobium*-specific *nifH* does not necessarily indicate the occurrence of a complete *nif* island, it is tempting to suggest that *Bradyrhizobium* members carrying the free-living *nif* island i) are distributed in a variety of ecosystems and geographic locations (Supplementary Fig. S13), and ii) are in general more abundant than that of symbiotic strains in the genus in most habitats except plants, hinting at a previously unrecognized role of their *nif* island in the free-living lifestyle.

It is interesting that the relative abundance of *nifH* often exhibits a bimodal distribution (Fig. 4). To examine whether this is a potential result of combining different sample types in individual categories, e.g., soils contain bulk soils, rhizosphere and lake sediment, we analyzed the relative abundance of *nifH* in different subtypes of each environment category. As shown in Supplementary Fig. S14 stacked histogram, there is not a strong association between the relative abundance of a particular type of *nifH* and the specific environment subtype, although in some cases, for example, FL *nifH* appears specifically enriched in "coral metagenome" samples of marine amplicon sequencing datasets but are almost depleted from "mangrove metagenome" samples. Similarly, for the plant-associated datasets, *nifH* from symbiotic members (Sym) is enriched in "root metagenome" samples, while FL *nifH* exhibits a high proportion in phyllosphere samples.

Caveats and concluding remarks

Adding the 88 soil-dwelling strains greatly expands our knowledge of the ecology and evolution of *Bradyrhizobium*.

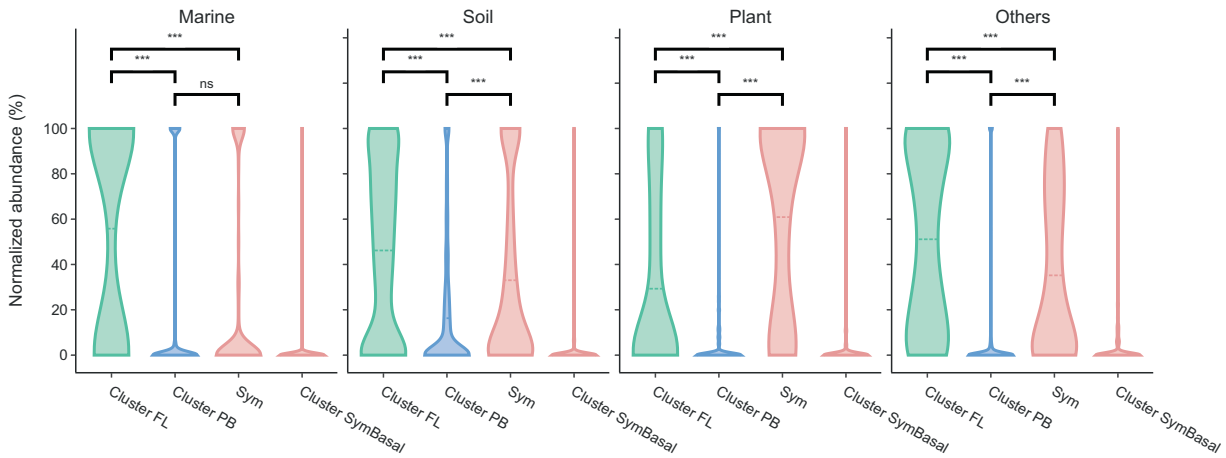

Fig. 4 Normalized abundance of free-living and symbiotic *nifH* of *Bradyrhizobium* in amplicon sequencing samples collected from different types of environments. The values shown as percentage in the y-axis in the violin plots denote the normalized abundance (see Materials and methods) of different types of *nifH*. The width of each curve represents the kernel estimation showing the frequency of samples at different normalized abundances. The *P* value is derived from a paired Wilcoxon–Mann–Whitney test (two-sided).

However, because of the bias towards symbiotic strains in previous studies and the difficulties in cultivating those dwelling in soil environments, the free-living strains analyzed here may underestimate the real diversity of wild diazotrophic *Bradyrhizobium* that adapt to a free-living lifestyle. For example, prior studies have shown that *Bradyrhizobium* is the dominant genus in the diazotrophic communities in many soil ecosystems, most of which have not been sampled for *Bradyrhizobium* cultivation however [60, 61]. Of the limited types of soil ecosystems sampled here, the *Bradyrhizobium* strains we collected are not necessarily representative of the wild populations in these ecosystems owing to the potential cultivation bias. For example, our cultivation process was under aerobic condition and the cultivation medium was rich in fixed nitrogen, conditions disfavoring the isolation of diazotrophic members. Therefore, it is not clear whether important free-living diazotrophic lineages occupying distinct phylogenetic positions are missing. Another caveat to the current conclusion is that some of our analyses built on an over-simplified classification of *Bradyrhizobium* lifestyles, which was based on the presence/absence of certain genes like *nif* and *nod*. In fact, isolates with a symbiosis island could be ineffective (i.e., those form nodules but fix little N_2 within nodules) [62]. Likewise, whether the free-living strains possessing the *nif* island could perform N_2 fixation, and if they could, the N_2 fixation efficiency under different O_2 concentrations, remains to be studied. Another overlooked issue related to lifestyle analysis is that lifestyle transitions may cover intermediate state, which cannot be accounted for by state-of-the-art ancestral reconstruction methods. Specifically, one possible route may include a non-*nif*-carrying free-living intermediate during the transition from a symbiotic ancestor to a *nif*-carrying free-living descendant, which is represented as $Sym \rightarrow FL_{nonnif} \rightarrow FL_{nif}$. An alternative trajectory can be $Sym \rightarrow (Sym + FL_{nif}) \rightarrow FL_{nif}$, in which the intermediate state was a free-living strain that carries a symbiosis island, as represented by the three strains carrying both a *nif* island and a symbiosis island or plasmid (AT1, DOA9, and p9-20). Besides, two of the *nif*-carrying strains that lack *nod* (SEMIA 6399 and P10 130) analyzed here were reported to nodulate legumes despite a high host specificity [63, 64]. However, this contradicts with the current understanding that *Bradyrhizobium* members that lack *nod* but can nodulate legumes are only found in the Photosynthetic supergroup [65]. Hence, whether they can use a *nod*-independent strategy for nodulation, and if so, how this mechanism works, remain further investigation.

It is also important to note that it is the relative abundance that was used to measure the prevalence of free-living *nifH* (Fig. 4). Hence, a high relative abundance of a certain type of *Bradyrhizobium nifH* only indicates its high abundance relative to other N_2-fixing members in the sampling sites. Actually, in some environments, the overall rates of N_2 fixation were very low [66]. Further, amplicon analysis suffers from biases ranging from sample preparation, primer selection, and chimeric sequences generated by sequencing and bioinformatics analysis. This could make the results based on different samples collected from various habitats difficult to compare. Besides, it would be interesting to compare the abundances of *nif*-carrying and non-*nif*-carrying *Bradyrhizobium* members from the same sampling sites, which could be done by measuring the abundances of both *nifH* and housekeeping genes that can provide a high taxonomic resolution.

In spite of the above caveats, our study reveals an interesting pattern of origins of free-living *nif*-carrying *Bradyrhizobium* from their symbiotic ancestors. Though the nested phylogenetic positions within symbiotic lineages (Fig. 1) might leave an impression that free-living *nif*-carrying members could have inherited the *nif* genes from their symbiotic ancestors, we provided compelling evidence that this is unlikely the case. Rather, it is the HGT of a conserved *nif* island that drives the independent transitions from symbiotic to free-living *nif*-carrying lineages. This might serve as a classic example of independent transitions in lifestyle driven by HGT of certain genes in bacteria, which, although mostly explored in pathogens or symbionts in prior studies [67–69], may play a prominent role in free-living bacteria. Given the global dominance of *Bradyrhizobium* in the soil microbiota, even a small proportion of them being diazotrophic members could potentially bring a considerable amount of fixed nitrogen to the bulk soils and other nitrogen-limited terrestrial ecosystems. Our results therefore have implications for understanding nitrogen fixation. Because one of the major aims of modern agricultural research is to transfer the N_2 fixation ability to non-legume crops [70, 71], the capacity of certain free-living *Bradyrhizobium* to fix N_2 and their association with non-leguminous plants may imply potential application in agriculture.

Data availability

The genomic sequences and raw reads of the 88 sequenced *Bradyrhizobium* and five *Afipia* strains are available at NCBI BioProject (accession: PRJNA698083).

Code availability

The Python [72] and Ruby [73] codes used for phylogenomic and comparative genomics analyses are deposited in the online repository https://github.com/luolab-cuhk/Bradyrhizobium-Nif-HGT.

Acknowledgements We thank Xiaoyuan Feng for help with genome assembly. We appreciate Shuang Wang and Jie Liu for providing soil

samples and Jianjun Xu for assisting in sample collection. We are grateful to Xingqin Lin for her helpful suggestions in medium design. We also thank Qin Li, Yan Li, Xiaojun Wang, Xiao Chu, Hao Zhang, Danli Luo and Mei Xie for their helpful discussion. This work was supported by the National Natural Science Foundation of China (92051113), the Hong Kong Research Grants Council Area of Excellence Scheme (AoE/M-403/16), the Direct Grant of CUHK (4053495), and The CUHK Impact Postdoctoral Fellowship Scheme to SW.

Compliance with ethical standards

Conflict of interest The authors declare no competing interests.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Ormeno-Orrillo E, Martinez-Romero E. A genomotaxonomy view of the Bradyrhizobium genus. Front Microbiol. 2019;10:1334.
2. Avontuur JR, Palmer M, Beukes CW, Chan WY, Coetzee MPA, Blom J, et al. Genome-informed Bradyrhizobium taxonomy: where to from here? Syst Appl Microbiol. 2019;42:427–39.
3. Andrews M, Andrews ME. Specificity in legume-rhizobia symbioses. Int J Mol Sci. 2017;18:705.
4. Parker MA. The spread of Bradyrhizobium lineages across host legume clades: from Abarema to Zygia. Micro Ecol. 2015;69:630–40.
5. Garrido-Oter R, Nakano RT, Dombrowski N, Ma KW, McHardy AC, Schulze-Lefert P, et al. Modular traits of the Rhizobiales root microbiota and their evolutionary relationship with symbiotic rhizobia. Cell Host Microbe. 2018;24:155–67.
6. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-Gonzalez A, Eldridge DJ, Bardgett RD, et al. A global atlas of the dominant bacteria found in soil. Science. 2018;359:320–5.
7. Hollowell AC, Regus JU, Gano KA, Bantay R, Centeno D, Pham J, et al. Epidemic spread of symbiotic and non-symbiotic Bradyrhizobium genotypes across California. Micro Ecol. 2016;71:700–10.
8. VanInsberghe D, Maas KR, Cardenas E, Strachan CR, Hallam SJ, Mohn WW. Non-symbiotic Bradyrhizobium ecotypes dominate North American forest soils. ISME J. 2015;9:2435–41.
9. Wang S, Meade A, Lam H, Luo H. Evolutionary timeline and genomic plasticity underlying the lifestyle diversity in Rhizobiales. Msystems. 2020;5:e00438–00420.
10. Peoples MB, Craswell ET. Biological nitrogen fixation: Investments, expectations and actual contributions to agriculture. Plant Soil. 1992;141:13–39.
11. Cleveland CC, Townsend AR, Schimel DS, Fisher H, Howarth RW, Hedin LO, et al. Global patterns of terrestrial biological nitrogen (N_2) fixation in natural ecosystems. Glob Biogeochem Cy. 1999;13:623–45.
12. Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. Plant Soil. 2008;311:1–18.
13. Vitousek PM, Menge DN, Reed SC, Cleveland CC. Biological nitrogen fixation: rates, patterns and ecological controls in terrestrial ecosystems. Philos Trans R Soc Lond B Biol Sci. 2013;368:20130119.
14. Reed SC, Cleveland CC, Townsend AR. Relationships among phosphorus, molybdenum and free-living nitrogen fixation in tropical rain forests: results from observational and experimental analyses. Biogeochemistry. 2013;114:135–47.
15. Matson AL, Corre MD, Burneo JI, Veldkamp E. Free-living nitrogen fixation responds to elevated nutrient inputs in tropical montane forest floor and canopy soils of southern Ecuador. Biogeochemistry. 2015;122:281–94.
16. Lee KB, De Backer P, Aono T, Liu CT, Suzuki S, Suzuki T, et al. The genome of the versatile nitrogen fixer Azorhizobium caulinodans ORS571. BMC Genomics. 2008;9:1–14.
17. Dreyfus B, Garcia JL, Gillis M. Characterization of Azorhizobium Caulinodans gen. nov,, sp. nov., a stem-nodulating nitrogen-fixing bacterium isolated from Sesbania rostrata. Int J Syst Bacteriol. 1988;38:89–98.
18. Dreyfus BL, Elmerich C, Dommergues YR. Free-living Rhizobium strain able to grow on N2 as the sole nitrogen source. Appl Environ Microbiol. 1983;45:711–3.
19. Alazard D. Nitrogen fixation in pure culture by rhizobia isolated from stem nodules of tropical Aeschynomene species. FEMS Microbiol Lett. 1990;68:177–82.
20. Terakado-Tonooka J, Fujihara S, Ohwaki Y. Possible contribution of Bradyrhizobium on nitrogen fixation in sweet potatoes. Plant Soil. 2013;367:639–50.
21. Wongdee J, Boonkerd N, Teaumroong N, Tittabutr P, Giraud E. Regulation of nitrogen fixation in Bradyrhizobium sp. strain DOA9 involves two distinct NifA regulatory proteins that are functionally redundant during symbiosis but not during free-living growth. Front Microbiol. 2018;9:1644.
22. Hunt S, Layzell DB. Gas exchange of legume nodules and the regulation of nitrogenase activity. Annu Rev Plant Phys. 1993;44:483–511.
23. Gallon JR. Reconciling the incompatible: N_2 fixation and O_2. N. Phytol. 1992;122:571–609.
24. Coombs JT, Franco CMM. Isolation and identification of actinobacteria from surface-sterilized wheat roots. Appl Environ Microbiol. 2003;69:5603–8.
25. Sachs JL, Kembel SW, Lau AH, Simms EL. In situ phylogenetic structure and diversity of wild Bradyrhizobium communities. Appl Environ Microbiol. 2009;75:4727–35.
26. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, Seo H, et al. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. Int J Syst Evol Microbiol. 2017;67:1613–7.
27. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012;19:455–77.
28. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014;30:2068–9.
29. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015;25:1043–55.
30. Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015;32:268–74.
31. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019;20:1–14.
32. Maddison WP, Maddison DR Mesquite: a modular system for evolutionary analysis. Version 3.61. 2019. http://www.mesquiteproject.org/.
33. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, et al. QIIME allows analysis of high-throughput community sequencing data. Nat Methods. 2010;7:335–6.
34. Berger SA, Stamatakis A. Aligning short reads to reference alignments and trees. Bioinformatics. 2011;27:2068–75.
35. Barbera P, Kozlov AM, Czech L, Morel B, Darriba D, Flouri T, et al. EPA-ng: massively parallel evolutionary placement of genetic sequences. Syst Biol. 2019;68:365–9.
36. Durrant MG, Li MM, Siranosian BA, Montgomery SB, Bhatt AS. A bioinformatic analysis of integrative mobile genetic elements

highlights their role in bacterial adaptation. Cell Host Microbe. 2020;27:140–53.
37. Starikova EV, Tikhonova PO, Prianichnikov NA, Rands CM, Zdobnov EM, Ilina EN, et al. Phigaro: high-throughput prophage sequence annotation. Bioinformatics. 2020;36:3882–4.
38. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E, Avarre JC, et al. Legumes symbioses: absence of nod genes in photosynthetic bradyrhizobia. Science. 2007;316:1307–12.
39. Dixon R, Kahn D. Genetic regulation of biological nitrogen fixation. Nat Rev Microbiol. 2004;2:621–31.
40. Outten FW, Djaman O, Storz G. A suf operon requirement for Fe-S cluster assembly during iron starvation in Escherichia coli. Mol Microbiol. 2004;52:861–72.
41. Edgren T, Nordlund S. The fixABCX genes in Rhodospirillum rubrum encode a putative membrane complex participating in electron transfer to nitrogenase. J Bacteriol. 2004;186:2052–60.
42. Ledbetter RN, Costas AMG, Lubner CE, Mulder DW, Tokmina-Lukaszewska M, Artz JH, et al. The electron bifurcating FixABCX protein complex from Azotobacter vinelandii: generation of low-potential reducing equivalents for nitrogenase catalysis. Biochemistry. 2017;56:4177–90.
43. Okubo T, Piromyou P, Tittabutr P, Teaumroong N, Minamisawa K. Origin and evolution of nitrogen fixation genes on symbiosis islands and plasmid in Bradyrhizobium. Microbes Environ. 2016;31:260–7.
44. Okubo T, Tsukui T, Maita H, Okamoto S, Oshima K, Fujisawa T, et al. Complete genome sequence of Bradyrhizobium sp. S23321: insights into symbiosis evolution in soil oligotrophs. Microbes Environ. 2012;27:306–15.
45. Nouwen N, Arrighi JF, Cartieaux F, Chaintreuil C, Gully D, Klopp C, et al. The role of rhizobial (NifV) and plant (FEN1) homocitrate synthases in Aeschynomene/photosynthetic Bradyrhizobium symbiosis. Sci Rep. 2017;7:1–10.
46. Hakoyama T, Niimi K, Watanabe H, Tabata R, Matsubara J, Sato S, et al. Host plant genome overcomes the lack of a bacterial gene for symbiotic nitrogen fixation. Nature. 2009;462:514–7.
47. Liu C, He Y, Chang Z. Truncated hemoglobin o of Mycobacterium tuberculosis: the oligomeric state change and the interaction with membrane components. Biochem Bioph Res Co. 2004;316:1163–72.
48. Pathania R, Navani NK, Rajomohan G, Dikshit KL. Mycobacterium tuberculosis Hemoglobin HbO associates with membranes and stimulates cellular respiration of recombinant Escherichia coli. J Biol Chem. 2002;277:15293–302.
49. Ascenzi P, De Marinis E, Coletta M, Visca P. H_2O_2 and NO scavenging by Mycobacterium leprae truncated hemoglobin O. Biochem Bioph Res Co. 2008;373:197–201.
50. Shimuta T, Nakano K, Yamaguchi Y, Ozaki S, Fujimitsu K, Matsunaga C, et al. Novel heat shock protein HspQ stimulates the degradation of mutant DnaA protein in Escherichia coli. Genes Cells. 2004;9:1151–66.
51. Jimenez-Vicente E, Yang ZY, del Campo JSM, Cash VL, Seefeldt LC, Dean DR. The NifZ accessory protein has an equivalent function in maturation of both nitrogenase MoFe protein P-clusters. J Biol Chem. 2019;294:6204–13.
52. Langille MGI, Hsiao WWL, Brinkman FSL. Detecting genomic islands using bioinformatics approaches. Nat Rev Microbiol. 2010;8:372–82.
53. Hacker J, Kaper JB. Pathogenicity islands and the evolution of microbes. Annu Rev Microbiol. 2000;54:641–79.
54. Krupovic M, Gribaldo S, Bamford DH, Forterre P. The evolutionary history of archaeal MCM helicases: a case study of vertical evolution combined with hitchhiking of mobile genetic elements. Mol Biol Evol. 2010;27:2716–32.
55. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T, Sasamoto S, et al. Complete genomic sequence of nitrogen-fixing symbiotic bacterium Bradyrhizobium japonicum USDA110. DNA Res. 2002;9:189–97.
56. Melnyk RA, Hossain SS, Haney CH. Convergent gain and loss of genomic islands drive lifestyle changes in plant-associated Pseudomonas. ISME J. 2019;13:1575–88.
57. Dupuy P, Sauviac L, Bruand C. Stress-inducible NHEJ in bacteria: function in DNA repair and acquisition of heterologous DNA. Nucleic Acids Res. 2019;47:1335–49.
58. Cury J. Evolutionary genomics of conjugative elements and integrons. Paris, France: PhD dissertation, Université Sorbonne Paris Cité; 2017.
59. Gaby JC, Buckley DH. A global census of nitrogenase diversity. Environ Microbiol. 2011;13:1790–9.
60. Han LL, Wang Q, Shen JP, Di HJ, Wang JT, Wei WX, et al. Multiple factors drive the abundance and diversity of the diazotrophic community in typical farmland soils of China. FEMS Microbiol Ecol. 2019;95:fiz113.
61. Meng H, Zhou Z, Wu R, Wang Y, Gu J. Diazotrophic microbial community and abundance in acidic subtropical natural and revegetated forest soils revealed by high-throughput sequencing of nifH gene. Appl Microbiol Biot. 2019;103:995–1005.
62. Gano-Cohen KA, Wendlandt CE, Stokes PJ, Blanton MA, Quides KW, Zomorrodian A, et al. Interspecific conflict and the evolution of ineffective rhizobia. Ecol Lett. 2019;22:914–24.
63. Helene LCF, Delamuta JRM, Ribeiro RA, Hungria M. Bradyrhizobium mercantei sp nov., a nitrogen-fixing symbiont isolated from nodules of Deguelia costata (syn. Lonchocarpus costatus). Int J Syst Evol Micr. 2017;67:1827–34.
64. Toniutti MA, Fornasero LV, Albicoro FJ, Martini M, Draghi W, Alvarez F, et al. Nitrogen-fixing rhizobial strains isolated from Desmodium incanum DC in Argentina: Phylogeny, biodiversity and symbiotic ability. Syst Appl Microbiol. 2017;40:297–307.
65. Okazaki S, Tittabutr P, Teulet A, Thouin J, Fardoux J, Chaintreuil C, et al. Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. ISME J. 2016;10:64–74.
66. Nelson MB, Martiny AC, Martiny JBH. Global biogeography of microbial nitrogen-cycling traits in soil. Proc Natl Acad Sci USA. 2016;113:8033–40.
67. Remigi P, Zhu J, Young JPW, Masson-Boivin C. Symbiosis within symbiosis: Evolving nitrogen-fixing legume symbionts. Trends Microbiol. 2016;24:63–75.
68. Novick RP, Ram G. The Floating (Pathogenicity) Island: A Genomic Dessert. Trends Genet. 2016;32:114–26.
69. Gal-Mor O, Finlay BB. Pathogenicity islands: a molecular toolbox for bacterial virulence. Cell Microbiol. 2006;8:1707–19.
70. Soumare A, Diedhiou AG, Thuita M, Hafidi M, Ouhdouch Y, Gopalakrishnan S, et al. Exploiting biological nitrogen fixation: a route towards a sustainable agriculture. Plants-Basel. 2020;9:1011.
71. Santi C, Bogusz D, Franche C. Biological nitrogen fixation in non-legume plants. Ann Bot-Lond. 2013;111:743–67.
72. Cock PJA, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics. 2009;25:1422–3.
73. Goto N, Prins P, Nakao M, Bonnal R, Aerts J, Katayama T. BioRuby: bioinformatics software for the Ruby programming language. Bioinformatics. 2010;26:2617–9.
74. Zulkower V, Rosser S. DNA Features Viewer: a sequence annotation formatting and plotting library for Python. Bioinformatics. 2020;36:4350–2.