Genome-informed *Bradyrhizobium* taxonomy: where to from here?

Juanita R. Avontuur[1], Marike Palmer[1], Chrizelle W. Beukes[1], Wai Y. Chan[2], Martin P.A. Coetzee[1], Jochen Blom[3], Tomasz Stępkowski[4], Nikos C. Kyrpides[5], Tanja Woyke[5], Nicole Shapiro[5], William B. Whitman[6], Stephanus N. Venter[1], Emma T. Steenkamp[1]*

[1] Department of Biochemistry, Genetics and Microbiology (BGM), Forestry and Agricultural Biotechnology Institute (FABI), University of Pretoria, Pretoria, South Africa

[2] Biotechnology Platform (BTP), Agricultural Research Council Onderstepoort Veterinary Institute (ARC-OVI), Onderstepoort, 0110, South Africa

[3] Bioinformatics and Systems Biology, Justus-Liebig-University Giessen, Giessen, Germany

[4] Autonomous Department of Microbial Biology, Faculty of Agriculture and Biology, Warsaw University of Life Sciences (SGGW), Poland

[5] DOE Joint Genome Institute, Walnut Creek, CA, United States

[6] Department of Microbiology, University of Georgia, Athens, GA, United States

* Corresponding author:

E.T. Steenkamp; Private Bag X20, Hatfield, 0028, South Africa; Email: emma.steenkamp.up.ac.za; Tel.: +27 012 420 3262

Abstract

Bradyrhizobium is thought to be the largest and most diverse rhizobial genus, but this is not reflected in the number of described species. Although it was one of the first rhizobial genera, its taxonomy remains complex. Various contemporary studies are showing that genome sequence information may simplify taxonomic decisions. Therefore, the growing availability of genomes for *Bradyrhizobium* will likely aid in delineation and characterization of new species. In this study, we addressed two aims: first, we reviewed the availability and quality of available genomic resources for *Bradyrhizobium*. This was achieved by comparing genome sequences in terms of sequencing technologies used and estimated level of completeness for inclusion in genome-based phylogenetic analyses. Secondly, we utilized these genomes to investigate the taxonomic standing of *Bradyrhizobium* in light of its diverse lifestyle. Although genome sequences differed in terms of their quality and completeness, our data indicate that the use of these genome sequences is adequate for taxonomic purposes. By using these resources, we inferred a fully resolved, well-supported phylogeny. It separated *Bradyrhizobium* into seven lineages, three of which corresponded to the so-called supergroups. Wide distribution of key lifestyle traits such as nodulation, nitrogen fixation and photosynthesis revealed that these traits have complicated evolutionary histories. This is the first robust *Bradyrhizobium* species phylogeny based on genome sequence information for investigating the evolution of this important assemblage of bacteria. Furthermore, this study provides the basis for using genome sequence information as a resource to make important taxonomic decisions, particularly at the species and genus levels.

Keywords

Phylogenomics, *Bradyrhizobium japonicum*, *Bradyrhizobium elkanii*, photosynthetic *Bradyrhizobium*, nodulation, photosynthesis

1. Introduction

The genus *Bradyrhizobium* was introduced in the early nineteen-eighties to accommodate the so-called slow-growing rhizobia (i.e., bacteria capable of nodulating and fixing atmospheric nitrogen in symbiosis with legumes) [25, 51]. The first validly described species was *B. japonicum* [51] which was followed nearly a decade later with the description of *Bradyrhizobium elkanii* [57] and *Bradyrhizobium liaoningense* [122]. Since then, the number of descriptions for new *Bradyrhizobium* species have increased drastically. Currently, the genus contains 53 species [3, 15, 49, 104], of which the names of 39 have been validly published and have standing in nomenclature ([32, 83, 84] www.bacterio.net). However, contemporary diversity studies suggest that most *Bradyrhizobium* species still remain to be discovered and formally described (e.g., [90, 108]).

Like most bacterial genera, the taxonomic status of *Bradyrhizobium* is constantly modified and updated as new experimental evidence emerges. For example, various species previously designated as belonging to other genera have now been transferred to *Bradyrhizobium* as a result of such studies. Based on 16S rRNA gene sequence similarity with *B. japonicum*, *Blastobacter denitrificans*, an aquatic budding bacterium capable of nodulating *Aeschynomene indica*, is now recognized as *Bradyrhizobium denitrificans* [111]. Similarly, *Agromonas oligotrophica*, isolated from paddy soils in Japan, was renamed as *B. oligotrophicum* [89]. Most recently, *Rhizobium lupini* was transferred to *Bradyrhizobium* based on empirical data [86]. Such data have also been used to amend or correct the taxonomy of described *Bradyrhizobium* species. For example, genotypic and phenotypic data were used to delineate and describe *B. diazoefficiens* [27], which is a soybean symbiont previously known as *B. japonicum* Group Ia [46].

As currently circumscribed, *Bradyrhizobium* includes species with diverse lifestyles [18, 28, 50, 112]. They can live symbiotically or endophytically with legume and non-legume hosts [18, 28, 69] and some are photosynthetic [25, 99]. Many are free-living in soil [112], while some apparently thrive in the clinical environment [11, 64]. When *Bradyrhizobium* associates with legumes, it usually has a broad geographic distribution, especially in tropical and subtropical areas [28, 69, 101]. Furthermore, these rhizobial members of the genus can fix nitrogen in symbiosis with early and later emerging legume clades in the Papilionoideae and Caesalpinioideae subfamilies of the Fabaceae [6, 10, 34, 36, 80, 92, 101, 113]. *Bradyrhizobium* strains are also known to be the predominant

symbionts of most nodulating legumes [80]. It is therefore not surprising that *Bradyrhizobium* is widely believed to have been the ancestral legume symbiont [80, 101] .

Bradyrhizobium is particularly important from an agricultural perspective. Many agriculturally important legumes can establish symbioses with members of this genus [48]. Also, *Bradyrhizobium* nodulates native grain legume crops that are commonly farmed by local smallholder and subsistence farmers in India and Africa [48, 75]. Additionally, countries such as Brazil and China rely on *Bradyrhizobium* strains as commercial inoculants for soybean production [4, 116]. However, sustainable agricultural practices and improved legume crop production are severely constrained by changes in environmental, climatic and social conditions [4, 43, 48]. Research should therefore aim to improve our understanding of biological nitrogen fixation, particularly regarding the *Bradyrhizobium*-legume symbiosis [4, 43, 48]. Detailed knowledge of the diversity and species richness of this genus would undoubtedly contribute to the discovery and development of well-adapted inoculants for more cost-effective and sustainable approaches, especially for local smallholder farmers and surrounding communities [43].

Phylogenetic analyses of *Bradyrhizobium* typically recover three main lineages, which are informally referred to as supergroups or superclades [63, 104, 114]. These supergroups are represented by *B. japonicum, B. elkanii* and the photosynthetic *Bradyrhizobium* species [63, 72, 114]. Members of the *B. japonicum* supergroup have been isolated from various continents and climatic zones [104, 114] as opposed to the *B. elkanii* supergroup that is predominantly recovered from legumes occurring in subtropical and tropical regions [28, 69]. As a result, the *B. japonicum* supergroup is thought to be associated with a much broader geographical distribution (and possibly host range) than the *B. elkanii* supergroup [107, 114]. However, the Photosynthetic supergroup has so far only been associated with *Aeschynomene* species, which are aquatic legumes found in tropical regions [72]. This vast phenotypic difference has previously been proposed as a potential differentiating trait for moving the Photosynthetic supergroup to a new genus [97, 110, 119].

Uncertainties pertaining to the sub-generic structure of *Bradyrhizobium*, particularly in terms of the Photosynthetic supergroup [96, 97, 119], can only be unambiguously resolved with the aid of a robust evolutionary hypothesis for the genus as shown in other studies [7, 79, 87]. Previous reconstructions of the phylogenetic history of the genus were mostly incongruent with one another

(e.g., [115, 119]) or lacked statistical support at the deeper nodes, especially those leading to the supergroups (e.g., [10, 63, 106]). These include studies that utilized the standard bacterial taxonomic marker, 16S rRNA gene sequence, which lacks resolving power in *Bradyrhizobium* (e.g., [28, 47, 63, 105]), as well as multilocus sequence analysis (MLSA) approaches based on the use of three to five housekeeping gene sequences (e.g., [47, 63, 93]). Although inadequate taxon sampling (among ingroup and outgroup taxa) could account for some of these spurious results [44, 126], the main cause was likely the lack of sufficient phylogenetic signal for reconstructing deeper phylogenetic nodes in such small datasets [19, 35, 45].

Various recent studies have demonstrated that a large number of gene sequences, shared within a bacterial genus and its outgroup taxa, is required for unequivocally inferring its phylogeny [9, 31, 79]. This approach is currently regarded as the most reliable method for generating species trees [19, 61], because it employs large datasets with sufficient phylogenetic signal for overcoming the non-phylogenetic signal or noise that dominate smaller datasets [19, 35, 79]. The development of such large datasets is increasingly feasible with the growing availability of whole genome sequences for bacteria [65, 117]. This is also true for *Bradyrhizobium*, from only *B. diazoefficiens* (previously known as *B. japonicum* USD110) having a genome sequence available in 2002 [27, 53], there are now many whole genome sequences for isolates of this genus available in public domain databases such as those of the National Centre for Biotechnology Information (NCBI, http://www.ncbi.nlm.nih.gov/) and the Joint Genome Institute (JGI, https://jgi.doe.gov).

Due to the vast increase in available genomic data for *Bradyrhizobium* the aim of this study was twofold: i) to review the taxonomic breadth and quality of the current genomic resources available for this genus and its relatives in the public domain, and ii) to utilize these genomes for investigating the taxonomic standing of *Bradyrhizobium* in light of the published literature and the diverse lifestyles reported for these bacteria. To address the first aim, all available genome sequences of *Bradyrhizobium* species and closely related genera were obtained and compared in terms of sequencing technologies used, completeness level and value for inclusion in genome-based phylogenetic analyses. We also complemented these existing resources by sequencing the genomes of eight additional *Bradyrhizobium* isolates, some potentially representing novel species [10, 103]. The second aim was achieved by first utilizing these genomic resources (combined with adequate taxon sampling, especially among the outgroups) to generate a robust phylogeny for the genus. This

genome-based evolutionary hypothesis was then compared to those inferred using conventional MLSA. Finally, we investigated the distribution of key predicted lifestyle traits (particularly, nodulation, diazotrophy and photosynthesis) across the *Bradyrhizobium* phylogeny. Our findings would thus be fundamental, not only for understanding evolutionary processes, but also for providing clarity regarding the taxonomic status of this important genus and its species.

2. Materials and Methods

2.1 Whole-genome Sequences

All *Bradyrhizobium* strains with publicly available genome sequences were included in this study. For outgroup purposes, we used the genome sequences of species from genera that are closely related to *Bradyrhizobium* as indicated by the most recent version of the 16S rRNA-based All Living Species tree (https://www.arb-silva.de/projects/living-tree/) [123]. These included genera in the Bradyrhizobiaceae (*Nitrobacter, Rhodopseudomonas*, *Afipia*, *Oligotropha, Tardiphaga, Bosea* and *Variibacter*), Methylobacteriaceae (*Methylobacterium*) and Xanthobacteraceae (*Pseudolabrys* and *Labrys*), as well as *Pseudorhodoplanes*. We also included species from rhizobial genera in the Rhizobiaceae (*Rhizobium*, *Mesorhizobium* and *Sinorhizobium*). All the genome sequences and associated information were obtained from NCBI (1 February 2018).

The genome sequences of eight additional *Bradyrhizobium* isolates (*B. arachidis* LMG 26795^{T}, *Bradyrhizobium* spp. GHa, GHvi, R5, RC2d, RC3b, Leo170 and Leo121) were determined in this study. All isolates were obtained from the South African Rhizobium Culture Collection (SARCC; Agricultural Research Council, Plant Health and Protection Institute, Roodeplaat). The only exception was *B. arachidis* LMG 26795^{T} that was obtained from the Belgian Coordinated Collections of Microorganisms (University of Gent, Belgium). For each isolate, high quality genomic DNA was extracted from 4 to 7-day old cultures (grown at 28 °C on Yeast Mannitol Agar) using hexadecyltrimethylammonium bromide (CTAB) and following the protocol of Cleenwerck et al. [21].

Whole-genome sequencing for *B. arachidis* LMG 26795^{T}, as well as strains GHa, GHvi, R5, RC2d and RC3b, was performed by JGI as part of Phase III (KMG-III) of the Genomic Encyclopedia of Bacteria and Archaea (GEBA) project [118]. For this purpose, 300-base pair (bp) insert libraries

were constructed and sequenced using the Illuminia HiSeq 2500-1TB sequencing platform [8]. Raw sequences were trimmed and filtered using BBTools (Bestus Bioinformaticus Tools; http://sourceforge.net/projects/bbmap) [16] to remove low quality regions, Illumina artifacts and PhiX sequences. The filtered sequences were then assembled with Velvet version 1.2.07 [125]. Using the draft assemblies generated from Velvet, simulated pair-end reads were created using wgsim (https://github.com/Ih3/wgsim.). Both the filtered sequences and the simulated pair-end reads were then used to construct the final assembly with Allpaths-LG version r46652 [40].

Whole-genome sequencing for *Bradyrhizobium* strains Leo170 and Leo121 was performed at the Central Analytical Facilities (CAF; Stellenbosch University, South Africa) using the Ion PI™ Chip on the Ion Torrent™ PGM (Thermo Fisher Scientific) platform with 200-bp chemistry. The single-ended reads produced were quality-filtered using the FASTX Toolkits (i.e., reads with an 85% coverage cut-off and a quality score below 20 were removed) [79]. The filtered sequences were assembled with Newbler v 2.9 [68], using the infrastructure of the Bioinformatics and Computational Biology Unit at the University of Pretoria.

2.2 Dataset Construction

Four datasets were constructed in this study. The first dataset contained all of the shared protein-coding gene sequences for *Bradyrhizobium*, as well as members of the Bradyrhizobiaceae, Methylobacteriaceae, Xanthobacteraceae and Rhizobiaceae (genome information provided in Supplementary Table S1). The second dataset contained gene sequences for *Bradyrhizobium* and members of the Bradyrhizobiaceae only (Supplementary Table S1). The third and fourth datasets included amino acid sequences of the six genes routinely used for MLSA in *Bradyrhizobium* (i.e., *atpD, dnaK, glnII, gyrB, recA* and *rpoB*), with their respective taxon compositions corresponding to those of the two genome-based datasets mentioned above.

For the two genome-based datasets, the set of single-copy orthologous genes shared among all the genomes examined were identified using the EDGAR (Efficient Database framework for comparative Genome Analyses using BLAST score Ratios) server (https://edgar.computational.bio.uni-giessen.de) [12, 124]. The two conventional MLSA datasets were built using EDGAR's ortholog retrieval function to extract sequences for the genes *atpD, dnaK, glnII, gyrB, recA* and *rpoB*. For the two genome-based datasets, we also investigated the

genomic distribution of individual genes by using the genome of *B. japonicum* USDA 6^T as a reference in Geneious v7 [54]. Additionally, the functional roles of each of the putative gene products included in the two genome-based datasets were determined using the Kyoto Encyclopedia of Genes and Genomes (KEGG) database by making use of the BlastKoala mapping tool [52]; www.kegg.jp/blastkoala/; accessed 13 June 2018) and the predicted gene sequences from *B. japonicum* USDA 6^T.

2.3 Phylogenetic Analysis

Multiple sequence alignment was achieved by first splitting the respective fasta files obtained from EDGAR into individual gene sequence files. The individual sequence matrices were then aligned using MUSCLE (Multiple Sequence Comparison by Log-Expectation [30] implemented in CLC Main Workbench 7.6 (CLC Bio; Aarhus, Denmark). In each alignment, missing or poorly aligned data were removed with GBLOCKS 0.91b [17]. All the aligned sequence sets were then concatenated and partitioned with the appropriate amino acid model of evolution, as indicated by ProtTest 3.4 [1] and implemented in FASconCAT-G v1.02 [56]. Finally, all four of the datasets constructed in this way were subjected to maximum likelihood (ML) analysis using RAxML 8.2.1 [102]. Branch support was also estimated with RAxML by performing non-parametric bootstrap analyses of 1000 repetitions.

2.4 Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI)

All genome sequences investigated in this study were subjected to Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI) analyses using the EDGAR server [12, 124]. These analyses involved pairwise comparisons across shared genomic regions and protein sequences, respectively, to obtain similarity values between taxon pairs [41, 91].

2.5. Distribution of Predicted Lifestyle Traits

The genes commonly involved in the nodulation, nitrogen fixation and photosynthetic processes were identified and selected from literature [50, 77]. Each of the individual genes was identified and retrieved from all genomes using the ortholog retrieval function in EDGAR, as previously mentioned. In case of the *nod* and *nif* genes, the genome of *B. japonicum* USDA 6^T was used as reference, while that of *Bradyrhizobium* sp. S23321 was used as reference for the identification of

the genes involved in photosynthesis. All of the genes identified was then subjected to Blast2GO v3.3.0 analysis to ensure that they harboured the expected functional domains [22].

3. Results

3.1. Whole-genome Sequences

Collectively, the data currently available for *Bradyrhizobium* species indicate that the genome sizes for these bacteria vary substantially (Supplementary Table S1). When known species were considered, their genome sizes ranged from 7.8 million bases (Mb) for *B. yuanmingense* CCBAU 35157 [108] to 10.1 Mb for *B. centrolobii* BR 10245^T [70]. Inclusion of other *Bradyrhizobium* isolates, as well as those obtained from soil diversity studies indicated that the genomes sizes for this genus can be as small as 6.8 Mb for *Bradyrhizobium* isolate URHD0069 (see Proposal ID: 653 in the JGI Genome portal; https://genome.jgi.doe.gov/portal/) and as large as 11.7 Mb for *Bradyrhizobium* isolate GAS478 (Proposal ID: 2044 in the JGI Genome portal). Because of the range of technologies used for sequencing followed by manual finishing with various methods, the *Bradyrhizobium* genomes further differed greatly in terms of their quality and estimated level of completeness. For example, those that were initially released and were reported as mainly employing Sanger-based shotgun sequencing were typically complete and assembled into whole replicons, e.g., *Bradyrhizobium* isolates BTAi1 and ORS 278 [38]. However, these approaches were extremely costly and time-consuming. Assembly to replicon level was generally also possible for genome sequencing projects that included third-generation sequencing approaches such as the Pacific Biosciences (PacBio) sequencing technology, e.g., *B. icense* LMTR13^T [78] and *Bradyrhizobium* isolate CCGE_LA001 [98]. Those genomes sequenced using only second-generation technologies (e.g., 454 and Illumina) were typically not complete and consisted of numerous contigs due to the very short-insert libraries used, which are known to be limited in their ability to span repetitive regions. On average these assemblies contained around 200 contigs, but could include over 1000 contigs, e.g., *B. elkanii* isolates 587 and TnphoA33 [26, 66].

The same general trends were observed for the draft genomes of the eight isolates sequenced in this study. Their genome sizes ranged from 7.9 Mb for isolate GHvi to 9.7 Mb for *B. arachidis* LMG 26795^T (Supplementary Table S2). The six genomes sequenced using the Illumina HiSeq platform had sequencing depths ranging from 59x to 121x and were assembled into 44-128 contigs. The two

isolates (Leo170 and Leo121) sequenced using Ion Torrent respectively had sequencing depths of 76x and 104x, and their assemblies consisted of 727 and 1114 contigs. Furthermore, the G+C content for these eight isolates was 62-64%, which is comparable to those of other *Bradyrhizobium* isolates and species (Supplementary Table S1). All eight of these draft genome assemblies have been deposited in the NCBI database (see Supplementary Table S2 for accession numbers).

Genome sequences were available for most of the genera closely related to *Bradyrhizobium*. These were available on EDGAR in the public database and were completely assembled (Supplementary Table S1). The remainder of the outgroup taxa were available on the NCBI database as draft genomes and selected based on, either being the only representative genome of that taxon or, in the case where multiple genomes were available, the type species of outgroup taxa closely related to *Bradyrhizobium*. The genomes for these outgroups were assembled into 3-140 contigs. On average the genome sizes of the outgroups were smaller than those of *Bradyrhizobium*, with the smallest being *Oligotropha*, averaging 3.6 Mb, and *Methylobacterium* ranging between 6-8 Mb.

3.2 Dataset Construction

To avoid accidental inclusion of phylogenetic noise (e.g., paralogy, and inconsistencies regarding sequencing and gene annotation) into our datasets, we used EDGAR's strict orthology estimation [12] to identify the gene sequences used in this study. This was particularly important given the fact that the 204 genomes examined here differed considerably regarding their completeness estimates and the annotation approaches utilized (Supplementary Table S1). During the orthology estimation process, the bit score ratio value cutoff was 33% and BLAST hits with a mean Expect(E)-value of 1e-05 were regarded as orthologs. Due to this conservative approach, the number of sequences included in the two genome-based datasets depended strongly on the number of taxa investigated and the estimated completeness level of the genomes investigated. Also, a set of four genomes were not utilized in this study because their genomes were incomplete (see Supplementary Table S1). The final dataset that contained only *Bradyrhizobium* and Bradyrhizobiaceae sequences consisted of 180 taxa and 400 genes (82 878 aligned amino acid residues). The dataset containing *Bradyrhizobium* sequences together with those from Bradyrhizobiaceae, Methylobacteriaceae, Xanthobacteraceae, and Rhizobiaceae consisted of 200 taxa and included 128 genes (56 454 aligned amino acid residues). The corresponding MLSA datasets respectively consisted of 4 131 and 4 054 aligned amino acid residues.

Analysis with Geneious indicated that all the genes included in both of the genome-based datasets mapped to the chromosome of *B. japonicum* USDA 6^T where they were widely distributed and mostly not clustered in specific regions (Supplementary Figure S1). The KEGG analysis with BlastKOALA further indicated that the genes in both datasets were involved in diverse putative functions (Supplementary Table S3). Of those genes predicted to have known functions (i.e., 73% of the 400 genes were identified), most encoded for products involved in functional classes relating to Metabolism (e.g., Amino acid, Nucleotide, Energy, Carbohydrate, Cofactors and Vitamins), Genetic Informational Processes (DNA replication and repair, transcription, and translation), Environmental Information Processes (such as membrane transport) and Cellular Processes (e.g., Cell growth and death), with some genes involved in multiple functional classes.

3.3 Phylogenetic Analysis

ML analysis of the two genome-based datasets produced phylogenies with highly congruent topologies (Figure 1 and Supplementary Figure S2). Both recovered *Bradyrhizobium* as monophyletic and separated it into seven unique clades or lineages. Of these, three corresponded to the *B. japonicum*, *B. elkanii* and Photosynthetic supergroups. The remaining four, were designated according to the species or isolates they contain, i.e., *B. jicamae* supergroup, Kakadu supergroup, Soil 1 supergroup and the Soil 2 supergroup (see below). The Soil 1 supergroup was represented by a single taxon in our phylogenies, while the remaining six clades all had 100% bootstrap support in both trees. The only exception was the *B. elkanii* supergroup that received 99% bootstrap support in the 128-gene, 200-taxon phylogeny that included sequences of *Bradyrhizobium*, Bradyrhizobiaceae and other bacterial families (Figure 1).

Within the 400-gene, 180-taxon phylogeny, the Photosynthetic and Kakadu supergroups were sister taxa, with the *B. japonicum* supergroup representing their closest neighbour (see the inset on Figure 1 and Supplementary Figure S2). The *B. elkanii* and *B. jicamae* supergroups were sister taxa, with the Soil 1 supergroup being their closest neighbour. Overall, these two large assemblages grouped sister to each other with the Soil 2 supergroup at their base. All of these relationships received full or nearly full statistical support (i.e., all the deeper nodes in the tree had bootstrap values of 100%, except for the *B. elkanii* supergroup + *B. jicamae* supergroup clade that received 98%). The 128-gene, 200-taxon phylogeny generally supported the same relationships but lacked support for similar relationships among the *B. elkanii*, *B. jicamae* and Soil 1 supergroups (Figure 1).

a)
B. japonicum supergroup
Kakadu supergroup
Photosynthetic supergroup
Soil 1 supergroup
B. jicamae supergroup
B. elkanii supergroup
Soil 2 supergroup
Outgroup
KEY
Bradyrhizobium japonicum supergroup
Kakadu supergroup
Photosynthetic supergroup
Bradyrhizobium elkanii supergroup
Bradyrhizobium jicamae supergroup
Soil 1 supergroup
Soil 2 supergroup
Outgroup taxa
Root nodule
Soil
Clinical
Endophyte
Other
b)
Bradyrhizobium sp. Leo121
B. elkanii WSM2783
Bradyrhizobium sp. Leo170
Bradyrhizobium sp. Ai1a-2
Bradyrhizobium sp. LTSP885
Bradyrhizobium sp. LTSPM299
B. macuxiense BR 10303T
Bradyrhizobium sp. th.b2
B. elkanii UASWS1015
Bradyrhizobium sp. OHSUIII
Bradyrhizobium sp. SK17
Bradyrhizobium sp. DFCI-1
Bradyrhizobium sp. PARBB1
B. viridifuturi SEMIA 690T
Bradyrhizobium sp. UFLA03-84
B. tropiciagri SEMIA 6148T
Bradyrhizobium sp. NAS96.2
B. mercantei SEMIA 6399T
B. embrapense SEMIA 6208T
B. erythrophlei MT12
Bradyrhizobium sp. C9
B. pachyrhizi BR 3262
B. brasilense UFLA03-321T
Bradyrhizobium sp. R5
B. pachyrhizi PAC48T
B. elkanii USDA 3254
B. elkanii USDA 3259
B. elkanii USDA 94
B. elkanii CCBAU 05737
B. elkanii BLY6-1
B. elkanii BLY3-8
B. elkanii CCBAU 43297
B. elkanii USDA 76T
B. elkanii 587
B. elkanii TnphoA33
B. erythrophlei GAS478
B. ottawaense GAS524
B. lablabi GAS522
B. lablabi MT34
Bradyrhizobium sp. URHA0002
B. jicamae PAC68T
B. paxllaeri LMTR21T
B. lablabi CCBAU 23086T
B. valentinum LmjM3T
B. icense LMTR13T
Bradyrhizobium sp. LMTR3
B. retamae Ro19T
B. elkanii WSM1741
B. lablabi GAS165
B. canariense GAS369
B. erythrophlei GAS138
Bradyrhizobium sp. URHD0069
B. erythrophlei GAS242
B. lablabi GAS499
Nitrobacter hamburgensis X14T
Nitrobacter winogradskyi Nb-255T
Nitrobacter vulgaris Ab1
Bradyrhizobium sp. NFR13
Tardiphaga robiniae Vaf-07T
Rhodopseudomonas palustris BisB18
Rhodopseudomonas palustris BisA53
Rhodopseudomonas palustris HaA2
Rhodopseudomonas palustris BisB5
Rhodopseudomonas palustris DX.1
Rhodopseudomonas palustris TIE.1
Rhodopseudomonas palustris CGA009T
Oligotropha carboxidovorans OM5
Oligotropha carboxidovorans OM5 ATCC 49405
Oligotropha carboxidovorans OM4
Afipia broomeae ATCC 49717T
Afipia massiliensis LC387
Bradyrhizobium sp. CCH10-C7
Afipia birgiae 34632T
Pseudorhodoplanes sinuspersici RIPI110T
Pseudolabrys sp. Root1462
Variibacter gotjawalensis GJW 30T
Labrys sp. WJW
Methylobacterium zatmanii PSBB041
Methylobacterium phyllosphaerae CBMB27
Methylobacterium nodulans ORS2060T
Bosea lupini LMG 26383T
Bosea robiniae DSM 26672
Bosea thiooxidans DSM 9653
Bradyrhizobium sp. CCH1-B1
Bradyrhizobium sp. CCH5-A9
Mesorhizobium loti MAFF303099
Mesorhizobium opportunistum WSM2075T
Mesorhizobium australicum WSM2073T
Sinorhizobium medicae WSM419
Sinorhizobium meliloti 1021
Sinorhizobium americanum CCGM7
B. lupini HPC(L)
Rhizobium tropici CIAT899
Rhizobium etli CFN42T
Rhizobium leguminosarum bv. trifoli WSM2304
B. canariense UBMA192
B. canariense UBMA182
B. canariense UBMA181
B. canariense UBMA183
B. canariense UBMA195
B. canariense UBMA171
B. canariense UBMAN05
B. canariense UBMA510
B. canariense UBMA122
B. canariense UBMA050
B. canariense UBMA051
B. canariense UBMA060
B. canariense UBMA061
B. canariense UBMA052
Bradyrhizobium sp. WSM1253
Bradyrhizobium sp. WSM1417
Bradyrhizobium sp. WSM471
Bradyrhizobium sp. BF49
Bradyrhizobium sp. URHA0013
Bradyrhizobium sp. LTSP849
Bradyrhizobium sp. LTSP857
Bradyrhizobium sp. OK095
Bradyrhizobium sp. YR681
Bradyrhizobium sp. S23321
B. japonicum CCBAU 15517
B. japonicum CCBAU 15354
B. japonicum USDA 123
B. japonicum UBMA197
B. japonicum J5
Bradyrhizobium sp. G22
B. japonicum USDA 38
B. japonicum Is-34
B. japonicum CCBAU 25435
B. japonicum SEMIA 5079
B. japonicum CCBAU 15618
B. japonicum E109
B. japonicum USDA 6T
B. japonicum FN1
Bradyrhizobium sp. WSM3983
Bradyrhizobium sp. GHa
B. manausense BR 3351T
Bradyrhizobium sp. GHvi
B. japonicum USDA124
B. sacchari BR10280T
B. stylosanthis BR 446T
B. arachidis LMG 26795T
Bradyrhizobium genosp. SA-4 CB756
Bradyrhizobium sp. USDA 3384
B. japonicum USDA 4
Bradyrhizobium sp. CCBAU 15615
B. ottawaense L2
B. ottawaense OO99T
Bradyrhizobium sp. CCBAU 15635
Bradyrhizobium sp. CCBAU 15544
B. shewense ERR11T
Bradyrhizobium sp. RC3b
Bradyrhizobium sp. WSM2793
Bradyrhizobium sp. 85S1MB
Bradyrhizobium sp. 39S1MB
Bradyrhizobium sp. AT1
Bradyrhizobium sp. Y36
B. nitroreducens TSA1
B. japonicum 22
Bradyrhizobium sp. DOA1
Bradyrhizobium sp. CCGE-LA001
Bradyrhizobium sp. RC2d
B. japonicum USDA 135
B. liaoningense CCNWSX0360
Bradyrhizobium sp. CCBAU 43298
B. forestalis INPA54BT
Bradyrhizobium sp. WSM1743
B. japonicum is5
B. japonicum in8p8
Bradyrhizobium sp. Leaf396
Bradyrhizobium sp. CCH5-F6
Bradyrhizobium sp. DOA9
B. yuanmingense CCBAU 05623
B. yuanmingense BR 3267
B. yuanmingense CCBAU 35157
B. yuanmingense CCBAU 10071T
B. yuanmingense CCBAU 25021
Bradyrhizobium sp. cf659
Bradyrhizobium sp. WSM2254
Bradyrhizobium sp. CCBAU 41267
B. diazoefficiens USDA 110T
B. diazoefficiens Y21
B. diazoefficiens NK6
B. diazoefficiens SEMIA 5080
B. japonicum Is-1
B. diazoefficiens USDA 122
Bradyrhizobium sp. NAS80.1
Bradyrhizobium sp. AS23.2
B. centrolobii BR 10245T
Bradyrhizobium sp. cp5.3
B. neotropicale BR10247T
Bradyrhizobium sp. Ec3-3
Bradyrhizobium sp. ARR65
Bradyrhizobium sp. Tv2a-2
Bradyrhizobium sp. STM3843
B. oligotrophicum S58T
Bradyrhizobium sp. BTAi1
Bradyrhizobium sp. ORS285
Bradyrhizobium sp. ORS375
Bradyrhizobium sp. ORS278
Bradyrhizobium sp. STM3809
0.1

Fig. 1. Phylogenies inferred from protein-coding genes for the genus *Bradyrhizobium* and its close relatives. (a) Summary tree inferred from the maximum likelihood phylogeny generated from the 400-gene, 180-taxon dataset (Supplementary Fig. S2). Node support values of 100% are indicated by thicker branches. (b) ML phylogeny of the 128-gene, 200-taxon dataset that included sequences from *Bradyrhizobium*and other Bradyrhizobiaceae, as well as Methylobacteriaceae, Xanthobacteraceae and Rhizobiaceae. Strains from Rhizobiaceae (*Sinorhizobium*, *Mesorhizobium* and *Rhizobium*) were used as an outgroup. Tree branches and isolate source are colour coded and indicated according to the key provided. Most branches received 100% bootstrap support and only those with support values lower than 100% are indicated. The scale bar indicates the number of amino acid changes per site.

The *B. japonicum*, *B. elkanii*, *B. jicamae* and Photosynthetic supergroups all contained the type strains of validly published *Bradyrhizobium* species together with known isolates whose taxonomy still requires investigation (Table 1). All the isolates included in the three remaining supergroups are yet to be described. The Kakadu supergroup were represented by *Bradyrhizobium* isolates Tv2a-2 and ARR65 from the root nodules of legumes growing in Australia and the Barro Colorado Island of Panama [90, 109]. Previous research has shown that ARR65 is likely conspecific with a number of other isolates originating from diverse legumes occurring in Australia's Kakadu National Park [106], hence our designation for this supergroup. The Soil 1 supergroup was represented by isolate GAS165 that is part of a genomic diversity study of forest soil (see Proposal ID: 2044 in the JGI Genome portal; https://genome.jgi.doe.gov/portal/). Isolates from this diversity study was also included in the Soil 2 supergroup, which contained isolates from a similar study of Mediterranean grassland soil (see Proposal ID: 653 in the JGI Genome portal).

Based on our analyses, most of the *Bradyrhizobium* supergroups included numerous instances of where incorrect names have been applied to particular isolates. For example, the type strain of *B. lablabi* (CCBAU 233086^T) is a member of the *B. jicamae* supergroup, yet other isolates designated as conspecifics also occur in the Soil 1 and Soil 2 supergroups (Figure 1). Also, some isolates bearing this species name occur in the *B. jicamae* supergroup where they appear to be only distantly related to the type strain of *B. lablabi*. Various other instances where revision of an isolate's species name is required are indicated in Supplementary Table S1. In addition, a number of the isolates for which whole genome sequences are available, are indicated in the public domain databases as members of *Bradyrhizobium*, although our data show that they belong to other genera (i.e., isolates CCH1-B1 and CCH5-A9 belong to *Bosea*, CCH10-C7 belongs to *Afipia*, NFR13 belongs to *Tardiphaga*, and

HPC(L) does not belong to *Bradyrhizobium*; See Supplementary Table S1 for NCBI accession numbers).

Taxon selection in the two conventional MLSA datasets corresponded to those of the 128-gene, 200-taxon dataset and the 400-gene, 180-taxon dataset, but included the sequences for only six protein-coding genes. ML analyses of the MLSA datasets generated phylogenies with similar topologies, which both recovered *Bradyrhizobium* as monophyletic (Supplementary Figures S3 and S4). However, neither were fully congruent with those of the two genome-based ML trees. The most prominent discordances were in terms of the *B. elkanii*, *B. jicamae* and Soil 2 supergroups that were not recovered from the two small datasets. The only relationships or clusters that remained consistent among the various phylogenies were the *B. japonicum*, Photosynthetic and Kakadu supergroups, and the relationships among them. The MLSA-based relationships within these supergroups were also similar to those observed in the genome-based phylogenies.

In the two MLSA trees, the *B. elkanii* supergroup was distributed across two lineages where one contained only four root-nodule isolates (Leo121, Leo170 and WSM2783, respectively obtained from nodules of *Leobordea divaricate*, *L. lanceolata* and *L. carinata* and Ai1a-2 obtained from *Andira inermis*). The *B. jicamae* supergroup was mostly also distributed across multiple lineages in the MLSA trees, with one containing mainly type strains, another containing only soil associated isolates, while *B. valentinum* LmjM3^{T} and *B. erythrophlei* GAS478 grouped separately as singletons. The same was true for the Soil 2 supergroup, where one of its three MLSA-based lineages included the Soil 1 supergroup (Supplementary Figure S3 and S4).

3.4 Average Nucleotide Identity (ANI) and Average Amino Acid Identity (AAI)

ANI and AAI analysis supported separation of *Bradyrhizobium* into the seven identified main clades or lineages (Figure 2, Supplementary Table S4 and S5). Within each, interspecies ANI values for the *B. japonicum* and *B. elkanii* supergroups were above 82% and 79% respectively, while those for the photosynthetic and Kakadu, *B. jicamae* and Soil 2 supergroups were slightly lower, but always above 79%. For each supergroup, the interspecies AAI values were generally higher than the ANI values and, in all cases, exceeded 80%. Members of the same species had ANI and AAI values >96 and >97%, respectively (Figure 2, Supplementary Table S4 and S5). The ANI and AAI values for all the taxa included in *Bradyrhizobium* were both >75%. Between the different supergroups, ANI

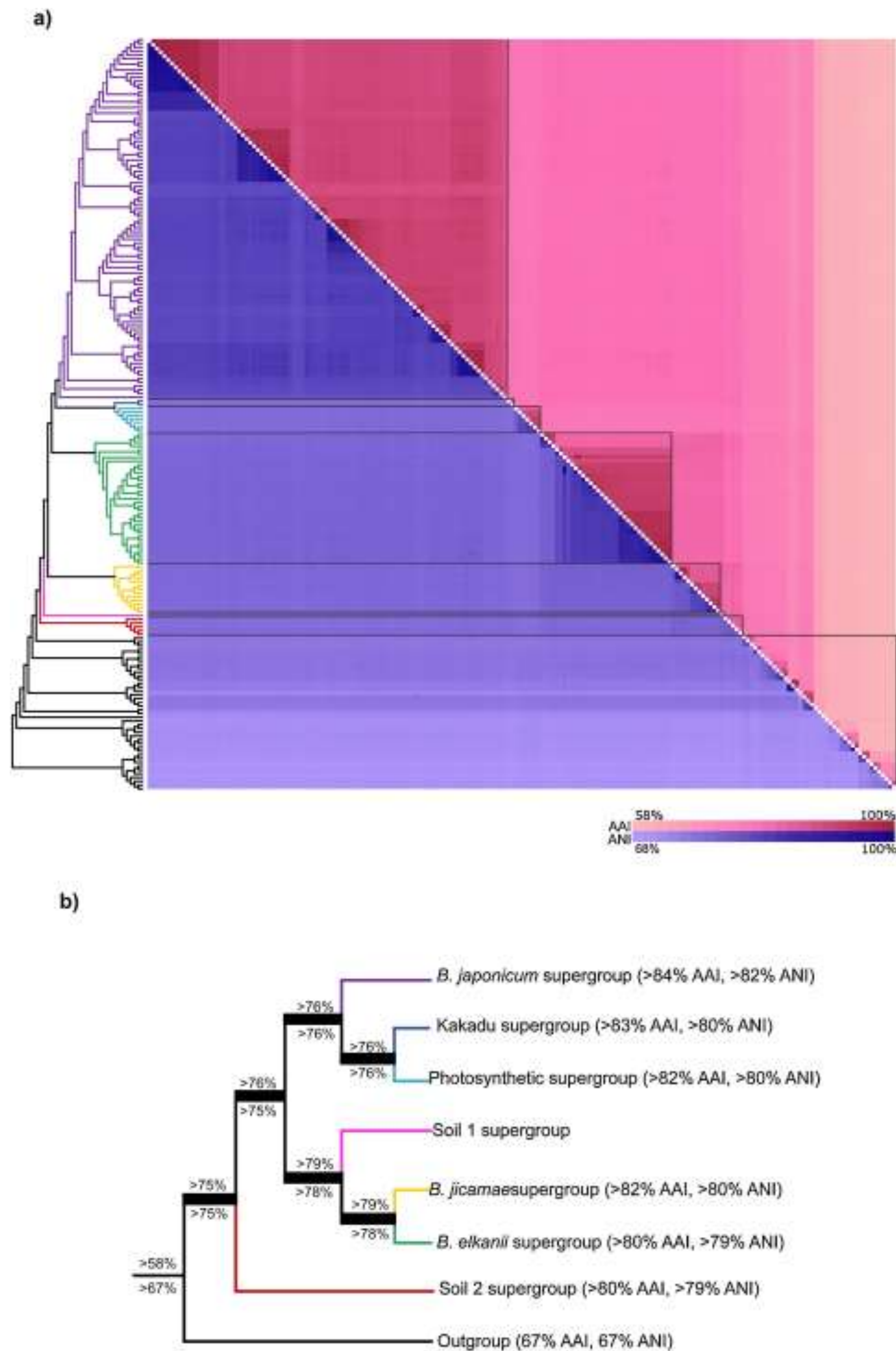

Fig. 2. Average Nucleotide Identity (ANI) and Amino Acid Identity (AAI) between members of *Bradyrhizobium* and its close relatives. (a) A heatmap illustrating similarity values relative to intra-generic relationships inferred from the maximum likelihood topology (Fig. 1). The ANI values are indicated in the lower triangle and the AAI values are indicated in the upper triangle of the heatmap. For specific values see Supplementary Table S4 and S5. (b) Summary tree of the 400-gene, 180 phylogeny (Supplementary Fig. S2) showing AAI and ANI values for the intra-generic relationships (AAI values are indicated above branches and ANI values below branches).

values were 76-79% and AAI values were 75-79%. Between the different genera of the Bradyrhizobiaceae (*Afipia*, *Bradyrhizobium, Nitrobacter*, *Rhodopseudomonas*, *Oligotropha* and *Tardiphaga*) ANI values ranged from 73% to 76% and AAI values ranged from 70% to 75%.

3.5 Distribution of Predicted Lifestyle Traits

In terms of source or origin, the *B. japonicum*, *B. elkanii*, *B. jicamae* and Kakadu supergroups all included root/stem nodule isolates and free-living soil bacteria. The *B. japonicum* and *B. elkanii* supergroups also included isolates originating from the clinical environment and strains that have been designated as plant endophytes (Figure 3). While the Soil 1 and Soil 2 supergroups included soil bacteria only, the Kakadu and Photosynthetic supergroups contained root/stem nodule bacteria only (Figure 3).

To obtain a more detailed picture of the bacterial lifestyles associated with *Bradyrhizobium*, the presence of genes commonly involved in the nodulation, nitrogen fixation and photosynthetic processes were investigated across all taxa included in the study. A set of 28 genes were used for this purpose and included *nodABCDIJ* (nodulation), *nifABDEHK* (nitrogen fixation), *bchCGPXYZ* (bacteriochlorophyll), *crtBCDEFI* (carotenoids) and *pufBALM* (light harvesting complex) [50, 77]. Although all three sets of genes only occurred in some members of *Bradyrhizobium*, they were completely absent from certain individuals in the *B. japonicum* and *B. jicamae* supergroups, as well as all members of the Soil 1 and Soil 2 supergroups. There were several instances where single *nod*, *nif* or photosynthetic genes were missing from certain taxa (Figure 3 and Supplementary Table S6). Most of these were likely due to the lower completeness level of the genomes included and might be identified when the quality of genome sequences of these bacteria is improved. Among the *Bradyrhizobium* sequences, single *nod*, *nif* and photosynthetic genes, not accompanied by other genes involved in these processes, were also detected in a few instances. Analysis with Blast2GO showed that these all contained the expected functional motifs and similarities to the reference sequences (Supplementary Table S7).

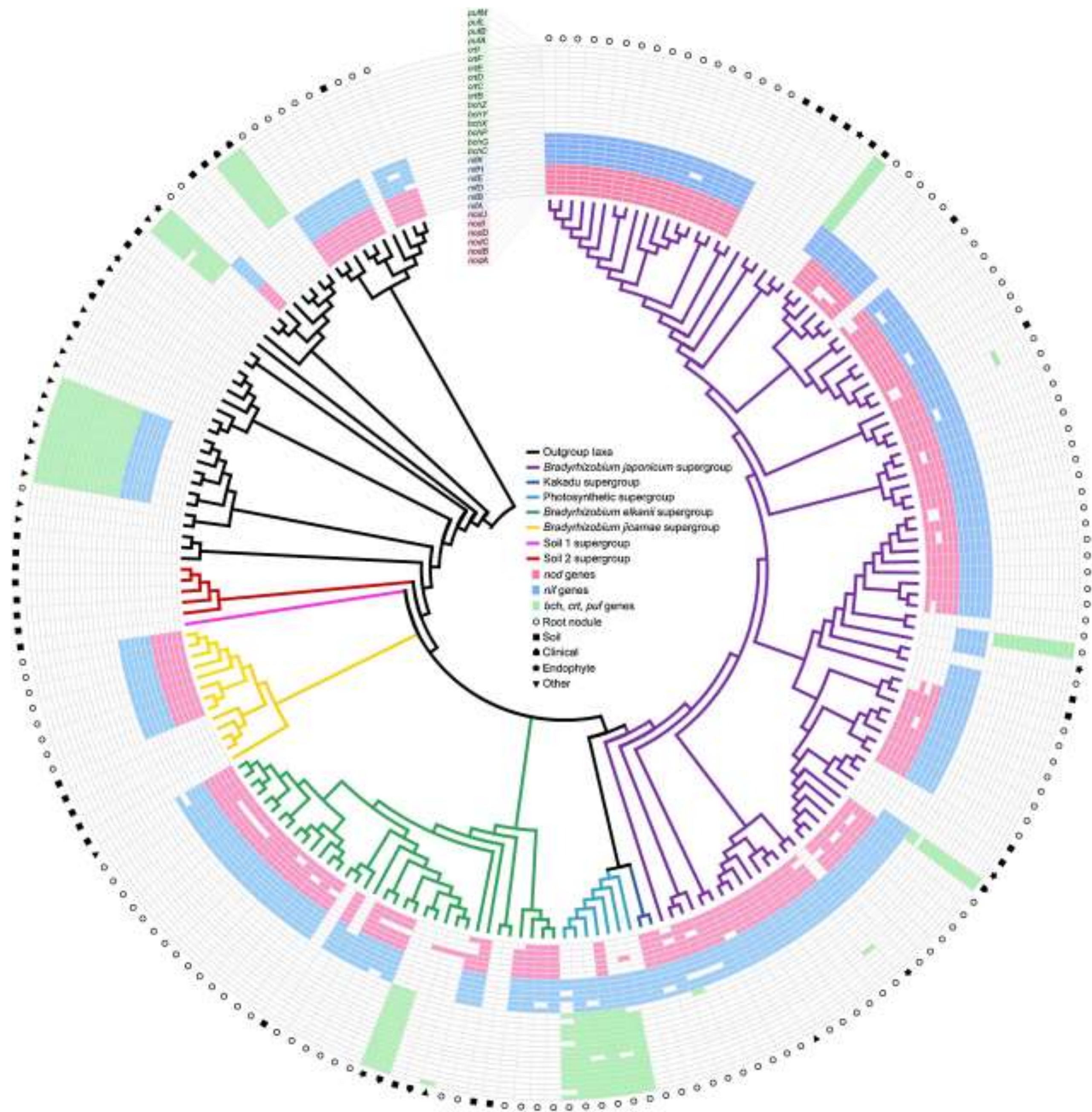

Fig. 3. Cladogram generated from the 128-gene, 200-taxon maximum likelihood topology (Fig. 1) depicting the absence and presence of nodulation (*nod*), nitrogen fixation (*nif*) and photosynthetic (*bch*, *crt* and *puf*) genes in each of the strains. Absence and presence of genes are indicated by a coloured and uncoloured box, respectively. The colour of tree branches, presence and absence of genes and strain source are indicated according to the key provided.

Based on our analyses, complete or near-complete sets of *nod* and *nif* genes were distributed throughout the different lineages of *Bradyrhizobium*, as well as in *Methylobacterium* and members of the Rhizobiaceae (Figure 3 and Supplementary Table S6). All members from the Soil 1 and Soil 2 supergroups lacked both the *nod* and *nif* gene sets, and these genes were also absent from some

individuals of the *B. japonicum*, *B. elkanii* and *B. jicamae* supergroups. Furthermore, *nif* genes were present in all isolates with *nod* genes, although a number of isolates possessed *nif* genes in the absence of any *nod* genes. These included most members of the Photosynthetic supergroup, some individuals of the *B. japonicum* and *B. elkanii* supergroups and, among the outgroup taxa, all *Rhodopseudomonas* isolates.

As expected, complete or near-complete sets of the *puf*, *crt* and *bch* genes were found in all but the known non-photosynthetic isolate STM3843 of the Photosynthetic supergroup [72]. They were also found in isolates of *Rhodopseudomonas*, *Bosea* and *Methylobacterium* (Figure 3 and Supplementary Table S6). However, within *Bradyrhizobium*, all three sets of genes were present in two members of the *B. elkanii* supergroup, i.e., isolate DFCI-1 originating from a human tissue sample [11] and isolate PARBB1 originating from a study of bacteria associated with the freshwater diatom *Asterionella formosa* (see NCBI accession NKIX00000000). The photosynthetic gene sets were further also present in three members of the *B. japonicum* supergroup, i.e., S23321 originating from soil and known to contain these genes [77], isolate 39S1MB originating from *Glycine max* root nodules [14], and isolate CCH5-F6 that originates from a hospital biofilm (see NCBI accession LSIC01000099). Among the taxa examined, the photosynthetic genes were only present in isolates lacking *nod* genes, with the only exception being the Photosynthetic supergroup isolate ORS285, which originates from *Aeschynomene* stem nodules [71].

4. Discussion

Numerous genomes are currently available in the public domain for a diverse set of *Bradyrhizobium* species. Although the completeness of the available genome sequences varies substantially due to the variety of sequencing technologies employed and the level of sequencing efforts, our data indicates that most are adequate and useful for taxonomic studies. This is particularly true for those seeking to determine genome sequence similarity metrics such as ANI or to infer genome-based phylogenies. As little as 20% of the genome sequence for two individuals are needed for meaningful ANI comparisons [5, 91], while robust phylogenies require the sequences of as few as 31 randomly selected protein-coding marker genes [20]. For example, the eight *Bradyrhizobium* genomes determined in this study were sequenced using either the Ion Torrent or Illumina HiSeq platforms and accordingly produced genomes of substantially different completeness estimates, but all of them

met the minimal standards for their use in bacterial taxonomy [20]. Our results were also consistent with those of a previous study, which used analyses with lineage-specific marker sets to demonstrate the completeness and contaminant-free nature of bacterial genomes [81, 82]. Therefore, most of the *Bradyrhizobium* genome sequences in the public domain are of adequate or high quality and this number will likely increase due to large scale genome sequencing projects such as GEBA (Genomic Encyclopedia of Bacteria and Archaea) [58, 59, 74, 121] and wide access to reliable and affordable second- and third generation sequencing technologies [62].

By making use of publicly available genomic resources, this study generated the first resolved phylogeny for *Bradyrhizobium*. This was done using two datasets, one containing the sequences for 128 genes and 200 taxa, and the other 400 genes and 180 taxa. Despite the overall concordance between these phylogenies, the use of more genes (i.e. 400 genes) increased the statistical support in the backbone of the phylogeny to unequivocally resolve the relationships among all clades. This is not unusual when large datasets, comprising of informative sequence data, shared among all the taxa of interest, are brought into account for phylogenetic inference [19, 23]. As shown in previous studies [19, 61, 79], the larger 400 gene dataset, unlike the 128-gene dataset, contained sufficient phylogenetic signal for "overshadowing" the non-phylogenetic signal inherent to the sequence data, thus allowing more reliable inference of the evolutionary history of *Bradyrhizobium*.

The *Bradyrhizobium* phylogeny includes several deep divergences that separate the genus into six distinct clades and one, as yet, single taxon lineage. Three of these were previously known (i.e., the *B. japonicum*, *B. elkanii* and Photosynthetic supergroups) [63, 104, 114], while the existence of other such groups was only hinted at in the past (e.g., [63]). The current study is therefore the first to unambiguously resolve these other groups, which we named the *B. jicamae*, Kakadu, Soil 1 and Soil 2 supergroups. Known symbiotic isolates occur within all, bar the Soil 1 and Soil 2 supergroups. Our results further suggest that the latter supergroups and a subset of isolates from the *B. jicamae* supergroup occupy environments that are functionally distinct from those typically associated with *Bradyrhizobium* [28, 63, 73]. These isolates mostly originate from forest and Mediterranean grassland soil and have been shown to lack symbiotic and photosynthetic abilities [50, 112]. It is thus likely that this genus includes many more non-symbiotic or non-photosynthetic strains, which is consistent with the notion that *Bradyrhizobium* has a long and complex evolutionary history [10, 101, 106].

Conventional MLSA supports many of the groups recovered in our genome-based phylogeny for *Bradyrhizobium*, but not their relationships with one another. MLSA studies (those conducted here and previously) [10, 63, 93, 106] typically recover the *B. japonicum*, *B. elkanii* and Photosynthetic supergroups. In previous studies, the *B. jicamae* supergroup has been recovered as distinct but closely related to the *B. elkanii* supergroup [10, 63, 104, 106]. The Kakadu supergroup was first observed in a study of Australian *Bradyrhizobium* [106] and later in a South African study [10], while ours is the first study to report the Soil 1 and Soil 2 supergroups. Comparison with the genome-based phylogeny for *Bradyrhizobium* suggests that the major failing of MLSA pertains to the relationships among supergroups. In the current study, this was particularly evident for the *B. jicamae, B. elkanii* and Soil 2 supergroups that lacked support in MLSA trees. Little to no statistical support in MLSA trees was also previously reported for the *B. japonicum*, *B. elkanii* and Photosynthetic supergroups [10, 63, 106]. The primary factor that likely detracts from MLSA phylogenies is the high degree of conservation in housekeeping gene sequences used [39, 85], which causes these datasets to lack sufficient phylogenetic signal for overcoming stochastic error [35, 79].

Our results showed that the genes required for nodulation and nitrogen fixation are not uniformly distributed across the *Bradyrhizobium* phylogeny. Nodulation and nitrogen fixation are dependent on the presence of symbiotic islands found either on the chromosome of *Bradyrhizobium* isolates [25, 53] or, as recently reported, on a plasmid [76]. Many isolates of *Bradyrhizobium*, including those in the *B. jicamae*, Soil 1 and Soil 2 supergroups, completely lacked *nif* and *nod* genes. This has also been seen in recent studies where some individuals did not possess the machinery required for either or both of these processes [38, 50, 77, 112]. Such isolates are often described as having abandoned the symbiotic lifestyle [77, 94], which has led to suggestions that symbiotic nitrogen fixation may not be a dominant lifestyle of *Bradyrhizobium* or even rhizobia in general [112].

Our results further showed that the photosynthetic trait is not limited to the known photosynthetic isolates of *Bradyrhizobium*, but that it can also be found in other members of this genus. Among rhizobia, this trait is thought to be rare as it has only been reported in certain *Methylobacterium* and *Bradyrhizobium* isolates associated with the aquatic legume *Aeschynomene* [37, 42]. In these photosynthetic *Bradyrhizobium* isolates, the trait requires a 45-50 kilobase pair chromosomal gene cluster [37, 38, 72]. The only other member of this genus previously known to have this trait is the oligotrophic isolate S23321 that originates from paddy field soil [77]. In the current study, the set

of photosynthesis genes were also identified in isolates associated with *G. max* root nodules and a freshwater diatom, as well as two isolates from the clinical environment. Overall, such complete or near complete sets of photosynthesis genes were present in *B. japonicum*, *B. elkanii* and the Photosynthetic supergroups, while certain isolates of the *B. japonicum* and *B. elkanii* supergroups also harboured one or two of the individual photosynthesis genes. Although it was previously suggested that the photosynthetic trait in *Bradyrhizobium* evolved from a photosynthetic ancestor common to *Rhodopseudomonas* [37, 77], our data additionally point to the potential involvement of horizontal gene transfer (HGT). For example, the photosynthesis genes found in isolates from biofilms of hospital hose showers could have acquired them via HGT from other members of the biofilm community (e.g., see [33, 100]). Future work should therefore seek to understand the evolutionary processes responsible for the distribution patterns of photosynthesis genes in *Bradyrhizobium* and their role in the biology of this genus.

To the best of our knowledge our study is the first to investigate the presence of photosynthesis related genes in the genus *Bosea*, which is a close relative of *Bradyrhizobium*. Isolates of *Bosea* harbored complete sets of the photosynthesis genes but lacked *nod* and *nif* genes. Members of this genus are commonly isolated from hospital water supplies, agricultural soil or as endophytic root-nodule bacteria [24, 25, 60]. That they lack *nod* and *nif* genes is consistent with the fact that they have never been shown capable of nodulation and symbiotic nitrogen fixation [24, 95]. Although some isolates of *Bosea* have been reported to encode the *nifH* gene [25], this gene does not share homology with any of the genes investigated within this study (results not shown). The role of this gene and the photosynthesis related genes in *Bosea* thus remains unknown.

The taxonomic placement of the Photosynthetic supergroup within *Bradyrhizobium* has been dealt with cautiously [99, 115]. Despite, previous suggestions that the Photosynthetic supergroup should be described as a novel genus with the proposed name '*Photorhizobium'* [29, 120], our results confirms the monophyletic nature of *Bradyrhizobium* and the nested position of the Photosynthetic supergroup within this assemblage. The previous proposals were based on 16S rRNA gene phylogenies and photosynthesis as a distinguishing trait [99, 110, 119]. However, by employing genome data we showed that this trait is not limited to the Photosynthetic supergroup, and that its recognition as a separate genus would render the remainder of *Bradyrhizobium* paraphyletic. In fact, none of the accessory traits typically associated with a particular lifestyle allow resolution of the

supergroups of *Bradyrhizobium.* The same is true for whole genome-based metrics such as AAI and ANI. For example, ANI and AAI values for the genus *Bradyrhizobium* is in the proposed range for generic boundaries ranging between 70-95 % for ANI [55] and 60-85% for AAI [67]. Further, based on a study by Parks and colleagues [81], *Bradyrhizobium* is seen as a single unit. Taken together, these data thus indicate that *Bradyrhizobium*, as it is currently circumscribed, represents a single and distinct, albeit large and diverse, genus.

It is noteworthy that the presence of the photosynthetic genes found in *Bradyrhizobium* occurred predominantly in isolates lacking *nod* genes. This phenomenon is not widely studied, but from work done on isolates from the Photosynthetic supergroup, it appears to be related to the nodulation pathway employed by nodulating *Bradyrhizobium* isolates [13]. In these bacteria, nodulation may or may not require Nod-factors (i.e., lipochitooligosaccharides compounds whose production is dependent on *nod* genes) [38, 72]. In the Nod-factor independent system, photosynthesis presumably stimulates nitrogen fixation while at the same time saving energy for the plant [42]. Based on work using the endophytic, *nod*-gene lacking *Bradyrhizobium* S23321, photosynthesis is thought to be one of the metabolic adaptations allowing its survival in paddy soil [42]. The distribution of photosynthetic genes among extant taxa might also be linked to the evolution of *Bradyrhizobium*, which is thought to have involved transformation of phototrophic, non-symbiotic, free-living diazotrophs into nitrogen-fixing symbionts [72, 88]. Such phototrophic forms of *Bradyrhizobium* could have evolved from an ancestor common to *Bradyrhizobium* and its sister taxon *Rhodopseudomonas.* Maintenance of these symbionts in environments where light is limited would then allow loss of photosynthesis genes over time [88]. As mentioned above, alternative hypotheses involving HGT may also explain the patchy occurrence of photosynthesis genes across the *Bradyrhizobium* phylogeny. Although HGT also affects the *nod* and *nif* genes of *Bradyrhizobium*, *nod* and *nif* phylogenies are often congruent [10, 104]. This suggests that both HGT and vertical inheritance from parent to offspring influence evolution of the nodulation and nitrogen-fixation traits. More information is thus needed to resolve the ancestry of the genus.

In this study, the use of available genomic resources allowed us to address numerous instances where incorrect names were applied to isolates in public domain databases. For example, several strains clustered with genera outside of *Bradyrhizobium* (i.e., *Bradyrhizobium* sp. NFR that grouped with *Tardiphaga*, *Bradyrhizobium* sp. CCH10-C7 that grouped with *Afipia*, and *Bradyrhizobium* sp.

CCH1-B1 and CCH5-A9 grouping with *Bosea*). In addition, our data and a previous study by Agarwal and Purohit [2] showed that *B. lupini* HPC(L) is not conspecific to *R. lupini* USDA 3051^T which was recently moved to *Bradyrhizobium* [86]. Our results also suggested that incorrect species names were applied within each of the supergroups and that these require revision. These findings fully support those of a recent study where genus and species affiliations were determined using relative evolutionary divergence estimates [81]. The invaluable source of information that is captured and retrievable through the use of genome data in taxonomy, not only provide us the opportunity to construct a robust hierarchy based on evolutionary history, but also provide insight into the understanding and unravelling of complex traits associated with lifestyles. For instance, comparative genomics together with detailed ecological studies might shed light on the genetic background responsible for maintaining certain lifestyles within this genus.

Acknowledgements

We would like to acknowledge the Ion Torrent Sequencing Facility at the University of Pretoria and the Central Analytical Facilities (CAF) at Stellenbosch University for genome sequencing. The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Conflict of interest

The authors have no conflicts of interest to declare.

Funding information

We thank the South African National Research Foundation (NRF) and the Department of Science and Technology for the support via the Centre of Excellence (CoE), South African Research Chairs Initiative (SARChI) and Competitive Programme for Rated Researchers (CPRR) funding instruments. We also thank the NRF for student funding, Grant SFH150721128526.

5. References

[1] Abascal, F., Zardoya, R., Posada, D., (2005) ProtTest: selection of best-fit models of protein evolution, Bioinformatics 21, 2104-2105.

[2] Agarwal, L., Purohit, H.J., (2013) Genome sequence of *Rhizobium lupini* HPC (L) isolated from saline desert soil, Kutch (Gujarat), Genome Announc. 1, e00071-00012.

[3] Ahnia, H., Bourebaba, Y., Durán, D., Boulila, F., Palacios, J.M., Rey, L., Ruiz-Argüeso, T., Boulila, A., Imperial, J., (2018) Bradyrhizobium algeriense sp. nov., a novel species isolated from effective nodules of Retama sphaerocarpa from Northeastern Algeria, Syst. Appl. Microbiol.

[4] Alves, B.J., Boddey, R.M., Urquiaga, S., (2003) The success of BNF in soybean in Brazil, Plant Soil. 252, 1-9.

[5] Arahal, D.R., (2014) Whole-genome analyses: average nucleotide identity, in: Methods in Microbiology, Elsevier, pp. 103-122.

[6] Azani, N., Babineau, M., Bailey, C.D., Banks, H., Barbosa, A.R., Pinto, R.B., Boatwright, J.S., Borges, L.M., Brown, G.K., Bruneau, A., (2017) A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny The Legume Phylogeny Working Group (LPWG), Taxon 66, 44-77.

[7] Bennett, J.S., Jolley, K.A., Earle, S.G., Corton, C., Bentley, S.D., Parkhill, J., Maiden, M.C., (2012) A genomic approach to bacterial taxonomy: an examination and proposed reclassification of species within the genus *Neisseria*, Microbiology 158, 1570-1580.

[8] Bennett, S., (2004) Solexa Ltd Pharmocogenomics 5.

[9] Beukes, C.W., Palmer, M., Manyaka, P., Chan, W.Y., Avontuur, J.R., van Zyl, E., Huntemann, M., Clum, A., Pillay, M., Palaniappan, K., (2017) Genome data provides high support for generic boundaries in *Burkholderia* sensu lato, Front. Microbiol. 8, 1154.

[10] Beukes, C.W., Stępkowski, T., Venter, S.N., Cłapa, T., Phalane, F.L., le Roux, M.M., Steenkamp, E.T., (2016) Crotalarieae and Genisteae of the South African Great Escarpment are nodulated by novel *Bradyrhizobium* species with unique and diverse symbiotic loci, Mol. Phylogenet. Evol. 100, 206-218.

[11] Bhatt, A.S., Freeman, S.S., Herrera, A.F., Pedamallu, C.S., Gevers, D., Duke, F., Jung, J., Michaud, M., Walker, B.J., Young, S., (2013) Sequence-based discovery of *Bradyrhizobium enterica* in cord colitis syndrome, N. Engl. J. Med. 369, 517-528.

[12] Blom, J., Kreis, J., Spänig, S., Juhre, T., Bertelli, C., Ernst, C., Goesmann, A., (2016) EDGAR 2.0: an enhanced software platform for comparative gene content analyses, Nucleic Acids Res. 44, W22-W28.

[13] Bonaldi, K., Gargani, D., Prin, Y., Fardoux, J., Gully, D., Nouwen, N., Goormachtig, S., Giraud, E., (2011) Nodulation of Aeschynomene afraspera and A. indica by photosynthetic Bradyrhizobium sp. strain ORS285: the nod-dependent versus the nod-independent symbiotic interaction, Mol. Plant Microbe Interact. 24, 1359-1371.

[14] Bromfield, E.S., Cloutier, S., Tambong, J.T., Thi, T.V.T., (2017) Soybeans inoculated with root zone soils of Canadian native legumes harbour diverse and novel *Bradyrhizobium* spp. that possess agricultural potential, Syst. Appl. Microbiol. 40, 440-447.

[15] Bünger, W., Grönemeyer, J.L., Sarkar, A., Reinhold-Hurek, B., (2018) Bradyrhizobium ripae sp. nov., a nitrogen-fixing symbiont isolated from nodules of wild legumes in Namibia, Int. J. Syst. Evol. Microbiol.

[16] Bushnell, B., (2014) BBTools software package, URL http://sourceforge. net/projects/bbmap.

[17] Castresana, J., (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17, 540-552.
[18] Chaintreuil, C., Giraud, E., Prin, Y., Lorquin, J., Bâ, A., Gillis, M., De Lajudie, P., Dreyfus, B., (2000) Photosynthetic bradyrhizobia are natural endophytes of the African wild rice *Oryza breviligulata*, Appl. Environ. Microbiol. 66, 5437-5447.
[19] Chan, J.Z., Halachev, M.R., Loman, N.J., Constantinidou, C., Pallen, M.J., (2012) Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*, BMC Microbiol. 12, 302.
[20] Chun, J., Oren, A., Ventosa, A., Christensen, H., Arahal, D.R., da Costa, M.S., Rooney, A.P., Yi, H., Xu, X.-W., De Meyer, S., (2018) Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes, Int. J. Syst. Evol. Microbiol. 68, 461-466.
[21] Cleenwerck, I., Vandemeulebroecke, K., Janssens, D., Swings, J., (2002) Re-examination of the genus *Acetobacter*, with descriptions of *Acetobacter cerevisiae* sp. nov. and *Acetobacter malorum* sp. nov, Int. J. Syst. Evol. Microbiol. 52, 1551-1558.
[22] Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research, Bioinformatics 21, 3674-3676.
[23] Daubin, V., Gouy, M., Perriere, G., (2002) A phylogenomic approach to bacterial phylogeny: evidence of a core of genes sharing a common history, Genome Res. 12, 1080-1090.
[24] De Meyer, S.E., Willems, A., (2012) Multilocus sequence analysis of Bosea species and description of *Bosea lupini* sp. nov., *Bosea lathyri* sp. nov. and *Bosea robiniae* sp. nov., isolated from legumes, Int. J. Syst. Evol. Microbiol. 62, 2505-2510.
[25] de Souza, J.A.M., Alves, L.M.C., de Mello Varani, A., de Macedo Lemos, E.G., (2014) The family bradyrhizobiaceae, in: The Prokaryotes, Springer, pp. 135-154.
[26] de Souza, J.A.M., Tieppo, E., de Souza Magnani, G., Alves, L.M.C., Cardoso, R.L., Cruz, L.M., de Oliveira, L.F., Raittz, R.T., de Souza, E.M., de Oliveira Pedrosa, F., (2012) Draft genome sequence of the nitrogen-fixing symbiotic bacterium *Bradyrhizobium elkanii* 587, J. Bacteriol. 194, 3547-3548.
[27] Delamuta, J.R.M., Ribeiro, R.A., Ormeno-Orrillo, E., Melo, I.S., Martínez-Romero, E., Hungria, M., (2013) Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov, Int. J. Syst. Evol. Microbiol. 63, 3342-3351.
[28] Durán, D., Rey, L., Sanchez-Cañizares, C., Jorrin, B., Imperial, J., Ruiz-Argüeso, T., (2013) Biodiversity of slow-growing Rhizobia: the genus *Bradyrhizobium*, CRC Press, Boca Raton, FL: .
[29] Eaglesham, A.R., Ellis, J.M., Evans, W.R., Fleischman, D.E., Hungria, M., Hardy, R.W., (1990) The first photosynthetic N 2-fixing *Rhizobium*: characteristics, in: Nitrogen Fixation, Springer, pp. 805-811.
[30] Edgar, R.C., (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32, 1792-1797.
[31] Estrada-de los Santos, P., Palmer, M., Chávez-Ramírez, B., Beukes, C., Steenkamp, E., Briscoe, L., Khan, N., Maluk, M., Lafos, M., Humm, E., (2018) Whole Genome Analyses Suggests that Burkholderia *sensu lato* Contains Two Additional Novel Genera (*Mycetohabitans* gen. nov., and *Trinickia* gen. nov.): Implications for the Evolution of Diazotrophy and Nodulation in the Burkholderiaceae, Genes 9, 389.
[32] Euzéby, J.P., (1997) List of Bacterial Names with Standing in Nomenclature: a folder available on the Internet, Int. J. Syst. Bacteriol. 47, 590-592.
[33] Flemming, H.-C., Wingender, J., (2010) The biofilm matrix, Nat. Rev. Microbiol. 8, 623.

[34] Fonseca, M.B., Peix, A., de Faria, S.M., Mateos, P.F., Rivera, L.P., Simões-Araujo, J.L., França, M.G.C., dos Santos Isaias, R.M., Cruz, C., Velázquez, E., (2012) Nodulation in *Dimorphandra wilsonii Rizz.*(Caesalpinioideae), a threatened species native to the Brazilian Cerrado, PLoS One 7, e49520.
[35] Galtier, N., Daubin, V., (2008) Dealing with incongruence in phylogenomic analyses, Philos. Trans. Royal Soc. B 363, 4023-4029.
[36] Germano, M.G., Menna, P.m., Mostasso, F.L., Hungria, M., (2006) RFLP analysis of the rRNA operon of a Brazilian collection of bradyrhizobial strains from 33 legume species, Int. J. Syst. Evol. Microbiol. 56, 217-229.
[37] Giraud, E., Fleischman, D., (2004) Nitrogen-fixing symbiosis between photosynthetic bacteria and legumes, Photosynth. Res. 82, 115-130.
[38] Giraud, E., Moulin, L., Vallenet, D., Barbe, V., Cytryn, E., Avarre, J.-C., Jaubert, M., Simon, D., Cartieaux, F., Prin, Y., (2007) Legumes symbioses: absence of Nod genes in photosynthetic bradyrhizobia, Science 316, 1307-1312.
[39] Glaeser, S.P., Kämpfer, P., (2015) Multilocus sequence analysis (MLSA) in prokaryotic taxonomy, Syst. Appl. Microbiol. 38, 237-245.
[40] Gnerre, S., MacCallum, I., Przybylski, D., Ribeiro, F.J., Burton, J.N., Walker, B.J., Sharpe, T., Hall, G., Shea, T.P., Sykes, S., (2011) High-quality draft assemblies of mammalian genomes from massively parallel sequence data, PNAS 108, 1513-1518.
[41] Goris, J., Konstantinidis, K.T., Klappenbach, J.A., Coenye, T., Vandamme, P., Tiedje, J.M., (2007) DNA–DNA hybridization values and their relationship to whole-genome sequence similarities, Int. J. Syst. Evol. Microbiol. 57, 81-91.
[42] Gourion, B., Bonaldi, K., Giraud, E., (2015) Metabolism of Photosynthetic Bradyrhizobia during Root and Stem Symbiosis with *Aeschynomene* Legumes, Biological Nitrogen Fixation 2, 283.
[43] Grönemeyer, J.L., Reinhold-Hurek, B., (2018) Diversity of Bradyrhizobia in Subsahara Africa: A Rich Resource, Front. Microbiol. 9.
[44] Heath, T.A., Hedtke, S.M., Hillis, D.M., (2008) Taxon sampling and the accuracy of phylogenetic analyses, J Syst Evol 46, 239-257.
[45] Hillis, D.M., (1998) Taxonomic sampling, phylogenetic accuracy, and investigator bias, Syst. Biol. 47, 3-8.
[46] Hollis, A., Kloos, W., Elkan, G., (1981) DNA: DNA hybridization studies of *Rhizobium japonicum* and related Rhizobiaceae, Microbiology 123, 215-222.
[47] Hungria, M., Menna, P., Delamuta, J.R.M., (2015) *Bradyrhizobium*, the ancestor of all rhizobia: phylogeny of housekeeping and nitrogen-fixation genes, Biological Nitrogen Fixation 2, 191-202.
[48] Jaiswal, S.K., Dakora, F.D., (2019) Widespread distribution of highly adapted *Bradyrhizobium* species nodulating diverse legumes in Africa, Front. Microbiol. 10, 310.
[49] Jang, J., Ashida, N., Kai, A., Isobe, K., Nishizawa, T., Otsuka, S., Yokota, A., Senoo, K., Ishii, S., (2018) Presence of Cu-Type (NirK) and cd1-Type (NirS) Nitrite Reductase Genes in the Denitrifying Bacterium Bradyrhizobium nitroreducens sp. nov, Microbes Environ. 33, 326-331.
[50] Jones, F.P., Clark, I.M., King, R., Shaw, L.J., Woodward, M.J., Hirsch, P.R., (2016) Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes–a genome comparison, Sci. Rep. 6, 25858.
[51] Jordan, D., (1982) Transfer of *Rhizobium japonicum* Buchanan 1980 to Bradyrhizobium gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants, Int. J. Syst. Evol. Microbiol. 32, 136-139.

[52] Kanehisa, M., Sato, Y., Morishima, K., (2016) BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences, J. Mol. Biol. 428, 726-731.
[53] Kaneko, T., Nakamura, Y., Sato, S., Minamisawa, K., Uchiumi, T., Sasamoto, S., Watanabe, A., Idesawa, K., Iriguchi, M., Kawashima, K., (2002) Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110, DNA Res. 9, 189-197.
[54] Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data, Bioinformatics 28, 1647-1649.
[55] Konstantinidis, K.T., Tiedje, J.M., (2007) Prokaryotic taxonomy and phylogeny in the genomic era: advancements and challenges ahead, Curr. Opin. Microbiol. 10, 504-509.
[56] Kück, P., Longo, G.C., (2014) FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies, Front. Zool. 11, 81.
[57] Kuykendall, L., Saxena, B., Devine, T., Udell, S., (1992) Genetic diversity in *Bradyrhizobium japonicum* Jordan 1982 and a proposal for *Bradyrhizobium elkanii* sp. nov, Can. J. Microbiol. 38, 501-505.
[58] Kyrpides, N.C., Hugenholtz, P., Eisen, J.A., Woyke, T., Göker, M., Parker, C.T., Amann, R., Beck, B.J., Chain, P.S., Chun, J., (2014) Genomic encyclopedia of bacteria and archaea: sequencing a myriad of type strains, PLoS Biol. 12, e1001920.
[59] Kyrpides, N.C., Woyke, T., Eisen, J.A., Garrity, G., Lilburn, T.G., Beck, B.J., Whitman, W.B., Hugenholtz, P., Klenk, H.-P., (2014) Genomic Encyclopedia of Type Strains, Phase I: the one thousand microbial genomes (KMG-I) project, Stand. Genomic Sci. 9, 1278.
[60] La Scola, B., Mallet, M.-N., Grimont, P.A., Raoult, D., (2003) *Bosea eneae* sp. nov., *Bosea massiliensis* sp. nov. and *Bosea vestrisii* sp. nov., isolated from hospital water supplies, and emendation of the genus *Bosea* (Das et al. 1996), Int. J. Syst. Evol. Microbiol. 53, 15-20.
[61] Lang, J.M., Darling, A.E., Eisen, J.A., (2013) Phylogeny of bacterial and archaeal genomes using conserved genes: supertrees and supermatrices, PLoS One 8, e62510.
[62] Levy, S.E., Myers, R.M., (2016) Advancements in next-generation sequencing, Annu. Rev. Genomics Hum. G. 17, 95-115.
[63] Lindström, K., Aserse, A.A., Mousavi, S.A., (2015) Evolution and taxonomy of nitrogen-fixing organisms with emphasis on rhizobia, John Wiley and Sons, Inc, Hoboken, NJ, USA.
[64] Lo, S.-C., Hung, G.-C., Li, B., Lei, H., Li, T., Nagamine, K., Zhang, J., Tsai, S., Bryant, R., (2013) Isolation of novel *Afipia septicemium* and identification of previously unknown bacteria *Bradyrhizobium* sp. OHSU_III from blood of patients with poorly defined illnesses, PLoS One 8, e76142.
[65] Loman, N.J., Pallen, M.J., (2015) Twenty years of bacterial genome sequencing, Nat. Rev. Microbiol. 13, 787.
[66] Lopes, E.M., Castellane, T.C.L., Moretto, C., Lemos, E., Souza, J., (2014) Emulsification properties of bioemulsifiers produced by wild-type and mutant *Bradyrhizobium elkanii* strains, J. Bioremediat. Biodegrad. 5, 1-6.
[67] Luo, C., Rodriguez-r, L.M., Konstantinidis, K.T., (2014) MyTaxa: an advanced taxonomic classifier for genomic and metagenomic sequences, Nucleic Acids Res. 42, e73-e73.
[68] Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., (2005) Genome sequencing in microfabricated high-density picolitre reactors, Nature 437, 376.
[69] Menna, P.m., Hungria, M., (2011) Phylogeny of nodulation and nitrogen-fixation genes in *Bradyrhizobium*: supporting evidence for the theory of monophyletic origin, and spread and maintenance by both horizontal and vertical transfer, Int. J. Syst. Evol. Microbiol. 61, 3052-3067.

[70] Michel, D.C., Passos, S.R., Simões-Araujo, J.L., Baraúna, A.C., da Silva, K., Parma, M.M., Melo, I.S., De Meyer, S.E., O'Hara, G., Zilli, J.E., (2017) *Bradyrhizobium centrolobii* and *Bradyrhizobium macuxiense* sp. nov. isolated from *Centrolobium paraense* grown in soil of Amazonia, Brazil, Arch. Microbiol. 199, 657-664.
[71] Molouba, F., Lorquin, J., Willems, A., Hoste, B., Giraud, E., Dreyfus, B., Gillis, M., de Lajudie, P., Masson-Boivin, C., (1999) Photosynthetic bradyrhizobia from *Aeschynomene* spp. are specific to stem-nodulated species and form a separate 16S ribosomal DNA restriction fragment length polymorphism group, Appl. Environ. Microbiol. 65, 3084-3094.
[72] Mornico, D., Miché, L., Béna, G., Nouwen, N., Verméglio, A., Vallenet, D., Smith, A.A., Giraud, E., Médigue, C., Moulin, L., (2011) Comparative genomics of *Aeschynomene* symbionts: insights into the ecological lifestyle of nod-independent photosynthetic bradyrhizobia, Genes 3, 35-61.
[73] Moulin, L., Béna, G., Boivin-Masson, C., Stępkowski, T., (2004) Phylogenetic analyses of symbiotic nodulation genes support vertical and lateral gene co-transfer within the *Bradyrhizobium* genus, Mol. Phylogenet. Evol. 30, 720-732.
[74] Mukherjee, S., Seshadri, R., Varghese, N.J., Eloe-Fadrosh, E.A., Meier-Kolthoff, J.P., Göker, M., Coates, R.C., Hadjithomas, M., Pavlopoulos, G.A., Paez-Espino, D., (2017) 1,003 reference genomes of bacterial and archaeal isolates expand coverage of the tree of life, Nat. Biotechnol. 35, 676.
[75] Ojha, A., Tak, N., Rathi, S., Chouhan, B., Rao, S.R., Barik, S.K., Joshi, S.R., Sprent, J.S., James, E.K., Gehlot, H.S., (2017) Molecular characterization of novel *Bradyrhizobium* strains nodulating *Eriosema chinense* and *Flemingia vestita*, important unexplored native legumes of the sub-Himalayan region (Meghalaya) of India, Syst. Appl. Microbiol. 40, 334-344.
[76] Okazaki, S., Noisangiam, R., Okubo, T., Kaneko, T., Oshima, K., Hattori, M., Teamtisong, K., Songwattana, P., Tittabutr, P., Boonkerd, N., (2015) Genome analysis of a novel *Bradyrhizobium* sp. DOA9 carrying a symbiotic plasmid, PLoS One 10, e0117392.
[77] Okubo, T., Tsukui, T., Maita, H., Okamoto, S., Oshima, K., Fujisawa, T., Saito, A., Futamata, H., Hattori, R., Shimomura, Y., (2012) Complete genome sequence of *Bradyrhizobium* sp. S23321: insights into symbiosis evolution in soil oligotrophs, Microbes Environ. 27, 306-315.
[78] Ormeño-Orrillo, E., Rogel, M.A., Zúñiga-Dávila, D., Martínez-Romero, E., (2018) Complete Genome Sequence of the Symbiotic Strain *Bradyrhizobium icense* LMTR 13T, Isolated from Lima Bean (*Phaseolus lunatus*) in Peru, Genome Announc. 6, e00146-00118.
[79] Palmer, M., Steenkamp, E.T., Coetzee, M.P., Chan, W.-Y., Van Zyl, E., De Maayer, P., Coutinho, T.A., Blom, J., Smits, T.H., Duffy, B., (2017) Phylogenomic resolution of the bacterial genus *Pantoea* and its relationship with *Erwinia* and *Tatumella*, Antonie Leeuwenhoek 110, 1287-1309.
[80] Parker, M.A., (2015) The spread of *Bradyrhizobium* lineages across host legume clades: from *Abarema* to *Zygia*, Microb. Ecol. 69, 630-640.
[81] Parks, D.H., Chuvochina, M., Waite, D.W., Rinke, C., Skarshewski, A., Chaumeil, P.-A., Hugenholtz, P., (2018) A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life, Nat. Biotechnol.
[82] Parks, D.H., Imelfort, M., Skennerton, C.T., Hugenholtz, P., Tyson, G.W., (2015) CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, Genome Res. gr. 186072.186114.
[83] Parte, A.C., (2013) LPSN—list of prokaryotic names with standing in nomenclature, Nucleic Acids Res. 42, D613-D616.

[84] Parte, A.C., (2018) LPSN-List of Prokaryotic names with Standing in Nomenclature (bacterio. net), 20 years on, Int. J. Syst. Evol. Microbiol. 68, 1825-1829.
[85] Pascual, J., Macián, M.C., Arahal, D.R., Garay, E., Pujalte, M.J., (2010) Multilocus sequence analysis of the central clade of the genus *Vibrio* by using the 16S rRNA, *recA, pyrH, rpoD, gyrB, rctB* and *toxR* genes, Int. J. Syst. Evol. Microbiol. 60, 154-165.
[86] Peix, A., Ramírez-Bahena, M.H., Flores-Félix, J.D., de la Vega, P.A., Rivas, R., Mateos, P.F., Igual, J.M., Martínez-Molina, E., Trujillo, M.E., Velázquez, E., (2015) Revision of the taxonomic status of the species *Rhizobium lupini* and reclassification as *Bradyrhizobium lupini* comb. nov, Int. J. Syst. Evol. Microbiol. 65, 1213-1219.
[87] Prasanna, A.N., Mehra, S., (2013) Comparative phylogenomics of pathogenic and non-pathogenic mycobacterium, PLoS One 8, e71248.
[88] Provorov, N., Andronov, E., (2016) Evolution of root nodule bacteria: reconstruction of the speciation processes resulting from genomic rearrangements in a symbiotic system, Microbiology 85, 131-139.
[89] Ramírez-Bahena, M.-H., Chahboune, R., Peix, A., Velázquez, E., (2013) Reclassification of *Agromonas oligotrophica* into the genus *Bradyrhizobium* as *Bradyrhizobium oligotrophicum* comb. nov, Int. J. Syst. Evol. Microbiol. 63, 1013-1016.
[90] Reeve, W., Ardley, J., Tian, R., Eshragi, L., Yoon, J.W., Ngamwisetkun, P., Seshadri, R., Ivanova, N.N., Kyrpides, N.C., (2015) A genomic encyclopedia of the root nodule bacteria: Assessing genetic diversity through a systematic biogeographic survey, Stand. Genomic Sci. 10, 14.
[91] Richter, M., Rosselló-Móra, R., (2009) Shifting the genomic gold standard for the prokaryotic species definition, PNAS 106, 19126-19131.
[92] Rivas, R., García-Fraile, P., Velázquez, E., (2009) Taxonomy of bacteria nodulating legumes, Microbiol. Insights 2, MBI. S3137.
[93] Rivas, R., Martens, M., De Lajudie, P., Willems, A., (2009) Multilocus sequence analysis of the genus *Bradyrhizobium*, Syst. Appl. Microbiol. 32, 101-110.
[94] Sachs, J., Ehinger, M., Simms, E., (2010) Origins of cheating and loss of symbiosis in wild *Bradyrhizobium*, J. Evol. Biol. 23, 1075-1089.
[95] Safronova, V.I., Kuznetsova, I.G., Sazanova, A.L., Kimeklis, A.K., Belimov, A.A., Andronov, E.E., Pinaev, A.G., Chizhevskaya, E.P., Pukhaev, A.R., Popov, K.P., (2015) *Bosea vaviloviae* sp. nov., a new species of slow-growing rhizobia isolated from nodules of the relict species *Vavilovia formosa* (Stev.) Fed, Antonie Leeuwenhoek 107, 911-920.
[96] Sahgal, M., Johri, B., (2003) The changing face of rhizobial systematics, Curr. Sci. 43-48.
[97] Sawada, H., Kuykendall, L.D., Young, J.M., (2003) Changing concepts in the systematics of bacterial nitrogen-fixing legume symbionts, J. Gen. Appl. Microbiol. 49, 155-179.
[98] Servín-Garcidueñas, L.E., Rogel, M.A., Ormeño-Orrillo, E., Zayas-del Moral, A., Sánchez, F., Martínez-Romero, E., (2016) Complete genome sequence of *Bradyrhizobium* sp. strain CCGE-LA001, isolated from field nodules of the enigmatic wild bean *Phaseolus microcarpus*, Genome Announc. 4, e00126-00116.
[99] So, R.B., Ladha, J.K., Young, J.P.W., (1994) Photosynthetic symbionts of *Aeschynomene* spp. form a cluster with bradyrhizobia on the basis of fatty acid and rRNA analyses, Int. J. Syst. Evol. Microbiol. 44, 392-403.
[100] Soto-Giron, M.J., Rodriguez-R, L.M., Luo, C., Elk, M., Ryu, H., Hoelle, J., Santo Domingo, J.W., Konstantinidis, K.T., (2016) Characterization of biofilms developing on hospital shower hoses and implications for nosocomial infections, Appl. Environ. Microbiol. AEM. 03529-03515.
[101] Sprent, J.I., Ardley, J., James, E.K., (2017) Biogeography of nodulated legumes and their nitrogen- fixing symbionts, New Phytol. 215, 40-56.

[102] Stamatakis, A., (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics 30, 1312-1313.
[103] Steenkamp, E.T., Stępkowski, T., Przymusiak, A., Botha, W.J., Law, I.J., (2008) Cowpea and peanut in southern Africa are nodulated by diverse *Bradyrhizobium* strains harboring nodulation genes that belong to the large pantropical clade common in Africa, Mol. Phylogenet. Evol. 48, 1131-1144.
[104] Stępkowski, T., Banasiewicz, J., Granada, C.E., Andrews, M., Passaglia, L.M., (2018) Phylogeny and Phylogeography of Rhizobial Symbionts Nodulating Legumes of the Tribe Genisteae, Genes 9, 163.
[105] Stępkowski, T., Czaplińska, M., Miedzinska, K., Moulin, L., (2003) The variable part of the *dnaK* gene as an alternative marker for phylogenetic studies of rhizobia and related alpha Proteobacteria, Syst. Appl. Microbiol. 26, 483-494.
[106] Stępkowski, T., Watkin, E., McInnes, A., Gurda, D., Gracz, J., Steenkamp, E.T., (2012) Distinct *Bradyrhizbium* communities nodulate legumes native to temperate and tropical monsoon Australia, Mol. Phylogenet. Evol. 63, 265-277.
[107] Stępkowski, T., Żak, M., Moulin, L., Króliczak, J., Golińska, B., Narożna, D., Safronova, V.I., Mądrzak, C.J., (2011) *Bradyrhizobium canariense* and *Bradyrhizobium japonicum* are the two dominant rhizobium species in root nodules of lupin and serradella plants growing in Europe, Syst. Appl. Microbiol. 34, 368-375.
[108] Tian, C.F., Zhou, Y.J., Zhang, Y.M., Li, Q.Q., Zhang, Y.Z., Li, D.F., Wang, S., Wang, J., Gilbert, L.B., Li, Y.R., (2012) Comparative genomics of rhizobia nodulating soybean suggests extensive recruitment of lineage-specific genes in adaptations, PNAS 109, 8629-8634.
[109] Tian, R., Parker, M., Seshadri, R., Reddy, T., Markowitz, V., Ivanova, N., Pati, A., Woyke, T., Baeshen, M.N., Baeshen, N.A., (2015) High-quality permanent draft genome sequence of Bradyrhizobium sp. Tv2a. 2, a microsymbiont of Tachigali versicolor discovered in Barro Colorado Island of Panama, Stand. Genomic Sci. 10, 27.
[110] Van Berkum, P., Eardly, B.D., (2002) The aquatic budding bacterium *Blastobacter denitrificans* is a nitrogen-fixing symbiont of *Aeschynomene indica*, Appl. Environ. Microbiol. 68, 1132-1136.
[111] Van Berkum, P., Leibold, J.M., Eardly, B.D., (2006) Proposal for combining *Bradyrhizobium* spp.(*Aeschynomene indica*) with *Blastobacter denitrificans* and to transfer *Blastobacter denitrificans* (Hirsch and Muller, 1985) to the genus *Bradyrhizobium* as *Bradyrhizobium denitrificans* (comb. nov.), Syst. Appl. Microbiol. 29, 207-215.
[112] VanInsberghe, D., Maas, K.R., Cardenas, E., Strachan, C.R., Hallam, S.J., Mohn, W.W., (2015) Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils, ISME J. 9, 2435.
[113] Vega- Hernández, M., Pérez- Galdona, R., Dazzo, F., Jarabo- Lorenzo, A., Alfayate, M., León- Barrios, M., (2001) Novel infection process in the indeterminate root nodule symbiosis between *Chamaecytisus proliferus* (tagasaste) and *Bradyrhizobium* sp, New Phytol. 150, 707-721.
[114] Vinuesa, P., Rojas-Jiménez, K., Contreras-Moreira, B., Mahna, S.K., Prasad, B.N., Moe, H., Selvaraju, S.B., Thierfelder, H., Werner, D., (2008) Multilocus sequence analysis for assessment of the biogeography and evolutionary genetics of four *Bradyrhizobium* species that nodulate soybeans on the Asiatic continent, Appl. Environ. Microbiol. 74, 6987-6996.
[115] Vinuesa, P., Silva, C., Werner, D., 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.

[116] Wang, J.Y., Wang, R., Zhang, Y.M., Liu, H.C., Chen, W.F., Wang, E.T., Sui, X.H., Chen, W.X., (2013) *Bradyrhizobium daqingense* sp. nov., isolated from soybean nodules, Int. J. Syst. Evol. Microbiol. 63, 616-624.
[117] Whitman, W.B., (2014) The need for change: embracing the genome, in: Methods in Microbiology, Elsevier, pp. 1-12.
[118] Whitman, W.B., Woyke, T., Klenk, H.-P., Zhou, Y., Lilburn, T.G., Beck, B.J., De Vos, P., Vandamme, P., Eisen, J.A., Garrity, G., (2015) Genomic encyclopedia of bacterial and archaeal type strains, phase III: the genomes of soil and plant-associated and newly described type strains, Stand. Genomic Sci. 10, 26.
[119] Willems, A., Coopman, R., Gillis, M., (2001) Phylogenetic and DNA-DNA hybridization analyses of *Bradyrhizobium* species, Int. J. Syst. Evol. Microbiol. 51, 111-117.
[120] Willems, A., Doignon-Bourcier, F., Goris, J., Coopman, R., de Lajudie, P., De Vos, P., Gillis, M., (2001) DNA-DNA hybridization study of *Bradyrhizobium* strains, Int. J. Syst. Evol. Microbiol. 51, 1315-1322.
[121] Wu, D., Hugenholtz, P., Mavromatis, K., Pukall, R., Dalin, E., Ivanova, N.N., Kunin, V., Goodwin, L., Wu, M., Tindall, B.J., (2009) A phylogeny-driven genomic encyclopaedia of Bacteria and Archaea, Nature 462, 1056.
[122] Xu, L., Ge, C., Cui, Z., Li, J., Fan, H., (1995) *Bradyrhizobium liaoningense* sp. nov., isolated from the root nodules of soybeans, Int. J. Syst. Evol. Microbiol. 45, 706-711.
[123] Yarza, P., Richter, M., Peplies, J., Euzeby, J., Amann, R., Schleifer, K.-H., Ludwig, W., Glöckner, F.O., Rosselló-Móra, R., (2008) The All-Species Living Tree project: a 16S rRNA-based phylogenetic tree of all sequenced type strains, Syst. Appl. Microbiol. 31, 241-250.
[124] Yu, J., Blom, J., Glaeser, S., Jaenicke, S., Juhre, T., Rupp, O., Schwengers, O., Spänig, S., Goesmann, A., (2017) A review of bioinformatics platforms for comparative genomics. Recent developments of the EDGAR 2.0 platform and its utility for taxonomic and phylogenetic studies, J. Biotechnol. 261, 2-9.
[125] Zerbino, D., Birney, E., (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs, Genome Res. gr. 074492.074107.
[126] Zwickl, D.J., Hillis, D.M., (2002) Increased taxon sampling greatly reduces phylogenetic error, Syst. Biol. 51, 588-598.

Table 1. Summary of the supergroup status, host and geographic origin of the isolates included in this study [1].

Supergroup [2]	**Described species and type strains**	**Other isolates**	**Host or Substrate**	**Geographic origin**
B. japonicum	*B. arachidis* LMG 26795^T, *B. centrolobii* BR 10245^T, *B. diazoefficiens* USDA 110^T, *B. forestalis* INPA54B^T, *B. japonicum* USDA 6^T, *B. neotropicale* BR10247^T, *B. ottawaense* OO99^T, *B. sacchari* BR 10280^T, *B. shewense* ERR11^T, *B. stylosanthis* BR 446^T, *B. yuanmingense* CCBAU 10071^T	22, 39S1MB, 85S1MB, AS23.2, AT1, BF49, BR 3267, CB756, CCBAU 05623, CCBAU 15354, CCBAU 15517, CCBAU 15544, CCBAU 15615, CCBAU 15618, CCBAU 15635, CCBAU 25021, CCBAU 25435, CCBAU 35157, CCBAU 41267, CCBAU 43298, CCGE_LA001, CCH5-F6, CCNWSX0360, Cf659, Cp5.3, DOA1, DOA9, E109, Ec3.3, FN1, G22, GHa, GHvi, In8p8, Is-1, Is-34, Is5, J5, L2, Leaf396, LTSP849, LTSP857, NAS80.1, NAS96.2, NK6, OKO95, Rc2d, Rc3b, S23321 , SEMIA 5079, SEMIA 5080, Th.b2, TSA1, UBMA050, UBMA051, UBMA052, UBMA060, UBMA061, UBMA122, UBMA171, UBMA181, UBMA182, UBMA183, UBMA192, UBMA195, UBMA197, UBMA510, UBMAN05, URHA0013, USDA 122, USDA 123, USDA 124, USDA 135, USDA 3384, USDA 38, USDA 4, WSM1253, WSM1417, WSM1743, Y21, WSM2254, WSM2793, WSM3983, WSM471, Y36, YR681	*Acacia dealbata, Aeschynomene americana, Amphicarpaea bracteata, Arabidopsis thaliana, Arachis hypogaea, Centrolobium paraense, Centrosema pubescens, Crotalaria paulina, Erythrina brucei, Erythrina costaricensis, Glycine max, Indigofera* sp., *Ipomoea batatas, Kennedia coccinea, Lespedeza cuneata, Lupinus albescens, Lupinus angustifolius, Lupinus* sp., *Macrotyloma africanum, Ornithopus compressus, Ornithopus pinnatus, Phaseolus microcarpus, Phaseolus vulgaris, Populus deltoides, Populus* sp., *Rhynchosia totta, Saccharum* sp., *Stylosanthes guianensis, Vigna unguiculata,* Grassland soil, Forest soil, Hospital biofilm, Paddy soil, Soil	Algeria. Argentina, Australia, Barro Colorado Island, Botswana, Brazil, Canada, Chile, China, Ethiopia, Greek Island of Sifnos, Japan, Mexico, Panama, South Africa, Switzerland, Thailand, United Kingdom, USA, Zimbabwe
B. elkanii	*B. brasilense* UFLA 03-321^T, *B. elkanii* USDA 76^T, *B. embrapense* SEMIA 6208^T, *B. macuxiense* BR 10303^T, *B. manausense* BR3351^T, *B. mercentei* SEMIA 6399^T, *B. pachyrhizi* PAC48^T, *B. tropiciagri* SEMIA 6148^T, *B. viridifuturi* SEMIA 690^T	587, Ai1a-2, BLY3-8, BLY6-1, BR3262, C9, CCBAU 05737, CCBAU 43297, DFCI-1, Leo121, Leo170, LTSPM299, LTSP885, MT12, OHSU_III, PARBB1, R5, SK17, TnphoA33, UASWS1015, USDA 94, USDA 3254, USDA 3259, WSM2783	*Andira inermis, Asterionella formosa, Centrolobium paraense, Centrosema pubescens, Deguelia costata, Desmodium heterocarpon, Glycine max, Leobordea carinata, Leobordea divaricata, Leobordea lanceolata, Neonotonia wightii, Pachyrhizus erosus, Phaseolus acutifolius, Phaseolus lunatus, Phaseolus vulgaris, Vigna unguiculata,*	Argentina, Brazil, Canada, China, Costa Rica, Myanmar, South Africa, South Korea, Switzerland, United Kingdom, USA

			Agricultural soil, Forest soil, Sewage, Human blood, Human tissue specimen	
Photosynthetic	*B. oligotrophicum* S58^{T}	BTAi1, ORS 278, ORS 285, ORS 375, STM 3809, STM 3843	*Aeschynomene afraspera*, *Aeschynomene indica*, *Aeschynomene sensitive*	French Guiana, Senegal, USA
Kakadu	None	ARR65, Tv2a-2	*Stylosanthes viscosa*, *Tachigali versicolor*	Australia, Barro Colorado Island of Panama
B. jicamae	*B. icense* LMTR13^{T}, *B. jicamae* PAC68^{T}, *B. lablabi* CCBAU 23086^{T}, *B. paxllaeri* LMTR21^{T}, *B. retamae* Ro19^{T}, *B. valentinum* LmjM3^{T}	GAS478, GAS522, GAS524, LMTR3, MT34, URHA0002, WSM1741	*Lablab purpureus, Lupinus mariae-josephae, Pachyrhizus erosus, Phaseolus lunatus, Retama monosperma, Rhynchosia minima,* Forest soil, Soil	Australia, China, Honduras, Morocco, Peru, Spain, USA
Soil 1	None	GAS165	Forest soil	USA
Soil 2	None	GAS138, GAS242, GAS369, GAS499, URHD0069	Forest soil, Soil	USA

[1] For full details of the individual isolates, see supplementary Table S2.

[2] These designations are according to Figure 1 and Supplementary Figure S2.

Supplementary material

Supplementary Figure S1. Chromosomal positions of the 128 and 400 protein-coding genes used for the construction of the maximum likelihood phylogenies (Figure 1 and Supplementary Figure S2) distributed on the chromosome, using the type species *B. japonicum* USDA 6^T as reference. Genes are a representation of the shared genes of respective taxa investigated in this study. For more information see Supplementary Table S3.

Supplementary Figure S2. Phylogenies based on amino acid sequences for the genus *Bradyrhizobium* with appropriate outgroup taxa. a) Summary tree inferred from the 400-gene, 180-taxa maximum likelihood phylogeny, illustrating the intra-generic relationships of *Bradyrhizobium*. Branch support was inferred from 1000 bootstrap replicates, support values of 100% are indicated by thicker tree branches. b) A maximum likelihood phylogeny of the amino acid sequences of 400 genes shared by 180 taxa (*Bradyrhizobium*+Bradyrhizobiaceae). *Pseudorhodoplanes sinepersici* RIPI110^T was used as an outgroup. Most of the tree branches received 100% bootstrap support and are colour coded same as insert. The scale bar indicates the number of amino acid changes per site.

Supplementary Figure S3. Phylogenies based on amino acid sequences of six concatenated genes for the genus *Bradyrhizobium* and its close relatives. a) Summary tree inferred from the 180-taxa MLSA maximum likelihood phylogeny, illustrating the intra-generic relationships of *Bradyrhizobium*. Branch support was inferred from 1000 bootstrap replicates, support values of >80% are indicated by thicker tree branches. b) Maximum likelihood (ML) phylogeny of 200 in- and outgroup taxa (*Bradyrhizobium*+Bradyrhizobiaceae+ Methylobacteriaceae+Xanthobacteraceae+Rhizobiaceae) based on an *atpD*, *dnaK*, *glnII*, *gyrB*, *recA* and *rpoB* concatenated dataset. Strains from Rhizobiaceae (*Sinorhizobium*, *Mesorhizobium* and *Rhizobium*) were used as an outgroup. The scale bar indicates the number of amino acid changes per site. Only ML bootstrap support of ≥60% is indicated.

Supplementary Figure S4. Phylogenies based on amino acid sequences of six concatenated genes for the genus *Bradyrhizobium* and its close relatives. a) Summary tree inferred from the 180-taxa MLSA maximum likelihood phylogeny, illustrating the intra-generic relationships of *Bradyrhizobium*. Branch support was inferred from 1000 bootstrap replicates, support values of

>80% are indicated by thicker tree branches. b) Maximum likelihood (ML) phylogeny of 180 in- and outgroup taxa (*Bradyrhizobium*+Bradyrhizobiaceae) based on *atpD*, *dnaK*, *glnII*, *gyrB*, *recA* and *rpoB* concatenated dataset. *Pseudorhodoplanes sinepersici* RIPI110^T was used as an outgroup. The scale bar indicates the number of amino acid changes per site. Only ML bootstrap support of ≥60% is indicated.

List of supplementary tables

Table S1. Genome properties of all bacterial taxa used in this study, as well as their *Bradyrhizobium* supergroup/clade status.

Table S2. Genome sequencing and assembly information of eight additional *Bradyrhizobium* strains.

Table S3. Functional roles of each of the 400 putative gene products used for inferring *Bradyrhizobium* phylogeny.

Table S4. Average Nucleotide Identity (ANI) calculated using JSpecies implemented in EDGAR.

Table S5. Average Amino Acid Identity (AAI) computed using the BLASTP alignment method in EDGAR.

Table S6. Absence and presence of nodulation, nitrogen fixation and photosynthetic genes investigated in all taxa included in this study.

Table S7. Functional Annotations of photosynthetic genes using Blast2GO.