The Structure and Evolution of Beta-Rhizobial Symbiotic Genes Deduced from Their Complete Genomes


1Key Laboratory of Microbial Diversity Research and Application of Hebei Province, College of Life Sciences, Hebei University, Baoding, China
2The James Hutton Institute, Invergowrie, Dundee DD2 5DA, UK

Abstract

Nitrogen-fixing Rhizobia were discovered more than 100 years ago. They are classified into two clades, Alpha- and Beta-rhizobia. Their symbiotic function is remarkable, but its origin and evolution has been confusing from a phylogenetic perspective. In this study, we make use of 33 publicly available complete genome sequences downloaded from NCBI, which consist of bacteria and archaea, and focus on 10 strains, constructing symbiotic structural maps for them based on their genomes and previous gene annotations. Phylogenies of the symbiosis-essential genes nodA and nifH were examined. Although large incongruities with some hypotheses from previous studies were detected by the present study, we support the general concept that Beta-rhizobia were the original symbionts of legumes, but that their symbiotic genes originated from a common ancestor to the Alpha-rhizobia. We also confirm that the spread and maintenance of symbiotic genes occurred mainly through vertical transmission, with lateral transfer playing a significant, albeit supporting, role.

Keywords: Rhizobia; Symbiosis genes; Origin; Evolution

Introduction

Rhizobia are soil bacteria able to establish a nitrogen-fixing symbiosis with leguminous plants. Most of them belong to the Alphaproteobacteria based on the sequences of the gene coding for 16S rRNA [1,2]. However, over the last 15 years studies have reported the presence of legume-nodulating bacteria in the genera Burkholderia and Cupriavidus (Burkholderiaceae) in the Betaproteobacteria [3-21]. Nodulation and nitrogen fixation capacity are very important factors in understanding the evolution of Rhizobia. Burkholderia and Cupriavidus strains were previously reported as exclusively non-symbiotic bacteria before these genera were discovered to contain Rhizobia, being isolated from soil, water, plants, rhizosphere and from infected humans [4,22-25]. This extreme diversity in habitats and ecological lifestyles illustrates their remarkable capacity for adaptation [26,27]. Beta-rhizobial symbionts have different geographical distributions, with South America and South Africa as their main centers of diversity. Mimosoid-nodulating Burkholderia symbionts have been isolated from native and invasive Mimoso species across Brazil, Uruguay, North America, Taiwan, China and Australia [5-10,28-32], as well as from related legumes in the Mimosoideae that are native and endemic to South America, particularly those in the “Piptadenia Group” [13,33,34]. Mimosoid-nodulating Cupriavidus symbionts were initially found in Taiwan, India, China and other parts of the tropics [4,6-28,29,35-39] and later isolated from the native ranges of their invasive hosts, M. pigna and M. pudica, in Costa Rica and Texas [9,39], and in recent years from various native Mimosoid hosts in French Guyana, Brazil and Uruguay [20,22,32,34]. Parallel studies on strains from South Africa revealed that Burkholderia symbionts were widespread in native and endemic papilionoid legumes in the tribes Podaierae, Crotalariae, Phaseoleae and Indigofereae [12,33,40-44]. In this context, it should be noted that the Burkholderia strains originated from South Africa are in different species to those so far described for the Mimosoideae-nodulating strains from South America, and that they are largely incapable of nodulating each other’s hosts. The only species so far shown to be in common between the two continents is B. tuberum [45], which exists in two symbiobars, sv. mimosae in South America and sv. papilionoideae in South Africa [32]. Interestingly, nodulating strains isolated from the invasive South African legume, Dipogon lignosus (Phaseoleae) in New Zealand and Australia have been shown to be largely Burkholderia and these are capable of nodulating many South African native legumes [46,47]. Taken together, this evidence indicates that South America and South Africa are centres of diversity of nodulating Burkholderia from Mimosoid and Papilionoid legumes, respectively, and this might indicate that the two continents, which were conjoined in the Cambrian period, share a symbiotic Burkholderia ancestor. Over evolutionary and geological time, the separation of the continents has resulted in a geographical distribution of Beta-rhizobia which implies that each group of symbionts has a special evolutionary history which has resulted in particular selection mechanisms between them and their legume hosts.

In order to form an effective symbiosis, Rhizobia require specific genes, which are usually located in regions within symbiotic plasmids (pSym) or in mobile genomic regions called symbiotic islands; these include nodulation genes (nod, nol and noe) and nitrogen-fixation genes (nif, fix and fdx genes). The nod genes specify the synthesis of lipo-oligosaccharide signals (LCOs), the so-called Nod factors (NFs), which are responsible for determining infection, nodule formation and the control of host-specificity [48]. Different types of nodulation genes were found within Rhizobia which can be divided into two sets. The first of these are the structural nod genes: the nodABC and nodIF genes, termed “common” because they are present in almost all rhizobial species, and the second group are the regulatory nod genes, such as nodD, whose gene-product, the nodD protein, activates the transcription of structural nod genes, and regulates the initial infection events. Other
nod genes, such as nodFE, nodH, nodI, nodP, nodQ, nodSU, nodX and nodZ, are present in various combination in rhizobial species and are called host-specific nod genes [49]. With regard to nitrogen fixation, which it should be stressed is not confined only to symbiotic bacteria, but is also widespread in free-living bacteria, the nitrogenase protein complex is an ATP-hydrolyzing, redox active complex of two main proteins, whose various components are encoded by a large set of genes. nif genes are found within all N-fixing bacteria (diazotrophs), and these encode the subunits of the functional nitrogenase protein and a suite of proteins involved with regulation, activation, metal transport, and cluster biosynthesis [50], such as the nifA and nifL genes (encoding regulators), the nitrogenase structural genes nifHDK, and other genes (nifX, nifV/W/fx/ABCX, nifB/fdx/N/nifZ/fixU). Other nitrogen fixation genes are denoted as fixation genes, which are related to respiration (fixNOQP), the nitrogen electron transport chain (fixABCX), and other regulatory genes, such as fixL, fixK and fixGHST [50].

The large symbiotic plasmid was included in the first analysis of a complete genome sequence in Rhizobium, that of Ensifer (syn. Sinorhizobium) fredii USDA110 [51]. Until now, a total of seven 90 rhizobial genomes have been sequenced (http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html), but the complete genome data for rhizobial strains are not sufficient for proper genome-level taxonomic and phylogenetical analyses, even though so many draft genome sequences are available. Nevertheless, in spite of this paucity in information, genome sequence analyses are increasingly being used in rhizobial taxonomy studies. For example, the first sequence of Rhizobium, that of R. leguminosarum sv. vicieae strain 3841, which is the only strain of R. leguminosarum which has been sequenced to date, shows that it harbors a circular chromosome and six circular plasmids [52]. Moreover, in the case of the Rhizobium/Agrobacterium genera which are clustered together in their 16S rDNA phylogenies, their complete genomes are highly supportive of them belonging to separate clades which could correspond to distinct genera [53]; this is further supported by the fact that two strains from the same species also displayed different genome traits, especially in their mobile symbiotic genes [54]. The first genomic study of a β-rhizobium, Cupriavidus taiwanensis LMG19424, revealed characteristics of a minimal rhizobium, including the most compact (35 kb) symbiotic island (nod and nif) identified so far in any rhizobium, suggesting that this Beta-rhizobial species evolved recently [55]. Betarhizobia belong to the versatile and environmentally diverse genera Burkholderia and Cupriavidus, some of which are opportunistic pathogens, but recent studies have suggested that nodulating bacteria differ from the pathogens in these genera (e.g. Burkholderia) in several aspects including secretion systems and other traits, and suggest that beta-rhizobia have the potential for safe application as beneficial plant inoculants [56].

Several studies have hypothesized the horizontal transfer of symbiotic genes between Alpha- and Beta-rhizobia, or between Burkholderia and Cupriavidus on the basis of phylogenies using sequences of their symbiosis-related genes, such as nodA, nodC and nifH [19,20,30,32,57] but both vertical and horizontal transfer occur in Burkholderia [13] their phylogenies displaying signs of the origin of Beta-rhizobia to some degree. Although it is widely reported [58], some reports have indicated that horizontal gene transfer (HGT) has not been common even within Alpharhizobia, as revealed by the nodA and nodC phylogenies of some Ensifer and Rhizobium symbionts [8,31]. This lack of clarity as to the origin and evolution of symbiotic genes (nodulation and nitrogen fixation genes) in rhizobium sensu lato means that their origin/evolution in the Beta-rhizobia are still a subject of much debate.

With regard to the origins of Beta-rhizobial symbiosis-related genes, Mimosa- (and other Mimosoideae)-nodulating Burkholderia strains are somewhat separate from Cupriavidus strains, such as C. taiwanensis and C. necator-like strains in their nodA and nodC phylogeny, but are still clearly related to them, and both are very different from Alpharhizobia, including those which can nodulate Mimosa [31,35,37,38]. However, B. tuberum STM678 and related South African strains which nodulate papilionoid legumes and which cannot nodulate Mimosa, appear to be more closely related to Alpha-rhizobia and are distant from other beta-rhizobia in terms of their nod genes [5,12-16,41]. Burkholderia species with nod genes that are related to B. tuberum STM678 include B. sprentiae, B. rhynchosiae, B. diworthii and B. dipogonis, as well as several other strains from papilionoid legumes from South Africa; the similarity in their nod genes suggests that they may have an origin common to some Alpha-rhizobia from papilionoid legumes, such as Bradyrhizobium. Indeed, it is clear that the nodulating Burkholderia have divided into two groups according to their very different nod genes: the mimosoid nodulators and the papilionoid nodulators. This is exemplified by the division of the species B. tuberum into the papilionoid-nodulating sv. papilionoidae (e.g. STM678) and the mimosoid-nodulating sv. mimosae (e.g. strain CCGE1002), depending on which type of nod gene they harbor [32]. Interestingly, the nod gene phylogeny of B. tuberum STM678 is in conflict with its nifH phylogeny, as it is grouped with all the other symbiotic (and free-dwelling diazotrophic) Burkholderia strains, which form a monophyletic group. This demonstrates that the nod genes evolved according to geographical and host factors, and are the basis of the symbiotype concept which states that it is the mobile nodule genes and not the core genome which determines host range in Rhizobia [59]. The genome sequence of B. phymatum STM815 has recently been published [38], and this shows that it has some similarities with C. taiwanensis in the structure of its symbiotic genes [55]. Moreover, the draft genome sequence of B. mimosarum strain LMG23256 and Cupriavidus sp. strain UYPR2.512 were announced, and these have demonstrated some different chromosome properties from Alpha-rhizobia [60,61]. To better understand the nodulating bacteria and their relationship with their geographical distribution, the project of sequencing several model LNB (legume-nodulating bacteria) genomes has been carried out to provide valuable insights into the genetic evolution of symbiotic nitrogen fixation [62].

With regard to transfer of symbiotic genes between the two classes of Rhizobium, [63] used complete genome sequences to study the origin of the rhizobial nodulation genes nodI, and showed that the entire nodII clade is included in the Burkholderiaceae DRA-ATPase/permease gene family, suggesting that the nodII genes originated from gene duplication in a lineage of the Betaproteobacterial class, and further suggests that Betaproteobacterial symbiosis genes were originally transferred to Alphaproteobacteria. However, the nodII sequences of B. tuberum STM678 were not included in the clade of β-rhizobial genes used in the study of [63], and yet there are discrepancies between the nodA and nodII phylogenies based on their partial sequences. In this study, we attempt to elucidate the evolutionary origin of nodulation and nitrogen-fixation genes by comparing structural maps of symbiotic regions between Alpha- and Beta-rhizobia, and we analyze phylogenies constructed using nodA and nifH sequences, and then examine the B. tuberculosis symbiosis genes based on complete genomes. The analysis showed that the nifH and nodA sequences of another B. tuberum strain (CCGE1002), which was isolated from Mimosa occidentalis in Mexico, and which belongs to the mimosoid symbiobair of B. tuberum, were grouped within clades of Beta-rhizobial genes. Indeed, [63] declared
that the nodA aa sequence of CCGE1002 clustered with Beta-rhizobial genes, which is consistent with relationships deduced using nodI sequences. However, the partial sequences are inaccurate in some aspects, and in the present study, we have found that there is little interaction between the two rhizobial clades, and we further suggest that the nod genes of Alpha- and (Mimoso-nodulating) Beta-rhizobia evolved independently, but we also lend support to the concept that lateral gene transfer has occurred in some clusters.

Methods

Data assembly and nucleotide sequence accession numbers

All complete genomes were accessed from NCBI (http://www.ncbi.nlm.nih.gov). Nodulation (nod) genes and nitrogen fixation genes (nif, fix and fdx) were screened from genomic sequences, and then selected for further phylogenetic analysis and for structural mapping of symbiotic regions. NCBI accession numbers for the 12 complete genomes in the present study are as follows: *Cupriavidus taiwanensis* LMG19424\(^1\) (NC_010528.1, NC_010530.1, NC_010529.1), *Burkholderia phymatum* STM815\(^2\) (NC_010622.1, NC_010623.1, NC_010625.1, NC_010627.1), *Burkholderia phenoliruptrix* BR3459a (NC_018695.1, NC_018672.1, NC_018696.1), *Burkholderia* sp. CCGE1002 (NC_011417.1, NC_011418.1, NC_014119.1, NC_014120.1), *Sinorhizobium medique* SMW 414 (NC_009636.1, NC_009620.1, NC_009621.1, NC_009622.1), *Rhizobium etli* sv. mimosae Mim-1 (NC_021905.1, NC_021906.1, NC_021907.1, NC_021910.1, NC_021908.1, NC_021909.1, NC_021911.1), *Rhizobium etli* CFN42 (NC_007761.1, NC_007762.1, NC_007763.1, NC_007764.1, NC_004041.1, NC_007765.1, NC_007766.1), *Rhizobium leguminosarum* sv. viciae 3841 (NC_000380.1, NC_000382.1, NC_000383.1, NC_000379.1, NC_000381.1, NC_000384.1, NC_000378.1), *Mesorhizobium loti* MAFF303099 (NC_002678.2, NC_002679.1, NC_002682.1), *Azorhizobium caulinodans* ORS571\(^3\) (NC_009937.1), *Bradyrhizobium japonicum* USDA6\(^4\) (NC_017249.1), *Methylbacterium nodulans* ORS2060 (NC_011894.1, NC_011892.1, NC_011887.1, NC_011893.1, NC_011895.1, NC_011888.1, NC_011889.1, NC_011890.1). The 10 genomes utilized which are annotated as follows: *C. taiwanensis* LMG19424\(^5\) [55], *Burkholderia phymatum* STM815\([38,86]\) *B. phenoliruptrix* BR3459a [64], *Burkholderia* sp. CCGE1002 [65], *Sinorhizobium medique* SMW419 [66], *Mesorhizobium loti* MAFF303099 [67], *Azorhizobium caulinodans* ORS571 [68], *Bradyrhizobium japonicum* USDA6\([9,97]\), *Rhizobium leguminosarum* sv. viciae 3841 [52], *R. etli* CFN 42 [70], *Methylbacterium nodulans* ORS 2060 (https://www.ncbi.nlm.nih.gov/genome?term=Methylbacterium++nodulans) and *R. etli* sv. mimosae Mim-1 (https://www.ncbi.nlm.nih.gov/genome?term=Rhizobium+etli++by+mimosas+str++.Mim1). The structural map of symbiotic genes for *Cupriavidus taiwanensis* LMG19424\(^4\) was then referenced. Finally, we confirmed that our structural map (using the other 8 strains) is consistent with symbiotic gene clusters from the NCBI database (www.ncbi.nlm.nih.gov/genbank).

Whole nodA and nifH gene sequences were downloaded from NCBI. The dataset of nodA genes contained 19 Alpha-rhizobial strains and 4 Beta-rhizobial strains. We also collected the nifH genes of 17 Alphaproteobacteria, 5 Betaproteobacteria and 11 other nitrogen fixing strains. These datasets were used for phylogenetic analysis using the distance method.

Phylogenetic profiling analysis

For the primary analysis, we searched for the largest dataset of nodA and nifH genes preferentially associated with *Rhizobia* using the distance method for phylogenetic profiling analysis. All available complete nodA and nifH sequences were aligned using the ClustalX program [71] with default parameters. Multi-alignments were visually corrected and used to draw phylogenetic trees using the genetic distance-based neighbor-joining algorithms of the MEGA 6.0 software [72] with partial deletion and an 80% coverage cut-off. Bootstrap analyses were performed using 1000 replicates for distance. The MEGA 6.0 model test was performed to select a model of nucleotide substitution, and the “best” model with the lowest Bayesian information criterion (BIC) score) was used for each gene. The neighbor-joining phylogenetic trees were visualized by using the TREEVIEW program (Page, 1996). For phylogenetic analysis of nodA, a dataset of 338 nucleotide sequence sites was analyzed using the NJ model, whereas 822 nucleotide sequence sites from 33 species were used in the phylogenetic analysis of nifH. *Rhodopseudomonas palustris* CGA009 was used as an out-group in the nifH tree. As only partial nodA sequences could be obtained from *B. tuberum* strains STM678\(^3\) and WSM4176, and from *B. sprentiae* strain WSM5005 and *B. dilworthii* strain WSM3556, phylogenetic trees based only on these partial nodA sequences were also constructed.

Structural map of symbiotic regions

In this study, the structural map of symbiotic genes of *C. taiwanensis* LMG19424\(^2\) are used [55] in an examination of 12 complete annotated genomes, during which the symbiotic regions and the entire nod and nif genes were screened. For each specific symbiotic gene, it was located in NCBI, and its genomic context, genomic regions, transcripts, size, and products were acquired. By comparing the reference strains and analyzing the particular gene location, a map of symbiotic regions was drawn using CoreDRAW X7\(^3\) software by inputting the size and location of each gene. All the symbiotic genes were essentially analyzed via the map of symbiotic regions.

Results

Characteristics of chromosomes and symbiotic genes in Beta-rhizobia

It was reported that the first complete genome sequence of a legume-nodulating Betaproteobacterium, *C. taiwanensis* LMG19424\(^2\), consists of two chromosomes and a large symbiotic plasmid. The genome displays an unexpected high similarity with the genome of the saprophytic bacterium *C. eutrophus* H16, and reveals a most compact (35 kb) symbiotic island (nod and nif) (Table 1) [55]. *Burkholderia phymatum* STM815\(^5\) and *B. phenoliruptrix* BR3459a each harbor two chromosomes (3.48/2.70 Mb and 4.15/2.71 Mb, respectively), while STM815\(^5\) contains two plasmids, pPHY01 (1.90 Mb) and pPHY02 (0.60 Mb), and strain BR3459a has a single large symbiotic plasmid, pSYMB3459 (0.79 Mb) [38,64]. The *Burkholderia* sp. CCGE1002 genome comprises three chromosomes (3.52, 2.59 and 1.28 Mb) and one plasmid (489 kb) [65]. Interestingly, there are different genomic sizes between the rhizobial clades, with those of Beta-rhizobia ranging from 6.48 to 8.68 Mb, while Alpha-rhizobia range from 5.37 to 9.21 Mb, with the genome of *Azorhizobium caulinodans* ORS571\(^3\) being the smallest of the strains in our study (Table 1). The organization of the symbiotic genes is also different within the two clades: Alpha-rhizobia either combine related genes on relatively mobile chromosomal islands in the case of *Mesorhizobium, Azorhizobium, Bradyrhizobium* and *Methylbacterium* or on symbiotic plasmids in the case of *Rhizobium* and *Ensifer* (*Sinorhizobium*). Beta-rhizobia also appear to exclusively contain an independent transmissible plasmid containing all their symbioses genes, although [61] have recently suggested that the nod
<table>
<thead>
<tr>
<th>Strains</th>
<th>Host Name</th>
<th>Isolation Country</th>
<th>Genome Size (Mb)</th>
<th>RefSeq Genes</th>
<th>Chromosomes</th>
<th>Plasmids</th>
<th>GC%</th>
<th>Strains</th>
<th>Genotype</th>
<th>Genome Size (Mb)</th>
<th>RefSeq Genes</th>
<th>Chromosomes</th>
<th>Plasmids</th>
<th>GC%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cupriavidus taiwanensis</td>
<td>LMG19424</td>
<td>Taiwan</td>
<td>6.48</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkholderia phymatum STM815</td>
<td>Mimosa pudica</td>
<td>China</td>
<td>8.68</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkholderia phymatum BR42</td>
<td>Mimosa spp</td>
<td>France, Guiana</td>
<td>8.68</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Shizosaccharomyces mitis SW419</td>
<td>Mimosa trifida</td>
<td>South America</td>
<td>4.15</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rhizobium etli</td>
<td>Mimosa affinis</td>
<td>Mexico</td>
<td>7.19</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Methylobacterium nodulans ORS 2060</td>
<td>Crotalaria sp.</td>
<td>Senegal</td>
<td>8.84</td>
<td>T</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
genes of papilionoid-nodulating Beta-rhizobia seem to be chromosomal, but it should be stressed that this has not yet been confirmed by further detailed analyses.

The symbiotic gene organizations were determined in different organisms (Figure 1), each exhibiting significant characteristics. Beta-rhizobia have 9 nodulation genes in common with an arrangement of nodUSAJICBD. In addition, Cupriavidus uniquely harbors other nod genes, such as nodQ, whereas only Burkholderia contains nodT and nodW. Next to the nod genes, the regulator nifA was tightly combined with nod genes and separated from other nif genes in Cupriavidus, which contrasts with Burkholderia where nifA is closely organized with

**Figure 1:** The Structural Map of symbiotic regions in Rhizobia. The datal length of genes, is available from GenBank Database (except Burkholderia tuberum STM678, Genome Institute (JGI) website). Genes are colored according to their name: Green(nod genes), Yellow(nif genes). Blue (fix genes); Beta-rhizobia shows in a, b, c, d and e, Alfa-rhizobia shows in f, g, h, i, j, k, l and m.
nifED(N)XQ, nifN replacing nifD in *Cupriavidus*. This demonstrates that the two genera of Beta-rhizobia have different nitrogen fixation mechanisms. *FixCBA* are present in the genomes of three Beta-rhizobial strains, but strain CCGE1002 lacks *fixX* and *nifWV* and the nitrogenase structural genes *nifHKD* are present in all four Beta-rhizobial strains examined. The genomes of *B. phymatum* STM815 and *B. phenoliruptrix* BR3459a share another two common copies of *nifZT* and *fixLB*. The organization of symbiosis genes between Alpha- and Beta-rhizobia is different: Beta-rhizobia have less complexity in their nodulation and nitrogen fixation gene structure, suggesting the possibility that they have evolved more recently than Alpha-rhizobia.

**The divergence in symbiosis genes between Alpha- and Beta-rhizobia**

We found that the nitrogenase regulator genes *nifEN(D), nifQ(N)* and the nitrogenase structural genes *nifHKD* are common between Alpha- and Beta-rhizobia. Moreover, *nodN* is present in Alpha-rhizobia and *nodQ* in Beta-rhizobia (ATP sulfurylase, APS kinase, respectively), and *nifN* and *nodP* are present together in *Mesorhizobium*, *Azorhizobium* and *Bradyrhizobium*. However, the remarkable discrepancies are that *fixNOQ*, derived from *fixLJ*-like genes in many Alpha-rhizobia, were found to be located on Beta-rhizobial chromosomes, instead of plasmids, and Beta-rhizobia only have *fixABCDXN* without any gene modifications.

From arrangements of *nod* genes in different *Rhizobia* we discovered that *nodABC*N is common in all, but that the host-specific *nod* genes are diverse. Some host-specific genes are common between some *Rhizobius* species e.g. *nodUS* are present in three of the Beta-rhizobial strains, which is consistent with their ability to nodulate the same host (*Mimosa pudica*), and these three genes are also shared with other *Rhizobius*, such as *Rhizobium* and *Bradyrhizobium*. Interestingly, *Sinorhizobium* and two *Burkholderia* strains also shared the *nodH* and *nodQ* genes, which may be linked with the fact that *Sinorhizobium* strains nodulate some Indian and Mexican *Mimosa* spp. [31,37]. We also found that *nodH* was located on the *Methyllobacterium* symbiosis island, which may be significant in terms of recent studies showing that related genera in the *Crotalariaeae* (*Aspalathus*, *Rafinia* and *Lebeckia*) are associated with two very different clades of bacteria i.e. *Burkholderia* and *Mesorhizobium/Rhizobium* [14-16]. The *nodZ* gene (Nod factor fucosyl transferase) was observed in *Mesorhizobium loti*, and was not found to be present in Beta-rhizobia. Although *Mesorhizobium* has not been isolated from *Mimosa pudica*, it has been reported to be isolated from *Pithecellobium hysterolefolium*, which is also within the *Mimosoideae* [9]. In addition, *R. leguminosarum* has two *nodT* genes, which it shares with three *Burkholderia* strains, which is interesting in consideration that *Rhizobium* strains are often isolated from *Mimosa pudica* e.g. *R. etli*, *R. tropici*, *R. leguminosarum* and *R. altiplani* [31,35,59,73,74]. Moreover, the *Mimosa*-nodulating *R. etli* *Mimosa* strain Mim-1 has more *nod* genes than *R. etli*, and *R. etli* *Mimosa* strains have a broader host range than *Mimosa* spp. [54]. In contrast, we also found that *nodU* and *nodW* of *Rhizobium* are organized as in *Azorhizobium caulindians*, but it is not yet known if the latter share hosts legumes with Beta-rhizobia. With respect to Alpha-rhizobia, *nodEHH* is common in *Rhizobium* and *Sinorhizobium*, which is corroborated by the fact that these two closely-related genera generally exhibit wide host ranges, albeit ones which rarely intersect. Finally, we determined that three Alpha-rhizobia genera, *Bradyrhizobium*, *Rhizobium* and *Azorhizobium* shared *nodU* and *nodS* (except for *R. leguminosarum* and *R. etli*), and in this context it is interesting that *Sinorhizobium* spp. nodulate with *Rhizobium* and *Azorhizobium* [58], but so far are not reported to do so with *Bradyrhizobium*. However, *Bradyrhizobium* housed the widest range of host-specific genes, sharing them with most other *Rhizobium* discovered in our study, so it is possible that strains of *Bradyrhizobium* that can nodulate *Sesbania* spp. will eventually be isolated. We can, therefore, conclude that rhizobial host range is related to the different host-specific genes organized either on sym-plasmids or on symbiosis islands, and that the wide host range of some rhizobial strains is due to their production of many kinds of Nod Factors [49] i.e., that broad host range *Rhizobia* harbor a wider array of host-specific genes than more specific and less promiscuous *Rhizobia*.

**Phylogenetic Analysis of *nodA* and *nifH* genes based on complete genomes**

Phylogenies using partial sequences (338 bp) of *nodA* (Figure 2) and entire sequences (822 bp) of *nifH* (Figure 3) from complete genomes of Alpha- and Beta-rhizobia were constructed (Table 1). The Beta-rhizobial strains examined formed two groups in the *nodA* dendrogram, one group clustering with the majority of the Beta-rhizobia, all of which are *Mimosa*-nodulators. Within this group, *Burkholderia* strains STM815 and *Burkholderia* phenoliruptrix BR3459a were close to each other with a similarity of 97.0%; *B. tuberum* CCGE1002 also clustered with them with 77.4-77.6% similarities, but the *nodA* sequence of *C. tainwansensis* LMG19424 was more distant from the afore mentioned three strains with only 70.3-74.0% similarities to them. The two *Cupriavidus* strains, *C. tainwansensis* and *Cupriavidus* sp. UYPR2.512, were 86.4% similar to each other. These Beta-rhizobia were all very distant from the Alpha-rhizobia with low similarities of 50.6%-70.6%, in contrast to the *Mimosa*-nodulators, the *nodA* sequences of the other group of legume-nodulating Beta-rhizobia which comprises papilionoid-nodulating *Burkholderia* strains, were very close to Alpha-rhizobia, with 86.4% similarity to *Methyllobacterium nodulans* ORS 2060. They were also close to *Mesorhizobium*, *Bradyrhizobium* and *Rhizobium*, with more than 73.5% similarities. There is 62.8-71% similarity between the two groups of nodulating *Burkholderia*.

Nitrogen-fixing organisms are not restricted to *Rhizobia*, and the ability to fix nitrogen is widely distributed in the bacterial and archaeal domains. The phylogeny based on *nifH* genes (33 genomes in total) reveals several separate clusters within *Rhizobius* (Figure 3); the first group consists of four *Mimosa*-nodulating Beta-rhizobial species and the relationships between them are similar to their *nodA* sequence phylogenies. Strains STM815 and BR3459a are closely related with 99.9% similarity, and CCGE1002 clustered with them with 83.3-83.5% similarity, but the *nifH* sequence of *C. tainwansensis* LMG19424 is more distant from the three *Burkholderia* strains with 76-81.8% similarity. A separate group of Betaproteobacteria is comprised of free-living and plant-associated *Burkholderia* species and these have 86% similarity with another Beta-rhizobial strain, the papilionoid nodulator *B. tuberum* STM678. The two groups of *Burkholderia* (i.e. the *Mimosa*-nodulators and the free-living diazotrophs plus *B. tuberum* STM678) have a closer relationship to each other in terms of *nifH* (73-85.7% similarity) than they have with Alpha-rhizobia. Interestingly, in spite of it belonging to the Beta-rhizobia, all the *Burkholderia* Beta-rhizobia had *nifH* sequences that were closer to the clade of free-living *Burkholderia* diazotrophs, with 85.7% similarity, than to *C. tainwansensis* (e.g. its *nifH* sequence similarity with CCGE1002 was 76%). The single Beta-rhizobial strain in our study which was isolated from papilionoid legumes, *B. tuberum* STM678 also had a *nifH* sequence which was closer to free-living *Burkholderia* strains (86-87.4% similarity) than to Alpha-rhizobia (68.2-80.2% similarity), which is in contrast to the relationship revealed by its *nodA* phylogeny. It is particularly interesting...
that the *nifH* of *B. tuberum* STM678 is closer to free-living strains than to Beta-rhizobia from *Mimosa* (86.3% and 80.8% similarity, respectively), and that it was quite close to the photosynthetic symbiont *Bradyrhizobium* sp. BTAi1 (which does not possess *nod* genes; [75] with 80.6% similarity. The two types of *Bradyrhizobium* Beta-rhizobia have divergent traits in *nifH*, as although they are still distant from most Alpha-rhizobia, the Beta-rhizobia from papilionoids were closer to Alpha-rhizobia (80.2% similarity) than they were to their *Mimosa*-nodulating cousins (74% similarity).

**Discussion**

From the arrangement of symbiotic genes in structural maps of the Beta-rhizobia, we suggest that they have evolved recently. Although they are generally considered to be symbionts of *Mimosa*, the Beta-rhizobia, especially the *Burkholderia* members, are versatile and the legume nodulation host range of this group has recently been extended into the papilionideae sub-family [12,14-16,40,41]. This group contains *B. tuberum*, the only nodulating *Burkholderia* species so far described that is common to both Africa and America, and which exists as two symbiovars that can nodulate either mimosoid (sv. mimosae in South America) or papilionoid (sv. papilionoideae in South Africa) legumes depending on which *nod* genes they possess. Accordingly, [61] have recently deduced that the symbiotic genes of the South African strain *B. tuberum* sv. papilionoideae STM678 are similar to Beta-rhizobia from other South African papilionideae (e.g. *Lebeckia* spp.), such as *B. dilworthii* and *B. spreitae*, which was to be expected in consideration of other reports [14-16,41,61]. Also reported that the *fixNOQP* and *fixGHIS* nitrogenase production and assembly genes are missing in all the *Burkholderia* strains that they examined, but after further analysis of the products of these genes we found their protein products were annotated on the published genome for *B. phymatum* STM815, but there were no relevant annotation gene names in GenBank. *Burkholderia tuberum* strains CCGE1002 and STM678 from mimosoid and papilionoid hosts, respectively, showed...
different symbiotic gene arrangements, but they belong to two different symbiovars, so this would be expected. Indeed, with rapidly increasing numbers of rhizobial strains having their genomes published, and with sequences becoming more accurate and reliable, some relationships within genera will inevitably change, as has already been shown for R. etli CFN42 [76]. Previous reports have discussed the origin of Rhizobia using partial sequences of symbiosis genes but the present study is the first to examine them in terms of whole genome sequences [13,30,31,77-80]. Using partial sequences, [63] supposed that the nodII genes originated from gene duplication in a lineage of the Betaproteobacterial family, and suggested that Betaproteobacteria symbiosis genes were transferred to Alphaproteobacteria, but made no conclusions about the evolutionary origin of symbiotic nitrogen fixation. After examining the ACC deaminase (acds) genes among Alpha- and Beta-rhizobia from the Cape Fynbos, recombination and horizontal transfer of nodulation genes (HGT) were suggested [14,81]. The acds gene is often located on transferable elements such as plasmids in Rhizobium and Sinorhizobium/Ensifer, and has been reported to be prone to HGT, most likely through symbiosis island and plasmid exchange, and is a common and important plant-beneficial property among Fynbos rhizobia. In the present study, we examined the divergence and mutual characteristics from the Cape Fynbos, recombination and horizontal transfer of the ACC deaminase (acds) and we conclude that although the common nodABC and nodI genes are present in all symbiotic Rhizobia, with the exception of some

Figure 3: nifH phylogeny. The tree was reconstructed by using NJ approach based on an 785-bp alignment matrix; Bootstrap support values were calculated by 1000 replications; Values along branches indicate bootstrap percentages; nifH sequences for published bacteria are available from GenBank.
photosynthetic bradyrhizobia [75], there is significant distance between the two clades (Alpha- and Beta-rhizobia) in their noda phylogeny, and also that some discrepancies could be detected only through the full genome and partial sequence analyses conducted in the present study. Previously, the noda gene of B. tumerum sv. papilionoideae STM678 was found to be more closely related to Methylobacterium nodulans (Alphaproteobacteria) based on its DNA and amino acid sequences, but its nifH sequence is closer to free-living Burkholderia (Betaproteobacteria) (Figure 3) [67,63]. As stated earlier, in contrast to STM678, which nodulates papilionoideae, in our complete genome study we found that B. tumerum sv. mimoseoideae strain CCGE1002 was slightly distant from the other two Mimosa-nodulating Burkholderia strains; this might relate to its reported ineffectiveness as a symbiont compared to (for example) B. phymatum STM815 [31].

Previous estimates as to the origin of Beta-rhizobia mainly stem from phylogenetic analyses of partial noda and nif genes sequences. For example, [30] postulated that a symbiotic noda sequence of Burkholderia is old and stable but that the horizontal gene transfer of nodulation genes likely occurred from Alpha- to Betaproteobacteria. nif genes are known to have higher similarity between Alpha- and Betaproteobacteria than noda genes, with 36.4%-77.4% similarity between the two clades and with free-living nitrogen fixation bacteria. Indeed, in our study nifH was shown to be particularly close within certain genera e.g. between two Burkholderia strains (99.9% similarity) and three Sinorhizobium strains (97.4%) which suggests strongly that the gene transfer often occurred within the same phylogenetic lineages. A unique origin of common noda genes and their horizontal transfer from Alpha- to Betaproteobacteria has been hypothesized i.e., the Alpha-rhizobial origin of nodulation genes [5,30,55,78,79,82]. This is supported by the papilionoideae-nodulating Beta-rhizobial strain B. tumerum STM678 which is close to Alpha-rhizobia in its partial noda sequence, but in the present study we found that Alpha- and Beta-rhizobia are distant from each other, with no distinct limit between the two clades, or even within each clade. In the noda phylogeny we found that Beta-rhizobia from Mimosa are very distant from the Alpha-rhizobia (including Alpha-rhizobia that nodulate Mimosa) with low similarities of 50.6%-70.6%, but we also found that the noda sequence of Burkholderia strain CCGE1002 (from Mimosa) was quite close to the Alpha-rhizobial strain Mesorhizobium loti MAF303099 (from Lotus spp.) and also to C. taiwanensis LMG19424 with similarities with both strains being around 70.3-70.6%. Moreover, two groups of Mimosa- and papilionoid nodulators in the Beta-rhizobia exhibited 62.8-71% similarity, but the noda sequences of two B. tumerum strains from different hosts (CCGE1002 and STM678) have a similarity of 71%, which is slightly closer than the distance between the two Mimosa-nodulators, CCGE1002 and C. taiwanensis (70.3%). Therefore, we suggest that noda genes generally evolved based on the lineages of their host rhizobial genera e.g. most noda sequences are closer to species/strains within their genera and distant from species/strains in other genera, even though they nodulate the same or a similar host legume. This phenomenon is also apparent in the nifH phylogeny of Rhizobia, as we found that Mimosa-nodulating Burkholderia Rhizobia are closer to free-living Burkholderia strains and to the papilionoid-nodulating B. tumerum STM678 than to Mimosa-nodulating Rhizobia in the other Beta-rhizobial genus, C. taiwanensis.

After examination of the noda and nifH phylogeny and their distance based on complete genomes, with respect to the common noda genes, nodaABCJ, we suggest that Beta- and Alpha-rhizobial symbiotic genes originated independently. Burkholderia Beta-rhizobia clearly have a common nif gene origin with free-living diazotrophic burkholderias, but we also found that Bradyrhizobium (particularly Bradyrhizobium sp. BTA11) has a lower similarity to Rhizobium, Sinorhizobium and Mesorhizobium than it does to free-living N-fixers. However, it should be strongly underlined that each genus has one nif origin, although the possibility of HGT between (and within) phylogenetic groups remain [63] suggested noda gene transfer from Beta- to Alphaproteobacteria, and a remarkable example of gene transfer known to have occurred in nature is that of the 500 kb symbiosis island in the chromosome of M. loti [83], which is transmissible and has insertion sequences. Therefore, we could suggest that the symbiotic genes evolved from gene duplication, and gene transfer occurred later within two or one clades resulting from the interaction between Rhizobia, legumes, and their environment. In support of this, a 410 kb symbiosis-relevant region of the Bradyrhizobium japonicum chromosome was suggested to be comprised of DNA fragments from different origins by comparing it with other free-living bacteria [76,84-88]. On the other hand, there are rhizobial strains where the chromosom-borne symbiosis genes are conserved and stable. Two legumes: Astragalus sinicus and chickpea (Cicer arietinum) exhibited conserved nodule genes in conditions of chromosomal diversity, and demonstrated that these two legumes which host Mesorhizobium species, such as M. ciceri and M. mediterraneum, have identical noda and nodC sequences in spite of their diversity in geographical origins, host and chromosomal backgrounds. Furthermore, [89-91] found that symbiotic genes (noda, nodC, nodH and nifH) within Robinia pseudoacacia mesorhizobia from Poland and Japan were highly similar, suggesting that the symbiotic apparatus evolved under strong host plant constraints.

The environment, biotic and abiotic conditions may strongly influence the selection of bacterial strains or species that are able to live in the soil. In addition, host selective pressures and lateral gene transfer in the soil are the main mechanisms that shape the genetic structure of symbiotic microorganisms [14-16,31,79,92] and these confound the use of 16S rDNA phylogenies to describe symbiotic bacterial relationships with their hosts and each other, thus making study of the evolutionary history of symbiosis difficult. On the other hand, nif and nod genes are selectively lost, duplicated, and horizontally transferred [93]. Even those located on genomic islands in Mesorhizobium and Bradyrhizobium may be transferred across divergent chromosomal lineages; duplication of nif genes in several rhizobial types as shown by sequencing of multiple copies has demonstrated that they are identical in many cases [94-96]. This is also corroborated by our study; both nif and nod gene products have well defined functions, and so it might be speculated that the symbiotic region of B. japonicum was located originally on a plasmid similar to the Sym plasmids of S. meliloti and then became part of the chromosome by integration at one stage during evolution. Alternatively, the symbiotic plasmids of S. meliloti (and other Rhizobia) may have evolved by excision of a chromosomal region [97].

Nitrogen fixation is undoubtedly an ancient innovation that is not only crucial for extant life, but played a critical role during the early expansion of microbial life as abiotic nitrogen sources became scarce. Hence the idea that nitrogen fixation had originated in the last common ancestor of the three domains (bacteria, archaea and eukaryotes), at least as inferred by the presence of nitrogenase in the two major prokaryote domains. In addition, there is a lack of nitrogenase homologs in eukaryotes and most prokaryotes, which may be because of gene loss after the atmosphere became oxic [98]. Nitrogenase genes may share a common evolutionary history [99]; in order to survive in ancient surroundings, bacteria possibly inherited their nod genes directly.
Rhizobial diversity provides a pool of symbiotic bacteria to be selected by compatible host legumes; a single, few, or many bacterial cells may fit individual plant variability and also survive different environmental conditions fluctuating over time and space [100], and they will evolve in response to any selection pressures they may exert on each other. The two partners in the symbiosis also become mutually influenced. Taken together, this can result in symbiosis genes being lost or acquired by HGT. Some rhizobial genera, such as the genus *Bradyrhizobium* has a very wide host range exemplified by its ability to nodulate legumes from all three legume sub-families (Papilionoideae, Mimosoideae, and Caesalpinioideae), whereas others, such as *R. leguminosarum* and *Neorhizobium galegae* have a very narrow host range.

In conclusion, we strongly support the contention that vertical transmission played an important role in the spread and maintenance of symbiotic genes in Beta-rhizobia, as demonstrated by nodII [63], but HGT also played a significant role [8,70,101-102] as a result of the loss and acquisition of symbiosis genes under the pressure of the environment. Although the ancestor of symbiotic genes and whether their transfer was from Alpha- to Beta-rhizobia or vice versa is still controversial, with our increasing knowledge about Beta-rhizobial diversity, we and others have established from their nodA gene homology that there are two main centers of Beta-rhizobia: those associated with Mimosoideae in Brazil (S. America) and those with Papilionoideae in the Fynbos (S. Africa). The two nodulating *Burkholderia* groups, mimosoid- and papilionoid nodulators, are not entirely independent on each other; however, as the South American and African plates were integrated within the Gondwana supercontinent until 200 Mya [102]. It is possible that a common ancestor of *Burkholderia* was already present in soils on Gondwana, and when the supercontinent broke up separate populations of these bacteria were established in the newly-formed continents of South America and Africa. *Burkholderia* is known to be at least 50 Mya old [30,33], and may be much older, and is likely to have been present (in acidic soils) when the legumes first emerged approx. 60 Mya, although we cannot be sure if they or a similarly ancient nodulating Alpha-rhizobial type, such as *Bradyrhizobium* were the first microsymbionts that they encountered [102]. Later in the evolutionary history of the legumes (33 Mya; [33,102]) and after the mainly non-nodulating Burkholderias encountered these emerging plants (e.g. those in the genus *Mimosa*) as they colonized and speciated within the acidic soils of the seasonally dry highland regions of central South America (e.g. the Cerrado). In parallel, in South Africa, the papilionoid tribes associated with the South African Burkholderias, Crotalariae and Podalyrieae, arose 44–46 Mya [103], and these plants also presumably encountered the acid-loving *Burkholderia* as they colonized and speciated within the acidic soils of the Fynbos. The main differences between the nodulating Burkholderias in South Africa and those of their South African cousins is that the former have very different nod genes to local Alpha-rhizobia [61], whereas the latter nodulate a wide range of Fynbos legume genera which are often also capable of nodulating with Alpha-rhizobia, such as *Mesorhizobium* [14-16]. This most likely explains why the nod genes of the South African Burkholderias are so similar to those of Alpha-rhizobia, but it does not tell us which came first and who transferred them to WHO. On the other hand, the South American mimosoideae-nodulating Burkholderias appear to have emerged quite separately from their local Alpha-rhizobial populations, and thus it is difficult to hypothesise from whence they obtained their nodulation genes, but it is possible that their ancestors nodulated a now extinct group of legumes which preceded the mimosoids. Certainly, further and wider sampling of symbionts from other legume sub-families, tribes and genera in South America (and South Africa) will assist in helping us answer these questions.

Acknowledgements

This research was financially supported by the National Natural Science Foundation of China (grant Nos. 30970005 and 30370051), Key Bioengineering Discipline of Hebei Province (No.1050-5003002), the Public service sectors (agriculture) research and special funds (grant No. 201103050-07).

References


