Phylogenetic distribution and evolutionary dynamics of nod and T3SS genes in the genus Bradyrhizobium

Albin Teulet, Djamel Gully, Zoe Rouy, Alicia Camuel, Ralf Koebnik, Eric Giraud, Florent Lassalle

MICROBIAL GENOMICS

RESEARCH ARTICLE

Teulet *et al.*, *Microbial Genomics* 2020;6

DOI 10.1099/mgen.0.000407

Phylogenetic distribution and evolutionary dynamics of *nod* and T3SS genes in the genus *Bradyrhizobium*

Albin Teulet[1], Djamel Gully[1], Zoe Rouy[2], Alicia Camuel[1], Ralf Koebnik[3], Eric Giraud[1,*,†] and Florent Lassalle[4,5†,‡]

Abstract

Bradyrhizobium are abundant soil bacteria and the major symbiont of legumes. The recent availability of *Bradyrhizobium* genome sequences provides a large source of information for analysis of symbiotic traits. In this study, we investigated the evolutionary dynamics of the nodulation genes (*nod*) and their relationship with the genes encoding type III secretion systems (T3SS) and their effectors among bradyrhizobia. Based on the comparative analysis of 146 *Bradyrhizobium* genome sequences, we identified six different types of T3SS gene clusters. The two predominant cluster types are designated RhcIa and RhcIb and both belong to the RhcI-T3SS family previously described in other rhizobia. They are found in 92/146 strains, most of them also containing *nod* genes. RhcIa and RhcIb gene clusters differ in the genes they carry: while the translocon-encoding gene *nopX* is systematically found in strains containing RhcIb, the *nopE* and *nopH* genes are specifically conserved in strains containing RhcIa, suggesting that these last two genes might functionally substitute *nopX* and play a role related to effector translocation. Phylogenetic analysis suggests that bradyrhizobia simultaneously gained *nod* and RhcI-T3SS gene clusters via horizontal transfer or subsequent vertical inheritance of a symbiotic island containing both. Sequence similarity searches for known Nop effector proteins in bradyrhizobial proteomes revealed the absence of a so-called core effectome, i.e. that no effector is conserved among all *Bradyrhizobium* strains. However, NopM and SUMO proteases were found to be the main effector families, being represented in the majority of the genus. This study indicates that bradyrhizobial T3SSs might play a more significant symbiotic role than previously thought and provides new candidates among T3SS structural proteins and effectors for future functional investigations.

DATA SUMMARY

All the sequence genomes used in this study were downloaded from the National Center for Biotechnology Information (NCBI) RefSeq or GenBank databases. These genomes, which were annotated *de novo* by the MicroScope pipeline, are also publicly available from the MicroScope platform (http://www.genoscope.cns.fr/agc/microscope). Online Supporting Data is available from the Figshare repository at the following http://dx.doi.org/10.6084/m9.figshare.11378772, http://dx.doi.org/10.6084/m9.figshare.12202202 and http://dx.doi.org/10.6084/m9.figshare.12191103.

INTRODUCTION

A recent analysis of microbial communities across the globe indicated that rhizobia of the genus *Bradyrhizobium* are the most ubiquitous and abundant bacteria in soils [1]. They thus appear at the top of the list of the 'most wanted' bacteria to

Received 20 December 2019; Accepted 26 June 2020; Published 12 August 2020

Author affiliations: [1]IRD, Laboratoire des Symbioses Tropicales et Méditerranéennes (LSTM), UMR IRD/SupAgro/INRA/Université de Montpellier/CIRAD, TA-A82/J – Campus de Baillarguet 34398, Montpellier cedex 5, France; [2]LABGeM, Génomique Métabolique, CEA, Genoscope, Institut François Jacob, Université d'Évry, Université Paris-Saclay, CNRS, Evry, France; [3]IRD, CIRAD, Université de Montpellier, IPME, Montpellier, France; [4]Department of Infectious Disease Epidemiology. Imperial College London, St Mary's Hospital Campus, Praed Street, London W2 1NY, UK; [5]Pathogen and Microbes Program, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.
***Correspondence:** Eric Giraud, eric.giraud@ird.fr

Keywords: symbiosis; legume; rhizobium; effector; nodulation; type III secretion system.

Abbreviations: Bp, base pair; CDS, protein-coding sequence; DTL, horizontal transfer and loss; ErnA, effector required for nodulation A; ETI, effector-triggered immunity; GBDP, genome-to-genome blast distance phylogeny; HGT, horizontal gene transfer; Mbp, mega base pairs; MIIA, metal ion-inducible autocleavage; ML, maximum likelihood; MSLA, multilocus sequence analysis; NCBI, national center for biotechnology information; NEL, novel E3 ligase; NF, Nod factors; Nop, nodulation outer protein; SUMO, small ubiquitin-like modifier; T3E, type III effector; T3SS, type III secretion system.

‡Present address: Pathogen and Microbes Program, Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK.
†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. Supplementary material is available with the online version of this article.

investigate for a better understanding of the contribution of microbes in the functioning of soil ecosystems. Additionally, these bacteria are the most widely used in agriculture due to their ability to interact symbiotically with legumes of high agronomic importance, including staple crops such as soybean, peanut and cowpea. This symbiosis leads to the formation a new organ, the nodule, generally found on the root system, in which the bacteria fix dinitrogen for the benefit of the plant and receive a carbon source in return. Bradyrhizobia therefore play a very important role in the preservation of the environment by limiting the use of nitrogen fertilizers for the promotion of the growth of legume crops and by contributing to nitrogen enrichment of terrestrial ecosystems.

Not all bradyrhizobia are able to nodulate legumes, as some strains isolated from soils or other environments lack the symbiotic *nod* genes that are necessary for the synthesis of Nod factors (NFs) that govern nodulation [2, 3]. Upon recognition by the plant, this key signal molecule triggers the genetic program leading to nodule formation and its intracellular infection by the bacteria [4]. However, some bradyrhizobia have evolved alternative strategies to interact symbiotically with legumes in the absence of the NF signal. This is the case for some photosynthetic *Bradyrhizobium* strains (including ORS278 and BTAi1), which lack the canonical *nodABC* genes necessary for the synthesis of NFs but are able to induce the formation of nitrogen-fixing nodules on the root and the stem of some legume species of the genus *Aeschynomene* [5]. The bacterial signals necessary for the establishment of this NF-independent symbiosis remain to be discovered.

The *nod* genes, which remain the main determinants conferring the ability to nodulate plants to rhizobia, can be transferred from strain to strain by horizontal gene transfer (HGT) [6, 7]. While these symbiotic genes are typically harboured by a plasmid in other rhizobial genera, these genes are generally found in a chromosomal region of *Bradyrhizobium* strains, called the symbiotic island. In addition, this region generally harbours the *nif* and *fix* genes, which are necessary for the formation and functioning of the nitrogenase complex. This region can be easily recognized, as it is characterized by lower G+C content, different codon usage for the identified protein-coding sequences (CDSs) and the presence of numerous transposases and insertion sequences that strongly promote genomic rearrangements in this region. The high plasticity of the symbiotic island incidentally promotes the selection of symbiotically adaptive variants in bradyrhizobial populations [8]. The presence of genes that code for the type III secretion system (T3SS) has also been frequently reported within this region of *Bradyrhizobium* genomes [9–12].

The T3SS, which was initially found in pathogenic bacteria, is a complex secretory machinery composed of more than 20 conserved proteins that form a syringe-like structure and enable bacteria to deliver type III effector proteins (T3Es) into eukaryotic cells [13]. These T3Es are highly diverse in their structure and mode of action and play a crucial role in bacteria–host interaction by modulating key host cell functions, in most cases related to the suppression of host immunity [14]. Functional T3SSs encoded by 'Rhizobium-conserved' (*rhc*) gene clusters have been reported in several rhizobial species [15, 16]. However, only a few rhizobial T3Es, including those called Nop (nodulation outer proteins), have been characterized and their mode of action generally remains to be determined [16]. The rhizobial T3SS machinery contains most of the core components that were described in the T3SS of pathogenic bacteria, but the genetic organization of the *rhc* cluster and its transcriptional regulation are different from its pathogenic counterpart [17]. Notably, the expression of both *rhc* and *nop* genes is controlled by the regulator TtsI, which is itself under the control of the *nod* gene regulator, NodD [18, 19].

While the virulence of plant pathogenic bacteria largely depends on the functioning of the T3SS [14], the importance of rhizobial T3SS for symbiosis remains unclear. Indeed, the presence of the T3SS among nodulating rhizobia is not a rule, indicating that the T3SS is dispensable for the establishment of the symbiosis. Furthermore, the T3SS can play a neutral, a positive or a negative role, depending on the host plant [16, 20]. On one hand, some of the secreted T3Es may repress the plant immune system, favouring the establishment of the infection and the nodulation. On the other hand, the same effectors may, in other plant contexts, be recognized by receptor proteins [resistant (R) proteins] activating an effector-triggered immunity (ETI), which blocks the

Impact Statement

The nitrogen-fixing symbiosis between legume plants and their bacterial partners (rhizobia) is of immense biological, ecological and agronomic importance, as it is a major contributor to global nitrogen cycling. It has been accepted for a long time that the establishment of the rhizobium–legume symbiosis always relies on the bacterial synthesis of Nod factors, which trigger nodule organogenesis and infection programs. However, recent studies revealed that some rhizobia from the genus *Bradyrhizobium* are able to nodulate some legume species in the absence of Nod factor, relying instead on a type III secretion system (T3SS), a common virulence factor of pathogenic bacteria. In this study, we explored and compared publicly available *Bradyrhizobium* genome sequences to obtain new insights into the significance of the T3SS in the symbiotic adaptation of bradyrhizobia. Our data indicate that the nodulation genes *nod* and T3SS genes have been co-inherited both vertically and horizontally in many instances, and are present together in almost all nodulating *Bradyrhizobium* strains, suggesting an important role of the T3SS in the symbiotic process. Furthermore, we highlighted several genetic determinants that are probably required for the T3SS function and identified newly putative effectors that may contribute to the symbiotic properties of bradyrhizobia.

infection and renders the interaction incompatible [21, 22]. More recently, it has been reported that the T3SS can play a more prominent symbiotic role by directly activating the nodulation in the absence of NFs [23]. This NF-independent, T3SS-dependent symbiotic process was initially discovered between *Bradyrhizobium elkanii* USDA61 and *Glycine max* cv. *Enrei*. This function of the T3SS could be more widespread than originally thought, given that a large and number of diverse non-photosynthetic bradyrhizobia strains were shown to nodulate *Aeschynomene indica* thanks to their T3SS [24]. A recent study conducted on the strain ORS3257 using *A. indica* as host plant revealed that this T3SS-dependent symbiotic process relies on a cocktail of at least five T3Es playing distinct and complementary roles in nodule organogenesis, infection and repression of host immune functions [25]. Among these T3Es, a new one named ErnA (effector required for nodulation A) was shown to act as a key factor to form nodules.

A large number of *Bradyrhizobium* genomes have been sequenced, including those from strains isolated from nodules of several legume species, but also from various other ecological niches. In this study, we took advantage of this wealth of genome sequences to obtain new insights into the significance of the T3SS in the symbiotic adaptation of bradyrhizobia. We examined the phylogenetic distribution of T3SS and *nod* genes as well as some *nop* genes reported to have a symbiotic role to answer to the following questions. What is the distribution of the *nod* and T3SS genes among strains isolated from nodules or other environments? What are the different events of acquisition or loss of the T3SS that occurred during the evolutionary history of bradyrhizobia? Is the acquisition of the T3SS systematically associated with the acquisition of *nod* genes through lateral transfer of the symbiotic island? How congruent is the evolution of the *nod* and T3SS genes? How widespread are the T3Es reported to have a symbiotic role?

METHODS

Phylogenomic analysis of all *Bradyrhizobium* genomes used in this study

A custom set of 155 genomes was gathered covering 28 different *Bradyrhizobium* species, as well as selected genomes from the genera *Nitrobacter* and *Rhodopseudomonas* as outgroups. The genomes listed in Table S1 (available in the online version of this article) were downloaded from the National Center for Biotechnology Information (NCBI) RefSeq or GenBank databases and used as input for the bioinformatic pipeline Pantagruel [26] to build a phylogenomic database. In short, coding sequences (CDSs) and the corresponding protein sequences were extracted from the RefSeq or GenBank annotation and then clustered into homologous gene families using MMSeqs2 [27]. Homologues were aligned with CLUSTAL Omega [28] and reverse-translated into CDS alignment with PAL2NAL [29]. Four hundred and fifty-three single-copy core gene families were selected based on their presence in at least 153 of the 155 genomes, thus allowing for rare losses or incomplete genome sequences. CDS alignments for these families were concatenated, resulting in 378951 aligned nucleotide positions. From this core alignment, a maximum-likelihood (ML) tree was inferred using RAxML (v8.2.4) [30] under the GTRCATX model, and branch lengths were then refined under the GTRGAMMAX model, while branch supports were estimated from 200 rapid bootstrap trees. Bayesian samples of gene trees were inferred from CDS alignments with MrBayes (v 3.2.6) [31] using a Metropolis-coupled Monte Carlo Markov chain (MCMCMC) approach, with two independent sets of four chains (one cold, three heated) running for 2000000 generations and sampled every 500; the presented trees are majority-rule consensus trees, discarding the first 25% of the chain as burn-in. The consensus of Bayesian gene trees for Rhc-T3SS cluster gene families is available as part of the online Supporting Data at the Figshare data repository (http://dx.doi.org/10.6084/m9.figshare.11378772).

Prediction of *nod* and T3SS clusters

All genome sequences used in this study (Table S1) were downloaded using the MicroScope platform (http://www.genoscope.cns.fr/agc/microscope). Genome annotations were performed *de novo* using the MicroScope pipeline [32]. A screen of *nod* and T3SS clusters in each genome was performed by searching the key genes *nodA* and *rhcN*. Reference CDSs were extracted from the *Bradyrhizobium diazoefficiens* USDA110^T genome (blr2025 and blr1816, respectively) and searched against the PkGDB genome database using the BLASTN algorithm as implemented on the MicroScope platform with the following parameters: query coverage ≥70%, identity ≥40%. The syntenic organizations of regions surrounding *nodA* and *rhcN* homologues found in bradyrhizobial genomes were then compared with each other using the MaGe genomic visualization system available on the MicroScope platform.

Prediction of known Nops repertoires

The T3Es known to be secreted through the T3SS or to affect the symbiotic interaction were identified from the literature (Table S2) and then searched in all strains using the BLASTP algorithm as implemented on the MicroScope platform with the same similarity constraints as described above. The absence of *nopX* and *nopE*/*nopH* genes in RhcIa- and RhcIb-containing genomes, respectively, were confirmed by TBLASTN search of the NopX proteins from USDA61 strain and NopE/NopH proteins from USDA110^T. The search of genes containing known functional domains conserved in some T3Es was performed by interrogating the precomputed InterProScan annotations hosted by the MicroScope platform for each strain genome, using the InterPro identifier of the functional domain as a query.

Phylogenetic analysis

The concatenated NodABC and RhcC2NV amino acid sequences were aligned with CLUSTALW using MEGA X software [33]. Phylogenetic trees were constructed from the resulting alignments using the ML method as implemented in MEGA

X software. Branch support was also estimated using MEGA X by performing bootstrap analyses of 500 replicates. Final editing was carried out using Figtree (http://tree.bio.ed.ac.uk/software/figtree/) and further editing was performed with graphics software. The same procedure was used to construct the SctN phylogenetic tree using the SctN homologues from other T3SS-containing bacteria (*Rhizobiales*, *Enterobacteriales*, *Aeromonadales*, *Pseudomonadales*, *Rhodobacterales*, *Burkholderiales*, *Oceanospirillales*, *Xanthomonadales* and *Chlamydiales*) that are available in the NCBI database.

Prediction of RhcI-T3SS clusters gain and loss events

Evolutionary analyses of the presence/absence of RhcIa- and RhcIb-T3SS clusters were performed using the program Gloome [34]. We used the Mixture model, which allows one to study differences in both gain and loss probabilities. The expectations and probabilities for both gain and loss events were estimated using stochastic mapping and a predicted event was considered as valid if the probability was at least 0.5.

Estimation of individual gene family evolution scenarios and co-evolution

We used gene tree/species tree reconciliation to obtain a finer estimation of the scenarios of events that marked the evolutionary history of individual genes within the RhcI-T3SS clusters and *nod* operons. This approach integrates the information from gene phylogenies to infer events of gene duplication, horizontal transfer and loss (DTL) that occurred over the diversification of the gene family [35]. In short, we used the software GeneRax to estimate ML scenarios as well as to make a Bayesian sample of close-to-optimal scenarios [36]. We applied this procedure to the gene families *nodA*, *nodB*, *nodC*, *rhcC2*, *rhcD*, *rhcN*, *rhcV*, *nopC*, *nopM*, *nopL* and *nopT*, as well as to a sample of control gene families from the bradyrhizobial pangenome; we used the core-genome-based tree of 155 bradyrhizobia and outgroups described above as a reference species tree. Bayesian scenario samples were used to compute co-evolution scores between gene families, using an approach adapted from Lassalle *et al.* [26] and detailed further in File S1. ML and Bayesian samples of scenarios, summaries of the inferred transfer events and of the co-evolution analysis are available as part of online Supporting Data at the Figshare data repository (http://dx.doi.org/10.6084/m9.figshare.12191103). Species tree inference and reconciliation computations were performed on the MRC Cloud Infrastructure for Microbial Bioinformatics (MRC CLIMB) cloud-based computing servers [37].

RESULTS AND DISCUSSION

General features of *Bradyrhizobium* genomes and phylogenomic analysis of the genus

In this study we considered 146 bradyrhizobial genomes and 9 genomes from related genera that were available in the public databases. A large proportion (106 out of 146) of the brayrhizobia strains sequenced were isolated from the nodules of various legumes species. The plants from which the most isolates were sequenced are: (i) *G. max* (36 strains), (ii) *Vigna unguiculata* (12 strains), (iii) *Aeschynomene* spp. (9 strains), (iv) *Phaseolus* spp. (9 strains) and (v) *Lupinus* spp. (7 strains). The other strains originated from various environments (different types of soils, water, biofilm, plant or animal tissue) (Table S1). The majority of bradyrhizobial genomes have sizes ranging from 7 to 10 Mb, with an average of 8.5 Mb. However, a few strains (CCH10-C7, CCH4-A6, NFR13 and SG-6C) have smaller genomes with sizes between 4 and 6 Mb. Among the 33 strains with a complete or near-complete (i.e. <10 contigs) genome sequence, four strains (BTAi1, NK6, DOA9, G22) contain one or several plasmids. Remarkably, strain DOA9 is the only *Bradyrhizobium* strain carrying a symbiotic plasmid harbouring *nod, nif* and T3SS genes [38] and strain NK6 contains four plasmids [39].

To better understand the evolutionary origins of *nod* and T3SS genes in bradyrhizobia, we first aimed at establishing a robust phylogeny of these organisms, which may later be used as a reference to infer evolutionary scenarios for the gain and loss of symbiotic genes. We built a phylogenetic tree based on the concatenated core genome (453 single-copy genes), which we rooted using several representatives from the neighbouring genera *Nitrobacter* and *Rhodopseudomonas* as an outgroup. This analysis yielded six phylogroups within *Bradyrhizobium*, each supported by 100% bootstrap values (Fig. 1). Phylogroup 1 clustered with more than half of the studied genomes. A large proportion of the bradyrhizobia belonging to this phylogroup were isolated from soybean, which is the most important legume crop in the world, thus explaining the over-representation of this phylogroup. Phylogroup 1 gathers strains from the species *Bradyrhizobium japonicum* and *B. diazoefficiens*, as well as strains assigned to various other taxa: *Bradyrhizobium yuanmingense* CCBAU10071^T, *Bradyrhizobium vignae* LMG28791^T, *Bradyrhizobium forestalis* INPA54B^T, *Bradyrhizobium shewense* ERR11^T, *Bradyrhizobium sacchari* BR10280^T, *Bradyrhizobium stylosanthis* BR446^T, *Bradyrhizobium manausense* BR3351^T, *Bradyrhizobium neotropicale* BR10247^T and *Bradyrhizobium centrolobii* BR10245^T. Phylogroup 2 clustered six photosynthetic bradyrhizobia strains, including *Bradyrhizobium oligotrophicum* S58^T, as well as the non-photosynthetic strain STM3843. All seven strains in phylogroup 2 were reported to nodulate some *Aeschynomene* species in an NF-independent manner; they lack *nod* genes, with the exception of strain ORS285, which is able to use both NF-dependent and NF-independent strategies, depending on the host plant [5, 40, 41]. Phylogroup 3 only included two strains, ARR65 and Tv2a-2, which were isolated from nodules of *Stylosanthes viscosa* and *Tachigali versicolor*, respectively [42, 43]. These two strains notably differ from their closest relatives in phylogroup 2 through the presence of *nod* genes. Phylogroup 4 gathered 29 strains from various species, including the type strains *B. elkanii* USDA76^T, *Bradyrhizobium pachyrhizi* PAC48^T, *Bradyrhizobium brasilense* UFLA03-321^T, *Bradyrhizobium viridifuturi* SEMIA690^T, *Bradyrhizobium tropiciagri* SEMIA6148^T, *Bradyrhizobium mercantii* SEMIA6399^T, and *Bradyrhizobium*

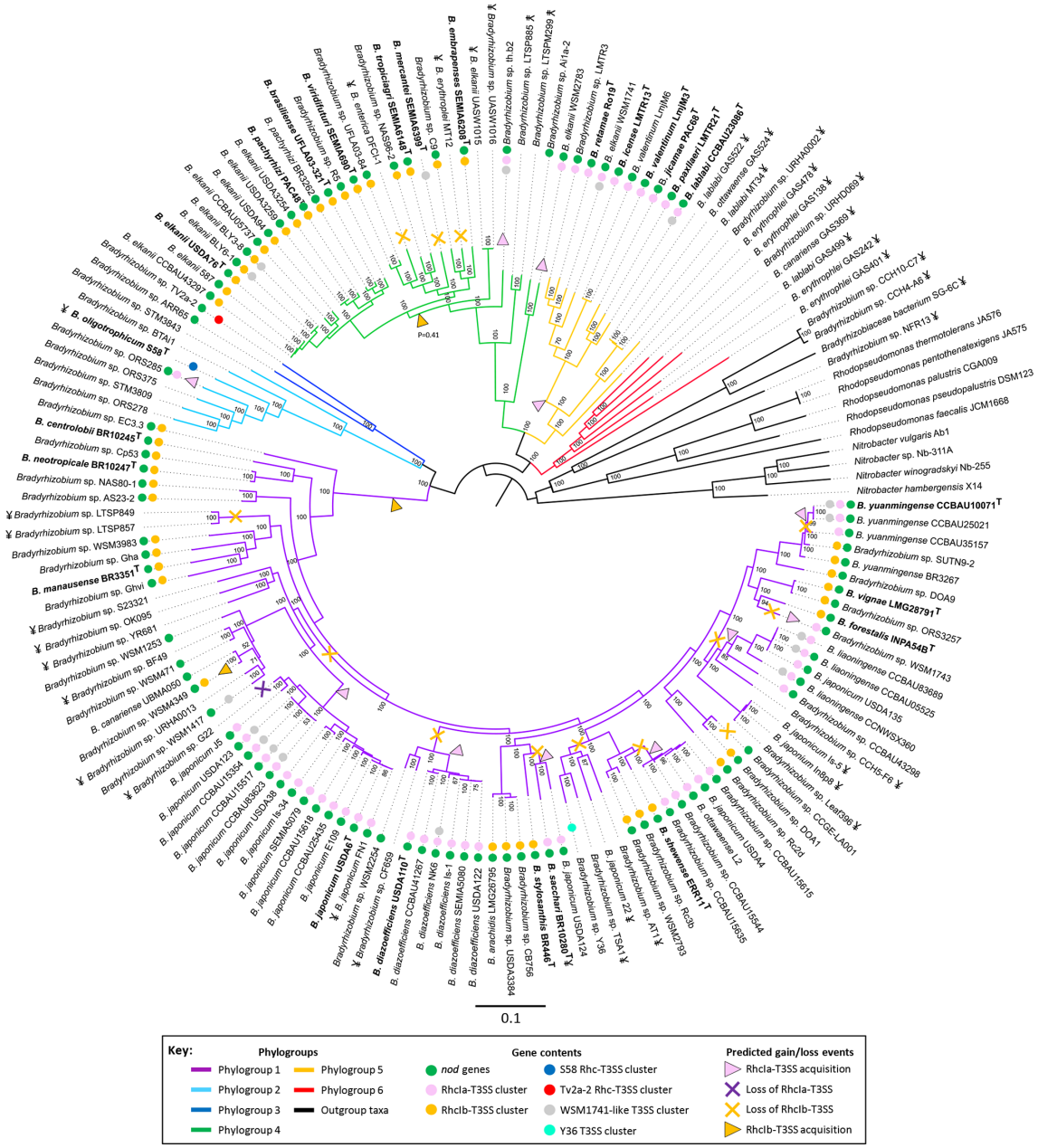

Fig. 1. Phylogenomic tree and RhcI-T3SS cluster gain–loss prediction inferred for the genus *Bradyrhizobium* and its close relatives. The phylogeny was inferred from 453 single-copy core genes for 146 bradyrhizobia strains, and several species of the neighbouring genera *Rhodopseudomonas* and *Nitrobacter*. Node support values are indicated at each branch. The phylogroup to which each strain belongs is indicated by the colour of the branch according to the key provided. The presence of *nod* or T3SS clusters in each strain is shown by colour circles referenced in the key. Probable RhcI-T3SS cluster acquisition and loss events are represented by a colour-coded arrowhead and a cross, respectively, according to the key provided. For all these predicted events, the probability was at least 0.5 except for the gain event of phylogroup 4 for which the probability was 0.41. The scale bar indicates the number of amino acid changes per site. The symbol ¥ denotes the strains that were not isolated from root nodules.

embrapense SEMIA6208^T. Similarly, phylogroup 5 gathered 14 strains from several species, such as *Bradyrhizobium retamae* Ro19^T, *Bradyrhizobium icense* LMTR13 ^{T}, *Bradyrhizobium valentinum* LmjM3^T, *Bradyrhizobium jicamae* PAC68^T, *Bradyrhizobium paxllaeri* LMTR21^T and *Bradyrhizobium lablabi* CCBAU23086^T. Phylogroup 6 contained six strains, all of which were isolated from soils. Finally, the four strains with the smallest genome (CCH10-C7, CCH4-A6, NFR13 and SG-6C) formed a clade that branches with representatives of the genus *Rhodopseudomonas*. These groupings were all supported by whole genome-to-genome distances (Fig. S1).

According to this phylogenetic analysis, the taxonomic classification of several strains needs to be revised. Firstly, the four small-genome strains branched with *Rhodopseudomonas*, which indicates that they do not belong to the genus *Bradyrhizobium* but rather to another genus within the family *Bradyrhizobiaceae*. With this in mind, the genome size of these four strain genomes appears to fit well within the range of sizes observed for *Rhodospeudomonas* (3.7–5.4 Mbp) and *Nitrobacter* (3.6–5.0 Mbp) genomes (NCBI genome database, last accessed September 2019), contradicting the idea that they had undergone some recent genome reduction event. Within the genus *Bradyrhizobium*, several strains do not belong to the species to which they were assigned. This is the case for several strains of *B. japonicum* (USDA4; USDA124; USDA135; 22; is5, in8P8) and *B. elkanii* (UASWS1015; WSM2783; WSM1747), which were found outside of the clade containing the type strain of their respective species. In phylogroup 6, strain GAS499 was wrongly assigned to *B. lablabi*, the type strain of which belongs to another phylogroup. The other strains in this phylogroup are *Bradyrhizobium erythroplei* or *Bradyrhizobium canariense*; the genome sequences of their type strains were not yet available and are not covered by this study, but have representatives in other phylogroups, suggesting that a revision is needed for the taxonomy of either phylogroup 6 strains or the distant strains sharing their name. The incorrect naming of some strains was based on their 16S rRNA and/or the internally transcribed spacer (ITS) sequences, while for the strains belonging to the phylogroup 6 and for WSM2783 and WSM1747 the criteria for their taxonomic classification remain elusive [44–46]. It is now recognized that these two genetic markers are not sufficiently discriminating to distinguish *Bradyrhizobium* strains due to limited sequence polymorphisms in this group. It was therefore proposed to use an alternative molecular method based on multilocus sequence analysis (MSLA) for taxonomic assignments [47]. Our phylogenetic analysis implements a genome-based taxonomy approach, which can be seen as a 'super MSLA', which is not limited to a few partial sequences of housekeeping genes but comprises several hundred conserved genes, as recommended by taxonomic authorities [48, 49].

Notably, during the compilation of this paper, two phylogenomic analyses focusing on the genus *Bradyrhizobium* were published, but they did not include exactly the same taxa as our study. One study was also based on the core genome [50], whereas the other was based on the genome-to-genome BLAST distance phylogeny (GBDP) [51]. The phylogenetic trees of these two studies show a very similar topology to that presented here (Fig. 1), which confirms the robustness of these different phylogenomic analyses. In our tree, the four 'non-*Bradyrhizobium*' small-genome strains form a clade that groups together with the *Rhodopseudomonas* and *Nitrobacter* genomes we used as outgroups. In the study by Avontuur *et al.* [50], which included a wider diversity of outgroup, CCH10-C7 was found to be associated with the genus *Afipia*, while NFR13 belonged to *Tardiphaga*. We explored the phylogenetic position of these strains using the Type Strain Genome Server (https://tygs.dsmz.de/) based on GBDP distances, and this suggested that the strains CCH4-16, CCH10-C7 and SG-6C belong to the genus *Afipia*. A detailed taxonomic study is needed to delineate a formal taxonomic name for these mistyped strains, but this is beyond the scope of this study.

Presence of multiple and diverse T3SS clusters among bradyrhizobia

The *rhcN* gene was used in BLAST searches to survey for the presence of a T3SS among our bradyrhizobial genome collection. This gene encoding an ATPase, which is essential for the function of the secretion machinery, is one of the most conserved core T3SS genes and has often been used for phylogenetic analyses [52, 53].

Among the 146 bradyrhizobial genomes analysed, 96 contain at least 1 *rhcN* homologue and among them, 16 contain 1 additional homologue, and 2 additional homologues are only found in the genome of strain WSM1741 (Fig. 2). This single-gene search analysis suggests that a large proportion (66%) of the bradyrhizobia strains may contains a T3SS gene cluster, and that some of them may even carry several such clusters – a situation already observed in *Ensifer fredii* strains NGR234, HH103 and USDA257, which each contain two T3SS clusters borne on separate plasmids [54–56].

Previous phylogenetic analyses based on various T3SS core proteins, including RhcN, indicated that the T3SS evolved into seven distinct families that spread between Gram-negative bacteria by HGT [52, 57]. So far, the characterized T3SSs from the different rhizobia all clustered in the same family, Rhc, and were further sub-divided into four sub-groups (annotated as α-RhcI to α-Rhc-III and ß-Rhc, which includes the T3SSs from the β-rhizobium *Cupriavidus taiwanensis*) [17]. To better understand how the different bradyrhizobial RhcN homologues are distributed among these different families, an ML tree was constructed including representatives of all the families and sub-groups that were previously defined. The great majority of bradyrhizobial RhcN homologues clustered in sub-group I of the Rhc family (α-RhcI), forming a large clade (annotated A on Fig. 2), which also contains homologues from *Ensifer* and *Mesorhizobium*. Other RhcN homologues seem to be related to the α-RhcI group, but are different from previously defined subgroups [17]; they branch at the base of the α-RhcI group and form a distinct clade (annotated B), grouping genes found in the strains *B. oligotrophicum* S58^T and *Bradyrhizobium* sp. Tv2a-2. Genes from a group of 17 strains, including the type strains *Bradyrhizobium mercantei* SEMIA6399^T, *Bradyrhizobium lablali* CCBAU23086^T and *B. yuanmingense* CCBAU10071^T, form a separate clade (annotated C) that branched next to the Ysc family, which includes proteins from animal pathogenic and symbiotic bacteria. The gene from strain Y36 (annotated D) clearly falls within the Ysc family. When a genome carries several RhcN homologues, at least one copy systematically belongs to the α-RhcI group, with the other copies being associated with clade C, except for strains LmjM3, LmjM6 and WSM1741, whose extra RhcN genes form a clade (annotated E) that is basal to

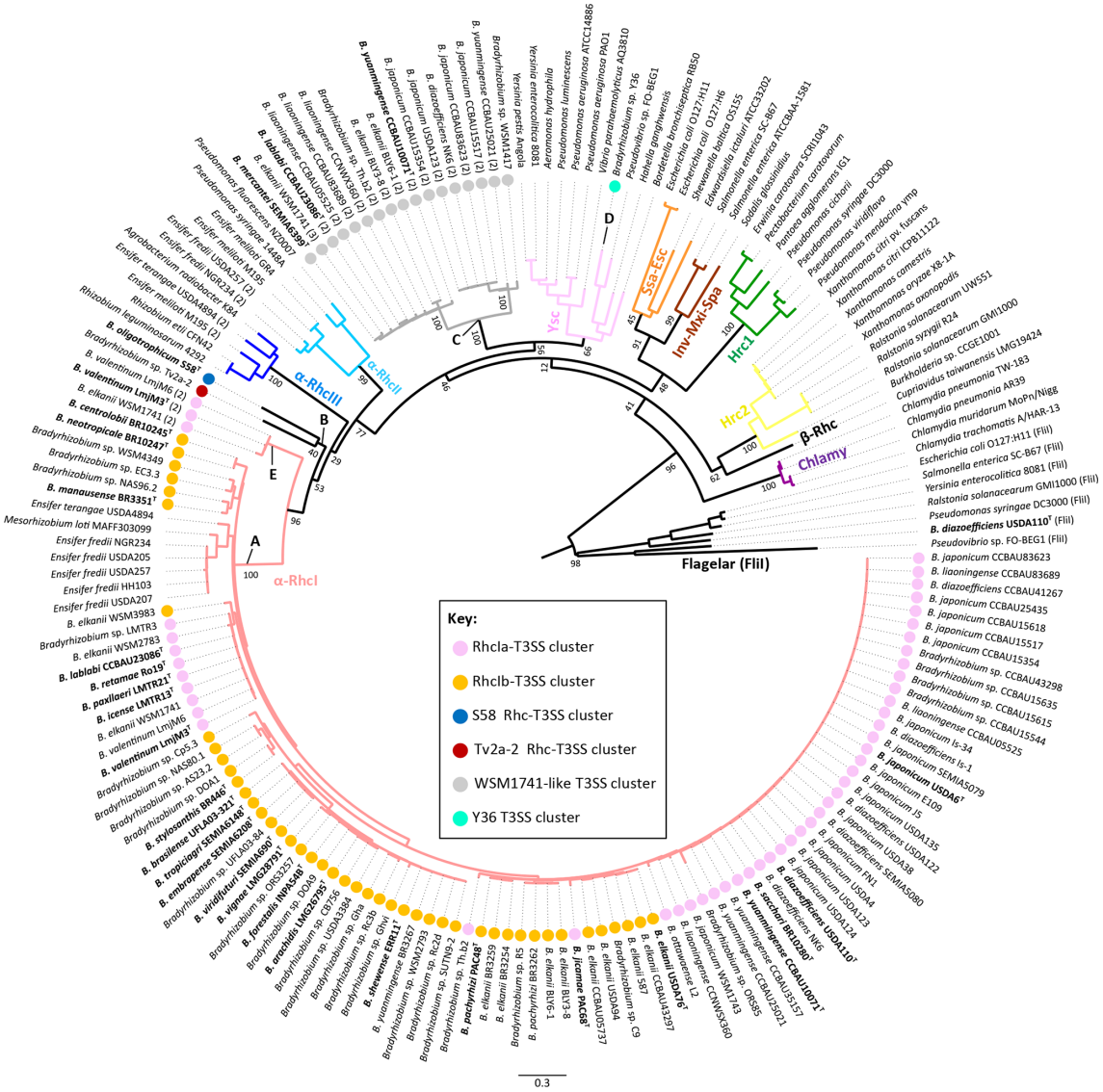

Fig. 2. ML phylogenetic tree of the SctN/RhcN proteins from bradyrhizobia and several representatives of symbiotic, pathogenic and plant-associated bacteria. The ML analysis was performed using mega X and 500 bootstrap replicates. The different SctN families defined according to Troisfontaine and Cornelis [52] are differently coloured. For the *Bradyrhizobium* strains, the different coloured circles (see the key provided) correspond to the distinct genetic organization of the T3SS clusters to which the sctN gene belongs (see Fig. 3). The scale bar indicates the number of amino acid changes per site. The numbers indicated within the parentheses are used to discriminate between the different SctN/RhcN homologues identified in the same strain.

the α-RhcI family. To distinguish the RhcN homologues that do not fall in one of the Rhc families that have already been defined, i.e. those that are related to the Ysc family (clade C and D), we chose to name them SctN, with respect to the unified nomenclature for T3SS proteins (with the Sct prefix referring to secretion) [58].

To further refine the classification of the T3SS systems in bradyrhizobia, we analysed the genetic organization of the gene clusters that encode the T3SS machinery. In total, we identified six different arrangements of T3SS genes among the bradyrhizobial strains (Fig. 3). The majority of T3SS gene clusters belong to the α-RhcI family, with two sub-families named RhcIa and RhcIb, which correspond to the gene clusters already described for *B. diazoefficiens* USDA110^{T} and *B. elkanii* USDA61, respectively (Fig. 3) [9, 59]. In both sub-families, all genes encoding the core components of the T3SS machinery are arranged in three distinct regions (Fig. 3). The main central region (I) contains in the following order the *rhcC1* gene in one direction and the *nopB*, *rhcJ*, *nolU*, *nolV*, *rhcN*, *rhcZ*, *rhcQ*, *rhcR*, *rhcS*, *rhcT* and *rhcU* genes, which are transcribed in the opposite direction. This central region is perfectly conserved in the two sub-families, whereas two additional conserved regions at the extremities of the gene clusters (region II and region III) differ in the details

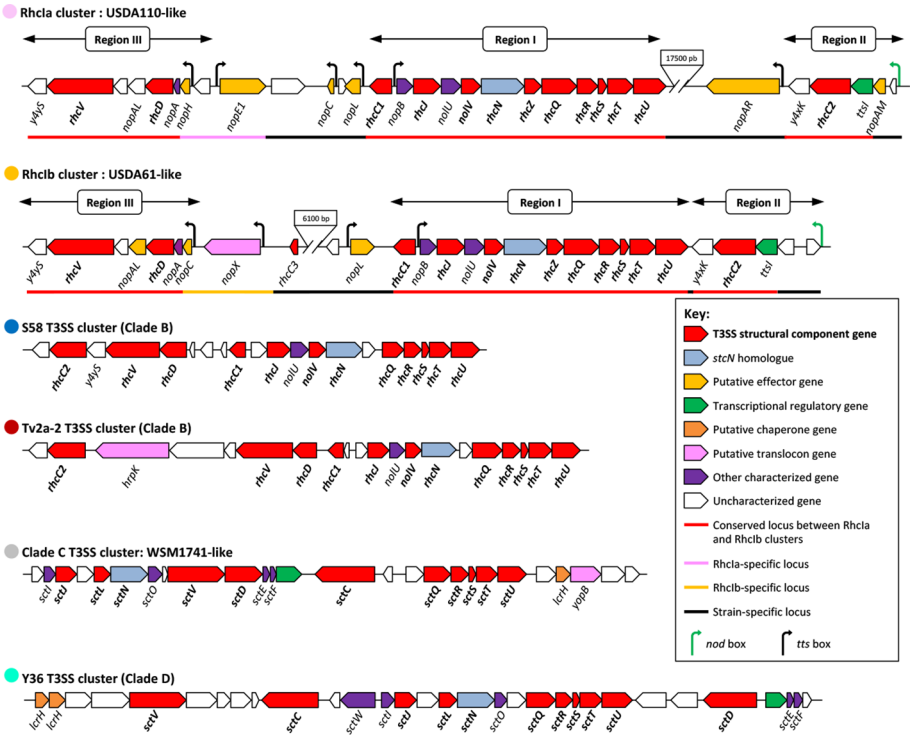

Fig. 3. Genetic organization of the different T3SS gene clusters identified in the genus *Bradyrhizobium*. The six different T3SS gene clusters identified are designated by the same colour circles used in Fig. 2 and the name of a representative strain in which the corresponding T3SS cluster is identified. Open reading frames are represented by coloured arrows according to the key provided and showing the transcription sense. The *nod* and *tts* boxes are represented by green and black arrows, respectively. The two-way arrows above the RhcIa and RhcIb-T3SS clusters represent the three genetic regions defined by Tampakaki [17]. The most common names previously used to characterize the T3SS gene cluster in rhizobia were applied for the T3SS gene cluster found to belong to the Rhc family (according to Fig. 2), while for both the WSM1741-like and Y36-like T3SS clusters, which do not belong to the Rhc family, the *sct* gene nomenclature was used.

of their gene contents. Notably, in region II, the predicted CDSs between the *nod* box and the *ttsI* regulator differ and are strain-specific. In strain USDA110^T, the RhcIa cluster starts with a CDS (bsl1845) encoding a protein of unknown function, followed by the putative effector gene *nopAM* [60], while in USDA61, the RhcIb clusters are predicted to contain two divergently transcribed CDSs encoding hypothetical proteins. Region III starts with two different small genes encoding secreted proteins of unknown function: *nopC* (RhcIb cluster) or *nopH* (RhcIa cluster) [61, 62]. Notably, the region upstream of the RhcIb cluster codes for the NopX protein, which is homologous to the putative translocon HrpF in *Xanthomonas* [63]; the corresponding gene is absent in all the strains with a RhcIa cluster, such as strain USDA110^T.

One of these two gene arrangements is systematically found in all the strains containing a *rhcN* gene of the α-RhcI family, but some minor variations occur among strains: (i) gene predictions in the region upstream of the *ttsI* gene are often ambiguous, sometimes with a prediction of a CDS showing some similarities with the *nopAM* gene identified in strain USDA110^T; (ii) a rearrangement of the region II provoked by transposase insertions and leading to a separation of the *ttsI* and *rhcC2* genes (observed only in ORS285 strain); and (iii) the occasional presence of an intergenic region of around 650 bp between *rhcS* and *rhcT*. Interestingly, this 'intergenic' region is predicted to contain a *tts* box, suggesting that the genes of such a region I might be transcribed in two distinct operons.

Notably, the T3SS gene clusters from clade E share the same genetic organization as RhcIa and should therefore be considered to be a bona fide RhcI-T3SS cluster. In contrast, T3SS gene clusters from strains S58^T and Tv2a-2 (clade B) both have a different genetic organization than RhcI types and should therefore constitute new types of Rhc cluster (Figs 2 and 3). In particular, the T3SS cluster in the Tv2a-2 strain harbours a homologue of the *hrpK* gene, which is found in some non-canonical T3SS gene clusters from *Ensifer* strains and phytopathogenic bacteria, where it is thought to encode a translocon component [13, 17].

Analysis of the regions surrounding the *rhcN*/*sctN* homologues of strain Y36 (clade D) or of the strains belonging to clade C revealed additional types of T3SS gene clusters, suggesting a distinct origin of these unconventional T3SSs (Figs 2 and 3). Determining whether these atypical T3SS clusters are functional and can affect the symbiotic properties of the strains requires further investigation.

A striking difference of the T3SS clusters from clades B to D in comparison to the RhcI clusters is the absence of the *ttsI* regulatory gene. The fact that *ttsI* is under the control of the master regulator NodD [18, 19, 64] supports the hypothesis that the ancestor of the RhcI-T3SS cluster was formed in the presence of the *nod* gene cluster, and that later propagation of the T3SS would rely on its co-transfer with *nod* genes, most probably via a symbiotic island gathering these two gene clusters.

The *nod* and T3SS genes share a common history in the genus *Bradyrhizobium*

To test this hypothesis, we scrutinized the genomes of the 146 *Bradyrhizobium* strains for the presence of *nodABC* genes using BLASTN and observed that 100 strains contained these canonical *nod* genes (Fig. 1). Strikingly, more than 90% of them (92 out of the 100) also contain the RhcI-T3SS cluster. In fact, all the strains containing a RhcI-T3SS cluster contain the *nodABC* genes, except for strain *Bradyrhizobium liaoningense* CCBAU83689 (Fig. 1). We then examined the genome of the 33 strains with a complete or near-complete sequence to see whether the *nod* and RhcI-T3SS gene clusters are co-localized or not. For the 12 strains containing a RhcI-T3SS cluster, this cluster is systematically found inside the symbiotic island containing the *nod* and *nif* genes (Table S3). Conversely, for the three strains (S58^T, NK6 and WSM1417) with an atypical T3SS cluster, this cluster is not found on the symbiotic island but on another horizontal acquired island for S58^T and WSM1417 strains, or on a plasmid for NK6 strain (Table S3).

To verify that the RhcI-T3SS and *nod* gene clusters follow the same evolutionary history, we compared the phylogenetic trees based on the concatenation of the NodA, NodB and NodC proteins or on the concatenation of the RhcC2, RhcN and RhcV proteins — each corresponding gene belonging to one of the three regions composing the T3SS cluster. Several incongruences occur between the trees, which likely result from ancient recombination events between symbiotic islands, as may have occurred after symbiotic islands from different origins were acquired independently and co-inhabited in the same rhizobial genome. This hypothesis of recombination among clusters is further supported by the evolutionary scenarios that we estimated for component genes of these clusters. Inferred scenarios show that most HGT events were inferred uniquely in each of the *nod* or *rhc* gene trees, indicating that independent importation of foreign sequences at each gene locus occurred regularly over the evolution of the bradyrhizobial symbiotic islands (online Supporting Data, http://dx.doi.org/10.6084/m9.figshare.12191103). The (rare) observation of two RhcI-T3SS clusters with divergent *rhcN* genes within the extant genomes of strains WSM1741, LmjM3 and LmjM6 finally proves the possibility of a co-habitation being required for such a recombination event (Fig. 2).

Nevertheless, the phylogenetic trees for NodABC and RhcC2VN present many matching clades grouping corresponding sequences from the same strains (Fig. 4), supporting the hypothesis of co-evolution between these two gene clusters, at least in recent times.

We thus aimed to quantify the co-evolution of single-gene families within and between the *nod* or RhcI-T3SS gene clusters. To do so, we estimated evolution scenarios describing the events of vertical and horizontal dissemination that marked the diversification of key genes present in these gene clusters. We then computed pairwise co-evolution scores based on how much each pair of gene families had shared events in their scenarios. As expected, genes show the strongest co-evolution association with those that are co-located with them in modern genomes: genes within the *nod* operon (i.e. *nodA*, *nodB* and *nodC*) and those within the Rhc-T3SS gene cluster (i.e. *rhcC2*, *rhcD*, *rhcN*, *rhcV* and *nopC*) each form modules of genes with significantly strong co-evolution history. Co-evolution between the *nod* and RhcI-T3SS modules was also strong and significant, even though slightly lower than within these modules (Fig. S3, http://dx.doi.org/10.6084/m9.figshare.12191103). This is evident when considering all shared evolutionary events, including both vertical and horizontal modes of co-evolution (Fig. S3 a, c), or when considering only co-transfer events, i.e. only events of joint acquisitions via HGT (Fig. S3b, d); this suggest that these genes were often co-acquired and subsequently conserved together in genomes.

Taken together, these data support the hypothesis that bradyrhizobia gained *nod* and RhcI-T3SS gene clusters simultaneously through the acquisition of a symbiotic island, while the atypical T3SS clusters were gained independently. The subsequent vertical inheritance and preservation of the RhcI-T3SS in most of the strains containing *nod* genes – notwithstanding a few recombination events – also indicates that this secretory machinery plays an important role in their symbiotic lifestyle.

T3SS was gained and lost multiple times in the genus *Bradyrhizobium*

The distribution of the RhcIa and RhcIb clusters is not random within the genus *Bradyrhizobium*, with the cluster types occurring in homogeneous patches within clades of the core genome tree (Fig. 1). The distribution of the RhcI-T3SS cluster types is not uniform with regard to their host genome phylogroup: (i) strains belonging to phylogroups 6 and 2 (with the exception of the strain ORS285), as well as the two members of phylogroup 3 have no RhcI cluster; (ii) phylogroup 5 strains possessing a T3SS all have the RhcIa cluster type; (iii) all phylogroup 4 strains possessing a T3SS, except strain th.b2, have the RhcIb cluster; (iv) RhcIa and RhcIb clusters are irregularly distributed within phylogroup 1.

To determine the origins of this heterogeneous distribution of RhcIa and RhcIb cluster types, we first explored the evolution scenarios computed for each component gene. However, each gene's sample of scenarios presented a set of many probable events, distributed over the reference *Bradyrhizobium* tree, and while statistical agreement between those sets can be computed (see the co-evolution scores above), it is difficult

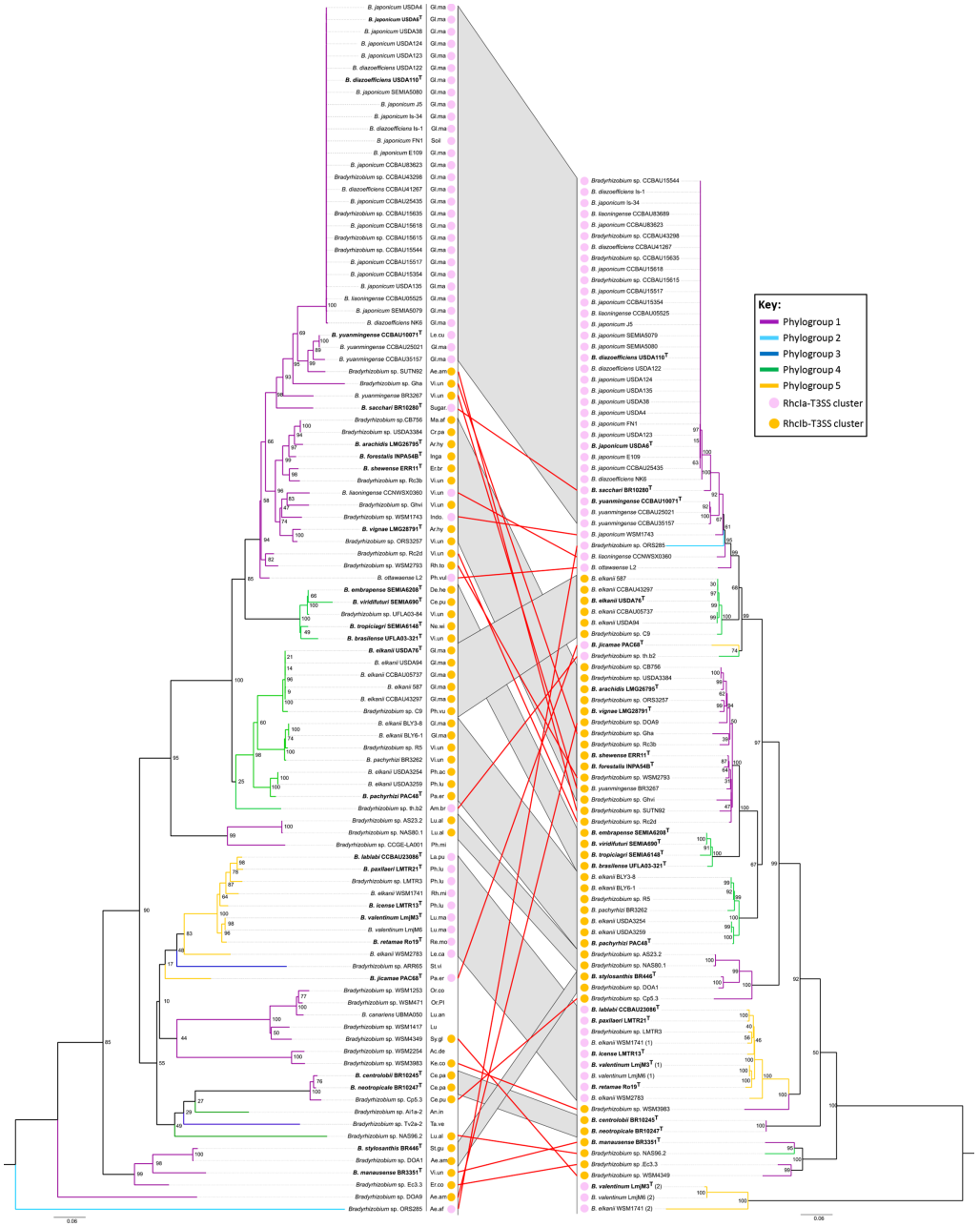

Fig. 4. Comparison of phylogenetic trees for concatenated NodABC (left) and RhcC2NV (right) proteins from *Bradyrhizobium*. The two ML phylogenetic trees were constructed using MEGA X and 500 bootstrap replicates. *Bradyrhizobium* sp. UFLA03-84 is not included in the RhcI-T3SS tree, since this strain lacks the region III containing the *rhcV* gene. The scale bar represents the number of amino acid substitutions per site. The colouring scheme for tree branches and circles is indicated in the key provided. The phylogenetic congruence between the two trees is represented by grey bands linking the different clades, while incongruence is represented by red lines. The environment for isolation of the strains is indicated in the field next to the strain names on the left-hand tree *Acacia dealbata* (Ac.de), *Aeschynomene afraspera* (Ae.af), *Aeschynomene americana* (Ae.am), *Amphicarpaea bracteata* (Am.br), *Andira inermis* (An.in), *Arachis hypogaea* (Ar.hy), *Centrolobium paraense* (Ce.pa), *Centrosema pubescens* (Ce.pu), *Crotalaria paulina* (Cr.pa), *Desmodium heterocarpon* (De.he), *Erythrina brucei* (Er.br), *Erythrina costaricensis* (Er.co), *G. max* (Gl.ma), *Indigofera* sp. (Indo.), *Inga* sp. (In.sp), *Kennedia coccinea* (Ke.co), *Lablab purpureus* (La.pu), *Leobordea carinata* (Le.ca), *Lespedeza cuneata* (Le.cu), *Lupinus albescens* (Lu.al), *Lupinus angustifolius* (Lu.an), *Lupinus maria-josephae* (Lu.ma), *Lupinus* sp. (Lu.sp), *Macrotyloma africanicus* (Ma.af), *Neonotonia wighti* (Ne.wi), *Ornithopus compressus* (Or.co), *Ornithopus pinnatus* (Or.pi), *Pachyrhizus erosus* (Pa.er), *Phaseolus acutifolius* (Ph.ac), *Phaseolus lunatus* (Ph.lu), *Phaseolus microcarpus* (Ph.mi), *Phaseolus vulgaris* (Ph.vu), *Retama monosperma* (Re.mo), *Rhynchosia minima* (Rh.mi), *Rhynchosia totta* (Rh.to), Sugar cane (Su.ca), *Stylosanthes guianensis* (St.gu), Stylosanthes viscosa (St.vi), *Syrmatium glabrum* (Sy.gl), *Tachigali versicolor* (Ta.ve) and *Vigna unguiculata* (Vi.un).

to summarize that information in a consensus scenario for the whole gene cluster (see high-probability co-events in the online Supporting Data; http://dx.doi.org/10.6084/m9.figshare.12191103). For greater ease of interpretation, we thus chose to complement our analysis with a simpler method of inference of the history of events of gain and loss of Rhc clusters. We used the program Gloome, which relies on the profile of presence and absence of the whole RhcI clusters in genomes and performs a stochastic mapping of events over the reference tree, to estimate a single scenario of evolution for each of the RhcI cluster types along the *Bradyrhizobium* tree [34]. According to the scenario estimated with Gloome, an equal number of gain and loss events (14 each) occurred across the genus tree, but with different frequencies depending on the phylogroup.

A few seminal events of independent acquisition of the RhcI-T3SS cluster define the main structure of their distribution: the acquisition of the RhcIa cluster in the ancestor of the phylogroup 1 clade corresponding to the species *B. japonicum* and *B. diazoefficiens*; another acquisition of the RhcIa cluster in the ancestor of the phylogroup 5 clade regrouping *B. lablabi*, *B. paxllaeri*, *B. jicamae*, *B. valentinum*, *B. icense* and *B. retamae*; and the acquisition of the RhcIb cluster in the ancestor of the phylogroup 4 clade regrouping the species *B. elkanii, Bradyrhizobium pachirizi*, *Bradyrhizobium brazilense*, *B. viridifuturi*, *Bradyrhizobium tropicagri*, *B. mercanti* and *Bradyrhizobium embrapenses*. These seminal horizontal transfer events were followed by the conservation of the acquired cluster, leaving patterns of locally homogeneous cluster distribution, suggesting that strong purifying selection acted on these acquired genes. Another acquisition of the RhcIb cluster was inferred at the root of the phylogroup 1, but this event was later followed by independent losses of the cluster: it was lost by the ancestor of the group formed by *B. japonicum* and unclassified *Bradyrhizobium* strains such as WSM1417 and WSM1253. The ancestor of *B. japonicum* later regained a T3SS cluster of the type RhcIa. Other loss events are associated with gains of the other cluster type on the same branch of the tree, and six such replacements of the RhcIb cluster by the RhcIa type occurred, all within phylogroup 1 (Fig. 1).

This scenario of repeated replacements is reflected in the phylogeny of the RhcC2VN proteins, where clades of Rhc proteins associated with the RhcIa organization recurrently emerge from a larger clade of proteins associated mostly with the RhcIb type (Fig. 4). The RhcIa clade at the top of Fig. 4 regroups many almost identical sequences, suggesting a more recent spread of this cluster type, with repeated events of horizontal dissemination into distinct clades of phylogroup 1. Interestingly, most of these RhcIa phylogroup 1 strains were isolated from *G. max* nodules. It is very likely that a strong selective pressure, probably driven by the interaction with the plant, has favoured the exchange of RhcI-T3SS clusters within this phylogroup. The selective advantage of switching to a RhcIa cluster type may be specific to phylogroup 1, as *B. elkanii* strains of phylogroup 4 also isolated from *G. max* nodules all harbour a RhcIb cluster.

The tree based on concatenated RhcC2VN proteins presents the evolutionary point of view of the genes encoded in the cluster (Fig. 4), which shows that the RhcIa type evolved twice from a clade associated with the RhcIb type. These events correspond to a gain by the ancestor of *B. japonicum* (phylogroup 1), and another gain in the ancestor of phylogroup 5 Rhc+ strains (Fig. 1). The lineages of Rhc proteins involved in these two gain events are not related, with those found in phylogroup 5 branching deeply in the RhcC2VN phylogenetic tree. Interestingly, the diversity of proteins within this deep clade mirrors that of their host genomes. This suggests that in phylogroup 5, the *rhc* genes (or at least large segments of the locus including the *rhcC2*, *rhcV* and *rhcN* genes) were transmitted vertically after the gain of the ancestral RhcIa type cluster. This unique gain event in phylogroup 5 is confirmed by reconciliation-based scenarios (see ML reconciliations in the Online Supporting Data; http://dx.doi.org/10.6084/m9.figshare.12191103), but those scenarios additionally implied that it was followed by occasional events of homologous recombination of individual *rhc* genes among phylogroup 5 strains.

To evaluate whether these repeated changes of RhcI type were regularly associated with such a reshuffling of the alleles of core *rhc* genes, we then studied the individual trees of genes located in different regions of the RhcI locus: *rhcV* in region III, *rhcN* in region I and *rhcC2* in region II (Figs 3 and S2). The evolutionary histories of these genes appear to be mostly congruent, as they share the majority of ancient splits, including those leading to the emergence of the clades that are homogeneous for their RhcI type. This strongly indicates that the whole RhcI-T3SS cluster locus mostly evolved as a linked genetic unit, being transferred from one genomic background to another as a single fragment. In addition, it shows that the diversification of *rhc* genes is not associated with switches in RhcI cluster type; it is only the genetic organization of the locus that changed when changes of RhcI type occurred. Each switch of cluster type could in fact have been mediated through a single homologous recombination anchored in the conserved regions, from upstream of *nopA* to downstream of *rhcC1* (Fig. 3, red underlined regions), converting the type-specific region located in between (Fig. 3, gold and pink underlined regions). This single-step scenario for conversion of a RhcIb-type cluster into a RhcIa cluster, and vice versa, makes it a potentially frequent event.

Indeed, a few RhcI gain and loss events were observed in recent lineages, mostly limited to single genomes in our sample. These recent changes may represent random flickers in the evolutionary history with no particular selective value, or could alternatively be the sign of positive selection for the acquisition of the T3SS, or relaxation of purifying selection leading to its loss. The type-specific variant regions are also the site of the accumulation of strain-specific genetic material – mostly insertion sequences (Fig. 3, black underlined regions) – indicating that these regions are foci of rapid gene content evolution, and that strain-specific genetic apparatus have not yet undergone strong selective pressure.

On a longer evolutionary timescale, the fact that sporadic T3SS acquisitions (of either RhcIa or RhcIb type) always occurred in a Nod+ background suggests that the T3SS may provide some symbiotic function that is under positive selection in strains that already produce the Nod signal. Similarly, sporadic losses occurred in some lineages or recent clades, including independent losses in strains DFCI-1, SEMIA6399^T and MT12 in phylogroup 4, and a loss in the ancestor of strains LTSP849 and LTSP857 in phylogroup 1. Notably, in all these cases, the loss of the T3SS is correlated with the loss of the *nod* genes on the same branch. This suggests a scenario where loss of the *nod* genes led to a relaxation of the purifying selection acting on the existing RhcI-T3SS cluster, which finally resulted in its loss. However, another explanation (that does not involve natural selection) is that the *nod* and *rhc* genes could have been acquired concomitantly or lost via homologous recombination between the surrounding core genes, thus resulting in insertion or deletion of the whole symbiotic island.

Altogether, these data indicate that the T3SS have been acquired a few times by ancestors of bradyrhizobial lineages and that the distribution of the T3SS types was later disturbed by a dynamic history of gain via horizontal gene transfer and loss, perhaps driven by selective pressure related to symbiosis with the plant host.

Finally, we also analysed the phylogenetic distribution of the atypical T3SS clusters and observed that these clusters are distributed patchily among phylogroups 1 to 5 and absent in phylogroup 6 (see Fig. 1). The patterns of conservation of atypical clusters among some groups of phylogenetically related strains within the *B. japonicum*, *B. lianoningense*, *B. yuanmingense* and *B. elkanii* species suggest separate events of acquisition in their respective common ancestors. Other than that, no correlation can be drawn between the presence of one of these atypical T3SS cluster and any other symbiotic or lifestyle features.

Possible new cases of *Bradyrhizobium* strains nodulating through their T3SS in the absence of NF

Considering the environment of isolation of the strains, it appears that 38 out of the 40 strains isolated from soils or other environments, but not nodules, lack both the *nod* and RhcI-T3SS genes. The only two exceptions are *B. sacchari* BR10280^T and *B. japonicum* FN1, isolated respectively from sugarcane roots and from Manitoba soils. These 38 strains are distributed in 5 out the 6 phylogroups and notably include all the 6 representative strains of phylogroup 6 represented in our dataset (Fig. 1). This suggests that the presence of the symbiotic genes could have a negative impact on bacterial fitness in the absence of an interaction with a plant, and be counter-selected outside of nodules. Furthermore, the strains from the divergent phylogroup six may not have a genomic background allowing them to nodulate legume plants upon acquisition of a symbiotic island.

Conversely, most isolates from plant nodules do contain *nodABC* genes, which reinforces the idea that the NF-dependent symbiotic process is the predominant type of interaction, even though some exceptions (8 isolates out of 106) exist. As previously indicated, five strains that were isolated from *Aeschynomene* species (*A. indica* and *A. evenia*) and lack *nod* and RhcI-T3SS clusters nodulate *Aeschynomene* species in an NF-independent process [5, 40, 41]. More surprisingly, three strains (*B. mercantei* SEMIA6399^T, *B. liaoningense* CCBAU83689 and *Bradyrhizobium* sp. Y36) isolated from plant species that do not belong to the genus *Aeschynomene* were also found to lack *nod* genes. The strain *B. mercantei* SEMIA6399^T was isolated from an effective nodule of the tropical legume *Deguelia costata* (syn. *Lonchocarpus costatus*), an important plant species belonging to the tribe *Millettieae* that is native to eastern Brazil [65, 66]. Interestingly, this strain, which is phylogenetically distant to the phylogroup 2 bradyrhizobia nodulating *Aeschynomene,* was confirmed to form effective nodules on its plant of origin but also on *Macroptilium atropurpureum* [65]. This strain is even used as a commercial inoculant for *D. costata*. If the symbiotic properties of the sequenced strain are confirmed, this strain would be the first example of a *Bradyrhizobium* outside phylogroup 2 that is able to use an NF-independent pathway to naturally nodulate legumes other than *Aeschynomene* spp. Strain SEMIA6399^T also lacks a RhcI-T3SS cluster but contains an atypical T3SS cluster. It would be interesting to elucidate whether this T3SS plays a role in the nodulation abilities of this bacterium.

The two others nod-lacking strains (*B. liaoningense* CCBAU83689 and *Bradyrhizobium* sp. Y36) were isolated from *G. max* and from *Phaseolus vulgaris*, respectively, two legume species known to be nodulated by rhizobia with *nod* genes. One may wonder whether these two strains were opportunist bacteria that took advantage of rhizobia excreting NFs to co-infect the same nodules. It would be interesting to verify the ability of these strains to nodulate their original host plants in an NF-independent manner. As previously stated, the strain *B. liaoningense* CCBAU83689 is the only bradyrhizobium that has a RhcI-T3SS cluster but lacks *nod* genes. We can therefore assume that this strain has previously acquired a symbiotic island containing both *nod* and T3SS genes and subsequently lost the *nod* genes but maintained the RhcI-T3SS cluster. While this strain indeed nodulates *G. max,* it remains to be determined whether its RhcI-T3SS is the major determinant governing the nodulation. Similarly, strain Y36 also contains an atypical T3SS cluster whose symbiotic role cannot be excluded.

The absence of *nopX* is correlated with the presence of *nopH* and *nopE*

The translocator proteins form a pore (called a translocon) in the host membrane that allows the passage of effectors into the eukaryotic host cell [67]. The components forming the translocon in rhizobial T3SS remain unknown, but the homology of NopX with the putative translocator HrpF from *Xanthomonas* species suggests that it could be a translocon component [68]. The absence of a gene encoding a translocator protein in all the strains containing a RhcIa cluster

is enigmatic and raises the question of which gene(s) may play this function. A similar case has been reported in the genus *Xanthomonas*, where members of the ancestral, early-branching clade lack *hrpF* [69]. Comparative genomics permitted us to identify specifically in this group of strains two candidate translocator genes (*hpaT* and *hpaH*) for which a role in effector translocation has been subsequently confirmed [69]. Following this approach, we speculated that all the RhcIa+ strains should contain (a) specific gene(s) that could functionally substitute *nopX*. Comparison of the bradyrhizobial genomes highlighted two genes that were specifically associated with the presence of a RhcIa cluster (Figs 3 and 5): *nopH*, which encodes a protein of unknown function of about 100 amino acids, and *nopE*, which encodes a protein of about 500 amino acids that is assumed to be a translocated effector [70]. These two genes are found in the same genetic organization and closely associated with the core T3SS gene clusters in all the RhcIa+ strains; *nopH* is always found upstream of *nopA,* and *nopE* is always found upstream of region III in a genetic organization that is very similar to *nopX* in the RhcIb cluster (Fig. 3).

To the best of our knowledge, the function of NopH is not known, except that this protein was found in the secretome of USDA110^T [61]. In contrast, NopE has been functionally studied, in particular in strain USDA110^T, which harbours two NopE homologues. It has been shown using CyaA reporter fusion that both NopE variants were translocated into plant cells and that they are most probably functionally redundant [70]. Indeed, only a *nopE* double mutant is significantly (positively) affected in its ability to nodulate *Vigna radiata*, suggesting that NopE proteins or a process relying on NopE activity play a negative role in the nodulation of this plant species [70].

NopH and NopE display no sequence similarity with NopX or with known translocators from other bacteria. However, the translocon proteins are not conserved among different bacterial species [71]. The specific presence of *nopH* and *nopE* only in the RhcIa+ strains make them attractive candidates for functional studies of the T3SS. It would be interesting to determine if they play a role in the translocation of effectors, keeping in mind that the functional features already reported for NopE are not incompatible with such a role.

Drastic variation between the effectome of *Bradyrhizobium* strains

In order to determine the content and diversity of T3Es in the genus *Bradyrhizobium*, we screened for the presence of a set of T3Es in the predicted proteomes of all strains using BLASTP (% of identity ≥40% over ≥70% of the length of the protein). We restricted this analysis to rhizobial effectors that have been demonstrated to be secreted through the T3SS and/or shown to affect the symbiosis positively or negatively (Table S2). As shown in Fig. 5 and as expected from previous results, T3E candidates were generally not found in genomes that do not harbour a RhcI-T3SS cluster, except for six strains (CCGE-LA001, WSM2254, ARR65, Tv2a-2, SEMIA6399^T and Ai1a-2). Two of these strains (SEMIA6399^T and Tv2a-2) present an atypical T3SS cluster, so it cannot be excluded that the identified effectors are translocated by the secretory machinery encoded by these atypical T3SS gene clusters. In addition, these two strains and strain Ai1a-2 present a few genes encoding components of the RhcI-T3SS in their genomes, which raises the possibility that their ancestor's genome once harboured a RhcI-T3SS cluster with its cortege of T3Es but lost most of the T3SS genes since (Fig. 2).

For all the strains containing a RhcI-T3SS cluster, the number of identified T3Es is highly variable, ranging from 2 for WSM3983 to 24 for *B. diazoefficiens* USDA122 (Fig. 5). Notably, *B. diazoefficiens* and *B. japonicum* strains isolated from soybean nodules have the largest set of effectors, with an average of 20 T3Es. Nevertheless, this number of T3Es also varies considerably within these 2 species, from 14 to 24. The repertoire of effectors, also called the effectome, is frequently remodelled by the acquisition and loss of genes, a process most likely shaped by the selective pressure of the host plant.

It is worthy of note that there is not a single T3E that is found in all strains with a RhcI-T3SS cluster. While no core effectome can therefore be defined within the genus *Bradyrhizobium*, four T3Es (NopM, NopP, NopT, NopC) are nevertheless present in a majority of genomes and are often found in multiple copies. Interestingly, these four T3Es are also conserved in other rhizobial genera [60], which leads to the speculation that they were part of an 'early' effectome that was acquired horizontally by the ancestors of bradyrhizobial lineages, possibly at the same time as the RhcI-T3SS clusters, and later lost in some specific lineages, perhaps following different selective pressures.

This hypothesis is in line with our reconciliation-based inference of evolutionary scenarios for the T3E genes *nopC*, *nopM*, *nopL* and *nopT*, and with our estimation of how much they co-evolved between themselves and with genes from the *nod* and RhcI-T3SS clusters (Fig. S3). As expected, the gene *nopC* was found to co-evolve strongly with the *rhc* genes encoded within the same locus (for RhcI-type T3SS clusters). Surprisingly, *nopM*, *nopL* and *nopT* formed a co-evolving gene module, even though they are encoded at distant loci. These three T3Es also showed some association with *nod* and *rhc* genes. Specifically, these T3Es are only loosely associated – sometimes not significantly – with either *nod* or *rhc* gene modules when considering both vertical and horizontal modes of co-evolution (Fig. S3a, c), whereas they showed a strong association with them (and *nod* genes in particular) when considering only horizontal co-evolution (Fig. S3b, d). A possible interpretation of these patterns is that these three T3Es were frequently co-acquired with *nod* genes (and with *rhc* genes to a lesser extent), but would often have been lost afterwards, leading to rare vertical co-transmission with either *nod* or *rhc* genes. This supports the 'early effectome' hypothesis, where these T3Es were acquired by bradyrhizobial ancestors as part of the same 'informational package' containing the *nod* and *rhc* genes (possibly on the same symbiotic island) but were

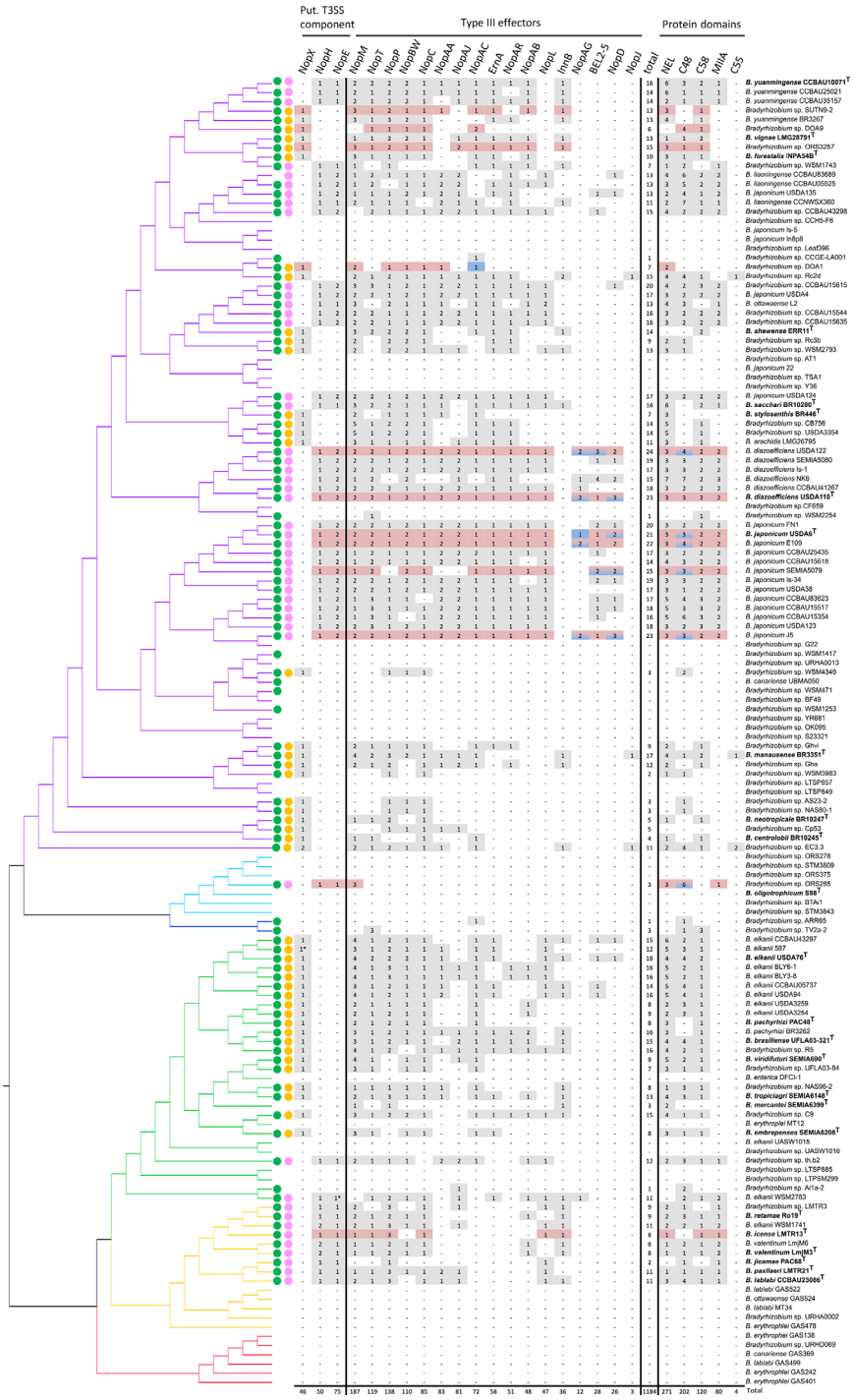

Fig. 5. Representation of the effector candidate repertoire in the genus *Bradyrhizobium*. The number and distribution of some T3SS components, some known Nops and some proteins containing known functional domains identified in some rhizobia effectors are shown. Numbers represent the number of proteins found in each strain according the research parameters defined in the Materials section. NEL, novel E3-ubiquitine ligase. MIIA, metal ion-inducible autocleavage protein. C55, YopJ-type protein. C48, SUMO protease. C58, YopT-type protein. The asterisk (*) indicates frameshift mutations in the corresponding gene. The colouring scheme of Fig. 1 is used to indicate which phylogroup the strain belongs to and the coloured circles indicate the presence of a *nod* gene cluster (in green), a RhcIa-T3SS cluster (in pink) or a RhcIb-T3SS cluster (in yellow). For the 12 strains containing a RhcI-T3SS cluster and for which the sequence genome is complete or almost complete, the boxes containing the number are respectively highlighted in red or blue when the genes are on the symbiotic island or another horizontal acquired region. When the cases are bicoloured, this indicates that different effector homologues can be found on both the symbiotic island and on another horizontal acquired region.

differentially lost or retained in descendant lineages, owing to varied selective pressures.

We also examined whether the distribution of any T3E was associated with the presence of the RhcIa or the RhcIb cluster types. No strict correlation except those discussed before for NopX, NopH and NopE was observed, indicating that the ancestral RhcIa or RhcIb clusters most probably shared similar T3E repertoires. However, at a lower frequency, two other effectors, NopAG and NopJ, both encoding a putative acetyltransferase, were found to be specifically associated with the RhcIa and RhcIb cluster, respectively. NopAG is present in seven strains of *B. japonicum* and *B. diazoefficiens* and the distant strain WSM2783, with no homologue being found outside the genus *Bradyrhizobium*. This exclusive distribution suggests that this T3E has been specifically acquired or evolved within the *B. diazoefficiens*/*B. japonicum* group and later horizontally transferred to the strain WSM2793. The NopJ effector was only found in three strains that are phylogenetically very distant from each other. This T3E, which belongs to the YopJ effector family in pathogenic bacteria, had previously only been reported in one rhizobial genome, *Ensifer fredii* strain NGR234 [60, 72]. It is therefore possible that NGR234 and these three *Bradyrhizobium* strains acquired NopJ independently by lateral gene transfer.

For the 12 strains containing a RhcI-T3SS cluster and for which the sequence genome is complete or almost complete, we have also examined the localization on the genome of the predicted T3Es. As shown in Fig. 5, most of the T3Es are associated with the symbiotic island. The only exceptions are observed for homologues of *nopAG*, *nopD*, *bel2-5* or other genes encoding C48 SUMO proteases, which are found outside of the symbiotic island. However, a *tts* box can be identified upstream of most of these genes, reinforcing the idea that they would be bona fide T3Es. Furthermore, these genes are also found within regions with low GC content and different codon usage with respect to the rest of the chromosome. This suggests that their localization away from symbiotic genes could be the result of a fragmentation of the initially acquired symbiotic island (such as has been described for the NK6 strain [39]) or from another horizontal acquisition event.

The diversity and the effectors number are most probably underestimated

T3Es are often modular proteins constituted of multiple conserved domains [73]. This particular architecture complicates the search for homologues because of the low level of similarity between effector proteins outside of these conserved domains. Thus, the search for specific functional domains has been proposed as a more sensitive approach to identify new candidate effector proteins [74]. Among the rhizobial effectors, five functional domains have been identified: two eukaryotic domains, NEL (for novel E3 ligase/IPR029487) identified in NopM [75, 76] and the SUMO protease domain (peptidase family C48/IPR003653) identified in NopD and BEL2-5 [77, 78]; the YopT-type domain (peptidase family C58/IPR006473) identified in NopT [79, 80], the YopJ-type domain (peptidase family C55/IPR005083) identified in NopJ [72] and the recent identified MIIA domain (metal ion-inducible autocleavage/IPR011086) identified in NopE [81]. In order to highlight new putative T3Es, we performed an InterProScan analysis on the annotated bradyrhizobial genomes using the MaGe interface on the MicroScope platform (Fig. 5). All proteins identified as containing one of these domains were found specifically in strains that have a RhcI-T3SS cluster or in five of the six strains (WSM2254, ARR65, Tv2a-2, SEMIA6399^T and Ai1a-2) that were discussed above and are supposed to have lost their T3SS. The search for YopT and YopJ-type domains as well as for the MIIA domain retrieved 120, 4 and 80 proteins, respectively, which correlates well with the number of homologues identified by BLASTP for NopT (119 homologues), NopJ (three homologues) and NopE (75 homologues).

In contrast, the search for the NEL and the SUMO protease domains returned a far higher number of putative T3Es (271 and 202 proteins) than the BLASTP search for homologues of NopM protein or the two SUMO proteases NopD and BEL2-5, which retrieved 187, 28 and 26 homologous proteins, respectively. We thus identified 84 additional NopM-family T3Es and 180 additional proteins containing a SUMO protease domain. This suggests that there are still many T3Es to be discovered in the genus *Bradyrhizobium*, some of which may play an important symbiotic role.

Conclusion

The importance of the T3SS in the symbiotic ability of rhizobia remains unclear. In this study, we took advantage of the large repertoire of genome sequences that is available for the genus *Bradyrhizobium* to reveal that the RhcI-T3SS gene clusters and the *nod* genes largely share a common evolutionary history and have both been conserved in almost all nodulating *Bradyrhizobium* strains. This concomitant maintenance of *nod* and T3SS clusters is most likely driven by host plant-related selective pressure and suggests a much more important role of the T3SS in the symbiotic process than hitherto suspected. The recent identification that a T3SS effector (ErnA) that is conserved among bradyrhizobia is able to induce nodule formation in the absence of NFs strengthens this idea [25].

This comparative genomic study provides new opportunities to fill gaps in our sparse knowledge of the role of the T3SS in rhizobial symbiosis. Indeed, we propose new genetic determinants that are probably required for the T3SS symbiotic function based on their phylogenetic distribution: *nopE* and *nopH* distribution complements that of *nopX*, in association with their respective RhcI gene cluster types, suggesting a possible role for these two proteins that is related to the translocation function. Furthermore, we highlighted the high diversity of NopM and SUMO protease effector families, revealing a much larger effectome than anticipated in some strains. Some of the newly identified putative effectors may hence possibly contribute to symbiotic efficiency and/

or the modulation of the host spectrum of these strains. The discovery of several atypical T3SS clusters in some strains raises the question of whether these gene clusters contribute to the symbiotic properties of bradyrhizobia. Finally, the realization that some strains isolated from nodules lack *nod* genes but contain a T3SS cluster raises the possibility that the T3SS evolved as the main determinant governing the symbiotic interaction. These strains might constitute new models for future studies aiming at a better understanding of T3SS's symbiotic function.

Funding information
This work was supported by grants from the French national research agency (grant 'SymEffectors' number ANR-16-CE20-0013). A. T. was supported by a PhD fellowship from the French Ministry of National Education, Higher Education and Research. F. L. was supported by a UK Medical Research Council (MRC) grant (MR/N010760/1).

Acknowledgements
The LABGeM (CEA/Genoscope and CNRS UMR8030), the France Génomique and French Bioinformatics Institute national infrastructures (funded as part of the Investissement d'Avenir programme managed by Agence Nationale pour la Recherche, contracts ANR-10-INBS-09 and ANR-11-INBS-0013) are acknowledged for support with the MicroScope annotation platform. Computational calculations were performed on the Imperial College high-performance computing (HPC) cluster and on the MRC Cloud Infrastructure for Microbial Bioinformatics (MRC CLIMB) cloud-based computing servers.

Author contributions
A.T and E.G. devised the study, with input from F.L., Z.R., D.G., A.C and R. K. Bioinformatic analyses were performed by A.T and F.L. with advice from R.K., A.T., F.L and E.G. wrote the manuscript, and all authors provided critical review and commentary.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-González A, Eldridge DJ *et al*. A global atlas of the dominant bacteria found in soil. *Science* 2018;359:320–325.
2. VanInsberghe D, Maas KR, Cardenas E, Strachan CR, Hallam SJ *et al*. Non-symbiotic *Bradyrhizobium* ecotypes dominate North American forest soils. *ISME J* 2015;9:2435–2441.
3. Jones FP, Clark IM, King R, Shaw LJ, Woodward MJ *et al*. Novel European free-living, non-diazotrophic *Bradyrhizobium* isolates from contrasting soils that lack nodulation and nitrogen fixation genes - a genome comparison. *Sci Rep* 2016;6:25858.
4. Oldroyd GED, Speak OGE. Speak, Friend, and enter: signalling systems that promote beneficial symbiotic associations in plants. *Nat Rev Microbiol* 2013;11:252–263.
5. Giraud E, Moulin L, Vallenet D, Barbe V, Cytryn E *et al*. Legumes symbioses: absence of NOD genes in photosynthetic bradyrhizobia. *Science* 2007;316:1307–1312.
6. Masson-Boivin C, Giraud E, Perret X, Batut J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium recipes? *Trends Microbiol* 2009;17:458–466.
7. Masson-Boivin C, Sachs JL. Symbiotic nitrogen fixation by rhizobia-the roots of a success story. *Curr Opin Plant Biol* 2018;44:7–15.
8. Zhao R, Liu LX, Zhang YZ, Jiao J, Cui WJ *et al*. Adaptive evolution of rhizobial symbiotic compatibility mediated by co-evolved insertion sequences. *ISME J* 2018;12:101–111.
9. Kaneko T, Nakamura Y, Sato S, Minamisawa K, Uchiumi T *et al*. Complete genomic sequence of nitrogen-fixing symbiotic bacterium *Bradyrhizobium japonicum* USDA110. *DNA Res* 2002;9:189–197.
10. Reeve W, van Berkum P, Ardley J, Tian R, Gollagher M *et al*. High-quality permanen¹draf¹ genome sequence of the *Bradyrhizobium elkanii* type strain USDA 76^T, isolated from *Glycine max* (L.) Merr. *Stand Genomic Sci* 2017;12:26.
11. Gully D, Teulet A, Busset N, Nouwen N, Fardoux J *et al*. Complete genome sequence of *Bradyrhizobium* sp. ORS285, a photosynthetic strain able to establish Nod factor-dependent or Nod factor-independent symbiosis with *Aeschynomene* legumes. *Genome Announc* 2018;5:e00421–17.
12. Le Quéré A, Gully D, Teulet A, Navarro E, Gargani D *et al*. Complete genome sequence of *Bradyrhizobium* sp. strain ORS3257, an efficient nitrogen-fixing bacterium isolated from cowpea in Senegal. *Microbiol Resour Announc* 2019;8:e01449–18.
13. Deng W, Marshall NC, Rowland JL, McCoy JM, Worrall LJ *et al*. Assembly, structure, function and regulation of type III secretion systems. *Nat Rev Microbiol* 2017;15:323–337.
14. Büttner D. Behind the lines-actions of bacterial type III effector proteins in plant cells. *FEMS Microbiol Rev* 2016;40:894–937.
15. Deakin WJ, Broughton WJ. Symbiotic use of pathogenic strategies: rhizobial protein secretion systems. *Nat Rev Microbiol* 2009;7:312–320.
16. Staehelin C, Krishnan HB. Nodulation outer proteins: double-edged swords of symbiotic rhizobia. *Biochem J* 2015;470:263–274.
17. Tampakaki AP. Commonalities and differences of T3SSs in rhizobia and plant pathogenic bacteria. *Front Plant Sci* 2014;5:114.
18. Krause A, Doerfel A, Göttfert M. Mutational and transcriptional analysis of the type III secretion system of *Bradyrhizobium japonicum*. *Mol Plant Microbe Interact* 2002;15:1228–1235.
19. Wassem R, Kobayashi H, Kambara K, Le Quéré A, Walker GC *et al*. TtsI regulates symbiotic genes in *Rhizobium* species NGR234 by binding to *tts* boxes. *Mol Microbiol* 2008;68:736–748.
20. Miwa H, Okazaki S. How effectors promote beneficial interactions. *Curr Opin Plant Biol* 2017;38:148–154.
21. Yang S, Tang F, Gao M, Krishnan HB, Zhu H. R gene-controlled host specificity in the legume-rhizobia symbiosis. *Proc Natl Acad Sci USA* 2010;107:18735–18740.
22. Sugawara M, Takahashi S, Umehara Y, Iwano H, Tsurumaru H *et al*. Variation in bradyrhizobial NopP effector determines symbiotic incompatibility with Rj2-soybeans via effector-triggered immunity. *Nat Commun* 2018;9:3139.
23. Okazaki S, Kaneko T, Sato S, Saeki K. Hijacking of leguminous nodulation signaling by the rhizobial type III secretion system. *Proc Natl Acad Sci USA* 2013;110:17131–17136.
24. Okazaki S, Tittabutr P, Teulet A, Thouin J, Fardoux J *et al*. Rhizobium-legume symbiosis in the absence of Nod factors: two possible scenarios with or without the T3SS. *ISME J* 2016;10:64–74.
25. Teulet A, Busset N, Fardoux J, Gully D, Chaintreuil C *et al*. The rhizobial type III effector eRNA confers the ability to form nodules in legumes. *Proc Natl Acad Sci USA* 2019;116:21758–21768.
26. Lassalle F, Veber P, Jauneikaite E, Didelot X. Automated reconstruction of all gene histories in large bacterial pangenome datasets and search for co-evolved gene modules with Pantagruel. *bioRxiv* 2019;586495.
27. Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat Biotechnol* 2017;35:1026–1028.
28. Sievers F, Higgins DG, Desmond GH. Clustal omega for making accurate alignments of many protein sequences. *Protein Sci* 2018;27:135–145.
29. Suyama M, Torrents D, Bork P. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. *Nucleic Acids Res* 2006;34:W609–W612.
30. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Phylogenetics* 2013;30:1312–1313.

31. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A *et al*. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol* 2012;61:539–542.
32. Vallenet D, Calteau A, Cruveiller S, Gachet M, Lajus A *et al*. Microscope in 2017: an expanding and evolving integrated resource for community expertise of microbial genomes. *Nucleic Acids Res* 2017;45:D517–D528.
33. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
34. Cohen O, Ashkenazy H, Belinky F, Huchon D, Pupko T. GLOOME: gain loss mapping engine. *Bioinformatics* 2010;26:2914–2915.
35. Szöllősi GJ, Tannier E, Daubin V, Boussau B. The inference of gene trees with species trees. *Syst Biol* 2015;64:e42–e62.
36. Morel B, Kozlov AM, Stamatakis A, Szöllősi GJ. GeneRax: a tool for species tree-aware maximum likelihood based gene tree inference under gene duplication, transfer, and loss. *bioRxiv* 2019;779066.
37. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J *et al*. CLIMB (the cloud infrastructure for microbial bioinformatics): an online resource for the medical microbiology community. *Microb Genom* 2016;2:e000086.
38. Okazaki S, Noisangiam R, Okubo T, Kaneko T, Oshima K *et al*. Genome analysis of a novel *Bradyrhizobium* sp. DOA9 carrying a symbiotic plasmid. *PLoS One* 2015;10:e0117392.
39. Iida T, Itakura M, Anda M, Sugawara M, Isawa T *et al*. Symbiosis island shuffling with abundant insertion sequences in the genomes of extra-slow-growing strains of soybean bradyrhizobia. *Appl Environ Microbiol* 2015;81:4143–4154.
40. Miché L, Moulin L, Chaintreuil C, Contreras-Jimenez JL, Munive-Hernández J-A *et al*. Diversity analyses of *Aeschynomene* symbionts in Tropical Africa and Central America reveal that *nod*-independent stem nodulation is not restricted to photosynthetic bradyrhizobia. *Environ Microbiol* 2010;12:2152–2164.
41. Okubo T, Fukushima S, Itakura M, Oshima K, Longtonglang A *et al*. Genome analysis suggests that the soil oligotrophic bacterium *Agromonas oligotrophica* (*Bradyrhizobium oligotrophicum*) is a nitrogen-fixing symbiont of *Aeschynomene indica*. *Appl Environ Microbiol* 2013;79:2542–2551.
42. Tian R, Parker M, Seshadri R, Reddy T, Markowitz V *et al*. 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* 2015;10:27.
43. Stępkowski T, Watkin E, McInnes A, Gurda D, Gracz J *et al*. Distinct *Bradyrhizobium* [corrected] communities nodulate legumes native to temperate and tropical monsoon Australia. *Mol Phylogenet Evol* 2012;63:265–277.
44. van Berkum P, Fuhrmann JJ. Evolutionary relationships among the soybean bradyrhizobia reconstructed from 16S rRNA gene and internally transcribed spacer region sequence divergence. *Int J Syst Evol Microbiol* 2000;50 Pt 6:2165–2172.
45. Benzine J, Shelobolina E, Xiong MY, Kennedy DW, McKinley JP *et al*. Fe-phyllosilicate redox cycling organisms from a redox transition zone in Hanford 300 area sediments. *Front Microbiol* 2013;4:388.
46. Crovadore J, Calmin G, Cochard B, Chablais R, Lefort F. Whole-genome sequence of *Bradyrhizobium elkanii* strain UASWS1015, a highly ammonia-tolerant nitrifying bacterium. *Genome Announc* 2016;4:e00111–00116.
47. van Berkum P, Elia P, Song Q, Eardly BD. Development and application of a multilocus sequence analysis method for the identification of genotypes within genus *Bradyrhizobium* and for establishing nodule occupancy of soybean (*Glycine max* L. Merr). *Mol Plant Microbe Interact* 2012;25:321–330.
48. Chun J, Oren A, Ventosa A, Christensen H, Arahal DR *et al*. Proposed minimal standards for the use of genome data for the taxonomy of prokaryotes. *Int J Syst Evol Microbiol* 2018;68:461–466.
49. de Lajudie PM, Andrews M, Ardley J, Eardly B, Jumas-Bilak E *et al*. Minimal standards for the description of new genera and species of rhizobia and agrobacteria. *Int J Syst Evol Microbiol* 2019;69:1852–1863.
50. Avontuur JR, Palmer M, Beukes CW, Chan WY, Coetzee MPA *et al*. Genome-informed *Bradyrhizobium* taxonomy: where to from here? *Syst Appl Microbiol* 2019;42:427–439.
51. Garrido-Sanz D, Redondo-Nieto M, Mongiardini E, Blanco-Romero E, Durán D *et al*. Phylogenomic analyses of *Bradyrhizobium* reveal uneven distribution of the lateral and subpolar flagellar systems, which extends to *Rhizobiales*. *Microorganisms* 2019;7:50.
52. Troisfontaines P, Cornelis GR. Type III secretion: more systems than you think. *Physiology* 2005;20:326–339.
53. Hu Y, Huang H, Cheng X, Shu X, White AP *et al*. A global survey of bacterial type III secretion systems and their effectors. *Environ Microbiol* 2017;19:3879–3895.
54. Schmeisser C, Liesegang H, Krysciak D, Bakkou N, Le Quéré A *et al*. *Rhizobium* sp. strain NGR234 possesses a remarkable number of secretion systems. *Appl Environ Microbiol* 2009;75:4035–4045.
55. Weidner S, Becker A, Bonilla I, Jaenicke S, Lloret J *et al*. Genome sequence of the soybean symbiont *Sinorhizobium fredii* HH103. *J Bacteriol* 2012;194:1617–1618.
56. Schuldes J, Rodriguez Orbegoso M, Schmeisser C, Krishnan HB, Daniel R *et al*. Complete genome sequence of the broad-host-range strain *Sinorhizobium fredii* USDA257. *J Bacteriol* 2012;194:4483.
57. Gazi AD, Sarris PF, Fadouloglou VE, Charova SN, Mathioudakis N *et al*. Phylogenetic analysis of a gene cluster encoding an additional, rhizobial-like type III secretion system that is narrowly distributed among *Pseudomonas syringae* strains. *BMC Microbiol* 2012;12:188.
58. Hueck CJ. Type III protein secretion systems in bacterial pathogens of animals and plants. *Microbiol Mol Biol Rev* 1998;62:379–433.
59. Okazaki S, Zehner S, Hempel J, Lang K, Göttfert M. Genetic organization and functional analysis of the type III secretion system of *Bradyrhizobium elkanii*. *FEMS Microbiol Lett* 2009;295:88–95.
60. Kimbrel JA, Thomas WJ, Jiang Y, Creason AL, Thireault CA *et al*. Mutualistic co-evolution of type III effector genes in *Sinorhizobium fredii* and *Bradyrhizobium japonicum*. *PLoS Pathog* 2013;9:e1003204.
61. Hempel J, Zehner S, Göttfert M, Patschkowski T. Analysis of the secretome of the soybean symbiont *Bradyrhizobium japonicum*. *J Biotechnol* 2009;140:51–58.
62. Jiménez-Guerrero I, Pérez-Montaño F, Medina C, Ollero FJ, López-Baena FJ. NopC is a rhizobium-specific type 3 secretion system effector secreted by *Sinorhizobium* (*Ensifer*) *fredii* HH103. *PLoS One* 2015;10:e0142866.
63. Büttner D, Nennstiel D, Klüsener B, Bonas U. Functional analysis of HrpF, a putative type III translocon protein from *Xanthomonas campestris* pv. *vesicatoria*. *J Bacteriol* 2002;184:2389–2398.
64. Marie C, Deakin WJ, Ojanen-Reuhs T, Diallo E, Reuhs B *et al*. TtsI, a key regulator of *Rhizobium* species NGR234 is required for type III-dependent protein secretion and synthesis of rhamnose-rich polysaccharides. *Mol Plant Microbe Interact* 2004;17:958–966.
65. Helene LCF, Delamuta JRM, Ribeiro RA, Hungria M. *Bradyrhizobium mercantei* sp. nov., a nitrogen-fixing symbiont isolated from nodules of *Deguelia costata* (syn. Lonchocarpus costatus). *Int J Syst Evol Microbiol* 2017;67:1827–1834.
66. Ribeiro RA, Helene LCF, Delamuta JRM, Hungria M. Genome sequence of *Bradyrhizobium mercantei* strain SEMIA6399^T, isolated from nodules of *Deguelia costata* in Brazil. *Genome Announc* 2017;5:e00943–17.
67. Ji H, Dong H. Key steps in type III secretion system (T3SS) towards translocon assembly with potential sensor at plant plasma membrane. *Mol Plant Pathol* 2015;16:762–773.
68. Saad MM, Staehelin C, Broughton WJ, Deakin WJ. Protein-protein interactions within type III secretion system-dependent pili of *Rhizobium* sp. strain NGR234. *J Bacteriol* 2008;190:750–754.
69. Pesce C, Jacobs JM, Berthelot E, Perret M, Vancheva T *et al*. Comparative genomics identifies a novel conserved protein, HpaT, in proteobacterial type III secretion systems that do not

possess the putative translocon protein HrpF. *Front Microbiol* 2017;8:1177.

70. Wenzel M, Friedrich L, Göttfert M, Zehner S. The type III-secreted protein NopE1 affects symbiosis and exhibits a calcium-dependent autocleavage activity. *Mol Plant Microbe Interact* 2010;23:124–129.
71. Büttner D, Bonas U. Port of entry--the type III secretion translocon. *Trends Microbiol* 2002;10:186–192.
72. Kambara K, Ardissone S, Kobayashi H, Saad MM, Schumpp O *et al*. Rhizobia utilize pathogen-like effector proteins during symbiosis. *Mol Microbiol* 2009;71:92–106.
73. Dean P. Functional domains and motifs of bacterial type III effector proteins and their roles in infection. *FEMS Microbiol Rev* 2011;35:1100–1125.
74. Gomez-Valero L, Rusniok C, Carson D, Mondino S, Pérez-Cobas AE *et al*. More than 18 000 effectors in the *Legionella* genus genome provide multiple, independent combinations for replication in human cells. *Proc Natl Acad Sci USA* 2019;116:2265–2273.
75. Xin D-W, Liao S, Xie Z-P, Hann DR, Steinle L *et al*. Functional analysis of NopM, a novel E3 ubiquitin ligase (Nel) domain effector of Rhizobium sp. strain NGR234. *PLoS Pathog* 2012;8:e1002707.
76. Xu C-C, Zhang D, Hann DR, Xie Z-P, Staehelin C. Biochemical properties and in planta effects of NopM, a rhizobial E3 ubiquitin ligase. *J Biol Chem* 2018;293:15304–15315.
77. Rodrigues JA, López-Baena FJ, Ollero FJ, Vinardell JM, Espuny MDR *et al*. NopM and NopD are rhizobial nodulation outer proteins: identification using LC-MALDI and LC-ESI with a monolithic capillary column. *J Proteome Res* 2007;6:1029–1037.
78. Faruque OM, Miwa H, Yasuda M, Fujii Y, Kaneko T *et al*. Identification of *Bradyrhizobium elkanii* genes involved in incompatibility with soybean plants carrying the *Rj4* allele. *Appl Environ Microbiol* 2015;81:6710–6717.
79. Dai W-J, Zeng Y, Xie Z-P, Staehelin C. Symbiosis-promoting and deleterious effects of NopT, a novel type 3 effector of *Rhizobium* sp. strain NGR234. *J Bacteriol* 2008;190:5101–5110.
80. Fotiadis CT, Dimou M, Georgakopoulos DG, Katinakis P, Tampakaki AP. Functional characterization of NopT1 and NopT2, two type III effectors of *Bradyrhizobium japonicum*. *FEMS Microbiol Lett* 2012;327:66–77.
81. Hoyer E, Knöppel J, Liebmann M, Steppert M, Raiwa M *et al*. Calcium binding to a disordered domain of a type III-secreted protein from a coral pathogen promotes secondary structure formation and catalytic activity. *Sci Rep* 2019;9:7115.