

Bayesian Phylogenetic Analysis of Rhizobia Isolated From Root-Nodules of Three Tunisian Wild Legume Species of the Genus *Sulla*

Rim Chriki-Adeeb and Ali Chriki*

Laboratory of Genetics, Faculty of Sciences of Bizerte, Zarzouna 7021, Tunisia

Abstract

We used a relaxed-clock program (*Multidivtime*) to estimate the phylogenetic relatedness and substitution rates by comparative analyses of 16S rRNA gene and ITS region sequences. Four rhizobial strains isolated from three Tunisian wild *Sulla* legumes were used together with reference strains to obtain the 16S rRNA gene and ITS region alignments. Bayesian inferred trees were congruent, and showed a clear split between *Agrobacterium* and *Rhizobium* species. Among the four Tunisian isolates, Hsc01 belonged to *Rhizobium* species and formed a monophyletic clade with *Rhizobium sullae*. The strain Hc04 was placed into the *Agrobacterium* group, but it belonged to *Rhizobium* species. Two remaining strains (Hc01 and Hcar01) belonged to typically *Agrobacterium* species. Divergence times between these four strains were earlier the existence of legumes. Using sequence alignments of the 16S rRNA genes and ITS regions, we inferred different rates of nucleotide substitutions across these two molecular markers. The ITS region evolutionary rate was 15-fold higher than the 16S rRNA gene rate, suggesting that the ITS region represented an appropriate molecular marker for inferring phylogenies and divergence times in bacteria.

Keywords: *Agrobacterium*; *Rhizobium*; 16S rRNA gene; ITS region; Molecular phylogeny; *Sulla*

Introduction

The family *Leguminosae* (or *Fabaceae*) is the third largest family of angiosperms with approximately 730 genera and about 19400 species [1]. It is most remarkable for its wide evolutionary diversification and cosmopolitan distribution [2]. Within this family, the genus *Hedysarum* L. consists of about 150 species [3]. Herbaceous legumes of this genus present a natural distribution, particularly in the Mediterranean, temperate Europe, North and South Africa, Asia Minor, Siberia, North America from Arizona into Alaska and the Arctic regions of Canada [4]. In the Mediterranean basin, nine *Hedysarum* species were classified in the section *Spinossissima* [5]. Referring to morphological characters, Choi and Ohashi [6] treated this section as a distinct genus *Sulla*.

In Tunisia, the genus *Sulla* is represented by five native species (*S. capitata*, *S. carnosa*, *S. coronaria*, *S. pallida* and *S. spinosissima*) widely distributed throughout the country [5]. The interest in these wild legumes arises from their important forage value under all climate types (sub humid, semiarid and arid) of Tunisia. However, little is known about their root-nodule symbiotic bacteria which were investigated in only the two spontaneous species, *S. spinosissima* and *S. carnosa*, of the arid regions of the country [7,8]. Up today, there is no information regarding nitrogen-fixing symbioses diversity from the other *Sulla* wild species distributed in the semiarid and sub humid areas of Tunisia.

The establishment of nitrogen-fixing symbiosis with some soil bacteria, collectively named rhizobia, is a unique feature of the plants of the *Fabaceae* family [9]. According to Bergey's Manual of Systematic Bacteriology, rhizobia are included in the well-known genera *Rhizobium*, *Bradyrhizobium*, *Agrobacterium* and *Phyllobacterium* within the family of *Rhizobiaceae* in the *Alphaproteobacteria* subphylum [10]. Since then, rhizobial and agrobacterial taxa have been revised at the genus and species levels. The genus *Sinorhizobium* was proposed for *Rhizobium fredii* [11,12]. *Mesorhizobium* genus was proposed by Jarvis *et al.* [13] for other rhizobial taxa, e.g. *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum* and *Rhizobium tianshanense*. Also, Agrobacteria were classified as *Agrobacterium tumefaciens*, *Agrobacterium radiobacter*, *Agrobacterium rhizogenes*, *Agrobacterium rubi* and *Agrobacterium vitis*. Young *et al.* [14] proposed that the genus *Rhizobium* be emended to include the species within the genus *Agrobacterium*. Part of the justification for this proposal was the observation that the 16S rRNA sequences of certain species of *Rhizobium* (e.g. *galegae*) are more similar to *Agrobacterium*

sequences than they are to *Rhizobium* sequences. This proposal has been controversial, however, because of the ecological and genomic differences that exist between the two genera [15,16].

Rhizobium, *Bradyrhizobium*, *Mesorhizobium*, *Sinorhizobium* and *Agrobacterium* were traditionally classified on the basis of phenotypic characteristics such as nodulation, and pathogenic and physiological properties. However, phenotypic traits have become less important in the taxonomic evaluation of rhizobia [17]. Due to its universal distribution and slow rate of sequence evolution, 16S rRNA has been the most widely used gene for the identification and taxonomic assignment of bacteria [18]. However, high sequence variation in the internally transcribed spacer (ITS) region has been shown to be more informative for the taxonomic evaluation and phylogenetic classification of rhizobia [19,20].

In this study, rhizobial strains were isolated from root-nodules of *Sulla* legumes grown in semiarid and sub humid areas of Tunisia. Their ITS regions and 16S rRNA genes were sequenced and compared with those of reference strains. Phylogenetic analysis of these sequences was performed to compare the phylograms generated from these two types of sequences and to evaluate the use of the ITS region as a taxonomic marker. Evolutionary rates and divergence times between the different rhizobial species included in this study were also estimated using several Bayesian approaches.

Materials and Methods

Collection of root-nodules and isolation of bacteria

The investigated strains and the host plants are listed in Table 1. Root-nodules were collected in November-December 2009 from young *Sulla* plants with about three or four ramifications. The distance between sampling sites was at least 50 km from each site, 4-6 plants separated at

*Corresponding author: Ali Chriki, Laboratory of Genetics, Faculty of Sciences of Bizerte, Zarzouna 7021, Tunisia, Tel: +21626306523; Fax: +216 72 59 05 66; E-mail: ali.chriki@gmail.com

Received March 07, 2015; Accepted March 24, 2015; Published March 31, 2015

Citation: Chriki-Adeeb R, Chriki A (2015) Bayesian Phylogenetic Analysis of Rhizobia Isolated From Root-Nodules of Three Tunisian Wild Legume Species of the Genus *Sulla*. J Phylogen Evolution Biol 3: 149. doi:10.4172/2329-9002.1000149

Copyright: © 2015 Chriki-Adeeb R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

least by 30 m between each other were chosen randomly and uprooted. Three or four undamaged, healthy root-nodules of similar size were excised from the lateral roots of each plant, brushed free of soil debris and then immediately either directly used for bacterial isolation or stored dried in tubes containing CaCl₂ [21]. Isolation was performed on each nodule separately. Upon use, nodules were rehydrated in sterile water and surface-sterilized by immersion in 95% ethanol for 20 s followed by 5% sodium hypochlorite for 3 min, and finally washed seven times with sterile distilled water. Nodules were then individually crushed in a drop of sterile water and the suspension was streaked on yeast-mannitol agar (YMA) medium [21] in Petri dishes. Isolates were incubated at 28°C, and the purity of the isolates was guaranteed by repeated streaking of single colonies on the same media. Pure cultures were preserved at 4°C for temporary storage or in 30% v/v glycerol at -20°C for longer time storage.

DNA extraction

DNAs were extracted according to the method of Muresu et al. [22] with little modifications. Cells were lysed by resuspending a loopful of plate-grown isolated colonies in 50 µL of lysis buffer [0.5% sodium dodecyl sulphate (SDS), 0.1M NaOH] in a 1.5-mL polypropylene tube, followed by stirring for 60 s on a vortex and heating at 95°C for 15 min. The lysate was centrifuged for 15 min, and 40 µL of the supernatant was mixed with 160 µL of sterile water. Lysates were stored at 4°C prior to PCR.

PCR amplification of the 16S rRNA gene and ITS region

For the 16S rRNA gene amplification, one microliter of the lysate containing the total DNA of each bacterial isolate was treated in a PCR Eppendorf Cycler using the two 16S rRNA gene-target universal bacterial primers 63F [23] and 1389R [24] at 1 µM each in a 25-µL reaction volume, using the following program: initial denaturation at 95°C for 2 min; 35 cycles at 95°C for 30 s, 55°C for 1 min, 72°C for 4 min; and a final extension at 72°C for 10 min. Primers FGPS1490 [25] and FGPS132' [26] were used to amplify the ITS region, using the following program: initial denaturation at 94°C for 5 min; 35 cycles at 94°C for 1 min, 55°C for 55 s, 72°C for 2 min; and a final extension at 72°C for 10 min. For both PCR, the reaction mixture contained 20 mM Tris-HCl (pH 8.4), 50 mM KCl, 1.5 mM MgCl₂, 0.2 mM of each of dATP, dCTP, dGTP and dTTP, 1 µM of each primer, and 2.5 U Taq DNA polymerase, recombinant (Invitrogen Life Technologies).

rDNA sequencing

PCR products were electrophoresed on 1.5% agarose gel. Bands were eluted and purified with QIAquick gel extraction kit (Qiagen) and sequenced. Sequencing reactions were performed in the DNA Engine Tetrad 2 Peltier Thermal Cycler (BIO-RAD) using the ABI BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems), following the protocols supplied by the manufacturer. Single-pass sequencing was performed on each template using the above-described primers. CodonCode Aligner (CodonCode Corporation) was used to edit and reconcile sequences. The 16S rRNA and ITS region sequences were then deposited in GenBank under the accession numbers indicated in Table 1, and compared with those held in GenBank using the BLASTN program [27]. Percent identity between nucleotide sequences was estimated by using the FASTA program version 36.3.5d [28].

Phylogenetic analyses

Two data sets were constructed from the BLASTN researches. The first data set was composed of 20 16S rRNA sequences from *Agrobacterium* and *Rhizobium* species, including the four 16S rRNA sequences obtained in this study. The second data set contained 21

ITS sequences from the same *Agrobacterium* and *Rhizobium* groups, also including the four ITS sequences from this study. These two data sets were aligned by using ClustalX version 2.0.10 [29]. We used 1513 and 1636 nucleotide positions for phylogenetic analysis of the first and second data sets, respectively. The best-fit model of nucleotide substitution for these data sets was selected by sampling across the entire general time reversible model space using the Bayesian Markov chain Monte Carlo (MCMC) analysis [30] using MrBayes program V3.2 [31].

The substitution model was set to HKY [32] for tree construction. To test the robustness of inferred topologies, posterior probabilities were determined by a Bayesian MCMC method also implemented in the program MrBayes V3.2. Two independent runs were conducted for one million generations using the HKY matrix and model parameters (gamma shape and proportion invariant), and trees were sampled every 500 generations (default value). Agreement of the estimated parameters between the two independent runs was assessed by the average standard deviation of split frequencies (a value below 0.01 was considered as a very good indication of convergence). After discarding the first 25% of the MCMC chains as burn-in, the sampled trees from both runs were pooled. The tree with maximum posterior probability was assessed using a consensus of the pooled trees. Posterior probability of >90% was considered to identify supported nodes. We also tested two alternative molecular clock hypotheses (strict-clock hypothesis versus relaxed-clock model) by calculating the Bayes Factors (BF) using the following equation [33]:

$2\ln(\text{BF}_{21}) = 2[\ln(\text{HM}_2) - \ln(\text{HM}_1)]$ where HM₂ is the harmonic mean of the posterior sample of likelihoods from the second model and HM₁ is the harmonic mean of the posterior sample of likelihoods from the first model. Positive values of $2\ln(\text{BF}_{21})$ greater than 10 are indicative of support for the second model over the first one.

Divergence time estimating

Molecular dating analyses were conducted separately for each data set, using a Bayesian relaxed molecular clock approach [34]. This Bayesian dating approach uses MCMC methods to derive posterior distributions of rates and times given an assumed evolutionary topology. For each data set, the phylogenetic tree generated by MrBayes program was used as a framework for this study. The Thorne et al. [34] method could be achieved in three steps. First, the program Baseml of the PAML package version 3.14 [35] was used to obtain estimates of the transition/transversion ratio, and the rates for site classes under the discrete-gamma model of rates among sites under the HKY+G substitution model. The slow-growing rhizobial species, *Bradyrhizobium japonicum*, was used as outgroup to investigate divergence times within- and between *Agrobacterium/Rhizobium* groups. Second, the program Estbranches [34] was used to estimate branch lengths of the topology. Finally, the program Multidivtime [34] was used to estimate divergence times. After a burn-in phase including 100 000 cycles, the Markov chain was sampled 10 000 times with 100 cycles between samples. Convergence of the MCMC algorithm was assessed by repeating each analysis at least twice. The outgroup is not considered in the Multidivtime output files.

Multidivtime requires a value for the mean of the prior distribution of the time separating the ingroup root from the present (rttm). A value of 250 million years (Myr) was used on the basis of the divergence time between *Rhizobium* and *Agrobacterium*, estimated to be in the range 110-415 Myr [36,37]. Because of the uncertainty in determining this prior, a standard deviation (rttmSD) of 125 Myr was used. Other parameters such as the mean of the prior distribution for the rate

Isolate	Host plant	GeneBank accession number	
		16S rRNA gene	ITS region
Hc01, Hc04	<i>S. coronarium</i>	JN944178, JN944189	JN944179, JN944182
Hcar01	<i>S. carmosa</i>	JN944192	JN944185
Hsc01	<i>S. capitata</i>	JN944190	JN944183

Table 1: Bacterial strains isolated from *Sulla* species in Tunisia.

of evolution (r_{rate}) and the mean of the prior distribution for the autocorrelation parameter (brown mean) were calculated specially for each alignment using the branch lengths information from the Estbranches program and the in group root prior; r_{rate} was given by the median of the root-to-tip branch lengths divided by r_{ttm} , whereas brown mean was obtained dividing a constant value of 1.5 by r_{ttm} [38]. Standard deviations of these parameters (r_{rateSD} and brownSD) were set equal to the parameters themselves [39,40].

Results and Discussion

Sequence analyses

For the 16S rRNA sequence analyses, 16 reference strains and four Tunisian isolates were included. For the ITS region sequence analysis, four Tunisian isolates and 17 reference strains were studied. All ITS region sequences analysed in this study contained two deduced tRNA genes, tRNA^{Ala} and tRNA^{Ala}. It is interesting to note that the four Tunisian strains produced one band in PCR amplification of the ITS region. Kwon et al. [19] also signalled that most rhizobial strains contained one type of ITS region sequence, suggesting that this region may be useful as a marker for phylogenetic analysis.

Among the four Tunisian strains included, the percent sequence identity between the 16S rRNA genes of Hc04 and those of Hc01, Hcar01, and Hsc01 averaged 94.5%, with values ranging from 94.2 to 94.8%. The percent sequence identity between the 16S rRNA genes of Hsc01 and those of Hc01, Hcar01, and Hc04 was slightly less (94.4%), with values ranging from 94.1 to 94.6%. In contrast, the percent sequence identity between the 16S rRNA genes of Hc01 and Hcar01 was high (98.6%), indicating that these two strains could be assigned to the same species. It is in fact generally admitted that 16S rRNA sequences with greater than 97% identity are typically assigned to the same species, those with >95% identity are typically assigned to the same genus, and those with > 80% identity are typically assigned to the same phylum [41]. According to these conventionally sequence identity limits, the four Tunisian rhizobial isolates represented different strains that could be placed in three genera of the same phylum.

The sequence identity search from the database showed that the strain Hsc01 was most similar (99.4%) to *R. sultae* (Y10170). The strains Hcar01 and Hc01 were most similar to *R. pusense* FJ969841 (99.9%) and *A. tumefaciens* EF620461 (99.6%), respectively. The strain Hc04 was most similar (96.3%) to *R. huautlense* (AM237359), indicating that it may represent a new species to be further characterised by DNA: DNA hybridisations.

A similar pattern but with lower percent sequence identity values emerged in parallel comparisons among the ITS region sequences. For the four Tunisian strains, the average percent sequence identity was 72.5%, with values ranging from 67.2% (between Hsc01 and Hc04) to 78.9% (between Hc01 and Hcar01). After a pairwise analysis, the ITS regions of strain Hcar01 and *A. tumefaciens* (AF541973) showed 92.5% sequence identity. The strain Hc01 was most similar (84.9%) to *A. rubi* (AF345277). The strain Hsc01 was similar to *A. rhizogenes* AF345275 (79.0%) and to *R. hainannense* AF321872 (76.2%), whereas Hc04 was most similar (72.0%) to *R. giardinii* (EU288738).

Both molecular markers used in this study confirmed the assignment of Hsc01 and Hc04 to the genus *Rhizobium*, whereas Hc01 and Hcar01 could be considered as *Agrobacterium*-like bacteria. The substantial percent sequence identity (79%) between the ITS regions of Hsc01 and *A. rhizogenes* confirmed the relationship of the later *Agrobacterium* with some *Rhizobium* species, such as *R. tropici* [42]. Also, the relatively high percent sequence identity (72%) between the ITS regions of Hc04 and *R. giardinii* is especially interesting in rhizobial taxonomy because *R. giardinii* is separated from rhizobial genera and may represent a novel genus [43]. The identification of *Agrobacterium*-like bacteria (Hc01, Hcar01) isolated from root-nodules of *Sulla* species is another interesting result because such nodule endophyte may have a positive impact on nodulation [44], and on the growth of host plants [43]. However, *Agrobacterium* was not isolated from root-nodules of *Sphaerophysa salsula* [9], and did not represent a predominant nodule endophyte [45].

Despite the important number of nodules (about 50) examined in this study, only four *Rhizobium*-*Agrobacterium* strains were isolated. Previous studies indicated the difficulty to isolate rhizobia from root-nodules of wild *Sulla* species [46]. This experimental difficulty may be due to the viable but non-culturable state of rhizobia within nodules [22].

Best substitution model testing

Phylogenetic analysis is first based on the choice of the model of DNA substitution that best describes each alignment. The models typically used in phylogenetic analysis are all special cases of the general time-reversible (GTR) model. We applied the reversible jump MCMC approach [30] by using MrBayes program to select the best DNA substitution model for our 16S rRNA and ITS alignments. In this Bayesian analysis, model choice is guided by the posterior probability criterion. For the two data sets analysed in this study, the posterior probability is spread among a handful of models. Table 2 shows the best models selected for each alignment. For the two data sets analysed, we observe that there is considerable uncertainty concerning the correct substitution model. Models with the highest posterior probabilities were (121131) and (121321) for 16S rRNA and ITS alignments, respectively. However, these two models, which are not available in MrBayes v3.2, were not used for further analyses. So, the HKY model, referred as (121121) and associated to posterior probability values greater than 0.05 (Table 2), was retained for both alignments. It should be noted that the most complicated model (GTR: (123456)) was not found to be best in any of our two analyses. This model was however chosen with a high probability in other analyses as in the mammalian mitochondrial alignment [30], and for the molecular markers *glnII* and *recA* in *Bradyrhizobium* populations [47].

Molecular clock testing

The HKY model with invariant sites and gamma rate distribution (HKY + I + G) was used for all phylogenetic analyses conducted in this study to test the molecular clock models using MrBayes program. On the basis of the Bayes factors (BF), for which the 2ln (BF21) values were positive and greater than 10 (Table 3), the strict-clock hypothesis was rejected for both alignments in favour of the relaxed-clock hypothesis.

Strict (also called global) molecular clock uses the assumption of a constant substitution rate across the phylogenetic tree for each gene, but it does not make any assumption on rate constancy across genes [48]. Reject of the global molecular clock hypothesis indicated a deviation from evolutionary rate constancy among lineages [49]. In contrast, the relaxed molecular clock allows substitution rate to vary

16S rRNA gene		ITS region	
Model'	Posterior probability	Model'	Posterior probability
[121131]	0.222	[121321]	0.179
[123343]	0.085	[121121]	0.171
[121134]	0.054	[121131]	0.108
[121121]	0.053	[121341]	0.082
-	-	[121323]	0.078
-	-	[123121]	0.056

(') Model symbolisation: For equal substitution rates, ($r_{AC} = r_{AG} = r_{AT} = r_{CG} = r_{CT} = r_{GT}$), the model is noted [111111]; for all other models, the index number for the first rate is always 1, and indices are labelled sequentially [30].

Table 2: Best models for the 16S rRNA and ITS sequence alignments (only models with posterior probabilities above 0.050 were retained).

Alignment	ln (HM ₁)	ln (HM ₂)	2ln (BF ₂₁)'
16S rRNA gene	-4495.52	-4437.01	117.02
ITS region	-13764.25	-13703.41	121.68

(') $2\ln (BF_{21}) = 2 [\ln (HM_2) - \ln (HM_1)]$

Table 3: Testing of the molecular clock hypotheses under the 'HKY + I + G' substitution model (HM₁: harmonic mean for the strict-clock model; HM₂: harmonic mean for the relaxed-clock model; BF: Bayes factor).

between lineages [34,50,51]. In the present study, the tree topologies for the 16S rRNA gene and for the ITS region were constructed under the relaxed molecular clock model by using MrBayes program.

16S rRNA gene phylogeny

The Bayesian phylogeny of the twenty 16S rRNA gene sequences showed two main groups (Figure 1). The *Bradyrhizobium japonicum* sequence (AY904772) was used as the outgroup. One group consisted of 11 isolates mainly related to *Agrobacterium* clade, and another group included the remaining 8 isolates related to *Rhizobium* clade. Posterior probability support values for the *Agrobacterium* and *Rhizobium* lineages were 0.8 and 0.7, respectively. These same two lineages have been identified in previous analyses of 16S rRNA sequences [52] regardless of the method of phylogenetic analysis.

There was sufficient sequence variation in the 16S rRNA gene to resolve relationships within these two major groups. The median number of pairwise nucleotide differences among the 11 isolates in *Agrobacterium* was 56 (about 4% of the approximately 1300 bp 16S rRNA region analysed), and producing in the consensus tree two monophyletic clades supported by high posterior probability values, at least of 0.8. One clade was represented by three *Rizobium* species (*R. huautlense*: AM273359, *R. soli*: EF363715, and *Rhizobium* sp.: AB453869), and included our isolate Hc04 (JN944189). The second clade was composed by all sensu stricto *Agrobacterium* species (*A. vitis*: CP000634, *A. tumefaciens*: AJ389901, *A. tumefaciens*: EF620461), and two other *Rhizobium* species (*R. giardinii*: U86344 and *R. pusense*: FJ969841). Our Rhizobial isolates Hc01 (JN944178) and Hcar01 (JN944192) belonged to *Agrobacterium tumefaciens* (EF620461), and to *Rhizobium pusense* (FJ969841), respectively. *Rhizobium pusense*, which formed a monophyletic clade with the strain Hcar01 (Figure1), may be an *Agrobacterium* sp. because it is unable to nodulate chickpea or to induce tumour in tobacco plant [53]. The same situation was described for *Agrobacterium* sp. H13-3 (the former *Rhizobium* lupine H13-3) which is unable to nodulate *Lupinus* under laboratory conditions [54].

Unlike for the *Agrobacterium* group, the median number of pairwise nucleotide differences among the 8 isolates of the *Rhizobium* group (Figure 1) was only 24 (<2% of the 1300 bp 16S rRNA region

analysed). Despite the low sequence variation in the 16S rRNA gene in the *Rhizobium* group, three monophyletic clades were well distinguishable, and supported by high posterior probabilities. One monophyletic clade was represented by isolate Hsc01 (JN944190) and *Rhizobium sullae* (Y10170), or the well known *R. 'hedysari'* isolated from the European cultivated *H. coronarium* [55]. This later clade was a sister to a second monophyletic clade formed by *R. leguminosarum* (FJ595998), *R. tropici* (U89832), and *A. rhizogenes* (GU320290). A third monophyletic clade was composed by three *Rhizobium* species (*R. leguminosarum*: FJ172678, *R. sullae*: EU723702, and *R. alamii*: GU552885).

It should be noted here that in the 16S rRNA phylogram (Figure 1), *R. giardinii* (U86344) as well as the *R. huautelense* clade (AM273359, EF363715, AB453869, and JN944189) were with the members of the *Agrobacterium* clade. Inversely, *Agrobacterium rhizogenes* (GU320290) was included in the *Rhizobium* clade. These "*R. huautelense*-*Agrobacterium*" and "*A. rhizogenes*-*Rhizobium*" clusters were in agreement with previous studies on the 16S rRNA phylogenies [19,20,42,52]. Despite these exceptions, the 16S rRNA gene topology (Figure1) showed a clear split between *Agrobacterium* and *Rhizobium* clades. Among our four 16S rRNA sequences included in the present study, one sequence (Hc04: JN944189) was positioned in the *R. huautelense* cluster, two sequences (Hc01: JN944178 and Hcar01: JN944192) were placed with the members of the *Agrobacterium* (sensu stricto) clade, and one sequence (Hsc01: JN944190) was included in the *Rhizobium* clade.

ITS region phylogeny

A similar topology was obtained from the ITS sequences analysis (Figure 2). The split between the *Agrobacterium* and *Rhizobium* clades was supported by moderate posterior probability values (0.5 and 0.6, respectively). In the *Agrobacterium* group, *A. vitis* (CP000634) branched at the basal node. Isolate Hsc01 (JN944183) was also branched at the basal node in the *Rhizobium* clade.

As for the 16S rRNA analysis, the ITS region phylogenetic tree (Figure 2) showed that the *Agrobacterium* group was more heterogeneous than the *Rhizobium* clade. A monophyletic clade, grouping the *R. giardinii* strains (EU288738, AF345268, AF321871) and Hc04 isolate (JN944182), was included in the *Agrobacterium* group. Isolates Hc01 (JN944179) and Hcar01 (JN944185) belonged to *A. rubi* and *A. tumefaciens*, respectively. In the *Rhizobium* group, *A. rhizogenes* (AF345275) formed a monophyletic clade with *R. hainanense* (AF321872).

There was also sufficient sequence variation in the ITS region to resolve relationships within the two main rhizobial groups. The median number of pairwise nucleotide differences among the 11 isolates in the *Agrobacterium* group was 204, against 193 among the remaining sequences in the *Rhizobium* clade. These values, which were higher than those obtained for 16S rRNA sequences, confirmed that ITS sequencing is an efficient tool for analyzing relatedness between closely related rhizobial strains [19].

Divergence time estimation

The estimation of the dates when species diverged is often perceived to be as important as estimating the phylogeny itself [56]. Due to the absence of an informative fossil record, dating bacterial divergence times is very difficult [57]. Some estimates have been made based on the assumption that eukaryotes and bacteria have comparable substitution rates for all genes [36,37]. Based on 16S rDNA, these authors dated the *Agrobacterium*-*Rhizobium* split of at about 250 million years ago

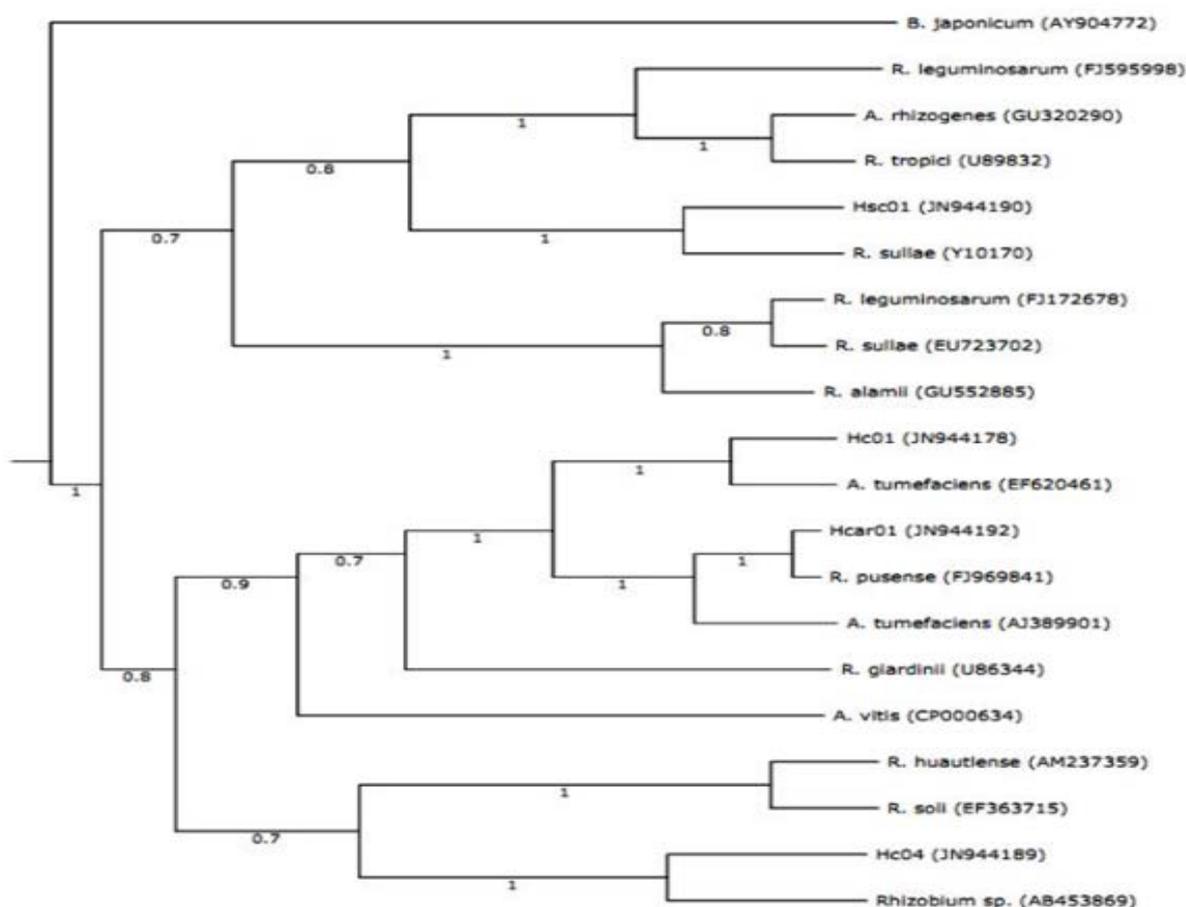


Figure 1: Bayesian inference tree topology based on 16S rRNA gene sequences from *Sulla* nodule isolates and reference strains from GenBank. The Bayesian phylogeny was constructed under the 'HKY+I+G' substitution model, and the molecular relaxed-clock. The tree was rooted on *B. japonicum*. Numbers on nodes indicate posterior probabilities of the Bayesian method. GenBank accession numbers are in parentheses. Abbreviations: *R.*, *Rhizobium*; *A.*, *Agrobacterium*; *B.*, *Bradyrhizobium*.

(110-415 Myr). This information has been used in the present study to estimate divergence times between *Agrobacterium* and *Rhizobium* species, using the 16S rRNA gene (Figure 3) and the ITS region (Figure 4). We particularly estimated the age of the following nodes:

- Node 1: divergence between *Agrobacterium* and *Rhizobium* clades;
- Node 2: divergence of the Hsc01 lineage;
- Node 3: divergence of the Hc04 lineage;
- Node 4: divergence between Hc01 and Hcar01 clades.

Divergence times for these four nodes were estimated separately for the 16S rRNA gene and the ITS region (Table 4). Standard deviations (SD) for the age of nodes, and 95% credible intervals (CI) were also given. Table 4 showed that the variation among date estimates for individual nodes was large between the different molecular data sets. For example, the divergence time of the root (node 1) varied from 329.64 ± 92.88 Myr (for the ITS marker) to 360.41 ± 92.73 Myr (for the 16S rRNA gene). The posterior time estimates obtained from the 16S rRNA gene analysis were slightly older than those from the ITS marker (except for node 4), but the CIs overlap substantially between the two analyses. Overall, the date estimates from both molecular markers (Tables 4) were all earlier than the oldest known legume fossils with an

age of 56 Myr [58]; this confirmed the possibility that rhizobia started to diverge before the existence of leguminous plants [59].

Substitution rate estimation

It has been well established that Multidivtime method by using the relaxed-clock model produced reliable point estimates of time for large numbers of genes [38]. The Multidivtime approach was used as a framework for MCMC estimation of evolutionary rates for the 16S rRNA gene and the ITS region. The date of split between *Agrobacterium* and *Rhizobium* clades was used as the divergence time prior in Multidivtime analyses for calibrating rates of both molecular markers (see Materials and Methods). The Rato program [60] was used to read the Multidivtime main output file, in particular by extracting the average substitution rate for each alignment used. Rate estimates obtained for the 16S rRNA gene and the ITS region were 0.005 and 0.076 substitutions/site per 50 Myr, respectively. For the 16S rRNA gene, our substitution rate estimation was lower than the average substitution rate (0.01 substitutions/site per 50 Myr) in eubacteria [36], but not very different from an estimated substitution rate of 0.012-0.018 substitutions/site per 100 Myr in non-endosymbiotic bacteria [61]. In contrast, the current study showed a substantial difference in evolutionary rates between the 16S rRNA and ITS markers. For the ITS region, the substitution rate was approximately 15-fold higher (0.076

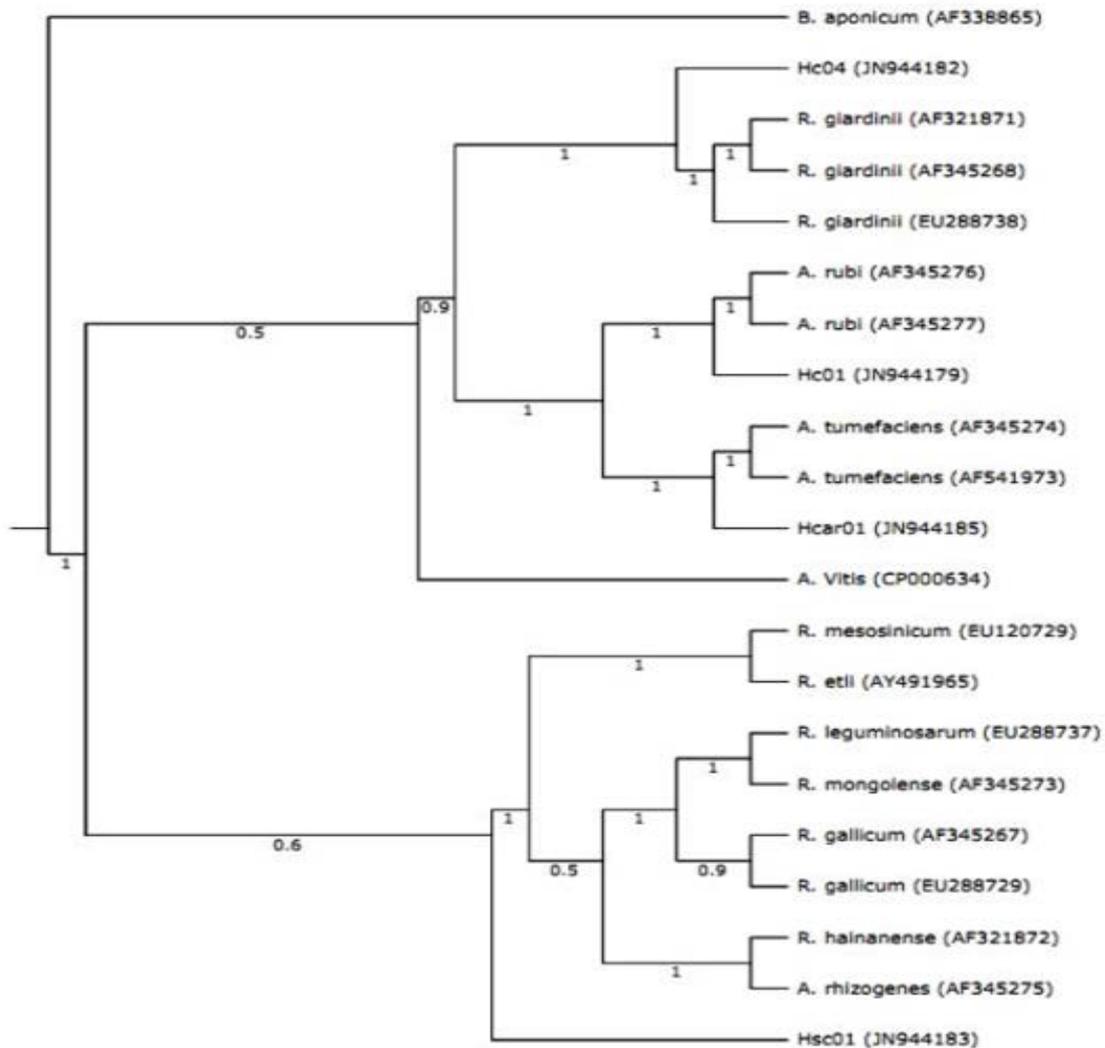


Figure 2: Bayesian inference tree topology based on Internally Transcribed Spacer (ITS) region sequences from *Sulla* nodule isolates and reference strains from GenBank. The Bayesian phylogeny was constructed under the 'HKY+I+G' substitution model, and the molecular relaxed-clock. The tree was rooted on *B. japonicum*. Numbers on nodes indicate posterior probabilities of the Bayesian method. GenBank accession numbers are in parentheses.

versus 0.005 substitutions/sites per 50 Myr for the 16S rRNA marker), indicating that most sites in rRNA genes are under strong selective constraint. This finding confirmed the wide range of substitution rates across genes and bacterial taxa suggested by [18]. In this context, Parker et al. [62] showed that the *nifD* marker had more than six-fold higher evolutionary rate than 16S rRNA. However, the 16S rRNA gene evolved slightly faster in obligate endosymbiotic bacteria, and averaged 0.029-0.045 substitutions per site per 50 Myr in some gamma proteobacterial symbionts [63,64].

Conclusion

This is the first report of rhizobial strains being the predominant root-nodule symbionts for three wild species of the genus *Sulla* in their Tunisian native geographic range. Four isolates, sampled from *S. coronaria* (Hc01 and Hc04), *S. carnososa* (Hcar01), and *S. capitata* (Hsc01), were studied together with reference strains for 16S rRNA gene and ITS region sequences. In both the 16S rRNA gene and

ITS region phylogenies, Tunisian rhizobial strains were placed into different groups. The Hsc01 strain was always placed in the *Rhizobium* group in both molecular marker phylogenies. In the 16S rRNA gene tree, this strain formed a strongly supported monophyletic clade with the well known *R. sulae* of the European cultivated *H. coronarium* (namely *S. coronaria*).

Among the *Agrobacterium* group, the Hc04 strain belonged to *Rhizobium* species such as *R. soli* and *R. huautlense* in the 16S rRNA gene tree, or *R. giardinii* in the ITS marker phylogeny. This strain may represent a novel species of the genus *Rhizobium*. In phylogenetic trees from molecular markers, the strains Hc01 and Hcar01 were always included in the genus *Agrobacterium* sensu stricto composed by the classical species *A. vitis*, *A. rubi* and *A. tumefaciens*.

Despite the heterogeneity within the *Agrobacterium* group, the Bayesian phylogenetic inference showed a clear split between *Rhizobium/Agrobacterium* clades for both 16S rRNA gene and ITS

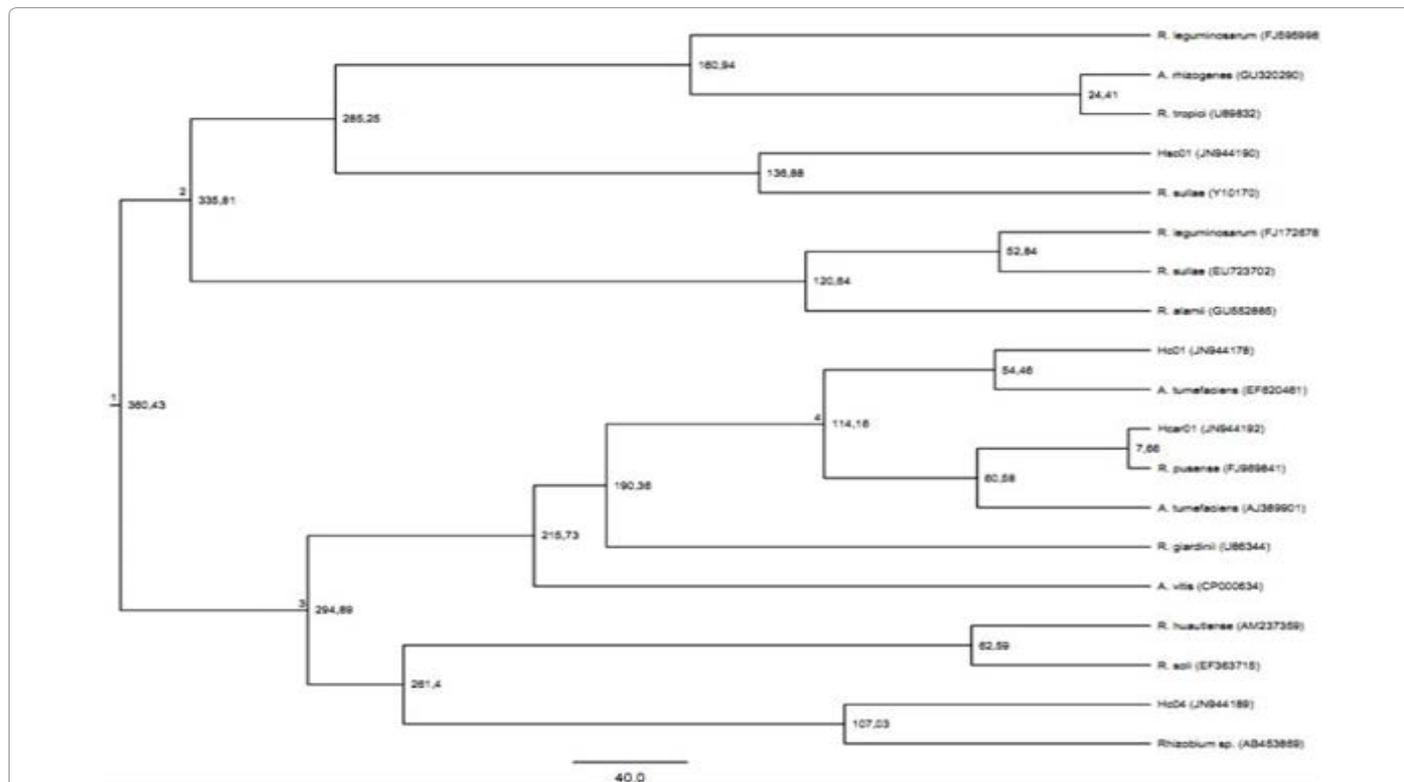


Figure 3: Divergence times among 16S rRNA gene sequences from *Sulla* nodule isolates and reference strains from GenBank. Sequence accession numbers are in parentheses. The *Multidivtime* approach was with a mean prior of 250 ± 125 million years (Myr) for the *Rhizobium-Agrobacterium* split. Date estimates are given for nodes 1-4 in Table 4. The scale bar represents 40 Myr.

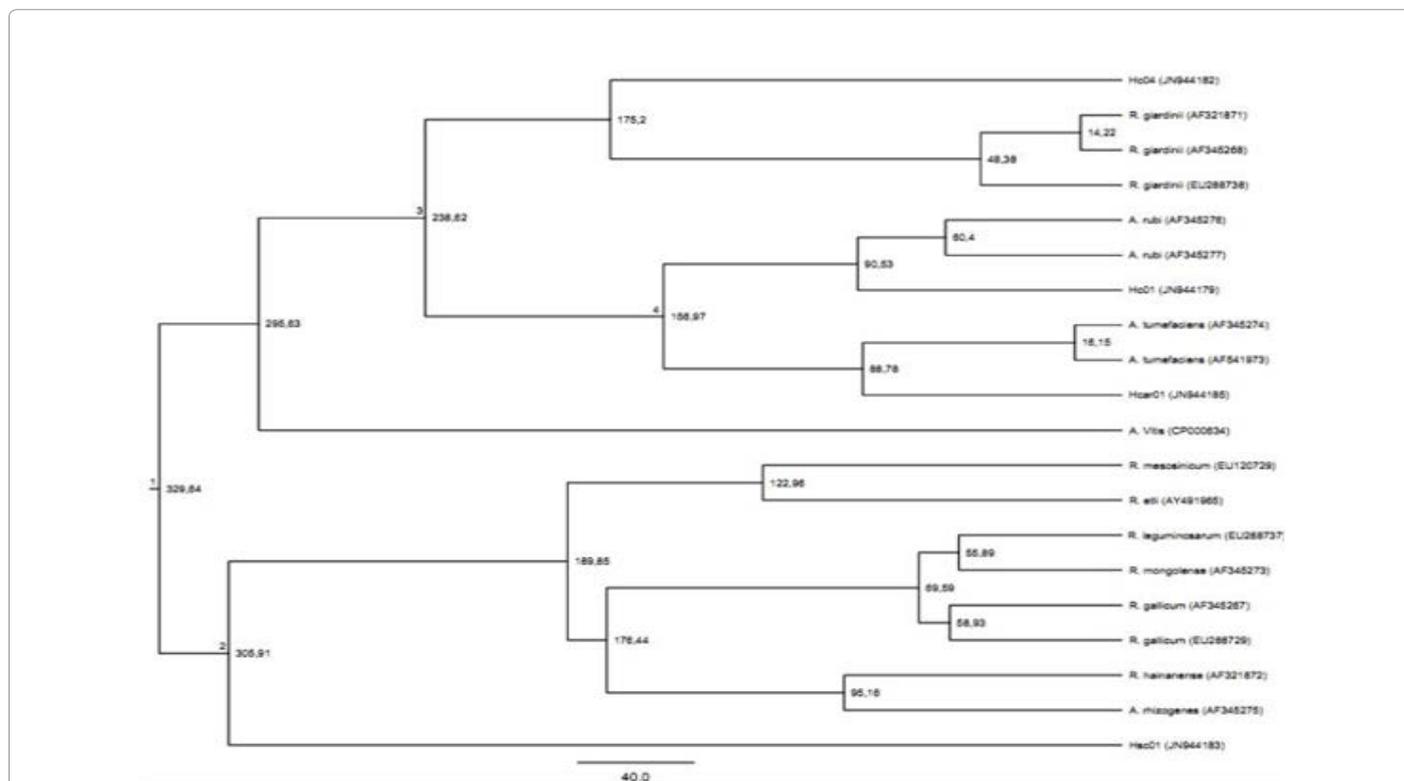


Figure 4: Divergence times among ITS region sequences from *Sulla* nodule isolates and reference strains from GenBank. Sequence accession numbers are in parentheses. The *Multidivtime* approach was with a mean prior of 250 ± 125 million years (Myr) for the *Rhizobium-Agrobacterium* split. Date estimates are given for nodes 1-4 in Table 4. The scale bar represents 40 Myr.

Node	16S rRNA gene			ITS region		
	Age	SD	95% CI	Age	SD	95% CI
1	360.43	92.73	[177.60, 517.09]	329.64	92.88	[164.06, 507.40]
2	335.81	90.23	[162.59, 494.11]	305.91	88.37	[149.95, 480.02]
3	294.89	84.54	[135.72, 454.17]	238.62	72.95	[111.87, 388.08]
4	114.16	53.28	[25.29, 229.16]	156.97	57.24	[64.15, 281.76]

Table 4: Node age, standard deviation (SD), and 95% credible interval (CI) estimates based on 16S rRNA gene and ITS region sequences.

region. By using the approximately date of 250 million years ago as a divergence time prior between these rhizobial clades, the evolutionary rate of the ITS marker was estimated to be about 15-fold than the 16S rRNA gene rate (0.076 versus 0.005 substitutions/site per 50 Myr). This indicated that the ITS marker can serve as a reliable chronometer for bacterial evolution. However, it will be important in future work to extend analyses to additional loci that evolve in more consistent manners such as symbiotic genes (e.g., *nod* and *nif* genes).

References

- Wojciechowski MF, Lavin M, Sanderson MJ (2004) A phylogeny of legumes (Leguminosae) based on analysis of the plastid *matK* gene resolves many well-supported subclades within the family. *Am J Bot* 91: 1846-1862.
- Lafay B, Burdon JJ (1998) Molecular diversity of rhizobia occurring on native shrubby legumes in southeastern Australia. *Appl Environ Microbiol* 64: 3989-3997.
- Wei GH, Zhang ZX, Chen C, Chen WM, Ju WT (2008) Phenotypic and genetic diversity of rhizobia isolated from nodules of the legume genera *Astragalus*, *Lepedeza* and *Hedysarum* in northwestern China. *Microbiol Res* 163: 651-662.
- Kishinevsky BD, Nandasena KG, Yates RJ, Nemas C, Howieson JG (2003) Phenotypic and genetic diversity among rhizobia isolated from three *Hedysarum* species: *H. spinosissimum*, *H. coronarium* and *H. flexuosum*. *Plant Soil* 251: 143-153.
- Chennaoui-Kourda H, Marghali S, Marrakchi M, Trifi-Farah N (2007) Genetic diversity of *Sulla* genus (*Hedysaraceae*) and related species using inter-simple sequence repeat (ISSR) markers. *Bioch Syst Ecol* 35: 682-688.
- Choi B-H, Ohashi H (2003) Generic criteria and an infrageneric system for *Hedysarum* and related genera (*Papilionoideae*-*Leguminosae*). *Taxon* 52: 567-576.
- Zakhia F, Jeder H, Domergue O, Willems A, Cleyet-Marel JC, et al. (2004) Characterisation of wild legume nodulating bacteria (LNB) in the infra-arid zone of Tunisia. *Syst Appl Microbiol* 27: 380-395.
- Mahdhi M, Fterich A, Rejili M, Rodriguez-Llorente ID, Mars M (2012) Legume-nodulating bacteria (LNB) from three pasture legumes (*Vicia sativa*, *Trigonella maritima* and *Hedysarum spinosissimum*) in Tunisia. *Ann Microbiol* 62: 61-68.
- Deng ZS, Zhao LF, Kong ZY, Yang WQ, Lindström K, et al. (2011) Diversity of endophytic bacteria within nodules of the *Sphaerophysa salsula* in different regions of Loess Plateau in China. *FEMS Microbiol Ecol* 76: 463-475.
- Jordan DC (1984) Family III Rhizobiaceae Conn 1938: *Bergeys Manual of Systematic Bacteriology*. Williams and Wilkins, Baltimore, MD, USA.
- Chen WX, Yan GH, Li JL (1988) Numerical taxonomic study of fast-growing soybean rhizobia and proposal that *Rhizobium fredii* be assigned to *Sinorhizobium* gen. nov. *Int J Syst Bacteriol* 38: 392-397.
- de Lajudie P, Willems A, Nick G, Moreira F, Molouba F, et al. (1998) Characterization of tropical tree rhizobia and description of *Mesorhizobium plurifarum* sp. nov. *Int J Syst Bacteriol* 48 Pt 2: 369-382.
- Jarvis BDW, van Berkum P, Chen WX, Nour SM, Fernandez MP et al. (1997) Transfer of *Rhizobium loti*, *Rhizobium huakuii*, *Rhizobium ciceri*, *Rhizobium mediterraneum*, and *Rhizobium tianshanense* to *Mesorhizobium* gen. nov. *Int J Syst Bacteriol* 47: 895-898.
- Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H (2001) A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium undicola* de Lajudie et al 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola*, and *R. vitis*. *Int J Syst Evol Microbiol* 51: 89-103.
- Farrand SK, Van Berkum PB, Oger P (2003) *Agrobacterium* is a definable genus of the family Rhizobiaceae. *Int J Syst Evol Microbiol* 53: 1681-1687.
- Young JM, Kuykendall LD, Martínez-Romero E, Kerr A, Sawada H (2003) Classification and nomenclature of *Agrobacterium* and *Rhizobium*. *Int J Syst Evol Microbiol* 53: 1689-1695.
- Young JM, Park DC (2007) Probable synonymy of the nitrogen-fixing genus *Azotobacter* and the genus *Pseudomonas*. *Int J Syst Evol Microbiol* 57: 2894-2901.
- Kuo CH, Ochman H (2009) Inferring clocks when lacking rocks: the variable rates of molecular evolution in bacteria. *Biol Direct* 4: 35.
- Kwon SW, Park JY, Kim JS, Kang JW, Cho YH, et al. (2005) Phylogenetic analysis of the genera *Bradyrhizobium*, *Mesorhizobium*, *Rhizobium* and *Sinorhizobium* on the basis of 16S rRNA gene and internally transcribed spacer region sequences. *Int J Syst Evol Microbiol* 55: 263-270.
- Iglesias O, Rivas R, García-Fraile P, Abril A, Mateos PF, et al. (2007) Genetic characterization of fast-growing rhizobia able to nodulate *Prosopis alba* in North Spain. *FEMS Microbiol Lett* 277: 210-216.
- Date RA (1982) Collection, isolation, characterisation and conservation of *Rhizobium*: Nitrogen Fixation in Legumes. Academic, New York.
- Muresu R, Polone E, Sulas L, Baldan B, Tondello A, et al. (2008) Coexistence of predominantly nonculturable rhizobia with diverse, endophytic bacterial taxa within nodules of wild legumes. *FEMS Microbiol Ecol* 63: 383-400.
- Marchesi JR, Sato T, Weightman AJ, Martin TA, Fry JC, et al. (1998) Design and evaluation of useful bacterium-specific PCR primers that amplify genes coding for bacterial 16S rRNA. *Appl Environ Microbiol* 64: 795-799.
- Osborn AM, Moore ER, Timmis KN (2000) An evaluation of terminal-restriction fragment length polymorphism (T-RFLP) analysis for the study of microbial community structure and dynamics. *Environ Microbiol* 2: 39-50.
- Navarro E, Simonet P, Normand P, Bardin R (1992) Characterization of natural populations of *Nitrobacter* spp. using PCR/RFLP analysis of the ribosomal intergenic spacer. *Arch Microbiol* 157: 107-115.
- Ponsonnet C, Nesme X (1994) Identification of *Agrobacterium* strains by PCR-RFLP analysis of *pTi* and chromosomal regions. *Arch Microbiol* 161: 300-309.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *J Mol Biol* 215: 403-410.
- Pearson WR, Lipman DJ (1988) Improved tools for biological sequence comparison. *Proc Natl Acad Sci U S A* 85: 2444-2448.
- Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25: 4876-4882.
- Huelsenbeck JP, Larget B, Alfaro ME (2004) Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol* 21: 1123-1133.
- Ronquist F, Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572-1574.
- Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J Mol Evol* 22: 160-174.
- Brown JM, Lemmon AR (2007) The importance of data partitioning and the utility of Bayes factors in Bayesian phylogenetics. *Syst Biol* 56: 643-655.
- Thorne JL, Kishino H, Painter IS (1998) Estimating the rate of evolution of the rate of molecular evolution. *Mol Biol Evol* 15: 1647-1657.
- Yang Z (1997) PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* 13: 555-556.
- Ochman H, Wilson AC (1987) Evolution in bacteria: evidence for a universal substitution rate in cellular genomes. *J Mol Evol* 26: 74-86.

37. Moran NA, Munson MA, Baumann P, Ishikawa H (1993) A molecular clock in endosymbiotic bacteria is calibrated using the insect hosts. *Proc R Soc Lond B* 253: 167-171.
38. Battistuzzi FU, Filipinski A, Hedges SB, Kumar S (2010) Performance of relaxed-clock methods in estimating evolutionary divergence times and their credibility intervals. *Mol Biol Evol* 27: 1289-1300.
39. Rutschmann F (2005) Bayesian Molecular Dating Using PAML/Multidivtime A Step-by-Step Manual, Version 15 University of Zurich, Switzerland.
40. Battistuzzi FU, Billing-Ross P, Paliwal A, Kumar S (2011) Fast and slow implementations of relaxed-clock methods show similar patterns of accuracy in estimating divergence times. *Mol Biol Evol* 28: 2439-2442.
41. Schloss PD, Handelsman J (2005) Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Appl Environ Microbiol* 71: 1501-1506.
42. Valverde A, Igual JM, Peix A, Cervantes E, Velázquez E (2006) *Rhizobium lusitanum* sp. nov. a bacterium that nodulates *Phaseolus vulgaris*. *Int J Syst Evol Microbiol* 56: 2631-2637.
43. Wang FQ, Wang ET, Zhang YF, Chen WX (2006) Characterization of rhizobia isolated from *Albizia* spp. in comparison with microsymbionts of *Acacia* spp. and *Leucaena leucocephala* grown in China. *Syst Appl Microbiol* 29: 502-517.
44. Liu J, Wang ET, Ren da W, Chen WX (2010) Mixture of endophytic *Agrobacterium* and *Sinorhizobium meliloti* strains could induce nonspecific nodulation on some woody legumes. *Arch Microbiol* 192: 229-234.
45. Zakhia F, Jeder H, Willems A, Gillis M, Dreyfus B, et al. (2006) Diverse bacteria associated with root nodules of spontaneous legumes in Tunisia and first report for *nifH*-like gene within the genera *Microbacterium* and *Starkeya*. *Microb Ecol* 51: 375-393.
46. Benhizia Y, Benhizia H, Benguedouar A, Muresu R, Giacomini A, et al. (2004) Gamma proteobacteria can nodulate legumes of the genus *Hedysarum*. *Syst Appl Microbiol* 27: 462-468.
47. Stepkowski T, Moulin L, Krzyżńska A, McInnes A, Law IJ, et al. (2005) European origin of *Bradyrhizobium* populations infecting *Lupinus* and *Serradella* in soils of Western Australia and South Africa. *Appl Environ Microbiol* 71: 7041-7052.
48. Blair Hedges S, Kumar S (2003) Genomic clocks and evolutionary timescales. *Trends Genet* 19: 200-206.
49. Toleno DM, Durbin ML, Lundy KE, Clegg MT (2010) Extensive evolutionary rate variation in floral color determining genes in the genus *Ipomoea*. *Plant Species Biol* 25: 30-42.
50. Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol* 7: 214.
51. Yang Z, Rannala B (2006) Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol* 23: 212-226.
52. Eardly BD, Nour SM, van Berkum P, Selander RK (2005) Rhizobial 16S rRNA and *dnaK* genes: mosaicism and the uncertain phylogenetic placement of *Rhizobium galegae*. *Appl Environ Microbiol* 71: 1328-1335.
53. Panday D, Schumann P, Das SK (2011) *Rhizobium pusense* sp. nov., isolated from the rhizosphere of chickpea (*Cicer arietinum* L.). *Int J Syst Evol Microbiol* 61: 2632-2639.
54. Wibberg D, Blom J, Jaenicke S, Kollin F, Rupp O, et al. (2011) Complete genome sequencing of *Agrobacterium* sp. H13-3, the former *Rhizobium lupini* H13-3, reveals a tripartite genome consisting of a circular and a linear chromosome and an accessory plasmid but lacking a tumor-inducing Ti-plasmid. *J Biotechnol* 155: 50-62.
55. Squartini A, Struffi P, Döring H, Selenska-Pobell S, Tola E, et al. (2002) *Rhizobium sullae* sp. nov. (formerly '*Rhizobium hedysari*'), the root-nodule microsymbiont of *Hedysarum coronarium* L. *Int J Syst Evol Microbiol* 52: 1267-1276.
56. Aris-Brosou S, Xia X (2008) Phylogenetic analyses: A toolbox expanding towards Bayesian methods. *Int J Plant Genomics* 2008: 683509.
57. Turner SL, Young JP (2000) The glutamine synthetases of rhizobia: phylogenetics and evolutionary implications. *Mol Biol Evol* 17: 309-319.
58. Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. *Syst Biol* 54: 575-594.
59. Young JP, Johnston AW (1989) The evolution of specificity in the legume-rhizobium symbiosis. *Trends Ecol Evol* 4: 341-349.
60. Erikson T, Rutschmann F (2005) Rato program version 13.
61. Clark MA, Moran NA, Baumann P (1999) Sequence evolution in bacterial endosymbionts having extreme base compositions. *Mol Biol Evol* 16: 1586-1598.
62. Parker MA, Lafay B, Burdon JJ, van Berkum P (2002) Conflicting phylogeographic patterns in rRNA and *nifD* indicate regionally restricted gene transfer in *Bradyrhizobium*. *Microbiology* 148: 2557-2565.
63. Clark MA, Moran NA, Baumann P, Wernegreen JJ (2000) Cospeciation between bacterial endosymbionts (*Buchnera*) and a recent radiation of aphids (*Uroleucon*) and pitfalls of testing for phylogenetic congruence. *Evolution* 54: 517-525.
64. Bressan A, Arneodo J, Simonato M, Haines WP, Boudon-Padiou E (2009) Characterization and evolution of two bacteriome-inhabiting symbionts in cixiid planthoppers (Hemiptera: Fulgoromorpha: Pentastirini). *Environ Microbiol* 11: 3265-3279.