Genomic identification of rhizobia-related strains and threshold of ANI and core-genome for family, genus and species

Qian Wang[1], Wentao Zhu[2], Entao Wang[3], Linshuang Zhang[4], Xiangyang Li[5], Gejiao Wang[6]*

[1,2,4,5,6]State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, P. R. China
*Email: gejiao@mail.hzau.edu.cn
[3]Departamento de Microbiología, Escuela Nacional de Ciencias Biológicas, Instituto Politécnico Nacional, México D. F. 11340, Mexico
Email: entaowang@yahoo.com.mx

Abstract*—Aiming at accurately and rapidly identifying our heavy metal resistant rhizobial strains, genomic average nucleotide identity (ANI) and core genome analyses were performed to investigate the phylogenetic relationships among 45 strains in the families of Rhizobiaceae and Bradyrhizobiaceae. The results showed that both of the ANI and core-genome phylogenetic trees revealed similar relationship. In ANI analysis, the 90%, 75% and 70% ANI values could be the thresholds for species, genus and family, respectively. Analyzing the genomes using multi-dimensional scaling and scatter plot showed highly consistent with the ANI and core-genome phylogenetic results. With these thresholds, the 45 strains were divided into 24 genomic species within the genera Agrobacterium, Allorhizobium, Bradyrhizobium, Sinorhizobium and a putative novel genus represented by Ag. albertimagni AOL15. The ten arsenite-oxidizing and antimonite tolerant strains were identified as Ag. radiobacter, and two Sinorhizobium genomic species differing from S. fredii. In addition, the description of Pararhizobium is questioned because ANI values greater than 75% were detected between P. giardinii H152T and Sinorhizobium strains. Also, reversion of the species definition for several strains in R. etli and R. leguminosarum was suggested. Our results demonstrate that analyses of ANI and core-genome are rapid and confident methods to identify the rhizobial strains, and it will be also convenient when more genome data are accumulated.*

***Keywords*—Antimonite tolerance, arsenite-oxidation, genome, phylogeny, Rhizobia.**

I. INTRODUCTION

It is well known that the symbiotic bacteria (rhizobia) and the tumor-inducing phytopathogenic bacteria (agrobacteria) in Rhizobiaceae family are phylogenetically intermingled in some genera, even in the same species. Originally, the symbiotic bacteria were all grouped within the genus *Rhizobium*, which was established in 1890 with *Rhizobium leguminosarum* as the type species [1, 2]; and the tumor-inducing phytopathogenic bacteria were designed as the genus *Agrobacterium* which was first proposed by Conn including *Agrobacterium tumefaciens* (tumor-inducing), *Agrobacterium radiobacter* (no tumor) and *Agrobacterium rhizogenes* (hairy root) based on their phytopathogenic symptoms [3]. Later, *Agrobacterium rubi* (from *Rubiaceae* plants), *Agrobacterium vitis* (from *Vitis* plants) and *Agrobacterium larrymoorei* (from *Ficus* plants) were established [4-6], which were divided into Biovars I, II and III [7]. Based upon the phylogeny of 16S rRNA gene, the genus *Agrobacterium* and a later described genus *Allorhizobium* [8] were officially immerged into Rhizobium [9]. However, this combination caused frequently argument because their different affection on plants, and their divergent phylogenetic relationships of 16S rRNA, 23S rRNA and *recA* genes [10-14], as well as the fatty acid profiles [15]. With description of more and more symbiotic and non-symbiotic species in the combined genus *Rhizobium*, its polyphylic feature was further apparent.

Meanwhile, some novel molecular techniques have been developed for estimating the phylogenetic relationships, such as the multilocus sequence analysis (MLSA) and whole genome sequencing. Recently, the taxonomy of *Agrobacterium/Rhizobium* group was dramatically revised again based upon the MLSA data of four or six protein-coding housekeeping genes [16-17], which led the split of *Agrobacterium*/*Rhizobium* group into five sister genera, *Agrobacterium*, *Allorhizobium*, *Neorhizobium*, *Pararhizobium* and *Rhizobium*. In the recently emended *Agrobacterium* genus, *Ag. radiobacter* and *Ag. rubi* are phytopathogenic species, while *Ag. nepotum*, *Ag. pusense*, and *Ag. skierniewicense* were new combinations transferred from the former *Rhizobium* species. The emended *Allorhizobium* covered the phytopathogenic species *Al. vitis* (formerly *Agrobacterium vitis*), and the symbiotic or endophytic species *Al. taibaishanense*, *Al. paknamense*, *Al. oryzae*, *Al. pseudoryzae* and *Al. borbori*. The genus *Neorhizobium* included the species *N. galegae*, *N. vignae*, *N. huautlense* and *N. alkalisoli* transferred from the former *Rhizobium* species [16]. *Pararhizobium* included *P. giardinii*, *P. capsulatum*, *P. herbae* and *P. sphaerophysae* [17], which were all transferred from the former *Rhizobium* species. After the reversion, the species

represented by *Rhizobium leguminosarum* are maintained in the genus *Rhizobium*, and the phytopathogenic species *R. rhizogenes* (former *Agrobacterium rhizogenes*) was also included in this genus.

Despite the nomenclature change or taxonomic reversion, the pathogenic (for plants and human being), symbiotic, endophytic and saprophytic bacterial species are intermingled in the five *Agrobacterium*/*Rhizobium* sister genera [16-18]. Furthermore, these four living states or characters even can be found in the single species *Ag. radiobacter* [19] or in the same strains of *R. rhizogenes* [20]. Although the recent reversions have resolved the nomenclature argument about the symbiotic *Rhizobium* species and the phytopathogenic *Agrobacterium* species, the phylogenetic relationships between the symbiotic species and phytopathogenetic species were still not sufficiently revealed because only several housekeeping genes have been considered [16-17]. To obtain an insight view in the phylogenetic relationships among the members of *Agrobacterium*/*Rhizobium*, the whole genome comparison would be valuable.

Previously, we isolated some arsenite-oxidizing or antimonite tolerant strains and they were primitively identified as unnamed species within *Agrobacterium* and *Sinorhizobium* based on the 16S rRNA gene sequence analyses [21-23]. Aiming at further identifying them, as well as developing a rapid, confident/stable, high-throughput identification method, we performed this study by using the genome data. In particular, the average nucleotide identity (ANI) and core-genome [24] were estimated to ascertain the phylogenetic relationships among the 45 strains in the family Rhizobiaceae. The results offered accurate identification of our test strains and generated some valuable taxonomic clues.

II. MATERIAL AND METHOD

2.1 Genomic information

In total, 45 available genome sequences were used in this study (See Supplementary Table S1 for details), in which 34 were extracted in January, 2015 from the NCBI GenBank, including 31 *Rhizobium-Agrobacterium* strains, one *Sinorhizobium* strain, and two *Bradyrhizobium* strains, which were originally isolated from agricultural soils, root nodules, plant tumors, heavy metal-contaminated soil, or saline desert soil (Table S1). In addition, 11 genomes covering nine arsenite-oxidizing strains of *Agrobacterium* (6) and *Sinorhizobium* (3), and an antimonite tolerant *Sinorhizobium* strain isolated in our previous studies [21-23], and a type strain *Agrobacterium radiobacter* DSM30147^T were sequenced in this study in Shanghai Majorbio Bio-Pharm Technology Co., Ltd. The NCBI GenBank accession numbers for the genomic sequences of the 45 strains are shown in the supplementary Table S1. Genome annotations of these strains were performed through the NCBI Prokaryotic Genome Annotation Pipeline (http://www.ncbi.nlm.nih.gov/genome/ annotation_prok).

2.2 Phylogenetic analysis based on 16S rRNA genes (rrs)

To determine the phylogenetic relationship among the 45 selected strains, the *rrs* sequences were either taken from single *rrs* gene in the GenBank or retrieved from the genome sequences. The distance between strains was calculated using the neighbor-joining (NJ) method and a phylogenetic tree was reconstructed with the Mega 5.05 software [25].

2.3 Phylogenomic analysis based on core-genome sequences

To assess genome diversity, all the coding sequences (CDSs) of the 45 genomes were merged together and the core-genome sequences were searched against themselves based on the BlastP algorithm, with a cutoff of 50% protein identity and 70% of the whole sequences [26]. For the phylogenomic analysis, each set of the converged core CDSs was aligned with ClustalW. Then, all alignments were cascaded into a string of amino acid sequences, and a NJ tree with 1,000 bootstrap was assembled using the Mega 5.05 program [25].

2.4 Phylogenomic analysis based on average nucleotide identity (ANI) values

The ANI values between each pair of genomes among the 45 strains were calculated by the JSpecies software [27] according to the instructions. Based on the pairwise ANI values, a lower left matrix was constructed to represent the pairwise distance (defined as 100% - ANI) and the matrix was used to assemble an ANI divergence dendogram with the method of neighbor-joining (NJ) in the Mega 5.05 program [25].

2.5 Multidimensional scaling (MDS) and scatter plot analyses based on pairwise ANI values

It is widely accepted that high ANI values represent close relationships in taxonomy [27]. Using the SPSS program [28] the MDS [29] algorithm was applied to place each object in 45-dimensional spaces and to ensure that the pairwise distances were well preserved. Each point was then assigned coordinates in each of the 45 dimensions, and, finally, the perceptual mapping was shown in two dimensions. The scatter diagram, which was based on the coordinates calculated by MDS, was constructed

using the Excel program. In addition, another scatter diagram was created, which was based on the pairwise average genome size versus the pairwise ANI values, using the Excel program.

TABLE S1
GENERAL GENOMIC INFORMATION OF THE 45 STRAINS USED IN THIS STUDY.

Species	Isolution source	Genome size	GC content	Predicted CDs	Accession No.	Level
Agrobacterium sp. C13*	Soil	5.64	59.8	5303	ASYD00000000	draft
Agrobacterium sp. D14*	Arsenic-enriched soil	5.54	59.8	5186	ASXX00000000	draft
Agrobacterium sp. JL28*	Antimony mine	5.65	59.8	5326	ASXZ00000000	draft
Agrobacterium sp. LY4*	Soil	5.64	59.8	5324	ASYA00000000	draft
Agrobacterium sp. TS43*	Arsenic-enriched soil	5.65	59.8	5368	ASYB00000000	draft
Agrobacterium sp. TS45*	Arsenic-enriched soil	5.64	59.8	5310	ASYC00000000	draft
Ag. tumefaciens 5A	Arsenic-enriched soil	5.74	58.6	5520	AGVZ00000000	draft
Ag. tumefaciens GW4	Arsenic polluted soil	5.64	59.8	5131	AWGV01000000	draft
***Ag. radiobacter* DSM 30147^{T*}**	Soil	7.18	59.9	6834	ASXY00000000	draft
Ag. tumefaciens C58	Cherry tree tumor	5.67	59.1	5355	GCA_000092025	complete
Ag. tumefaciens Cherry 2E-2-2	Crown gall	5.43	59.9	5045	APCC00000000	draft
Ag. tumefaciens CCNWGS0286	Zinc-lead mine tailing	5.21	59.5	4985	AGSM00000000	draft
Ag. tumefaciens F2	Soil	5.47	59.5	5321	AFSD00000000	draft
Agrobacterium sp. ATCC 31749	Soil	5.46	59	5536	AECL00000000	draft
Agrobacterium sp. H13-3	Rhizosphere	5.57	58.5	5345	GCA_000192635	complete
Agrobacterium sp. 224MTsu3.1	Soil	4.8	59.7	4593	ARQL00000000	draft
Ag. albertimagni AOL15	Arsenite oxidizing biofilm	5.09	61.2	4811	ALJF00000000	draft
Allorhizobium vitis S4	Vitis vinifera nodule	6.32	57.5	5389	GCA_000016285	complete
R. etli 8C-3	Root nodule	3.47	61.1	5076	ABRA00000000	draft
***R. etli* CFN 42^{T}**	Phaseolus vulgaris nodule	6.53	61.1	5963	GCA_000092045	complete
R. etli Kim 5	Root nodule	4.14	61.1	5963	ABQY00000000	draft
R. freirei PRF 81	Bean nodule	7.08	59.9	6271	AQHN00000000	draft
***R. grahamii* CCGE 502^{T}**	Root nodule	7.15	59.4	7368	AEYE00000000	draft
***R. gallicum bv. gallicum* R602^{T}**	Phaseolus vulgaris nodule	7.22	60.8	7152	GCA_000816845	complete
R. lupini HPC(L)	Saline desert soil	5.27	59.2	4615	AMQQ00000000	draft
R. leguminosarum bv. trifolii WSM597	Trifolium polymorphum nodule	7.63	61	7159	AKHZ00000000	draft
R. leguminosarum bv. trifolii SRDI565	Trifolium polymorphum nodule	6.91	60.7	6870	AQUD00000000	draft
R. leguminosarum bv. phaseoli 4292	Bean nodule	7.35	60.7	7255	AQZR00000000	draft
R. leguminosarum bv. viciae TOM	Legume root nodule	7.36	60.8	7298	AQUC00000000	draft

R. leguminosarum bv. viciae WSM1481	Legume root nodule	7.56	61	7548	AQUM00000000	draft
R. leguminosarum bv. viciae 3841	Legume root nodule	7.75	60.9	7131	GCA_000009265	complete
R. phaseoli Ch24-10	Root nodule	6.62	61.3	6512	AHJU00000000	draft
R. rhizogenes K84	Plant root soil	7.27	59.9	6285	GCA_000016265	complete
***R. tropici* CIAT 899^{T}**	Phaseolus vulgaris nodule	6.69	59.5	6287	GCA_000330885	complete
Rhizobium sp. 42MFCr.1	Arabidopsis thaliana rhizosphere	6.21	59.9	6332	ARHV00000000	draft
Rhizobium sp. AP16	Populus deltoides rhizosphere	6.5	60.2	6123	AJVM00000000	draft
Rhizobium sp. CF142	Populus deltoides rhizosphere	7.46	60.1	7229	AJWE00000000	draft
***Pararhizobium giardinii* H152^{T}**	Phaseolus vulgaris nodule	6.81	60.7	6782	ARBG00000000	draft
***S. fredii* USDA 205^{T}**	Soybean nodule	7.01	62.3	6436	AUTC00000000	draft
Sinorhizobium sp. GL2*****	Arsenic polluted soil	7.05	62.1	7586	AUTB00000000	draft
Sinorhizobium sp. GL28*	Arsenic polluted soil	8.45	61.6	7431	AUSZ00000000	draft
Sinorhizobium sp. GW3*	Arsenic polluted soil	7.36	62	7450	AUSY00000000	draft
Sinorhizobium sp.Sb3*	Coalmine	6.08	61.6	7706	AUTA00000000	draft
Bradyrhizobium diazoefficiens USDA 110	Soybean nodule	9.11	64.1	8373	GCA_000011365	complete
Bradyrhizobium japonicum USDA 6	Soybean nodule	9.21	63.7	8826	GCA_000284375	complete

****The strains isolated and sequenced in this study. The type strains are in bold.***

III. RESULTS

3.1 General genomic features of the involved strains

For the 18 strains previously classified into the genus *Agrobacterium*, three complete genomes (*Ag. tumefaciens* C58, *Agrobacterium*-like sp. H13-3 and *Al. vitis* S4) and 15 draft genomes (including six obtained in this study) were obtained. For the 19 *Rhizobium* strains five complete genomes (*R. etli* CFN 42^{T}, *R. leguminosarum bv. viciae* 3841, *R. tropici* CIAT 899^{T}, *R. rhizogenes* K84, and *R. gallicum* R602sp^{T}), and 14 draft genomes (including the type strain *R. grahamii* CCGE 502^{T}) were found. In addition, draft genomes were also obtained for *P. giardinii* H152^{T}, five *Sinorhizobium* strains (included the type strain *S. fredii* USDA205^{T}) and two *Bradyrhizobium* strains *B. diazoefficiens* USDA110^{T} and *B. japonicum* USDA6^{T}. The GC content range of the 45 strains is 57.5 - 64.1%. The genome sizes vary from 3.47 (*R. etli* 8C-3) to 9.21 Mb (*Bradyrhizobium japonicum* USDA6^{T}), whereas the number of predicted CDSs vary from 4593 (*Agrobacterium* sp. 224MTsu3.1) to 8826 (*Bradyrhizobium japonicum* USDA6^{T}).

3.2 Phylogenetic relationship based on rrs sequences

A NJ phylogenetic tree based on the *rrs* genes of the 45 strains (available as Fig. S1) revealed that the strains belonging to *Agrobacterium tumefaciens* were separated into two branches and *Allorhizobium vitis* S4 was interfused among the *Ag. tumefaciens* strains. In addition, the *Ag. radiobacter* DSM 30147^{T} was clustered with *Rhizobium* sp. PRF 81, *R. tropici* CIAT 899^{T}, *Rhizobium* sp. AP16 and *R. rhizogenes* K84 (Fig. S1). The *Rhizobium* sp. CF142 was clustered in genus *Agrobacterium*, while *Rhizobium lupini* HPC(L) was grouped into *Bradyrhizobium* (Fig. S1).

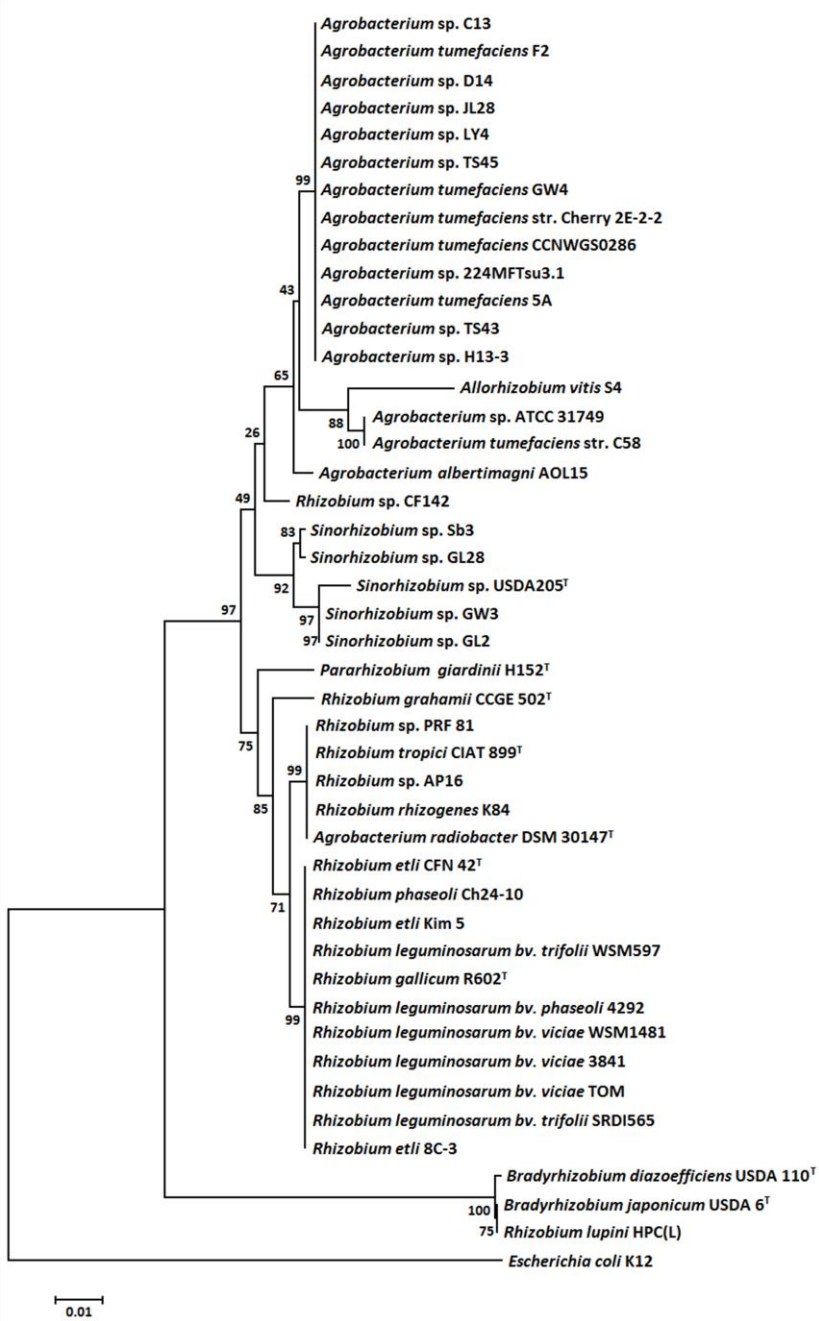

FIG. S1. A NJ PHYLOGENETIC TREE BASED ON 16S RRNA GENE SEQUENCES (RRS). THE TREE WAS BUILT FOR 45 *RHIZOBIUM* FAMILY STRAINS, WHICH INCLUDES SIX TYPE STRAINS. THE AVERAGE LENGTH OF THESE 16S RRNA GENE SEQUENCES IS 1,389 BP. HORIZONTAL BRANCH LENGTHS ARE PROPORTIONAL TO THE ESTIMATED NUMBER OF NUCLEOTIDE SUBSTITUTIONS, AND BOOTSTRAP PROBABILITIES (AS PERCENTAGES) ARE DETERMINED FROM 1000 RE-SAMPLINGS. THE 16S RRNA GENE SEQUENCE OF *ESCHERICHIA COLI* K12 WAS USED AS THE REFERENCE.

3.3 Phylogenomic relationship based on the core-genome sequences

Using the cutoff of 50% protein identity and 70% of the whole sequences, 313 core-genome CDSs were identified for the 45 strains. In the phylogenomic tree based on the core-genome (Fig. 1), the tested strains were grouped into six lineages, including 1) *Ag. radiobacter*/*tumefaciens* (Biovar I)-*R. lupini* HPC(L) lineage; 2) *Ag. albertimagni* (Biovar III) lineage; 3) *Allorhizobium vitis* (former *Ag. vitis*) lineage; 4) *Rhizobium* lineage covering *R. leguminosarum*, *R. etli*, *R. phaseoli*, *R. gallicum*, *R. tropici*, *R. freirei*, *R. grahamii* and *R. rhizogenes* (former *Ag. rhizogenes*); 5) *Pararhizobium giardinii* (former *R. giardinii*) and *Sinorhizobium* lineage; and 6) *Bradyrhizobium* lineage.

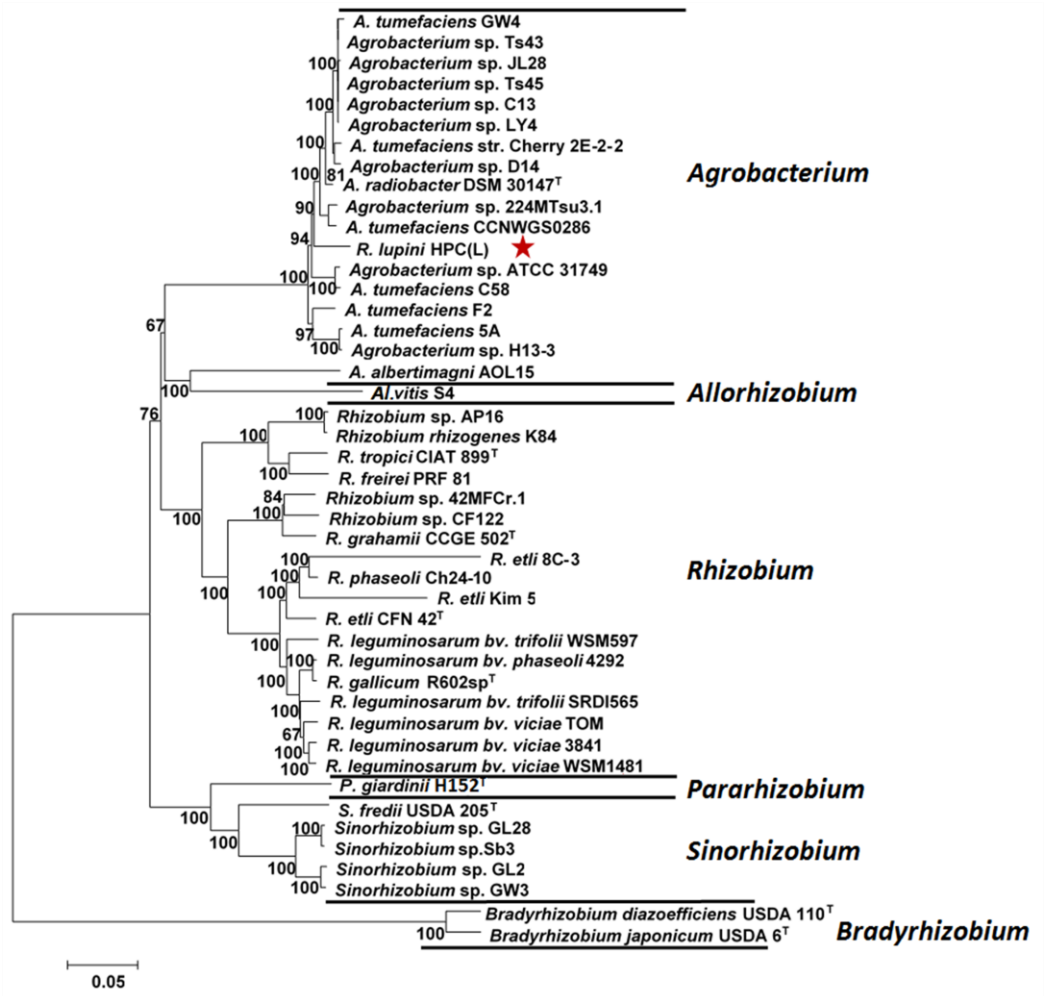

FIG. 2. A NJ PHYLOGENOMIC TREE BASED ON THE AVERAGE NUCLEOTIDE IDENTITY (ANI) FOR THE TESTED STRAINS IN THIS STUDY (TABLE S2). ALL OF THE STRAINS COULD BE CLEARLY DIVIDED INTO *AGROBACTERIUM*, *RHIZOBIUM*, *ALLORHIZOBIUM*, *PARARHIZOBIUM*, *SINORHIZOBIUM* AND *BRADYRHIZOBIUM* BRANCH. THE STRAINS *R. LUPINI* HPC(L) WAS LABELED BY RED STARS, SINCE THEY CLUSTERED WITH AGROBACTERIUM GROUP.

3.4 Phylogenomic relationship based on ANI values

The ANI values between each pair of genomes were calculated and 990 ANI values were obtained for the 45 strains (Table S2). In the NJ phylogenomic tree constructed with the ANI data, the 45 strains were also divided into six lineages (Fig. 2), same as the lineages defined with the core-genome (Fig. 1). The members in distinct families, Rhizobiaceae and Bradyrhizobiacea, showed 66.00-68.01 % ANI and the strains within Rhizobiaceae presented ANI >70.54. The ANI values were lower than 75% among different genera in family Rhizobiaceae, except *Pararhizobium* that presented 75.16-76.22% ANI with the *Sinorhizobium* strains (Table S2). At 90% ANI value, all the type strains for the defined species in the genus *Rhizobium* were separated and the 45 strains could be delineated into 24 genomic species (Fig. 2, also Table S2). 1) Among the 17 strains belonging to *Agrobacterium*, 11 were identified as *Ag. radiobacter*, including all the six tested arsinite-oxidizing strains; while 5 strains and *R. lupini* HPC(L) represented six distinct *Agrobacterium* genomic species (ANI < 90% with the other *Agrobacterium* strains); and the last strain *Ag. albertimagni* AOL15 was a very divergent lineage sharing ANI of 72.42-73.18% with the other *Agrobacterium* strains. 2) For the 18 *Rhizobium* strains (except the *R. lupini* strain), *R. phaseoli* Ch24-10, *R. etli* 8C-3 and *R. etli* Kim5 formed a genomic species; the six *R. leguminosarum* strains and *R. gallicum* R602^T formed another genomic species; *R. rhizogenes* K84 and *Rhizobium* sp. AP16 represented the third genomic species; while the other six strains were single lineages corresponding to *R. etli*, *R. tropici*, *R. freirei*, *R. grahamii* and 2 unnamed species. 3) For the genus *Sinorhizobium*, strains GL28 and Sb3 form the sp. I; while GW3 and GL2 formed sp. II; both were different from the type strain *S. fredii* USDA 205^T. 4) *Pararhizobium giardinii* H152^T was grouped in *Sinorhizobium* as the most divergent lineage (ANI > 75% with the *Sinorhizobium* strains). 5) The two *Bradyrhizobium* strains were two lineages corresponding to *B. japonicum* and *B. diazoefficiens*, respectively. 6) The remaining genospecies were *Allorhizobium vitis* S4 (Figs. 1 and 2).

3.5 Similarity levels using MDS and scatter plot analyses based on pairwise ANI values

In the MDS scatter diagram (Fig. 3), the 45 genomes (represented by 45 spots) were clearly separated into five groups. 1) Eighteen strains within the *Rhizobium* formed a group located on the upper right side (except *R. lupini*); 2) 16 strains within *Agrobacterium* group (except *Ag. albertimagni* AOL15) and *Rhizobium lupini* HPC(L) are located on the upper left side of the vertical axis; 3) five strains of *Sinorhizobium* group together with *P. giardinii* are distributed near the vertical axis; *Ag. albertimagni* is near them; 4) two *Bradyrhizobium* strains are a group located on the bottom right side of the vertical axis (Fig. 3); 5) *Al. vitis* S4 occupied a unique position differed from all the other groups (Fig. 3).

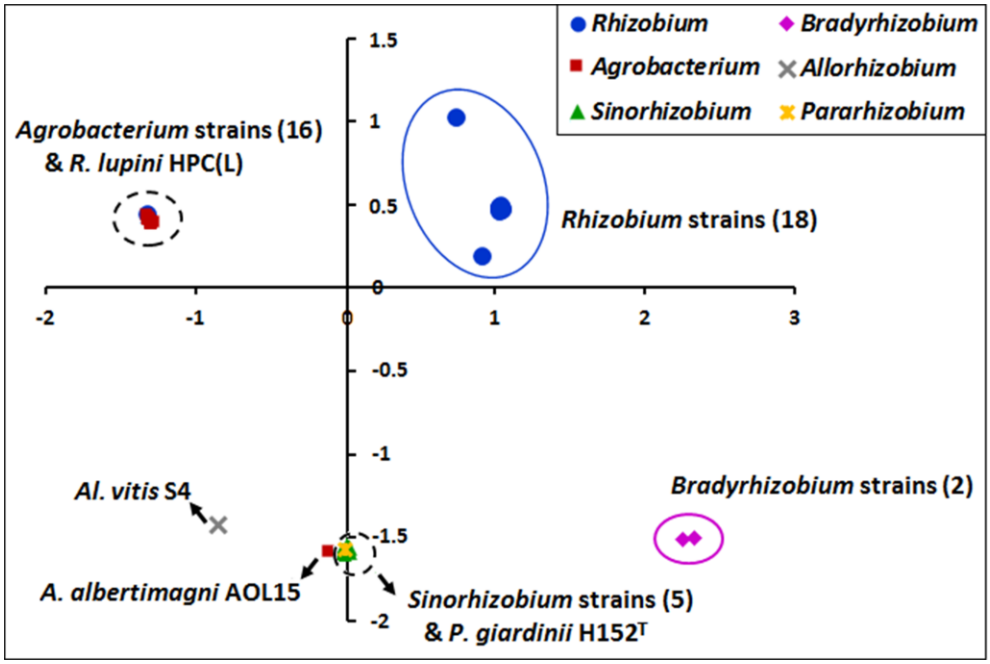

FIG. 3. THE MULTIDIMENSIONAL SCALING (MDS) ANALYSIS BASED ON THE PAIRWISE ANI VALUES. EACH POINT REPRESENTS A SINGLE STRAIN, AND THE DISTANCE BETWEEN TWO POINTS REPRESENTS THE RELATIVE GENETIC DISTANCE BETWEEN THE TWO STRAINS. THE STRAINS ARE DIVIDED INTO SIX GROUPS, *AGROBACTERIUM*, *RHIZOBIUM*, *ALLORHIZOBIUM*, *PARARHIZOBIUM*, *SINORHIZOBIUM* AND *BRADYRHIZOBIUM*, WHICH ARE INDICATED BY BLUE (*RHIZOBIUM*), RED (*AGROBACTERIUM*), GREY (*ALLORHIZOBIUM*), YELLOW (*PARARHIZOBIUM*), GREEN (*SINORHIZOBIUM*) AND PINK (*BRADYRHIZOBIUM*), RESPECTIVELY. THE NUMBER OF STRAINS IN EACH GROUP WAS LABELED.

To further determine the similarity level of strains within each genus, another scatter plot analysis of the 45 strains was performed based on the 990 pairwise ANI values (Fig. 4A). Since only one strain belonged to each of the genera *Allorhizobium* and *Pararhizobium*, the similarity cannot be compared in this test. Meanwhile, the strains within the genus *Rhizobium* possess a wide range of ANI values (approximately 72-98%), which indicated the diverse genetic distance among the strains within this genus (Fig. 4A). In contrast, the strains belonged to *Agrobacterium*, *Sinorhizobium* and *Bradyrhizobium* showed relatively narrow range of ANI value (86-100% for *Agrobacterium*; 78-98% for *Sinorhizobium*; and 89% for *Bradyrhizobium*) (Fig. 4A). The strains within *Bradyrhizobium* shared lowest ANI similarity with the strains belonging to the other genera (approximately 67%, Fig. 4, yellow); and most of the strains within *Rhizobium*, *Agrobacterium*, and *Sinorhizobium* groups shared 71-75% ANI similarities with each other (Fig. 4A), except that the ANI similarities between *R. lupini* HPC(L) and the *Agrobacterium* strains were higher (~ 85-88%, Fig. 4A) than those with the strains in other genera (~71-75%, Fig. 4A). Moreover, without *R. lupini* HPC(L), the *Rhizobium* strains showed relatively narrow range of ANI value (76-98%, Fig. 4B), which indicating that *R. lupini* HPC(L) may be more appropriate to be re-classified into *Agrobacterium*.

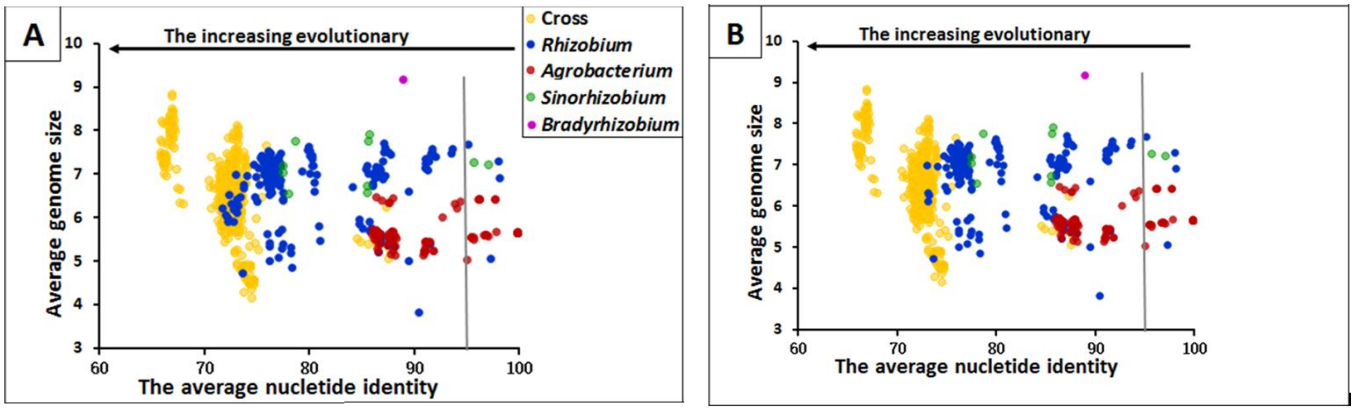

FIG. 4. PLOTTED RESULTS (990 POINTS) OF PAIRWISE AVERAGE GENOME SIZE VERSUS PAIRWISE ANI VALUES. EACH POINT REPRESENTS THE PAIRWISE ANI VALUES OF TWO STRAINS. (A) THE BLUE (171 POINTS), RED (153 POINTS), GREEN (10 POINTS), PINK (1 POINTS) AND YELLOW (636 POINTS) PLOTS INDICATE THE PAIRWISE AVERAGE GENOME SIZE VERSUS PAIRWISE ANI VALUES OF THE STRAINS WITHIN *RHIZOBIUM*, *AGROBACTERIUM*, *SINORHIZOBIUM* AND *BRADYRHIZOBIUM*, AND AMONG THE FOUR GROUPS, RESPECTIVELY. SINCE THERE IS ONLY ONE STRAIN BELONGED TO *ALLORHIZOBIUM* OR *PARARHIZOBIUM*, THE PAIRWISE AVERAGE GENOME SIZE VERSUS PAIRWISE ANI VALUES OF THESE TWO STRAINS CANNOT BE COMPARED. (B) WITHOUT *R. LUPINI* HPC(L), ONLY 153 BLUE POINTS REPRESENTED THE COMPARISON OF 18 *RHIZOBIUM* STRAINS WERE SHOWED.

3.6 The core genes among the studied strains in genera level

To further understand the similar and different genetic characteristics of the tested strains, the core genes in *Agrobacteirum*, *Rhizobium*, *Sinorhizobium* and *Bradyrhizobium* have been identified. *Allorhizobium* and *Pararhizobium* were excluded since only one strain was used in each of these two genera. The 18 strains previously classified to *Agrobacterium* shared 891 core genes, while the 20 strains within *Rhizobium* genus had 768 core genes. When the strain *R. lupini* HPC(L) was moved into *Agrobacterium* group, *Al. vitis* was separated from *Agrobacterium*, and *P. giardinii* H152^T was separated from *Rhizobium*, the core genes of *Agrobacterium* and *Rhizobium* groups would be 1,065 and 977, respectively, further proving the rationality of their reclassification.

IV. DISCUSSION

The taxonomy and nomenclature of genera in Rhizobiaceae have been changed dramatically in the last two decades associated with the development of taxonomic methods, especially the application of distinct molecular methods. Currently, *Agrobacterium*, *Allorhizobium*, *Ensifer* (*Sinorhizobium*), *Neorhizobium*, *Pararhizobium* and *Rhizobium* are described or emended based upon the phylogenetic relationships of 16S rRNA gene and multilocus sequencing analysis [16-17]. All these genera contained the symbiotic nitrogen-fixing and the tumor-inducing phytopathogenic bacteria [19-20], as well as saprophytic and endophytic bacteria [30]. Meanwhile, the genome sequencing data have been considered in description of novel genus and species in the family Rhizobiaceae, such as *Rhizobium lentis* and sister species [31] and *Pseudorhizobium pelagicum* [32]. These studies demonstrated that the genome analyses are valuable for the classification of *Rhizobium-Agrobacterium* related bacteria.

In the present study, the ten arsenite-oxidizing or antimonite tolerant strains were identified by comparing their genome sequences with other 35 related genome sequences available in the database. Our phylogenomic analyses of both the core-genome and the ANI supported the definition of *Agrobacterium*, *Allorhizobium*, *Sinorhizobium* (*Ensifer*), and *Rhizobium* (Figs. 1 and 2), and these groups were also supported by the MDS analysis and scatter plot based on pairwise ANI values (Figs. 3 and 4). These results demonstrated the analyses of ANI and core-genome are both convenient and confident taxonomy methods. From our data, the following threshold values could be drawn: 1) 70% for family (66.00-68.01 % between Bradyrhizobiaceae and Rhizobiaceae, >70.54 among the strains within Rhizobiaceae); 2) 75% for genus, which fits for definition of *Agrobacterium*, *Allorhizobium*, *Sinorhizobium* and *Rhizobium*; 3) 90% for species according to the differentiation of *R. etli*, *R. leguminosarum*, *R. rhizogenes*, *R. tropici*, *R. freirei* and the two species of *Bradyrhizobium*. Applying these threshold values, all the six arsenite-oxidizing *Agrobacterium* strains (C13, D14, JL28, LY4, TS43 and TS45) could be identified as *Ag. radiobacter* since they shared ANI >96.8 % with each other and >94.50 % with the type strain. As to the three antimonite-oxidizing and one antimonite tolerant *Sinorhizobium* strains, GL2 and GW3 could be identified as

Sinorhizobium sp. I, while GL28 and Sb3 as *Sinorhizobium* sp. II, both showed ANI values >78.21 % with *S. fredii* USDA 205^T. The exact taxonomic affiliation of the four *Sinorhizobium* strains can be further determined by comparing with other defined species in the genus.

In addition to the identification of our test strains, several taxonomic clues are worthy to discuss. 1) Except of the 11 strains of *Ag. radiobacter*, the sharing of ANI between 84.99% and 88.72% of the other five *Agrobacterium* strains and *R. lupini* HPC(L) with the *Ag. radiobacter* strains indicated that they might represent sister species of *Ag. radiobacter*, which were previously termed as *Agrobacterium* sensu stricto [33]. *Rhizobium lupini* HPC(L) is apparently a misnamed strain since it showed closer relationships with *R. etli* and *Rhizobium leguminosarum* in 16S rRNA analysis [34], and it should be reclassified as a member of *Ag. radiobacter* based on the analyses of ANI and core-genome. This change does not affect the nomenclature of the species, since the type strain of *R. lupini* USDA3051^T has been reclassified as *Bradyrhizobium lupini* based on the comparison of 16S rRNA, *recA* and *glnII* genes [35]. 2) The strain *Ag. albertimagni* AOL15, for whom the genus was reported as quite uncertain [36], seemed representing an independent genus based upon its ANI <74.29 % with the other strains involved in the study. 3) The strain *P. giardinii* $H152^T$ seemed belonging to the genus *Sinorhizobium* (ANI>75.16-76.22 %); therefore, the description of *Pararhizobium* based upon the MLSA results [17, 33] is questionable. 4) The classification of *R. phaseoli* Ch24-10, *R. etli* 8C-3 and *R. etli* Kim 5 should be re-examined since they formed a genospecies differed from the type strain of *R. etli*. 5) The species definition of the six *R. leguminosarum* strains should be revised since they presented ANI values greater than 90% with the type strain *R. gallicum* $R602^T$. 6) *Rhizobium* sp. AP16 could be identified as *R. rhizogenes*. All of these observations were supported by the core-genome analysis (Fig. 1), ANI tree (Fig. 2), ANI values (Table S2) and the MDS and scatter plot analyses (Figs. 3 and 4). In addition, the core genes number was increased when calculated without *R. lupini* HPC(L) and *Ag. albertimagni* AOL15, respectively (Fig. 5), which is also consistent with the analyses of ANI and core-genome.

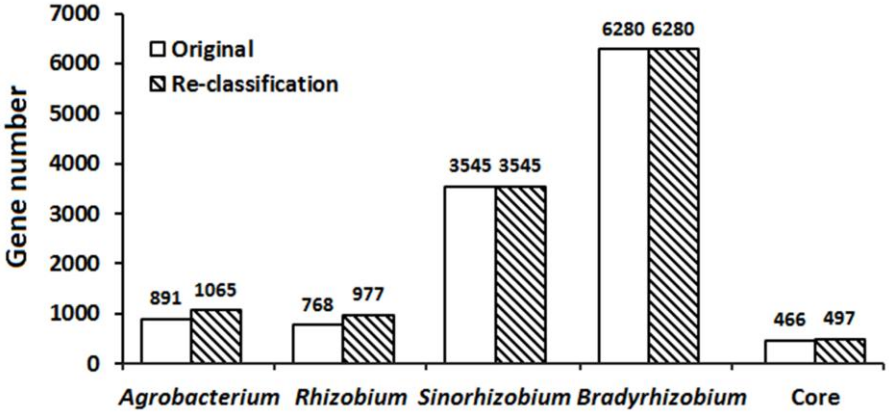

FIG. 5. THE COMPARISON OF CORE GENES AMONG *AGROBACTERIUM*, *RHIZOBIUM*, *SINORHIZOBIUM*, AND *BRADYRHIZOBIUM* GENERA. THE NUMBER OF THE CORE GENES IN *AGROBACTERIUM*, *RHIZOBIUM*, *SINORHIZOBIUM*, AND *BRADYRHIZOBIUM* WERE 891, 768, 3,545 AND 6,280, RESPECTIVELY (MARKED AS ORIGINAL). HOWEVER, IF THE *R. LUPINI* HPC(L) WAS CLUSTERED INTO *AGROBACTERIUM* GROUP, AND *ALLORHIZOBIUM VITIS* S4 AND *PARARHIZOBIUM GIARDINII* $H152^T$ WERE CLASSIFIED INTO THE NEW GENUS (MARKED AS RE-CLASSIFICATION), THE CORE GENES IN *AGROBACTERIUM* AND *RHIZOBIUM* GROUPS WOULD CHANGE TO 1,065 AND 977, RESPECTIVELY.

A considerable advantage of the ANI and core-genome over the MLSA or single gene analyses (16S rRNA or recA) for species identification is its stability and ease of access to information worldwide. In this study, we gathered genomic information for the 45 strains and constructed a mini-database of 990 pairwise ANI values (Table S2). This mini-database can provide a first-step ANI resource, which allows users to finish a genome-based ANI identification of the strains within the family Rhizobiaceae rapidly. In addition, the analysis of core-genome compared hundreds of common genes included housekeeping genes, such as16S rRNA gene and *recA*, which make the comparison more convincible. So far, sequencing bacterial genomes is cost-efficient, and good quality draft genomes are good enough for ANI or core-genome comparisons. Thus, the ANI and core-genome methodologies provide power tools for phylogenomic studies.

V. CONCLUSION

Conclusively, we propose the analyses of ANI and core-genome as convenient methods to estimate the phylogenetic relationship for the rhizobia-related strains, following the thresholds of 90%, 75% and 70% ANI values for species, genus and family, respectively. With these thresholds, we identified the ten arsenite-oxidizing and antimonite-tolerant strains as *Ag.*

radiobacter and two *Sinorhizobium* genomic species differing from *S. fredii*. In addition, the description of *Pararhizobium* is questioned because ANI values greater than 75% were detected between *P. giardinii* H152^T and *Sinorhizobium* strains. Also, reversion of the species definition for several strains in *R. etli* and *R. leguminosarum* was suggested. Our results demonstrate that analyses of ANI and core-genome are powerful supplemented methods to taxonomic identification of bacterial strains.

ACKNOWLEDGEMENTS

This work was supported by the National Natural Science Foundation, P. R. China (31670108 and J1103510).

REFERENCES

[1] B. Frank, B. Ueber die Pilzsymbiose der Leguminosen. Berlin: Parey, 1890.

[2] C. F. Tian, J. P. Young, E. T. Wang, S. M. Tamimi, and W. X. Chen, "Population mixing of *Rhizobium leguminosarum* bv. viciae nodulating *Vicia faba*: the role of recombination and lateral gene transfer," FEMS Microbiol. Ecol. vol. 73, pp. 563-576, 2010.

[3] H. J. Conn, "Validity of the genus *Alcaligenes*," J. Bacteriol. vol. 44, pp. 353, 1942.

[4] M. P. Starr, and J. E. Weiss, "Growth of phytopathogenic bacteria in a synthetic asparagin medium," Phytopathology. vol. 33, pp. 314-318, 1943.

[5] K. Ophel, and A. Kerr, "*Agrobacterium vitis* sp. nov. for strains of *Agrobacterium* biovar 3 from grapevines," Int. J. Syst. Bacteriol. vol. 40, pp. 236-241, 1990.

[6] H. Bouzar, and J. B. Jones, "*Agrobacterium larrymoorei* sp. nov., a pathogen isolated from aerial tumours of *Ficus benjamina*. Int. J. Syst. Evol. Microbiol. vol. 51, pp. 1023-1026, 2001.

[7] K. Kersters, and J. De Ley, Genus III. *Agrobacterium* Conn 1942, 359AL. In Bergey's Manual of Systematic Bacteriology, N. R. Krieg & J. G. Holt. Baltimore: Williams & Wilkins, pp. 244-254, 1984.

[8] P. de Lajudie, E. Laurent-Fulele, A. Willems, U. Torck, R. Coopman, M. D. Collins, K. Kersters, B. Dreyfus, and M. Gillis, "*Allorhizobium undicola* gen. nov., sp. nov., nitrogen-fixing bacteria that efficiently nodulate *Neptunia natans* in Senegal," Int. J. Syst. Bacteriol. vol. 48, pp. 1277-1290, 1998.

[9] J. Young, L. Kuykendall, E. Martinez-Romero, A. Kerr, H. Sawada, "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. vol. 51, pp. 89-103, 2001.

[10] S. K. Farrand, P. B. van Berkum, and P. Oger, "*Agrobacterium* is a definable genus of the family Rhizobiaceae," Int. J. Syst. Evol. Microbiol. vol. 53, pp. 1681-1687, 2003.

[11] K. Lindström, and E. Martínez-Romero, International Committee on Systematics of Prokaryotes. Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*: Minutes of the meeting, 23-24 July 2006, Århus, Denmark. Int. J. Syst. Evol. Microbiol. vol. 57, pp. 1365-1366, 2007.

[12] K. Lindström, and J. Young, International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of *Agrobacterium* and *Rhizobium*: Minutes of the meeting, 7 September 2010, Geneva, Switzerland. Int. J. Syst. Evol. Microbiol. vol. 61, pp. 3089-3093, 2011.

[13] J. Puławska, M. Maes, A. Willems, and P. Sobiczewski, "Phylogenetic analysis of 23S rRNA gene sequences of *Agrobacterium, Rhizobium* and *Sinorhizobium* strains," Syst. Appl. Microbiol. vol. 23, pp. 238-244, 2000.

[14] M. Shams, L. Vial, D. Chapulliot, X. Nesme, and C. Lavire, "Rapid and accurate species and genomic species identification and exhaustive population diversity assessment of *Agrobacterium* spp. using *recA*-based PCR. Syst. Appl. Microbiol. vol. 36, pp. 351-358, 2013.

[15] S. Tighe, P. De Lajudie, K. Dipietro, K. Lindström, G. Nick, and B. D. Jarvis, "Analysis of cellular fatty acids and phenotypic relationships of *Agrobacterium, Bradyrhizobium, Mesorhizobium, Rhizobium* and *Sinorhizobium* species using the Sherlock Microbial Identification System," Int. J. Syst. Evol. Microbiol. vol. 50, pp. 787-801, 2000.

[16] S. A. Mousavi, J. Osterman, N. Wahlberg, X. Nesme, C. Lavire, L. Vial, L. Paulin, P. de Lajudie, and K. Lindstrom, "Phylogeny of the *Rhizobium-Allorhizobium-Agrobacterium* clade supports the delineation of *Neorhizobium* gen. nov.," Syst. Appl. Microbiol. vol. 37, pp. 208-215, 2014.

[17] S. A. Mousavi, A. Willems, X. Nesme, P. de Lajudie, and K. Lindström, " Revised phylogeny of Rhizobiaceae: proposal of the delineation of *Pararhizobium* gen. nov., and 13 new species combinations," Syst. Appl. Microbiol. vol. 38(2), pp. 84-90, 2015.

[18] F. Aujoulat, H. Marchandin, I. Zorgniotti, A. Masnou, and E. Jumas-Bilak, "*Rhizobium pusense* is the main human pathogen in the genus *Agrobacterium*/*Rhizobium*. Clin. Microbiol. Infect. vol. 21(5), pp. 472.e1-5, 2015.

[19] W. Yang, Z. Kong, W. Chen, and G. Wei, "Genetic diversity and symbiotic evolution of rhizobia from root nodules of *Coronilla varia*," Syst. Appl. Microbiol. vol. 36, pp. 49-55, 2013.

[20] E. Velázquez, A. Peix, J. L. Zurdo-Piñiro, J. L. Palomo, P. F. Mateos, R. Rivas, E. Muñoz-Adelantado, N. Toro, P. García-Benavides, and E. Martínez-Molina, "The coexistence of symbiosis and pathogenicity-determining genes in *Rhizobium rhizogenes* strains enables them to induce nodules and tumors or hairy roots in plants," Mol. Plant. Microbe. Interact. vol. 18, pp. 1325-1332, 2005.

[21] L. Cai, G. Liu, C. Rensing, and G. Wang, "Genes involved in arsenic transformation and resistance associated with different levels of arsenic-contaminated soils," BMC. Microbiol. vol. 9, pp. 4, 2009.

[22] Q. Wang, D. Xiong, P. Zhao, X. Yu, B. Tu B, and G. Wang, "Effect of applying an arsenic-resistant and plant growth-promoting rhizobacterium to enhance soil arsenic phytoremediation by *Populus deltoides* LH05-17," J. Appl. Microbiol. 111, 1065-1074, 2011.

[23] Z. Shi, Z. Cao, D. Qin, W. Zhu, Q. Wang, M. Li, and G. Wang, "Correlation models between environmental factors and bacterial resistance to antimony and copper," PLoS. One. vol. 8, pp. e78533, 2013.

[24] Y. M. Zhang, C. F. Tian, X. H. Sui, W. F. Chen, and W. X. Chen, "Robust markers reflecting phylogeny and taxonomy of rhizobia," PloS One vol. 7, pp. e44936, 2012.

[25] K. Tamura, D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar, "MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods," Mol. Biol. Evol. vol. 28, pp. 2731-2739, 2011.

[26] JZ-M. Chan, M. Halachev, N. Loman, C., Constantinidou, and M. Pallen, "Defining bacterial species in the genomic era: insights from the genus *Acinetobacter*," BMC. Microbiol. vol. 12, pp. 302, 2012.

[27] J. Goris, K. T. Konstantinidis, J. A. Klappenbach, T. Coenye, P. Vandamme, and J. M. Tiedje, "DNA–DNA hybridization values and their relationship to whole-genome sequence similarities," Int. J. Syst. Evol. Microbiol. vol. 57, pp. 81-91, 2007.

[28] M. J. Norušis, and S. Inc, SPSS professional statistics 6.1. Prentice Hall, 1994.

[29] J. B. Kruskal, "Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis," Psychometrika. vol. 29, pp. 1-27, 1964.

[30] T. Gu, L.N. Sun, J. Zhang, X. H. Sui, and S. P. Li, "*Rhizobium flavum* sp. nov., a triazophos-degrading bacterium isolated from soil under the long-term application of triazophos," Int. J. Syst. Evol. Microbiol. vol. 64, pp. 2017-2022, 2014.

[31] M. H. Rashid, J. P. Young, I. Everall, P. Clercx, A. Willems, M. S. Braun, and M. Wink, "Average nucleotide identity of genome sequences supports the description of *Rhizobium lentis* sp. nov., *Rhizobium bangladeshense* sp. nov. and *Rhizobium binae* sp. nov. from lentil (Lens culinaris) nodules," Int. J. Syst. Evol. Microbiol vol. 65, pp. 3037-3045, 2015.

[32] N. E. Kimes, M. López-Pérez, J. D. Flores-Félix, M. H. Ramírez-Bahena, J. M. Igual, A. Peix, F. Rodriguez-Valera, and E. Velázquez, "*Pseudorhizobium pelagicum* gen. nov., sp. nov. isolated from a pelagic Mediterranean zone," Syst. Appl. Microbiol. vol. 38, pp. 293-299, 2015.

[33] E. Ormeño-Orrillo, L. E. Servín-Garcidueñas, M. A. Rogel, V. González, H. Peralta, J. Mora, J. Martínez-Romero, and E. Martínez-Romero, "Taxonomy of rhizobia and agrobacteria from the Rhizobiaceae family in light of genomics," Syst. Appl. Microbiol. vol. 38(4), pp. 287-291, 2015.

[34] L. Agarwal, and H. J. Purohit, "Genome sequence of *Rhizobium lupini* HPC(L) isolated from Saline Desert soil, Kutch (Gujarat)," Genome. Announc. vol. 1(1), e00071-12, 2013.

[35] A. Peix, M. H. Ramírez-Bahena, J. D. Flores-Félix, P. Alonso de la Vega, R. Rivas, P. F. Mateos, J. M. Igual, E. Martínez-Molina, M. E. Trujillo, and E. Velázquez, "Revision of the taxonomic status of the species *Rhizobium lupini* and reclassification as *Bradyrhizobium lupini* comb. nov," Int. J. Syst. Evol. Microbiol. vol. 65(4), pp. 1213-1219, 2015.

[36] W. L. Trimble, L. T. Phung, F. Meyer, J. A. Gilbert, and S. Silver, "Draft genome sequence of *Agrobacterium albertimagni* strain AOL15," J. Bacteriol. vol. 194, pp. 6986-6987, 2012.