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

Qian Wang¹, Wentao Zhu², Entao Wang³, Linhuang Zhang⁴, Xiangyang Li⁵, Gejiao Wang⁶*

¹²³⁴⁵⁶State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, P. R. China
*Email: gejiao@mail.hzau.edu.cn
²Departamento 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 polyphyllic 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. neopomum, Ag. pusense, and Ag. skirniewicinense were new combinations transferred from the former Rhizobium species. The emended Allorhizobium covered the phytopathogenic species Al. viitis (formerly Agrobacterium viitis), 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* DSM30147T 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**

<table>
<thead>
<tr>
<th>Species</th>
<th>Isolation source</th>
<th>Genome size</th>
<th>GC content</th>
<th>Predicted CDs</th>
<th>Accession No.</th>
<th>Level</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Agrobacterium</em> sp. C13*</td>
<td>Soil</td>
<td>5.64</td>
<td>59.8</td>
<td>5303</td>
<td>ASYD000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. D14*</td>
<td>Arsenic-enriched soil</td>
<td>5.54</td>
<td>59.8</td>
<td>5186</td>
<td>ASXX000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. JL28*</td>
<td>Antimony mine</td>
<td>5.65</td>
<td>59.8</td>
<td>5326</td>
<td>ASXZ000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. LY4*</td>
<td>Soil</td>
<td>5.64</td>
<td>59.8</td>
<td>5324</td>
<td>ASYA000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. TS43*</td>
<td>Arsenic-enriched soil</td>
<td>5.65</td>
<td>59.8</td>
<td>5368</td>
<td>ASYB000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. TS45*</td>
<td>Arsenic-enriched soil</td>
<td>5.64</td>
<td>59.8</td>
<td>5310</td>
<td>ASYC000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> 5A</td>
<td>Arsenic-enriched soil</td>
<td>5.74</td>
<td>58.6</td>
<td>5520</td>
<td>AGVZ000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> GW4</td>
<td>Arsenic polluted soil</td>
<td>5.64</td>
<td>59.8</td>
<td>5131</td>
<td>AWGV010000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Ag. radiobacter DSM 30147T</em></td>
<td>Soil</td>
<td>7.18</td>
<td>59.9</td>
<td>6834</td>
<td>ASXY000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> C58</td>
<td>Cherry tree tumor</td>
<td>5.67</td>
<td>59.1</td>
<td>5355</td>
<td>GCA_000092025</td>
<td>complete</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> Cherry 2E-2-2</td>
<td>Crown gall</td>
<td>5.43</td>
<td>59.9</td>
<td>5045</td>
<td>APCC000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> CCNWGS0286</td>
<td>Zinc-lead mine tailing</td>
<td>5.21</td>
<td>59.5</td>
<td>4985</td>
<td>AGSM000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>tumefaciens</em> F2</td>
<td>Soil</td>
<td>5.47</td>
<td>59.5</td>
<td>5321</td>
<td>AFSD000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. ATCC 31749</td>
<td>Rhizosphere</td>
<td>5.46</td>
<td>59.5</td>
<td>5536</td>
<td>AECL000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. H13-3</td>
<td>Soil</td>
<td>5.57</td>
<td>58.5</td>
<td>5345</td>
<td>GCA_000192635</td>
<td>complete</td>
</tr>
<tr>
<td><em>Agrobacterium</em> sp. 224MTsu3.1</td>
<td>Soil</td>
<td>4.8</td>
<td>59.7</td>
<td>4593</td>
<td>ARQL000000000</td>
<td>draft</td>
</tr>
<tr>
<td>Ag. <em>albertimagni</em> AOL15</td>
<td>Arsenite oxidizing biofilm</td>
<td>5.09</td>
<td>61.2</td>
<td>4811</td>
<td>AJF000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>Allorhizobium vitis</em> S4</td>
<td>Vitis vinifera nodule</td>
<td>6.32</td>
<td>57.5</td>
<td>5389</td>
<td>GCA_00016285</td>
<td>complete</td>
</tr>
<tr>
<td><em>R. etli</em> 8C-3</td>
<td>Root nodule</td>
<td>3.47</td>
<td>61.1</td>
<td>5076</td>
<td>ABRA000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. etli</em> CFN 42T</td>
<td>Phaseolus vulgaris nodule</td>
<td>6.53</td>
<td>61.1</td>
<td>5963</td>
<td>GCA_000092045</td>
<td>complete</td>
</tr>
<tr>
<td><em>R. etli</em> Kim 5</td>
<td>Root nodule</td>
<td>4.14</td>
<td>61.1</td>
<td>5963</td>
<td>ABQY000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. freirei</em> PRF 81</td>
<td>Bean nodule</td>
<td>7.08</td>
<td>59.9</td>
<td>6271</td>
<td>AQBH000000000</td>
<td>complete</td>
</tr>
<tr>
<td><em>R. grahamii</em> CCGE 502T</td>
<td>Root nodule</td>
<td>7.15</td>
<td>59.4</td>
<td>7368</td>
<td>AEYE000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. gallicum</em> bv. gallicum R602T</td>
<td>Phaseolus vulgaris nodule</td>
<td>7.22</td>
<td>60.8</td>
<td>7152</td>
<td>GCA_000816845</td>
<td>complete</td>
</tr>
<tr>
<td><em>R. lupini</em> HPC(L)</td>
<td>Saline desert soil</td>
<td>5.27</td>
<td>59.2</td>
<td>4615</td>
<td>AMQQ000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. leguminosarum</em> bv. trifolii WSM597</td>
<td>Trifolium polymorphum nodule</td>
<td>7.63</td>
<td>61</td>
<td>7159</td>
<td>AKHZ000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. leguminosarum</em> bv. trifolii SRD1565</td>
<td>Trifolium polymorphum nodule</td>
<td>6.91</td>
<td>60.7</td>
<td>6870</td>
<td>AQUSD000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. leguminosarum</em> bv. phaseoli 4292</td>
<td>Bean nodule</td>
<td>7.35</td>
<td>60.7</td>
<td>7255</td>
<td>AQZR000000000</td>
<td>draft</td>
</tr>
<tr>
<td><em>R. leguminosarum</em> bv. viciae TOM</td>
<td>Legume root nodule</td>
<td>7.36</td>
<td>60.8</td>
<td>7298</td>
<td>AQUC000000000</td>
<td>draft</td>
</tr>
</tbody>
</table>
### 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\textsuperscript{T}, *R. leguminosarum* bv. *viciae* 3841, *R. tropici* CIAT 899\textsuperscript{T}, *R. rhizogenes* K84, and *R. gallicum* R602sp\textsuperscript{T}), and 14 draft genomes (including the type strain *R. grahamii* CCGE 502\textsuperscript{T}), were found. In addition, draft genomes were also obtained for *P. giardinii* H152\textsuperscript{T}, five *Sinorhizobium* strains (including the type strain *S. fredii* USDA205\textsuperscript{T}) and two *Bradyrhizobium* strains. *B. diazoefficiens* USDA110\textsuperscript{T} and *B. japonicum* USDA6\textsuperscript{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\textsuperscript{T}), whereas the number of predicted CDSs vary from 4593 (*Agrobacterium* sp. 224MTsu3.1) to 8826 (*Bradyrhizobium* *japonicum* USDA6\textsuperscript{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\textsuperscript{T} was clustered with *Rhizobium* sp. PRF 81, *R. tropici* CIAT 899\textsuperscript{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).

---

For Table S2, You need to contact to our chief editor or Corresponding author: gejiao@mail.hzau.edu.cn  Page | 79
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 RESAMPLINGS. 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.
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 Bradyrhizobiaceae, 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 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 Sh3 form the sp. I; while GW3 and GL2 formed sp. II; both were different from the type strain S. fredii USDA 205. 4) Pararhizobium giardinii H152 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).


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*. 
The strain, 70.54 among the strains within Rhizobiaceae); 2) 75% for genus, which fits 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 puckei [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.

For Table S2, You need to contact to our chief editor or Corresponding author: gejiao@mail.hzau.edu.cn
Sinorhizobium sp. I, while GL28 and Sh3 as Sinorhizobium sp. II, both showed ANI values >78.21 % with S. fredii USDA 205T. 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 USDA3051T 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. giardini H152T 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 R602T. 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 GIARDINI H152T 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 convincing. 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* 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 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**


[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.
