Systematic and Applied Microbiology 38 (2015) 287–291

Contents lists available at ScienceDirect

Systematic and Applied Microbiology

journal homepage: www.elsevier.de/syapm

Taxonomy of rhizobia and agrobacteria from the *Rhizobiaceae* family in light of genomics

Ernesto Ormeño-Orrillo, Luis E. Servín-Garcidueñas, Marco A. Rogel, Víctor González, Humberto Peralta, Jaime Mora, Julio Martínez-Romero, Esperanza Martínez-Romero*

Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca, Morelos, Mexico

ARTICLE INFO

Keywords:
Rhizobia
Root nodule bacteria
Systematics
Phylogeny
Genome
Next generation sequencing

ABSTRACT

Phylogenomic analyses showed two major superclades within the family *Rhizobiaceae* that corresponded to the *Rhizobium*/*Agrobacterium* and *Shinella*/*Ensifer* groups. Within the *Rhizobium*/*Agrobacterium* group, four highly supported clades were evident that could correspond to distinct genera. The *Shinella*/*Ensifer* group encompassed not only the genera *Shinella* and *Ensifer* but also a separate clade containing the type strain of *Rhizobium giardinii*. *Ensifer adhaerens* (Casida A^T) was an outlier within its group, separated from the rest of the *Ensifer* strains. The phylogenomic analysis presented provided support for the revival of *Allorhizobium* as a *bona fide* genus within the *Rhizobiaceae*, the distinctiveness of *Agrobacterium* and the recently proposed *Neorhizobium* genus, and suggested that *R. giardinii* may be transferred to a novel genus. Genomics has provided data for defining bacterial-species limits from estimates of average nucleotide identity (ANI) and *in silico* DNA–DNA hybridization (DDH). ANI reference values are becoming the gold standard in rhizobial taxonomy and are being used to recognize novel rhizobial lineages and species that seem to be biologically coherent, as shown in this study.

Introduction

Rhizobia are soil and rhizospheric bacteria that may form nitrogen fixing symbioses in leguminous plants allowing their growth in poor nitrogen soils. Thus, rhizobia have been considered as bio-fertilizers and have been used as inoculants in agriculture for over 120 years. Rhizobial genetic diversity, as well as their plant–bacteria molecular interactions, has been well studied. In 1991, Graham et al. [18] published a set of recommendations for the description of novel rhizobial species on the "basis of both phylogenetic and phenotypic traits" using "genomic relationships to the greatest degree possible" and as "the culmination of considerable research". A large number of species have been reported since, based mainly on polyphasic analysis using a number of molecular-marker phylogenies, DNA–DNA hybridization (DDH) results and the description of different distinctive phenotypic features. Studies based on molecular marker sequences represented a significant advance in rhizobial taxonomy and have been included in most studies. Newly described species are periodically revised by the International Taxonomy Subcommittee on *Agrobacterium* and *Rhizobium* [24,25]. Several reviews on rhizobial taxonomy have been published [41,52,55] but none have been specifically oriented toward genomics.

The following genera within the family *Rhizobiaceae* include rhizobial members: *Rhizobium*, *Ensifer* (former *Sinorhizobium*), *Agrobacterium* and *Shinella* [5]. *Agrobacterium* includes tumor-forming bacteria as well as nitrogen-fixing nodule bacteria. An additional genus of the family *Rhizobiaceae*, *Carbophilus* [5], has not been described as containing nodule bacteria. A characteristic of rhizobia belonging to the family *Rhizobiaceae* and agrobacteria is their genome organization in multireplicons [17,19,21,26]. Furthermore, phenotypic distinctive characteristics in rhizobia may be encoded in extrachromosomal replicons (ERs) [33], a feature not normally recognized in novel species descriptions. In *Rhizobium*, *Ensifer* and *Agrobacterium*, almost half of the genome may be contained in ERs (reviewed in [26]), and some ERs even have roles in rhizobial growth rate and survival [4,16,19,20]. Two types of ERs have been recognized: plasmids and chromids [19]. ERs that carry "essential" genes with conserved gene sequences and sharing similar GC content and codon usage with the chromosome have been named chromids [19]. Chromids have been proposed as characteristic of a genus and contain many genus-specific genes [26]. Taxonomic phenotypic characteristics are encoded in chromids in *Rhizobium* [33]. Interestingly, chromids carry many genes

* Corresponding author. Tel.: +52 777 3131697.
E-mail address: emartine@ccg.unam.mx (E. Martínez-Romero).

http://dx.doi.org/10.1016/j.syapm.2014.12.002
0723-2020/

that are highly expressed by rhizobia on plant roots [33]. On the other hand, plasmids are highly variable and confer adaptive traits, such as nodulation and nitrogen fixation in legumes [6,8,19,26,29,38,43,49,51], or they may be transferred between bacteria [30,31,40].

Genomic impact on rhizobial taxonomy

Genomics has revolutionized microbiology and is having a significant impact on taxonomy. For many years, results from DDH experiments were the basis for circumscribing prokaryotic species [46]. However, alternatives for estimating DNA relatedness, such as whole-genome average nucleotide identity (ANI) [23] and *in silico* DDH [2], are currently much better than wet lab DDH which has been shown to produce highly variable results from lab to lab and from different DNA samples [27]. Additionally, the G+C content normally reported in novel bacteria descriptions may be calculated accurately from genomic data.

The novel quantitative genomic analyses are beginning to be used in rhizobial taxonomy. Species descriptions where ANI and/or *in silico* DDH were used to support or complement wet lab DDH values have been published [9,10,13,28]. Likewise, limits obtained from ANI and *in silico* DDH estimates have led to the discovery of novel rhizobial lineages [27,37]. Lack of genome sequences for most type strains of rhizobia is a limiting factor for the use of these novel approaches, although the reducing cost of whole genome sequencing will ease this restriction. As an example, two recent studies coupled genome sequencing of the rhizobia being characterized with that of all related type strains, which allowed complete replacement of wet lab DDH with ANI [14,15].

ANI values derived from only the conserved core genes of a group (referred to as ANIo) have also been proposed as a replacement for wet lab DDH, with approximately 96% ANIo corresponding to 70% DDH [22]. A minimum of three but a recommended number of six to eight genes can give a good estimate of ANIo [22]. Consequently, ANI values based on concatenated sequences of a few partial sequences of conserved core genes are being used to delineate putative rhizobial species [1,11,39]. Nevertheless, care must be exercised in not assuming that values obtained with partial sequences will correspond exactly to ANIo, instead, intra- and inter-species identity values must be evaluated in order to find a suitable cut-off value for species delineation that is appropriate for the set of genes being used. Recently, a set of three novel conserved genes has been proposed as a suitable tool for rhizobial taxonomy because the concatenated partial sequences produced ANI that were closely correlated with whole genome ANI [56].

A phylogenomic view of rhizobia and agrobacteria within the *Rhizobiaceae*

Besides providing quantitative values for species delineation, whole genomes allow the reconstruction of phylogenetic trees based on hundreds or thousands of genes that depict evolutionary relationships better than phylogenies based on a few markers including 16S rRNA genes. To date, 29 complete and 141 draft whole genome sequences (WGS) from members of the family *Rhizobiaceae* are available from the GenBank database (Supplementary Table S1). These genomes include 23 type strains, three of which are completely sequenced. Additionally, one complete and three draft WGS genomes sequenced at CCG-UNAM were included in the analysis (Supplementary Table S1). A total of 166 (66%) of the strains had genomes encoding *nodC*. Most strains lacking this gene are labeled as agrobacteria.

We checked the identity of all sequenced type strains by comparing their genomes against partial sequences of genes previously obtained for the same strains available from GenBank, and two anomalies were found. The *A. radiobacter* DSM30147^T genome (accession number ASXY01, Bioproject PRJNA212112) had identical sequences to several previously reported *A. radiobacter* DSM30147^T genes (*aptD*, *rpoB*, *mutS*, *gyrB*, *gltD*, *glnII*) but showed only 90–97% identity with others (*rpoD*, *chvA*, *hrcA*). The *R. gallicum* R602sp^T genome (accession number ARDC01, BioProject PRJNA169700) had divergent sequences in all the genes compared, which clustered within the *R. leguminosarum* clade (data not shown). Both genomes were excluded from further analyses, as they did not correspond to the designated type strains.

Except for one comparison, type strain genomes shared a maximum ANI value of 92%, thus supporting the proposed cut-off level of 95% as a species delineation threshold [23]. *R. gallicum* R602^T (newly sequenced at CCG) and *R. mongolense* USDA 1844^T shared an ANI value of 95.1% that validates their previously proposed synonymy [44]. Based on a 95% ANI threshold, the 172 sequenced strains would represent 77 genospecies (Supplementary Table S1). Given the scarcity of sequenced type strains, most of these genospecies could not be ascribed to described taxa solely by ANI and so were assigned arbitrary labels (GS1-G48) in Supplementary Table S1. To date, there are 27 genomes available from different *Ensifer meliloti* strains and eight genomes from distinct *E. fredii* strains, while the remaining geno(species) have from 1 to 6 sequenced strains. We used a sample of 113 genome sequences representing all possible (geno)species in order to construct a genome-based phylogeny with the aim of shedding light on uncertainties or controversies in the taxonomy of several clades within the family *Rhizobiaceae*. Due to the unreliable classification or naming of many sequenced strains (Supplementary Table S1) we chose to include a species designation only for type strains or strains that had been assigned to a known species on the basis of DNA-DNA hybridization or ANI analyses. As shown in Fig. 1, two major superclades were observed within the family *Rhizobiaceae*, which corresponded to the *Rhizobium*/*Agrobacterium* and *Shinella*/*Ensifer* groups.

Within the *Rhizobium*/*Agrobacterium* group, several highly supported clades were evident. One clade included *Agrobacterium* biovar 1 strains, as well as the type strains of *Agrobacterium rubi* and *Agrobacterium larrymoorei*. This clade, referred to here as *Agrobacterium sensu stricto*, had been previously revealed by *recA* sequence analysis [7] and includes strains whose genomes encode a protelomerase, which are characterized by possessing a linear replicon [36].

A second clade included strains of the recently proposed genus *Neorhizobium* [32]. This clade was previously known as the "*Rhizobium galegae* complex". The genome phylogeny supported the proposal of Mousavi et al. [32] for including *Rhizobium vignae* in *Neorhizobium*. However, ANI values between *R. vignae* and *N. galegae* strains were lower than 91%, indicating that *R. vignae* should not be included in the *N. galegae* species as suggested by Mousavi et al. [32] and must therefore be referred to as *Neorhizobium vignae*.

A third clade included the type strains of *Rhizobium undicola* (former *Allorhizobium undicola* [12,54]), as well as strain S4 of *Agrobacterium vitis*. The isolated position of this clade in relation to *Agrobacterium sensu stricto* and *Neorhizobium* could support the revival of *Allorhizobium* as a genus within the *Rhizobiaceae*, as has been recently suggested [36], and which includes the species *Allorhizobium vitis* (formerly *Agrobacterium vitis*) and *Allorhizobium taibaishanense* (former *Rhizobium taibaishanense*) [32].

A fourth clade included species closely related to *Rhizobium leguminosarum*, the type species of the genus *Rhizobium*, hence, we referred to this clade as *Rhizobium sensu stricto*. Within this clade, distinct groups of closely related species were observed. Hairy-root forming bacteria, originally described as *Agrobacterium rhizogenes* (biovar 2 agrobacteria) were found within the