University of Dundee

Rhizobium brockwellii sp. nov., Rhizobium johnstonii sp. nov. and Rhizobium beringeri sp. nov., three genospecies within the Rhizobium leguminosarum species complex

Young, J. Peter W.; Jorrin, Beatriz; Moeskjær, Sara; James, Euan K.

[Link to publication in Discovery Research Portal](#)

Rhizobium brockwellii sp. nov., *Rhizobium johnstonii* sp. nov. and *Rhizobium beringeri* sp. nov., three genospecies within the *Rhizobium leguminosarum* species complex

J. Peter W. Young[1,*], Beatriz Jorrin[2], Sara Moeskjær[3] and Euan K. James[4]

Abstract

Genomic evidence indicates that the *Rhizobium leguminosarum* species complex comprises multiple distinct species, perhaps 18 or more. Of the five earliest genospecies (gs) to be described, only two have formal names: *R. leguminosarum sensu stricto* (gsE) and *Rhizobium ruizarguesonis* (gsC). Here, we provide formal descriptions and names for the other three genospecies, based on the publicly available genome sequences for multiple strains of each species: *Rhizobium brockwellii* sp. nov. (gsA, 37 strains, type strain CC275e^T=LMG 6122^T = ICMP 2163^T=NZP 561^T = PDDCC 2163^T=HAMBI 13^T), *Rhizobium johnstonii* sp. nov. (gsB, 54 strains, type strain 3841^T = LMG 32736^T=DSM 114642^T) and *Rhizobium beringeri* sp. nov. (gsD, 8 strains, type strain SM51^T = LMG 32895^T = DSM 115206^T). Each species forms a well-supported clade in a phylogeny based on 120 concatenated core genes. All strains have average nucleotide identity (ANI) above 96% with the relevant type strain and below 96% with all other type strains. Each species is characterised by a number of genes that are absent or rare in other species.

DATA SUMMARY

Supplementary files can be found in Figshare https://doi.org/10.6084/m9.figshare.22734917.v1[1]

INTRODUCTION

Rhizobium leguminosarum (Frank 1879) Frank 1889^{AL} is the nomenclatural type species of the genus *Rhizobium* (Family *Rhizobiaceae*, Order *Hyphomicrobiales*, Class *Alphaproteobacteria*) [2]. Strains of this species form nitrogen-fixing nodules on the roots of certain legume plants. Three symbiovars (sv.) are known, defined by their distinct host specificity: sv. *viciae* nodulates plants in the tribe Fabeae (Vicieae), sv. *trifolii* nodulates members of the genus *Trifolium*, and sv. *phaseoli* nodulates members of the genus *Phaseolus*. The different host ranges are specified by homologous but distinct sets of nodulation genes that are plasmid-borne, frequently transferred horizontally, and hence of no taxonomic significance. It has been evident for some years that the genomic diversity within *R. leguminosarum*, as traditionally defined, is too great to be encompassed within a single species that meets currently accepted norms such as an average nucleotide identity (ANI) greater than 95%. It comprises a cluster of related species and can best be described as the *R. leguminosarum* species complex (Rlc). Initially, five genospecies (A to E) were defined based on low-coverage genome sequencing of 72 strains from *Trifolium repens* and *Vicia sativa* nodules at a single location in the UK [3], then the same five genospecies were confirmed in a set of 196 isolates from *T. repens* nodules from several localities in northern Europe [4], and a third study reported three of these genospecies plus two novel ones among isolates from five

European countries [5]. Considering all 429 publicly available genome sequences of Rlc strains, from a wide geographic range and various host plants, 18 clades were identified within the Rlc that could potentially be considered distinct species, as well as seven individual strains that fell outside these clades [6].

Some of these clades within the Rlc have already been assigned formal taxonomic names. The type strain of *R. leguminosarum*, USDA2370[T], is in genospecies gsE, so this genospecies remains *R. leguminosarum sensu stricto* [6]. The first new name proposed within the Rlc was *Rhizobium laguerreae* [7]. The type strain of this species is in clade R, but this is part of a group of closely related clades (N, O, P, Q, R) and further work is needed to decide how widely the name *R. laguerreae* should be applied within this group [6]. The name *Rhizobium ruizarguesonis* has been published for genospecies C [8]. Genospecies G is now *Rhizobium sophorae* [9], genospecies I is now '*Rhizobium indicum*' [10] and genospecies S is *Rhizobium changzhiense* [11]. There are several other clades within the Rlc that are equally deserving of formal species names. Our purpose here is to provide names for three of the five original genospecies that are still anonymous: genospecies A is named *Rhizobium brockwellii*, genospecies B becomes *Rhizobium johnstonii* and genospecies D becomes *Rhizobium beringeri*.

EVIDENCE FOR THE NOVEL SPECIES

The International Code of Nomenclature of Prokaryotes [12] specifies that "Descriptions of taxa should include the following information: (a) those characteristics which are essential for membership in the taxon, i.e., those characteristics which constitute the basic concept of the taxon; (b) those characteristics which qualify the taxon for membership in the next higher taxon; (c) the diagnostic characteristics, i.e., those characteristics which distinguish the taxon from closely related taxa; and (d) in the case of species, the total number of strains studied, and the strain designations should be given.". To fulfil these requirements, we provide (a) definitions of the species in terms of genomic similarity and phylogenetic coherence, (b) genome-based relatedness to the type strain of the type species of the genus, (c) diagnostic characteristics based on multiple strains of each species, (d) a table listing all strains with their genome accession details and relevant properties.

Descriptions of novel species (protologues) tend to focus on a description of the type strain, but it is actually much more important to consider the range of variation across all available examples of the proposed species. Indeed, the above quotation from the ICNP does not mention the type strain. We have genome sequences for numerous strains of each of the three novel species proposed here, so our descriptions will focus on the species-wide characteristics that can be deduced from these genomes. We base the analyses on 37 genomes of *R. brockwellii* (gsA), 54 genomes of *R. johnstonii* (gsB), and eight genomes of *R. beringeri* (gsD), listed in Table S1, available in the online version of this article. The key results, namely that strains of each species form a robust clade in a phylogeny based on 120 core genes and share appropriately high values of ANI, are already published [6].

While microbiologists have, in the past, used numerous phenotypic tests in attempting to describe bacterial species, these have seldom distinguished reliably among closely related species [13]. For instance, there are published data for utilisation of 95 carbon substrates by the proposed type strain of *R. johnstonii*, strain 3841, in Table S6 of ref [3]. However, the same table demonstrates that the utilisation profiles of other members of this species are all different and are no more similar to that of this type strain than are those of other species. The problem is that phenotypic differences between the type strains of closely related species are frequently encoded by accessory genes and will vary within as well as between species. Fortunately, genome sequences, and especially the availability of multiple genome sequences within each species, now provide much more robust characters to fulfil the ICNP requirements.

In principle, phenotypic properties can be predicted from genomic sequence data. For example, the authors of a tool called Protologger used it to predict metabolic pathways as well as to provide other genome-based information for descriptions of novel genera and species that met the ICNP criteria without any recourse to laboratory-based phenotypic characterisation [14]. We ran Protologger on each of the type strains of our three proposed novel species. The phenotypic predictions generated by the tool were that starch can be utilised as a carbon source; sulphide and L-serine are predicted to be utilised to produce L-cysteine and acetate; L-glutamate production from ammonia was predicted via L-glutamine; no antibiotic resistance genes were identified. These predictions were identical for all three type strains, so they did not provide diagnostic characters for describing the species. Protologger did, however, provide other information that was valuable, such as percentage of conserved proteins (POCP) compared with type strains of related species.

The sequence of the ribosomal small subunit RNA (16S rRNA) gene is normally included in species descriptions. However, we note that, within the genus *Rhizobium*, multiple distinct species share identical 16S rRNA sequences, while there is some variation in the sequence of this gene among strains that indisputably represent the same species [3, 6]. Hence, the 16S rRNA sequence is useful only in confirming that strains represent members of the genus *Rhizobium*. Any phylogeny based on a single gene is prone to artefacts due to a limited number of informative sites and the possibility of horizontal gene transfer, but the combined evidence of a large number of genes can yield a robust phylogeny. Fig. 1 shows a phylogeny, based on 120 core genes, that provides strong support for the monophyly of each of the three proposed species. It is modified from Figure 3 of reference [6], which should be consulted for full details.
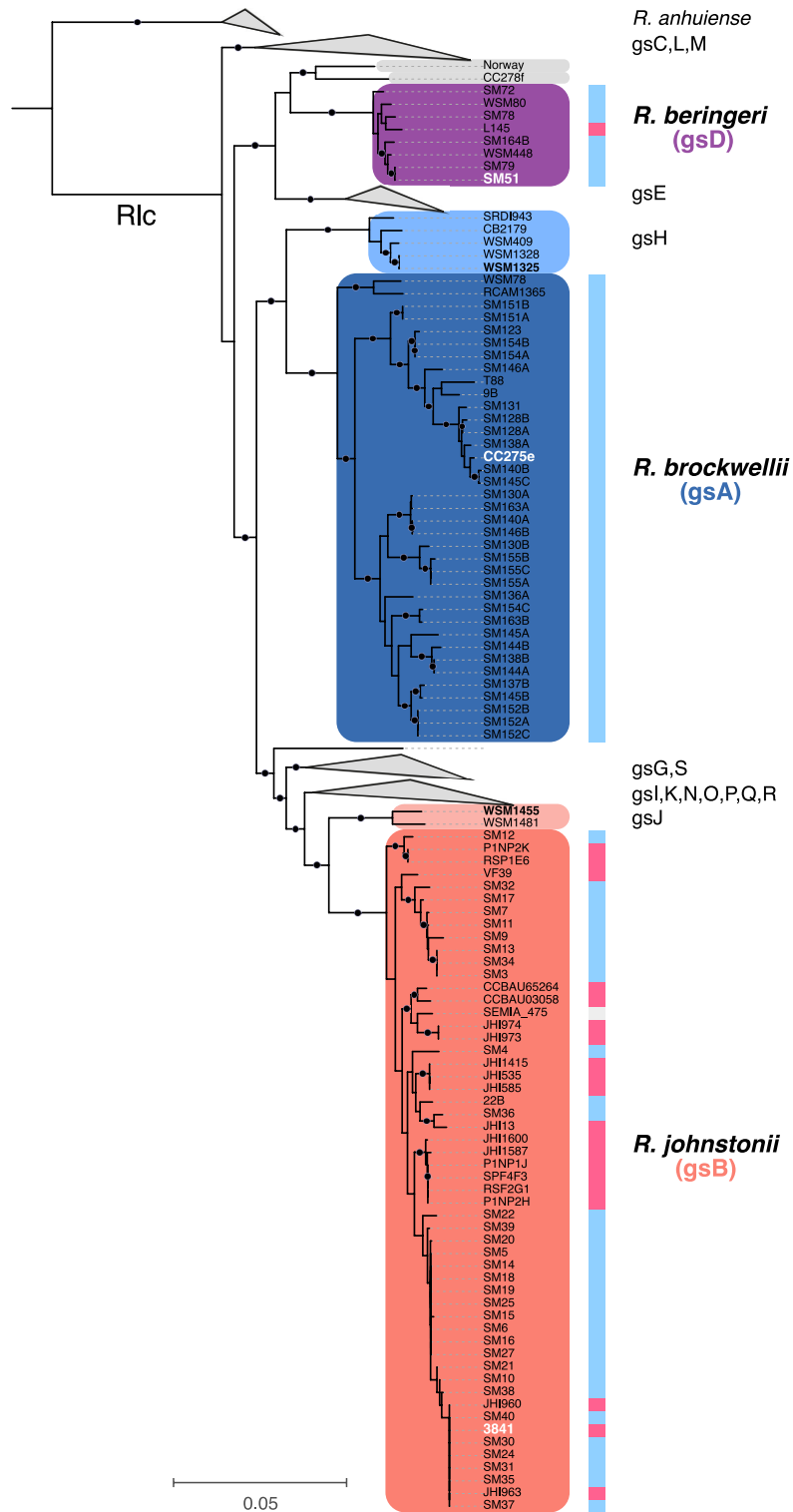
Fig. 1. Phylogeny of the three novel species and their immediate neighbours within the Rlc. Other genospecies are collapsed. Maximum likelihood tree based on concatenated sequences of 120 core genes; modified from Figure 3 of ref [6]. The tree was constructed using RAxML-NG [27] and displayed using iTOL [28]. Black dots indicate branches with 100% bootstrap support. Type strains are highlighted in white. The symbiovar of each strain is indicated by the vertical bar (*trifolii* blue, *viciae* red). Scale indicates substitutions per site.
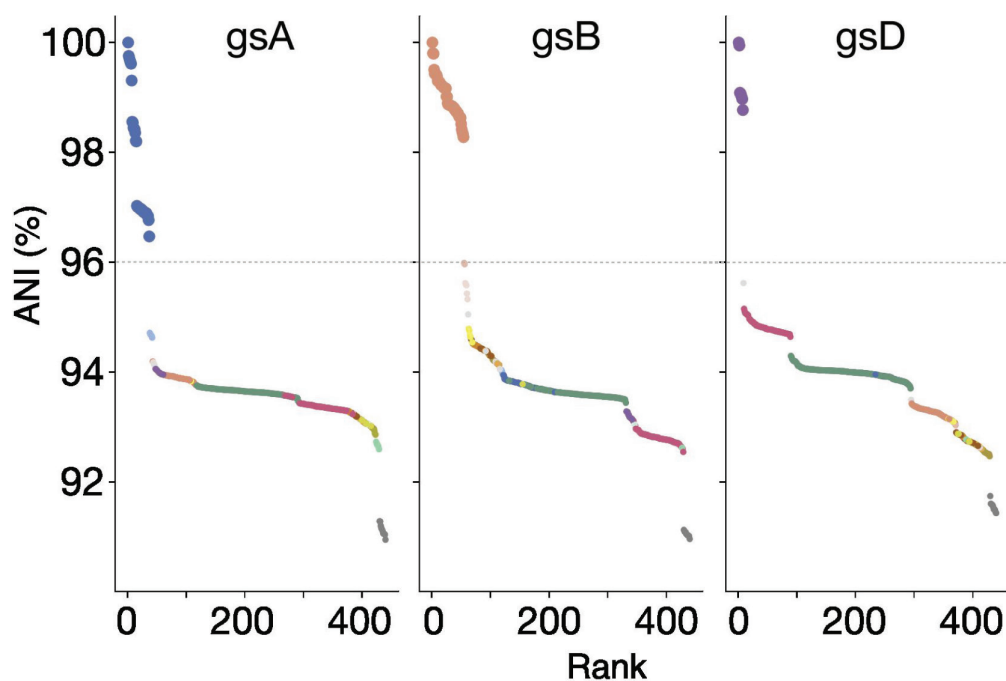
Fig. 2. Average nucleotide identity between the three type strains of the proposed species *R. brockwellii* (gsA), *R. johnstonii* (gsB) and *R. beringeri* (gsD) and 440 genomes of strains of gsA (blue), gsB (orange), gsD (purple), other genospecies within the Rlc (other colours) and the sister group R. anhuiense (grey). Based on published analysis using fastANI [18]; modified from Figure 5 of ref [6]. Full data in Table S1.

For genus delimitation, POCP is a more reliable indicator than 16S rRNA. It has been suggested that a 'prokaryotic genus can be defined as a group of species with all pairwise POCP values higher than 50%' [15]. However, given the primacy of the type in prokaryote taxonomy, a more consistent and stable criterion would be to require only that all type strains within the genus have POCP values above 50% with the type strain of the type species of the genus. The three novel species proposed here are so closely related to the type species that their POCP values are above 84% and their assignment to the genus *Rhizobium* is indisputable.

To delimit species within a genus, ANI based on whole genomes is the most widely accepted metric, with members of a species sharing a minimum of 95–96% ANI with the type strain [16]. While this threshold is a good guide across the whole of the prokaryotes, it is important to consider the distribution of the metric within the organisms of interest and, if possible, to choose a boundary that corresponds to a gap in the observed values. There are two justifications for this: (i) our taxonomic units should, as far as possible, correspond to natural units of biological diversity; (ii) ambiguity in taxonomic assignment is reduced if observations rarely fall close to the threshold. It should also be noted that there are many different methods to estimate ANI, and they yield slightly different values, so thresholds are specific to the chosen method [17]. We have used fastANI [18]. All the ANI values needed for our current purpose are already published [6], and the relevant plots are shown in Fig. 2. The minimum ANI of any gsA strain with the type strain is 96.49%, while the most closely related strain outside the species (in the sister group gsH) has an ANI of 94.77%. For gsB, the corresponding values are 98.28 and 95.99% (sister group gsJ), and for gsD they are 98.77 and 95.62% (sister strain Norway, not assigned to genospecies). In each case, therefore, there is a very clear gap around or above 96% ANI that provides a natural demarcation of the species.

It is worth noting that each of the proposed novel species is represented by strains isolated from a wide geographic range, so their narrow genomic diversity is not a result of limited sampling. The countries of origin of the sequenced strains are listed in the species descriptions below, and in Table S1.

It is usual to report the ranges of genome sizes and of DNA G+C nucleotide percentages as part of the description of novel species. These two metrics are plotted in Fig. 3, and it is striking that there is a strong negative correlation between them. This arises because the average DNA G+C content of accessory genes is lower than that of core genes [19]. It is also interesting to note that, despite their relatively close relationship as members of the Rlc, the three novel species are well separated when genome size and DNA G+C content are considered jointly.

All genomes of members of the genus *Rhizobium* include multiple large plasmids that can be classified by the *repABC* operons that maintain them [4]. Membership of a plasmid type is defined by an amino acid identity of the RepA protein above 90% to that of the type representative. We used a set of RepA, RepB and RepC protein sequences representing the
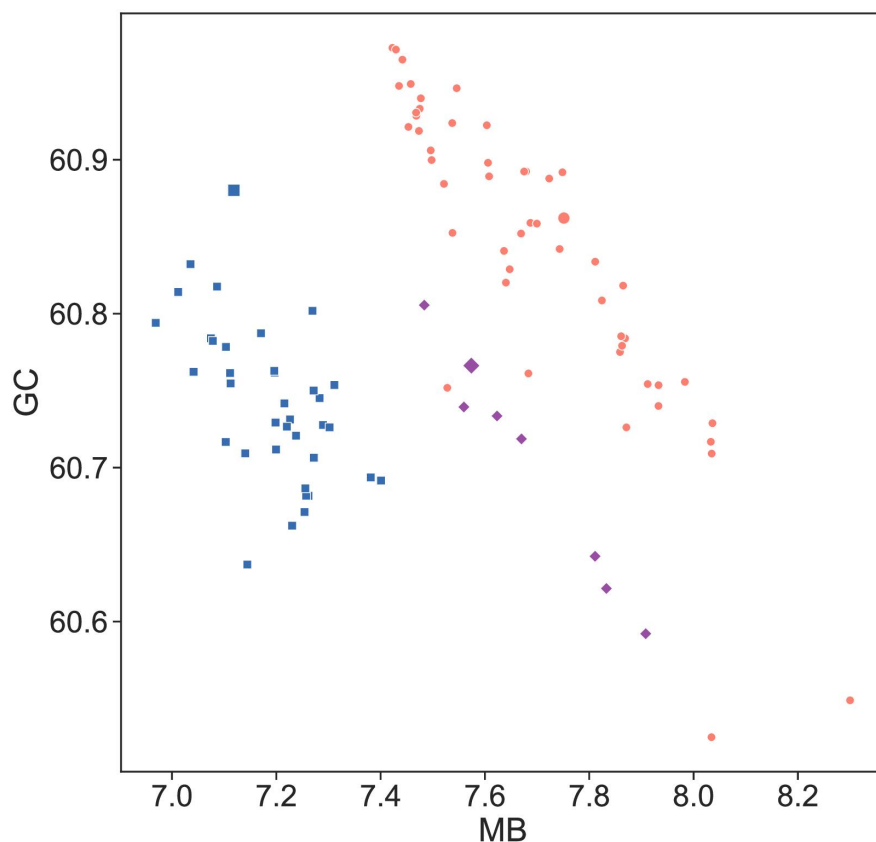
Fig. 3. Percentage of G+C nucleotides plotted against genome size for the available genomes of the three species *R. brockwellii* (gsA, blue squares), *R. johnstonii* (gsB, orange circles) and *R. beringeri* (gsD, purple diamonds). The type strains are indicated by larger symbols. Full data are available in Table S1.

known plasmid classes as queries in a tblastn search of the 440 genomes to identify the types present in each genome (Table S1). While most plasmid types are widely distributed throughout the Rlc, there are some genospecies-specific patterns. The Rh03 type is found in nearly all Rlc genomes, except in the clade that includes gsD and gsE. The only strains in this clade that have Rh03 are SM72, in gsD, and the two diverged strains Norway and CC278f that are the sister clade of gsD. The Rh05 plasmid type is almost entirely confined to two of the novel species defined here. It is in 100% of gsA and gsB genomes, and in their respective sister species gsH (4 out of 5 strains) and gsJ (2 out of 2 strains). Elsewhere in the Rlc, it is only found in five genomes out of 333.

To identify unique characteristics of each species, we sought genes that are present in all or nearly all strains of the species but absent from all or nearly all strains of other species in the Rlc, using the 440 Rlc and *Rhizobium anhuiense* genomes [6]. The set of proteins encoded by each genome was predicted using Prodigal 2.6.3 [20], then Diamond 2.0.15.153 [21] was used to search for homologs (minimum 70% amino acid identity) of all the proteins of each of the three novel type strains. The presence/absence matrices for these genes across all strains are given in Tables S2, S3 and S4, available from Figshare (https://doi.org/10.6084/m9.figshare.22734917.v1). The annotation for each gene is taken from the NCBI RefSeq version of the type strain genome sequence.

Table 1 lists 22 proteins of CC275e[T], the proposed type strain of *R. brockwellii* (gsA) that are diagnostic characteristics of the species. They were present in at least 95% of gsA genomes but no more than 5% of genomes in other genospecies of the Rlc, and were absent from gsH (the sister group of gsA). These proteins were predominantly encoded on the chromosome or the Rh01 chromid (pRltCC275eF), with one on the Rh05 plasmid (pRltCC275eD). A further 21 proteins (not shown) met these abundance thresholds but were shared by gsA and gsH, so represent synapomorphies of the gsA–gsH clade.

Similarly, Table 2 lists 21 proteins of strain 3841[T], the proposed type strain of *R. johnstonii* (gsB), that are diagnostic characteristics of the species. They were present in at least 95% of gsB genomes but no more than 5% of genomes in other genospecies of the Rlc, and were absent from the two genomes of gsJ, the sister group of gsB (Table 1). In the type strain, the corresponding genes were distributed in ten clusters across the chromosome, the Rh01 and Rh02 chromids (pRL12 and

Table 1. Genes that are specific to the species *R. brockwellii* (gsA)

Protein	Description	Replicon*	Percentage of gsA	Percentage of non-gsA
WP_033181355.1	type II toxin–antitoxin system ParD family antitoxin	chr	100	4.5
WP_033181354.1	type II toxin–antitoxin system RelE/ParE family toxin	chr	100	3.2
WP_050516830.1	hypothetical protein	chr	97	0.7
WP_033182791.1	response regulator transcription factor	chr	97	0.7
WP_050516831.1	histidine kinase	chr	97	0.7
WP_033182837.1	voltage-gated chloride channel family protein	chr	100	2.0
WP_033182838.1	EamA family transporter	chr	100	2.5
WP_082229827.1	dual specificity protein phosphatase family protein	chr	100	0.5
WP_245483449.1	hypothetical protein	chr	100	3.2
WP_050516784.1	esterase	chr	100	4.2
WP_033182350.1	hypothetical protein	chr	100	0.0
WP_130708130.1	PAS domain S-box protein	Rh05	97	0.0
WP_033183077.1	gamma-glutamyltransferase	Rh01	100	3.2
WP_033183076.1	ABC transporter permease	Rh01	100	3.2
WP_033183075.1	ABC transporter permease	Rh01	100	3.2
WP_033183074.1	ABC transporter ATP-binding protein	Rh01	100	3.2
WP_245483569.1	BMP family protein	Rh01	100	3.2
WP_033183072.1	cysteine hydrolase	Rh01	100	3.2
WP_050516841.1	amidase family protein	Rh01	100	3.2
WP_033183071.1	TetR family transcriptional regulator	Rh01	100	3.2
WP_245494088.1	HlyD family secretion protein	Rh01	100	2.5
WP_033184559.1	hypothetical protein	Rh01	100	4.2

*Location of the gene in the genome of the type strain.

pRL11, respectively) and the Rh05 plasmid (pRL9). A further 34 proteins met these abundance thresholds but were present in gsJ, so represent synapomorphies of the gsB–gsJ clade.

For *R. beringeri* (gsD), with the type strain SM51[T], there are 15 proteins that are diagnostic characteristics of the species. They are in all eight gsD genomes but in fewer than 5% of other genomes, and are completely absent from the closely related gsE and from the two unclassified strains Norway and CC278f that are the sister group of gsD. They are largely encoded on the chromosome and the Rh01-type chromid (Table 3). A further 29 proteins were in all gsD and fewer than 5% of other genomes, but were present in a few gsE or in Norway or CC278f.

In these tables, we have highlighted only those genes that are strongly associated with just one species; there are many other genes that are shared with some, but not all, of the other species of the Rlc. Taken together, these genes will provide each species with a distinctive set of phenotypic properties, but we have not speculated on what these might be. For taxonomic purposes, it is easier and more reliable to determine the presence of genes than to measure complex and potentially environment-specific phenotypes. We do not yet have a clear understanding of the interplay of genetic and ecological processes that lead to the creation and maintenance of well-defined but closely related bacterial species in a species complex such as the Rlc. An intriguing first indication of ecological differences between species in the Rlc is provided by a recent study of pairwise strain interactions that showed, for example, that gsA strains were inhibited by medium in which gsE strains have grown, but growth enhancement was seen in the reciprocal experiment [22]. There is much more to do, but an important first step is to provide a clear definition of the species involved. In this study, we have defined three widespread species, but there are other species within the Rlc that are yet to be named [6].

Table 2. Genes that are specific to the species *R. johnstonii* (gsB)

Protein	Description	Replicon*	Percentage of gsB	Percentage of non-gsB
WP_206778323.1	CopD family protein	chr	98	4.7
WP_155251664.1	hypothetical protein	chr	98	0.3
WP_011650960.1	arylsulfatase	chr	96	3.6
WP_011650961.1	arylsulfatase	chr	96	3.6
WP_086935541.1	arylsulfatase	chr	96	3.6
WP_222277485.1	DUF3482 domain-containing protein	chr	98	2.1
WP_011652379.1	PIG-L family deacetylase	chr	98	2.1
WP_028743325.1	AAA family ATPase	Rh05	100	2.8
WP_011655074.1	Cu(I)-responsive transcriptional regulator	Rh02	98	4.7
WP_011655075.1	heavy metal translocating P-type ATPase	Rh02	98	4.7
WP_011655224.1	patatin-like phospholipase family protein	Rh02	100	3.9
WP_011655225.1	GMC family oxidoreductase	Rh02	100	3.4
WP_011655226.1	hypothetical protein	Rh02	100	3.4
WP_011655227.1	hypothetical protein	Rh02	100	3.4
WP_011655228.1	hypothetical protein	Rh02	100	3.4
WP_041936931.1	hypothetical protein	Rh02	100	3.9
WP_011655230.1	NADH:flavin oxidoreductase	Rh02	100	3.9
WP_049778478.1	hypothetical protein	Rh02	100	2.8
WP_028742514.1	MFS transporter	Rh01	98	1.3
WP_028742513.1	LysR substrate-binding domain-containing protein	Rh01	98	1.3
WP_011649338.1	hypothetical protein	Rh01	100	0.5

*Location of the gene in the genome of the type strain.

DESCRIPTION OF *RHIZOBIUM BROCKWELLII* SP. NOV.

R. brockwellii (brock.well'i.i N.L. masc. gen. n. *brockwellii*, of Brockwell, to honour John Brockwell, who isolated the type strain and made major contributions to the development of rhizobial inoculation in Australia).

The type strain is Brockwell CC275e^T=ICMP 2163^T = NZP 561^T=LMG 6122^T= HAMBI 13^T = ATCC 35181^T. It was isolated from a nodule on *Trifolium repens* growing near Cape Grim, Tasmania, Australia in 1966 by John Brockwell (pers. comm.) and has been a widely used clover inoculant in Australia and New Zealand [23]. The genome sequence [23] is INSDC accession GCA_000769405.2 and the 16S rRNA gene sequence is OQ226180.

This description is based on 37 isolates with published genome sequences (Table S1). They were isolated in Denmark, Russia, Belarus, Colombia and Australia. The species has also been found in the UK [3].

Members of this species have ANI above 96% with the type strain, and, in a phylogeny based on core gene sequences, group with the type strain in a clade that excludes members of other species. The species is genospecies A of the *Rhizobium leguminosarum* species complex, as defined by Young *et al.* [6] and s__Rhizobium leguminosarum_E in the Genome Taxonomy Database release 07_RS207 [24].

The type strain has a POCP of 85.27% with USDA 2370^T, the type strain of *R. leguminosarum*, the type species of the genus. This confirms that the novel species is a member of the genus *Rhizobium*. Furthermore, the sequence of the 16S rRNA gene of the type strain is identical to that of USDA 2370^T. This 16S rRNA sequence is also found in 31 of the other genomes of strains of *R. brockwellii*, but a further five genomes share a single substitution, C1076T [6].

The genome size range is 6.97–7.40 Mb (that of the type strain is 7.12 Mb). The genomic DNA G+C content range is 60.64–60.88% (type strain 60.88%).

Table 3. Genes that are specific to the species *R. beringeri* (gsD)

Protein	Description	Replicon*	Percentage of gsD	Percentage of non-gsD
WP_130766944.1	orotate phosphoribosyltransferase	Rh01	100	0.0
WP_128403411.1	hypothetical protein	Rh01	100	2.5
WP_128403413.1	hypothetical protein	Rh01	100	2.3
WP_128403417.1	sulfurtransferase TusA family protein	Rh01	100	0.0
WP_245461731.1	DUF3396 domain-containing protein	Rh01	100	0.0
WP_233943409.1	DUF4150 domain-containing protein	Rh01	100	0.0
WP_128403146.1	nucleotidyl transferase AbiEii/AbiGii toxin family protein	chr	100	3.9
WP_130766704.1	type IV toxin-antitoxin system AbiEi family antitoxin	chr	100	3.5
WP_245504877.1	alpha/beta hydrolase	chr	100	4.2
WP_245504879.1	hypothetical protein	chr	100	0.5
WP_128400621.1	hypothetical protein	chr	100	0.0
WP_128403000.1	cell envelope integrity protein TolA	chr	100	3.2
WP_130766732.1	TetR/AcrR family transcriptional regulator	chr	100	0.9
WP_130660113.1	hypothetical protein	unplaced	100	0.0
WP_128401342.1	hypothetical protein	unplaced	100	0.0

*Location of the gene in the genome of the type strain.

In addition to the two chromids that are universal in the Rlc (plasmid types Rh01 and Rh02), strains of *R. brockwellii* carry plasmids of types Rh03 (37 out of 37 strains), Rh05 (37 out of 37) and Rh06 (36 out of 37). Other plasmids are present in some strains, though Rh07 has not been detected.

Table 1 lists 22 genes that are present in 97–100% of strains of *R. brockwellii* but are absent from the sister species gsH and are rare elsewhere in the Rlc (<5% of genomes). In the type strain, they are predominantly carried on the chromosome or the Rh01 chromid (pRltCC275eF). These genes are diagnostic characteristics of the species.

All 37 isolates characterised so far are sv. *trifolii* and nodulate clovers (*Trifolium* spp.), but this should not be considered a defining feature of the species.

DESCRIPTION OF *RHIZOBIUM JOHNSTONII* SP. NOV.

R. johnstonii (john.ston'i.i N.L. masc. gen. n. *johnstonii*, of Johnston, to honour Andrew Johnston, whose pioneering research into the genetics of *Rhizobium* made extensive use of the type strain.

The type strain is John Innes Centre strain 3841[T] = LMG 32736[T]=DSM 114642[T]. It is a spontaneous streptomycin-resistant mutant of strain 300, which was isolated in 1970 from a field nodule of pea (*Pisum sativum*) in Norfolk, UK by John Beringer (pers. comm.) [25]. The genome sequence [19] is INSDC accession GCA_000009265.1 and the 16S rRNA gene sequence is OQ226178.

This description is based on 54 isolates with published genome sequences (Table S1). They were isolated in the UK, France, Spain, Serbia, Russia, China and Argentina.

Members of this species have ANI above 96% with the type strain, and, in a phylogeny based on core gene sequences, group with the type strain in a clade that excludes members of other species. The species is genospecies B of the *Rhizobium leguminosarum* species complex, as defined by Young *et al.* [6], and is included in s__Rhizobium leguminosarum_L in the Genome Taxonomy Database release 07_RS207 [24], which also encompasses the related genospecies J and K.

The type strain 3841[T] has a POCP of 84.43% with USDA 2370[T], the type strain of *R. leguminosarum*, the type species of the genus. This confirms that the novel species is a member of the genus *Rhizobium*. The sequence of the 16S rRNA gene of the type strain differs from that of USDA 2370[T] by a single nucleotide substitution (C1076T) and all strains of the species have the same 16S sequence as the type strain [6].

The genome size range is 7.42–8.30 Mb (that of the type strain is 7.75 Mb). The genomic DNA G+C content range is 60.52–60.97% (type strain 60.86%).

In addition to the two chromids that are universal in the Rlc (plasmid types Rh01 and Rh02), strains of *R. johnstonii* carry plasmids of types Rh03 and Rh05 (54 out of 54 strains). Other plasmids are present in some strains.

Table 2 lists 21 genes that are present in 97–100% of strains of *R. johnstonii* but are absent from the sister species gsJ and are rare elsewhere in the Rlc (<5% of genomes). These genes are diagnostic characteristics of the species. The type strain is sv. *viciae*, but sv. *viciae* and sv. *trifolii* are both well represented among the strains of this species.

DESCRIPTION OF *RHIZOBIUM BERINGERI* SP. NOV.

R. beringeri (ber.in.ger'i N.L. masc. gen. n. *beringeri*, of Beringer, to honour John Beringer, who initiated the genetic study of *Rhizobium leguminosarum* and made important early discoveries.

The type strain is SM51^T = LMG 32895^T = DSM 115206^T. It was isolated from a field nodule of white clover (*Trifolium repens*) collected on 15 June 2015 in Store Heddinge, Denmark (55.334401, 12.380648) by Sara Moeskjær and colleagues [4]. The genome sequence [4] is INSDC accession GCA_004306515 and the 16S rRNA gene sequence is OQ226177.

This description is based on eight isolates with published genome sequences (Table S1). They were isolated in Denmark, France and Australia. The species has also been found in the UK [3].

Members of this species have ANI above 96% with the type strain, and, in a phylogeny based on core gene sequences, group with the type strain in a clade that excludes members of other species. The species is genospecies D of the *Rhizobium leguminosarum* species complex, as defined by Young *et al.* [6], and is included in s__Rhizobium leguminosarum_K in the Genome Taxonomy Database release 07_RS207 [24], which also encompasses the related strains CC278f and Norway that are not included in *R. beringeri*.

The type strain SM51^T has a POCP of 84.03% with USDA 2370^T, the type strain of *R. leguminosarum*, the type species of the genus. This confirms that the novel species is a member of the genus *Rhizobium*. The sequence of the 16S rRNA gene of the type strain differs from that of USDA 2370^T by a single nucleotide substitution (C1076A) and all strains of the species have the same 16S sequence as the type strain [6].

The genome size range is 7.48–7.91 Mb (that of the type strain is 7.57 Mb). The genomic DNA G+C content range is 60.59–60.81% (type strain 60.77%).

In addition to the two chromids that are universal in the Rlc (plasmid types Rh01 and Rh02), strains of *R. beringeri* carry plasmids of types Rh04 and (in seven out of eight strains) Rh08. Other plasmids are present in some strains.

Table 3 lists 15 genes that are present in 100% of strains of *R. beringeri* but are absent from the closely related species gsE and isolates Norway and CC278f and are rare elsewhere in the Rlc (<5% of genomes). These genes are diagnostic characteristics of the species. Among the eight characterised strains of *R. beringeri*, seven (including the type strain) are sv. *trifolii*, but one is sv. *viciae*.

NOTE ADDED IN PROOF: *RHIZOBIUM ACACIAE*

While the present study was under review, Hsouna *et al.* [26] published the name *Rhizobium acaciae* for Rlc gsJ and provided genome sequences and descriptions of three additional strains. Since this is the sister taxon of *R. johnstonii*, we carried out some additional analyses to include the three novel genomes. A phylogeny of the relevant part of the Rlc, using the same genes and methods as Fig. 1, confirmed that all five strains of *R. acaciae* form a well-resolved clade that is the sister group of *R. johnstonii* (Fig. S1). The lowest ANI between two strains of *R. acaciae* was 98.06%, while the highest ANI between strains of *R. acaciae* and *R. johnstonii* was 96.05%, confirming the clear gap between the two species. None of the 21 genes listed in Table 2 are present in any of the five *R. acaciae* genomes, so these genes remain specific to *R. johnstonii*. All the evidence therefore confirms that *R. acaciae* and *R. johnstonii* are distinct species.

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

References

1. Young JPW, Jorrin B, Moeskjær S, James EK. *Rhizobium brockwellii* sp. nov., *Rhizobium johnstonii* sp. nov. and *Rhizobium beringeri* sp. nov., three genospecies within the *Rhizobium leguminosarum* species complex. *Figshare* 2023. DOI: 10.6084/m9.figshare.22734917.

2. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.

3. Kumar N, Lad G, Giuntini E, Kaye ME, Udomwong P, et al. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. *Open Biol* 2015;5:140133.

4. Cavassim MIA, Moeskjær S, Moslemi C, Fields B, Bachmann A, et al. Symbiosis genes show a unique pattern of introgression and selection within a *Rhizobium leguminosarum* species complex. *Microb Genom* 2020;6:e000351.

5. Boivin S, Ait Lahmidi N, Sherlock D, Bonhomme M, Dijon D, et al. Host-specific competitiveness to form nodules in *Rhizobium leguminosarum* symbiovar viciae. *New Phytol* 2020;226:555–568.

6. Young JPW, Moeskjær S, Afonin A, Rahi P, Maluk M, et al. Defining the *Rhizobium leguminosarum* species complex. *Genes* 2021;12:111.

7. Saïdi S, Ramírez-Bahena M-H, Santillana N, Zúñiga D, Álvarez-Martínez E, et al. *Rhizobium laguerreae* sp. nov. nodulates *Vicia faba* on several continents. *Int J Syst Evol Microbiol* 2014;64:242–247.

8. Jorrin B, Palacios JM, Peix Á, Imperial J. *Rhizobium ruizarguesonis* sp. nov., isolated from nodules of *Pisum sativum* L. *Syst Appl Microbiol* 2020;43:126090.

9. Jiao YS, Yan H, Ji ZJ, Liu YH, Sui XH, et al. *Rhizobium sophorae* sp. nov. and *Rhizobium sophoriradicis* sp. nov., nitrogen-fixing rhizobial symbionts of the medicinal legume *Sophora flavescens*. *Int J Syst Evol Microbiol* 2015;65:497–503.

10. Rahi P, Giram P, Chaudhari D, diCenzo GC, Kiran S, et al. *Rhizobium indicum* sp. nov., isolated from root nodules of pea (*Pisum sativum*) cultivated in the Indian trans-Himalayas. *Syst Appl Microbiol* 2020;43:126127.

11. Zhang J, Peng S, Andrews M, Liu C, Shang Y, et al. *Rhizobium changzhiense* sp. nov., isolated from effective nodules of *Vicia sativa* L. in North China. *Int J Syst Evol Microbiol* 2019;71:004724.

12. Parker CT, Tindall BJ, Garrity GM. International Code of Nomenclature of Prokaryotes. *Int J Syst Evol Microbiol* 2019;69:S1–S111.

13. Ormeño-Orrillo E, Martínez-Romero E. Phenotypic tests in *Rhizobium* species description: an opinion and (a sympatric speciation) hypothesis. *Syst Appl Microbiol* 2013;36:145–147.

14. Hitch TCA, Riedel T, Oren A, Overmann J, Lawley TD, et al. Automated analysis of genomic sequences facilitates high-throughput and comprehensive description of bacteria. *ISME Commun* 2021;1:16.

15. Qin Q-L, Xie B-B, Zhang X-Y, Chen X-L, Zhou B-C, et al. A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol* 2014;196:2210–2215.

16. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. *Proc Natl Acad Sci* 2009;106:19126–19131.

17. Palmer M, Steenkamp ET, Blom J, Hedlund BP, Venter SN. All ANIs are not created equal: implications for prokaryotic species boundaries and integration of ANIs into polyphasic taxonomy. *Int J Syst Evol Microbiol* 2020;70:2937–2948.

18. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 2018;9:5114.

19. Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, et al. The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. *Genome Biol* 2006;7:R34.

20. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 2010;11:119.

21. Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* 2021;18:366–368.

22. Fields B, Moffat EK, Harrison E, Andersen SU, Young JPW, et al. Genetic variation is associated with differences in facilitative and competitive interactions in the *Rhizobium leguminosarum* species complex. *Environ Microbiol* 2022;24:3463–3485.

23. Perry BJ, Ferguson S, Laugraud A, Wakelin SA, Reeve W, et al. Complete genome sequences of *Trifolium* spp. inoculant strains *Rhizobium leguminosarum* sv. *trifolii* Ta1 and Cc275E: resources for genomic study of the *Rhizobium–Trifolium* symbiosis. *Mol Plant-Microbe Interactions* 2020.

24. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, et al. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 2022;50:D785–D794.

25. Johnston AWB, Beringer JE. Identification of the *Rhizobium* strains in pea root nodules using genetic markers. *J Gen Microbiol* 1975;87:343–350.

26. Hsouna J, Ilahi H, Han J-C, Gritli T, Ellouze W, et al. *Rhizobium acaciae* sp. nov., a new nitrogen-fixing symbiovar isolated from root nodules of *Acacia saligna* in Tunisia. *Int J Syst Evol Microbiol* 2023;73.

27. Kozlov AM, Darriba D, Flouri T, Morel B, Stamatakis A. RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. *Bioinformatics* 2019;35:4453–4455.

28. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Res* 2021;49:W293–W296.