High-quality permanent draft genome sequence of *Rhizobium leguminosarum* bv. *viciae* strain GB30; an effective microsymbiont of *Pisum sativum* growing in Poland

Andrzej Mazur¹, Sofie E. De Meyer², Rui Tian², Jerzy Wielbo¹, Kamil Zebracki¹, Rekha Seshadri³, TBK Reddy³, Victor Markowitz⁴, Natalia N. Ivanova³, Amrita Pati³, Tanja Woyke³, Nikos C. Kyrpides³,⁵ and Wayne Reeve²*

**Abstract**

*Rhizobium leguminosarum* bv. *viciae* GB30 is an aerobic, motile, Gram-negative, non-spore-forming rod that can exist as a soil saprophyte or as a legume microsymbiont of *Pisum sativum*. GB30 was isolated in Poland from a nodule recovered from the roots of *Pisum sativum* growing at Janow. GB30 is also an effective microsymbiont of the annual forage legumes vetch and pea. Here we describe the features of *R. leguminosarum* bv. *viciae* strain GB30, together with sequence and annotation. The 7,468,464 bp high-quality permanent draft genome is arranged in 78 scaffolds of 78 contigs containing 7,227 protein-coding genes and 75 RNA-only encoding genes, and is part of the GEBA-RNB project proposal.

**Keywords:** Root-nodule bacteria, Nitrogen fixation, Rhizobia, Alphaproteobacteria, GEBA-RNB

**Introduction**

The most efficient biological nitrogen fixation occurs when bacterial microsymbionts (rhizobia) form an effective symbiotic association with legume host plants. Legumes can develop these interactions with many different species of rhizobia belonging mainly to the Alphaproteobacteria, including *Azorhizobium*, *Allorhizobium*, *Bradyrhizobium*, *Ensifer*, *Mesorhizobium* and *Rhizobium* [1, 2]. The genus *Rhizobium* contains at the time of writing 71 species, and within a species there may be distinct symbiovars [3].

Within the species *Rhizobium leguminosarum*, there are three distinct symbiovars [4, 5] including bv. *phaseoli* that forms nodules with *Phaseolus vulgaris*, bv. *trifolii* that forms nodules with clover (*Trifolium*) and bv. *viciae* that forms nodules on vetch, pea and lentil (*Vicia, Lathyrus, Pisum* and *Lens*). In *R. leguminosarum* the *nod* genes that define these distinct host specificities are mostly located on the symbiotic plasmid, which has genetically been designated pSym. The genomes of *R. leguminosarum* strains are usually large and complex containing, in addition to pSym, a chromosomal replicon and extra-chromosomal low-copy-number replicons characterized by the presence of repABC replication systems [6–8]. Recent studies have revealed that substantial divergence can occur in this genome organization and in the metabolic versatility of *R. leguminosarum* isolates [5, 9–12]. Kumar et al. [5] demonstrated that the diversity of *R. leguminosarum* within a local population of nodule isolates was 10 times higher than that found for *Ensifer medicae*. It was noted that the abundance of a particular genotype within the population can vary significantly and adaptation to the edaphic environment is a sought after trait particularly for the development of inoculants [13, 14].

*R. leguminosarum* bv. *viciae* GB30 was isolated as the most abundant nodule inhabitant (>42 %) of *Pisum sativum* cv. Ramrod plants cultivated at a field site in Janow, Poland [10]. In contrast to other abundant isolates, GB30 formed nodules and fixed nitrogen with both *P. sativum* and *Vicia villosa* (cv. Wista). Preliminary investigation into the genome architecture using Eckhardt analysis has revealed that GB30 contained a multipartite genome consisting of six replicons with one chromosome.

* Correspondence: W.Reeve@murdoch.edu.au
²Centre for Rhizobium Studies, Murdoch University, Murdoch, Western Australia
Full list of author information is available at the end of the article

© 2015 Mazur et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly credited. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
and five plasmids [10]. The genome of this strain could therefore provide important insights into the mechanisms required by effective \textit{R. leguminosarum} microsymbionts to adapt to a particular edaphic environment. Here, we present a set of general features for \textit{Rhizobium leguminosarum} bv. \textit{viciae} GB30 together with the description of the complete genome sequence and annotation.

**Organism information**

**Classification and features**

\textit{R. leguminosarum} bv. \textit{viciae} strain GB30 is a motile, Gram-negative rod in the order \textit{Rhizobiales} of the class \textit{Alphaproteobacteria}. The rod-shaped form varies in size with dimensions of 0.8-1 \textmu m in width and 2.3-2.5 \textmu m in length (Fig. 1 Left and Center). It is fast growing, forming colonies within 3–4 days when grown on half strength Lupin Agar (½LA) [15] at 28 °C. Colonies on ½LA are white-opaque, slightly domed and moderately mucoid with smooth margins (Fig. 1 Right).

Figure 2 shows the phylogenetic relationship of \textit{Rhizobium leguminosarum} bv. \textit{viciae} strain GB30 in a 16S rRNA gene sequence based tree. This strain is phylogenetically most related to \textit{Rhizobium laguerreae} FB206\(^T\) and \textit{Rhizobium gallicum} R602sp\(^T\) based on the 16S rRNA gene alignment with sequence identities of 100 %, as determined using the EzTaxon-e server [16]. \textit{Rhizobium laguerreae} FB206\(^T\) was isolated from effective \textit{Vicia faba} root nodules in Tunisia [17], whereas \textit{Rhizobium gallicum} R602sp\(^T\) was isolated from effective \textit{Phaseolus vulgaris} root nodules in France [18]. Sequence similarity was also investigated with strains from the GEBA-RNB project [12] and GB30 was found to be closely related to \textit{R. leguminosarum} bv. \textit{trifolii} WSM1689 with 100 % 16S rRNA gene sequence identity. \textit{R. leguminosarum} bv. \textit{trifolii} WSM1689 is a highly effective microsymbiont of the perennial clover \textit{Trifolium uniflorum} and has been shown to have a remarkable narrow host range [19]. Minimum Information about the Genome Sequence (MIGS) is provided in Table 1 and Additional file 1: Table S1.

**Symbiotaxonomy**

\textit{R. leguminosarum} bv. \textit{viciae} strain GB30 was obtained from pea nodules (\textit{P. sativum} cv. Ramrod) growing in sandy loam (N:P:K 0.157:0.014:0.013 %) in Janow near Lublin (Poland). The soil contained a relatively high number of \textit{R. leguminosarum} bv. \textit{viciae}, bv. \textit{trifolii} and bv. \textit{phaseoli} cells i.e., \(9.2 \times 10^3, 4.2 \times 10^3\) and \(1.5 \times 10^3\) bacteria/g of soil, respectively, as determined by the most probable number (MPN) method [10]. Plants were grown on 1 m\(^2\) plot for six weeks between May and June, 2008. Five randomly chosen pea plants growing in each other’s vicinity were harvested; the nodules were collected, surface-sterilized and the microsymbionts isolated [10]. One of the most abundant isolates, GB30, formed nodules (Nod\(^+\)) and fixed N\(_2\) (Fix\(^+\)) with \textit{P. sativum} and \textit{Vicia villosa} (cv. Wista) increasing the wet mass weight by 54 and 38 %, respectively. Plants inoculated with GB30 also showed a 2.6 fold increase in nodule number and a 2.2 fold increase in seed pod number.

**Genome sequencing and annotation information**

**Genome project history**

This organism was selected for sequencing on the basis of its environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance, and is part of the Genomic Encyclopedia of Bacteria and Archaea, The Root Nodulating Bacteria chapter (GEBA-RNB) project at the U.S. Department of Energy, Joint Genome Institute [12]. The genome project is deposited in the Genomes OnLine Database [20] and the high-quality permanent draft genome sequence in IMG [21]. Sequencing, finishing and annotation were performed by the JGI using state of the art sequencing technology [22]. A summary of the project information is shown in Table 2.

**Growth conditions and genomic DNA preparation**

\textit{R. leguminosarum} bv. \textit{viciae} strain GB30 was grown to mid logarithmic phase in TY rich media [23] on a gyratory shaker at 28 °C. DNA was isolated from 60 mL of cells...
Fig. 2 Phylogenetic tree highlighting the position of *Rhizobium leguminosarum* bv. *viciae* GB30 (shown in blue print) relative to other type and non-type strains in the *Rhizobium* genus using a 901 bp internal region of the 16S rRNA gene. *Bradyrhizobium elkanii* ATCC 49852T was used as outgroup. All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [36]. The tree was built using the maximum likelihood method with the General Time Reversible model. Bootstrap analysis with 500 replicates was performed to assess the support of the clusters. Type strains are indicated with a superscript T. Strains with a genome sequencing project registered in GOLD [20] are shown in bold and have the GOLD ID mentioned after the strain number, otherwise the NCBI accession number has been provided. Finished genomes are designated with an asterisk.

*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0009658)  
*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0009658)  
*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0007437)  
*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0007356)  
*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0007391)  
*Rhizobium leguminosarum* bv. *viciae* GB30 (Gp0007351)  
*Rhizobium leguminosarum* bv. *viciae* WSM1455 (Gp0007354)  
*Rhizobium phaseoli* ATCC 14482T (EF141340)  
*Rhizobium fabae* LMG 23997T (DQ835306)  
*Rhizobium pisi* DSM 30132T (AY509899)  
*Rhizobium etli* USDA 9032T (U28916)  
*Rhizobium miluense* CCBAU 41251T (EF061096)  
*Rhizobium tropici* CIAT899T (Gp0006704)*  
*Rhizobium tropici* CIAT899T (Gp0006704)*  
*Rhizobium tibeticum* CCBAU 85039T (EU256404)  
*Rhizobium tubonense* CCBAU 85046T (EU256434)  
*Rhizobium mongolense* USDA1844T (Gp0010241)  
*Rhizobium yanglingense* SH22623T (AF003375)  
*Rhizobium alamii* LMG 24466T (AM931436)  
*Rhizobium mesosinicum* CCBAU25010T (DQ100063)  
*Rhizobium sutteri* WSM1592 (Gp0010240)  
*Rhizobium sutteri* IS 123T (Y10170)  
*Rhizobium loessense* CCBAU 71908T (AF364069)  
*Rhizobium galegae* LMG 6214T (X67226)  
*Rhizobium vignae* CCBAU 05176T (GU128881)  
*Rhizobium vitis* LMG 8750T (X67225)  
*Rhizobium taibaishanense* CCNWSX 0483T (HM776997)  
*Rhizobium soli* DS-42T (EF363715)  
*Rhizobium larrymoorei* LMG 21410T (NR 026519)  
*Rhizobium radiobacter* ATCC 19358T (AJ389904)  
Bradyrhizobium elkanii ATCC 49852T (AF362942)
using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [24].

**Genome sequencing and assembly**

The draft genome of *Rhizobium leguminosarum* bv. *viciae* GB30 was generated at the DOE Joint Genome Institute [22]. An Illumina Std shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 25,943,396 reads totaling 3,891.5 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found at the JGI web site [25]. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artefacts (Mingkun L, Copeland A, Han J. unpublished). Following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet version 1.1.04 [26] (2) 1–3 Kbp simulated paired end reads were created

### Table 1

| MIGS ID | Property | Term                          | Evidence code |
|---------|----------|-------------------------------|---------------|
|         |          | Domain Bacteria               | TAS [39]      |
|         |          | Phylum Proteobacteria         | TAS [40, 41]  |
|         |          | Class Alphaproteobacteria     | TAS [42, 43]  |
|         |          | Classification Order          | TAS [44]      |
|         |          | Family Rhizobiaceae           | TAS [45]      |
|         |          | Genus Rhizobium               | TAS [46]      |
|         |          | Species Rhizobium leguminosarum | TAS [47–49] |
|         | Gram stain | Negative                      | IDA           |
|         | Cell shape | Rod                           | IDA           |
|         | Motility   | Motile                        | IDA           |
|         | Sporulation| Non-sporulating                | NAS           |
|         | Temperature range | Mesophile                 | NAS           |
|         | Optimum temperature | 28 °C                      | TAS [9]       |
|         | pH range; Optimum | Not reported                |               |
|         | Carbon source | Not reported                 |               |
|         | Habitat    | Soil, root nodule, on host   | TAS [9]       |
| MIGS-6  | Salinity   | Non-halophile                 | NAS           |
| MIGS-6.3| Oxygen requirement | Aerobic                   | TAS [49]      |
| MIGS-15 | Biotic relationship | Free living, symbiotic     | TAS [10]      |
| MIGS-14 | Pathogenicity | Non-pathogenic              | TAS [50]      |
| MIGS-4  | Geographic location | Janow, near Lublin, eastern Poland | TAS [10] |
| MIGS-5  | Sample collection | Between May and June, 2008   | TAS [10]      |
| MIGS-4.1| Latitude   | 51.387638                     | TAS [10]      |
| MIGS-4.2| Longitude  | 22.369194                     | TAS [10]      |
| MIGS-4.3| Altitude   | 185 m                         | IDA           |

Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature); NAS: Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [51].

### Table 2

| MIGS ID | Property               | Term                                      |
|---------|------------------------|-------------------------------------------|
|         | Finishing quality      | High-quality permanent draft              |
| MIGS-28 | Libraries used         | Illumina Std PE                           |
| MIGS-29 | Sequencing platforms   | Illumina HiSeq 2000                       |
| MIGS-31.2| Fold coverage        | 121.9 x Illumina                          |
| MIGS-30 | Assemblers            | Velvet version 1.1.04; ALLPATHS v. h1043 |
| MIGS-32 | Gene calling methods  | Prodigal 1.4                              |
|         | Locus Tag              | A3A3                                      |
|         | GenBank ID             | ATTP00000000                               |
|         | GenBank Date of Release| July 9, 2013                               |
|         | GOLD ID                | Gp0009658 [52]                            |
|         | BIOPROJECT             | PRJNA165299                                |
| MIGS-13 | Source Material Identifier | GB30                               |
|         | Project relevance      | Symbiotic N₂ fixation, agriculture       |
from Velvet contigs using wgsim [27] (3) Illumina reads were assembled with simulated read pairs using Allpaths–LG (version r41043) [28]. Parameters for assembly steps were: 1) Velvet (velveth: 63 –shortPaired and velvetg: –very_clean yes –export-Filtered yes –min_contig_lgth 500 –scaffolding no –cov_cutoff 10) 2) wgsim (–e 0 –1 100 –2 100 –r 0 –R 0 –X 0) 3) Allpaths–LG (PrepareAllpaths: PHRED_64 = 1 PLOIDY = 1 FRAG_COVERAGE = 125 JUMP_COVERAGE = 25 LONG_JUMP_COV = 50, RunAllpathsLG: THREADS = 8 RUN = std_shredpairs TARGETS = standard VAPI_WARN_ONLY = True OVERWRITE = True). The final draft assembly contained 78 contigs in 78 scaffolds. The total size of the genome is 7.5 Mb and the final assembly is based on 910.4 Mb of Illumina data, which provides an average of 121.9× coverage.

Genome annotation
Genes were identified using Prodigal [29], as part of the DOE-JGI genome annotation pipeline [30, 31]. The predicted CDSs were translated and used to search the National Centre for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRfam, Pfam, KEGG, COG, and InterPro databases. The tRNAscanSE tool [32] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [33]. Other non–coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL [34]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) system [35] developed by the Joint Genome Institute, Walnut Creek, CA, USA.

Genome Properties
The genome is 7,468,464 nucleotides with 60.81 % GC content (Table 3) and comprised of 78 scaffolds of 78 contigs. From a total of 7,302 genes, 7,227 were protein encoding and 75 RNA only encoding genes. The majority of genes (79.57 %) were assigned a putative function whilst the remaining genes were annotated as hypothetical. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3 Genome Statistics for Rhizobium leguminosarum bv. viciae strain GB30

| Attribute                  | Value         | % of Total |
|----------------------------|---------------|------------|
| Genome size (bp)           | 7,468,464     | 100.00     |
| DNA coding (bp)            | 6,497,898     | 87.00      |
| DNA G + C (bp)             | 4,541,558     | 60.81      |
| DNA scaffolds              | 78            | 1.00       |
| Total genes                | 7,302         | 100.00     |
| Protein coding genes       | 7,227         | 98.97      |
| RNA genes                  | 75            | 1.03       |
| Pseudo genes               | 0             | 0          |
| Genes in internal clusters | 470           | 6.44       |
| Genes with function prediction | 5,810     | 79.57      |
| Genes assigned to COGs     | 5,182         | 70.97      |
| Genes with Pfam domains    | 6,025         | 82.51      |
| Genes with signal peptides | 634           | 8.68       |
| Genes with transmembrane proteins | 1,646 | 22.54      |
| CRISPR repeats             | 1             |            |

Table 4 Number of genes associated with the general COG functional categories.

| Code | Value | % age | Description                                      |
|------|-------|-------|--------------------------------------------------|
| J    | 233   | 3.90  | Translation, ribosomal structure and biogenesis  |
| A    | 0     | 0.00  | RNA processing and modification                  |
| K    | 597   | 9.98  | Transcription                                    |
| L    | 128   | 2.14  | Replication, recombination and repair            |
| B    | 2     | 0.03  | Chromatin structure and dynamics                 |
| D    | 35    | 0.59  | Cell cycle control, Cell division, chromosome partitioning |
| V    | 119   | 1.99  | Defense mechanisms                               |
| T    | 285   | 4.77  | Signal transduction mechanisms                   |
| M    | 310   | 5.18  | Cell wall/membrane/envelope biogenesis           |
| N    | 93    | 1.56  | Cell motility                                    |
| U    | 58    | 0.97  | Intracellular trafficking, secretion, and vesicular transport |
| O    | 206   | 3.44  | Posttranslational modification, protein turnover, chaperones |
| C    | 325   | 5.43  | Energy production and conversion                 |
| G    | 644   | 10.77 | Carbohydrate transport and metabolism            |
| E    | 689   | 11.52 | Amino acid transport and metabolism              |
| F    | 116   | 1.94  | Nucleotide transport and metabolism              |
| H    | 270   | 4.52  | Coenzyme transport and metabolism                |
| I    | 241   | 4.03  | Lipid transport and metabolism                   |
| P    | 317   | 5.30  | Inorganic ion transport and metabolism           |
| Q    | 186   | 3.11  | Secondary metabolite biosynthesis, transport and catabolism |
| R    | 695   | 11.62 | General function prediction only                 |
| S    | 381   | 6.37  | Function unknown                                 |
| -    | 2,120 | 29.03 | Not in COGs                                     |

The total is based on the total number of protein coding genes in the genome.

Conclusion
Rhizobium leguminosarum bv. viciae GB30 belongs to a group of Alpha-rhizobia strains isolated from Pisum sativum in Poland. Strain GB30 is part of the GEBARNB project that sequenced 24 R. leguminosarum strains.
and 12 *R. leguminosarum* bv. *viciae* strains [12]. Phylogenetic analysis revealed that GB30 is most closely related to *Rhizobium leguminosarum* bv. *trifolii* CB782 and WSM1689, both part of the GEBEA-RNB project [12]. Full genome comparison of GB30 and WSM1689 [19] revealed that GB30 has the largest genome (7.4 Mbp), with the highest COG count (5,182), the lowest Pfam % (82.51) and the lowest TIGRfam % (22.13 %). The genome attributes of *R. leguminosarum* bv. *viciae* GB30, in conjunction with the other *R. leguminosarum* genomes, will be important for on-going comparative and functional analyses of the plant microbe interactions required for the successful establishment of agricultural crops.

**Additional file**

**Additional file 1: Table S1.** Associated MIGS record.

**Abbreviations**

GEBEA-RNB: Genomic Encyclopedia of Bacteria and Archaea – Root Nodule Bacteria; JGI: Joint Genome Institute; TY: Trypton Yeast; CTAB: Cetyl trimethyl ammonium bromide; WSM: Western Australian Soil Microbiology; MPN: Most probable number; IMG-ER: Integrated Microbial Genomes-Expert Review; NCBI: National Centre for Biotechnology Information.

**Authors’ contribution**

AM supplied the strain and background information for this project, JW initially characterized the strain, TR supplied DNA to JGI and performed all imaging. SDM drafted the paper, AM provided financial support and all other authors were involved in sequencing the genome and/or editing the final manuscript. All authors read and approved the final manuscript.

**Competing interests**

The authors declare that they have no competing interests.

**Acknowledgements**

This work was performed under the auspices of the US Department of Energy’s Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under contract No. DE-AC52-07CH1123, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

**Author details**

1. Department of Genetics and Microbiology, Maria Curie Sklodowska University, Lublin, Poland.
2. Centre for Rhizobium Studies, Murdoch University, Murdoch, Western Australia.
3. DOE Joint Genome Institute, Walnut Creek, California, USA.
4. Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA.
5. Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 18 March 2015 Accepted: 5 June 2015

**Published online:** 16 July 2015

**References**

1. Franche C, Lindstrom K, Elmerich C. Nitrogen-fixing bacteria associated with leguminous and non-leguminous plants. Plant Soil. 2009;213:55–9.
2. Masson-Bovin C, Graud E, Perret X, Batur J. Establishing nitrogen-fixing symbiosis with legumes: how many rhizobium strains? Trends Microbiol. 2009;17:458–66.
3. List of prokaryotic names with standing in nomenclature [http://www.bacterio.net/rhizobium.html](http://www.bacterio.net/rhizobium.html).
4. Rogel MA, Ormeño-Orrillo E, Martínez Romero E. Symbiovars in rhizobia reflect bacterial adaptation to legumes. Syst Appl Microbiol. 2011;34:96–104.
5. Kumar N, Lad G, Guinetti E, Kaye ME, Udomwong P, Shamsani NJ, Young JPW, Bally X. Bacterial genospecies that are not ecologically coherent: population genomics of *Rhizobium leguminosarum*. Open biol 2015, 5.
6. Young JPW, Crossman LC, Johnston AWB, Thomson NR, Ghazoui ZF, Hull KH, Wexler M, Carson ARI, Todd JD, Poole PG, Mauchline TH, East AK, Quail MA, Churcher C, Arrowsmith C, Cheereiac J, Chillingworth T, Clarke K, Cronin A, Davis P, Fraser A, Hance Z, Hauser H, Jagels K, Moule S, Mungall K, Norbertczak H, Rabinowitsch E, Sanders M, Simmonds M, et al.: The genome of *Rhizobium leguminosarum* has recognizable core and accessory components. Genome Biol 2006, 7.
7. Reeves W, O’Hara G, Chain P, Arldley J, Brau L, Nandasena K, et al. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. Stand Genomic Sci. 2010;2:347–56.
8. Reeves W, O’Hara G, Chain P, Arldley J, Brau L, Nandasena K, et al. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM2304, an effective microsymbiont of the South American clover *Trifolium polymorphum*. Stand Genomic Sci. 2010;2:66–76.
9. Welbo J, Marek-Kozczuk M, Mazur A, Kubik-Komar A, Skorupska A. Genetic and metabolic divergence within a *Rhizobium leguminosarum* bv. *trifolii* population recovered from clover nodules. Appl Environ Microbiol. 2010;76:4593–600.
10. Welbo J, Marek-Kozczuk M, Mazur A, Kubik-Komar A, Skorupska A. The structure and metabolic diversity of population of pea microsymbionts isolated from root nodules. BMRI. 2011;1:55–69.
11. Mazur A, Stasiak G, Welbo J, Kubik-Komar A, Marek-Kozczuk M, Skorupska A: Intrageneric diversity of *Rhizobium leguminosarum* bv. *trifolii* clover nodule isolates. BMC Microbiol 2011, 11.
12. Reeves W, Arldley J, Tian R, Eshrai L, Yoon J, Ngamwisetkin P, et al. A genomic encyclopedia of the root nodule bacteria: Assessing genetic diversity through a systematic biogeographic survey. Stand Genomic Sci. 2015;10:114.
13. Howieson J, Ballard R. Optimising the legume symbiosis in stressful and competitive environments within southern Australia - some contemporary thoughts. Soil Biol Biochem. 2004;36:1261–73.
14. Howieson JG, Yates RJ, Foster K, Besier B. Prospects for the future use of legumes. In: Dilworth MJ, James EK, Sprent JI, Newton WE, editors. Leguminous Nitrogen-Fixing Symbioses. London, UK: Elsevier; 2008. p. 363–94.
15. Howieson JG, Ewing MA, D’Antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. Plant Soil. 1988;105:179–88.
16. Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. J Syst Evol Microbiol. 2012;62:716–21.
17. Said S, Ramirez-Bahena M-H, Santillana N, Zúñiga D, Alvarez-Martinez E, Peix A, et al. *Rhizobium leguminosarum* sp. nov. nodulates Vicia faba on several continents. Int J Syst Evol Microbiol. 2014;64:292–2.
18. Amarger N, Macheret V, Laguerre G, Rhizobium gillicum sp. nov. and *Rhizobium gardnii* sp. nov., from *Phaseolus vulgaris* nodules. Int J Syst Bacteriol. 1997;47:996–1006.
19. Terpolilli J, Rui T, Yates R, Howieson J, Poole P, Munk C, et al. Genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689, the microsymbiont of the one flowered clover *Trifolium uniflorum*. Stand Genomic Sci. 2014;9:527–39.
20. Pagani I, Licilios K, Jansson J, Chen IM, Smimova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012;40:D571–579.
21. Markowitz VM, Shen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42:D560–7.
22. Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. PLoS ONE. 2012;7:e48837.
23. Beining JE. R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol. 1974;84:198–9.
24. CTAB DNA extraction protocol [http://jgi.doe.gov/collaborate-with-jgi/promo/overview/protocols-sample-preparation-information/]
25. JGI Website [http://www.jgi.doe.gov]
26. Zerbino D, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9.
27. wgsim [https://github.com/lh3/wgsim]
28. Gnerre S, MacCallum L, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. PNAS USA. 2011;108:1513–8.

29. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:199.

30. Mavromati K, Ivanova NN, Chen IM, Szezo E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard Operating Procedure for the annotations of microbial genomes. Stand Genomic Sci. 2009;1:63–7.

31. Chen IM, Markowitz VM, Chu K, Anderson I, Mavromati K, Kyrpides NC, et al. Improving microbial genome annotations in an integrated database context. PLoS ONE. 2013;8:e54859.

32. Lowe TM, Eddy SR. RfamScan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:955–64.

33. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res. 2007;35:7188–96.

34. INFERNAL. Inference of RNA alignments [http://infernal.janelia.org].

35. Markowitz VM, Mavromati K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG-ER: a system for microbial genome annotation expert review and curation. Bioinformatics. 2009;25:2271–8.

36. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGAS: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9.

37. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. Towards a richer description of our complete collection of genomes and metagenomes. "Minimum Information about a Genome Sequence" (MIGS) specification. Nat Biotechnol. 2008;26:541–7.

38. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, et al. The Genomic Standards Consortium. PLoS Biol. 2011;9:e1001088.

39. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: the domains Bacteria, Archaea and Eucarya. Proc Natl Acad Sci USA. 1990;87:4576–7.

40. Chen WX, Wang ET, Kuykendall LD. The Proteobacteria. New York: Springer-Verlag; 2005.

41. Young JM, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H. Genus I. Rhizobium. In Bergey’s Manual of Systematic Bacteriology. Second ed. New York: Springer-Verlag; 2005.

42. Chen WX, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H. Approvals and taxonomic notes published outside the IJSEM. Int J Syst Evol Microbiol. 2005;55:2235–2238.

43. Young JM, Kuykendall LD, Martinez-Romero E, Kerr A, Sawada H. A revision of Rhizobium Frank 1889, with an emended description of the genus, and the inclusion of all species of Agrobacterium Conn 1942 and Alphazobium undicola de Lajudie et al. 1998 as new combinations. Rhizobium radiobacter, R-rhizogenes, R-rubi, R-undicola and R-viti. Int J Syst Evol Microbiol. 2001;51:89–103.

44. Ramirez-Bahena MH, Garcia-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, et al. Revision of the taxonomic status of the species Rhizobium leguminosarum (Frank 1879) Frank 1889(AL), Rhizobium phaseoli Dangeard 1926(AL) and Rhizobium trifolii Dangeard 1926(AL). R-trifolii is a later synonym of R-leguminosarum. Reclassification of the strain R-leguminosarum DSM 30132 (= NCMB 11478) as Rhizobium pisi sp nov. Int J Syst Evol Microbiol. 2008;58:2484–90.

45. Biological Agents: Technical rules for biological agents. TRBA 466.

46. Skerman V, McGowan V, Sneath P. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225.

47. Skerman V, McGowan V, Sneath P. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–227.

48. Ramirez-Bahena MH, Garcia-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, et al. Revision of the taxonomic status of the species Rhizobium leguminosarum (Frank 1879) Frank 1889(AL), Rhizobium phaseoli Dangeard 1926(AL) and Rhizobium trifolii Dangeard 1926(AL). R-trifolii is a later synonym of R-leguminosarum. Reclassification of the strain R-leguminosarum DSM 30132 (= NCMB 11478) as Rhizobium pisi sp nov. Int J Syst Evol Microbiol. 2008;58:2484–90.