

# Complete genome sequence of *Rhizobium leguminosarum* bv *trifolii* strain WSM2304, an effective microsymbiont of the South American clover *Trifolium polymorphum*

Wayne Reeve<sup>1\*</sup>, Graham O'Hara<sup>1</sup>, Patrick Chain<sup>2,3</sup>, Julie Ardley<sup>1</sup>, Lambert Bräu<sup>1</sup>, Kemanthi Nandesena<sup>1</sup>, Ravi Tiwari<sup>1</sup>, Stephanie Malfatti<sup>2,3</sup>, Hajnalka Kiss<sup>2,3</sup>, Alla Lapidus<sup>2</sup>, Alex Copeland<sup>2</sup>, Matt Nolan<sup>2</sup>, Miriam Land<sup>2,4</sup>, Natalia Ivanova<sup>2</sup>, Konstantinos Mavromatis<sup>2</sup>, Victor Markowitz<sup>5</sup>, Nikos Kyrpides<sup>2</sup>, Vanessa Melino<sup>1</sup>, Matthew Denton<sup>6</sup>, Ron Yates<sup>1,7</sup> & John Howieson<sup>1,7</sup>.

<sup>1</sup> Centre for *Rhizobium* Studies, Murdoch University, Western Australia, Australia

<sup>2</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>3</sup> Lawrence Livermore National Laboratory, Livermore, California, USA

<sup>4</sup> Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA

<sup>5</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>6</sup> Department of Primary Industries, Victoria, Australia

<sup>7</sup> Department of Agriculture and Food, Western Australia, Australia

\*Corresponding author: Wayne Reeve

**Keywords:** microsymbiont, non-pathogenic, aerobic, Gram-negative rod, root-nodule bacteria, nitrogen fixation, *Alphaproteobacteria*

---

*Rhizobium leguminosarum* bv *trifolii* is the effective nitrogen fixing microsymbiont of a diverse range of annual and perennial *Trifolium* (clover) species. Strain WSM2304 is an aerobic, motile, non-spore forming, Gram-negative rod, isolated from *Trifolium polymorphum* in Uruguay in 1998. This microsymbiont predominated in the perennial grasslands of Glencoe Research Station, in Uruguay, to competitively nodulate its host, and fix atmospheric nitrogen. Here we describe the basic features of WSM2304, together with the complete genome sequence, and annotation. This is the first completed genome sequence for a nitrogen fixing microsymbiont of a clover species from the American center of origin. We reveal that its genome size is 6,872,702 bp encoding 6,643 protein-coding genes and 62 RNA only encoding genes. This multipartite genome was found to contain 5 distinct replicons; a chromosome of size 4,537,948 bp and four circular plasmids of size 1,266,105 bp, 501,946 bp, 308,747 bp and 257,956 bp.

---

## Introduction

Since ancient times, crop fields have been regularly rotated with legumes, and this continues in the modern world because of the recognition that the productivity of agricultural systems is nitrogen dependent [1]. Legumes may redress nitrogen deficiency through the fixation of atmospheric nitrogen by rhizobia in root nodules [2]. Today, despite the ready availability of nitrogen-fertilizer manufactured through the Haber-Bosch process, globally in excess of 400 million ha of agricultural land are sustained by nitrogen derived from forage legumes [3]. These forages are grown for animal

feed, for rotation with cereal crops, as disease breaks or as cover crops for tree plantations. Amongst the forage legumes, the *Trifolium* spp. (clovers) are acknowledged as one of the most important genera, with 237 species distributed across the temperate and sub-tropical regions of North and South America, Europe, Africa and Australasia [4].

These clovers are nodulated by *R. leguminosarum* bv *trifolii*, which is one of the most exploited species of root-nodule bacteria in world agriculture. However, because clovers are geographically

widely distributed, and phenologically variable (they may be either annual [e.g. *T. subterraneum*] or perennial [e.g. *T. pratense*, *T. raepens* and *T. polymorphum*]), it is rare that a single strain of *R. leguminosarum* bv *trifolii* can effectively fix nitrogen across a wide diversity of clovers, especially those from different geographical and phenological backgrounds [5].

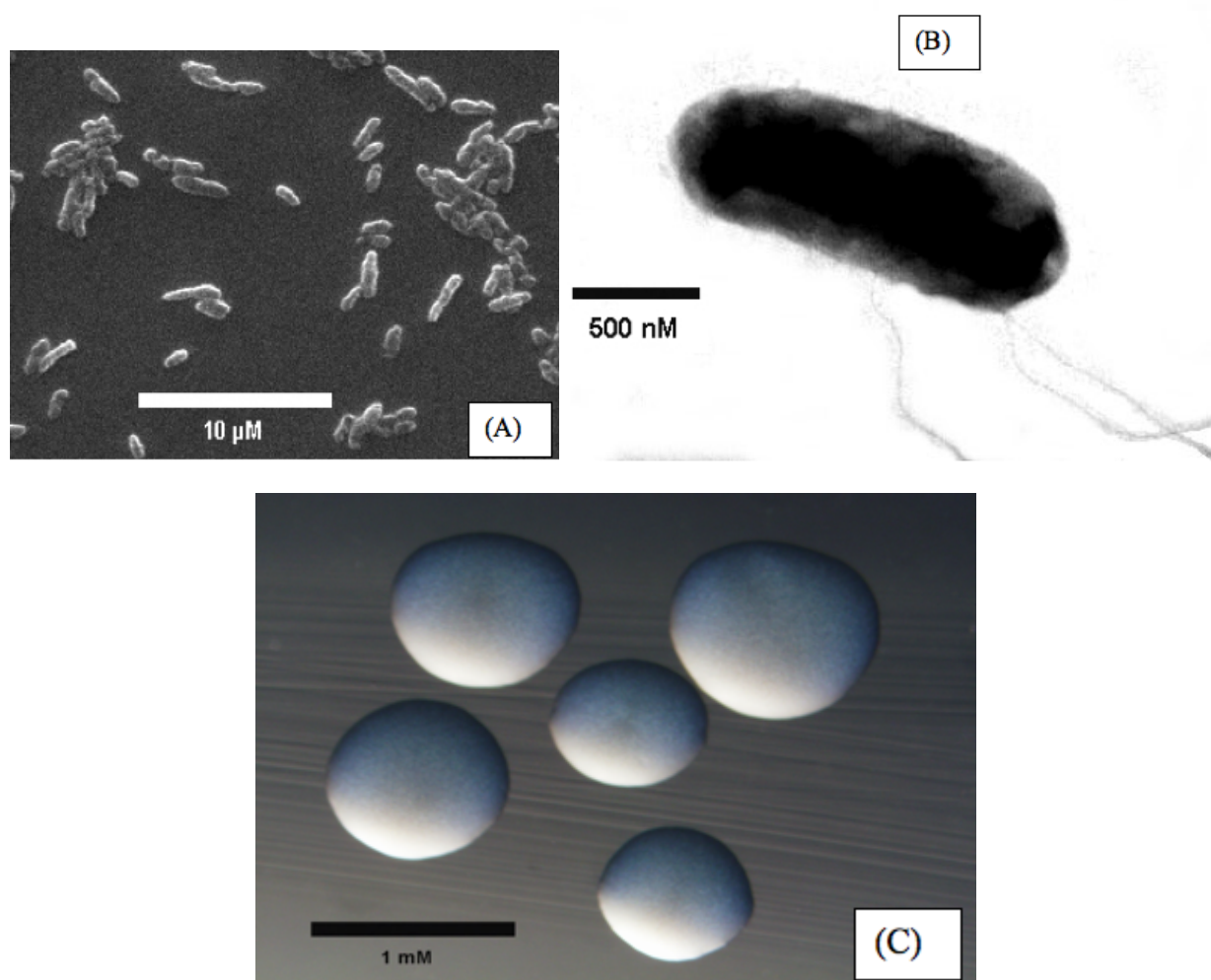
*Rhizobium leguminosarum* bv *trifolii* strain WSM2304 was isolated from a nodule recovered from the roots of the perennial clover *Trifolium polymorphum* growing at Glencoe Research Station near Tacuarembó, Uruguay in December 1998. WSM2304 is of particular interest because it is a highly effective microsymbiont of a perennial clover of South American origin, has a narrow, specialized host range for nitrogen fixation [5], and is highly competitive for nodulation of *T. polymorphum* in the acid, infertile soils of Uruguay

[6]. WSM2304 has also been implicated in host mediated selection for an effective microsymbiont under competitive conditions for nodulation [7].

Here we present a summary classification and a set of features for *R. leguminosarum* bv *trifolii* strain WSM2304 (Table 1), together with the description of the complete genome sequence and annotation.

## Classification and features

*R. leguminosarum* bv *trifolii* strain WSM2304 is a motile, Gram-negative, non-spore-forming rod (Figure 1 A & B) in the *Rhizobiaceae* family of the class *Alphaproteobacteria* that forms mildly mucoid colonies (Figure 1C) on solid media [24]. It has a mean generation time of 3.5 h in rich medium at the optimal growth temperature of 28°C [7].

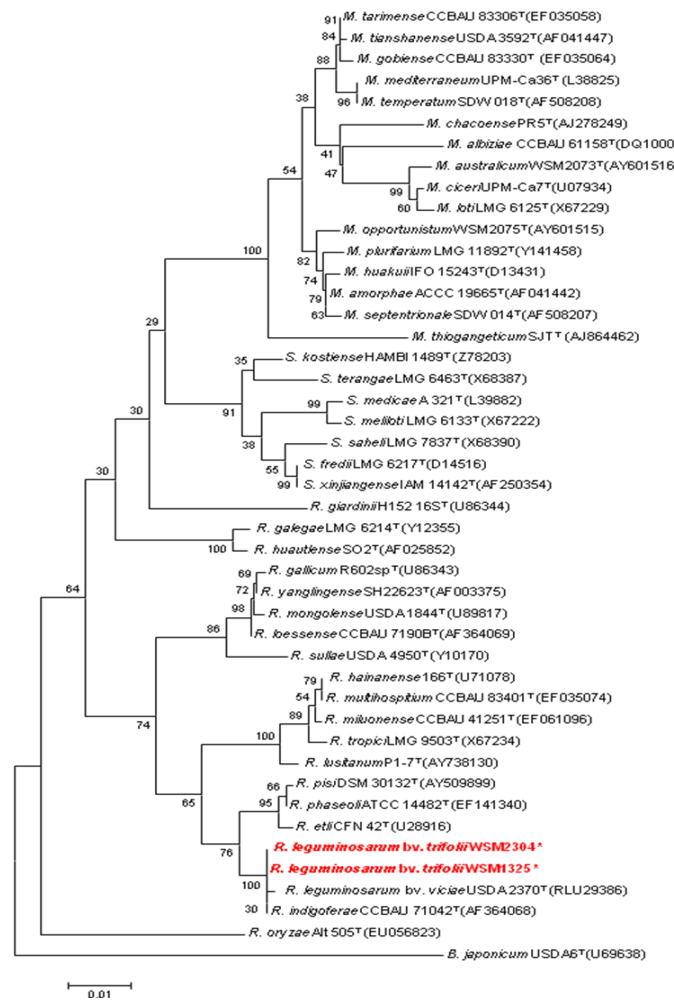


**Figure 1.** Images of *R. leguminosarum* bv *trifolii* strain WSM2304 using scanning (A) and transmission electron microscopy (B). The appearance of colony morphology on solid media (C).

Figure 2 shows the phylogenetic neighborhood of *R. leguminosarum* bv *trifolii* strain WSM2304 in a 16S rRNA-based tree. An intragenic fragment of 1,440 bp was chosen since the 16S rRNA gene has not been completely sequenced in many type strains. A comparison of the entire 16S rRNA gene of WSM2304 to completely sequenced 16S rRNA genes of other rhizobia revealed 100% gene sequence identity with *R. leguminosarum* bv *trifolii* strain WSM1325 but a 1 bp difference from the 16S rRNA gene of *R. leguminosarum* bv *viciae* strain 3841.

### Symbiotaxonomy

*R. leguminosarum* bv *trifolii* WSM2304 nodulates (Nod<sup>+</sup>) and fixes nitrogen effectively (Fix<sup>+</sup>) with the South American perennial clover *T. polymorphum* [5]. WSM2304 is Nod<sup>+</sup>, Fix<sup>-</sup> with Mediterranean annual clovers *T. subterraneum* and *T. glanduliferum*, in contrast to *R. leguminosarum* bv *trifolii* WSM1325 [5,29]. When inoculated onto perennial clovers of either North American or Mediterranean origin WSM2304 is variably Nod<sup>+</sup>, but always Fix<sup>-</sup> [5,6,30]. Under conditions of competitive nodulation, WSM2304 may preferentially nodulate *T. polymorphum* even when outnumbered 100:1 by WSM1325 [7].



**Figure 2.** Phylogenetic tree showing the relationships of *R. leguminosarum* bv *trifolii* strain WSM2304 with the type strains of *Rhizobiaceae* based on aligned sequences of the 16S rRNA gene (1,440 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 3.1 [25]. Kimura two-parameter distances were derived from the aligned sequences [26] and a bootstrap analysis [27] as performed with 500 replicates in order to construct a consensus unrooted tree using the neighbor-joining method [28] for each gene alignment separately. The genera in this tree include *Bradyrhizobium* (B.), *Mesorhizobium* (M), *Rhizobium* (R); *Ensifer* (*Sinorhizobium*) (S). Type strains are indicated with a superscript T. Strains with a genome sequencing project registered in GOLD [22] are in bold red print. Published genomes are designated with an asterisk.

**Table 1.** Classification and general features of *R. leguminosarum* bv *trifolii* WSM2304 according to the MIGS recommendations [8].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [5-7,9]
		Phylum <i>Proteobacteria</i>	TAS [5-7,10]
		Class <i>Alphaproteobacteria</i>	TAS [5-7,11,12]
		Order <i>Rhizobiales</i>	TAS [5-7,11,13]
		Family <i>Rhizobiaceae</i>	TAS [5-7,14]
		Genus <i>Rhizobium</i>	TAS [5-7,14-18]
	Classification	Species <i>Rhizobium leguminosarum</i> bv <i>trifolii</i> Strain WSM2304	TAS [5-7,14,16,18,19]
	Gram stain	negative	TAS [20]
	Cell shape	rod	TAS [20]
	Motility	motile	TAS [20]
	Sporulation	non-sporulating	TAS [20]
	Temperature range	mesophile	TAS [20]
	Optimum temperature	28°C	TAS [20]
	Salinity	unknown	TAS [20]
MIGS-22	Oxygen requirement	aerobic	TAS [20]
	Carbon source	glucose, mannitol	TAS [5-7]
	Energy source	chemoheterotroph	TAS [20]
MIGS-6	Habitat	Soil, root nodule, host	TAS [5-7]
MIGS-15	Biotic relationship	Free living, Symbiotic	TAS [5-7]
MIGS-14	Pathogenicity	none	TAS [20]
	Biosafety level	1	TAS [21]
	Isolation	<i>Trifolium polymorphum</i> root nodule	TAS [22]
MIGS-4	Geographic location	Glencoe Research Station, INIA, Uruguay	TAS [22]
MIGS-5	Sample collection time	December 1 <sup>st</sup> , 1998	TAS [22]
MIGS-4.1	Latitude	-56	
MIGS-4.2	Longitude	-31.41	TAS [22]
MIGS-4.3	Depth	5cm soil depth	NAS [2]
MIGS-4.4	Altitude	130m	TAS [22]

Evidence codes - IDA: Inferred from Direct Assay (first time in publication); 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 [23]. If the evidence code is IDA, then the property was directly observed for a living isolate by one of the authors or an expert mentioned in the acknowledgements.

## Genome sequencing and annotation

### 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 Community Sequencing Program at the Department of Energy Joint Genome Institute (JGI) for projects of relevance to DOE missions. The genome project is deposited in

the Genomes OnLine Database [22] and the complete genome sequence in GenBank. Sequencing, finishing and annotation were performed by the DOE Joint Genome Institute (JGI). A summary of the project information is shown in Table 2 and sequence data statistics from the trace archive for this project are presented in Table 3.

**Table 2.** Genome sequencing project information for *R. leguminosarum* bv *trifolii* WSM2304.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Four genomic libraries: three Sanger libraries; 1-2 kb pTH1522, 6-8 kb pMCL200, fosmid pcc1Fos and one 454 pyrosequencing standard library
MIGS-29	Sequencing platforms	ABI3730xl, 454 GS FLX
MIGS-31.2	Sequencing coverage	21.3x Sanger; 10.1x pyrosequencing
MIGS-30	Assemblers	Newbler version 1.1.02.15, Phrap
MIGS-32	Gene calling method	Prodigal
	Genbank ID	CP001191 (Chromosome) <sup>a</sup> CP001192 (pRLG201) <sup>b</sup> CP001193 (pRLG202) <sup>c</sup> CP001194 (pRLG204) <sup>d</sup> CP001195 (pRLG205) <sup>e</sup>
	Genbank Date of Release	October 16, 2008
	GOLD ID	Gc00870 <sup>f</sup>
	NCBI project ID	20179
	Database: IMG	643348569 <sup>g</sup>
	Project relevance	Symbiotic nitrogen fixation, agriculture

<sup>a</sup> <http://www.ncbi.nlm.nih.gov/nucore/209533368>

<sup>b</sup> <http://www.ncbi.nlm.nih.gov/nucore/209537694>

<sup>c</sup> <http://www.ncbi.nlm.nih.gov/nucore/209538856>

<sup>d</sup> <http://www.ncbi.nlm.nih.gov/nucore/209539307>

<sup>e</sup> <http://www.ncbi.nlm.nih.gov/nucore/209539531>

<sup>f</sup> [http://genomesonline.org/GOLD\\_CARDS/Gc00870.html](http://genomesonline.org/GOLD_CARDS/Gc00870.html)

<sup>g</sup> [http://img.jgi.doe.gov/taxon\\_oid=64173348569](http://img.jgi.doe.gov/taxon_oid=64173348569)

## Growth conditions and DNA isolation

*R. leguminosarum* bv *trifolii* WSM2304 was grown to mid logarithmic phase in TY medium (a rich medium) [31] on a gyratory shaker at 28°C. DNA was isolated from 60 ml of cells using a CTAB (Cetyl trimethylammonium bromide) bacterial genomic DNA isolation method (<http://my.jgi.doe.gov>).

## Genome sequencing and assembly

The genome was sequenced using a combination of Sanger and 454 sequencing platforms. All general aspects of library construction and sequencing performed at the JGI can be found at the JGI website (<http://www.jgi.doe.gov/>). 454 Pyrosequencing reads were assembled using the Newbler assembler version 1.1.02.15 (Roche). Large Newbler contigs were broken into 5,676 fragments of 1,500 bp with 100 bp overlap and entered into the assembly as pseudo-reads. The se-

quences were assigned quality scores based on Newbler consensus q-scores with modifications to account for overlap redundancy and to adjust inflated q-scores. A hybrid 454/Sanger assembly was made using the phrap assembler. Possible mis-assemblies were corrected and gaps between contigs were closed by custom primer walks from sub-clones or PCR products. A total of 1,826 Sanger finishing reads were produced. Illumina reads were used to improve the final consensus quality using an in-house developed tool (the Polisher). The final assembly consists of 168,617 Sanger reads in addition to 5,663 454 pseudo reads. The error rate of the completed genome sequence is less than 1 in 100,000. Together all sequence types provided about 31.4× coverage of the genome.

**Table 3.** Production sequence for the finished genome of *R. leguminosarum* bv *trifolii* WSM2304, JGI project 4024175

Vector/Type	Library id	Insert size (kb)	Reads	Mb	q20 (Mb)
pMCL200	FHOO	7.0 ± 0.9	74,398	66.6	51.6
pcc1Fos	FHTU	36 ± 3.4	15,776	11.8	7.8
pTH1522	FNNZ	1.8 ± 0.3	79,386	68.6	53.5
454-std	FHTW	NA	719,338	69.9	NA

### Genome annotation

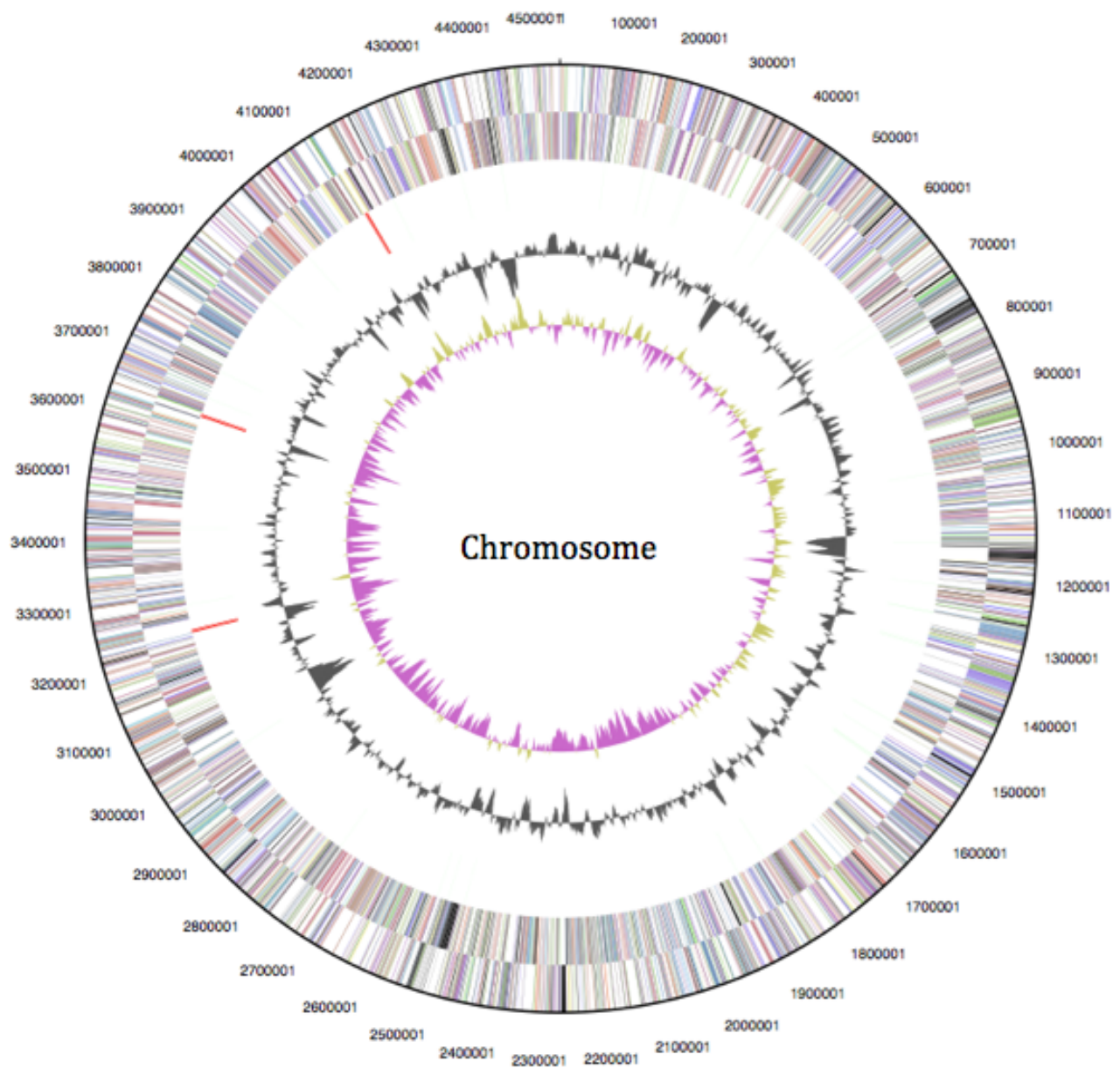
Genes were identified using Prodigal [32] as part of the Oak Ridge National Laboratory genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [33]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. Additional gene prediction analyses and functional annotation were performed within the Integrated Microbial Genomes platform (<http://img.jgi.doe.gov/er>) [34].

### Genome properties

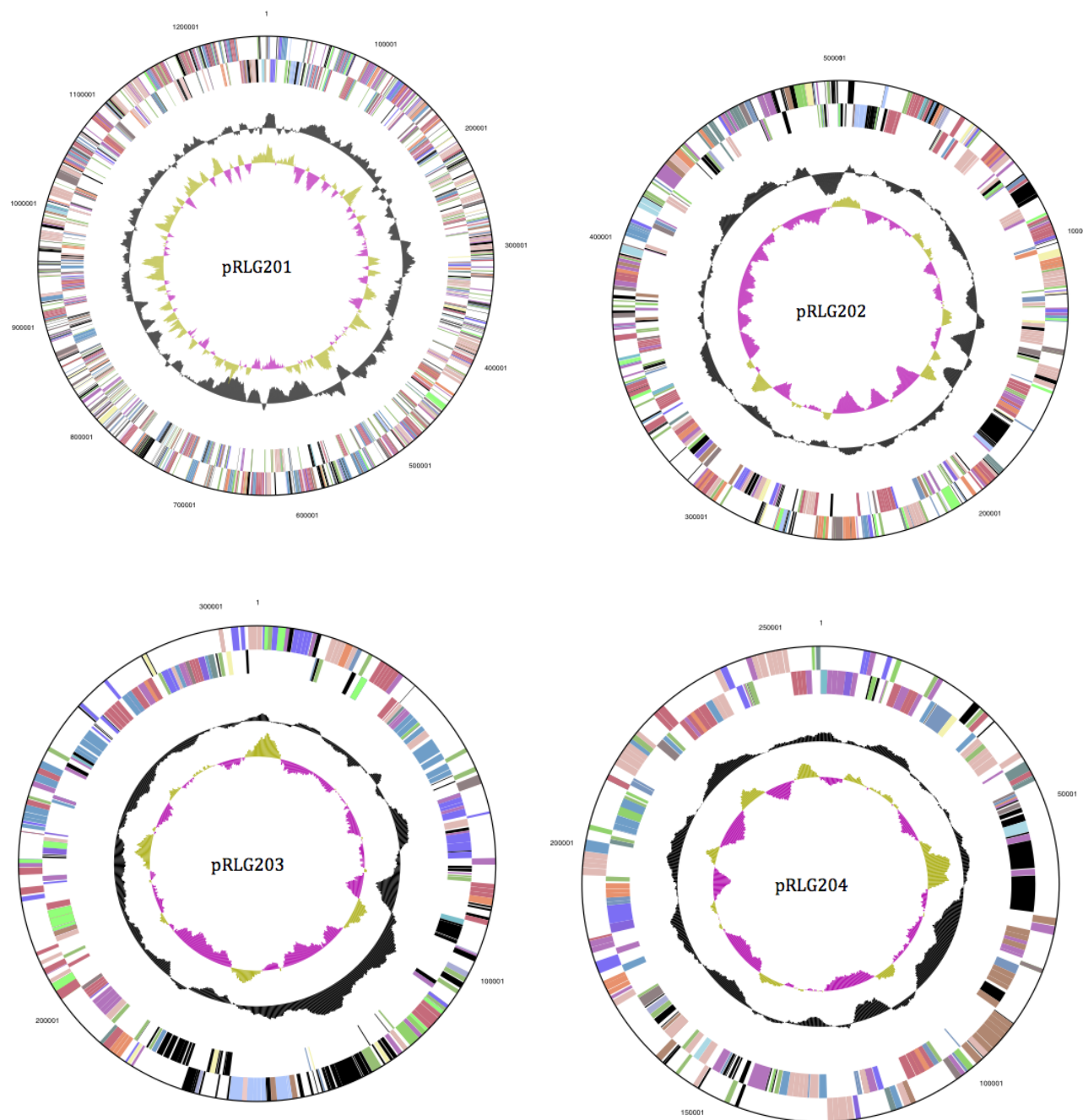
The genome is 6,872,702 bp long with a 61.18% GC content, (Table 4) and comprised of 5 replicons; 1 circular chromosome of size 4,537,948 bp (Figure 3) and 4 circular plasmids of size 4,537,948, 1,266,105, 501,946, 308,747 and 257,956 bp (Figure 4). Of the 6,643 genes predicted, 6,581 were protein coding genes, and 62 RNA only encoding genes. In addition, 166 pseudogenes were identified. The majority of the genes (72.44%) were assigned a putative function whilst the remaining ones were annotated as hypothetical proteins. The distribution of genes into COGs functional categories is presented in Table 5.

**Table 4.** Genome Statistics for *R. leguminosarum* bv *trifolii* WSM2304

Attribute	Value	% of Total
Genome size (bp)	6,872,702	100.00%
DNA coding region (bp)	6,053,973	88.09%
DNA G+C content (bp)	4,204,577	61.18%
Number of replicons	5	100.00%
Extrachromosomal elements	4	80.00%
Total genes	6,643	100.00%
RNA coding genes	62	0.93%
rRNA operons	3	
Protein-coding genes	6,581	99.07%
Pseudo genes	166	2.49%
Genes with function prediction	4,812	72.44%
Genes in paralog clusters	4,104	61.78%
Genes assigned to COGs	5,105	76.85%
Genes assigned Pfam domains	5,149	77.51%
Genes with signal peptides	2,247	33.83%
Genes with transmembrane helices	1,495	22.50%
CRISPR repeats	0	



**Figure 3.** Graphical circular map of the chromosome of *R. leguminosarum* bv *trifolii* WSM2304. From outside to the center: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew. Chromosome is not drawn to scale relative to the plasmids in Figure 4.



**Figure 4.** Graphical circular map of the plasmids of *R. leguminosarum* by *trifolii* WSM2304. From outside to the center: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew. Plasmids pRLG201, pRLG202, pRLG203 and pRLG204 are not drawn to scale relative to each other or to the chromosome in Figure 3.



**Table 5.** Number of genes associated with the general COG functional categories

Code	value	%age	Description
J	194	2.95	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	558	8.48	Transcription
L	164	2.49	Replication, recombination and repair
B	2	0.03	Chromatin structure and dynamics
D	38	0.58	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	66	1.00	Defense mechanisms
T	317	4.82	Signal transduction mechanisms
M	309	4.70	Cell wall/membrane biogenesis
N	91	1.38	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	88	1.34	Intracellular trafficking and secretion
O	165	2.51	Posttranslational modification, protein turnover, chaperones
C	314	4.77	Energy production and conversion
G	599	9.10	Carbohydrate transport and metabolism
E	687	10.44	Amino acid transport and metabolism
F	109	1.66	Nucleotide transport and metabolism
H	180	2.74	Coenzyme transport and metabolism
I	238	3.62	Lipid transport and metabolism
P	278	4.22	Inorganic ion transport and metabolism
Q	156	2.37	Secondary metabolites biosynthesis, transport and catabolism
R	710	10.79	General function prediction only
S	532	8.08	Function unknown
-	1,476	22.43	Not in COGs

## 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-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396. We thank

Gordon Thompson (Murdoch University) for the preparation of SEM and TEM photos. We gratefully acknowledge the funding received from Murdoch University Strategic Research Fund through the Crop and Plant Research Institute (CaPRI), and the Grains Research and Development Corporation (GRDC), to support the National *Rhizobium* Program (NRP) and the Centre for *Rhizobium* Studies (CRS) at Murdoch University.

## References

1. Hamblin J. Preface, 1998, pp xi-xiii. In: Lupins as Crop Plants. *Biology, Production and Utilization*. Gladstones JS, Atkins CA, Hamblin J (Eds.). CAB International, Madison, NY.
2. Sprent JI. *Legume nodulation: a global perspective*. 2009. Oxford, Wiley-Blackwell.
3. Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. *Marschner Review. Plant Soil* 2008; **311**:1-18. [doi:10.1007/s11104-008-9668-3](https://doi.org/10.1007/s11104-008-9668-3)
4. Zohary M, Heller D. *The Genus Trifolium*. The Israel Academy of Sciences and Humanities, Ahva Printing Press 1984, Jerusalem.

5. Howieson JG, Yates RJ, O'Hara GW, Ryder M, Real D. The interactions of *Rhizobium leguminosarum* biovar *trifolii* in nodulation of annual and perennial *Trifolium* spp from diverse centres of origin. *Aust J Exp Agric* 2005; **45**:199-207. [doi:10.1071/EA03167](https://doi.org/10.1071/EA03167)
6. Yates RJ, Howieson JG, Real D, Reeve WG, Vivas-Marfisi A, O'Hara GW. Evidence of selection for effective nodulation in the *Trifolium* spp. symbiosis with *Rhizobium leguminosarum* biovar *trifolii*. *Aust J Exp Agric* 2005; **45**:189-198. [doi:10.1071/EA03168](https://doi.org/10.1071/EA03168)
7. Yates RJ, Howieson JG, Reeve WG, Brau L, Speijers J, Nandasena K, Real D, Sezmis E, O'Hara GW. Host-strain mediated selection for an effective nitrogen-fixing symbiosis between *Trifolium* spp. and *Rhizobium leguminosarum* biovar *trifolii*. *Soil Biol Biochem* 2008; **40**:822-833. [doi:10.1016/j.soilbio.2007.11.001](https://doi.org/10.1016/j.soilbio.2007.11.001)
8. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen MJ, Angiuoli SV. Towards a richer description of our complete collection of genomes and metagenomes: the "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol* 2008; **26**:541-547. [PubMed doi:10.1038/nbt1360](https://pubmed.ncbi.nlm.nih.gov/1631360/)
9. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, Bacteria, and Eucarya. *Proc Natl Acad Sci USA* 1990; **87**: 4576-4579. [PubMed doi:10.1073/pnas.87.12.4576](https://pubmed.ncbi.nlm.nih.gov/1073/pnas.87.12.4576)
10. Garrity GM, Holt JG. The Road Map to the Manual. In: Garrity GM, Boone DR, Castenholz RW (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 1, Springer, New York, 2001, p. 119-169.
11. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 1.
12. List editor. Validation List No. 107. List of new names and new combinations previously effectively, but not validly, published. *Int J Syst Evol Microbiol* 2006; **56**: 1-6. [PubMed doi:10.1099/ijs.0.64188-0](https://pubmed.ncbi.nlm.nih.gov/101099/ijs.0.64188-0)
13. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 2, Part C, Springer, New York, 2005, p. 324.
14. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**: 225-420.
15. Frank B. Über die Pilzsymbiose der Leguminosen. *Ber Dtsch Bot Ges* 1889; **7**: 332-346.
16. Jordan DC, Allen ON. Genus I. *Rhizobium* Frank 1889, 338; Nom. gen. cons. Opin. 34, Jud. Comm. 1970, 11. In: Buchanan RE, Gibbons NE (eds), *Bergey's Manual of Determinative Bacteriology*, Eighth Edition, The Williams and Wilkins Co., Baltimore, 1974, p. 262-264.
17. Young JM, Kuykendall LD, Martínez-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 *Allorhizobium undicola* de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi*, *R. undicola* and *R. vitis*.. *Int J Syst Evol Microbiol* 2001; **51**: 89-103. [PubMed](https://pubmed.ncbi.nlm.nih.gov/11111111/)
18. Editorial Secretary (for the Judicial Commission of the International Committee on Nomenclature of Bacteria). OPINION 34: Conservation of the Generic Name *Rhizobium* Frank 1889. *Int J Syst Bacteriol* 1970; **20**: 11-12; [doi:10.1099/00207713-20-1-11](https://doi.org/10.1099/00207713-20-1-11).
19. Ramírez-Bahena MH, García-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, Mateos PF, Martínez-Molina E, Velázquez E. Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889<sup>AL</sup>, *Rhizobium phaseoli* Dangeard 1926AL and *Rhizobium trifolii* Dangeard 1926AL. *R. trifolii* is a later synonym of *R. leguminosarum*. Reclassification of the strain *R. leguminosarum* DSM 30132 (=NCIMB 11478) as *Rhizobium pisi* sp. nov.. *Int J Syst Evol Microbiol* 2008; **58**: 2484-2490. [PubMed doi:10.1099/ijs.0.65621-0](https://pubmed.ncbi.nlm.nih.gov/101099/ijs.0.65621-0)
20. Kuykendall LD, Hashem F, Wang ET. Genus VII. *Rhizobium*, 2005, pp 325-340. In: *Bergey's Manual of Systematic Bacteriology*. Second Edition. Volume 2 The Proteobacteria. Part C The *Alpha*-, *Delta*-, and *Epsilonproteobacteria*. Brenner DJ, Krieg NR, Staley JT (Eds.), Garrity GM (Editor in Chief) Springer Science and Business Media Inc, New York, USA.
21. Biological Agents. Technical rules for biological agents [www.baua.de TRBA 466](http://www.baua.de/TRBA466).
22. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes OnLine Database (GOLD) in 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic*

- Acids Res* 2008; **36**:D475-D479.  
[PubMed doi:10.1093/nar/gkm884](#)
23. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, *et al.* The Gene Ontology Consortium. Gene ontology: tool for the unification of biology. *Nat Genet* 2000; **25**:25-29.  
[PubMed doi:10.1038/75556](#)
24. Howieson JG, Ewing MA, D'Antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil* 1988; **105**:179-188. [doi:10.1007/BF02376781](#)
25. Kumar S, Tamura K, Nei M. MEGA3: integrated software for molecular evolutionary genetics analysis and sequence alignment. *Brief Bioinform* 2004; **5**:150-163. [PubMed](#)  
[PubMed doi:10.1093/bib/5.2.150](#)
26. Kimura M. A simple model for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol* 1980; **16**:111-120.  
[PubMed doi:10.1007/BF01731581](#)
27. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. [doi:10.2307/2408678](#)
28. Saitou N, Nei M. Reconstructing phylogenetic trees. *Mol Biol Evol* 1987; **4**:406-425. [PubMed](#)
29. Bullard GK, Roughley RJ, Pulsford DJ. The legume inoculant industry and inoculant quality control in Australia: 1953–2003. *Aust J Exp Agric* 2005; **45**:127-140. [doi:10.1071/EA03159](#)
30. Centre for *Rhizobium* Studies. Annual Report. JG Howieson (Ed). 2001. Murdoch University Print, Perth, Australia.
31. Reeve WG, Tiwari RP, Worsely PS, Dilworth MJ, Glenn AR, Howieson JG. Constructs for insertion mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* 1999; **145**:1307-1316. [PubMed doi:10.1099/13500872-145-6-1307](#)
32. Anonymous. Prodigal Prokaryotic Dynamic Programming Genefinding Algorithm. Oak Ridge National Laboratory and University of Tennessee 2009. <http://compbio.ornl.gov/prodigal>
33. Pati A, Ivanova N, Mikhailova N, Ovchinikova G, Hooper SD, Lykidis A, Kyrpides NC. GenePRIMP: A Gene Prediction Improvement Pipeline for microbial genomes. (Submitted) 2009
34. Markowitz VM, Szeto E, Palaniappan K, Grechkin Y, Chu K, Chen IMA, Dubchak I, Anderson I, Lykidis A, Mavromatis K, *et al.* The Integrated Microbial Genomes (IMG) system in 2007: data content and analysis tool extensions. *Nucleic Acids Res* 2008; **36**:D528-D533.  
[PubMed doi:10.1093/nar/gkm846](#)