

# Genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689, the microsymbiont of the one flowered clover *Trifolium uniflorum*

Jason Terpolilli<sup>1</sup>, Tian Rui<sup>1</sup>, Ron Yates<sup>1,2</sup>, John Howieson<sup>1</sup>, Philip Poole<sup>3,4</sup>, Christine Munk<sup>5</sup>, Roxanne Tapia<sup>5</sup>, Cliff Han<sup>5</sup>, Victor Markowitz<sup>6</sup>, Reddy Tatiparthi<sup>7</sup>, Konstantinos Mavrommatis<sup>6</sup>, Natalia Ivanova<sup>7</sup>, Amrita Pati<sup>7</sup>, Lynne Goodwin<sup>5</sup>, Tanja Woyke<sup>7</sup>, Nikos Kyrpides<sup>7</sup> & Wayne Reeve<sup>\*1</sup>.

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

<sup>2</sup> Department of Agriculture and Food, Western Australia, Australia

<sup>3</sup> Department of Plant Sciences, University of Oxford, UK

<sup>4</sup> Sir Walter Murdoch Adjunct Professor, Murdoch University, Perth, Australia

<sup>5</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

<sup>6</sup> Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

<sup>7</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

\*Correspondence: Wayne Reeve (W.Reeve@murdoch.edu.au)

**Keywords:** root-nodule bacteria, nitrogen fixation, lupin-nodulating, rhizobia, *Alphaproteobacteria*

---

*Rhizobium leguminosarum* bv. *trifolii* is a soil-inhabiting bacterium that has the capacity to be an effective N<sub>2</sub>-fixing microsymbiont of *Trifolium* (clover) species. *R. leguminosarum* bv. *trifolii* strain WSM1689 is an aerobic, motile, Gram-negative, non-spore-forming rod that was isolated from a root nodule of *Trifolium uniflorum* collected on the edge of a valley 6 km from Eggares on the Greek Island of Naxos. Although WSM1689 is capable of highly effective N<sub>2</sub>-fixation with *T. uniflorum*, it is either unable to nodulate or unable to fix N<sub>2</sub> with a wide range of both perennial and annual clovers originating from Europe, North America and Africa. WSM1689 therefore possesses a very narrow host range for effective N<sub>2</sub> fixation and can thus play a valuable role in determining the geographic and phenological barriers to symbiotic performance in the genus *Trifolium*. Here we describe the features of *R. leguminosarum* bv. *trifolii* strain WSM1689, together with the complete genome sequence and its annotation. The 6,903,379 bp genome contains 6,709 protein-coding genes and 89 RNA-only encoding genes. This multipartite genome contains six distinct replicons; a chromosome of size 4,854,518 bp and five plasmids of size 667,306, 518,052, 341,391, 262,704 and 259,408 bp. This rhizobial genome is one of 20 sequenced as part of a DOE Joint Genome Institute 2010 Community Sequencing Program.

---

## Introduction

The nitrogen (N) cycle is one of the most important biogeochemical processes underpinning the existence of life on Earth. A key step in this cycle is to convert relatively inert atmospheric dinitrogen (N<sub>2</sub>) into a bioaccessible form such as ammonia (NH<sub>3</sub>) through a process referred to as biological nitrogen fixation (BNF). BNF is performed only by a specialized subset of *Bacteria* and *Archaea* that possess the necessary cellular machinery to enzymatically reduce N<sub>2</sub> into NH<sub>3</sub>.

Some of these bacteria (termed rhizobia or root nodule bacteria) have evolved non-obligatory symbiotic relationships with legumes whereby the bacteria receive a carbon source from the plant and in return supply fixed N to the host [1]. Harnessing this association can boost soil N-inputs and therefore production yields of legumes, or non-legumes grown in subsequent years, without the need for supplementation with industrially synthesized N-based fertilizers [2].

Some of the most widely cultivated pasture legumes are members of the legume genus *Trifolium* (clover). The natural distribution of these species spans three centers of diversity, with an estimated 28% of species in the Americas, 57% in Eurasia and 15% in sub-Saharan Africa [3]. Approximately 30 species of clover, predominately of Eurasian origin, are widely grown as annual and perennial species in pasture systems in Mediterranean and temperate climatic zones [3]. Globally-important perennial species of clover include *T. repens* (white clover), *T. pratense* (red clover), *T. fragiferum* (strawberry clover) and *T. hybridum* (alsike clover). While clovers are known to form N<sub>2</sub>-fixing symbiotic associations with *Rhizobium leguminosarum* bv. *trifolii*, there exists wide variation in symbiotic compatibility across different strains and hosts from ineffective (non-N<sub>2</sub>-fixing) nodulation to fully effective N<sub>2</sub>-fixing partnerships.

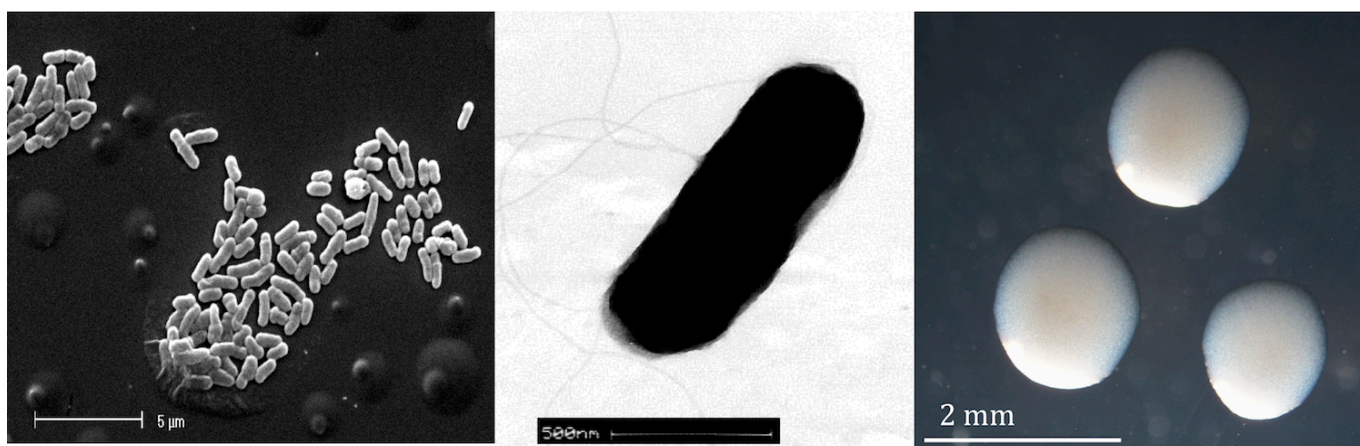
*Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 was isolated in 1995 from a nodule of the perennial clover *Trifolium uniflorum* collected on the edge of a valley 6 km from Eggares on the Greek Island of Naxos. *T. uniflorum* is one of small number of perennial *Trifolium* spp. found in the dry, Mediterranean basin. While WSM1689 has been shown to be either ineffective or unable to nodulate a range of annual and perennial *Trifolium* sp., it is a highly effective N<sub>2</sub>-fixing microsymbiont of *T. uniflorum* [4]. Therefore, *R. leguminosarum* bv. *trifolii* WSM1689 has a very narrow host range and thus represents a good isolate to study the genetic basis of symbiotic

specificity. The availability of this sequence data also complements the already published genomes of the clover-nodulating *R. leguminosarum* bv. *trifolii* WSM1325 [5] and WSM2304 [6]. Here we present a summary classification and a set of general features for *R. leguminosarum* bv. *trifolii* strain WSM1689 together with the description of the complete genome sequence and its annotation.

## Classification and features

*R. leguminosarum* bv. *trifolii* strain WSM1689 is a motile, non-sporulating, non-encapsulated, Gram-negative rod in the order *Rhizobiales* of the class *Alphaproteobacteria*. The rod-shaped form varies in size with dimensions of approximately 0.25-0.5 µm in width and 2.0 µm in length (Figure 1 Left and 1 Center). It is fast growing, forming colonies within 3-4 days when grown on half strength Lupin Agar (½LA) [7], tryptone-yeast extract agar (TY) [8] or a modified yeast-mannitol agar (YMA) [9] at 28°C. Colonies on ½LA are opaque, slightly domed and moderately mucoid with smooth margins (Figure 1 Right). Minimum Information about the Genome Sequence (MIGS) is provided in Table 1.

Figure 2 shows the phylogenetic neighborhood of *R. leguminosarum* bv. *trifolii* strain WSM1689 in a 16S rRNA gene sequence based tree. This strain shares 100% (1362/1362 bp) sequence identity to the 16S rRNA gene of *R. leguminosarum* bv. *trifolii* strain WSM1325 [5] and *R. leguminosarum* bv. *trifolii* strain WSM2304 [6].

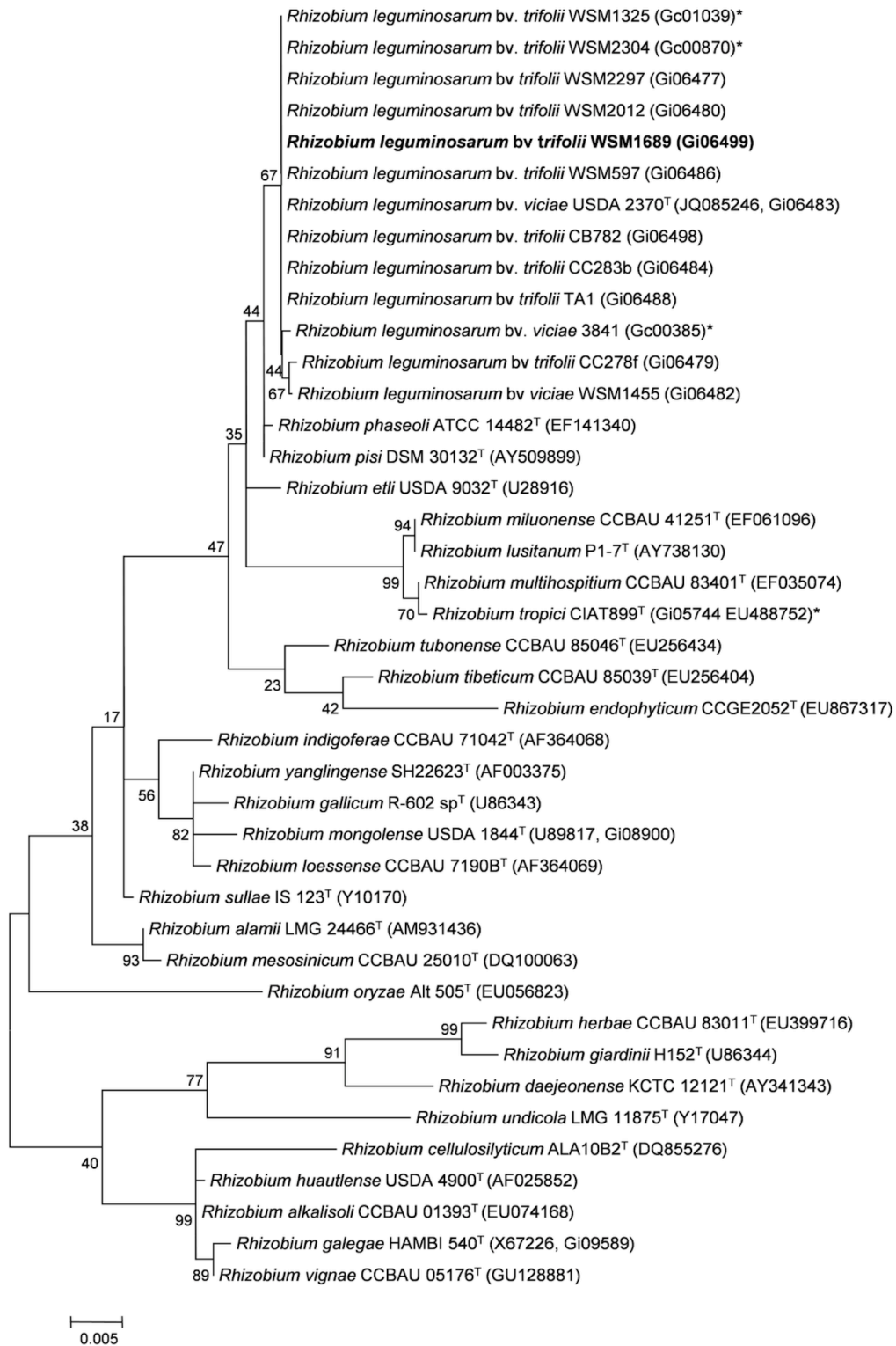


**Figure 1.** Images of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 using scanning (Left) and transmission (Center) electron microscopy and the appearance of colony morphology on ½LA (Right).

**Table 1.** Classification and general features of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 according to the MIGS recommendations [10,11].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [11]
		Phylum <i>Proteobacteria</i>	TAS [12]
		Class <i>Alphaproteobacteria</i>	TAS [13,14]
	Current classification	Order <i>Rhizobiales</i>	TAS [14,15]
		Family <i>Rhizobiaceae</i>	TAS [16,17]
		Genus <i>Rhizobium</i>	TAS [16,18-21]
		Species <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	TAS [16,18,21,22]
		Strain WSM1689	TAS [4]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	NAS
	Salinity	Not reported	NAS
MIGS-22	Oxygen requirement	Aerobic	TAS [4]
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-6	Habitat	Soil, root nodule, host	TAS [4]
MIGS-15	Biotic relationship	Free living, symbiotic	TAS [4]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	NAS [23]
	Isolation	Root nodule	TAS [4]
MIGS-4	Geographic location	Naxos, Greece	IDA
MIGS-5	Nodule collection date	1995	IDA
MIGS-4.1	Latitude	37.128333	IDA
MIGS-4.2	Longitude	25.443333	IDA
MIGS-4.3	Depth	Not reported	
MIGS-4.4	Altitude	Not reported	

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 [24].



**Figure 2.** Phylogenetic tree showing the relationship of *Rhizobium leguminosarum* bv *trifolii* WSM1689 (shown in bold print) to other root nodulating *Rhizobium* spp. in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1,180 bp internal region). All positions containing gaps and missing data were eliminated. All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5 [25]. The tree was built using the Maximum-Likelihood method with the General Time Reversible model [26]. Bootstrap analysis [27] with 500 replicates was performed to assess the support of the clusters. Type strains are indicated with a superscript T. Brackets after the strain name contain a DNA database accession number and/or a GOLD ID (beginning with the prefix G) for a sequencing project registered in GOLD [28]. Published genomes are indicated with an asterisk.

**Table 2.** Compatibility of WSM1689 with both perennial and annual *Trifolium* genotypes for nodulation (Nod) and N<sub>2</sub>-Fixation (Fix). Data compiled from [4].

Species Name	Cultivar	Origin	Growth habit	Nod	Fix	Comment
<i>T. uniflorum</i>	Nil	Europe	Perennial	Nod	Fix	Highly effective
<i>T. tumens</i>	1986267	Europe	Perennial	Nod	Fix	Ineffective
<i>T. tumens</i>	16758246	Europe	Perennial	Nod	Fix	Ineffective
<i>T. medium</i>	21881154	Europe	Perennial	Nod	Fix	Ineffective
<i>T. repens</i>	037701	Europe	Perennial	Nod	Fix	Ineffective
<i>T. repens</i>	036120	Europe	Perennial	Nod	Fix	Ineffective
<i>T. pratense</i>	Russian no 9	Europe	Perennial	Nod	Fix	Ineffective
<i>T. pratense</i>	Redquin	Europe	Perennial	Nod	Fix	Ineffective
<i>T. ambiguum</i>	Endura	Europe	Perennial	Nod	Fix	Ineffective
<i>T. canescens</i>	PL4188661999	Europe	Perennial	Nod	Fix	Ineffective
<i>T. fragiferum</i>	C1212	Europe	Perennial	Nod		No nodulation
<i>T. polymorphum</i>	87102	South America	Perennial	Nod	Fix	Ineffective
<i>T. longipes</i>	A2436817	North America	Perennial	Nod		No nodulation
<i>T. subterraneum</i>	York	Europe	Annual	Nod	Fix	Ineffective
<i>T. glanduliferum</i>	CP187182	Europe	Annual	Nod	Fix	Ineffective
<i>T. mulinerve</i>	87259	Africa	Annual	Nod		No nodulation
<i>T. tridentatum</i>	CQ1263	North America	Annual	Nod	Fix	Ineffective

## Symbiotaxonomy

*R. leguminosarum* bv. *trifolii* WSM1689 is a highly effective microsymbiont of the perennial Eurasian clover *Trifolium uniflorum* (Table 2). In contrast, WSM1689 does not nodulate the perennial *T. fragiferum* and forms white ineffective (Fix<sup>-</sup>) nodules with other perennial and annual clovers of Eurasian origin. Moreover, WSM1689 is either Nod<sup>-</sup> or Fix<sup>-</sup> on clovers of North American or African origin. Therefore, WSM1689 is unusual in having an extremely narrow clover host range for the establishment of effective N<sub>2</sub>-fixing symbiosis.

## 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 U.S. Department of Energy, Joint Genome Institute (JGI) for projects of relevance to agency missions. The genome project is deposited in the Genomes OnLine Database [28] and a finished genome sequence in IMG/GEBA. Sequencing, finishing and annotation were performed by the JGI. A summary of the project information is shown in Table 3.

### Growth conditions and DNA isolation

*Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 was grown to mid logarithmic phase in TY rich medium on a gyratory shaker at 28°C [29]. DNA was isolated from 60 mL of cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [30].

**Table 3.** Genome sequencing project information for *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Finished
MIGS-28	Libraries used	Illumina GAii shotgun and paired end 454 libraries
MIGS-29	Sequencing platforms	Illumina GAii and 454 GS FLX Titanium technologies
MIGS-31.2	Sequencing coverage	8.3x 454, 774.6x Illumina
MIGS-30	Assemblers	VELVET, version 1.1.05; Newbler, version 2.6; phrap, version SPS - 4.24
MIGS-32	Gene calling methods	Prodigal 1.4, GenePRIMP
	Genbank ID	Not yet available
	Genbank Date of Release	Not yet released
	GOLD ID	Gi06499
	NCBI project ID	62289
	Database: IMG-GEBA	2510065019
	Project relevance	Symbiotic nitrogen fixation, agriculture

### Genome sequencing and assembly

The genome of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 was sequenced at the Joint Genome Institute (JGI) using a combination of Illumina [31] and 454 technologies [32]. An Illumina GAii shotgun library which generated 73,565,648 reads totaling 5,591 Mbp, and a paired end 454 library with an average insert size of 12 Kbp which generated 376,185 reads totaling 93.4 Mbp of 454 data were generated for this genome. All general aspects of library construction and sequencing performed at the JGI can be found at [30]. The initial draft assembly contained 100 contigs in 4 scaffolds. The 454 paired end data was assembled with Newbler, version 2.6. The Newbler consensus sequences were computationally shredded into 2 Kbp overlapping fake reads (shreds). Illumina sequencing data was assembled with VELVET, version 1.1.05 [33], and the consensus sequence computationally shredded into 1.5 Kbp overlapping fake reads (shreds). We integrated the 454 Newbler consensus shreds, the Illumina VELVET consensus shreds and the read pairs in the 454 paired end library using parallel phrap, version SPS - 4.24 (High Performance Software, LLC). The software Consed [34-36] was used in the following finishing process. Illumina data was used to correct potential base errors and increase consensus quality using the software Polisher developed at JGI (Alla Lapidus, unpublished). Possible mis-assemblies were corrected using gapResolution (Cliff Han, unpublished), Dupfinisher [37], or sequencing cloned bridging PCR fragments with subcloning. Gaps between contigs were closed by editing in Consed, by PCR and by Bubble PCR (J-F Cheng, unpublished) primer walks. A total of 93 additional reactions were necessary to close gaps and to raise the quality of the finished sequence.

The total genome size is 6.9 Mbp and the final assembly is based on 57.3 Mbp of 454 draft data which provides an average 8.3x coverage of the genome and 5,345 Mbp of Illumina draft data which provides an average 774.6x coverage of the genome.

### Genome annotation

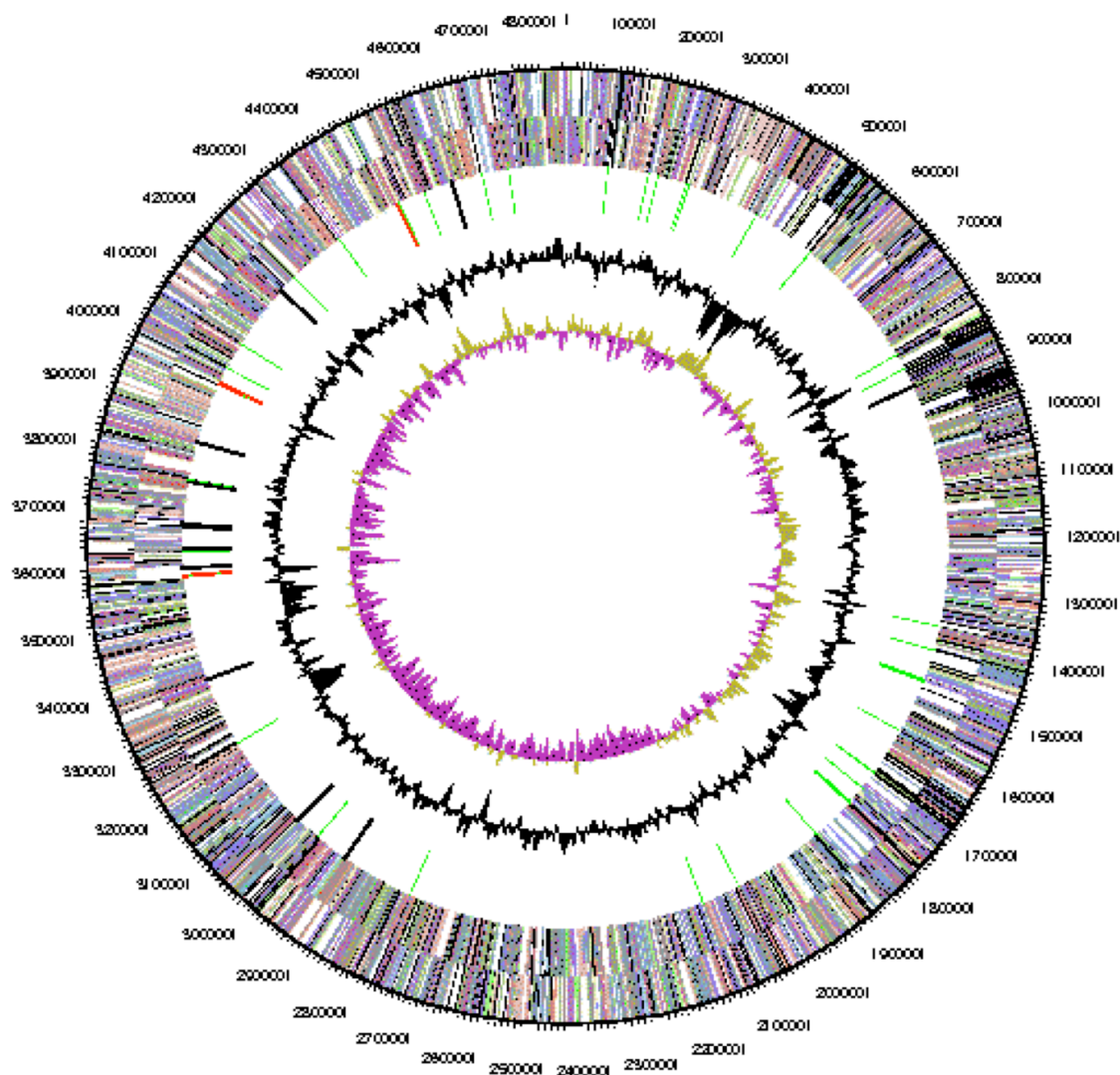
Genes were identified using Prodigal [38] as part of the DOE-JGI genome annotation pipeline, followed by a round of manual curation using the JGI GenePRIMP pipeline [39]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. These data sources were combined to assert a product description for each predicted protein. Non-coding genes and miscellaneous features were predicted using tRNAscan-SE [40], RNAMmer [41], Rfam [42], TMHMM [43], and SignalP [44]. Additional gene prediction analyses and functional annotation were performed within the Integrated Microbial Genomes (IMG-ER) platform [45,46].

### Genome properties

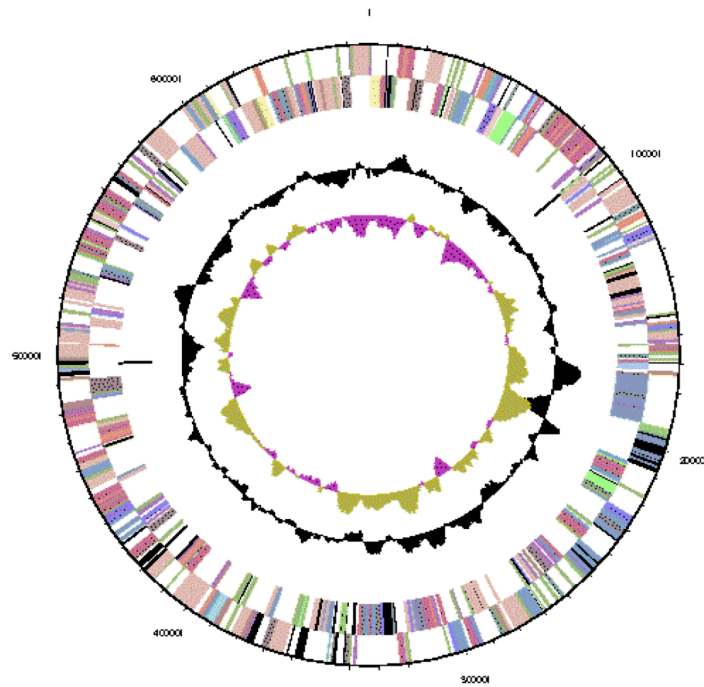
The genome is 6,903,379 nucleotides with 60.94% GC content (Table 4 and Figures 3a,3b,3c,3d,3e and Figure 3f), and comprised of 6 replicons. From a total of 6,798 genes, 6,709 were protein encoding and 89 RNA only encoding genes. Within the genome, 206 pseudogenes were also identified. The majority of genes (79.52%) 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 5.

**Table 4.** Genome Statistics for *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689.

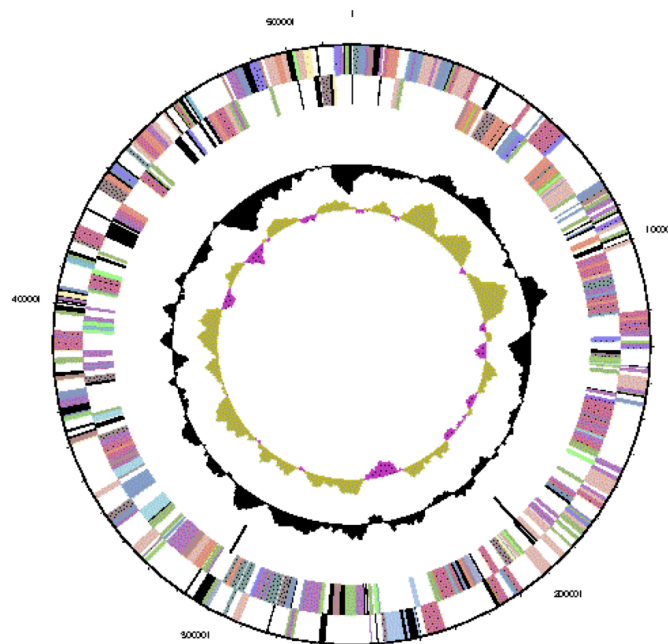
Attribute	Value	% of Total
Genome size (bp)	6,903,379	100.00
DNA coding region (bp)	6,004,795	86.98
DNA G+C content (bp)	4,206,909	60.94
Number of replicons	6	
Total genes	6,798	100.00
RNA genes	89	1.31
Protein-coding genes	6,709	98.69
Genes with function prediction	5,406	79.52
Genes assigned to COGs	5,400	79.44
Genes assigned Pfam domains	5,618	82.64
Genes with signal peptides	591	8.69
Genes coding transmembrane proteins	1,524	22.42
CRISPR repeats	0	



**Figure 3a.** Graphical circular map of Replicon WSM1689\_Rleg3\_Contig1814.1 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. 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.

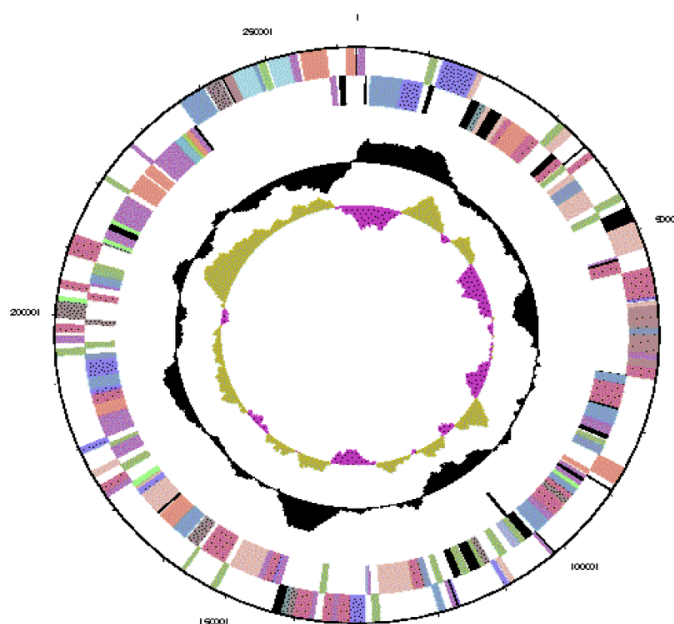


**Figure 3b.** Graphical circular map of replicon WSM1689\_Rleg3\_Contig1813.2 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. From outside to the center: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

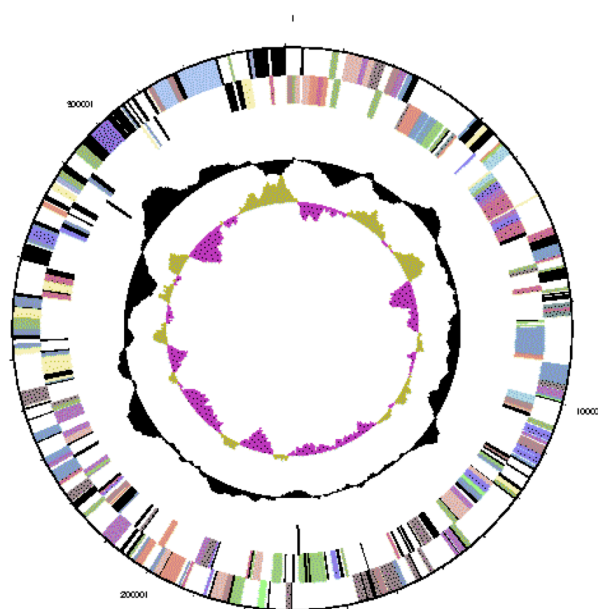


**Figure 3c.** Graphical circular map of replicon WSM1689\_Rleg3\_Contig1812.3 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. From outside to the center: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

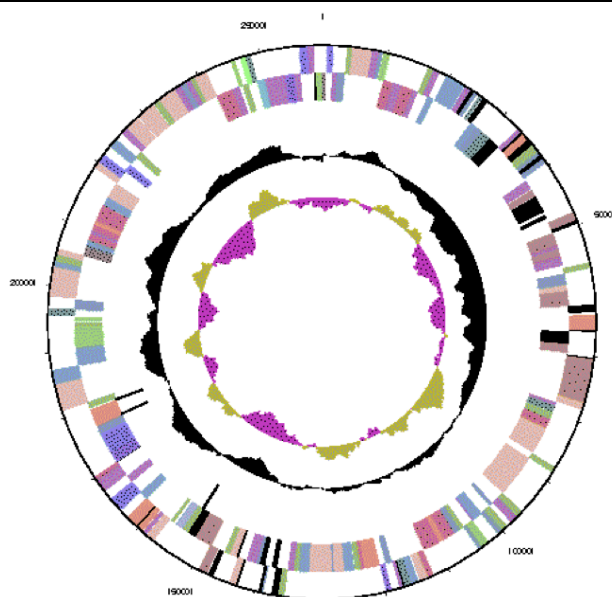




**Figure 3d.** Graphical circular map of replicon WSM1689\_Rleg3\_Contig1810.5 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. From outside to the center: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.



**Figure 3e.** Graphical circular map of replicon WSM1689\_Rleg3\_Contig1811.4 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. 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.



**Figure 3f.** Graphical circular map of replicon WSM1689\_Rleg3\_Contig1809.6 of the *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 genome. 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.

**Table 5.** Number of protein coding genes of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 associated with the general COG functional categories.

Code	Value	%age	COG Category
J	205	3.40	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	581	9.62	Transcription
L	153	2.53	Replication, recombination and repair
B	2	0.03	Chromatin structure and dynamics
D	39	0.65	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	66	1.09	Defense mechanisms
T	311	5.15	Signal transduction mechanisms
M	329	5.45	Cell wall/membrane biogenesis
N	81	1.34	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	82	1.36	Intracellular trafficking and secretion
O	187	3.10	Posttranslational modification, protein turnover, chaperones
C	311	5.15	Energy production conversion

**Table 5 (cont.).** Number of protein coding genes of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1689 associated with the general COG functional categories.

Code	Value	%age	COG Category
G	683	11.31	Carbohydrate transport and metabolism
E	629	10.42	Amino acid transport metabolism
F	105	1.74	Nucleotide transport and metabolism
H	192	3.18	Coenzyme transport and metabolism
I	222	3.68	Lipid transport and metabolism
P	297	4.92	Inorganic ion transport and metabolism
Q	147	2.43	Secondary metabolite biosynthesis, transport and catabolism
R	795	13.17	General function prediction only
S	620	10.27	Function unknown
-	1,398	20.56	Not in COGS
-	6,037	-	Total

## 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 gratefully acknowledge the funding received from the Murdoch University Sir Walter Murdoch Adjunct Professor Scheme for Professor Philip Poole.

## References

1. Terpolilli JJ, Hood GA, Poole PS. What determines the efficiency of N<sub>2</sub>-fixing *Rhizobium*-Legume symbioses? *Adv Microb Physiol* 2012; **60**: 325-389. [PubMed](#)  
<http://dx.doi.org/10.1016/B978-0-12-398264-3.00005-X>
2. Howieson JG, O'Hara GW, Carr SJ. Changing roles for legumes in Mediterranean agriculture: developments from an Australian perspective. *Field Crops Res* 2000; **65**:107-122.  
[http://dx.doi.org/10.1016/S0378-4290\(99\)00081-Z](http://dx.doi.org/10.1016/S0378-4290(99)00081-Z)
3. Lamont EJ, Zoghalmi A, Hamilton RS, Bennett SJ. Clovers (*Trifolium* L.). In: Maxted N, Bennett SJ, editors. *Plant Genetic Resources of Legumes in the Mediterranean*. Dordrecht: Kluwer Academic Publishers; 2001. p 79-98.
4. Howieson J, Yates R, O'Hara G, 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.  
<http://dx.doi.org/10.1071/EA03167>
5. Reeve W, O'Hara G, Chain P, Ardley J, Brau L, Nandesena K, Tiwari R, Copeland A, Nolan M, Han C, et al. Complete genome sequence of *Rhizobium leguminosarum* bv. *trifolii* strain WSM1325, an effective microsymbiont of annual Mediterranean clovers. *Stand Genomic Sci* 2010; **2**:347-356. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.852027>
6. Reeve W, O'Hara G, Chain P, Ardley J, Brau L, Nandesena K, Tiwari R, Malfatti S, Kiss H, Lapidus A, 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. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.44642>
7. Howieson JG, Ewing MA, D'antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil* 1988; **105**:179-188.  
<http://dx.doi.org/10.1007/BF02376781>
8. Beringer JE. R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol* 1974; **84**:188-198. [PubMed](#)  
<http://dx.doi.org/10.1099/00221287-84-1-188>

9. Terpolilli JJ. Why are the symbioses between some genotypes of *Sinorhizobium* and *Medicago* suboptimal for N<sub>2</sub> fixation? Perth: Murdoch University; 2009. 223 p.
10. Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, Tatusova T, Thomson N, Allen M, Angiuoli SV, 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**:541-547. [PubMed](#) <http://dx.doi.org/10.1038/nbt1360>
11. 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](#) <http://dx.doi.org/10.1073/pnas.87.12.4576>
12. Garrity GM, Bell JA, Lilburn T. Phylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT (eds), *Bergey's Manual of Systematic Bacteriology*, Second Edition, Volume 2, Part B, Springer, New York, 2005, p. 1.
13. 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.
14. 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](#) <http://dx.doi.org/10.1099/ijs.0.64188-0>
15. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's Manual of Systematic Bacteriology*. Second ed: New York: Springer - Verlag; 2005. p 324.
16. Skerman VBD, McGowan V, Sneath PHA. Approved Lists of Bacterial Names. *Int J Syst Bacteriol* 1980; **30**:225-420. <http://dx.doi.org/10.1099/00207713-30-1-225>
17. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.
18. Frank B. Über die Pilzsymbiose der Leguminosen. *Ber Dtsch Bot Ges* 1889; **7**:332-346.
19. 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.
20. 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](#)
21. 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. <http://dx.doi.org/10.1099/00207713-20-1-11>
22. 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 1889AL, *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](#) <http://dx.doi.org/10.1099/ijs.0.65621-0>
23. Agents B. Technical rules for biological agents. TRBA (<http://www.baua.de>):466.
24. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* 2000; **25**:25-29. [PubMed](#) <http://dx.doi.org/10.1038/75556>
25. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* 2011; **28**:2731-2739. [PubMed](#) <http://dx.doi.org/10.1093/molbev/msr121>
26. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
27. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. <http://dx.doi.org/10.2307/2408678>
28. Liolios K, Mavromatis K, Tavernarakis N, Kyrpides NC. The Genomes On Line Database (GOLD) in

- 2007: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res* 2008; **36**:D475-D479. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm884>
29. Reeve WG, Tiwari RP, Worsley PS, Dilworth MJ, Glenn AR, Howieson JG. Constructs for insertional mutagenesis, transcriptional signal localization and gene regulation studies in root nodule and other bacteria. *Microbiology* 1999; **145**:1307-1316. [PubMed](#) <http://dx.doi.org/10.1099/13500872-145-6-1307>
30. DOE Joint Genome Institute. <http://my.jgi.doe.gov/general/index.html>
31. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. [PubMed](#) <http://dx.doi.org/10.1517/14622416.5.4.433>
32. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005; **437**:376-380. [PubMed](#)
33. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics* 2010;Chapter 11:Unit 11 5.
34. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:186-194. [PubMed](#) <http://dx.doi.org/10.1101/gr.8.3.175>
35. Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998; **8**:175-185. [PubMed](#) <http://dx.doi.org/10.1101/gr.8.3.175>
36. Gordon D, Abajian C, Green P. Consed: a graphical tool for sequence finishing. *Genome Res* 1998; **8**:195-202. [PubMed](#) <http://dx.doi.org/10.1101/gr.8.3.195>
37. Han C, Chain P. Finishing repeat regions automatically with Dupfinisher. In: Valafar HRAH, editor. *Proceeding of the 2006 international conference on bioinformatics & computational biology*: CSREA Press; 2006. p 141-146.
38. 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**:119. [PubMed](#) <http://dx.doi.org/10.1186/1471-2105-11-119>
39. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kypides NC. GenePRIMP: a gene prediction improvement pipeline for prokaryotic genomes. *Nat Methods* 2010; **7**:455-457. [PubMed](#) <http://dx.doi.org/10.1038/nmeth.1457>
40. Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 1997; **25**:955-964. [PubMed](#)
41. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* 2007; **35**:3100-3108. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm160>
42. Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res* 2003; **31**:439-441. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkg006>
43. Krogh A, Larsson B, von Heijne G, Sonnhammer EL. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* 2001; **305**:567-580. [PubMed](#) <http://dx.doi.org/10.1006/jmbi.2000.4315>
44. Bendtsen JD, Nielsen H, von Heijne G, Brunak S. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol* 2004; **340**:783-795. [PubMed](#) <http://dx.doi.org/10.1016/j.jmb.2004.05.028>
45. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kypides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#) <http://dx.doi.org/10.1093/bioinformatics/btp393>
46. Integrated Microbial Genomes (IMG-ER) platform. <http://img.jgi.doe.gov/er>