

## Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain TA1

Wayne Reeve<sup>\*1</sup>, Rui Tian<sup>1</sup>, Sofie De Meyer<sup>1</sup>, Vanessa Melino<sup>1</sup>, Jason Terpolilli<sup>1</sup>, Julie Ardley<sup>1</sup>, Ravi Tiwari<sup>1</sup>, John Howieson<sup>1</sup>, Ronald Yates<sup>1,2</sup>, Graham O'Hara<sup>1</sup>, Mohamed Ninawi<sup>1</sup>, Hazuki Teshima<sup>3</sup>, David Bruce<sup>3</sup>, Chris Detter<sup>3</sup>, Roxanne Tapia<sup>3</sup>, Cliff Han<sup>3</sup>, Chia-Lin Wei<sup>3</sup>, Marcel Huntemann<sup>3</sup>, James Han<sup>3</sup>, I-Min Chen<sup>5</sup>, Konstantinos Mavromatis<sup>3</sup>, Victor Markowitz<sup>5</sup>, Natalia Ivanova<sup>3</sup>, Galina Ovchinnikova<sup>3</sup>, Ioanna Pagani<sup>3</sup>, Amrita Pati<sup>3</sup>, Lynne Goodwin<sup>4</sup>, Sam Pitluck<sup>3</sup>, Tanja Woyke<sup>3</sup> & Nikos Kyrpides<sup>3</sup>.

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

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

<sup>3</sup> DOE Joint Genome Institute, Walnut Creek, California, USA

<sup>4</sup> Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

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

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

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

---

*Rhizobium leguminosarum* bv. *trifolii* strain TA1 is an aerobic, motile, Gram-negative, non-spore-forming rod that is an effective nitrogen fixing microsymbiont on the perennial clovers originating from Europe and the Mediterranean basin. TA1 however is ineffective with many annual and perennial clovers originating from Africa and America. Here we describe the features of *R. leguminosarum* bv. *trifolii* strain TA1, together with genome sequence information and annotation. The 8,618,824 bp high-quality-draft genome is arranged in a 6 scaffold of 32 contigs, contains 8,493 protein-coding genes and 83 RNA-only encoding genes, and is one of 20 rhizobial genomes sequenced as part of the DOE Joint Genome Institute 2010 Community Sequencing Program.

---

### Introduction

Biological fixation of inert atmospheric dinitrogen gas is a process that can only be performed by certain prokaryotes in the domains *Archaea* and *Bacteria*. By far the greatest amounts of nitrogen (N) are fixed by specialized soil bacteria (root nodule bacteria or rhizobia) that form proto-cooperative, non-obligatory symbiotic relationships with legumes [1]. Indeed, these symbioses contribute ~40 million tonnes of N annually to support global food production [2].

Species of the legume genus *Trifolium* (clovers) are amongst the most widely cultivated pasture legumes. Naturally, this genus inhabits three distinct centers of diversity with approximately 28% of species in the Americas, 57% in Eurasia and 15% in Sub-Saharan Africa [3]. A smaller subset of about 30 species, almost all of Eurasian origin, are widely grown as annual and perennial species in pasture systems in Mediterranean and temperate

regions [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). Clovers usually form N<sub>2</sub>-fixing symbioses with the common soil bacterium *Rhizobium leguminosarum* bv. *trifolii*, and different combinations of *Trifolium* hosts and strains of *R. leguminosarum* bv. *trifolii* can vary markedly in symbiotic compatibility [4], resulting in a broad range of symbiotic developmental outcomes ranging from ineffective (non-nitrogen fixing) nodulation to fully effective N<sub>2</sub>-fixing partnerships [5].

In Australia, *Rhizobium leguminosarum* bv. *trifolii* strain TA1 (initially designated BA-Tas) has a long history of use as a commercial inoculant for *Trifolium* spp. [6]. TA1 was originally isolated from a root nodule on the annual species *T. subterraneum* in Bridport, Tasmania in the early

1950's [6]. This isolate is likely to be a naturalized strain of European origin that arrived by chance in Tasmania in the 1800's. Although widely used as a microsymbiont of European clovers, it became evident that this soil saprophyte is not acid tolerant [7] and survives poorly when coated onto clover seed with a peat based carrier [8-10]. Nevertheless, TA1 remains the commercial inoculant in Australia for perennial (*T. repens*, *T. pratense*, *T. fragiferum*, *T. hybridum*, *T. tumens* (talish clover)) and annual (*T. alexandrinum* (berseem clover), *T. glomeratum* (cluster clover) and *T. dubium* (suckling clover)) clovers of European origin [11]. Furthermore, this *R. leguminosarum* bv. *trifolii* strain has been adopted by the international community as a model organism to investigate the biology of the *Trifolium-Rhizobium* symbiosis [12]. Here we present a summary classification and a set of general features for *R. leguminosarum* bv. *trifolii* strain TA1 together with the description of the complete genome sequence and its annotation.

## Classification and general features

*R. leguminosarum* bv. *trifolii* strain TA1 is a motile, Gram-negative, non-spore-forming rod (Figure 1 Left and Center) in the order *Rhizobiales* of the class *Alphaproteobacteria*. It is slow growing, forming 1-4 mm diameter colonies within 3-5 days grown on half Lupin Agar ( $\frac{1}{2}$ LA) [13] at 28°C. Colonies on  $\frac{1}{2}$ LA are white-opaque, slightly domed, 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 TA1 in a 16S rRNA sequence based tree. This strain clusters closest to *R. leguminosarum* bv. *trifolii* T24 and *R.*

*leguminosarum* bv. *phaseoli* RRE6 with 99.9% and 99.8% sequence identity, respectively.

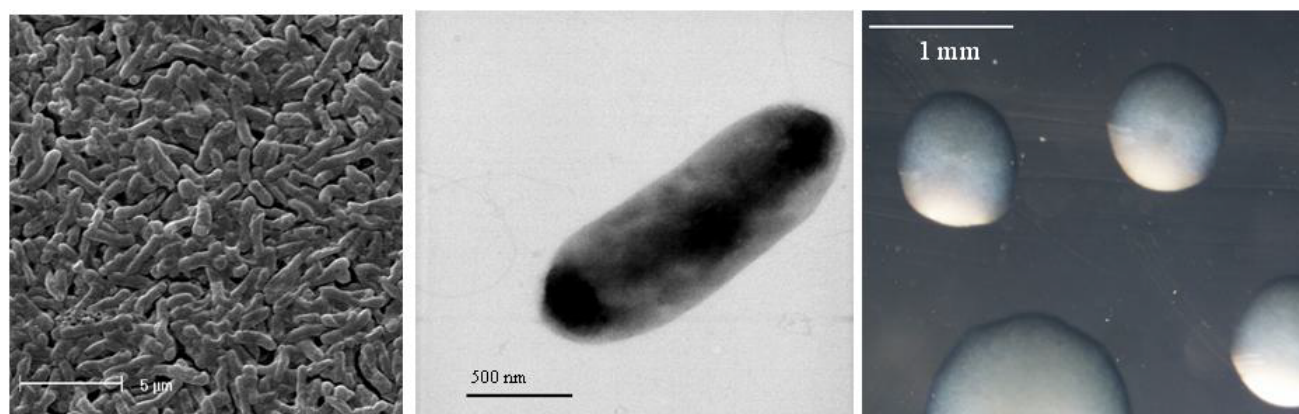
## Symbiotaxonomy

*Rhizobium leguminosarum* bv. *trifolii* strain TA1 is currently the commercial inoculant for white (*Trifolium repens*), red (*Trifolium pratense*) and strawberry (*Trifolium fragiferum*) clovers in Australia. TA1 in general is not as effective for nitrogen fixation on annual clovers as other strains, such as WSM1325 [34,35]. However TA1 is of particular interest because it displays a broad host range for nodulation and nitrogen fixation across annual and perennial clovers originating from the European and Mediterranean centre of origin of clovers [1]. TA1 is generally able to nodulate but unable to fix with many annual and perennial clovers originating from Africa and America [34].

## 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 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 [33] and an improved-high-quality-draft genome sequence in IMG. Sequencing, finishing and annotation were performed by the JGI. A summary of the project information is shown in Table 2.



**Figure 1.** Images of *Rhizobium leguminosarum* bv. *trifolii* strain TA1 using scanning (Left) and transmission (Center) electron microscopy as well as light microscopy to visualize colony morphology on solid media (Right).

**Table 1.** Classification and general features of *Rhizobium leguminosarum* bv. *trifolii* strain TA1 according to the MIGS recommendations [14].

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [15]
		Phylum <i>Proteobacteria</i>	TAS [16]
		Class <i>Alphaproteobacteria</i>	TAS [17,18]
		Order <i>Rhizobiales</i>	TAS [17,19]
		Family <i>Rhizobiaceae</i>	TAS [20,21]
		Genus <i>Rhizobium</i>	TAS [20,22-25]
	Current classification	Species <i>Rhizobium leguminosarum</i> bv. <i>trifolii</i>	TAS [20,22,25,26]
	Gram stain	Negative	TAS [27]
	Cell shape	Rod	TAS [27]
	Motility	Motile	TAS [27]
	Sporulation	Non-sporulating	TAS [27]
	Temperature range	Mesophile	TAS [27]
	Optimum temperature	28°C	TAS [27]
	Salinity	Not reported	
MIGS-22	Oxygen requirement	Aerobic	TAS [27]
	Carbon source	Varied	
	Energy source	Chemoorganotroph	TAS [27]
MIGS-6	Habitat	Soil, root nodule, on host	IDA
MIGS-15	Biotic relationship	Free living, symbiotic	IDA
MIGS-14	Pathogenicity	Non-pathogenic	TAS [27]
	Biosafety level	1	TAS [28]
	Isolation	Root nodule of <i>Trifolium subterraneum</i>	TAS [29]
MIGS-4	Geographic location	Bridport, Tasmania	IDA
MIGS-5	Nodule collection date	1953	IDA
MIGS-4.1	Longitude	147.667	IDA
MIGS-4.2	Latitude	-41.0335	IDA
MIGS-4.3	Depth	Not recorded	
MIGS-4.4	Altitude	Not recorded	

Evidence codes – IDA: Inferred from Direct Assay; TAS: Traceable Author Statement (i.e., a direct report exists in the literature). These evidence codes are from the Gene Ontology project [30].

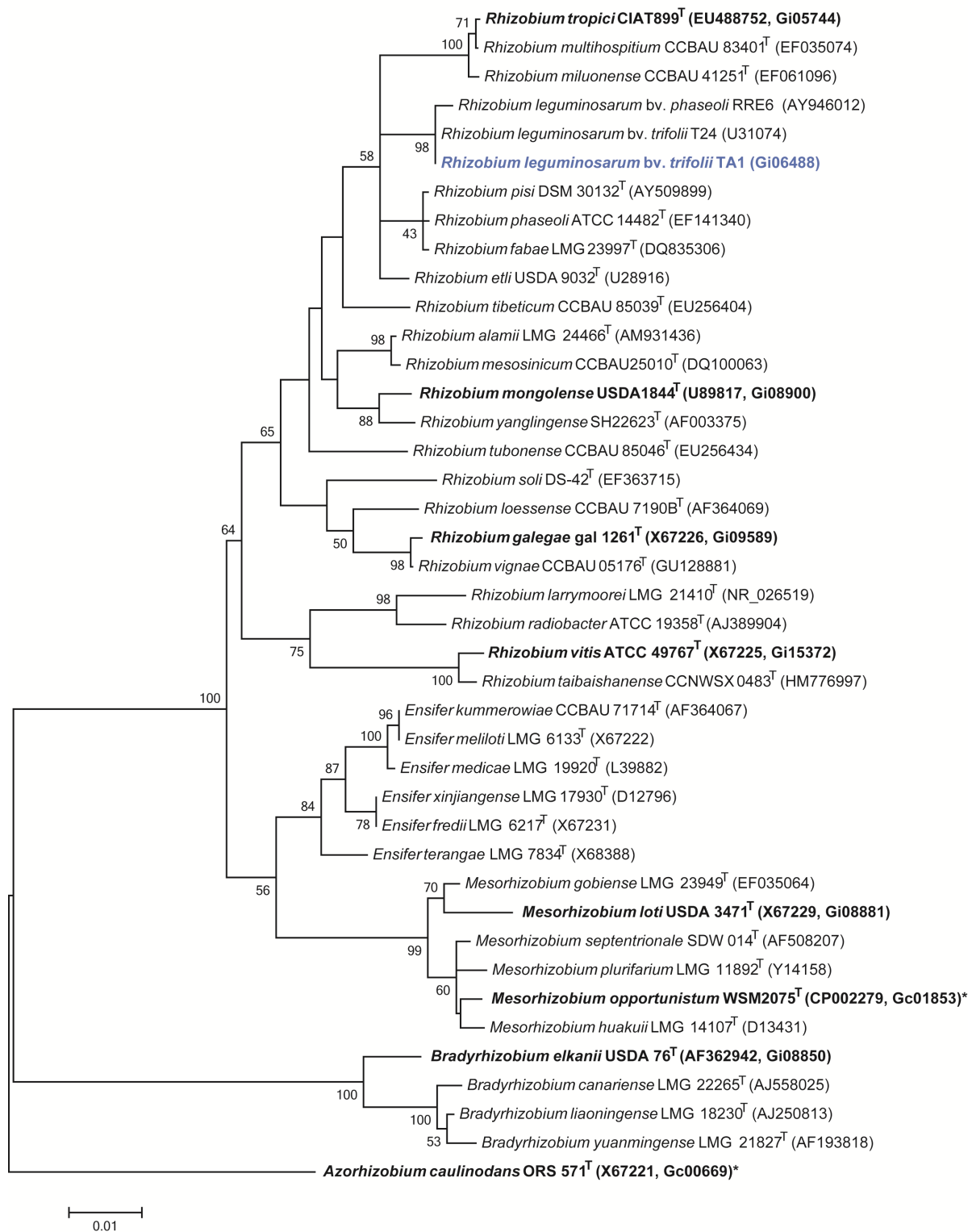


Figure 2. Phylogenetic tree showing the relationship of *Rhizobium leguminosarum* bv. *trifolii* strain TA1 (shown in blue print) with some of the root nodule bacteria in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1,307 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [31]. The tree was built using the maximum likelihood method with the General Time Reversible model. Bootstrap analysis [32] 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 [33] are in bold print and the GOLD ID is mentioned after the accession number. Published genomes are designated with an asterisk.

**Table 2.** Genome sequencing project information for *Rhizobium leguminosarum* bv. *trifolii* strain TA1.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
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	7.8× 454 paired end, 764.2× Illumina
MIGS-30	Assemblers	Velvet 1.0.13, Newbler 2.3, phrap 4.24
MIGS-32	Gene calling methods	Prodigal 1.4, GenePRIMP
	GOLD ID	Gi0648
	NCBI project ID	63831
	Database: IMG	2510461076
	Project relevance	Symbiotic N <sub>2</sub> fixation, agriculture

### Growth conditions and DNA isolation

*Rhizobium leguminosarum* bv. *trifolii* strain TA1 was grown to mid logarithmic phase in TY rich media [36] on a gyratory shaker at 28°C. DNA was isolated from 60 ml of cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [37].

### Genome sequencing and assembly

The genome of *Rhizobium leguminosarum* bv. *trifolii* strain TA1 was sequenced at the Joint Genome Institute (JGI) using a combination of Illumina [38] and 454 technologies [39]. An Illumina GAii shotgun library which generated 66,421,308 reads totaling 5,048 Mb, and a paired end 454 library with an average insert size of 13 kb which generated 393,147 reads totaling 100.1 Mb of 454 data were generated for this genome. All general aspects of library construction and sequencing performed at the JGI can be found at the JGI user homepage [40]. The initial draft assembly contained 199 contigs in 5 scaffolds. The 454 paired end data was assembled with Newbler, version 2.3. The Newbler consensus sequences were computationally shredded into 2 kb overlapping fake reads (shreds). Illumina sequencing data were assembled with VELVET, version 1.0.13 [41], and the consensus sequence were computationally shredded into 1.5 kb 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 [42-44] 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 (Han, 2006), 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 275 additional reactions were necessary to close gaps and to raise the quality of the finished sequence. The estimated genome size is 7.6 Mb and the final assembly is based on 65.3 Mb of 454 draft data which provides an average of 8.6× coverage of the genome and 4,864.7 Mb of Illumina draft data which provides an average 640.1× coverage of the genome.

### Genome annotation

Genes were identified using Prodigal [45] as part of the DOE-JGI Annotation pipeline [46], followed by a round of manual curation using the JGI GenePRIMP pipeline [47]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant 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 [48], RNAMMer [49], Rfam [50], TMHMM [51], and SignalP [52]. Additional gene prediction analyses and functional annotation were performed within the Integrated Microbial Genomes (IMG-ER) platform [37,53].

## Genome properties

The genome is 8,618,824 nucleotides with 60.74% GC content (Table 3) and comprised of 32 contigs in 6 scaffolds (Figure 3). From a total of 8,576 genes, 8,493 were protein encoding and 83 RNA only encoding genes. The majority of genes (77.85%) 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 sequencing project information for *Rhizobium leguminosarum* bv. *trifolii* strain SRDI943.

Attribute	Value	% of Total
Genome size (bp)	8,618,824	100.00
DNA coding region (bp)	7,407,820	85.95
DNA G+C content (bp)	5,234,677	60.74
Number of scaffolds	6	
Number of contigs	32	
Total genes	8,576	100.00
RNA genes	83	0.97
rRNA operons*	1	0.01
Protein-coding genes	8,493	99.03
Genes with function prediction	6,676	77.85
Genes assigned to COGs	6,673	77.81
Genes assigned Pfam domains	6,944	80.97
Genes with signal peptides	727	8.48
Genes with transmembrane helices	1,897	22.12
CRISPR repeats	0	

\*1 copy of 23S rRNA, 2 copies of 16S and 2 copies of 5S rRNA genes

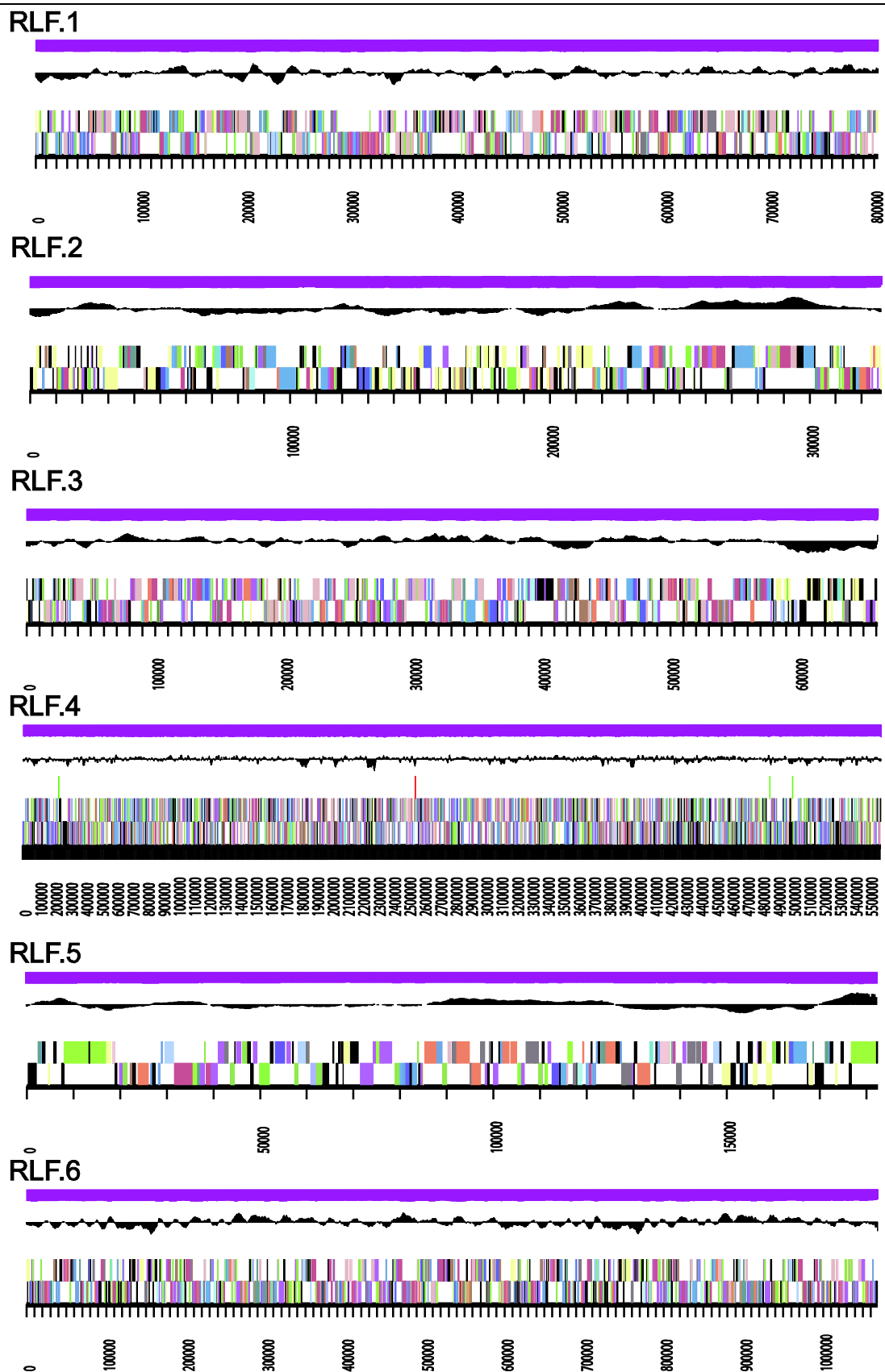


Figure 3. Graphical linear map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain TA1. From outside to the center: Genes on forward strand (color by COG categories), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

**Table 4.** Number of protein coding genes of *Rhizobium leguminosarum* bv. *trifolii* TA1 associated with the general COG functional categories.

Code	Value	%age	COG Category
J	247	3.29	Translation, ribosomal structure and biogenesis
A	1	0.01	RNA processing and modification
K	751	10.01	Transcription
L	317	4.23	Replication, recombination and repair
B	3	0.04	Chromatin structure and dynamics
D	44	0.59	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	92	1.23	Defense mechanisms
T	402	5.36	Signal transduction mechanisms
M	365	4.87	Cell wall/membrane biogenesis
N	100	1.33	Cell motility
Z	2	0.03	Cytoskeleton
W	0	0.00	Extracellular structures
U	114	1.52	Intracellular trafficking and secretion
O	217	2.89	Posttranslational modification, protein turnover, chaperones
C	384	5.12	Energy production conversion
G	746	9.95	Carbohydrate transport and metabolism
E	803	10.71	Amino acid transport metabolism
F	134	1.79	Nucleotide transport and metabolism
H	235	3.13	Coenzyme transport and metabolism
I	271	3.61	Lipid transport and metabolism
P	374	4.99	Inorganic ion transport and metabolism
Q	201	2.68	Secondary metabolite biosynthesis, transport and catabolism
R	976	13.02	General function prediction only
S	720	9.60	Function unknown
-	1,903	22.19	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 gratefully acknowledge the funding received from the Mur-

doch University Strategic Research Fund through the Crop and Plant Research Institute (CaPRI) and the Centre for Rhizobium Studies (CRS) at Murdoch University. The authors would like to thank the Australia-China Joint Research Centre for Wheat Improvement (ACCWI) and SuperSeed Technologies (SST) for financially supporting Mohamed Ninawi's PhD project.

## References

- O'Hara GW. The role of nitrogen fixation in crop production. *J Crop Prod* 1998; **1**:115-138. [http://dx.doi.org/10.1300/J144v01n02\\_06](http://dx.doi.org/10.1300/J144v01n02_06)
- Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil* 2008; **311**:1-18. <http://dx.doi.org/10.1007/s11104-008-9668-3>
- 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.
- 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>
- Melino VJ, Drew EA, Ballard RA, Reeve WG, Thomson G, White RG, O'Hara GW. Identifying abnormalities in symbiotic development between *Trifolium* spp. and *Rhizobium leguminosarum* bv. *trifolii* leading to sub-optimal and ineffective nodule phenotypes. *Ann Bot* 2012; **110**:1559-1572. [PubMed http://dx.doi.org/10.1093/aob/mcs206](http://dx.doi.org/10.1093/aob/mcs206)
- 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. <http://dx.doi.org/10.1071/EA03159>
- Watkin E, O'Hara G, Howieson JG, Glenn AR. Identification of tolerance to soil acidity in inoculant strains of *Rhizobium leguminosarum* bv. *trifolii*. *Soil Biol Biochem* 2000; **32**:1393-1403. [http://dx.doi.org/10.1016/S0038-0717\(00\)00057-2](http://dx.doi.org/10.1016/S0038-0717(00)00057-2)
- Deaker R, Roughley RJ, Kennedy IR. Legume seed inoculation technology: a review. *Soil Biol Biochem* 2004; **36**:1275-1288. <http://dx.doi.org/10.1016/j.soilbio.2004.04.009>
- Deaker R, Roughley RJ, Kennedy IR. Desiccation tolerance of rhizobia when protected by synthetic polymers. *Soil Biol Biochem* 2007; **39**:573-580. <http://dx.doi.org/10.1016/j.soilbio.2006.09.005>
- Hartley E, Gemell G, Deaker R. Some factors that contribute to poor survival of rhizobia on preinoculated legume seed. *Crop Pasture Sci* 2012; **63**:858-865. <http://dx.doi.org/10.1071/CP12132>
- Drew EA, Herridge DF, Ballard RA, O'Hara GW, Deaker R, Denton MD, Yates RJ, Gemell G, Hartley E, Phillips L, *et al.* *Inoculating legumes: a practical guide*. Grains Research and Development Corporation; 2012.
- Unkovich MJ. Nitrogen fixation in Australian dairy systems: review and prospect. *Crop Pasture Sci* 2012; **63**:787-804. <http://dx.doi.org/10.1071/CP12180>
- 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>
- 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](http://dx.doi.org/10.1038/nbt1360)
- 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](http://dx.doi.org/10.1073/pnas.87.12.4576)
- 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.
17. Editor L. 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>
  18. 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.
  19. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's Manual of Systematic Bacteriology*. Second ed: New York: Springer - Verlag; 2005.
  20. 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>
  21. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.
  22. Frank B. Über die Pilzsymbiose der Leguminosen. *Ber Dtsch Bot Ges* 1889; **7**:332-346.
  23. 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.
  24. 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](#)
  25. 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>
  26. 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>
  27. Kuykendall LD, Young JM, Martínez-Romero E, Kerr A, Sawada H. Genus I. *Rhizobium*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's Manual of Systematic Bacteriology*. Second ed. Volume 2. New York: Springer - Verlag; 2005.
  28. Agents B. Technical rules for biological agents. TRBA (<http://www.baua.de>):466.
  29. Brockwell J, McIlroy RA, Hebb DM. The Australian collection of *Rhizobium* strains for temperate legumes catalogue. CSIRO Plant Industry (Australia); 1998. 164 p.
  30. 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>
  31. 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>
  32. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. <http://dx.doi.org/10.2307/2408678>
  33. 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**(Database issue):D475-D479. [PubMed](#) <http://dx.doi.org/10.1093/nar/gkm884>
  34. 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>
  35. 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 *Rhi-*

- zobium 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>
36. 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>
  37. DOE Joint Genome Institute.  
<http://my.jgi.doe.gov/general/index.html>
  38. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. [PubMed](#)  
<http://dx.doi.org/10.1517/14622416.5.4.433>
  39. 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](#)
  40. DOE Joint Genome Institute.  
<http://my.jgi.doe.gov/general/index.html>
  41. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics* 2010;Chapter 11:Unit 11 5.
  42. 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>
  43. 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>
  44. 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>
  45. 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>
  46. Mavromatis K, Ivanova NN, Chen IM, Szeto E, Markowitz VM, Kyrpides NC. The DOE-JGI Standard operating procedure for the annotations of microbial genomes. *Stand Genomic Sci* 2009; **1**:63-67. [PubMed](#)  
<http://dx.doi.org/10.4056/sigs.632>
  47. Pati A, Ivanova NN, Mikhailova N, Ovchinnikova G, Hooper SD, Lykidis A, Kyrpides 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>
  48. 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](#)
  49. 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>
  50. 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>
  51. 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>
  52. 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>
  53. Markowitz VM, Mavromatis K, Ivanova NN, Chen IM, Chu K, Kyrpides NC. IMG ER: a system for microbial genome annotation expert review and curation. *Bioinformatics* 2009; **25**:2271-2278. [PubMed](#)  
<http://dx.doi.org/10.1093/bioinformatics/btp393>