Genome sequence of the clover-nodulating *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565.

Wayne Reeve[*1], Elizabeth Drew[2], Ross Ballard[2], Vanessa Melino[1], Rui Tian[1], Sofie De Meyer[1], Lambert Brau[3], Mohamed Ninawi[1], Hazuki Teshima[4], Lynne Goodwin[4], Patrick Chain[4], Konstantinos Liolios[5], Amrita Pati[5], Konstantinos Mavromatis[5], Natalia Ivanova[5], Victor Markowitz[6], Tanja Woyke[5] & Nikos Kyrpides[5].

[1] Centre for Rhizobium Studies, Murdoch University, Western Australia, Australia

[2] South Australian Research and Development Institute, Urrbrae, South Australia, Australia

[3] School of Life and Environmental Sciences, Faculty of Science & Technology, Deakin University, Melbourne, Victoria, Australia

[4] Los Alamos National Laboratory, Bioscience Division, Los Alamos, New Mexico, USA

[5] DOE Joint Genome Institute, Walnut Creek, California, USA

[6] Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA

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

Rhizobium leguminosarum bv. *trifolii* SRDI565 (syn. N8-J) is an aerobic, motile, Gram-negative, non-spore-forming rod. SRDI565 was isolated from a nodule recovered from the roots of the annual clover *Trifolium subterraneum* subsp. *subterraneum* grown in the greenhouse and inoculated with soil collected from New South Wales, Australia. SRDI565 has a broad host range for nodulation within the clover genus, however N_2-fixation is sub-optimal with some *Trifolium* species and ineffective with others. Here we describe the features of *R. leguminosarum* bv. *trifolii* strain SRDI565, together with genome sequence information and annotation. The 6,905,599 bp high-quality-draft genome is arranged into 7 scaffolds of 7 contigs, contains 6,750 protein-coding genes and 86 RNA-only encoding genes, and is one of 100 rhizobial genomes sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project.

Introduction

Plant available nitrogen is a precious commodity in many agricultural soils and the most commonly limiting nutrient in plant growth. The supply of plant available nitrogen to nitrogen (N)-deficient farming systems is thus vital to productivity [1]. The application of industrially fixed nitrogenous fertilizer can meet the demand for N. However, this is a costly option as the price of nitrogenous fertilizer is connected to the cost of fossil fuels required for its production. Furthermore, the use of nitrogenous fertilizer contributes to greenhouse gas emissions and pollution of the environment. A more environmentally sustainable option is to exploit the process of biological nitrogen fixation that occurs in the symbiosis between legumes and rhizobia [2].

In this symbiotic association, rhizobia reduce atmospheric dinitrogen (N_2) into bioavailable N that can be used by the plant for growth. Pasture legumes, including the clovers that comprise the *Trifolium* genus, are major contributors of biologically fixed N_2 to mixed farming systems throughout the world [3,4]. In Australia, soils with a history of growing *Trifolium* spp. have developed large and symbiotically diverse populations of *Rhizobium leguminosarum* bv. *trifolii* (*R. l. trifolii*) that are able to infect and form nodules on a range of clover species. The N_2-fixation capacity of the symbioses established by different combinations of clover hosts (*Trifolium* spp.) and strains of *R. l. trifolii* can vary from 10 to 130% when compared to an effective host-strain combination [3-9].

R. l. trifolii strain SRDI565 (syn. N8-J [10]) was isolated from a nodule recovered from the roots of the annual clover *Trifolium subterraneum* subsp. *subterraneum* that had been inoculated with soil collected from under a mixed pasture stand from Tumet, New South Wales, Australia and grown in N deficient media for four weeks after inoculation, in the greenhouse. SRDI565 was first noted for its sub-optimal N_2-fixation capacity on *T. subterraneum* cv. Campeda (<60% of that with strain WSM1325) and formation of white (Fix-) pseudo-nodules on *T. subterraneum* cv. Clare [10,11]. Here we present a preliminary description of the general features for *R. leguminosarum* bv. *trifolii* strain SRDI565 together with its genome sequence and annotation.

Classification and general features

R. l. trifolii strain SRDI565 is a motile, Gram-negative rod (Figure 1 Left and Center) in the order *Rhizobiales* of the class *Alphaproteobacteria*. It is fast growing, forming colonies within 3-4 days when grown on half strength Lupin Agar (½LA) [12] at 28°C. Colonies on ½LA are white-opaque, slightly domed and moderately mucoid with smooth margins (Figure 1 Right).

Symbiotaxonomy

R. l. trifolii SRDI565 forms nodules on (Nod+), and fixes N_2 (Fix+) with, a range of annual and perennial clover species of Mediterranean origin (Table 2). SRDI565 forms white, ineffective (Fix-) nodules with annual clovers *T. glanduliferum* and *T. subterraneum* cv. Clare, and with the perennial clovers *T. pratense* and *T. polymorphum*. SRDI565 does not form nodules on *T. vesiculosum*.

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 [30] 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 3.

Minimum Information about the Genome Sequence (MIGS) is provided in Table 1. Figure 2 shows the phylogenetic neighborhood of *R. l. trifolii* strain SRDI565 in a 16S rRNA sequence based tree. This strain clusters closest to *R. l. trifolii* T24 and *Rhizobium leguminosarum* bv. *phaseoli* RRE6 with 99.8% and 99.6% sequence identity, respectively.

Growth conditions and DNA isolation

Rhizobium leguminosarum bv. *trifolii* strain SRDI565 was cultured to mid logarithmic phase in 60 ml of TY rich media [31] on a gyratory shaker at 28°C. DNA was isolated from the cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [32].
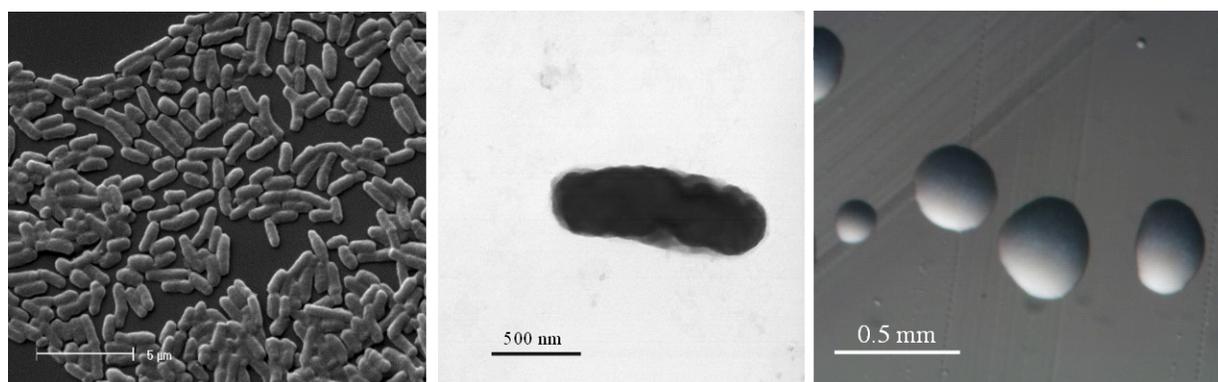

Figure 1. Images of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 using scanning (Left) and transmission (Center) electron microscopy as well as light microscopy to show the colony morphology on solid media (Right).

Table 1. Classification and general features of *Rhizobium leguminosarum* bv. *trifolii* SRDI565 according to the MIGS recommendations [13]

MIGS ID	Property	Term	Evidence code
		Domain *Bacteria*	TAS [13,14]
		Phylum *Proteobacteria*	TAS [15]
		Class *Alphaproteobacteria*	TAS [16]
	Current classification	Order *Rhizobiales*	TAS [17,18]
		Family *Rhizobiaceae*	TAS [19,20]
		Genus *Rhizobium*	TAS [19,21-24]
		Species *Rhizobium leguminosarum* bv. *trifolii*	TAS [19,21,24,25]
	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	Non-halophile	NAS
MIGS-22	Oxygen requirement	Aerobic	TAS [11]
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-6	Habitat	Soil, root nodule, on host	TAS [10]
MIGS-15	Biotic relationship	Free living, symbiotic	TAS [10]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [26]
	Isolation	Root nodule	TAS [10]
MIGS-4	Geographic location	NSW, Australia	TAS [10]
MIGS-5	Soil collection date	Dec, 1998	IDA
MIGS-4.1	Longitude	148.25	IDA
MIGS-4.2	Latitude	-35.32	IDA
MIGS-4.3	Depth	0-10cm	
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); 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 [27].
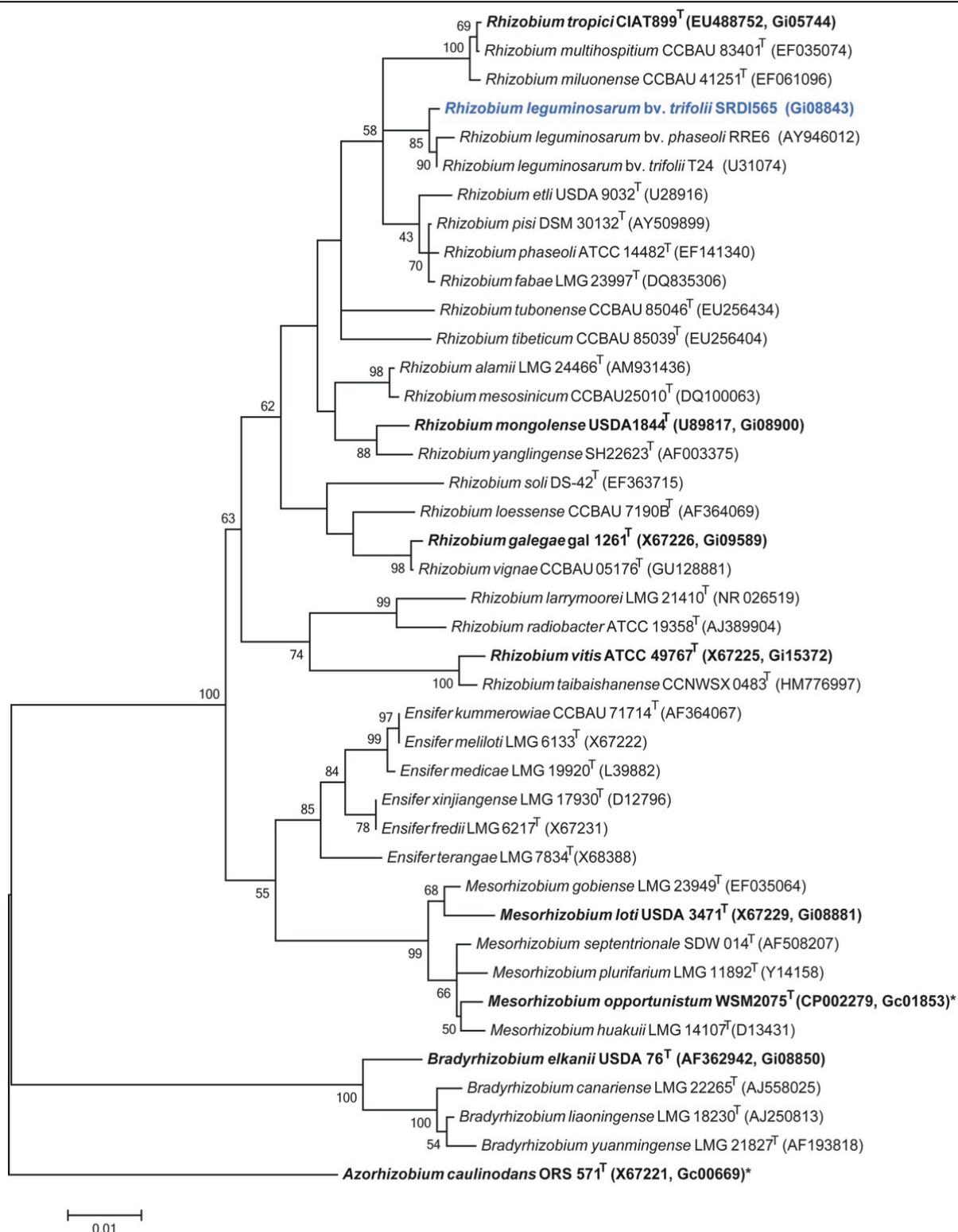
Figure 2. Phylogenetic tree showing the relationship of *Rhizobium leguminosarum* bv. *trifolii* SRDI565 (shown in blue print) with some of the root nodule bacteria in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1,307 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [28]. The tree was built using the maximum likelihood method with the General Time Reversible model. Bootstrap analysis [29] 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 [30] are in bold print and the GOLD ID is shown after the accession number. Published genomes are indicated with an asterisk.

Table 2. Compatibility of SRDI565 with eleven *Trifolium* genotypes for nodulation (Nod) and N_2-Fixation (Fix)

Species name	Cultivar	Common Name	Growth Type	Nod	Fix	Reference
T. glanduliferum Boiss.	Prima	Gland	Annual	+(w)	-	
T. michelianum Savi.	Bolta	Balansa	Annual	+	+	
T. purpureum Loisel	Paratta	Purple	Annual	+	+	[11]
T. resupinatum L.	Kyambro	Persian	Annual	+	+	
T. subterraneum L.	Campeda	Sub. clover	Annual	+	+	[10,11]
T. subterraneum L.	Clare	Sub. clover	Annual	+(w)	-	[10,11]
T. vesiculosum Savi.	Arrotas	Arrowleaf	Annual	-	-	
T. fragiferum L.	Palestine	Strawberry	Perennial	+	+	
T. polymorphum Poir	Acc.#087102	Polymorphous	Perennial	+(w)	-	[11]
T. pratense L.	-	Red	Perennial	+(w)	-	
T. repens L.	Haifa	White	Perennial	+	+	

(w) indicates nodules present were white.

Genome sequencing and assembly

The genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 was sequenced at the Joint Genome Institute (JGI) using Illumina [33] data. An Illumina short-insert paired-end library with an average insert size of 243 ± 58 bp was used to generate 18,700,764 reads and an Illumina long-insert paired-end library with an average insert size of 8,446 ± 2,550 bp was used to generate 21,538,802 reads totalling 6,036 Mbp of Illumina data (unpublished, Feng Chen).

All general aspects of library construction and sequencing performed at the JGI can be found at the JGI user homepage [34]. The initial draft assembly contained 22 contigs in 16 scaffolds. The initial draft data was assembled with Allpaths, version 39750, and the consensus was computationally shredded into 10 Kb overlapping fake reads (shreds). The Illumina draft data was also assembled with Velvet, version 1.1.05 [35], and the consensus sequences were computationally shredded into 1.5 Kb overlapping fake reads (shreds). The

Illumina draft data was assembled again with Velvet using the shreds from the first Velvet assembly to guide the next assembly. The consensus from the second VELVET assembly was shredded into 1.5 Kb overlapping fake reads. The fake reads from the Allpaths assembly and both Velvet assemblies and a subset of the Illumina CLIP paired-end reads were assembled using parallel phrap, version 4.24 (High Performance Software, LLC). Possible mis-assemblies were corrected with manual editing in Consed [36-38]. Gap closure was accomplished using repeat resolution software (Wei Gu, unpublished), and sequencing of bridging PCR fragments with PacBio (unpublished, Cliff Han) technology. For improved high quality draft, 4 PCR PacBio consensus sequences were completed to close gaps and to raise the quality of the final sequence. The estimated total size of the genome is 7 Mb and the final assembly is based on 6,036 Mb of Illumina draft data, which provides an average 862× coverage of the genome.

Genome annotation

Genes were identified using Prodigal [39] as part of the DOE-JGI annotation pipeline [40], followed by a round of manual curation using the JGI GenePRIMP pipeline [41]. 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 [42], RNAMMer [43], Rfam [44], TMHMM [45], and SignalP [46]. Addi-tional gene prediction analyses and functional annotation were performed within the Integrated Microbial Genomes (IMG-ER) platform [47,48].

Genome properties

The genome is 6,905,599 nucleotides with 60.67% GC content (Table 4) and comprised of 7 scaffolds (Figures 3,4,5,6,7,8,and 9) of 7 contigs. From a total of 6,836 genes, 6,750 were protein encoding and 86 RNA-only encoding genes. The majority of genes (77.98%) were assigned a putative function whilst the remaining genes were annotated as hy-pothetical. The distribution of genes into COGs functional categories is presented in Table 5.

Table 3. Genome sequencing project information for *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565.

MIGS ID	Property	Term
MIGS-31	Finishing quality	Improved high-quality draft
MIGS-28	Libraries used	2× Illumina libraries; Std short PE & CLIP long PE
MIGS-29	Sequencing platforms	Illumina HiSeq 2000, PacBio
MIGS-31.2	Sequencing coverage	862× Illumina
MIGS-30	Assemblers	with Allpaths, version 39750, Velvet 1.015, phrap 4.24
MIGS-32	Gene calling methods	Prodigal 1.4, GenePRIMP
	GOLD ID	Gi08843
	NCBI project ID	81743
	Database: IMG	2517287029
	Project relevance	Symbiotic N_2 fixation, agriculture

Table 4. Genome Statistics for *Rhizobium leguminosarum* bv. *trifolii* SRDI565

Attribute	Value	% of Total
Genome size (bp)	6,905,599	100.00
DNA coding region (bp)	5,960,775	86.32
DNA G+C content (bp)	4,189,855	60.67
Number of scaffolds	7	
Number of contigs	7	
Total gene	6,836	100.00
RNA genes	86	1.26
rRNA operons*	3	
Protein-coding genes	6,750	98.74
Genes with function prediction	5,331	77.98
Genes assigned to COGs	5,330	77.97
Genes assigned Pfam domains	5,535	80.97
Genes with signal peptides	603	8.82
Genes with transmembrane helices	1,552	22.70
CRISPR repeats	0	
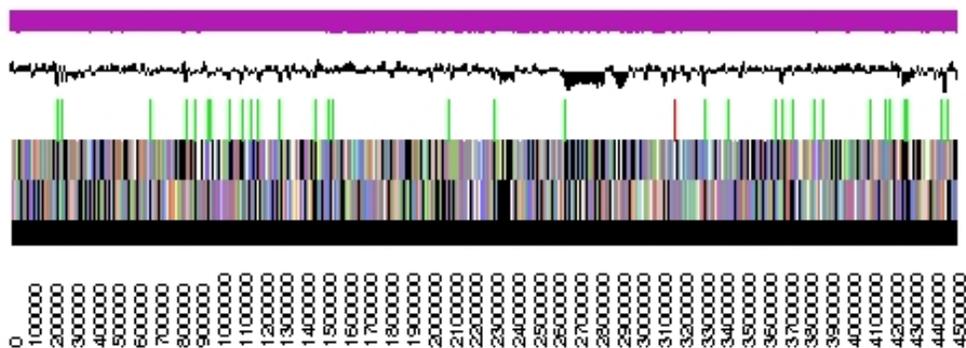
Figure 3. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (scaffold 1.1). From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.
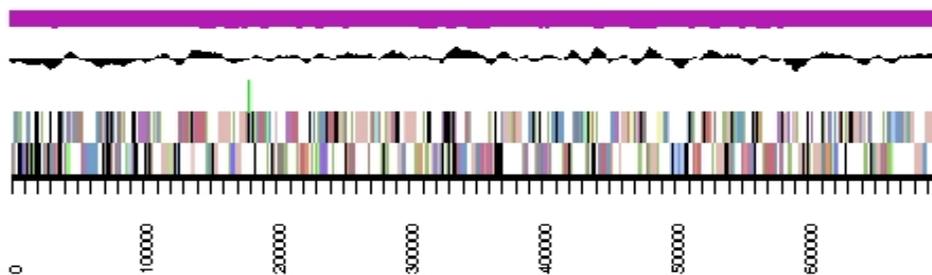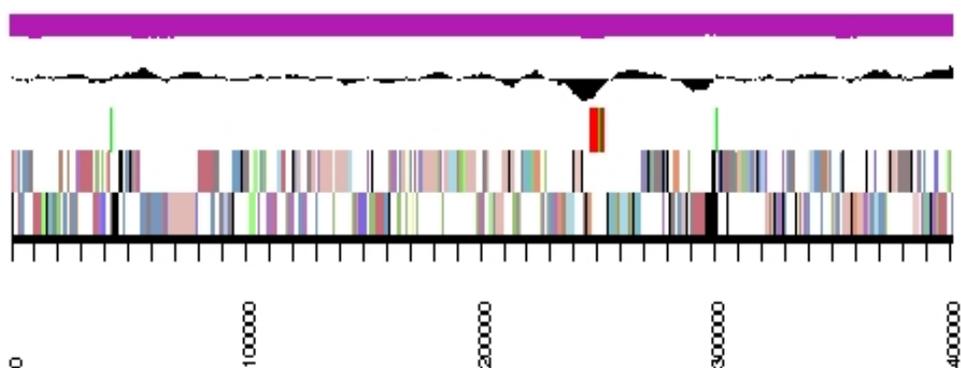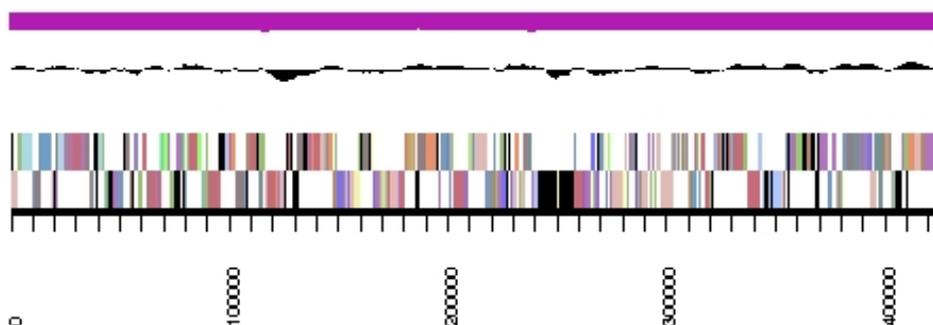

Figure 4. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (scaffold 2.2). From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.
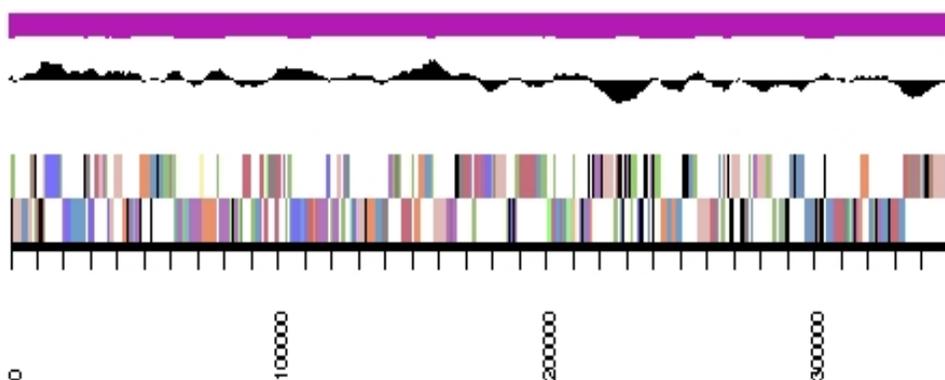
Figure 5. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (scaffold 3.3). From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Figure 6. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (scaffold 4.4). From bottom to the top of each scaffold: 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 7. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (scaffold 5.5). From bottom to the top of each scaffold: 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* SRDI565 associated with the general COG functional categories.

| Code	Value	%age	Description
J	191	3.22	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	574	9.67	Transcription
L	189	3.19	Replication, recombination and repair
B	3	0.05	Chromatin structure and dynamics
D	41	0.69	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	70	1.18	Defense mechanisms
T	320	5.39	Signal transduction mechanisms
M	315	5.31	Cell wall/membrane biogenesis
N	81	1.37	Cell motility
Z	0	0.00	Cytoskeleton
W	0	0.00	Extracellular structures
U	96	1.62	Intracellular trafficking and secretion
O	208	3.51	Posttranslational modification, protein turnover, chaperones
C	326	5.49	Energy production conversion
G	633	10.67	Carbohydrate transport and metabolism
E	591	9.96	Amino acid transport metabolism
F	109	1.84	Nucleotide transport and metabolism
H	193	3.25	Coenzyme transport and metabolism
I	216	3.64	Lipid transport and metabolism
P	272	4.58	Inorganic ion transport and metabolism
Q	148	2.49	Secondary metabolite biosynthesis, transport and catabolism
R	758	12.77	General function prediction only
S	600	10.11	Function unknown
-	1,506	22.03	Not in COGS
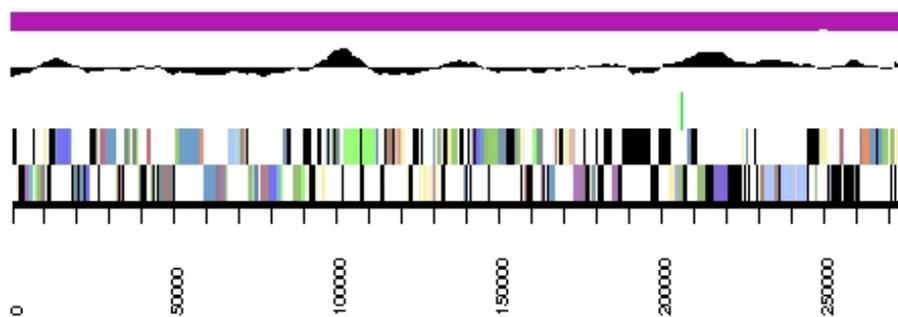
Figure 8. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (6.6). From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.
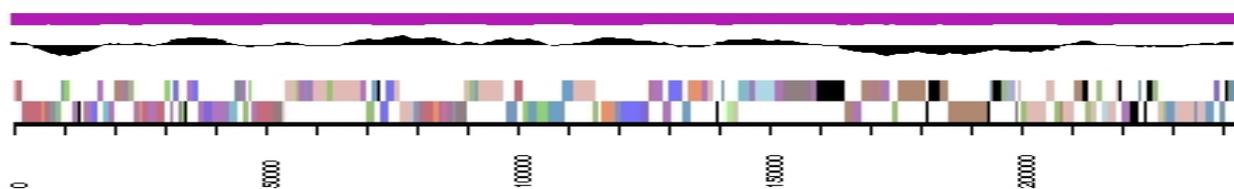
Figure 9. Graphical map of the genome of *Rhizobium leguminosarum* bv. *trifolii* strain SRDI565 (7.7). From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Acknowledgements

References

1. 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

2. Howieson JG, Yates RJ, Foster K, Real D, 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 363-394.

3. 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

4. Unkovich MJ, Baldock J, Peoples MB. Prospects and problems of simple linear models for estimating symbiotic N_2 fixation by crop and pasture legumes. *Plant Soil* 2010; **329**:75-89. http://dx.doi.org/10.1007/s11104-009-0136-5

5. Ballard RA, Shepherd BR, Charman N. Nodulation and growth of pasture legumes with naturalised soil rhizobia. 3. Lucerne (*Medicago sativa* L.). *Aust J Exp Agric* 2003; **43**:135-140. http://dx.doi.org/10.1071/EA02047

6. Denton MD, Coventry DR, Bellotti WD, Howieson JG. Distribution, abundance and symbiotic effectiveness of *Rhizobium leguminosarum* bv. *trifolii* from alkaline pasture soils in South Australia. *Anim Prod Sci* 2000; **40**:25-35. http://dx.doi.org/10.1071/EA99035

7. Drew EA, Charman N, Dingemanse R, Hall E, Ballard RA. Symbiotic performance of Mediterranean *Trifolium* spp. with naturalised soil rhizobia. *Crop Pasture Sci* 2011; **62**:903-913. http://dx.doi.org/10.1071/CP11047

8. Rys GJ, Bonish PM. Effectiveness of *Rhizobium trifolii* populations associated with *Trifolium* species in Taranaki, New Zealand. *New Zealand Journal of Experimental Agriculture* 1981; **9**:329-335. http://dx.doi.org/10.1080/03015521.1981.10425430

9. Slattery JF, Coventry DR. Acid-tolerance and symbiotic effectiveness of *Rhizobium leguminosarum* bv. *trifolii* isolated from subterranean clover growing in permanent pastures. *Soil Biol Biochem* 1995; **27**:111-115. http://dx.doi.org/10.1016/0038-0717(94)00143-O

10. Drew EA, Ballard RA. Improving N_2 fixation from the plant down: Compatibility of *Trifolium subterraneum* L. cultivars with soil rhizobia can influence symbiotic performance. *Plant Soil* 2010; **327**:261-277. http://dx.doi.org/10.1007/s11104-009-0052-8

11. 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 (Lond)* 2012; **110**:1559-1572. PubMed http://dx.doi.org/10.1093/aob/mcs206

12. 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

13. 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

14. 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

15. 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.

16. Garrity GM, Bell JA, Lilburn T. Class I. *Alphaproteobacteria* class. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. Second ed: New York: Springer - Verlag; 2005.

17. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. Bergey's Manual of Systematic Bacteriology. Second ed: New York: Springer - Verlag; 2005. p 324.

18. 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

19. 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

20. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.

21. Frank B. Über die Pilzsymbiose der Leguminosen. *Ber Dtsch Bot Ges* 1889; **7**:332-346.

22. 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.

23. 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

24. 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

25. 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

26. Agents B. Technical rules for biological agents. TRBA (http://www.baua.de):466.

27. 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

28. 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

29. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. http://dx.doi.org/10.2307/2408678

30. 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

31. 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

32. General Information for Collaborators. http://my.jgi.doe.gov/general/index.html

33. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. PubMed http://dx.doi.org/10.1517/14622416.5.4.433

34. http://my.jgi.doe.gov/general/index.html

35. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. Current Protocols in Bioinformatics 2010;Chapter 11:Unit 11 5.

36. Ewing B, Green P. Base-calling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* 1998; **8**:175-185. PubMed http://dx.doi.org/10.1101/gr.8.3.175

37. 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

38. 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

39. 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

40. 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

41. 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

42. 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

43. 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

44. 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

45. 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

46. Dyrløv 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

47. 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

48. DOE Joint Genome Institute. http://img.jgi.doe.gov/er