

## UC Merced

### UC Merced Previously Published Works

#### Title

High-quality permanent draft genome sequence of *Rhizobium sullae* strain WSM1592; a *Hedysarum coronarium* microsymbiont from Sassari, Italy

#### Permalink

<https://escholarship.org/uc/item/8dx8p3fn>

#### Journal

Environmental Microbiome, 10(1)

#### ISSN

2524-6372

#### Authors

Yates, Ron

Howieson, John

De Meyer, Sofie E

et al.

#### Publication Date

2015

#### DOI

10.1186/s40793-015-0020-2

Peer reviewed



SHORT GENOME REPORT

Open Access



# High-quality permanent draft genome sequence of *Rhizobium sulae* strain WSM1592; a *Hedysarum coronarium* microsymbiont from Sassari, Italy

Ron Yates<sup>1,2</sup>, John Howieson<sup>2</sup>, Sofie E. De Meyer<sup>2</sup>, Rui Tian<sup>2</sup>, Rekha Seshadri<sup>3</sup>, Amrita Pati<sup>3</sup>, Tanja Woyke<sup>3</sup>, Victor Markowitz<sup>4</sup>, Natalia Ivanova<sup>3</sup>, Nikos Kyrpides<sup>3,5</sup>, Angelo Loi<sup>1</sup>, Brad Nutt<sup>1</sup>, Giovanni Garau<sup>6</sup>, Leonardo Sulas<sup>7</sup> and Wayne Reeve<sup>2\*</sup>

## Abstract

*Rhizobium sulae* strain WSM1592 is an aerobic, Gram-negative, non-spore-forming rod that was isolated from an effective nitrogen (N<sub>2</sub>) fixing root nodule formed on the short-lived perennial legume *Hedysarum coronarium* (also known as *Sulla coronaria* or *Sulla*). WSM1592 was isolated from a nodule recovered from *H. coronarium* roots located in Ottava, bordering Sassari, Sardinia in 1995. WSM1592 is highly effective at fixing nitrogen with *H. coronarium*, and is currently the commercial *Sulla* inoculant strain in Australia. Here we describe the features of *R. sulae* strain WSM1592, together with genome sequence information and its annotation. The 7,530,820 bp high-quality permanent draft genome is arranged into 118 scaffolds of 118 contigs containing 7,453 protein-coding genes and 73 RNA-only encoding genes. This rhizobial genome is sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project.

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

## Introduction

The accessibility and supply of nitrogen fertilizer is an ever-increasing challenge that world agriculture faces [1]. Despite the fact the Earth's atmosphere consists of approximately 78 % dinitrogen, it is in a form that must be converted before it can be utilised by plants [2]. Conversion of N<sub>2</sub> can be achieved by the chemical synthesis of natural gas but these methods can be considered unsustainable because of the use of exhaustible and costly fossil fuel resources [3]. In addition, the manufacturing process not only increases the greenhouse gas emissions but also field N fertiliser application have been directly linked to contaminating and leading to detrimental effects in ecosystems and waterways. Alternatively, a more sustainable and environmentally friendly process of acquiring N is through the biological process of N fixation by diazotrophs [2]. Most biological fixation in world agriculture is provided from the process of symbiotic nitrogen fixation,

which occurs following the successful formation of an effective symbiosis by leguminous plants and bacterial microsymbionts [4].

The productive efficiencies of SNF in agricultural areas rely on considerable efforts by researchers and producers in matching suitable legume hosts with their compatible microsymbionts [5]. Some agricultural areas farm with indigenous legumes, while others embark on introducing exotic legumes and their compatible microsymbionts from different geographical locations that are edaphically and climatically suited to their own [4]. In Australia for instance, selection programs have enabled the domestication of new Mediterranean legume species and their microsymbionts [6]. One such grazing legume species commercially introduced into Australian and New Zealand agriculture includes the Papilionoid legume *Hedysarum coronarium* (also known as *Sulla coronaria* or *Sulla*). *Sulla* is a deep-rooted, short-lived perennial pasture legume that is grown throughout Mediterranean countries where it is fed green, used for silage or as hay [7]. It is noted that the microsymbionts of *Sulla* display a high level of specificity for nodulation and nitrogen fixation [8]. However,

\* Correspondence: W.Reeve@murdoch.edu.au

<sup>2</sup>Centre for Rhizobium Studies, Murdoch University, Murdoch, Western Australia, Australia

Full list of author information is available at the end of the article

when effectively nodulated *Sulla* plants have the ability to biologically fix large amounts of nitrogen for increased paddock fertility [9].

*Rhizobium sullae* strain WSM1592 is the current Australian commercial inoculant for *Sulla* after replacing strain CC1335 in 2006. This strain has also been deposited in the Western Australian Soil Microbiology collection and is available for research. WSM1592 was isolated in 1995 from a nodule collected from a *Sulla* plant sampled on a roadside in calcareous loamy sand near the Ottava agriculture research farm, east of Sassari in Sardinia, Italy. The location has a Mediterranean climate with a long-term mean seasonal rainfall of 547 mm. Here we present a preliminary description of the general features for *Rhizobium sullae* strain WSM1592 together with its genome sequence and annotation.

## Organism information

### Classification and features

*R. sullae* strain WSM1592 is a motile, Gram-negative rod (Fig. 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) [10], tryptone-yeast extract agar (TY) [11] or a modified yeast-mannitol agar (YMA) [12] at 28 °C. Colonies on ½LA are white-opaque, slightly domed and moderately mucoid with smooth margins (Fig. 1 Right).

Figure 2 shows the phylogenetic relationship of *R. sullae* strain WSM1592 in a 16S rRNA gene sequence based tree. This strain is phylogenetically the most related to *Rhizobium sullae* IS 123<sup>T</sup>, *Rhizobium leguminosarum* USDA 2370<sup>T</sup> and *Rhizobium phaseoli* ATCC 14482<sup>T</sup> with sequence identities to the WSM1592 16S rRNA gene sequence of 100 %, 99.84 % and 99.84 %, respectively, as determined using the EzTaxon-e server [13]. *Rhizobium sullae* IS 123<sup>T</sup> was isolated from a *Hedysarum coronarium* root nodule discovered in Southern Spain [14]. In contrast,

*R. leguminosarum* USDA 2370<sup>T</sup> was isolated from an effective nodule of *Pisum sativum* and is also able to nodulate *Trifolium repens* and *Phaseolus vulgaris* [15]. *R. phaseoli* ATCC 14482<sup>T</sup> was originally isolated from nodules of *Phaseolus vulgaris* and has been shown to nodulate *Trifolium repens*, but not *Pisum sativum* [15].

Minimum Information about the Genome Sequence [16] of WSM1592 is provided in Table 1 and Additional file 1: Table S1.

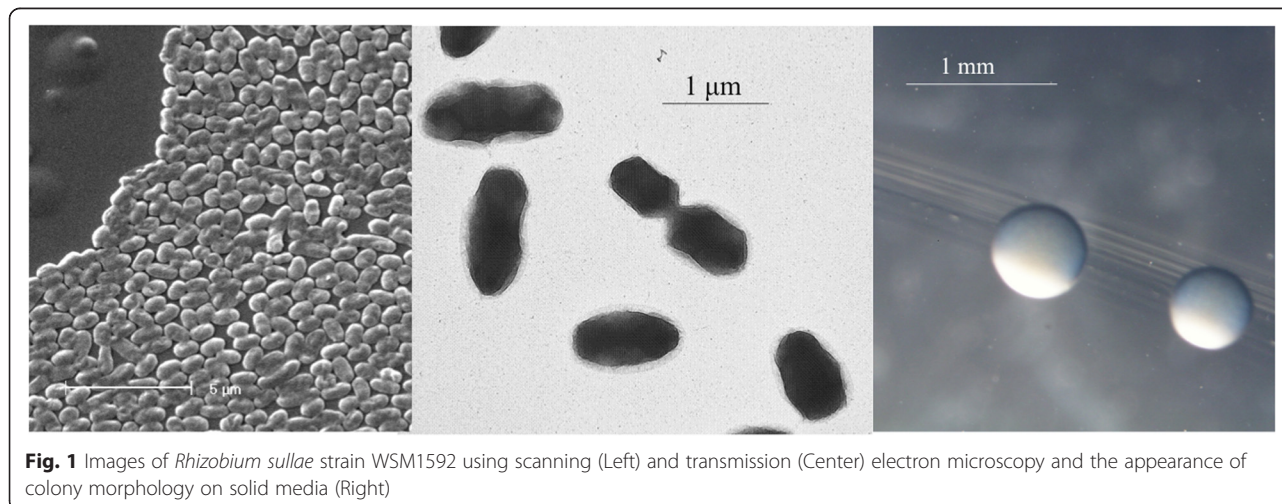
### Symbiotaxonomy

*Hedysarum coronarium* is a short-lived perennial pasture legume native to the Mediterranean basin and throughout the *Hedysarum* genus there is a large degree of specificity in symbiotic compatibility within this region [8]. *Rhizobium sullae* WSM1592 nodulates (Nod+) and fixes nitrogen effectively (Fix+) with *Hedysarum coronarium*. However, inoculation of *H. spinosissimum*, *H. flexuosum* and *H. carnosum* with WSM1592 results in mostly Nod- but always Fix-.

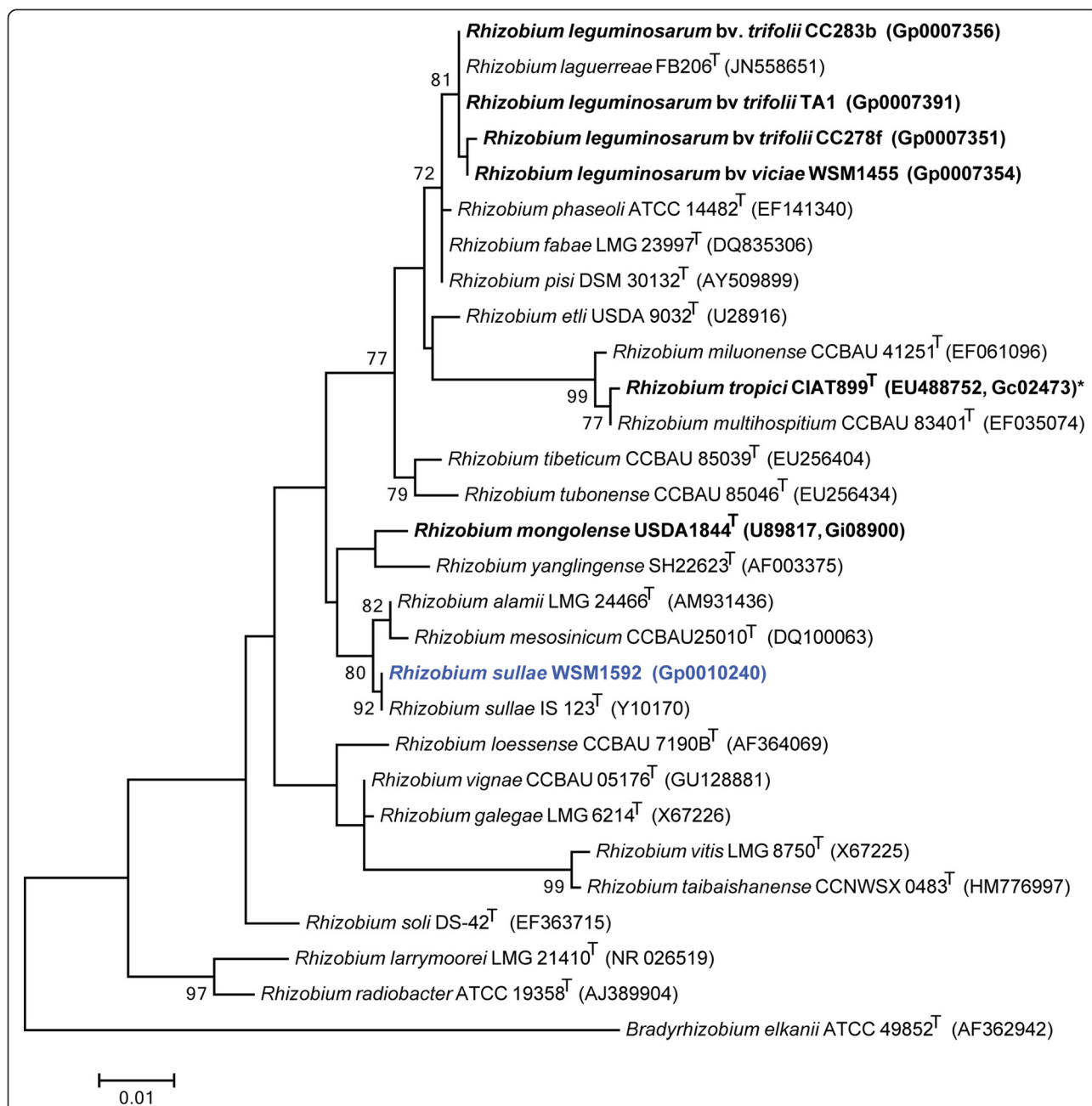
## Genome sequencing information

### Genome project history

This organism was selected for sequencing on the basis of its environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance, and is part of the *Genomic Encyclopedia of Bacteria and Archaea*, The Root Nodulating Bacteria chapter project at the U.S. Department of Energy, Joint Genome Institute [17]. The genome project is deposited in the Genomes OnLine Database [18] and the high-quality permanent draft genome sequence in IMG [19]. Sequencing, finishing and annotation were performed by the JGI using state of the art sequencing technology [20]. A summary of the project information is shown in Table 2.



**Fig. 1** Images of *Rhizobium sullae* strain WSM1592 using scanning (Left) and transmission (Center) electron microscopy and the appearance of colony morphology on solid media (Right)



**Fig. 2** Phylogenetic tree highlighting the position of *R. sullae* strain WSM1592 (shown in blue print) relative to other type and non-type rhizobia strains using a 901 bp internal region of the 16S rRNA gene. *Bradyrhizobium elkanii* ATCC 49852<sup>T</sup> was used as an outgroup. All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [33]. The tree was built using the maximum likelihood method with the General Time Reversible model. Bootstrap analysis with 500 replicates was performed to assess the support of the clusters. Type strains are indicated with a superscript T. Strains with a genome sequencing project registered in GOLD [18] have the GOLD ID mentioned after the strain number and represented in bold, otherwise the NCBI accession number is provided. Finished genomes are designated with an asterisk

#### Growth conditions and genomic DNA preparation

*R. sullae* WSM1592 was cultured to mid logarithmic phase in 60 ml of TY rich media [11] 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 [21].

#### Genome sequencing and assembly

The draft genome of *R. sullae* strain WSM1592 was generated at the DOE Joint Genome Institute using state of the art technology [20]. An Illumina Std shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 29,255,624 reads. All general

**Table 1** Classification and general features of *R. sultae* strain WSM1592 [16, 34]

MIGS ID	Property	Term	Evidence code
	Classification	Domain Bacteria	TAS [35]
		Phylum <i>Proteobacteria</i>	TAS [36, 37]
		Class <i>Alphaproteobacteria</i>	TAS [38]
		Order <i>Rhizobiales</i>	TAS [39]
		Family <i>Rhizobiaceae</i>	TAS [40]
		Genus <i>Rhizobium</i>	TAS [41]
		Species <i>Rhizobium sultae</i>	TAS [14]
		(Type) strain WSM1592	IDA
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Not reported	
	Optimum temperature	28 °C	NAS
	pH range; Optimum	Not reported	
	Carbon source	Not reported	
MIGS-6	Habitat	Soil, root nodule, on host	IDA
MIGS-6.3	Salinity	Non-halophile	NAS
MIGS-22	Oxygen requirement	Aerobic	IDA
MIGS-15	Biotic relationship	Free living, symbiotic	IDA
MIGS-14	Pathogenicity	Non-pathogenic	NAS
MIGS-4	Geographic location	Sassari, Italy	IDA
MIGS-5	Soil collection date	20 May 1995	IDA
MIGS-4.1	Latitude	8.465	IDA
MIGS-4.2	Longitude	40.777	IDA
MIGS-4.4	Altitude	83 m	IDA

Evidence codes – IDA Inferred from Direct Assay; TAS Traceable Author Statement (i.e., a direct report exists in the literature); NAS Non-traceable Author Statement (i.e., not directly observed for the living, isolated sample, but based on a generally accepted property for the species, or anecdotal evidence). These evidence codes are from the Gene Ontology project [42]

aspects of library construction and sequencing performed at the JGI can be found at the JGI's web site [22]. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artifacts (Mingkun L, Copeland A, Han J. unpublished). Following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet (version 1.1.04) [23] (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [24] (3) Illumina reads were assembled with simulated read pairs using Allpaths-LG (version r39750) [25].

**Table 2** Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality permanent draft
MIGS-28	Libraries used	Illumina Std PE (2x150bps)
MIGS-29	Sequencing platforms	Illumina HiSeq 2000
MIGS-31.2	Fold coverage	877x
MIGS-30	Assemblers	Velvet 1.1.04; Allpaths-LG r39750
MIGS-32	Gene calling methods	Prodigal 1.4
	Locus Tag	A3C1
	Genbank ID	ATZB00000000
	Genbank Date of Release	December 12, 2013
	GOLD ID	Gp0010240
	BIOPROJECT	PRJNA165333
MIGS-13	Source Material Identifier	WSM1592
	Project relevance	Symbiotic N <sub>2</sub> fixation, agriculture

Parameters for assembly steps were: 1) Velvet (velveth: 63 –shortPaired and velvetg: –very\_clean yes –export-Filtered yes –min\_contig\_lgth 500 –scaffolding no–cov\_cutoff 10) 2) wgsim (–e 0 –l 76 –r 0 –R 0 –X 0) 3) Allpaths-LG (PrepareAllpathsInputs: PHRED\_64 = 1 PLOIDY = 1 FRAG\_COVERAGE = 125 JUMP\_COVERAGE = 25 LONG\_JUMP\_COV = 50, RunAllpathsLG: THREADS = 8 RUN = std\_shredpairs TARGETS = standard VAPI\_WARN\_ONLY = True OVERWRITE = True). The final draft assembly contained 118 contigs in 118 scaffolds. The total size of the genome is 7.5 Mbp and the final assembly is based on 2,498,075,850 bp of Illumina data, which provides an average of 877× coverage of the genome.

**Table 3** Genome statistics

Attribute	Value	% of Total
Genome size (bp)	7,530,820	100.00
DNA coding (bp)	6,571,312	87.26
DNA G + C (bp)	4,508,646	59.87
DNA scaffolds	118	
Total genes	7,526	100.00
Protein coding genes	7,453	99.03
RNA genes	73	0.97
Pseudo genes	0	0
Genes in internal clusters	538	7.15
Genes with function prediction	5,902	78.42
Genes assigned to COGs	5,148	68.40
Genes with Pfam domains	6,174	82.04
Genes with signal peptides	659	8.76
Genes with transmembrane helices	1,699	22.58
CRISPR repeats	0	0



### Genome annotation

Genes were identified using Prodigal [26], as part of the DOE-JGI genome annotation pipeline [27, 28]. The predicted CDSs were translated and used to search the National Centre for Biotechnology Information non-redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool [29] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [30]. Other non-coding RNAs such as the RNA components of the protein secretion complex and the RNase P were identified by searching the genome for the corresponding Rfam profiles using INFERNAL [31]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes-Expert Review system [32] developed by the Joint Genome Institute, Walnut Creek, CA, USA.

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

Code	Value	% age	COG Category
J	186	3.25	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	554	9.67	Transcription
L	158	2.76	Replication, recombination and repair
B	2	0.03	Chromatin structure and dynamics
D	36	0.63	Cell cycle control, Cell division, chromosome partitioning
V	65	1.13	Defense mechanisms
T	211	3.68	Signal transduction mechanisms
M	293	5.11	Cell wall/membrane/envelope biogenesis
N	68	1.19	Cell motility
U	112	1.95	Intracellular trafficking, secretion, and vesicular transport
O	167	2.91	Posttranslational modification, protein turnover, chaperones
C	327	5.71	Energy production and conversion
G	614	10.71	Carbohydrate transport and metabolism
E	684	11.94	Amino acid transport and metabolism
F	107	1.87	Nucleotide transport and metabolism
H	174	3.04	Coenzyme transport and metabolism
I	196	3.42	Lipid transport and metabolism
P	318	5.55	Inorganic ion transport and metabolism
Q	138	2.41	Secondary metabolite biosynthesis, transport and catabolism
R	743	12.96	General function prediction only
S	578	10.09	Function unknown
-	2378	31.60	Not in COGS

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

### Genome properties

The genome is 7,530,820 nucleotides 59.87 % GC content (Table 3 and comprised of 118 scaffolds of 118 contigs). From a total of 7,526 genes, 7,453 were protein encoding and 73 RNA only encoding genes. The majority of genes (78.42 %) were assigned a putative function whilst the remaining genes were annotated as hypothetical. The distribution of genes into COG functional categories is presented in Table 4.

### Conclusions

*Rhizobium sllae* WSM1592 was isolated from a root nodule of *Hedysarum coronarium* (also known as *Sulla coronaria*). Phylogenetic analysis revealed that WSM1592 is the most closely related to *Hedysarum coronarium* IS 123<sup>T</sup>, which was also isolated from *Hedysarum coronarium* growing in Southern Spain. The genome of WSM1592 is the first to be described for a strain of *Rhizobium sllae* and is 7.5 Mbp, with a GC content of 59.87 %. As expected this genome contains the nitrogenase-RXN MetaCyc pathway characterized by the multiprotein nitrogenase complex and has been shown to fix effectively with *Hedysarum coronarium*. The genome attributes of WSM1592 will be important for the characterisation of the genetic determinants required for the establishment of an effective symbiosis with *Hedysarum*.

### Additional file

**Additional file 1: Table S1.** Associated MIGS record for WSM1592.

### Abbreviations

GEBA-RNB: *Genomic Encyclopedia of Bacteria and Archaea – Root Nodule Bacteria*; JGI: Joint Genome Institute; TY: Trypton Yeast; CTAB: Cetyl trimethyl ammonium bromide; WSM: Western Australian Soil Microbiology; SNF: Symbiotic Nitrogen Fixation.

### Competing interests

The authors declare that they have no competing interests.

### Authors' contributions

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

### 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.

### Author details

<sup>1</sup>Department of Agriculture and Food, Western Australia, Australia. <sup>2</sup>Centre for Rhizobium Studies, Murdoch University, Murdoch, Western Australia, Australia. <sup>3</sup>DOE Joint Genome Institute, Walnut, Creek, CA, USA. <sup>4</sup>Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>5</sup>Department of Biological Sciences, Faculty of

Science, King Abdulaziz University, Jeddah, Saudi Arabia. <sup>6</sup>Department of Agriculture, University of Sassari, Sardinia, Italy. <sup>7</sup>Institute for the Animal Production System in the Mediterranean Environment (ISPAAM), National Research Council (CNR), Sassari, Italy.

Received: 9 December 2014 Accepted: 19 May 2015

Published online: 24 July 2015

## References

- Herridge DF, Peoples MB, Boddey RM. Global inputs of biological nitrogen fixation in agricultural systems. *Plant Soil*. 2008;311:1–18.
- Sprent JI. Legume nodulation: A global perspective. Oxford, U.K: Wiley-Blackwell; 2009.
- Vance CP. Symbiotic nitrogen fixation and phosphorus acquisition. Plant nutrition in a world of declining renewable resources. *Plant Physiol*. 2001;127:390–7.
- 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. 363–94.
- Sessitsch A, Howieson JG, Perret X, Antoun H, Martinez-Romero E. Advances in *Rhizobium* research. *Crit Rev Plant Sci*. 2002;21:323–78.
- Howieson JG, Loi A, Carr SJ. *Biserrula pelecinus* L. - a legume pasture species with potential for acid, duplex soils which is nodulated by unique root-nodule bacteria. *Aust J Agric Res*. 1995;46:997–1009.
- Footo A. Local cultivar adaptation of Mediterranean Sulla. *N Z J Exp Agric*. 1988;153:25–7.
- Kishinevsky BD, Nandasena KG, Yates RJ, Nemas C, Howieson JG. Phenotypic and genetic diversity among rhizobia isolated from three *Hedysarum* species: *H. spinosissimum*, *H. coronarium* and *H. flexuosum*. *Plant Soil*. 2003;251:143–53.
- Sulas L, Seddaiu G, Muresu R, Roggero PP. Nitrogen fixation of Sulla under mediterranean conditions. *Agron J*. 2009;101:1470–8.
- Howieson JG, Ewing MA, D'antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. *Plant Soil*. 1988;105:179–88.
- Beringer JE. R factor transfer in *Rhizobium leguminosarum*. *J Gen Microbiol*. 1974;84:188–98.
- Terpolilli JJ. Why are the symbioses between some genotypes of *Sinorhizobium* and *Medicago* suboptimal for N<sub>2</sub> fixation? PhD thesis. Murdoch University: Microbiology; 2009.
- Kim O-S, Cho Y-J, Lee K, Yoon S-H, Kim M, Na H, et al. Introducing EzTaxon-e: a prokaryotic 16S rRNA gene sequence database with phylotypes that represent uncultured species. *Int J Syst Evol Microbiol*. 2012;62:716–21.
- Squartini A, Struffi P, Döring H, Selenska-Pobell S, Tola E, Giacomini A, et al. *Rhizobium sullae* sp. nov. (formerly *Rhizobium hedysari*), the root-nodule microsymbiont of *Hedysarum coronarium*. *Int J Syst Evol Microbiol*. 2002;52:1267–76.
- Ramirez-Bahena MH, Garcia-Fraile P, Peix A, Valverde A, Rivas R, Igual JM, et al. Revision of the taxonomic status of the species *Rhizobium leguminosarum* (Frank 1879) Frank 1889(AL), *Rhizobium phaseoli* Dangeard 1926(AL) and *Rhizobium trifolii* Dangeard 1926(AL). *R-trifolii* is a later synonym of *R-leguminosarum*. Reclassification of the strain *R-leguminosarum* DSM 30132 (= NCIMB 11478) as *Rhizobium pisi* sp nov. *Int J Syst Evol Microbiol*. 2008;58:2484–90.
- Field D, Garrity G, Gray T, Morrison N, Selengut J, Sterk P, et al. Towards a richer description of our complete collection of genomes and metagenomes "Minimum Information about a Genome Sequence" (MIGS) specification. *Nat Biotechnol*. 2008;26:541–7.
- Chen IM, Markowitz VM, Chu K, Anderson I, Mavromatis K, Kyrpides NC, et al. Improving microbial genome annotations in an integrated database context. *PLoS One*. 2013;8:e54859.
- 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;1:63–7.
- Lowe TM, Eddy SR. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res*. 1997;25:955–64.
- Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, et al. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res*. 2007;35:7188–96.
- INFERNAL. Inference of RNA alignments [http://infernal.janelia.org]
- 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;25:2271–8.
- 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;28:2731–9.
- Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, et al. The genomic standards consortium. *PLoS Biol*. 2011;9:e1001088.
- Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *Proc Natl Acad Sci U S A*. 1990;87:4576–9.
- Chen WX, Wang ET, Kuykendall LD. The *Proteobacteria*. New York: Springer-Verlag; 2005.
- Validation of publication of new names and new combinations previously effectively published outside the IJSEM. *Int J Syst Evol Microbiol*. 2005; 55:2235–2238.
- Garrity GM, Bell JA, Lilburn T, Class I. *Alphaproteobacteria*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's manual of systematic bacteriology*. Secondth ed. New York: Springer - Verlag; 2005.
- Kuykendall LD: Order VI. *Rhizobiales*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's Manual of Systematic Bacteriology*. Secondth ed. New York: Springer - Verlag; 2005. p. 324.
- Kuykendall LD: Family I. *Rhizobiaceae*. In: Brenner DJ, Krieg NR, Staley JT, editors. *Bergey's manual of systematic bacteriology*, vol. 2. 2nd ed. New York: Springer; 2005. p. 324–58. part C.
- 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*, vol. 2. Second ed. New York: Springer - Verlag; 2005.
- Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat Genet*. 2000;25:25–9.
- 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.
- Reeve W, Ardley J, Tian R, Eshraghi L, Yoon J, Ngamwisetkun P, et al. A genomic encyclopedia of the root nodule bacteria: Assessing genetic diversity through a systematic biogeographic survey. *Stand Genomic Sci*. 2015;10:14.
- Reddy TB, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res*. 2014.
- Markowitz VM, Chen I-MA, Palaniappan K, Chu K, Szeto E, Pillay M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. *Nucleic Acids Res*. 2014;42:D560–7.
- Mavromatis K, Land ML, Brettin TS, Quest DJ, Copeland A, Clum A, et al. The fast changing landscape of sequencing technologies and their impact on microbial genome assemblies and annotation. *PLoS One*. 2012;7:e48837.
- CTAB DNA extraction protocol [http://jgi.doe.gov/collaborate-with-jgi/pmo-overview/protocols-sample-preparation-information/]
- JGI Website [http://www.jgi.doe.gov]
- Zerbino D, Birney E. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res*. 2008;18:821–9.
- wgsm [https://github.com/lh3/wgsm]
- Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci U S A*. 2011;108:1513–8.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

