

Genome sequence of *Ensifer medicae* strain WSM1369; an effective microsymbiont of the annual legume *Medicago sphaerocarpos*

Jason Terpolilli¹, Giovanni Garau², Yvette Hill¹, Rui Tian¹, John Howieson¹, Lambert Bräu³, Lynne Goodwin⁴, James Han⁵, Konstantinos Liolios⁵, Marcel Huntemann⁵, Amrita Pati⁵, Tanja Woyke⁵, Konstantinos Mavromatis⁶, Victor Markowitz⁶, Natalia Ivanova⁵, Nikos Kyrpides⁵, Wayne Reeve^{1*}.

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

² Dipartimento di Agraria, S.T.A.A., University of Sassari, Italy

³ School of Life and Environmental Sciences, Deakin University, 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)

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

Ensifer medicae WSM1369 is an aerobic, motile, Gram-negative, non-spore-forming rod that can exist as a soil saprophyte or as a legume microsymbiont of *Medicago*. WSM1369 was isolated in 1993 from a nodule recovered from the roots of *Medicago sphaerocarpos* growing at San Pietro di Rudas, near Aggius in Sardinia (Italy). WSM1369 is an effective microsymbiont of the annual forage legumes *M. polymorpha* and *M. sphaerocarpos*. Here we describe the features of *E. medicae* WSM1369, together with genome sequence information and its annotation. The 6,402,557 bp standard draft genome is arranged into 307 scaffolds of 307 contigs containing 6,656 protein-coding genes and 79 RNA-only encoding genes. This rhizobial genome is one of 100 sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project.

Introduction

One of the key nutritional constraints to plant growth and development is the availability of nitrogen (N) in nutrient deprived soils [1]. Although the atmosphere consists of approximately 80% N, the overwhelming proportion of this is present in the form of dinitrogen (N₂) which is biologically inaccessible to most plants and other higher organisms. Before the development of the Haber-Bosch process, the primary mechanism for converting atmospheric N₂ into a bioaccessible form was via biological nitrogen fixation (BNF) [2]. In BNF, N₂ is made available by specialized microbes that possess the necessary molecular machinery to reduce N₂ into NH₃. Some plants, most of which are legumes, have harnessed BNF by evolving

symbiotic relationships with specific N₂-fixing microbes (termed rhizobia) whereby the host plant houses the bacteria in root nodules, supplying the microsymbiont with carbon and in return receives essential reduced N-containing products [3]. When BNF is exploited in agriculture, some of this N₂ fixed into plant tissues is ultimately released into the soil following harvest or senescence, where it can then be assimilated by subsequent crops. Compared to industrially synthesized N-based fertilizers, BNF is a low energy, low cost and low greenhouse-gas producing alternative and hence its application is crucial to increasing the environmental and economic sustainability of farming systems [4].

Forage and fodder legumes play vital roles in sustainable farming practice, with approximately 110 million ha under production worldwide [5], a significant proportion of which is made up by members of the genus *Medicago*. *Ensifer meliloti* and *E. medicae* are known to nodulate and fix N₂ with *Medicago* spp [6], although they have differences in host specificity. While *E. meliloti* strains do not nodulate *M. murex*, nodulate but do not fix N₂ with *M. polymorpha* and nodulate but fix very poorly with *M. arabica* [7,8], they are able to nodulate and fix N₂ with *Medicago* species originating from alkaline soils including the perennial *M. sativa* and the annuals *M. littoralis* and *M. tomata* [9,10]. In contrast, *E. medicae* strains can nodulate and fix N₂ with annuals well adapted to acidic soils, such as *M. murex*, *M. arabica* and *M. polymorpha* [7,8].

The *E. medicae* strain WSM1369 was isolated from a nodule collected from *M. sphaerocarpos* growing at San Pietro di Rudas, near Aggius in Sardinia (Italy). This strain nodulates and fixes N₂ effectively with *M. polymorpha* and *M. sphaerocarpos* [8]. Like *M. murex* and *M. polymorpha*, *M. sphaerocarpos* is an annual species which is tolerant of low pH soils [11], with studies suggesting that it only establishes N₂-fixing associations with *E. medicae* strains [8,9]. However, owing to a paucity of symbiotic information, it is not yet clear whether *M. sphaerocarpos* fixes N₂ with a wide range of *E. medicae* strains or if this ability is restricted to a smaller set of *E. medicae* accessions. Therefore, genome sequences of *E. medicae* strains effective with *M. sphaerocarpos* will provide a valuable genetic resource to further investigate the symbiotaxonomy of *Medicago*-nodulating rhizobia and will further enhance the existing available genome data for *Ensifer* microsymbionts [12-15]. Here we present a summary classification and a set of general features

for this microsymbiont together with a description of its genome sequence and annotation.

Classification and features

E. medicae WSM1369 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 1.0-1.5 µm in length (Figure 1 Left and 1 Center). It is fast growing, forming colonies within 3-4 days when grown on TY agar [16] or half strength Lupin Agar (½LA) [17] 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 *E. medicae* WSM1369 in a 16S rRNA sequence based tree. This strain shares 100% sequence identity (over 1290 bp) to the 16S rRNA of *E. medicae* A321^T and *E. medicae* WSM419 [13] and 99% sequence identity (1362/1366 bp) to the 16S rRNA of *E. meliloti* Sm1021 [12].

Symbiotaxonomy

E. medicae strain WSM1369 was isolated in 1993 from a nodule collected from the annual *M. sphaerocarpos* growing at San Pietro di Rudas, near Aggius, Sardinia in Italy (J. G. Howieson, pers. comm.). The site of collection was undulating grassland, with a soil derived from granite materials that had a depth of 20-40 cm and a pH of 6.0. The soil was a loamy-sand and *Lathyrus* and *Trifolium* spp. grew in association with *M. sphaerocarpos*. WSM1369 forms nodules (Nod⁺) and fixes N₂ (Fix⁺) with *M. polymorpha* and *M. sphaerocarpos* [8].

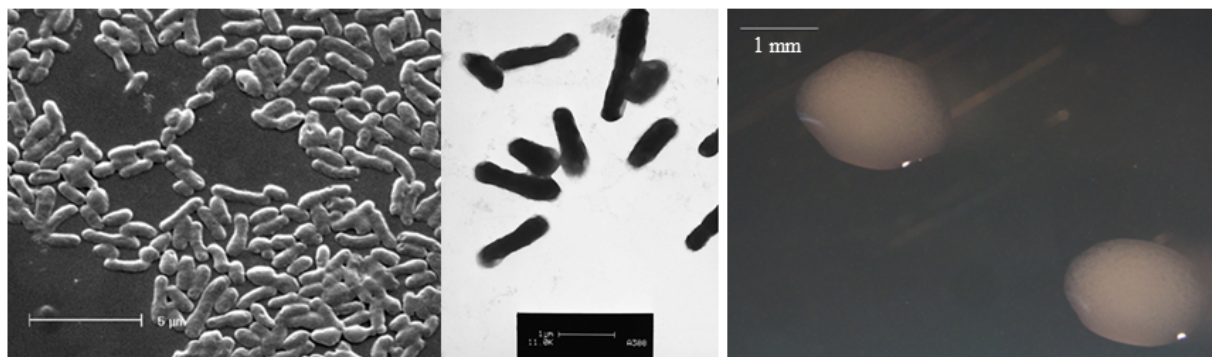


Figure 1. Images of *Ensifer medicae* WSM1369 using scanning (Left) and transmission (Center) electron microscopy and the appearance of colony morphology on half strength lupin agar (Right).

Table 1. Classification and general features of *Ensifer medicae* WSM1369 according to the MIGS recommendations [18]

MIGS ID	Property	Term	Evidence code
		Domain <i>Bacteria</i>	TAS [19]
		Phylum <i>Proteobacteria</i>	TAS [20]
		Class <i>Alphaproteobacteria</i>	TAS [21,22]
	Current classification	Order <i>Rhizobiales</i>	TAS [21,23]
		Family <i>Rhizobiaceae</i>	TAS [24,25]
		Genus <i>Ensifer</i>	TAS [26-28]
		Species <i>Ensifer medicae</i>	TAS [27]
		Strain WSM1369	TAS [8]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	IDA
	Salinity	Non-halophile	NAS
MIGS-22	Oxygen requirement	Aerobic	TAS [8]
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-6	Habitat	Soil, root nodule, on host	NAS
MIGS-15	Biotic relationship	Free living, symbiotic	TAS [8]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [29]
	Isolation	Root nodule	TAS [8]
MIGS-4	Geographic location	Sardinia, Italy	TAS [8]
MIGS-5	Soil collection date	28 April 1993	IDA
MIGS-4.1	Longitude	9.019167	IDA
MIGS-4.2	Latitude	40.971667	IDA
MIGS-4.3	Depth	0-10 cm	IDA
MIGS-4.4	Altitude	Not recorded	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 [30].

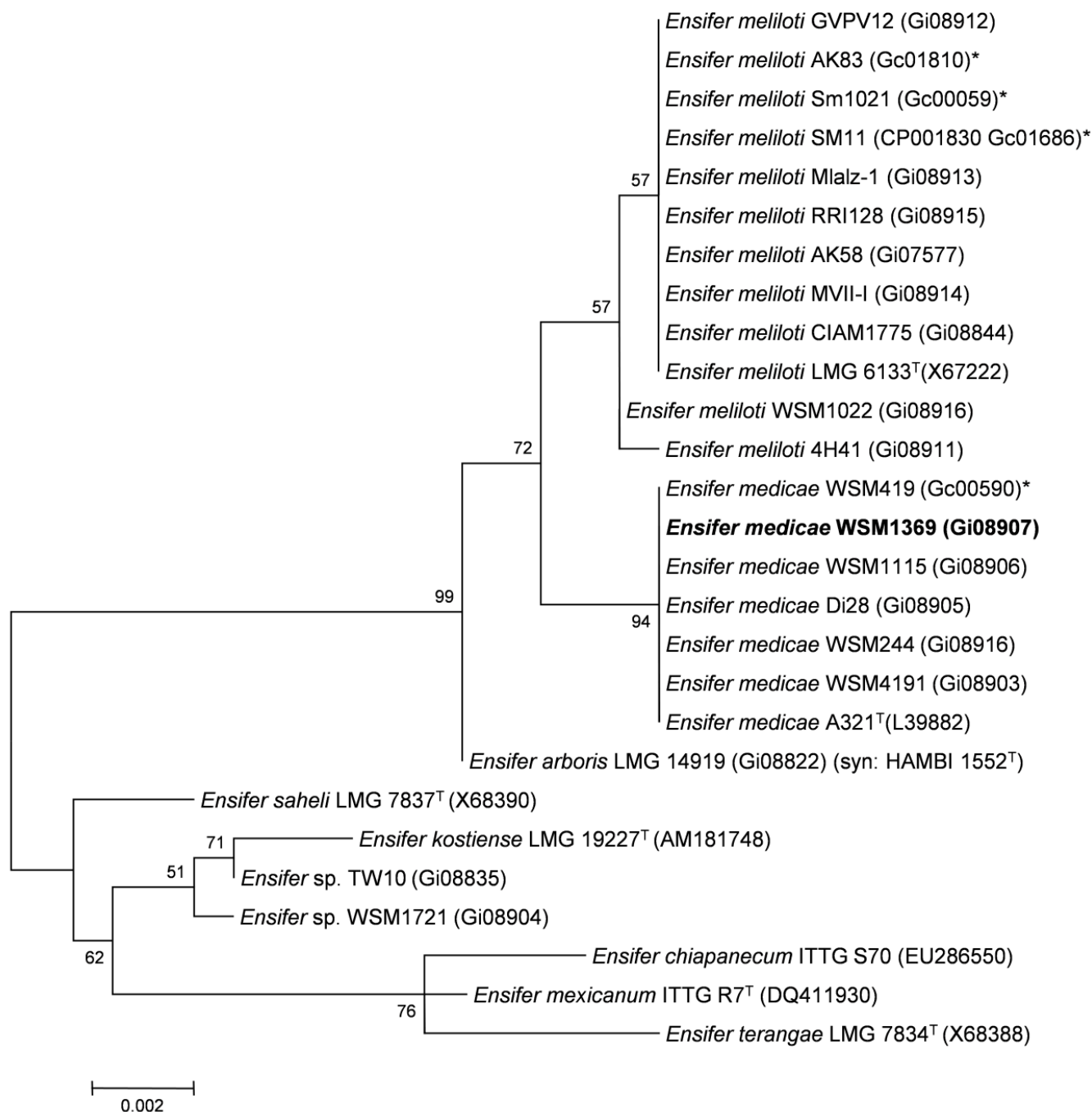


Figure 2. Phylogenetic tree showing the relationship of *Ensifer medicae* WSM1369 (shown in bold print) to other *Ensifer* spp. in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1,290 bp internal region). All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5 [31]. The tree was built using the Maximum-Likelihood method with the General Time Reversible model [32]. Bootstrap analysis [33] 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 [34]. Published genomes are indicated with an asterisk.

Table 2. Genome sequencing project information for *E. medicae* WSM1369

MIGS ID	Property	Term
MIGS-31	Finishing quality	Standard draft
MIGS-28	Libraries used	One Illumina fragment library
MIGS-29	Sequencing platforms	Illumina HiSeq 2000
MIGS-31.2	Sequencing coverage	Illumina: 321×
MIGS-30	Assemblers	Velvet version 1.1.04; Allpaths-LG version r39750
MIGS-32	Gene calling methods	Prodigal 1.4
	GenBank	AQUS00000000
	GenBank release date	August 28, 2013
	GOLD ID	Gi08907
	NCBI project ID	165337
	Database: IMG	2513237156
	Project relevance	Symbiotic N, fixation, agriculture

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 [34] and a standard 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.

Growth conditions and DNA isolation

E. medicae WSM1369 was cultured to mid logarithmic phase in 60 ml of TY rich medium on a gyratory shaker at 28°C [35]. DNA was isolated from the cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [36].

Genome sequencing and assembly

The genome of *Ensifer medicae* WSM1369 was sequenced at the Joint Genome Institute (JGI) using Illumina technology [37]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 13,712,318 reads totaling 2,057 Mbp.

All general aspects of library construction and sequencing performed at the JGI can be found at the JGI user home [36]. 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. and Han, J., unpublished). The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet [38] (version 1.1.04), (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [39], (3) Illumina reads were assembled with simulated read pairs using Allpaths-LG [40] (version r39750). Parameters for assembly steps were: 1) Velvet (velveth: 63 -shortPaired and velvetg: -veryclean yes -exportFiltered yes -mincontiglth 500 -scaffolding no-covcutoff 10) 2) wgsim (-e 0 -1 76 -2 76 -r 0 -R 0 -X 0) 3) Allpaths-LG (PrepareAllpathsInputs:PHRED64=1 PLOIDY=1 FRAGCOVERAGE=125 JUMPCOVERAGE=25 LONGJUMPCOV=50, RunAllpath-sLG: THREADS=8 RUN=stdshredpairs TARGETS=standard VAPIWARNONLY=True OVERWRITE=True). The final draft assembly contained 307 contigs in 307 scaffolds. The total size of the genome is 6.4 Mbp and the final assembly is based on 2,057 Mbp of Illumina data, which provides an average 321× coverage of the genome.

Genome annotation

Genes were identified using Prodigal [41] as part of the DOE-JGI annotation pipeline [42]. 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. The tRNAscanSE tool [43] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [44]. 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

[45]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes (IMG-ER) platform [46].

Genome properties

The genome is 6,402,557 nucleotides with 61.13% GC content (Table 3) and comprised of 307 scaffolds (Figure 3) of 307 contigs. From a total of 6,735 genes, 6,656 were protein encoding and 79 RNA only encoding genes. The majority of genes (74.14%) were assigned a putative function while the remaining genes were annotated as hypothetical. The distribution of genes into COGs functional categories is presented in Table 4.

Table 3. Genome Statistics for *Ensifer medicae* WSM1369

Attribute	Value	% of Total
Genome size (bp)	6,402,557	100.00
DNA coding region (bp)	5,536,774	86.48
DNA G+C content (bp)	3,913,921	61.13
Number of scaffolds	307	
Number of contigs	307	
Total gene	6,735	100.00
RNA genes	79	1.17
rRNA operons	1	0.01
Protein-coding genes	6,656	98.83
Genes with function prediction	4,993	74.14
Genes assigned to COGs	4,988	74.06
Genes assigned Pfam domains	5,185	76.99
Genes with signal peptides	508	7.54
Genes coding transmembrane proteins	1,424	21.14
CRISPR repeats	0	

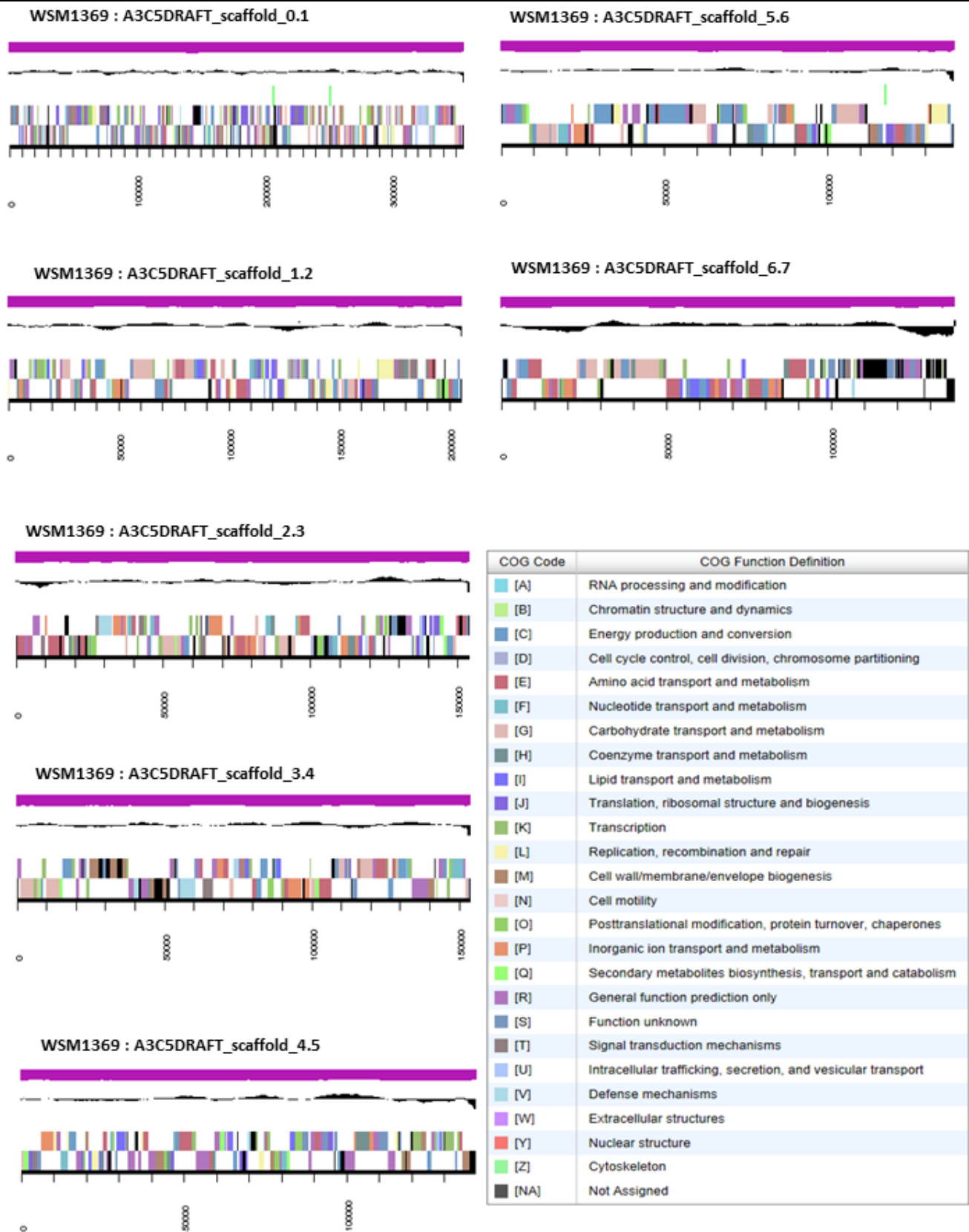


Figure 3. Graphical map of the genome of *Ensifer medicae* WSM1369 showing the seven largest scaffolds. From bottom to the top of each scaffold: Genes on forward strand (color by COG categories as denoted by the IMG platform), Genes on reverse strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.

Table 4. Number of protein coding genes of *Ensifer medicae* WSM1369 associated with the general COG functional categories.

Code	Value	% age	Description
J	193	3.48	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	486	8.77	Transcription
L	275	4.96	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	40	0.72	Cell cycle control, mitosis and meiosis
Y	0	0.00	Nuclear structure
V	54	0.97	Defense mechanisms
T	241	4.35	Signal transduction mechanisms
M	267	4.82	Cell wall/membrane biogenesis
N	77	1.39	Cell motility
Z	0	0.00	Cytoskeleton
W	1	0.02	Extracellular structures
U	124	2.24	Intracellular trafficking and secretion
O	184	3.32	Posttranslational modification, protein turnover, chaperones
C	308	5.56	Energy production conversion
G	510	9.21	Carbohydrate transport and metabolism
E	613	11.06	Amino acid transport metabolism
F	108	1.95	Nucleotide transport and metabolism
H	196	3.54	Coenzyme transport and metabolism
I	193	3.48	Lipid transport and metabolism
P	280	5.05	Inorganic ion transport and metabolism
Q	158	2.85	Secondary metabolite biosynthesis, transport and catabolism
R	662	11.95	General function prediction only
S	569	10.27	Function unknown
-	1,747	25.94	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 Murdoch University Strategic Research Fund through the Crop and Plant Research Institute (CaPRI) and the Centre for Rhizobium Studies (CRS) at Murdoch University.

References

- O'Hara GW. The role of nitrogen fixation in crop production. *J Crop Prod* 1998; (2):115-138. http://dx.doi.org/10.1300/J144v01n02_06
- Olivares J, Bedmar EJ, Sanjuan J. Biological nitrogen fixation in the context of global change. *Mol Plant Microbe Interact* 2013; **26**:486-494. [PubMed http://dx.doi.org/10.1094/MPMI-12-12-0293-CR](http://dx.doi.org/10.1094/MPMI-12-12-0293-CR)
- Terpolilli JJ, Hood GA, Poole PS. What determines the efficiency of N₂-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](http://dx.doi.org/10.1016/B978-0-12-398264-3.00005-X)
- 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)
- 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>
- Graham P. Ecology of the root-nodule bacteria of legumes. In: Dilworth MJ, James EK, Sprent JI, Newton WE, editors. *Nitrogen-Fixing Leguminous Symbioses*. Dordrecht: The Netherlands: Springer; 2008. p 23-43.
- Rome S, Fernandez MP, Brunel B, Normand P, Cleyet-Marel JC. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. *Int J Syst Bacteriol* 1996; **46**:972-980. [PubMed http://dx.doi.org/10.1099/00207713-46-4-972](http://dx.doi.org/10.1099/00207713-46-4-972)
- Garau G, Reeve WG, Brau L, Yates RJ, James D, Tiwari R, O'Hara GW, Howieson JG. The symbiotic requirements of different *Medicago* spp. suggest the evolution of *Sinorhizobium meliloti* and *S. medicae* with hosts differentially adapted to soil pH. *Plant Soil* 2005; **276**:263-277. <http://dx.doi.org/10.1007/s11104-005-0374-0>
- Terpolilli JJ, O'Hara GW, Tiwari RP, Dilworth MJ, Howieson JG. The model legume *Medicago truncatula* A17 is poorly matched for N₂ fixation with the sequenced microsymbiont *Sinorhizobium meliloti* 1021. *New Phytol* 2008; **179**:62-66. [PubMed http://dx.doi.org/10.1111/j.1469-8137.2008.02464.x](http://dx.doi.org/10.1111/j.1469-8137.2008.02464.x)
- Howieson JG, Nutt B, Evans P. Estimation of host-strain compatibility for symbiotic N-fixation between *Rhizobium meliloti*, several annual species of *Medicago* and *Medicago sativa*. *Plant Soil* 2000; **219**:49-55. <http://dx.doi.org/10.1023/A:1004795617375>
- Initiative IOC. Climate variability and change in southwest Western Australia. 2002. p 1-34.
- Galibert F, Finan TM, Long SR, Puhler A, Abola P, Ampe F, Barloy-Hubler F, Barnett MJ, Becker A, Boistard P, et al. The composite genome of the legume symbiont *Sinorhizobium meliloti*. *Science* 2001; **293**:668-672. [PubMed http://dx.doi.org/10.1126/science.1060966](http://dx.doi.org/10.1126/science.1060966)
- Reeve W, Chain P, O'Hara G, Ardley J, Nandesena K, Brau L, Tiwari R, Malfatti S, Kiss H, Lapidus A, et al. Complete genome sequence of the *Medicago* microsymbiont *Ensifer (Sinorhizobium) medicae* strain WSM419. *Stand Genomic Sci* 2010; **2**:77-86. [PubMed http://dx.doi.org/10.4056/sigs.43526](http://dx.doi.org/10.4056/sigs.43526)
- Terpolilli JJ, Hill YJ, Tian R, Howieson JG, Bräu L, Goodwin L, Han J, Liolios K, Huntemann M, Pati AWT, et al. Genome sequence of *Ensifer meliloti* strain WSM1022; a highly effective microsymbiont of the model legume *Medicago truncatula* A17. *Stand Genomic Sci* 2013; (In press). <http://dx.doi.org/10.4056/sigs.4838624>
- Tak N, Gehlot HS, Kaushik M, Choudhary S, Tiwari R, Tian R, Hill YJ, Bräu L, Goodwin L, Han J, et al. Genome sequence of *Ensifer* sp. TW10; a *Tephrosia wallichii* (Biyani) microsymbiont native to the Indian Thar Desert. *Stand Genomic Sci* 2013; (In press). <http://dx.doi.org/10.4056/sigs.4598281>
- 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](http://dx.doi.org/10.1099/00221287-84-1-188)
- 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)

19. 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>
20. 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.
21. 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>
22. 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.
23. 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.
24. 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>
25. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. *J Bacteriol* 1938; **36**:320-321.
26. Casida LE. *Ensifer adhaerens* gen. nov., sp. nov.: a bacterial predator of bacteria in soil. *Int J Syst Bacteriol* 1982; **32**:339-345. <http://dx.doi.org/10.1099/00207713-32-3-339>
27. Young JM. The genus name *Ensifer* Casida 1982 takes priority over *Sinorhizobium* Chen et al. 1988, and *Sinorhizobium morelense* Wang et al. 2002 is a later synonym of *Ensifer adhaerens* Casida 1982. Is the combination *Sinorhizobium adhaerens* (Casida 1982) Willems et al. 2003 legitimate? Request for an Opinion. *Int J Syst Evol Microbiol* 2003; **53**:2107-2110. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.02665-0>
28. Judicial Commission of the International Committee on Systematics of Prokaryotes. The genus name *Sinorhizobium* Chen et al. 1988 is a later synonym of *Ensifer* Casida 1982 and is not conserved over the latter genus name, and the species name '*Sinorhizobium adhaerens*' is not validly published. Opinion 84. *Int J Syst Evol Microbiol* 2008; **58**:1973. [PubMed](#) <http://dx.doi.org/10.1099/ijs.0.2008/005991-0>
29. Agents B. Technical rules for biological agents. TRBA (<http://www.baua.de>):466.
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. Nei M, Kumar S. *Molecular Evolution and Phylogenetics*. New York: Oxford University Press; 2000.
33. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 1985; **39**:783-791. <http://dx.doi.org/10.2307/2408678>
34. 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>
35. 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>
36. DOE Joint Genome Institute user home. <http://my.jgi.doe.gov/general/index.html>.
37. Bennett S. Solexa Ltd. *Pharmacogenomics* 2004; **5**:433-438. [PubMed](#) <http://dx.doi.org/10.1517/14622416.5.4.433>
38. Zerbino DR. Using the Velvet *de novo* assembler for short-read sequencing technologies. *Current Protocols in Bioinformatics* 2010;Chapter 11:Unit 11 5.
39. Wgsim sequence read simulator. <https://github.com/lh3/wgsim>.
40. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP,

- Sykes S, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. *Proc Natl Acad Sci USA* 2011; **108**:1513-1518. [PubMed](#)
<http://dx.doi.org/10.1073/pnas.1017351108>
41. 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>
42. 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>
43. 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](#)
44. Pruesse E, Quast C, Knittel K, Fuchs BdM, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. *Nucleic Acids Res* 2007; **35**:7188-7196. [PubMed](#)
<http://dx.doi.org/10.1093/nar/gkm864>
45. INFERNAL. <http://infernal.janelia.org>
46. 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>