

Binghamton University The Open Repository @ Binghamton (The ORB)

Biological Sciences Faculty Scholarship

Biological Sciences

5-2015

High-quality permanent draft genome sequence of *Bradyrhizobium* sp Tv2a.2, a microsymbiont of *Tachigali versicolor* discovered in Barro Colorado Island of Panama

Rui Tian

Matthew Parker

Binghamton University--SUNY

Rekha Seshadri

T. B.K. Reddy

Victor Markowitz

See next page for additional authors

Follow this and additional works at: https://orb.binghamton.edu/bio_fac

 Part of the [Biology Commons](https://biologycommons.org/)

Recommended Citation

Tian, Rui; Parker, Matthew; Seshadri, Rekha; Reddy, T. B.K.; Markowitz, Victor; Ivanova, Natalia; Pati, Amrita; Woyke, Tanja; Baeshen, Mohammed; Baeshen, Nabih; Kyrpides, Nikos; and Reeve, Wayne, "High-quality permanent draft genome sequence of *Bradyrhizobium* sp Tv2a.2, a microsymbiont of *Tachigali versicolor* discovered in Barro Colorado Island of Panama" (2015). *Biological Sciences Faculty Scholarship*. 11.
https://orb.binghamton.edu/bio_fac/11

This Article is brought to you for free and open access by the Biological Sciences at The Open Repository @ Binghamton (The ORB). It has been accepted for inclusion in Biological Sciences Faculty Scholarship by an authorized administrator of The Open Repository @ Binghamton (The ORB). For more information, please contact ORB@binghamton.edu.

Authors

Rui Tian, Matthew Parker, Rekha Seshadri, T. B.K. Reddy, Victor Markowitz, Natalia Ivanova, Amrita Pati, Tanja Woyke, Mohammed Baeshen, Nabih Baeshen, Nikos Kyrpides, and Wayne Reeve

SHORT GENOME REPORT

Open Access



High-quality permanent draft genome sequence of *Bradyrhizobium* sp. Ai1a-2; a microsymbiont of *Andira inermis* discovered in Costa Rica

Rui Tian¹, Matthew Parker², Rekha Seshadri³, TBK Reddy³, Victor Markowitz⁴, Natalia Ivanova³, Amrita Pati³, Tanja Woyke³, Mohammed Baeshen^{5,6}, Nabih Baeshen^{5,7}, Nikos Kyrpides^{3,7} and Wayne Reeve^{1*}

Abstract

Bradyrhizobium sp. Ai1a-2 is an aerobic, motile, Gram-negative, non-spore-forming rod that was isolated from an effective nitrogen fixing root nodule of *Andira inermis* collected from Tres Piedras in Costa Rica. In this report we describe, for the first time, the genome sequence information and annotation of this legume microsymbiont. The 9,029,266 bp genome has a GC content of 62.56% with 247 contigs arranged into 246 scaffolds. The assembled genome contains 8,482 protein-coding genes and 102 RNA-only encoding genes. This rhizobial genome was sequenced as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project proposal.

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

Introduction

Bradyrhizobium sp. strain Ai1a.2 is a representative of a distinctive lineage affiliated with the *Bradyrhizobium elkanii* superclade [1]. The *B. elkanii* superclade is one of the three main branches of the genus, together with the *B. japonicum*/*B. diazoefficiens* superclade [2,3], and the group encompassing photosynthetic *Aeschynomene* symbionts [4].

Members of the lineage represented by strain Ai1a.2 are readily diagnosed because they share a distinctive length variant in helix 9 within the 5' intervening sequence region of the 23S rRNA gene [5]. Strain Ai1a.2 and its relatives have an insertion of 16 nucleotides in this region in comparison to *B. elkanii* USDA76, which can be identified by a straightforward PCR assay [6]. In a survey of 420 *Bradyrhizobium* strains from 25 countries [1], only 2% of the strains had this 23S rRNA length variant. These strains all clustered together into a strongly supported clade based on concatenated data for 23S rRNA and five protein-coding genes [1].

This clade was placed as the most basally diverging lineage within the *B. elkanii* superclade, and it included

strains from three locations: Central America, the Caribbean, and South Africa. Strain Ai1a.2 was sampled in Costa Rica from the tree *Andira inermis* [6], and highly similar strains are also known to occur as symbionts of the same host legume in Panama [7]. Parker and Rousteau [8] also detected strains from this group in nodule samples from the beach legume *Canavalia rosea* in two Caribbean locations (Guadeloupe and Puerto Rico). Two *Bradyrhizobium* strains from distantly related legume hosts (*Leobordea* spp.) in South Africa (WSM2632, WSM2783) also belong to this clade [9].

Andira inermis, the host of strain Ai1a.2, is a large tree (up to 35 m height) commonly found in riparian habitats from southern Mexico through northern South America [10]. *Andira* was traditionally considered to be an early-diverging lineage within the Tribe Dalbergieae [11], but more recent phylogenetic analyses have suggested that it forms a separate lineage with unclear relationship to dalbergioid legumes [12]. Here we provide an analysis of the high-quality permanent draft genome sequence of *Bradyrhizobium* strain Ai1a.1. The fact that the genome of its close relative WSM2783 has also been sequenced as part of the Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) project [13] will enable detailed comparative analysis of this group.

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

¹Centre for Rhizobium Studies, Murdoch University, Murdoch, Australia
Full list of author information is available at the end of the article

Organism information

Classification and features

Bradyrhizobium sp. Ai1a-2 is a motile, non-sporulating, non-encapsulated, Gram-negative strain in the order *Rhizobiales* of the class *Alphaproteobacteria*. The rod shaped form has dimensions of approximately 0.5 μm in width and 1.5–2.0 μm in length (Figure 1 Left and Center). It is relatively slow growing, forming colonies after 6–7 days when grown on half strength Lupin Agar ($\frac{1}{2}$ LA) [14], tryptone-yeast extract agar (TY) [15] or a modified yeast-mannitol agar (YMA) [16] at 28°C. Colonies on $\frac{1}{2}$ LA are opaque, slightly domed and moderately mucoid with smooth margins (Figure 1 Right).

Figure 2 shows the phylogenetic relationship of *Bradyrhizobium* sp. Ai1a-2 in a 16S rRNA gene sequence based tree. The 16S rRNA gene sequence of Ai1a-2 (using a 1,370 bp intragenic sequence) is identical to that of *Bradyrhizobium* sp. WSM2783. *Bradyrhizobium* sp. Ai1a-2 is also closely related to *Bradyrhizobium* sp. Cp5.3 and *Bradyrhizobium* sp. Th.b2 with 16S rRNA gene sequence identities of 99.77% and 99.23%, respectively, as determined using NCBI BLAST analysis [17]. The highest identity (99.16%) of the 16S rRNA gene sequence of strain Ai1a-2 to type strain sequences occurs with *Bradyrhizobium icense* LMTR 13^T and *Bradyrhizobium paxllaeri* LMTR 21^T based on alignment using the EzTaxon-e server [18,19].

Minimum Information about the Genome Sequence (MIGS) is provided in Table 1 and Additional file 1: Table S1.

Symbiotaxonomy

Strain Ai1a.2 was isolated from the tree *Andira inermis*, Costa Rica [6]. The authentication of the symbiotic ability could not be performed using this host because seeds could not be accessed. The symbiotic capability of strain Ai1a.2 was tested on *Macroptilium atropurpureum* and this strain was able to nodulate this host. Acetylene reduction assays showed established nodules contained active nitrogenase, indicating an effective symbiosis with this host [6].

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, Root Nodulating Bacteria (GEBA-RNB) project at the U.S. Department of Energy, Joint Genome Institute (JGI). The genome project is deposited in the Genomes OnLine Database [20] and a high-quality permanent draft genome sequence in IMG [21]. Sequencing, finishing and annotation were performed by the JGI using state of the art sequencing technology [22]. A summary of the project information is shown in Table 2.

Growth conditions and genomic DNA preparation

Bradyrhizobium sp. Ai1a-2 was cultured to mid logarithmic phase in 60 ml of TY rich media on a gyratory shaker at 28°C [23]. DNA was isolated from the cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [24].

Genome sequencing and assembly

The draft genome of *Bradyrhizobium* sp. Ai1a-2 was generated at the DOE Joint Genome Institute (JGI) using the Illumina technology [25]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 21,669,974 reads totaling 3,250.5 Mbp. All general aspects of library construction and sequencing were performed at the JGI and details can be found on the JGI website [26]. 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) [27], (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [28], (3) Illumina reads were

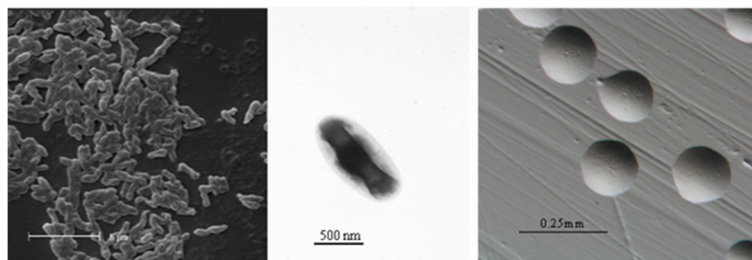


Figure 1 Images of *Bradyrhizobium* sp. Ai1a-2 using scanning (Left) and transmission (Center) electron microscopy as well as light microscopy to visualize colony morphology on solid media (Right).

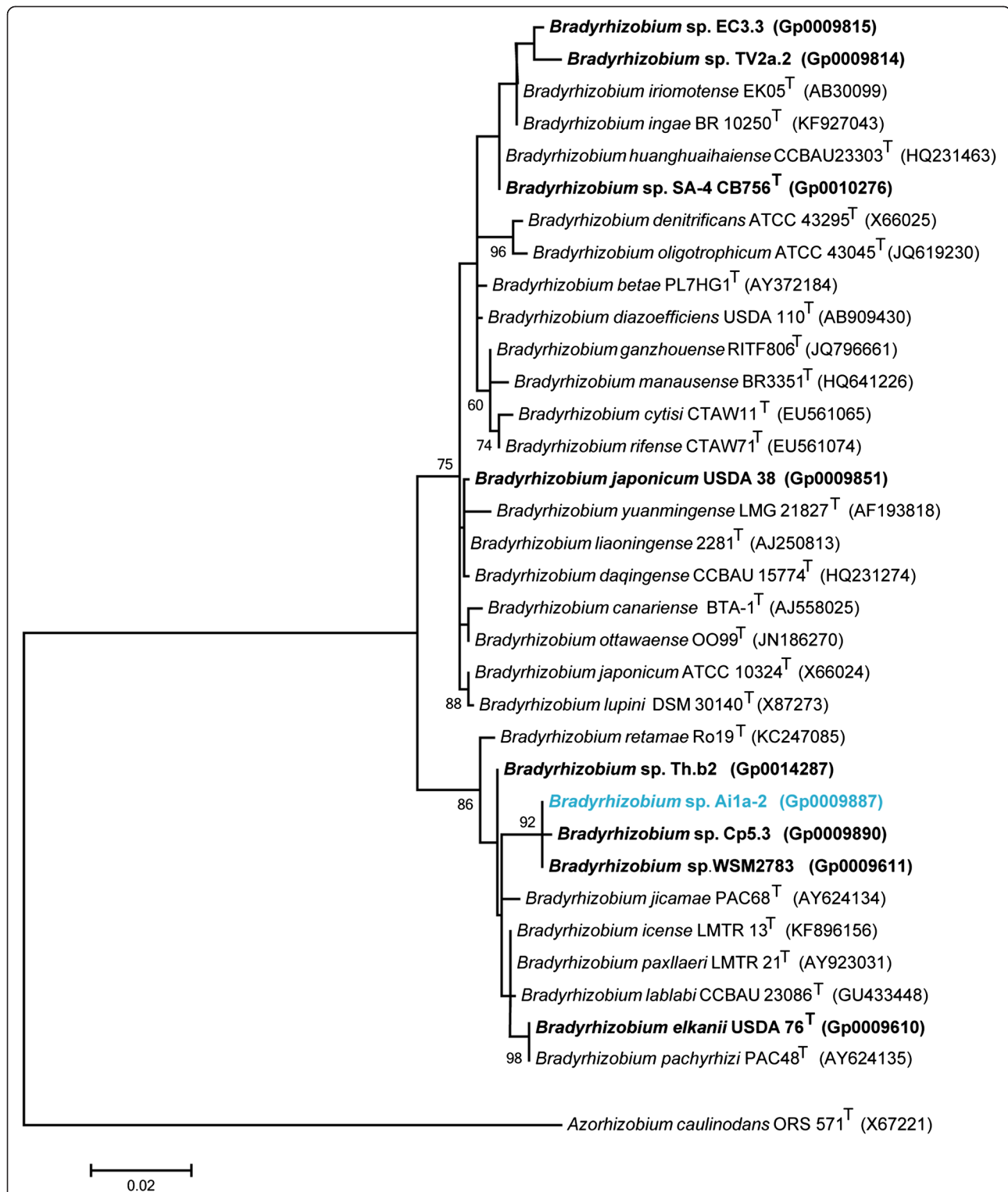


Figure 2 Phylogenetic tree showing the relationship of *Bradyrhizobium* sp. Ai1a-2 (shown in blue print) relative to other type and non-type strains in the *Bradyrhizobium* genus using a 1,310 bp intragenic sequence of the 16S rRNA gene. *Azorhizobium caulinodans* ORS 571^T sequence was used as an outgroup. All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5.05 [37]. 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 [20] have the GOLD ID mentioned after the strain number and are represented in bold, otherwise the NCBI accession number is provided.

Table 1 Classification and general features of *Bradyrhizobium* sp. Ai1a-2 in accordance with the MIGS recommendations [38] published by the Genome Standards Consortium [39]

MIGS ID	Property	Term	Evidence code
	Classification	Domain <i>Bacteria</i>	TAS [40]
		Phylum <i>Proteobacteria</i>	TAS [41,42]
		Class <i>Alphaproteobacteria</i>	TAS [42,43]
		Order <i>Rhizobiales</i>	TAS [44]
		Family <i>Bradyrhizobiaceae</i>	TAS [45]
		Genus <i>Bradyrhizobium</i>	TAS [46]
		Species <i>Bradyrhizobium</i> sp.	IDA
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Unknown	NAS
	Optimum temperature	28°C	NAS
	pH range; Optimum	Unknown	NAS
	Carbon source	Varied	NAS
	Energy source	Chemoorganotroph	NAS
MIGS-6	Habitat	Soil, root nodule, host	TAS [6]
MIGS-6.3	Salinity	Non-halophile	NAS
MIGS-22	Oxygen requirement	Aerobic	NAS
MIGS-15	Biotic relationship	Free living, symbiotic	TAS [6]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [47]
	Isolation	Root nodule of <i>Andira inermis</i>	TAS [6]
MIGS-4	Geographic location	Tres Piedras, Costa Rica	TAS [6]
MIGS-5	Sample collection	July 14, 2000	IDA
MIGS-4.1	Latitude	9.2835	IDA
MIGS-4.2	Longitude	-83.8533	IDA
MIGS-4.3	Depth	5 cm	IDA
MIGS-4.4	Altitude	50 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). Evidence codes are from the Gene Ontology project [48,49].

assembled with simulated read pairs using Allpaths-LG (version r42328) [29]. Parameters for assembly steps were: 1) Velvet (velveth: 63 -shortPaired and velvetg: -very_clean yes -exportFiltered yes -min_contig_lgth 500 -scaffolding no -cov_cutoff 10) 2) wgsim (-e 0 -1 100 -2 100 -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 247 contigs in 246 scaffolds. The total size of the genome is 9.0 Mbp and the final assembly is based on 1,081.2 Mbp

of Illumina data, which provides an average 119.7X coverage of the genome.

Genome annotation

Genes were identified using Prodigal [30], as part of the DOE-JGI genome annotation pipeline [31,32]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information (NCBI) non-redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAScanSE tool [33] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [34]. Other non-

Table 2 Project information

MIGS ID	Property	Term
MIGS-31	Finishing quality	High-quality permanent draft
MIGS-28	Libraries used	Illumina Standard PE
MIGS-29	Sequencing platforms	Illumina HiSeq2000
MIGS-31.2	Fold coverage	Illumina, 119.7x
MIGS-30	Assemblers	Velvet version 1.1.04; Allpaths-LG version r42328
MIGS-32	Gene calling method	Prodigal 1.4
	Locus Tag	K288
	GenBank ID	AUEZ00000000
	GenBank release date	June 12, 2014
	GOLD ID	Gp0009887 [50]
	BIOPROJECT	195749
MIGS-13	Source Material Identifier	Ai1a-2
	Project relevance	Symbiotic nitrogen fixation, agriculture

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 [35]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) system [36] developed by the Joint Genome Institute, Walnut Creek, CA, USA.

Genome properties

The genome is 9,029,266 nucleotides with 62.56% GC content (Table 3) and comprised of 246 scaffolds. From

Table 3 Genome statistics for *Bradyrhizobium* sp. Ai1a-2

Attribute	Value	% of total
Genome size (bp)	9,029,266	100.00
DNA coding (bp)	7,683,922	85.10
DNA G + C (bp)	5,648,849	62.56
DNA scaffolds	246	100
Total genes	8,584	100.00
Protein coding genes	8,482	98.81
RNA genes	102	1.19
Pseudo genes	0	0.00
Genes in internal clusters	837	9.75
Genes with function prediction	6,447	75.10
Genes assigned to COGs	5,111	59.54
Genes with Pfam domains	6,590	76.77
Genes with signal peptides	837	9.75
Genes with transmembrane helices	1,914	22.30
CRISPR repeats	0	0.00

a total of 8,584 genes, 8,482 were protein encoding and 102 RNA only encoding genes. The majority of genes (75.10%) 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.

Conclusions

Bradyrhizobium sp. Ai1a-2 is a member of a widely distributed *Bradyrhizobium* lineage, isolated from diverse legume hosts in North, Central and South America and South Africa. Little is currently known of the symbiotic associations of its host *Andira inermis*, apart from the discovery that the Puerto Rican isolate *Bradyrhizobium* sp. EC3.3 can also establish a symbiosis with this host [8]. The Costa Rican isolate Ai1a-2 16S rRNA gene sequence is distinct to that of EC3.3 but identical to the 16S rRNA sequence of South African isolate *Bradyrhizobium* sp.

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

Code	Value	% of total (5,698)	COG category
J	185	3.25	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	412	7.23	Transcription
L	223	3.91	Replication, recombination and repair
B	2	0.04	Chromatin structure and dynamics
D	33	0.58	Cell cycle control, cell division, chromosome partitioning
V	89	1.56	Defense mechanisms
T	234	4.11	Signal transduction mechanisms
M	277	4.86	Cell wall/membrane/envelope biogenesis
N	94	1.65	Cell motility
U	128	2.25	Intracellular trafficking, secretion, and vesicular transport
O	191	3.35	Posttranslational modification, protein turnover, chaperones
C	435	7.63	Energy production and conversion
G	340	5.97	Carbohydrate transport and metabolism
E	587	10.30	Amino acid transport and metabolism
F	77	1.35	Nucleotide transport and metabolism
H	198	3.47	Coenzyme transport and metabolism
I	311	5.46	Lipid transport and metabolism
P	364	6.39	Inorganic ion transport and metabolism
Q	256	4.49	Secondary metabolite biosynthesis, transport and catabolism
R	696	12.21	General function prediction only
S	566	9.93	Function unknown
-	3,473	40.46	Not in COGS

WSM2783. Phylogenetically, Ai1a-2 is closely related to *Bradyrhizobium* sp. Cp5.3 and *Bradyrhizobium* sp. Th.b2 from Panama and USA, respectively. The genome of *Bradyrhizobium* 1a-2 and Ai sp.WSM2783 were sequenced along with 23 other *Bradyrhizobium* genomes as a part of the GEBA-RNB project. Of these 25 sequenced strains, the *Bradyrhizobium* spp. Ai1a-2, WSM2783, Cp5.3, Th.b2 and *B. elkanii* USDA76^T are affiliated with the *Bradyrhizobium elkanii* superclade. The *Bradyrhizobium* Ai1a-2 genome has the 2nd lowest genome size (9 Mbp), gene count (8,584) and signal peptide percentage (9.75%) among these five strains. Comparing the genome attributes of *Bradyrhizobium* sp. Ai1a-2 along with other sequenced *Bradyrhizobium* genomes will be important for the understanding of the biogeography of *Bradyrhizobium* spp. interactions required for the successful establishments of effective symbioses with their diverse hosts.

Additional file

Additional file 1: Associated MIGS record. Table S1. Associated MIGS record for *Bradyrhizobium* sp. Ai1a-2.

Abbreviations

GEBA-RNB: Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria; JGI: Joint Genome Institute; ½LA: Half strength Lupin Agar; TY: Tryptone yeast; YMA: Yeast mannitol agar; CTAB: Cetyl trimethyl ammonium bromide.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

MP supplied the strain and background information for this project and the DNA to the JGI, TR performed all imaging, TR and WR drafted the paper, MNB and NAB provided financial support and all other authors were involved in sequencing the genome and/or 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. We thank Gordon Thompson (Murdoch University) for the preparation of SEM and TEM photos. We would also like to thank the Center of Nanotechnology at King Abdulaziz University for their support.

Author details

¹Centre for Rhizobium Studies, Murdoch University, Murdoch, Australia. ²Binghamton University, State University of New York, New York, USA. ³DOE Joint Genome Institute, Walnut Creek, California, USA. ⁴Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, California, USA. ⁵Center of Nanotechnology, King Abdulaziz University, Jeddah, Saudi Arabia. ⁶Department of Biological Sciences, Faculty of Science, Jeddah University, Jeddah, Saudi Arabia. ⁷Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

Received: 13 February 2015 Accepted: 16 April 2015

Published online: 14 June 2015

References

- Parker MA. The spread of *Bradyrhizobium* lineages across host legume clades: from Abarema to Zygia. *Microb Ecol.* 2014;69:630–640.
- Willems A, Coopman R, Gillis M. Phylogenetic and DNA-DNA hybridization analyses of *Bradyrhizobium* species. *Int J Syst Evol Microbiol.* 2001;51:111–7.
- Delamuta JR, Ribeiro RA, Ormeno-Orrillo E, Melo IS, Martinez-Romero E, Hungria M. Polyphasic evidence supporting the reclassification of *Bradyrhizobium japonicum* group Ia strains as *Bradyrhizobium diazoefficiens* sp. nov. *Int J Syst Evol Microbiol.* 2013;63:3342–51.
- Chaintreuil C, Arrighi JF, Giraud E, Miche L, Moulin L, Dreyfus B, et al. Evolution of symbiosis in the legume genus *Aeschynomene*. *New Phytol.* 2013;200:1247–59.
- Evguenieva-Hackenberg E, Klug G. RNase III processing of intervening sequences found in helix 9 of 23S rRNA in the alpha subclass of *Proteobacteria*. *J Bacteriol.* 2000;182:4719–29.
- Parker MA. rRNA and *dnaK* relationships of *Bradyrhizobium* sp. nodule bacteria from four papilionoid legume trees in Costa Rica. *Syst Appl Microbiol.* 2004;27:334–42.
- Parker MA. Symbiotic relationships of legumes and nodule bacteria on Barro Colorado Island, Panama: a review. *Microb Ecol.* 2008;55:662–72.
- Parker MA, Rousteau A. Mosaic origins of *Bradyrhizobium* legume symbionts on the Caribbean island of Guadeloupe. *Mol Phylogenet Evol.* 2014;77:110–5.
- Ardley JK, Reeve WG, O'Hara GW, Yates RJ, Dilworth MJ, Howieson JG. Nodule morphology, symbiotic specificity and association with unusual rhizobia are distinguishing features of the genus *Listia* within the Southern African crotalariaioid clade *Lotononis* s.l. *Ann Bot.* 2013;112:1–15.
- Weaver PL. *Andira inermis* (W. Wright) DC. Silvics of forest trees of the American tropics. SO-ITF-SM-20. New Orleans, LA: USDA Forest Service, Southern Forest Experiment Station; 1989. p. 7. 7.
- Lewis G SB, Mackinder B, Lock M. *Legumes of the world* Royal Botanic Gardens, Kew, UK; 2005.
- Cardoso D PR, de Queiroz LP, Boatwright JS, Van Wyk BE, Wojciechowski MF, Lavin M. Reconstructing the deep-branching relationships of the papilionoid legumes. *South Afr J Bot.* 2013;89:58–75.
- Reeve W, Ardley J, Tian R, Eshragi L, Yoon JW, 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.
- 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.
- Vincent JM. A manual for the practical study of the root-nodule bacteria. International Biological Programme. UK: Blackwell Scientific Publications, Oxford; 1970.
- NCBI BLAST [<http://blast.ncbi.nlm.nih.gov/Blast.cgi>]
- 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.
- EZTaxon [<http://eztaxon-e.ezbiocloud.net/>]
- Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes OnLine Database (GOLD) v. 4: status of genomic and metagenomic projects and their associated metadata. *Nucleic Acids Res.* 2012;40:D571–9.
- 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.
- 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;145:1307–16.
- Protocols and sample preparation information [<http://jgi.doe.gov/collaborate-with-jgi/pmo-overview/protocols-sample-preparation-information/>]
- Bennett S. Solexa Ltd. *Pharmacogenomics.* 2004;5:433–8.
- JGI: Joint Genome Institute [<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.

28. Reads simulator wgsim [<https://github.com/lh3/wgsim>]
29. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. List no. 106. *Int J Syst Evol Microbiol.* 2005;55:2235–38.
30. 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.
31. 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.
32. 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.
33. 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.
34. 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.
35. Infernal: inference of RNA alignments [<http://infernal.janelia.org/>]
36. 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.
37. 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.
38. 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. *Nature Biotechnol.* 2008;26:541–7.
39. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, et al. The Genomic Standards Consortium. *PLoS Biol.* 2011;9, e1001088.
40. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains *Archaea*, *Bacteria*, and *Eucarya*. *P Natl A Sci USA.* 1990;87:4576–9.
41. Garrity GM, Bell JA, Lilburn T. hylum XIV. *Proteobacteria* phyl. nov. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. *Bergey's manual of systematic bacteriology*. Volume 2. Second ed. New York: Springer - Verlag; 2005. p. 1.
42. 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. *P Natl A Sci.* 2011;108:1513–8.
43. 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*. Volume 2. Second ed. New York: Springer - Verlag; 2005. p. 1.
44. 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.
45. Garrity GM, Bell JA, Lilburn T. Family VII. *Bradyrhizobiaceae* fam. nov. In *Bergey's Manual of Systematic Bacteriology*. Volume 2. edition. Edited by Brenner DJ. New York: Springer - Verlag; 2005: 438
46. Jordan DC. Transfer of *Rhizobium japonicum* Buchanan 1980 to *Bradyrhizobium* gen. nov., a genus of slow-growing, root nodule bacteria from leguminous plants. *Int J Syst Bacteriol.* 1982;32:136–9.
47. Biological Agents: Technical rules for biological agents. *TRBA:466*.
48. 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.
49. Guide to GO evidence codes [<http://www.geneontology.org/GO.evidence.shtml>]
50. GOLD ID for *Bradyrhizobium elkanii* WSM1741 [<https://gold.jgi-psf.org/projects?id=9846>]

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

