Standards in Genomic Sciences

EXTENDED GENOME REPORT **Open Access**

CrossMark

High-quality draft genome sequence of *Rhizobium mesoamericanum* strain STM6155, a *Mimosa pudica* microsymbiont from New Caledonia

Agnieszka Klonowska[1,2], Aline López-López[2], Lionel Moulin[1,2], Julie Ardley[3], Margaret Gollagher[4], Dora Marinova[4], Rui Tian[3], Marcel Huntemann[5], T.B.K. Reddy[5], Neha Varghese[5], Tanja Woyke[5], Victor Markowitz[6], Natalia Ivanova[5], Rekha Seshadri[5], Mohamed N. Baeshen[7], Nabih A. Baeshen[8], Nikos Kyrpides[5,8] and Wayne Reeve[3*]

Abstract

Rhizobium mesoamericanum STM6155 (INSCD = ATYY01000000) is an aerobic, motile, Gram-negative, non-spore-forming rod that can exist as a soil saprophyte or as an effective nitrogen fixing microsymbiont of the legume *Mimosa pudica* L.. STM6155 was isolated in 2009 from a nodule of the trap host *M. pudica* grown in nickel-rich soil collected near Mont Dore, New Caledonia. *R. mesoamericanum* STM6155 was selected as part of the DOE Joint Genome Institute 2010 Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria (GEBA-RNB) genome sequencing project. Here we describe the symbiotic properties of *R. mesoamericanum* STM6155, together with its genome sequence information and annotation. The 6,927,906 bp high-quality draft genome is arranged into 147 scaffolds of 152 contigs containing 6855 protein-coding genes and 71 RNA-only encoding genes. Strain STM6155 forms an ANI clique (ID 2435) with the sequenced *R. mesoamericanum* strain STM3625, and the nodulation genes are highly conserved in these strains and the type strain of *Rhizobium grahamii* CCGE501[T]. Within the STM6155 genome, we have identified a *chr* chromate efflux gene cluster of six genes arranged into two putative operons and we postulate that this cluster is important for the survival of STM6155 in ultramafic soils containing high concentrations of chromate.

Keywords: Root-nodule bacteria, Nitrogen fixation, *Rhizobium*, *Alphaproteobacteria*, *Mimosa*

Introduction

The ability of legumes to engage in a dinitrogen fixing symbiosis with soil dwelling bacteria, collectively known as rhizobia, has contributed to their success in colonizing nitrogen deficient soils over a broad range of edaphic conditions. While legume crops and pastures make important contributions to agricultural productivity, invasive legume weeds such as *Mimosa pudica* L. have a negative impact on natural and agricultural ecological systems. *M. pudica* originates from America [1] and became a highly invasive pantropical weed. It has been identified as a pest species, associated with land degradation, biodiversity loss, and reduced agricultural and therefore economic productivity, with attendant social and health impacts [2]. It requires resource-intensive chemical and mechanical control methods [2]. Conversely, however, it has potential commercial value as a source of silver nanoparticles and pharmacologically active phytochemicals, and as a phytoremediant for arsenic-polluted soils [3–6]. Understanding the *Mimosa* symbiosis can therefore help to achieve outcomes such as preventing biodiversity loss and improving the use of terrestrial ecosystems, as well as promoting sustainable industry, which form part of the Sustainable Development Goals adopted in September 2015 as part of the UN's development agenda 'Transforming our world: the 2030 Agenda for Sustainable Development' [7].

M. pudica has the unusual property of interacting with microsymbionts belonging to both alpha- and beta-rhizobia

* Correspondence: W.Reeve@murdoch.edu.au
[3]School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia
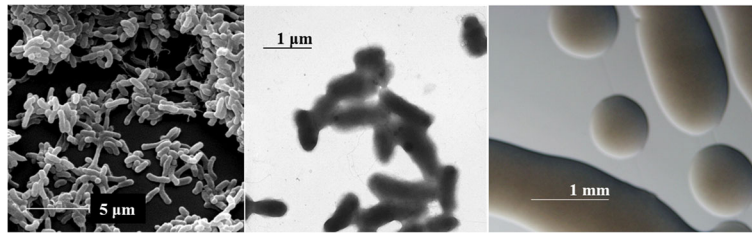Full list of author information is available at the end of the article

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 2 of 11

Fig. 1 Images of *Rhizobium mesoamericanum* STM6155 using scanning (*Left*) and transmission (*Center*) electron microscopy and the appearance of colony morphology on solid media (*Right*)
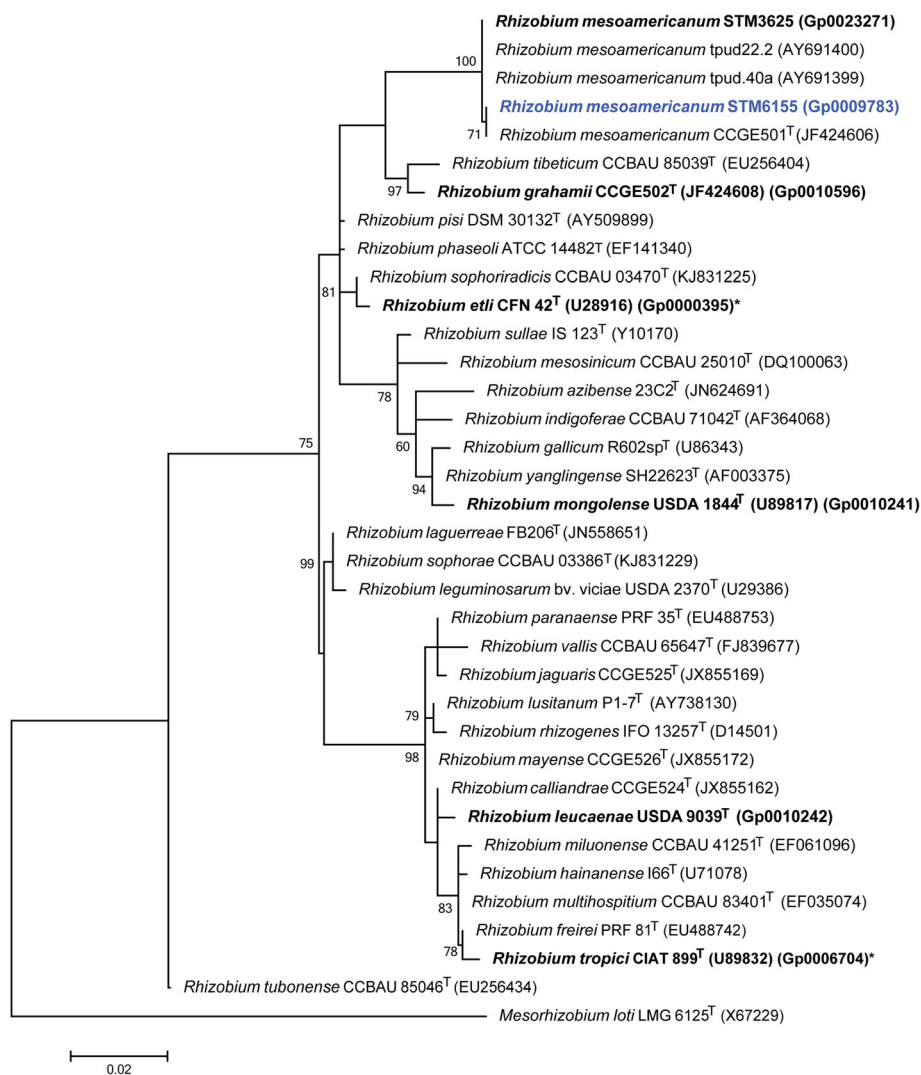

Fig. 2 Phylogenetic tree showing the relationship of *Rhizobium mesoamericanum* STM6155 (shown in *bold blue* print) to *Rhizobium* spp. and other root nodule bacteria in the order *Rhizobiales* based on aligned sequences of the 16S rRNA gene (1286 bp intragenic sequence). *Mesorhizobium loti* LMG6125[T] was used as an outgroup. All sites were informative and there were no gap-containing sites. Phylogenetic analyses were performed using MEGA, version 5 [53]. The tree was built using the Maximum-Likelihood method with the General Time Reversible model [54]. Bootstrap analysis [55] 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 [31] are in bold font and the GOLD ID is provided after the GenBank accession number, where this is available. Finished genomes are indicated with an asterisk

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 3 of 11

[8, 9]. Alpha-rhizobia are preferred symbionts of most legume species, but beta-rhizobia have a far narrower host range, with a particular affinity for the *Mimosa* genus in South America [10] and endemic papilionoid species in South Africa [11]. Diversity studies have shown that alpha-rhizobia are found less frequently than beta-rhizobia in the nodules of *M. pudica* [12–17], and nodulating species exhibit different competitive and symbiotic characteristics [18, 19]. *M. pudica* thus represents an interesting legume species for comparative analyses of symbiotic traits and plant-infection genetic programs in the two categories of symbionts.

M. pudica was introduced to New Caledonia at the end of the 19th century [15]. *Rhizobium mesoamericanum* STM6155 was isolated from nodules of *M. pudica* growing in soil characterized by neutral pH (6.8) and very high total nickel concentrations (10.1 g.kg^{-1}) that was collected near the abandoned nickel mining site of Mont Dore (S3: 22°15'16.51"S and 166°36'44.27"E) in New Caledonia [15].

The 16S rRNA and *recA* house-keeping genes of STM6155 showed 100 and 97% nucleotide identity with their orthologs in *Rhizobium mesoamericanum* CCGE501[T] from Mexico [20], and STM6155 was thus tentatively included in the same species. Among described alpha-rhizobial symbionts of *M. pudica* (*R. etli* bv. *mimosae*, *R. tropici* and *R. mesoamericanum*), *R. mesoamericanum* is the most frequently detected species, with a distribution on different continents (Central & South America, Asia) [17, 20]. In Mexico, endemic *Mimosa* spp. growing in weakly acidic, neutral or slightly alkaline soil are preferentially nodulated by Alphaproteobacterial rhizobia, including

Table 1 Classification and general features of *Rhizobium mesoamericanum* STM6155 in accordance with the MIGS recommendations [56] published by the Genome Standards Consortium [57]

MIGS ID	Property	Term	Evidence code[a]
	Classification	Domain Bacteria	TAS [58]
		Phylum *Proteobacteria*	TAS [59, 60]
		Class *Alphaproteobacteria*	TAS [59, 61]
		Order *Rhizobiales*	TAS [62]
		Family *Rhizobiaceae*	TAS [63]
		Genus *Rhizobium*	TAS [15]
		Species *mesoamericanum*	TAS [15, 20]
	Gram stain	Negative	IDA
	Cell shape	Rod	IDA
	Motility	Motile	IDA
	Sporulation	Non-sporulating	NAS
	Temperature range	Mesophile	NAS
	Optimum temperature	28°C	NAS
	pH range; Optimum	7.0	TAS [15, 20]
	Carbon source	Varied; includes mannitol	TAS [15, 20]
MIGS-6	Habitat	Soil, root nodule on host	TAS [15]
MIGS-6.3	Salinity	Up to 1.5% but not 3% NaCl (w/v)	TAS [15, 20]
MIGS-22	Oxygen requirement	Aerobic	TAS [15]
MIGS-15	Biotic relationship	Free-living/symbiont	TAS [15]
MIGS-14	Pathogenicity	Non-pathogenic	NAS
	Biosafety level	1	TAS [64]
	Isolation	Root nodule of *Mimosa pudica* L.	TAS [15]
MIGS-4	Geographic location	Proximity of Mont Dore, New Caledonia	TAS [15]
MIGS-5	Sample collection	2009	TAS [15]
MIGS-4.1	Latitude	166.612297	TAS [15]
MIGS-4.2	Longitude	−22.254586	TAS [15]
MIGS-4.4	Altitude	112 m	TAS [15]

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 [65, 66]

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 4 of 11

strains of *R. mesoamericanum* [21], whereas acid-tolerant *Burkholderia* spp. are favoured microsymbionts of endemic *Mimosa* spp., including *M. pudica*, in acidic Brazilian soils [14, 22]. *R. mesoamericanum* is much less effective for nitrogen fixation on *M. pudica* than *Burkholderia phymatum* STM815 or *Cupriavidus taiwanensis* STM6070 [12, 15], and much less competitive in comparison to *B. phymatum* and *B. tuberum* [19]. These data question how *R. mesoamericanum* can maintain itself as a symbiont of *M. pudica* despite its low competitiveness. Strain STM6155 has therefore been selected as part of the DOE Joint Genome Institute 2010 *Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria* (GEBA-RNB) sequencing project [23, 24], to investigate the genome traits that enable this species to adapt to a symbiotic and saprophytic lifestyle. Here we present a summary classification and a set of general features for *R. mesoamericanum* STM6155, together with a description of its genome sequence and annotation.

Organism information
Classification and features
Rhizobium mesoamericanum STM6155 is a motile, Gram-negative, non-spore forming strain in the order *Rhizobiales* of the class *Alphaproteobacteria*. The rod-shaped form has dimensions of 0.4–0.6 μm in width and 1.0–1.4 μm in length (Fig. 1 Left and Center). It is fast growing, forming colonies within 3–4 days when grown on half strength Lupin Agar (½LA) [25], tryptone-yeast extract agar (TY) [26] or a modified yeast-mannitol agar [27] 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. mesoamericanum* STM6155 in a 16S rRNA sequence based tree. This strain is the most similar to *R. mesoamericanum* CCGE501[T] based on the 16S rRNA gene alignment, with sequence identities of 100% over 1362 bp, as determined using the EzTaxon-e database, which contains the sequences of validly published type strains [28]. Minimum Information about the Genome Sequence for STM6155 is provided in Table 1 and Additional file 1: Table S1.

Symbiotaxonomy
R. mesoamericanum STM6155 was isolated from nodules of *M. pudica*, as were others members of this species including STM3625, STM3629, tpud40a and tpud22.2 [12, 15, 17]. However, the type strain of the species, CCGE501[T], originates from nodules of *Phaseolus vulgaris* L. [20]. Strain STM6155 forms nodules and fixes N_2 with several *Mimosa* species of American origin, including *M. pudica* and *Mimosa acustipulata* Benth. It forms white, ineffective nodules on *Mimosa pigra* L. and *Mimosa caesalpinifolia* Benth. but is unable to nodulate *Mimosa scabrella* Benth. STM6155 is also able to form

nitrogen-fixing nodules on *P. vulgaris* and on a legume, *Acacia spirorbis* Labill., which grows in the same area from which STM6155 originates [15]. The symbiotic characteristics of *R. mesoamericanum* STM6155 on a range of hosts are summarised in Additional file 1: Table S2. *R. mesoamericanum* STM6155 contains a full set of nodulation genes, and exhibits uncommon features, such as the presence of two alleles of the *nodA* gene in its genome, a feature that seems conserved in several strains of the species such as STM3625 [15, 17, 29].

Table 2 Genome sequencing project information for *Rhizobium mesoamericanum* STM6155

MIGS ID	Property	Term
MIGS 31	Finishing quality	High-quality draft
MIGS-28	Libraries used	1x Illumina Std PE library
MIGS 29	Sequencing platforms	Illumina HiSeq 2000
MIGS 31.2	Fold coverage	Illumina: 279x
MIGS 30	Assemblers	Velvet version 1.1.04; Allpaths-LG version r39750
MIGS 32	Gene calling method	Prodigal 1.4
	Locus Tag	YY3 [67]
	Genbank ID	ATYY00000000
	GenBank Date of Release	15th July 2013
	GOLD ID	Gp0009783
	NCBI BIOPROJECT	163057
MIGS 13	Source Material Identifier	STM6155, WSM4584
	Project relevance	Symbiotic N_2 fixation, agriculture

Table 3 Genome statistics for *Rhizobium mesoamericanum* STM6155

Attribute	Value	% of Total
Genome size (bp)	6,927,906	100.00
DNA coding (bp)	6,004,006	86.66
DNA G + C (bp)	4,080,584	58.90
DNA scaffolds	147	
Total genes	6926	100.00
Protein coding genes	6855	98.97
RNA genes	71	1.03
Pseudo genes	0	0.00
Genes in internal clusters	1382	19.95
Genes with function prediction	5265	76.02
Genes assigned to COGs	4585	66.20
Genes with Pfam domains	5490	79.27
Genes with signal peptides	538	7.77
Genes with transmembrane helices	1529	22.08
CRISPR repeats	0	0.00

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 5 of 11

Genome sequencing information
Genome project history
This organism was selected for sequencing at the U.S. Department of Energy funded Joint Genome Institute as part of the *Genomic Encyclopedia of Bacteria and Archaea-Root Nodule Bacteria* (GEBA-RNB) project [23, 24]. The root nodule bacteria in this project were selected on the basis of environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance. The genome project is deposited in the Genomes On-Line Database [30] and a high-quality permanent draft genome sequence is

deposited in IMG [31]. Sequencing, finishing and annotation were performed by the JGI. A summary of the project information is shown in Table 2.

Growth conditions and genomic DNA preparation
Rhizobium mesoamericanum STM6155 was streaked onto TY solid medium [26] and grown at 28 °C for 3 days to obtain well grown, well separated colonies, then a single colony was selected and used to inoculate 5 ml TY broth medium. The culture was grown for 48 h on a gyratory shaker (200 rpm) at 28 °C. Subsequently 1 ml
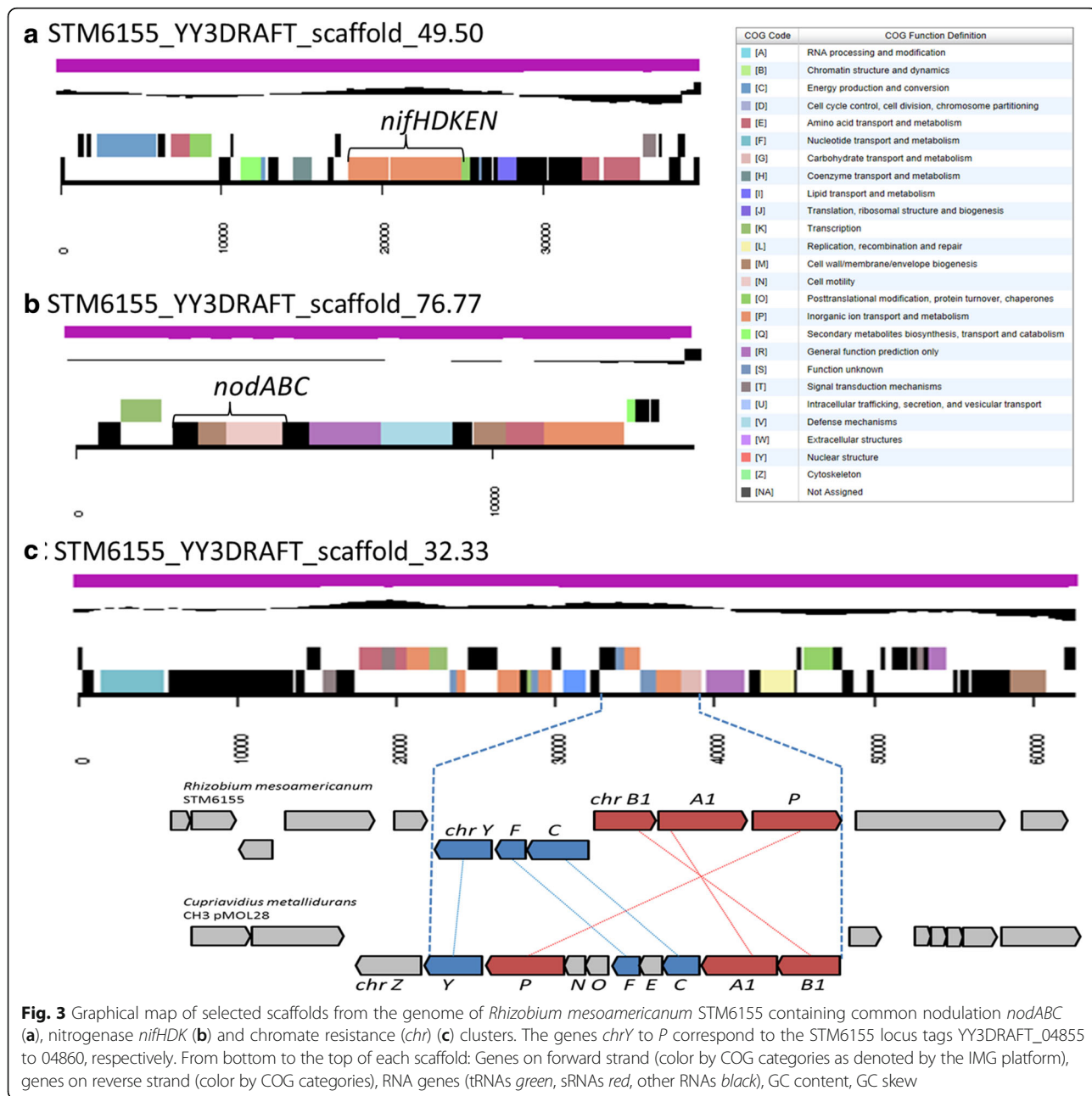

Fig. 3 Graphical map of selected scaffolds from the genome of *Rhizobium mesoamericanum* STM6155 containing common nodulation *nodABC* (**a**), nitrogenase *nifHDK* (**b**) and chromate resistance (*chr*) (**c**) clusters. The genes *chrY* to *P* correspond to the STM6155 locus tags YY3DRAFT_04855 to 04860, respectively. 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 genes of *Rhizobium mesoamericanum* STM6155 associated with general COG functional categories

Code	Value	%age	Description
J	215	4.13	Translation, ribosomal structure and biogenesis
A	0	0.00	RNA processing and modification
K	462	8.87	Transcription
L	141	2.71	Replication, recombination and repair
B	1	0.02	Chromatin structure and dynamics
D	41	0.79	Cell cycle control, cell division, chromosome partitioning
V	120	2.30	Defense mechanisms
T	236	4.53	Signal transduction mechanisms
M	283	5.43	Cell wall/membrane/envelope biogenesis
N	74	1.42	Cell motility
W	17	0.33	Extracellular structures
U	90	1.73	Intracellular trafficking, secretion and vesicular transport
O	191	3.67	Posttranslational modification, protein turnover, chaperones
C	324	6.22	Energy production and conversion
G	463	8.89	Carbohydrate transport and metabolism
E	593	11.38	Amino acid transport and metabolism
F	104	2.00	Nucleotide transport and metabolism
H	248	4.76	Coenzyme transport and metabolism
I	239	4.59	Lipid transport and metabolism
P	259	4.97	Inorganic ion transport and metabolism
Q	162	3.11	Secondary metabolites biosynthesis, transport and catabolism
R	556	10.67	General function prediction only
S	324	6.22	Function unknown
-	2341	33.80	Not in COGs

was used to inoculate 60 ml TY broth medium and the cells were incubated at 28 °C on a gyratory shaker at 200 rpm until an OD_{600nm} of 0.6 was reached. DNA was isolated from 60 ml of cells using a CTAB bacterial genomic DNA isolation method [32]. Final concentration of the DNA was set to 0.5 mg ml^{-1}.

Genome sequencing and assembly
The draft genome of *R. mesoamericanum* STM6155 was generated at the JGI using Illumina technology [33]. An

Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform which generated 14,034,164 reads totaling 2105 Mbp. All general aspects of library construction and sequencing performed at the JGI can be found on the JGI website [34]. 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), providing 12,829,288 trimmed reads totaling 1924 Mbp. The

Table 5 Percentage of Average Nucleotide Identities (ANI)[a] among *Rhizobium* genomes

Strain	CCGE501[T]	STM3625	STM6155	CFN 42[T]	Mim1	CIAT899[T]
R. mesoamericanum CCGE501[T]	—	**96.55**	**96.17**	84.28	84.6	84.69
R. mesoamericanum STM3625	**96.55**	—	**96.41**	84.4	85.19	85.03
R. mesoamericanum STM6155	**96.18**	**96.44**	—	84.45	85.31	84.97
R. etli CFN42[T]	84.25	84.4	84.42	—	**98.58**	84.45
R. etli bv. *mimosae* Mim1	84.59	85.16	85.3	98.6	—	84.71
R. tropici CIAT 899[T]	84.72	85.0	85.03	84.43	84.74	—

[a]ANI values were calculated with jSpecies (based on whole genome Mummer alignments) [68]. Genomes were downloaded from Genbank accessions when already published except *R. mesoamericanum* CCGE501[T] for which the draft genome was kindly provided by E. Martínez-Romero. Values in bold indicate values above the species cut-off (at least 95% on 69% of conserved DNA) [46]

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 7 of 11

following steps were then performed for assembly: 1) filtered Illumina reads were assembled using Velvet [35] (version 1.1.04); 2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim [36]; 3) Illumina reads were assembled with simulated read pairs using Allpaths–LG [37] (version r39750). Parameters for assembly steps were: 1) Velvet (velveth: –v –s 51 –e 71 –i 2 –t 1 –f "-shortPaired -fastq $FASTQ" –o "-ins_length 250 -min_contig_lgth 500"); 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, RunAllpathsLG: THREADS = 8 RUN = stdshredpairs TARGETS = standard VAPIWARNONLY = True OVERWRITE = True). The final draft assembly contained 152 contigs in 147 scaffolds. The total size of the genome is 6.9 Mbp and the final assembly is based on 1924 Mbp of Illumina data, which provides an average 279x coverage of the genome.

Genome annotation

Genes were identified using Prodigal [38] as part of the DOE-JGI annotation pipeline [39, 40]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information nonredundant database, UniProt, TIGRFam, Pfam, PRIAM, KEGG, COG, and InterPro databases. The tRNAScanSE tool [41] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [42]. 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 IN-FERNAL [43]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes – Expert Review platform [44] developed by the Joint Genome Institute, Walnut Creek, CA, USA. The annotated genome of *R. mesoamericanum* STM6155 is available in IMG (genome ID = 2513237088).

Genome properties

The genome is 6,927,906 nucleotides with 58.90% GC content (Table 3) and comprised of 147 scaffolds (selected scaffolds are shown in Fig. 3) of 152 contigs. The location of nodulation (Fig. 3a), nitrogenase (Fig. 3b) and chromate resistance (Fig. 3c) loci on genome scaffolds are shown.
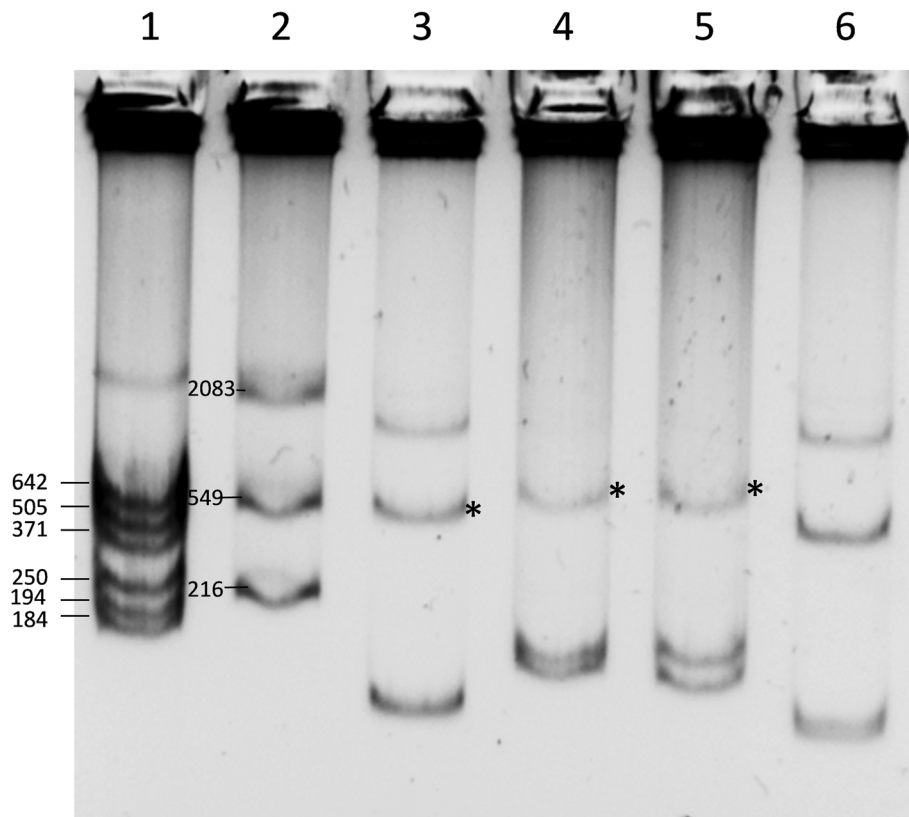

Fig. 4 Plasmid profiling of *Rhizobium* strains by the Eckhardt method. Plasmids were run on a 0.9% agarose gel at 5 Volts for 30 min then 60 Volts for 36h in a cold room. Lanes: 1: *R. etli* CFN42[T] (ladder); 2: *R. tropici* CIAT899[T] (ladder); 3: *R. mesoamericanum* STM3625 (French Guiana); 4: *R. mesoamericanum* STM3629 (French Guiana), 5: *R. mesoamericanum* STM6155 (New Caledonia); 6: *R. mesoamericanum* CCGE501[T] (Mexico). The * indicates the symbiotic plasmid

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 8 of 11

From a total of 6926 genes in the genome, 6855 were protein encoding and 71 RNA only encoding genes. The majority of genes (76.02%) 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.

Insights from the genome sequence

R. mesoamericanum STM6155 shares 100 and 99% sequence identity (over 1346 bp) to the 16S rRNA of the fully sequenced *R. mesoamericanum* type strain CCGE501[T] [45] and *R. mesoamericanum* strain STM3625 [29], respectively. Moreover the STM6155 genome shows 96.18% average nucleotide identity (ANI) (with 82% of conserved DNA), with the type strain of *R. mesoamericanum* CCGE501[T] [20], fitting with the species affiliation cut-off defined by Goris et al. (2007) [46] (Table 5).

Extended insights

We produced plasmid profiles of several *R. mesoamericanum* isolates by the Eckhardt method [47] to compare their plasmid content with genomic data. As shown in Fig. 4, the STM6155 plasmid profile differs from those of STM3625 and CCGE501[T]. Firstly, the STM6155 and STM3629 plasmid profiles suggested the absence of a 1.5 Mbp megaplasmid (P1) observed in CCGE501[T] and STM3625. The alignment of the megaplasmid P1 sequence of STM3625 with the draft genomes of STM6155 and CCGE501[T] (Fig. 5a) using progressive Mauve software [48] shows, however, the presence of P1 homologous regions in STM6155 and CCGE501[T] genomes. This suggests a putative integration of megaplasmid P1 into the bacterial chromosome in STM6155. This phenomenon was already reported in cell siblings of *Ensifer fredii* (formerly *Rhizobium* sp.) NGR234 [49]. The STM6155 plasmid profile suggests thus a diversity of genome architectures at the
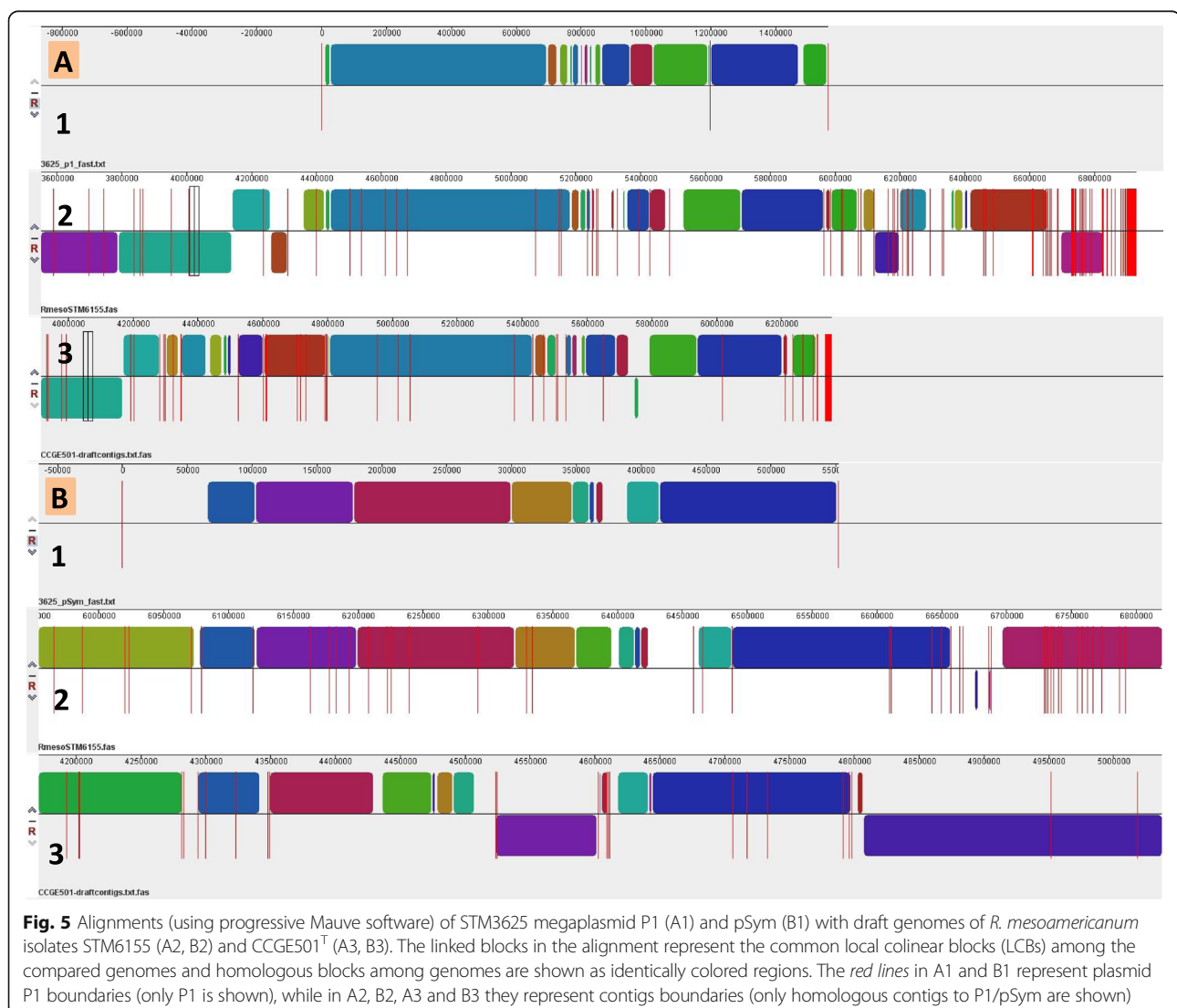

Fig. 5 Alignments (using progressive Mauve software) of STM3625 megaplasmid P1 (A1) and pSym (B1) with draft genomes of *R. mesoamericanum* isolates STM6155 (A2, B2) and CCGE501[T] (A3, B3). The linked blocks in the alignment represent the common local colinear blocks (LCBs) among the compared genomes and homologous blocks among genomes are shown as identically colored regions. The *red lines* in A1 and B1 represent plasmid P1 boundaries (only P1 is shown), while in A2, B2, A3 and B3 they represent contigs boundaries (only homologous contigs to P1/pSym are shown)

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 9 of 11

intra-species level in *R. mesoamericanum*. This diversity is observed among isolates originating from different continents like STM6155 (New Caledonia) and STM3625 (French Guiana), but also among isolates from the same country like STM3625 and STM3629 (both from French Guiana) [15, 17]. Secondly, Fig. 4 shows that STM6155 harbors a ca. 500 Kbp symbiotic plasmid (pSym) of a slightly larger size than those of STM3625 and CCGE501[T]. The alignment of the STM3625 pSym with the draft genomes of STM6155 and CCGE501[T] (using progressive Mauve, Fig. 5b) confirms the observed pSym size difference, with the presence of additional genomic regions in the STM3625 pSym. Althabegoiti and colleagues [45] have previously observed that there is only 61.4% of conserved DNA (with ANI of 98.07%) between the pSyms of CCGE501[T] and STM3625. Here we can extend this observation to the STM6155 pSym, which differs from both STM3625 and CCGE501[T] pSyms.
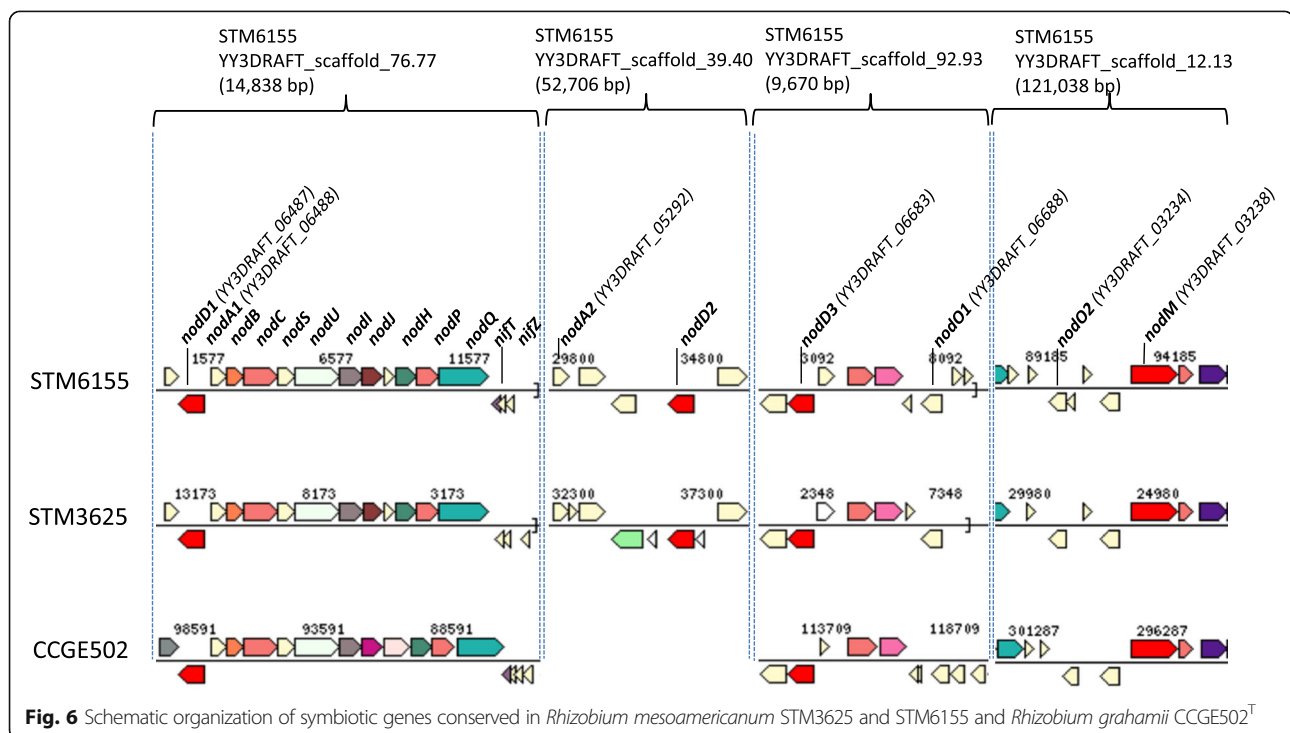
Despite the sequence diversity of the pSyms within *R. mesoamericanum* isolates, the STM6155 symbiosis nodulation genes are highly conserved with those of STM3625 and CCGE501[T]. The STM6155 nodulation genes include *nodA1BCSUIJHPQ*, an additional *nodA* (*nodA2*) gene, three *nodD* (*nodD1, 2* and *3*) transcriptional regulator genes, *nodM*, and 2 *nodO* (*nodO1, 2*) genes. The gene order is also conserved in *R. grahamii* CCGE502[T] but this strain does not contain the *nodA2* allele (Fig. 6).

Strain STM6155 was isolated from a nodule of *M. pudica* growing in ultramafic soil at a pH near neutral (pH 6.8) that contained high concentrations of heavy metals, and the highest concentrations of bioavailable chromate among four studied sites [15]. This strain was identified as being resistant to chromate concentrations up to 0.3 mM, that is comparable with chromate tolerance of *Cupriavidus metallidurans* CH34 [15, 50, 51]. Chromate resistance loci (*chr*) have been identified in the heavy-metal-tolerant *C. metallidurans* CH34 and we have discovered orthologs to these genes in STM6155 (Fig. 3c), that were absent from the more chromate sensitive strain *R. mesoamericanum* STM3625. MaGe [52] analysis has revealed synteny of six of the *C. metallidurans* CH34 plasmid-borne *chr* loci in STM6155. However, in contrast to CH3, the loci in STM6155 are arranged into two putative operons, *chrBAP* (locus tags YY3 DRAFT_04858 - YY3DRAFT_04860) and *chrCFY* (locus tags YY3DRAFT_04857 - YY3DRAFT_04855) located adjacent to one another on complementary strands.

Conclusions

R. mesoamericanum STM6155 is a microsymbiont of *Mimosa pudica* L. and *Phaseolus vulgaris* L. [9], both of which have centres of origin in central/south America. The genome size of STM6155 is 6.9 Mbp with 58.9% GC content. This strain forms a clique with the two other *R. mesoamericanum* strains STM3625 and CCGE501[T] based on average nucleotide identity comparisons (species cutoff above 95% on >69% of conserved DNA, as defined by Goris et al. [46]. However, the genome of STM6155 has a different architecture compared with the genomes of STM3625 and CCGE501[T], with STM6155 lacking a

Fig. 6 Schematic organization of symbiotic genes conserved in *Rhizobium mesoamericanum* STM3625 and STM6155 and *Rhizobium grahamii* CCGE502[T]

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 10 of 11

megaplasmid (P1) and containing a different sized pSym and small plasmid. Although STM6155 has a larger pSym, there is a notable symbiotic *nod* gene conservation between the three *R. mesoamericanum* strains, which is also shared with *Rhizobium grahamii* CCGE502[T] [20]. However, the genomes of the *R. mesoamericanum* strains contain two *nodA* alleles whereas *R. grahamii* CCGE502[T] genome has only one. Within the STM6155 genome, we have identified a *chr* chromate efflux gene cluster of six genes arranged into two putative operons and we postulate that this cluster is important for the survival of STM6155 in ultramafic soils containing high concentrations of chromate. The availability of sequenced genomes of *R. mesoamericanum* should provide further insights into rhizobial biogeographic distribution and should enable free-living and symbiotic attributes to be compared with those *Mimosa* symbioses induced by beta-rhizobia.

Additional file

Additional file 1: Table S1. Associated MIGS record for STM6155. **Table S2.** Nodulation and N$_2$ fixation properties of *Rhizobium mesoamericanum* STM6155 on selected legume hosts. (DOCX 50 kb)

Abbreviations
½LA: Half strength lupin agar; ANI: Average nucleotide identity; GEBA-RNB: Genomic encyclopedia for bacteria and Archaea-root nodule bacteria; IMG: Integrated microbial genomes; TY: Tryptone-yeast extract agar

Authors' contributions
LM supplied the strain, AK and LM the background information for this project and AK, JA, LM, TR and WR drafted the manuscript. TR provided the DNA to the JGI and performed all imaging, MB and NB provided financial support and ALL, MG, DM, MH, TBKR, NV, TW, VM, NI, RS and NK were involved in sequencing the genome and/or editing the final paper. All authors read and approved the final manuscript.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]IRD, Cirad, Univ. Montpellier, Interactions Plantes Microorganismes Environnement (IPME), 34394 Montpellier, France. [2]IRD, UMR LSTM-Laboratoire des Symbioses Tropicales et Méditerranéennes, 34398 Montpellier cedex 5, France. [3]School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia. [4]Curtin University Sustainability Policy Institute, Curtin University, Bentley, WA, Australia. [5]DOE Joint Genome Institute, Walnut Creek, CA, USA. [6]Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. [7]Department of Biology, Faculty of Science, University of Jeddah, Jeddah, Saudi Arabia. [8]Department of Biological Sciences, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia.

References
1. Simon MF, Grether R, de Queiroz LP, Särkinen TE, Dutra VF, Hughes CE. The evolutionary history of *Mimosa* (Leguminosae): toward a phylogeny of the sensitive plants. Am J Bot. 2011;98:1201–21.
2. Pacific Island Ecosystems at Risk (PIER) *Mimosa pudica*. http://www.hear.org/Pier/species/mimosa_pudica.htm. Accessed 21 Dec 2016.
3. Ganaie SU, Abbasi T, Abbasi SA. Green synthesis of silver nanoparticles using an otherwise worthless weed *Mimosa* (*Mimosa pudica*): Feasibility and process development toward shape/size control. Particul Sci Technol. 2015;33:638–44.
4. Ahmad H, Sehgal S, Mishra A, Gupta R. *Mimosa pudica* L. (Laajvanti): an overview. Pharmacogn Rev. 2012;6:115–25.
5. Joseph B, George J, Mohan J. Pharmacology and traditional uses of *Mimosa pudica*. Int J Pharm Sci Drug Res. 2013;5:41–4.
6. Visoottiviseth P, Francesconi K, Sridokchan W. The potential of Thai indigenous plant species for the phytoremediation of arsenic contaminated land. Environ Pollut. 2002;118:453–61.
7. UNDP 2015. http://www.undp.org/content/undp/en/home/sdgoverview/post-2015-development-agenda.html. Accessed 21 Dec 2016.
8. Moulin L, Munive A, Dreyfus B, Boivin-Masson C. Nodulation of legumes by members of the beta-subclass of *Proteobacteria*. Nature. 2001;411:948–50.
9. Gyaneshwar P, Hirsch AM, Moulin L, Chen WM, Elliott GN, Bontemps C, Estrada-de Los Santos P, Gross E, Dos Reis FB, Sprent JI, et al. Legume-nodulating betaproteobacteria: diversity, host range, and future prospects. Mol Plant Microb Interact. 2011;24:1276–88.
10. Bournaud C, de Faria SM, Dos Santos JMF, Tisseyre P, Silva M, Chaintreuil C, Gross E, James EK, Prin Y, Moulin L. *Burkholderia* species are the most common and preferred nodulating symbionts of the *Piptadenia* group (tribe Mimoseae). PLoS One. 2013;8:e63478.
11. Beukes CW, Venter SN, Law IJ, Phalane FL, Steenkamp ET. South African papilionoid legumes are nodulated by diverse *Burkholderia* with unique nodulation and nitrogen-fixation loci. PLoS One. 2013;8:e68406.
12. Barrett CF, Parker MA. Coexistence of *Burkholderia*, *Cupriavidus*, and *Rhizobium* sp. nodule bacteria on two *Mimosa* spp. in Costa Rica. Appl Environ Microb. 2006;72:1198–206.
13. Chen WM, Moulin L, Bontemps C, Vandamme P, Bena G, Boivin-Masson C. Legume symbiotic nitrogen fixation by beta-proteobacteria is widespread in nature. J Bacteriol. 2003;185:7266–672.
14. dos Reis Jr FB, Simon MF, Gross E, Boddey RM, Elliott GN, Neto NE, de Fatima LM, de Queiroz LP, Scotti MR, Chen W-M, et al. Nodulation and nitrogen fixation by *Mimosa* spp. in the Cerrado and Caatinga biomes of Brazil. New Phytol. 2010;186:934–46.
15. Klonowska A, Chaintreuil C, Tisseyre P, Miché L, Melkonian R, Ducousso M, Laguerre G, Brunel B, Moulin L. Biodiversity of *Mimosa pudica* rhizobial symbionts (*Cupriavidus taiwanensis*, *Rhizobium mesoamericanum*) in New Caledonia and their adaptation to heavy metal-rich soils. FEMS Microbiol Ecol. 2012;81:618–35.
16. Liu X, Wei S, Wang F, James EK, Guo X, Zagar C, Xia LG, Dong X, Wang YP. *Burkholderia* and *Cupriavidus* spp. are the preferred symbionts of *Mimosa* spp. in southern China. FEMS Microbiol Ecol. 2012;80:417–26.
17. Mishra RP, Tisseyre P, Melkonian R, Chaintreuil C, Miche L, Klonowska A, Gonzalez S, Bena G, Laguerre G, Moulin L. Genetic diversity of *Mimosa pudica* rhizobial symbionts in soils of French Guiana: investigating the origin and diversity of *Burkholderia phymatum* and other beta-rhizobia. FEMS Microbiol Ecol. 2012;79:487–503.
18. Elliott GN, Chou J-H, Chen W-M, Bloemberg GV, Bontemps C, Martínez-Romero E, Velázquez E, Young JPW, Sprent JI, James EK. *Burkholderia* spp. are the most competitive symbionts of *Mimosa*, particularly under N-limited conditions. Environ Microbiol. 2009;11:762–78.
19. Melkonian R, Moulin L, Béna G, Tisseyre P, Chaintreuil C, Heulin K, Rezkallah N, Klonowska A, Gonzalez S, Simon M, et al. The geographical patterns of symbiont diversity in the invasive legume *Mimosa pudica* can be explained by the competitiveness of its symbionts and by the host genotype. Environ Microbiol. 2014;16:2099–111.
20. López-López A, Rogel-Hernández M, Barois I, Ortiz Ceballos AI, Martínez J, Ormeño-Orrillo E, Martínez-Romero E. *Rhizobium grahamii* sp. nov. from *Dalea*

Klonowska *et al. Standards in Genomic Sciences* (2017) 12:7

Page 11 of 11

leporina, *Leucaena leucocephala*, *Clitoria ternatea* nodules, and *Rhizobium mesoamericanum* sp. nov. from *Phaseolus vulgaris*, siratro, cowpea and *Mimosa pudica* nodules. Int J Syst Evol Microbiol. 2012;62:2264–71.

21. Bontemps C, Rogel MA, Wiechmann A, Mussabekova A, Moody S, Simon MF, Moulin L, Elliott GN, Lacercat-Didier L, Dasilva C, et al. Endemic *Mimosa* species from Mexico prefer alphaproteobacterial rhizobial symbionts. New Phytol. 2016;209:319–33.

22. Stopnisek N, Bodenhausen N, Frey B, Fierer N, Eberl L, Weisskopf L. Genus-wide acid tolerance accounts for the biogeographical distribution of soil *Burkholderia* populations. Environ Microbiol. 2014;16:1503–12.

23. Reeve WG, Ardley J, Tian R, Eshragi L, Yoon JW, Ngamwisetkun P, Seshadri R, Ivanova NN, Kyrpides NC. A genomic encyclopedia of the root nodule bacteria: assessing genetic diversity through a systematic biogeographic survey. Stand Genom Sci. 2015;10:14.

24. Seshadri R, Reeve WG, Ardley JK, Tennessen K, Woyke T, Kyrpides NC, Ivanova NN. Discovery of novel plant interaction determinants from the genomes of 163 Root Nodule Bacteria. Sci Rep. 2015;5:16825.

25. Howieson JG, Ewing MA, D'Antuono MF. Selection for acid tolerance in *Rhizobium meliloti*. Plant Soil. 1988;105:179–88.

26. Beringer JE. R factor transfer in *Rhizobium leguminosarum*. J Gen Microbiol. 1974;84:188–98.

27. Vincent JM. A manual for the practical study of the root-nodule bacteria. International Biological Programme. Oxford: Blackwell Scientific Publications; 1970.

28. Kim O, Cho Y, Lee K, Yoon S, Kim M, Na H, Park S, Jeon Y, Lee J, Yi 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.

29. Moulin L, Mornico D, Melkonian R, Klonowska A. Draft genome sequence of *Rhizobium mesoamericanum* STM3625, a nitrogen-fixing symbiont of *Mimosa pudica* isolated in French Guiana (South America). Genome Announc. 2013;1:e00066–00012.

30. Reddy TBK, 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. 2015;43:D1099–106.

31. Markowitz VM, Chen IA, Palaniappan K, Chu K, Szeto E, Pillay M, Ratner A, Huang J, Woyke T, Huntemann M, et al. IMG 4 version of the integrated microbial genomes comparative analysis system. Nucleic Acids Res. 2014;42:D560–67.

32. Joint Genome Institute Protocols. http://jgi.doe.gov/user-program-info/pmo-overview/protocols-sample-preparation-information/. Accessed 21 Dec 2016.

33. Bennett S. Solexa Ltd. Pharmacogenomics. 2004;5:433–38.

34. Joint Genome Institute. http://jgi.doe.gov/. Accessed 21 Dec 2016.

35. Zerbino DR. Using the velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics. 2010;11(11):15.

36. GitHub - lh3/wgsim: Reads simulator. https://github.com/lh3/wgsim. Accessed 21 Dec 2016.

37. 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 U S A. 2011;108:1513–518.

38. Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.

39. Chen IMA, Markowitz VM, Chu K, Anderson I, Mavromatis K, Kyrpides NC, Ivanova NN. Improving microbial genome annotations in an integrated database context. PLoS One. 2013;8:e54859.

40. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen IM-A, Pati A, et al. The standard operating procedure of the DOE-JGI microbial genome annotation pipeline (MGAP v.4). Stand Genom Sci. 2015;10:86.

41. 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.

42. Pruesse E, Quast C, Knittel K, Fuchs BM, 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;35:7188–196.

43. Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–935.

44. 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;25:2271–78.

45. Althabegoiti MJ, Ormeño-Orrillo E, Lozano L, Torres Tejerizo G, Rogel MA, Mora J, Martínez-Romero E. Characterization of *Rhizobium grahamii* extrachromosomal replicons and their transfer among rhizobia. BMC Microbiol. 2014;14:6.

46. Goris J, Konstantinidis KT, Klappenbach J, Coenye T, Vandamme P, Tiedje JM. DNA-DNA hybridization values and their relationship to whole-genome sequence similarities. Int J Syst Evol Microbiol. 2007;57:81–91.

47. Eckhardt T. A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. Plasmid. 1978;1:584–88.

48. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One. 2010;5:e11147.

49. Mavingui P, Flores M, Guo X, Dávila G, Perret X, Broughton WJ, Palacios R. Dynamics of genome architecture in *Rhizobium* sp. strain NGR234. J Bacteriol. 2002;184:171–76.

50. Janssen PJ, Van Houdt R, Moors H, Monsieurs P, Morin N, Michaux A, Benotmane MA, Leys N, Vallaeys T, Lapidus A, et al. The complete genome sequence of *Cupriavidus metallidurans* strain CH34, a master survivalist in harsh and anthropogenic environments. PLoS One. 2010;5:e10433.

51. Juhnke S, Peitzsch N, Hübener N, Große C, Nies DH. New genes involved in chromate resistance in *Ralstonia metallidurans* strain CH34. Arch Microbiol. 2002;179:15–25.

52. Vallenet D, Labarre L, Rouy Z, Barbe V, Bocs S, Cruveiller S, Lajus A, Pascal G, Scarpelli C, Medigue C. MaGe: a microbial genome annotation system supported by synteny results. Nucleic Acids Res. 2006;34:53–65.

53. 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;28:2731–739.

54. Nei M, Kumar S. Molecular evolution and phylogenetics. New York: Oxford University Press; 2000.

55. Felsenstein J. Confidence limits on phylogenies: an approach using the bootstrap. Evolution. 1985;39:783–91.

56. 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;26:541–7.

57. Field D, Amaral-Zettler L, Cochrane G, Cole JR, Dawyndt P, Garrity GM, Gilbert J, Glöckner FO, Hirschman L, Karsch-Mizrachi I, et al. The genomic standards consortium. PLoS Biol. 2011;9:e1001088.

58. 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;87:4576–579.

59. Editor L. Validation of publication of new names and new combinations previously effectively published outside the IJSEM. IntJ Syst Evol Micr. 2005;55:2235–238.

60. Chen WX, Wang ET, Kuykendall LD. The *Proteobacteria*. New York: Springer - Verlag; 2005.

61. 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. 2nd ed. New York: Springer - Verlag; 2005.

62. Kuykendall LD. Order VI. *Rhizobiales* ord. nov. In: Garrity GM, Brenner DJ, Kreig NR, Staley JT, editors. Bergey's manual of systematic bacteriology. 2nd ed. New York: Springer - Verlag; 2005. p. 324.

63. Kuykendall LD. Family I. *Rhizobiaceae*. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Bergey's manual of systematic bacteriology. New York: Springer - Verlag; 2005.

64. Biological Agents: Technical rules for biological agents. http://www.baua.de/en/Topics-from-A-to-Z/Biological-Agents/TRBA/TRBA.html. Accessed 26 Dec 2016.

65. 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;25:25–9.

66. Guide to GO Evidence Codes. http://geneontology.org/page/guide-go-evidence-codes. Accessed 21 Dec 2016.

67. *Rhizobium mesoamericanum* STM6155 Genome sequencing and assembly. http://www.ncbi.nlm.nih.gov/bioproject/?term=YY3. Accessed 21 Dec 2016.

68. Richter M, Rosselló-Móra R. Shifting the genomic gold standard for the prokaryotic species definition. Proc Natl Acad Sci U S A. 2009;106:19126–131.