High-Quality draft genome sequence of the *Lotus* spp. microsymbiont *Mesorhizobium loti* strain CJ3Sym

Wayne Reeve1*, John Sullivan2, Clive Ronson2, Rui Tian1, Christine Munk3, Cliff Han3, T.B.K. Reddy4, Rekha Seshadri4, Tanja Woyke4, Amrita Pati4, Victor Markowitz5, Natalia Ivanova4 and Nikos Kyrpides4,6

Abstract

*Mesorhizobium loti* strain CJ3Sym was isolated in 1998 following transfer of the integrative and conjugative element ICEMlSymR7A, also known as the R7A symbiosis island, in a laboratory mating from the donor *M. loti* strain R7A to a nonsymbiotic recipient *Mesorhizobium* strain CJ3. Strain CJ3 was originally isolated from a field site in the Rocklands range in New Zealand in 1994. CJ3Sym is an aerobic, Gram-negative, non-spore-forming rod. This report reveals the genome of *M. loti* strain CJ3Sym currently comprises 70 scaffolds totaling 7,563,725 bp. The high-quality draft genome is arranged in 70 scaffolds of 71 contigs, contains 7,331 protein-coding genes and 70 RNA-only encoding genes, and is part of the GEBA-RNB project proposal.

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

Introduction

*Mesorhizobium loti* strain CJ3Sym was first described in work that showed that the symbiotic genes of *M. loti* strain R7A (a field reisolate of culture collection strain ICMP3153) were located on a large transmissible symbiosis island that could be transferred to nonsymbiotic mesorhizobia both in the laboratory and the environment [1, 2]. The symbiosis island was later classified as an integrative and conjugative element and renamed ICEMlSymR7A [3]. CJ3Sym was derived from a nonsymbiotic *Mesorhizobium* strain CJ3 by transfer of the symbiosis island from R7A in a laboratory mating experiment. The CJ3Sym progenitor strain CJ3 was a nonsymbiotic *Mesorhizobium* strain that was isolated from the rhizosphere of a *Lotus corniculatus* L. bird’s-foot trefoil cv. Grasslands Goldie (here after referred to as *Lotus corniculatus* cv. Grasslands Goldie) plant taken from a field site in the Rocklands range, Central Otago, New Zealand in 1994, near where ICEMlSymR7A was discovered [4]. The study was initiated to locate nonsymbiotic rhizobia that were postulated to be the likely progenitors of the diverse symbiotic strains that had received the symbiosis island through horizontal gene transfer at the field site.

Seven strains (CJ1 to CJ7) which had a similar colony morphology to *M. loti*, but which could not nodulate *Lotus corniculatus* cv. Grasslands Goldie and lacked *nod* and *nif* genes were isolated. The strains were shown to be closely related to the diverse symbiotic strains from the site by RFLP analysis, whole genome DNA-DNA hybridization analysis, full 16S rRNA gene sequencing and multilocus enzyme electrophoresis. The seven strains fell into four genomic species of nonsymbiotic mesorhizobia with strains CJ3, CJ1, CJ4 and CJ6 belonging to the same genomic species as the diverse symbiotic isolates.

When strains CJ1 to CJ7 were characterized it was noticed that they grew poorly, and only formed microcolonies after prolonged incubation on defined G/RDM agar media, in comparison to growth on rich YMA media. Auxanographic analysis revealed that all 7 strains were auxotrophic for thiamin and biotin and all but CJ5 were auxotrophic for nicotinate. In contrast to CJ3, strain CJ3Sym is prototrophic for all three vitamins and consistent with this the genes required for their biosynthesis are located on ICEMlSymR7A [5]. The CJ3Sym
sequence confirms that these are the only operons for the biosynthesis of the three vitamins in the genome.

**Organism information**

**Classification and features**

*Mesorhizobium loti* strain CJ3Sym is in the order *Rhizobiales* of the class *Alphaproteobacteria*. Cells are described as non-sporulating, Gram-negative, non-encapsulated, rods (Fig. 1 Left). The rod-shaped form varies in size with dimensions of 0.25-0.5 μm in width and 1.25-1.5 μm in length (Fig. 1 Left and Right). It forms 2 mm diameter colonies within 6 days and has a mean generation time of approximately 8 h when grown in TY broth at 28 °C [2]. Colonies on G/RDM agar [6] and half strength Lupin Agar (½LA) [7] are opaque, slightly domed, mucoid with smooth margins (Fig. 1 Right).

Strains of this organism are able to tolerate a pH range between 4 and 10. Carbon source utilization and fatty acid profiles of *M. loti* have been described previously [8–10]. Minimum Information about the Genome Sequence (MIGS) is provided in Table 1 and Additional file 1: Table S1.

Figure 2 shows the phylogenetic neighborhood of *M. loti* strain CJ3Sym in a 16S rRNA gene sequence based tree. This strain has 99.8 % (1,364/1,366 bp) 16S rRNA gene sequence identity to *M. loti* R88B (GOLD ID: Gi08827) and 99.6 % sequence identity (1,361/1,366 bp) to *M. australicum* WSM2073 (GOLD ID: Gc02468). *M. loti* strain R88B is a diverse symbiotic strain isolated from the same field site as CJ3Sym, confirming the close relationship between symbiotic and nonsymbiotic mesorhizobia isolated from the site. It is interesting to note that both of these strains cluster with *Mesorhizobium shangri-lense*, several *Mesorhizobium ciceri* strains and the type *M. loti* strain LMG 6125 (NZP2213) whereas *M. loti* strains R7A, NZP2037 and MAFF303099 form a separate cluster that shares only 98 % 16S rRNA gene sequence identity with CJ3Sym and R88B.

**Symbiotaxonomy**

*Mesorhizobium* sp. strain CJ3Sym was isolated from a laboratory mating experiment in which ICEMISymR7A was transferred from the donor strain R7A to the nonsymbiotic *Mesorhizobium* strain CJ3 [2]. The nonsymbiotic strain CJ3 was isolated from the rhizosphere of a *Lotus corniculatus* cv. Grasslands Goldie plant located at a field site that was an undeveloped tussock (*Festuca novae-zealandiae* and *Chionochloa rigida*) grassland located at an elevation of 885 m in Lammermoor, the Rocklands range, Otago, New Zealand in 1994 [4]. The soil was a dark brown silt loam with an acid pH (4.9) and a low (0.28 %) total nitrogen content [11]. CJ3 existed as a soil saprophyte that lacked symbiotic DNA. CJ3Sym forms effective nodules on *L. corniculatus* cv. Grasslands Goldie but has not yet been tested on any other *Lotus* species or ecotypes.

**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 project at the U.S. Department of Energy, Joint Genome Institute. The genome project is deposited in the Genomes OnLine Database [12] and a high-quality permanent draft genome sequence in IMG [13]. Sequencing, finishing and annotation were performed by the JGI using state of the art sequencing technology [14]. A summary of the project information is shown in Table 2.

**Growth conditions and genomic DNA preparation**

*M. loti* strain CJ3Sym was grown to mid logarithmic phase in TY rich medium [15] on a gyratory shaker at 28 °C.
DNA was isolated from 60 mL of cells using a CTAB (Cetyl trimethyl ammonium bromide) bacterial genomic DNA isolation method [16].

**Genome sequencing and assembly**

The draft genome of *M. loti* CJ3Sym was generated at the DOE Joint Genome Institute using Illumina technology [17]. An Illumina standard shotgun library was constructed and sequenced using the Illumina HiSeq 2000 platform, which generated 26,326,824 reads totaling 3,949 Mbp.

**Table 1** Classification and general features of *Mesorhizobium loti* strain CJ3Sym in accordance with the MIGS recommendations [30] published by the Genome Standards Consortium [31]

| MIGS ID | Property                  | Term                          | Evidence code*                |
|---------|---------------------------|-------------------------------|--------------------------------|
|         | Classification            | Domain Bacteria               | TAS [32]                      |
|         |                           | Phylum Proteobacteria         | TAS [23, 33]                  |
|         |                           | Class Alphaproteobacteria     | TAS [34]                      |
|         |                           | Order Rhizobiales             | TAS [35]                      |
|         |                           | Family Phyllobacteriaceae     | TAS [36]                      |
|         |                           | Genus Mesorhizobium           | TAS [9]                       |
|         |                           | Species *Mesorhizobium loti*  | TAS [8]                       |
|         |                           | Strain CJ3Sym                | TAS [2]                       |
|         | 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         | Unknown                       | NAS                            |
|         | Carbon source             | various                       | TAS [9]                       |
|         | Energy source             | chemoorganotroph              | TAS [9]                       |
|         | MIGS-6                    | Habitat                       | Soil, root nodule, host       | TAS [8] |
|         | MIGS-6.3                  | Salinity                      | Unknown                       | NAS    |
|         | MIGS-22                   | Oxygen requirement            | Aerobic                       | TAS [8] |
|         | MIGS-15                   | Biotic relationship           | Free living, Symbiotic        | TAS [8] |
|         | MIGS-14                   | Pathogenicity                 | None                          | NAS    |
|         | MIGS-4.1                  | Latitude                      | -45.864179                    | TAS [2] |
|         | MIGS-4.2                  | Longitude                     | 170.512551                    | TAS [2] |
|         | MIGS-4.3                  | Depth                         | 5-10 cm                       | IDA    |
|         | MIGS-4.4                  | Altitude                      | 50 m                          | IDA    |
|         | Isolation                 | Isolated following transfer of ICEMlSym<sup>R7A</sup> from the donor *M. loti* strain R7A to a nonsymbiotic recipient *Mesorhizobium* strain CJ3 in a laboratory mating | TAS [2] |

*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 [38, 39].

All general aspects of library construction and sequencing performed at the JGI can be found at the JGI’s web site [18]. 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). The following steps were then performed for assembly: (1) filtered Illumina reads were assembled using Velvet [19] (version 1.1.04), (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using
wgsim [20], (3) Illumina reads were assembled with simulated read pairs using Allpaths–LG [21] (version r41043). Parameters for assembly steps were: 1) Velvet –v –s 51 –e 71 –i 4 –t 1 –f ”-shortPaired -fastq $FASTQ” –o ”-ins_length 250 -min_contig_lgth 500”), 2) wgsim (−e 0–1 100–2 100 -r 0 -R 0 -X 0), 3) Allpaths–LG (STD_1,project,assembly,fragment,1,200,35,,inward,0,0). The final draft assembly contained 71 contigs in 70 scaffolds. The total size of the genome is 7.6 Mbp and the final assembly is based on 3,949 Mbp of Illumina data, which provides an average of 522x coverage of the genome.

**Genome annotation**

Genes were identified using Prodigal [22] as part of the DOE-JGI genome annotation pipeline [23], followed by a round of manual curation using the JGI GenePrimp pipeline [24]. The predicted CDSs were translated and used to search the National Center for Biotechnology Information non-redundant database, UniProt, TIGRFam, Pfam, KEGG, COG, and InterPro databases. The tRNAscanSE tool [25] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [26]. 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 [27]. Additional gene prediction analysis and manual functional annotation was performed within the Integrated Microbial Genomes-Expert Review (IMG-ER) system [28].

**Genome properties**

The genome is 7,563,725 nucleotides with 62.15 % GC content (Table 3) and is comprised of a single scaffold. From a total of 7,401 genes, 7,331 were protein encoding and 70 RNA-only encoding genes. The majority of genes (76.76 %) 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.

**Table 2** Project information

| MIGS ID | Property                | Term                          |
|---------|-------------------------|-------------------------------|
| MIGS-31 | Finishing quality       | High-quality permanent draft  |
| MIGS-28 | Libraries used          | One Illumina fragment library |
| MIGS-29 | Sequencing platforms    | Illumina HiSeq2000 technology |
| MIGS-31.2| Fold coverage          | Illumina 522x                 |
| MIGS-30 | Assemblers              | Velvet version 1.1.04, Allpaths-LG version r41043 |
| MIGS-32 | Gene calling methods    | Prodigal 1.4, GenePRIMP       |
| Locus Tag | A3A9                  |                               |
| GenBank ID | AXAL00000000          |                               |
| GenBank date of Release | September 30, 2013 |                               |
| GOLD ID | Gp00100000              |                               |
| BIOPROJECT | PRJNA165305            |                               |
| MIGS-13 | Source Material Identifier | CJ3Sym                      |
| Project relevance | Symbiotic nitrogen fixation, agriculture |
Conclusions

The *M. loti* strain CJ3Sym genome was completed to the stage where 70 scaffolds comprising 71 contigs and 7.56 Mb were obtained. A total of 7,401 genes were annotated. It is likely that the genome consists of a single chromosome and a single plasmid; however further assembly is required to confirm this. CJ3Sym is a strain that was derived from nonsymbiotic *Mesorhizobium* strain CJ3 by transfer of the symbiosis island ICE *Ml* Sym R7A from *M. loti* strain R7A in a laboratory mating experiment [2].

After the discovery of diverse *M. loti* strains containing ICE *Ml* Sym R7A at a New Zealand field site, a second adjacent field site was established and sampled to identify nonsymbiotic mesorhizobia that were the likely progenitors of the diverse symbiotic strains. Strain CJ3 was one of seven non-symbiotic *Mesorhizobium* strains isolated from the rhizosphere of *Lotus corniculatus* cv. Grasslands Goldie plants and one of the four that belonged to the same genomic species as the diverse symbiotic isolates that contained ICE *Ml* Sym R7A [4]. The genome of CJ3Sym is likely to contain a plasmid, as scaffold 17.18 contains a *trb* gene cluster (Locus tags 05060–05072 coordinates 16432–26076) and *traG* (locus tag 05072 coordinates 26704–28695) highly similar to genes on the *M. loti* strain MAFF303099 pMlb plasmid [29]. The same scaffold also contains likely plasmid replication genes.

Additional file

Table 4 Number genes associated with general COG functional categories

| Code | Value | % of total (5,809) | COG Category                                              |
|------|-------|-------------------|-----------------------------------------------------------|
| J    | 234   | 4.03              | Translation, ribosomal structure and biogenesis           |
| A    | 0     | 0.00              | RNA processing and modification                           |
| K    | 526   | 9.05              | Transcription                                             |
| L    | 139   | 2.39              | Replication, recombination and repair                      |
| B    | 5     | 0.09              | Chromatin structure and dynamics                          |
| D    | 33    | 0.57              | Cell cycle control, Cell division, chromosome partitioning|
| V    | 124   | 2.13              | Defense mechanisms                                        |
| T    | 216   | 3.72              | Signal transduction mechanisms                            |
| M    | 309   | 5.32              | Cell wall/membrane/envelope biogenesis                    |
| N    | 46    | 0.79              | Cell motility                                             |
| W    | 32    | 0.55              | Extracellular structures                                  |
| U    | 106   | 1.82              | Intracellular trafficking, secretion, and vesicular transport |
| O    | 205   | 3.53              | Posttranslational modification, protein turnover, chaperones|
| C    | 319   | 5.49              | Energy production and conversion                          |
| G    | 519   | 8.93              | Carbohydrate transport and metabolism                     |
| E    | 736   | 12.67             | Amino acid transport and metabolism                       |
| F    | 102   | 1.76              | Nucleotide transport and metabolism                       |
| H    | 274   | 4.72              | Coenzyme transport and metabolism                         |
| I    | 282   | 4.85              | Lipid transport and metabolism                            |
| P    | 286   | 4.92              | Inorganic ion transport and metabolism                    |
| Q    | 225   | 3.87              | Secondary metabolite biosynthesis, transport and catabolism|
| R    | 657   | 11.31             | General function prediction only                          |
| S    | 383   | 6.59              | Function unknown                                          |
| -    | 2,327 | 31.44             | Not in COGS                                               |

Table 4: Number genes associated with general COG functional categories

Additional file 1: Table S1. Associated MIGS record for *Mesorhizobium loti* CJ3Sym. (DOC 73 kb)

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’ contribution

JS and CR supplied the strain and background information for this project and contributed to the assembly of the manuscript with WR, TR supplied DNA to JGI and performed all imaging, WR coordinated the project and all other authors were involved in either sequencing the genome and/or editing the paper. All authors read and approved the final manuscript.

Acknowledgements

This work was performed under the auspices of the US Department of Energy’s Office of Science, Biological and Environmental Research Program, and by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC02-05CH11231, Lawrence Livermore National Laboratory under Contract No. DE-AC52-07NA27344, and Los Alamos National Laboratory under contract No. DE-AC02-06NA25396.

Author details

1Centre for Rhizobium Studies, Murdoch University, Perth, WA, Australia.
2Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand. 3Los Alamos National Laboratory, Bioscience Division, Los Alamos, NM, USA. 4DOE Joint Genome Institute, Walnut Creek, CA, USA.
References
1. Sullivan JT, Patrick HN, Lowther WL, Scott DB, Ronson CW. Nodulating strains of Rhizobium loti arise through chromosomal symbiotic gene transfer in the environment. Proc Natl Acad Sci U S A. 1995;92:8985–9.
2. Sullivan JT, Ronson CW. Evolution of rhizobia by acquisition of a 500-kb symbiosis island that integrates into a phe-tRNA gene. Proc Natl Acad Sci U S A. 1998;95:5145–9.
3. Ramsay JP, Sullivan JT, Stuart GG, Lamont IL, Ronson CW. Excision and transfer of the Mesorhizobium loti R7A symbiosis island requires an integrase IntS, a novel recombination directionality factor RdrS, and a putative relaxase RelS. Mol Biol. 2006;2:723–34.
4. Sullivan JT, Eardly BD, van Berkum P, Ronson CW. Four unnamed species of nonsymbiotic rhizobia isolated from the rhizosphere of Lotus corniculatus. Appl Environ Microbiol. 1996;62:2818–25.
5. Sullivan JT, Trezbiatowski JR, Cruickshank RW, Gozy J, Brown SD, Elliot RM, et al. Comparative sequence analysis of the symbiosis island of Mesorhizobium loti strain R7A. J Bacteriol. 2002;184:3086–95.
6. Ronson CW, Nixon BT, Albright LM, Ausbel FM. Rhizobium melloti ntrA (pAM) gene is required for diverse metabolic functions. J Bacteriol. 1987;169:2424–31.
7. Howieson JG, Evins MA, D’Antuono MF. Selection for acid tolerance in Rhizobium meliloti. Plant Soil. 1988;105:79–88.
8. Jarvis BDW, Panikhurst CE, Patel JJ. Rhizobium loti, a new species of legume root nodule bacteria. Int J Syst Bacteriol. 1982;32:378–80.
9. Jarvis BDW, Van Berkum P, Chen WX, Sour SM, Fernandez MP, Cleyet-Marechal JC, et al. Transfer of Rhizobium loti. Rhizobium huakuii, Rhizobium ciceri, Rhizobium mediterraneum, Rhizobium tianhansean to Mesorhizobium gen. nov. Int J Syst Evol Microbiol. 1997;47:895–9.
10. Tighe SW, de Lajudie P, Dipietro K, Lindstrom K, Nick G, Jarvis BDW. Analysis of cellular fatty acids and phenotypic relationships of Agrobacterium, Bradyrhizobium, Mesorhizobium, Rhizobium and Sinorhizobium species using the Sherlock Microbial Identification System. Int J Syst Evol Microbiol. 2000;50:787–801.
11. Chapman HM, Lowther WL, Trainor KD. Some factors limiting the success of Lotus corniculatus in hill and high country. Proc N Z Grassl Assoc. 1989;51:147–50.
12. Pagani I, Liolios K, Jansson J, Chen IM, Smirnova T, Nosrat B, et al. The Genomes Online Database (GOLD) v4: status of genomic and metagenomic projects and their associated metadata. Nucleic Acids Res. 2012;40:D571–579.
13. Markowitz VM, Chen I-M, 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.
14. 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.
15. Beringer JE. R factor transfer in Rhizobium leguminosarum. J Gen Microbiol. 1974;84:188–98.
16. Protocols and sample preparation information [http://www.jgi.doe.gov/collaborate-with-jgi/prmo-overview/protocols-sample-preparation-information/]
17. Bennett S. Solexa Ltd. Pharmacogenomics. 2004;5:433–8.
18. JGI: Joint Genome Institute [http://www.jgi.doe.gov]
19. Zerbino DR. Using the Velvet de novo assembler for short-read sequencing technologies. Current Protocols in Bioinformatics 2010, Chapter 11: Unit 11 15.
20. Reads simulator wgsim [http://github.com/lh3/wgsim]
21. Green S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108:1513–8.
22. 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.