High-quality draft genome sequence of *Ensifer meliloti* Mlalz-1, a microsymbiont of *Medicago laciniata* (L.) Miller collected in Lanzarote, Canary Islands, Spain

Wan Adnawani Meor Osman 1, Peter van Berkum 2, Milagros León-Barrios 3, Encarna Velázquez 4, Patrick Elia 2, Rui Tian 1, Julie Ardley 1, Margaret Gollagher 5, Rekha Seshadri 6, T. B. K. Reddy 6, Natalia Ivanova 6, Tanja Woyke 6, Amrita Pati 7, Victor Markowitz 7, Mohamed N. Baeshen 8, Naseebh Nabeeh Baeshen 8, Nikos Kyrpides 6 and Wayne Reeve 1

Abstract

*Ensifer meliloti* Mlalz-1 (INSDC = ATZD00000000) is an aerobic, motile, Gram-negative, non-spore-forming rod that was isolated from an effective nitrogen-fixing nodule of *Medicago laciniata* (L.) Miller from a soil sample collected near the town of Guatiza on the island of Lanzarote, the Canary Islands, Spain. This strain nodulates and forms an effective symbiosis with the highly specific host *M. laciniata*. 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) sequencing project. Here the features of *E. meliloti* Mlalz-1 are described, together with high-quality permanent draft genome sequence information and annotation. The 6,664,116 bp high-quality draft genome is arranged in 99 scaffolds of 100 contigs, containing 6314 protein-coding genes and 74 RNA-only encoding genes. Strain Mlalz-1 is closely related to *Ensifer meliloti* IAM 12611 T, *Ensifer medicae* A 321 T and *Ensifer numidicus* ORS 1407 T, based on 16S rRNA gene sequences. gANI values of ≥98.1% support the classification of strain Mlalz-1 as *E. meliloti*. Nodulation of *M. laciniata* requires a specific *nodC* allele, and the *nodC* gene of strain Mlalz-1 shares ≥98% sequence identity with *nodC* of *M. laciniata*-nodulating *Ensifer* strains, but ≤93% with *nodC* of *Ensifer* strains that nodulate other *Medicago* species. Strain Mlalz-1 is unique among sequenced *E. meliloti* strains in possessing genes encoding components of a T2SS and in having two versions of the adaptive acid tolerance response *lpiA-acvB* operon. In *E. medicae* strain WSM419, *lpiA* is essential for enhancing survival in lethal acid conditions. The second copy of the *lpiA-acvB* operon of strain Mlalz-1 has highest sequence identity (>96%) with that of *E. medicae* strains, which suggests genetic recombination between strain Mlalz-1 and *E. medicae* and the horizontal gene transfer of *lpiA-acvB*.

Keywords: Root-nodule bacteria, *Ensifer*, Geba-Rnb, *Medicago*, *lpiA-acvB* operon

Introduction

Symbiotic nitrogen fixation by pasture legumes and their associated root nodule bacteria provides a critical contribution to sustainable animal and plant production, and the maintenance of soil fertility in agricultural systems [1–3]. As such, it is of direct relevance to maintaining environmentally sustainable high agricultural yields, which significantly contributes to 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’ [4]. Medics (*Medicago* spp.) are some of the most important and extensively grown pasture legumes and their specific symbiosis with strains of rhizobia belonging to either *Ensifer* (synonym *Sinorhizobium*) *meliloti* or the closely related species *E. medicae* [5, 6] has been the subject of extensive research efforts [7].

* Correspondence: W.Reeve@murdoch.edu.au
1School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia
Full list of author information is available at the end of the article

© The Author(s). 2017 Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The Creative Commons Public Domain Dedication waiver (http://creativecommons.org/publicdomain/zero/1.0/) applies to the data made available in this article, unless otherwise stated.
Medicago laciniata (L.) Miller (cut leaf medic), an annual native of southern and eastern Mediterranean and Saharo-Sindian countries, is of importance because of its ability to grow in comparatively arid habitats and marginal cropping areas [8–11]. It is highly specific in its rhizobial requirements, forming a symbiosis only with a restricted subset of *E. meliloti* and not with strains that nodulate *Medicago sativa* L. (alfalfa) or *Medicago truncatula* Gaertn. [12, 13]. This symbiotic specificity has been linked to the rhizobial nod genes, in particular a specific nodC allele [14]. For example, van Berkum and colleagues found that most rhizobial strains isolated from Tunisian *M. truncatula* and *M. laciniata* shared chromosomal identity, but differed in their nodC alleles [15]. Based on these and other differing symbiotic traits, Villegas et al. [13] proposed two biovars within *E. meliloti*: bv. medicaginis for *Ensifer* strains that are symbiotically efficient on *M. laciniata* and bv. meliloti for the classical *E. meliloti* group that efficiently nodulates *M. sativa*. However, in subsequent studies the diversity observed within bv. medicaginis strains indicate that this group is certainly heterogeneous [16].

*M. laciniata* is native to the Canary Islands and is present on all of the islands of this archipelago, growing in environments that range from arid to subhumid. *Ensifer meliloti* strain Mlalz-1 was isolated from a N\(_2\)-fixing nodule of *M. laciniata* grown in alkaline soil (pH 9.0) collected in Guatiza, in the arid Northeast of Lanzarote Island, in 2007. This strain was one of the rhizobial genomes sequenced as part of the DOE Joint Genome Institute 2010 GEBA-RNB project proposal [17, 18]. Here an analysis of the complete genome sequence of *E. meliloti* Mlalz-1 is provided.

### Organism information

#### Classification and features

*E. meliloti* Mlalz-1 is a motile, non-sporulating, non-encapsulated, Gram-negative strain in the class **Alpha-proteobacteria**. The rod shaped form has dimensions of approximately 0.5 μm in width and 1.0–2.0 μm in length (Fig. 1* Left and Center*). It is fast growing, forming colonies after 3–5 days when grown on ½LA, TY, or a modified yeast-mannitol agar [19] at 28 °C. Colonies on ½LA are opaque, slightly domed and moderately mucoid with smooth margins (Fig. 1* Right*). Minimum Information about the Genome Sequence (MIGS) for strain Mlalz-1 is provided in Table 1 and Additional file 1: Table S1.

#### Symbiotaxonomy

*M. laciniata* is a highly specific host and its microsymbionts also appear to be highly specific since studies of *Medicago* isolates have shown that *M. laciniata* strains fail to nodulate a range of *Medicago* species [5, 12]. Bailly et al. [20] reported that isolates of *M. laciniata* nodulated and fixed nitrogen with *M. truncatula*, but also provided evidence that these were the progeny of horizontal transfer of the nodulation genes. Strain Mlalz-1 nodulates and is effective for nitrogen fixation with *M. laciniata*. We report here that strain Mlalz-1 is unable to nodulate *Medicago polymorpha* L., the definitive host for *E. medicae* strains [6].

#### Extended feature descriptions

Previous studies using multilocus sequence typing showed that *M. laciniata* rhizobia did not form a distinct chromosomal group [15]. Phylogenetic analysis of strain Mlalz-1 was performed by aligning the 16S rRNA sequence (1389 bp from scaffold 84.85) to the 16S rRNA gene sequences of *Ensifer* type strains (Fig. 2). Based on four variable sites within this 16S rRNA gene sequence alignment, strain Mlalz-1 is closely related to *E. meliloti* IAM 12611\(^T\) (= LMG 6133\(^T\)) [21], *E. medicae* A 321\(^T\) (= LMG 19920\(^T\)) [6] and *E. numidicus* ORS 1407\(^T\) [22]. The available IMG 16S rRNA sequence of strain Mlalz-1 gave alignment identities of 100% to *E. meliloti* IAM 12611\(^T\), 99.7% to *E. medicae* A 321\(^T\) and 99.5% to *E. numidicus* ORS 1407\(^T\). In contrast, *E. meliloti* IAM

---

**Fig. 1** Images of *Ensifer meliloti* Mlalz-1 using scanning (Left \(a\)) and transmission (Center \(b\)) electron microscopy as well as light microscopy to visualize colony morphology on solid media (Right \(c\))
1261\textsuperscript{T} and Ensifer terangae LMG 7834\textsuperscript{T} [23] were only 97.3\% similar.

**Genome sequencing information**

**Genome project history**

E. meliloti Mlalz-1 was selected for sequencing at the U.S. Department of Energy funded Joint Genome Institute as part of the GEBAB-RNB project [17, 18]. The root nodule bacteria in this project were selected based on environmental and agricultural relevance to issues in global carbon cycling, alternative energy production, and biogeochemical importance. In particular, strain Mlalz-1 was chosen since it has strict host specificity for M. laciniata, which is suited for cultivation in arid environments [11]. The E. meliloti Mlalz-1 genome project is deposited in the Genomes Online Database [24] and a high-quality permanent draft genome sequence (IMG Genome ID 2513237143) is deposited in IMG [25]. 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**

E. meliloti Mlalz-1 (= USDA 1984) was cultured on MAG solid media [26] for three days at 28 °C to obtain well grown, well separated colonies, then a single colony was selected from the plate and inoculated into 5 ml MAG broth media. The culture was grown for 48 h on a gyratory shaker (200 rpm) at 28 °C. Subsequently 1 ml was used to inoculate 50 ml of MAG and the cells were
incubated on a gyratory shaker (200 rpm) at 28 °C until an OD 600nm of 0.6 was reached. DNA was isolated from 50 ml of cells by Peter van Berkum according to the method described by van Berkum [26]. The final concentration of the DNA was set to 0.5 mg ml$^{-1}$.

Genome sequencing and assembly

The draft genome of *E. meliloti* Mlalz-1 was generated at the DOE Joint genome Institute (JGI) using Illumina technology [27]. An Illumina standard PE library was constructed and sequenced using the Illumina HiSeq 2000 platform that generated 35,720,836 reads totalling 4983 Mbp. All general aspects of library construction and sequencing were done at the JGI and details can be found on the JGI website [28]. All raw Illumina sequence data was passed through DUK, a filtering program developed at JGI, which removes known Illumina sequencing and library preparation artefacts (Mingkun L, Copeland A, Han J; unpublished). The following steps for assembly were: (1) filtered Illumina reads were assembled using Velvet (version 1.1.04) [29]; (2) 1–3 Kbp simulated paired end reads were created from Velvet contigs using wgsim (version 0.3.0) [30]; (3) Illumina reads were
assembled with simulated read pairs using Allpaths–LG (version r39750) [31]. Parameters for the assembly steps were 1) Velvet: -v -s 51 -e 71 -i 2 -t 1 -f “-shortPaired -fastq $FASTQ” -o “-ins_length 250 -min_contig_lgth 500” for Velvet and 2) wgsim: -e 0 -1 76 -2 76 -r 0 -R 0 -X 0. The final draft assembly contained 100 contigs in 99 scaffolds. The total size of the genome is 6.7 Mbp and the final assembly is based on 4983 Mbp of Illumina data, which provides an average of 748× coverage of the genome.

**Genome annotation**

Genes were identified using Prodigal [32], as part of the DOE-JGI genome annotation pipeline [33, 34]. 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 [35] was used to find tRNA genes, whereas ribosomal RNA genes were found by searches against models of the ribosomal RNA genes built from SILVA [36]. 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 RNA profiles using INFERNAL [37]. Additional gene prediction analysis and manual functional annotation was done within the Integrated Microbial Genomes-Expert Review platform [38] developed by the Joint Genome Institute, Walnut Creek, CA, USA.

**Genome properties**

The genome is 6,664,116 bp with 62.16% GC content (Table 3) and comprised of 99 scaffolds. From a total of 6388 genes, 6314 were protein encoding and 74 RNA only encoding genes. Most genes (79.52%) 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**

*E. meliloti* Mlalz-1 is one of seven strains of *E. meliloti* that have been sequenced from the GEBAY-RNB genome sequencing projects [17]. On the basis of 16S tRNA sequence identity, strain Mlalz-1 is closely related to *E. meliloti* IAM 12611T (=LMG 6133T), *E. medicae* A321T (=LMG 19920T) and *E. numidicus* ORS 1407T. As the genomes of these type strains have not been sequenced or are not publically available, gANI values [39]

### Table 4 Number of genes of *Ensifer meliloti* Mlalz-1 associated with the general COG functional categories

| Code | Value | %age | Description |
|------|-------|------|-------------|
| J    | 217   | 4.09 | Translation, ribosomal structure and biogenesis |
| A    | 0     | 0.00 | RNA processing and modification |
| K    | 466   | 8.77 | Transcription |
| L    | 122   | 2.3  | Replication, recombination and repair |
| B    | 1     | 0.02 | Chromatin structure and dynamics |
| D    | 39    | 0.73 | Cell cycle control, cell division, chromosome partitioning |
| Y    | 0     | 0.00 | Nuclear structure |
| V    | 117   | 2.20 | Defense mechanisms |
| T    | 216   | 4.07 | Signal transduction mechanisms |
| M    | 301   | 5.67 | Cell wall/membrane/envelope biogenesis |
| N    | 72    | 1.36 | Cell motility |
| Z    | 0     | 0.00 | Cytoskeleton |
| W    | 33    | 0.62 | Extracellular structures |
| U    | 74    | 1.39 | Intracellular trafficking, secretion, and vesicular transport |
| O    | 206   | 3.88 | Posttranslational modification, protein turnover, chaperones |
| C    | 358   | 6.74 | Energy production and conversion |
| G    | 555   | 10.45 | Carbohydrate transport and metabolism |
| E    | 584   | 10.99 | Amino acid transport and metabolism |
| F    | 116   | 2.18 | Nucleotide transport and metabolism |
| H    | 242   | 4.56 | Coenzyme transport and metabolism |
| I    | 220   | 4.14 | Lipid transport and metabolism |
| P    | 279   | 5.25 | Inorganic ion transport and metabolism |
| Q    | 159   | 2.99 | Secondary metabolite biosynthesis, transport and catabolism |
| R    | 551   | 10.37 | General function prediction only |
| S    | 348   | 6.55 | Function unknown |
| X    | 36    | 0.68 | Mobilome: prophages, transposons |
| Y    | 1729  | 27.07 | Not in COGs |

---

Osman et al. Standards in Genomic Sciences (2017) 12:58
had to be compared with other fully sequenced Ensifer strains (Table 5). E. meliloti Mlalz-1 currently forms a gANI clique with other E. meliloti strains (gANI values ≥98.14%), compared with gANI values of ≤87.9% with the finished genomes of other Ensifer strains. This supports the classification of strain Mlalz-1 as an E. meliloti strain, in accordance with the defined species affiliation cut-off value of 96.5% gANI [39]. The total genome size of strain Mlalz-1 is 6.6 Mb, which falls within the expected size range of 6.6–8.9 Mb for E. meliloti. The genome architecture of E. meliloti consists of a chromosome and the two symbiotic megaplasmids pSymA and pSymB [20]. Replication of a plasmid is initiated by the replication protein encoded by repC, which is present as a single copy on E. meliloti pSymA and pSymB. The E. meliloti Mlalz-1 genome carried 2 repC loci (A3CADRAFT_00120 and A3CADRAFT_01676) with highest encoded protein identity to RepC proteins of E. meliloti strains. Mlalz-1 A3CADRAFT_00120 and Mlalz-1 A3CADRAFT_01676 RepC2 had high identity (99.00%) to the RepC2 protein encoded by SMb20044 on pSymB of E. meliloti 1021. E. meliloti Mlalz-1 1021. E. meliloti Mlalz-1 A3CADRAFT_01676 RepC2 had highest identity (98.10%) to the RepC1 protein encoded by SMa2391 on pSymA of E. meliloti 1021. This indicated the presence of two megaplasmids in strain Mlalz-1, and that strain Mlalz-1 has a similar genome architecture to that of E. meliloti 1021.

Extended insights
All 29 E. meliloti strains within the gANI clique share a core set of 4948 orthologous genes, using cut off values of 1e-5 and 30% minimum protein identity. E. meliloti Mlalz-1 contains 176 unique genes, 96 (54.5%) of which encode hypothetical proteins. The unique genes include those encoding the components of a T2SS, located on scaffold A3CADRAFT_scaffold_5.6 (Fig. 3a), as well as genes that encode a DNA methyltransferase and a NifT/TauT family transport system. These T2SS components form part of a unique COG profile generated for Mlaz-1 (Table 6). The T2SS secretion system is used to translate a wide range of proteins from the periplasm across the outer membrane [40]. Although T2SS genes are not found in other E. meliloti strains or in the Ensifer fredii strains GR64 and USDA 257, they are present in the genomes of the E. fredii strains HH103 and NGR234, in a similar gene arrangement to that observed in E. meliloti Mlalz-1 [41, 42] (Fig. 3b). Generally, the T2SS gene cluster is comprised of 12–15 genes, and strain Mlalz-1 contains the 12 required genes gspDOGLMCKEFHIJ necessary for a functional T2SS, but lacks the gspS gene found only in certain genera [43] (Fig. 3c).

In common with some other E. meliloti strains, strain Mlalz-1 contains several genes encoding phage components. The PHASTER algorithm [44] was used to identify two resident prophages, present on scaffold A3CADRAFT_scaffold_4.5: one that was incomplete (Prophage Region 1) and one that was intact (Prophage Region 2) (Fig. 4). The proteins encoded by Prophage Region 1 (11.4 kb) and Prophage Region 2 (55 kb) were most closely related to the phage proteins of PHAGE_Mycobacterium_Catalina_NC031328 and PHAGE_Sinorhizobium_Medicae_NC_029046, respectively.

The Mlaz-1 genome also contains acid-tolerance or acid-responsive genes that are orthologous to the genes identified in the comparatively acid tolerant strain E. medicae WSM419. Acid-tolerance or acid-responsive genes identified in Mlaz-1 include actA (Int), actB, actR, actS, phrR, exoR, exoH, lpiA, acvB, degP1, mdh3, fbaB, groS, kdpB, kdpC, fixN2 and fixO2 [45–52] (Additional file 2: Table S2). It is notable that strain Mlalz-1 is unique among the sequenced Ensifer strains since it contains two versions of the highly acid-induced lpiA-acvB operon. One operon (A3CADRAFT_01189-A3CADRAFT_01190) is found on scaffold A3CADRAFT_scaffold_3.4, in a gene region that is conserved in other E. meliloti (sequence similarity >98%) and is located on the

### Table 5 Pairwise gANI comparisons of selected finished genomes of sequenced Ensifer strains

| Strain       | Gold ID: Gp | Casida A | USDA 257 | WSM 419 | 1021  | AK83  | BL225C | GR4   | Mlalz-1 | Rm41  | SM11  |
|--------------|-------------|----------|-----------|----------|-------|-------|--------|-------|---------|-------|-------|
| Eadhaeens    | 0094824     | 100      | 80.5      | 79.06    | 80.12 | 80.11 | 80.06  | 80.01 | 80.08   | 80.03 | 80.06 |
| E. fredii    | 0005169     | 80.5     | 100       | 81.89    | 83.26 | 83.24 | 83.25  | 83.20 | 83.14   | 83.33 | 83.22 |
| E. medicae   | 0000117     | 79.06    | 81.93     | 100      | 88.18 | 88.13 | 88.26  | 88.24 | 87.90   | 88.14 | 88.26 |
| E. meliloti  | 00000726    | 80.12    | 83.26     | 88.19    | 100   | 99.36 | 99.62  | 99.41 | 98.80   | 99.24 | 99.43 |
| E. meliloti  | 0006695     | 80.08    | 83.25     | 88.16    | 99.56 | 100   | 99.33  | 99.14 | 98.60   | 99.38 | 99.33 |
| E. meliloti  | 0006560     | 80.06    | 83.25     | 88.28    | 99.62 | 99.33 | 100    | 99.44 | 98.81   | 99.26 | 99.39 |
| E. meliloti  | 0020501     | 80.01    | 83.23     | 88.26    | 99.41 | 99.14 | 99.43  | 100   | 98.81   | 99.05 | 99.25 |
| E. meliloti  | 0010229     | 80.11    | 83.15     | 87.91    | **98.80** | 98.59 | **99.81** | 98.81 | **100** | **98.59** | **98.66** |
| E. meliloti  | 0025853     | 80.05    | 83.36     | 88.11    | 99.26 | 99.39 | 99.25  | 99.06 | **98.59** | 100   | **99.33** |
| E. meliloti  | 0006018     | 80.05    | 83.23     | 88.29    | 99.45 | 99.33 | 99.39  | 99.26 | **98.67** | 99.32 | 100   |

For E. meliloti Mlalz-1, gANI values above the microbial species delineation cutoff value of 96.5% [39] are in bold font
chromosome of the fully sequenced E. meliloti 1021. The second version of the lpiA-acvB operon (A3CADRAFT_05694-A3CADRAFT_05695) is located on A3CADRAFT_scaffold_47.48, in a gene region that is conserved in E. medicae genomes (sequence similarity >96%) and is located on the pSMED02 symbiotic plasmid of the fully sequenced E. medicae WSM419. The regulatory gene fsrR, required for the acid activated expression of lpiA in E. medicae WSM419 [53], is located upstream of A3CADRAFT_05694 in strain Mlalz-1. This regulatory gene is absent from the A3CADRAFT_01190 gene region, and from the lpiA-acvB gene regions of all other E. meliloti sequenced genomes. These findings suggest that E. meliloti Mlalz-1 acquired the plasmid-borne lpiA-acvB operon and associated fsrR regulatory gene by lateral transfer from an E. medicae strain.

Essential symbiotic (nod, nif and fix) genes identified in the E. meliloti Mlalz-1 genome (Additional file 2: Table S3 and S4) are located in several clusters on the following scaffolds: A3CADRAFT_scaffold_54.55 (Fig. 5a), A3CADRAFT_scaffold_61.62 (Fig. 5b), A3CADRAFT_scaffold_63.64 (Fig. 5c), A3CADRAFT_scaffold_71.72 (Fig. 5d)
and A3CADRAFT_scaffold_74.75 (Fig. 5e). Nodulation of
*M. laciniata* has been shown to require a specific *nodC* allele [14]. The *nodC* gene of strain Mla-z-1 has highest
sequence identity (≥ 98%) with *nodC* of other *M. laciniata-*nodulating *Ensifer* strains in the NCBI database, whereas
there is a lower sequence identity (≤ 93%) with *nodC* of *Ensifer* strains that nodulate other *Medicago*
species. Nodulation of *Medicago* hosts requires Nod factors that are
sulfated at the reducing terminus and acylated at the non-
reducing terminus, with a polyunsaturated fatty acyl tail
[54, 55]. The NodH sulfotransferase, together with the
NodP and NodQ sulfate-activating complex, are required
for Nod factor sulfation [56, 57]. Activity of NodL results in
O-acetylation of the Nod factor [58], while NodE and NodF
produce the specific polyunsaturated fatty acyl tail [55, 59].

Strain Mla-z-1 would appear to be typical of *Ensifer* strains
that nodulate *Medicago* species since the *nodEF*, *nodL*, and
*nodHPQ* genes that are required for these specific decorations
of the Nod factor are present in the genome. *E. meliloti* Mla-z-1 also possesses the three *nodD* genes that
mediate host-specific activation of *nodABC* in the symbiotic
interactions of *E. meliloti* with *Medicago* [60].

### Conclusions

*E. meliloti* Mla-z-1 is a rhizobial strain that is able to
nodulate and fix nitrogen with the highly specific host
*M. laciniata*. Although the 16S rRNA gene sequence di-
vergence was insufficient to differentiate strain Mla-z-1
from *E. meliloti*, *E. medicae* or *E. numidicus*, a gANI
value of 98.8% with the genome of *E. meliloti* 1021,
Fig. 5 Graphical map of the scaffolds: a A3CADRAFT_scaffold_54.55, b A3CADRAFT_scaffold_61.62, c A3CADRAFT_scaffold_63.64, d A3CADRAFT_scaffold_71.72 and e A3CADRAFT_scaffold_74.75 of Ensifer meliloti Mlalz-1 showing the location of common nodulation (nod) and fixation (nif and fix) genes within the symbiotic regions of this strain. From bottom to the top of the scaffold map: Genes on reverse strand (color by COG categories as denoted by the IMG platform), genes on forward strand (color by COG categories), RNA genes (tRNAs green, sRNAs red, other RNAs black), GC content, GC skew.
compared with 87.9% with the genome of *E. medicae* WSM419 identifies strain Mlalz-1 as *E. meliloti*. Nodulation of *M. laciniata* has been shown to be dependent on the presence of a specific *nodC* allele, which also is present in the genome of *E. meliloti* Mlalz-1, based on a 98% sequence identity with the *nodC* of other *M. laciniata*-nodulating *Ensifer* strains [14]. However, strain Mlalz-1 is unique among sequenced *E. meliloti* strains in possessing genes encoding components of a T2SS and in having two versions of the adaptive acid tolerance response *IpaA-acvB* operon. The second copy of the *E. meliloti* Mlalz-1 *IpaA-acvB* operon has highest sequence identity (>96%) with that of sequenced *E. medicae* strains, which infers horizontal gene transfer of this region from *E. medicae*.

**Additional files**

Additional file 1: Table S1. Associated MIGS record for *Ensifer meliloti* Mlalz-1. (DOCX 52 kb)

Additional file 2: Table S2-S4. Table S2. Acid responsive gene orthologs present in *Ensifer* strains. Table S3. The nodulation genes of *Ensifer meliloti* Mlalz-1. (DOCX 65 kb)

**Abbreviations**

1½LA: Half strength Lupin Agar; gANI: Genome-wide average nucleotide identity; GEBA-RNB: Genomic Encyclopedia for Bacteria and Archaea-Root Nodule Bacteria; IMG: Integrated Microbial Genomes; T2SS: Type II Secretion System; TY: Tryptone-yeast extract

**Acknowledgements**

We thank Gordon Thompson (Murdoch University) for the preparation of SEM and TEM photos. MLB thanks Alfredo Reyes-Betancourt, from the Orotava Botanical Garden (Tenerife), for providing *M. laciniata* seeds.

**Funding**

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 gratefully acknowledge the funding received from the Curtin University Sustainability Policy Institute and from Murdoch University’s Small Research Grants Scheme in 2016.

**Authors’ contributions**

MLB and EV isolated the strain and provided project metadata. PvB supplied the strain, the DNA and the background information for this project and participated in drafting the manuscript. PE curated the strain and performed sequence analysis of 16S rRNA and *nodC* genes. RT supplied DNA to JGI and performed all imaging. JA provided symbiotic phenotype data. WAMO, JA and WR performed bioinformatics analyses and drafted the paper, Mb and NB provided financial support, and MG, RS, TBKR, NI, TW, AP, VM 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.

**Publisher’s Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

1School of Veterinary and Life Sciences, Murdoch University, Murdoch, WA, Australia. 2U.S. Department of Agriculture, Soybean Genomics and Improvement Laboratory, Beltsville Agricultural Research Center, 10300 Baltimore Avenue, Bldg. 606, Beltsville, MD 20705, USA. 3Departamento de Bioquímica, Microbiología, Biología Celular y Genética, Universidad de La Laguna, Tenerife, Spain. 4Departamento de Microbiología y Genética e Instituto Hispanoluso de Investigaciones Agrarias (CIALÉ), Universidad de Salamanca, Salamanca, Spain. 5Curtin University Sustainability Policy Institute, Curtin University, Bentley, WA, Australia. 6DOE Joint Genome Institute, Walnut Creek, CA, USA. 7Biological Data Management and Technology Center, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. 8Department of Biology, Faculty of Science, University of Jeddah, Jeddah, Saudi Arabia.

Received: 30 March 2017 Accepted: 14 September 2017

Published online: 25 September 2017

**References**

1. Carlsson G, Huss-Deal K. Nitrogen fixation in perennial forage legumes in the field. Plant Soil. 2003;253:353–72.

2. Peoples MB, Brockwell J, Hunt JR, Swan AD, Watson L, Hayes RC, Li GD, Hackney B, Nutall J, Davies SL, Filey IRP. Factors affecting the potential contributions of *N* fixation by legumes in Australian pasture systems. Crop Pasture Sci. 2012;63:759–86.

3. Unkovich MJ, Baldock J, Peoples MB. Prospects and problems of simple linear models for estimating symbiotic *N* fixation by crop and pasture legumes. Plant Soil. 2010;329:75–89.

4. UNDP. 2015 http://www.undp.org/content/undp/en/home/sdgoverview/post-2015-development-agenda.html. Accessed 22 Sept 2017.

5. Béna G, Lyet A, Huguet T, Olivieri I. *Medicago–Sinorhizobium* symbiotic specificity evolution and the geographic expansion of *Medicago*. J Evol Biol. 2005;18:1547–58.

6. Rome S, Fernandez MP, Brunel B, Normand P, Cleyet-Mareel JC. *Sinorhizobium medicae* sp. nov., isolated from annual *Medicago* spp. Int J Syst Bacteriol. 1996;46:972–80.

7. Jones KM, Kobayashi H, Davies BW, Taga ME, Walker GC. How rhizobial symbions invade plants: the *Sinorhizobium–Medicago* model. Nature Rev Microbiol. 2007;5:619–33.

8. Jordan DC. Reduction of the nodulation barrier in *Medicago laciniata* by alteration of the root temperature. Plant Soil. 1981;61:93–111.

9. Small E; Alfalfa and relatives: Evolution and classification of *Medicago*. Ottawa: NRC Research Press; 2010.

10. Young RR, Croft PH, Sandral GA. Variation in flowering times and agronomic characteristics of *Medicago laciniata* (L.) Miller collected from diverse locations in new South Wales, Aust J Exp Agric. 1992;32:59–63.

11. Youssi N, Shem N, Ramazi A, Abbeld C. Growth, photosynthesis and water relations as affected by different drought regimes and subsequent recovery in *Medicago laciniata* (L.) populations. J Plant Biol. 2016;59:33–43.

12. Brockwell J, Hely FW. Symbiotic characteristics of *Rhizobium meliloti*: an appraisal of the systematic treatment of nodulation and nitrogen fixation interactions between hosts and *rhizobia* of diverse origins. Aust J Agr Econ. 1966;17:885–9.

13. Villegas MDC, Rome S, Mauré L, Domergue Q, Gardon L, Bailly X, Cleyet-Mareel J-C, Brunel B. Nitrogen-fixing *sinorhizobia* with *Medicago laciniata* constitute a novel biovar (bv. Medicago) of *S. meliloti*. Syst Appl Microbiol. 2006;29:36–38.

14. Baran LR, Bromfield ES, Brown DC. Identification and cloning of the bacterial nodulation specificity gene in the *Sinorhizobium meliloti-Medicago laciniata* symbiosis. Can J Microbiol. 2002;48:765–71.

15. van Berkum P, Badri Y, Elia P, Aouani ME, Eardy BD. Chromosomal and symbiotic relationships of *rhizobia* nodulating *Medicago truncatula* and *M. laciniata*. Appl Environ Microbiol. 2007;73:7597–604.

16. Mhatre B, Badri Y, Saidi S, de Lajudie P, Mhamdi R. Symbiotic diversity of *Ensifer meliloti* strains recovered from various legume species in Tunisia. Syst Appl Microbiol. 2009;32:583–92.

17. Reeves W, 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 Genomic Sci. 2015;10:1.

18. Seshadri R, Reeves 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.

19. Howieson JG, Dilworth MJ, editors. Working with Rhizobia. Canberra, Australia: Australian Centre for International Agricultural Research (ACIAR); 2016.
Bailly X, Olivier I, Brunel B, Cleyet-Marel JC, Béna G. Horizontal gene transfer and homologous recombination drive the evolution of the nitrogen-fixing symbionts of \textit{Medicago} species. J Bacteriol. 2007;189:5223–36.

Delajudie P, Willems A, Pot B, Devittinck D, Maestrojuan G, Neyra M, Collins MD, Dreyfus B, Kersters K, Gillis M. Polyphasic taxonomy of \textit{rhizobium} emendation of the genus \textit{Sinorhizobium} and description of \textit{Sinorhizobium melloti comb. nov}, \textit{Sinorhizobium trophimus sp. nov}, and \textit{Sinorhizobium targas sp. nov.} Int J Syst Bacteriol. 1994;44:715–33.

Young JM. The genus name \textit{Ensifer} Casida 1982 takes priority over \textit{Sinorhizobium} Chen et al. 1988, and \textit{Sinorhizobium marelense} Wang et al. 2002 is a later synonym of \textit{Ensifer adhaerens} Casida 1982. Is the combination "\textit{Sinorhizobium adhaerens}" (Casida 1982) Willems et al. 2003 legitimate? Request for an opinion. Int J Syst Evol Microbiol. 2003;53:2107–20.

Merabet C, Martens M, Mahdhi M, Zahkia F, Sy A, Le Roux C, Domergue O, Neyra M, Collins MD, Dreyfus B, Kersters K, Gillis M. Polyphasic taxonomy of \textit{rhizobium} emendation of the genus \textit{Sinorhizobium} and description of \textit{Sinorhizobium melloti comb. nov}, \textit{Sinorhizobium trophimus sp. nov}, and \textit{Sinorhizobium targas sp. nov.} Int J Syst Bacteriol. 1994;44:715–33.

Rebby TBK, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Markowitz VM, Chen IM, 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:D560–70.

van Berkum P. Evidence for a third uptake hydrogenase phenotype among the soybean Bradyrhizobium. Appl Environ Microbiol. 1990;56:3835–41.

Bennett S, Solexa Ltd. Pharmacogenomics. 2004;5:433.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall GS, 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;108:15131–8.

Zerbo DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genomes Proteomes Biol. 2009;7:101.

Gilbert J, Glöckner FO, Hirschman L, Karsch-Mizrachi I, et al. The genomic GI database context. PLoS One. 2013;8:e54859.

Chen IM, Markowitz VM, Chu K, Anderson I, Mavromatis K, Kyrpides NC, Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Páez-Espino D, Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: \textit{RNA} genes in genomic sequence. Nucleic Acids Res. 1997;25:955–62.

Krysciak D, Orbegoso MR, Schmeisser C, Streit WR. Molecular keys to broad host range \textit{Rhizobium} symbionts of \textit{Medicago} and homologous recombination drive the evolution of the nitrogen-fixing systems. Appl Environ Microbiol. 2009;75:4035–42.

Costa TR, Felisberto-Rodrigues C, Meir A, Prevost MS, Redzej A, Trokter M, Waksman G. Secretion systems in gram-negative bacteria: structural and mechanistic insights. Nature Rev Microbiol. 2015;13:343–59.

Ardt D, Grant JR, Nacu A, Sajed T, Pon A, Liang Y, Wihart DS. PHASTER: a better, faster version of the PHASTA phage search tool. Nucleic Acids Res. 2016;44:W16–21.

Tiwari RP, Reeve WG, Dhillon MJ, Glenn AR. An essential role for \textit{acrof} in acid tolerance of \textit{Rhizobium melloti}. Microbiology. 1996;142:693–704.

Tiwari RP, Reeve WG, Fenner BJ, Dhillon MJ, Glenn AR, Howison JG. Probing for \textit{phr} regulated genes in \textit{Sinorhizobium} medicae using transcriptional analysis. J Mol Microbiol Biotechnol. 2004;7:133–9.

Reddy TBK, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Markowitz VM, Chen IM, 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:D560–70.

van Berkum P. Evidence for a third uptake hydrogenase phenotype among the soybean Bradyrhizobium. Appl Environ Microbiol. 1990;56:3835–41.

Bennett S, Solexa Ltd. Pharmacogenomics. 2004;5:433.

Joint Genome Institute website [http://www.jgi.doe.gov/]. Accessed 22 Sept. 2017.

Zerbo DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genomes Proteomes Biol. 2009;7:101.

GitHub - lh3/wgsim: Reads simulator [https://github.com/lh3/wgsim]. Accessed 22 Sept. 2017.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall GS, 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;108:15131–8.

Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ. Prodigal: \textit{RNA} gene recognition and translation initiation site identification. BMC Bioinformatics. 2010;11:119.

Dénarié J, Debellé F, Promé JC, Heinemeyer I, Morgenstern B, Pommerening-Röser A, Flores M, et al. \textit{SILVA}: a comprehensive online resource for quality checked and aligned \textit{RNA} gene sequences. Nucleic Acids Res. 2005;33:563–9.

Schultze M, Staehelin C, Röhrig H, John M, Schmidt J, Kondorosi E, Schell J. \textit{Rhizobium} and \textit{Sinorhizobium} meliloti nodulation factors. J Biol Chem. 1993;268:20134–41.

Tiwari RP, Reeve WG, Fenner BJ, Dhillon MJ, Glenn AR, Rolfe BG, Djojoevica MA, Howison JG. Probing for \textit{phr}-regulated proteins in \textit{Sinorhizobium} medicae using proteomic analysis. J Mol Microbiol Biotechnol. 2004;7:140–7.

Krysciak D, Orbegoso MR, Schmeisser C, Streit WR. Molecular keys to broad host range \textit{Rhizobium} symbionts of \textit{Medicago} and homologous recombination drive the evolution of the nitrogen-fixing systems. Appl Environ Microbiol. 2009;75:4035–42.

Honma MA, Asomaning M, Ausubel FM. \textit{Rhizobium} nodulation factors. J Biol Chem. 1993;268:20134–41.

Kemble R, Glöckner FO, Hirschman L, Karsch-Mizrachi I, et al. The genomic GI database context. PLoS One. 2013;8:e54859.

Lowe TM, Eddy SR. RNAcons: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. 1997;25:965–6.

Pruesse E, Quast C, Knittel K, Böck D, Ludwig W, Peplies J, Glöckner FO. SILVA: a comprehensive online resource for quality checked and aligned \textit{RNA} sequence data compatible with ARB. Nucleic Acids Res. 2005;33:563–9.

Navrocck EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. Bioinformatics. 2013;29:2933–5.

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;25:2781–7.

Vargheese NJ, Mulkerjhe S, Ivanova N, Konstantinidis KT, Mavromatis K, Kyrpides NC, Patti A. Microbial species delineation using whole genome sequences. Nucleic Acids Res. 2015;43:6761–71.

Korotkov K, Sandkvist M, Hol WG. The type II secretion system: biogenesis, molecular architecture and mechanism. Nature Rev Microbiol. 2012;10:356–51.

Kryscia D, Opresea MR, Schmeisser C, Streit WR. Molecular keys to broad host range in \textit{Sinorhizobium} and \textit{Rhizobium} spp. Nucleic Acids Res. 2016;44:W16–21.

Schmeisser C, Lisesegur H, Kryscia D, Bakkou N, Le Quéré A, Wollherr A, Heinemeyer I, Morgenstern B, Pomerening-Röser A, Flores M, et al. \textit{Rhizobium} sp. strain NGR234 possesses a remarkable number of secretion systems. Appi Environ Microbiol. 2009;75:4035–45.
67. Woese CR, Kandler O, Wheelis ML. Towards a natural system of organisms: proposal for the domains Archaea, bacteria, and Eucarya. Proc Natl Acad Sci U S A. 1990;87:4576–9.

68. Garrity GM, Bell JA, Lilburn T. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Phylum XIV. Proteobacteria phyl. Nov. in Bergey’s manual of systematic bacteriology. Volume 2. Second edition. New York: Springer - Verlag; 2005. p. 1.

69. Garrity GM, Bell JA, Lilburn T. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Class I. Alphaproteobacteria class. In Bergey’s manual of systematic bacteriology. Second edition. New York: Springer - Verlag; 2005.

70. Euzéby J. Validation list no. 107. List of new names and new combinations previously effectively, but not validly, published. Int J Syst Evol Microbiol. 2006;56:1–6.

71. Kuykendall LD. In: Garrity GM, Brenner DJ, Krieg NR, Staley JT, editors. Order VI. Rhizobiales ord. Nov. in Bergey’s manual of systematic bacteriology. Second edition. New York: Springer - Verlag; 2005. p. 324.

72. Skerman VBD, McGowan V, Sneath PHA. Approved lists of bacterial names. Int J Syst Bacteriol. 1980;30:225–420.

73. Conn HJ. Taxonomic relationships of certain non-sporeforming rods in soil. J Bacteriol. 1938;36:320–1.

74. Casida LE. Ensifer adhaerens gen. Nov., sp. nov.: a bacterial predator of bacteria in soil. Int J Syst Evol Microbiol. 1982;32:339–45.

75. Biological Agents: Technical rules for biological agents [http://www.baua.de/en/Topics-from-A-to-Z/Biological-Agents/TRBA/TRBA.html]. Accessed 22 Sept 2017.

76. 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 Nature Genet. 2000;25:25–9.

77. Guide to GO Evidence Codes [http://geneontology.org/page/guide-go-evidence-codes]. Accessed 22 Sept 2017.

78. Locus Tag [https://www.ncbi.nlm.nih.gov/bioproject/?term=A3CA]. Accessed 22 Sept 2017.

79. GOLD ID for Ensifer meliloti Mlalz-1 [https://gold.jgi.doe.gov/projects?id=Gp0010229]. Accessed 22 Sept 2017.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at www.biomedcentral.com/submit