Data in Brief

Draft genome of *Leisingera aquaemixtae* CECT 8399T, a member of the *Roseobacter* clade isolated from a junction of fresh and ocean water in Jeju Island, South Korea

Lidia Rodrigo-Torres, María J. Pujalte, David R. Arahal *
Departamento de Microbiología y Ecología, Universidad de Valencia, Valencia, Spain
Colección Española de Cultivos Tipo (CECT), Universidad de Valencia, Valencia, Spain

**A B S T R A C T**

We report the draft genome sequence and annotation of *Leisingera aquaemixtae* CECT 8399T (DDBJ/EMBL/GenBank accession number CYSR00000000) which comprises 4,614,060 bp, 4313 protein coding genes, 54 tRNA coding genes and 7 rRNA coding genes. General findings of the annotated genome, such as pigment indigoidine operon, phenylacetate oxidation genes or predictable number of replicons, are commented in comparison to other *Leisingera* species. Average Nucleotide Identity between available genomes of type strains of species of *Leisingera* and *Phaeobacter* genera has been calculated to evaluate its current classification.

© 2016 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

**Keywords:** Rhodobacteraceae, Phaeobacter, Ectoine, Indigoidine, Phenylacetate degradation

### Specifications

| Organism/cell line/tissue | *Leisingera aquaemixtae* CECT 8399T |
|---------------------------|-----------------------------------|
| Strain                    | CECT 8399T                        |
| Sequencer                 | Illumina MiSeq                    |
| Data format               | Processed                         |
| Experimental factors      | Bacteria cells cultured in Marine Agar and DNA genomic extraction and sequencing |
| Experimental features     | Draft genome sequence of *Leisingera aquaemixtae* CECT 8399T, assembly and annotation |
| Consent                   | Reads and contig sequences and annotation are publicly available |
| Sample source location    | Jeju Island, South Korea, 33° 15’ 7” N, 126° 37’ 26” E. |

1. Direct link to deposited data

   [http://www.ebi.ac.uk/ena/data/view/CYSR00000000](http://www.ebi.ac.uk/ena/data/view/CYSR00000000)

The genus *Leisingera* is a member of the so-called *Roseobacter* group within the family *Rhodobacteraceae*, order *Rhodobacterales* of the class *Alphaproteobacteria*. The *Roseobacter* group is widely distributed in marine environments and contributes up to 20% of marine bacterioplankton [1]. This genus was first described by Schaefer et al. in 2002 [2], together with *Leisingera methylohalidivorans* as type species of the genus, and emended three times afterwards [3–5].

*Leisingera* and *Phaeobacter* spp. were intermixed and did not form monophyletic groups on 16S rRNA gene sequence trees. Brieder et al. (2014) [5] conducted a genome-scale study to better delineate species belonging to these two genera, and proposed the reclassification of *Phaeobacter aquaemixtae* [6] into *Leisingera*. Currently, four other species belong to the genus *Leisingera*: *L. methylohalidivorans* [2], *L. daeponensis* (formerly *Phaeobacter*) [5,7], *L. aquimarina* [4], and *L. caerulea* (formerly *Phaeobacter*) [5,8].

2. Experimental design, materials and methods

*Leisingera aquaemixtae* CECT 8399T is a Gram-negative, rod-shaped or ovoid bacterium isolated from a zone where the ocean and a freshwater spring meet at Jeju Island, South Korea. Colonies on Marine Agar are circular, smooth, convex, and circular yellowish white and some of them change to grayish as cultures ages. Optimal growth is at 30 °C, pH 7.0–7.5, and 2% (w/v) NaCl, but tolerates ranges of 10–40 °C temperature, pH 5.5–10.0 and 0.5–8% NaCl [6].

*L. aquaemixtae* CECT 8399T was cultured in marine agar (MA; Difco) at 26 °C under aerobic conditions during three days. Genomic DNA was isolated using Real Pure Spin kit (Durviz) following the standard
Protocol recommended by the manufacturer. The integrity of the extracted DNA was checked by visualization in a 2.0% agarose gel electrophoresis. Its purity and quantity was checked by measuring the absorbance at 260 and 280 nm with a spectrophotometer Nanodrop2000c (Thermo Scientific) and calculating the ratio A260/A280.

Genomic DNA was sequenced at Central Service of Support to Experimental Research (SCSIE) of the University of Valencia (Valencia, Spain) using an Illumina MiSeq platform with 2 × 250 paired-end reads. A total of 824,956 reads were obtained with 205,596,181 bp, which resulted in a sequencing coverage of 45 ×.

Reads were analyzed for quality control with the program FASTQC, developed by Babraham Bioinformatics, and wrapped in Galaxy Orione Server [9]. After filtering and trimming, the remaining reads were assembled using SPADES 3.0.0 [10] and MIRA [11]. SPADES scaffolds bigger than 1000 pb and with coverage larger than 10× were selected resulting in 53 scaffolds. MIRA contigs smaller than 1000 pb were discarded, 91 contigs remained. With these two sets of scaffolds and contigs, CISA integrator v1.0.1 [12] was used and a final set of 40 sequences was obtained. Tools used for filtering and trimming and assembly programs are also included in Galaxy Orione Web Server. The 40 contigs had a N50 of 339,773 bp and summarize 4,616,040 bp with a G + C content of 64.4%.

This draft genome was annotated with Prokka [13], within Galaxy Orione Server, and RAST v2.0 [14] using default parameters. Further analysis of annotated genome by Prokka was done with different web servers. WebMGA Server [15] was used to search for COGs, NCBI Batch CD-Search Tool [16] for Pfam domains, SignalP 4.1 Server [17] was used to predict signal peptides, TMHMM Server v2.0 [18] to predict transmembrane helix domains, antiSMASH 2.0.9 [19] to annotate secondary metabolites and CRISPRFinder [20] for finding CRISPR repeats. A total of 4313 protein coding genes, 54 tRNA coding genes and 7 rRNA coding genes were predicted by both Prokka and RAST. A resume of the genome sequence and annotation is in Table 1.

**Leisingera** species have in common a numerous quantity of replicons, thus *L. caerulea* DSM 25646T genome harbors three chromosomes [21] and *L. aquimarina* DSM 25465T seven plasmids [22]. Although the genome of *L. aquamixtae* CECT 8399T here presented is not in a complete level, some evidences show the same trend than its close relatives. Two dnaA genes coding for chromosome replication initiation proteins have been found. Three operons RepABC coding for RepC replicase and RepAB partitioning proteins, four RepAB modules, one RepA and one RepB coding genes are located in different contigs. One RepAB module is positioned in the same contig as a dnaA gene; a similar finding appeared in *L. caerulea* DSM 25646T genome [21] where genes usually located in chromosome were annotated together with a RepABC operon, suggesting a possible integration of a plasmid into the chromosome, also proposed for other members of *Roseobacter* clade [23]. RepA is encoded in this genome as sporulation inhibitor initiation protein Soj, chromosome partitioning protein, *para* ortholog, in *Bacillus subtilis* [24]. Two Post-Segregational Killing systems are also annotated in this draft genome, three ParE4 toxin and ParD4 antitoxin, located closely to a RepAB module, and a ParE1 toxin and ParD1 antitoxin which is not positioned next to any plasmid replication modules but to a conjugal coupling transfer protein TraG gene. Regarding conjugation, virB genes of the type IV secretion system were not found in this draft genome in contrast with other genomes of the genus, however, seven genes were found coding for Flp pilus assembly proteins CpaA, CpaB, TadB (2), TadD (3), four genes coding for type IV pilus biosynthesis proteins PIW(2) and PIP(2) and three genes coding for conjugation related proteins, suggesting this strain has the potential for pilus biogenesis but lacks virB genes and virD necessary to export DNA by conjugation.

Table 1

| Attribute                   | Value   | % of Total |
|-----------------------------|---------|------------|
| Genome size (bp)            | 4,614,060 | 100        |
| DNA coding (bp)             | 4,136,959 | 90.0       |
| DNA G + C (bp)              | 2,971,455 | 64.4       |
| DNA scaffolds               | 40       | 100        |
| Total genes                 | 4374     | 100        |
| Protein coding genes        | 4313     | 98.6       |
| RNA genes                   | 61       | 1.4        |
| Genes with function prediction | 3450     | 78.9       |
| Genes assigned to COGs      | 3410     | 78.0       |
| Genes with Pfam domains     | 3720     | 85.1       |
| Genes with signal peptides  | 397      | 9.1        |
| Genes with transmembrane helices | 953    | 21.8       |
| CRISPR repeats              | 0        | 0          |
synthase multienzyme complex genes are predicted. These findings suggest this strain can utilize siderophores as other Leisingera relatives do.

Fifteen Gene Transfer Agent genes have been annotated in this draft genome coding for phage genes that can transfer genomic DNA among prokaryotes. This set of genes is also found in L. daepenosensis DSM 23529\(^\mathrm{T}\) [31] and Phaeobacter inhibens T\(^{\mathrm{5}}\) genomes [32].

L. aquaemixtae CECT 8399\(^\mathrm{T}\) draft genome has revealed common traits with genomes of Leisingera species type strains and environmental and biotechnological potential. A final genome-based comparison was calculated with the recent released OrthoANI program [33] using genomes of type strains of species of Leisingera, Phaeobacter, Pseudophaeobacter and Sedimentitalea available at NCBI database. The resulting ANI values and the phylogenomic tree obtained (Fig. 1) reinforce recent reclassification carried by Breider et al. [5].

3. Nucleotide sequence accession number

The Whole Genome Shotgun project is deposited at DDBJ/EMBL/GenBank under accession number CYSR00000000.

Conflict of interest

The authors declare that there is no conflict of interests with respect to the work published in this paper.

Acknowledgments

This work was financially supported by projects TAXPROMAR 2010 (CGL2010-18134/BOS) of the Spanish Ministerio de Economía y Competitividad and GV PROMETEO/2012/040 of the Generalitat Valenciana. M.A. Ruvira is acknowledged for cultivation and DNA extraction.

References

[1] A. Buchan, J.M. González, M.A. Moran, Overview of the marine Roseobacter lineage. Appl. Environ. Microbiol. 71 (2005) 5665–5677, http://dx.doi.org/10.1128/AEM.71.10.5665–5677.2005.
[2] J.K. Schaefer, K.D. Goodwin, I.R. Mcdonald, J.C. Murrell, R.S. Oremland, Leisingera methylohalidivorans gen. nov., sp. nov., a marine methylotroph that grows on methyl bromide. Int. J. Syst. Evol. Microbiol. 52 (2002) 851–859, http://dx.doi.org/10.1099/ijs.0.01960-0 (The).
[3] T. Martens, T. Heidorn, R. Pukall, M. Simon, B.J. Tindall, T. Brinkhoff, Reclassification of Roseobacter gallaeciensis Ruiz-Ponte et al. 1998 as Phaeobacter gallaeciensis gen. nov., comb. nov., description of Phaeobacter inhibens sp. nov., reclassification of Ruegeria algicola (Lafay et al. 1995) Uchino et al. 1999 as Marinovum algicola gen. nov., comb. nov., and emended descriptions of the genera Roseobacter, Ruegeria and Leisingera. Int. J. Syst. Evol. Microbiol. 56 (2006) 1293-1304. http://dx.doi.org/10.1099/ijs.0.03724-0.
[4] I. Vandecandelaere, E. Segaert, A. Mollica, M. Faimali, P. Vandamme, Leisingera aquaemixtae sp. nov., isolated from a marine electroactive biofilm, and emended descriptions of Leisingera methylohalidivorans Schaefer et al. 2002, Pseudophaeobacter daepenosensis Yoon et al. 2007 and Phaeobacter inhibens Martens et al. 2006, Int. J. Syst. Evol. Microbiol. 58 (2008) 2788–2793, http://dx.doi.org/10.1099/ijs.0.063844-0.
[5] S. Breider, C. Scheuner, P. Schumann, A. Fiebig, J. Petersen, S. Pradella, et al., Genomesecale data suggest reclassifications in the Leisingera–Phaeobacter cluster including proposals for Sedimentitalea gen. nov. and Pseudophaeobacter gen. nov. Front. Microbiol. 5 (2014) 416, http://dx.doi.org/10.3389/fmicb.2014.00416.
[6] S. Park, D.S. Park, K.S. Bae, J.H. Yoon, Phaeobacter aquaemixtae sp. nov., isolated from the junction between the ocean and a freshwater spring. Int. J. Syst. Evol. Microbiol. 64 (2014) 1378–1383, http://dx.doi.org/10.1099/ijs.0.057646-0.
[7] J.-H. Yoon, S.-J. Kang, S.-Y. Lee, T.-K. Oh, Phaeobacter daepenosensis sp. nov., isolated from a tidal flat of the Yellow Sea in Korea. Int. J. Syst. Evol. Microbiol. 57 (2007) 856–861, http://dx.doi.org/10.1099/ijs.0.06477-0.
[8] I. Vandecandelaere, E. Segaert, A. Mollica, M. Faimali, P. Vandamme, Phaeobacter caeruleus sp. nov., a blue-coloured, colony-forming bacterium isolated from a marine electroactive biofilm. Int. J. Syst. Evol. Microbiol. 59 (2009) 1209–1214, http://dx.doi.org/10.1099/ijs.0.002642-0.
[9] G. Cuccuru, M. Orsini, A. Pinna, A. Stabellati, N. Soranzo, A. Travaglione, et al., Orione, a web-based framework for NGS analysis in microbiology. Bioinformatics 30 (2014) 1928–1929, http://dx.doi.org/10.1093/bioinformatics/btu135.
[10] A. Bankevich, S. Nurk, D. Antipov, AA. Gurevich, M. Dvorkin, AS. Kulikov, et al., SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (2012) 455–477, http://dx.doi.org/10.1089/cmb.2012.0021.
[11] B. Chevreux, T. Wetter, S. Suhai, Genome sequence assembly using trace signals and additional sequence information. Comput. Sci. Biol. Proc. Ger. Conf. Bioinforma. (1999) 45–56, http://dx.doi.org/10.1.1.23/7465.
