Genomics Data 7 (2016) 233–236

Contents lists available at ScienceDirect

Genomics Data

journal homepage: www.elsevier.com/locate/gdata

Draft genome of *Leisingera aquaemixtae* CECT 8399^T, 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

ARTICLE INFO

Article history:
Received 23 December 2015
Received in revised form 8 January 2016
Accepted 14 January 2016
Available online 15 January 2016

Keywords:
Rhodobacteraceae
Phaeobacter
Ectoine
Indigoidine
Phenylacetate degradation

ABSTRACT

We report the draft genome sequence and annotation of *Leisingera aquaemixtae* CECT 8399^T (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.

Specifications	
Organism/cell line/tissue	*Leisingera aquaemixtae* CECT 8399^T
Strain	CECT 8399^T
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 8399^T, 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

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 8399^T 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 8399^T 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

* Corresponding author at: Departamento de Microbiología y Ecología, Universidad de Valencia, Valencia, Spain.
E-mail addresses: lidia.rodrigo@uv.es (L. Rodrigo-Torres), maria.j.pujalte@uv.es (M.J. Pujalte), arahal@uv.es (D.R. Arahal).

http://dx.doi.org/10.1016/j.gdata.2016.01.010

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 4,614,060 bp with a G + C content of 64.4%.

This draft genome was annotated with Prokka [13], within Galaxy Orione Server, and RAST v.2.0 [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 [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 24564^T genome harbors three chromosomes [21] and *L. aquimarina* DSM 24565^T seven plasmids [22]. Although the genome of *L. aquaemixtae* CECT 8399^T 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 24564^T 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 subtillis* [24]. Two Post-Segregational Killing systems are also annotated in this draft genome, one 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), TadG(3), four genes coding for type IV pilus biogenesis proteins PilW(2) and PilP(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.

Type VI secretion system is present in this draft genome as in plasmids of *L. caerulea* DSM 24564^T, *L. aquimarina* DSM 24565^T and *L. methylohalidivorans* DSM 14366^T. This secretion system has been reported as a potent mediator in interbacterial interactions [25].

Colonies of *L. aquaemixtae* CECT 8399^T on Marine Agar are yellowish and turn grayish when the culture ages [6]. Pigmentation is also reported in other members of *Leisingera* genus [4,7,8] and is often related to secondary metabolite production in marine bacteria [26]. *Phaeobacter* sp. strain Y4I produces a blue pigment indigoidine via a nonribosomal peptide synthase (NRPS)-based biosynthetic pathway encoded by an operon *igiBCDFE* [27]. Indigoidine pigment biosynthesis operon was also reported to be present in *L. caerulea* DSM 24564^T and *L. daeponensis* DSM 24565^T [21,28]. Now we confirm that operon *igiRBCDFE* is also encoded in this draft genome (PHA8399_02441-6) with 92%–99% identity with operon genes in *Phaeobacter* sp. strain Y4I and also a ClpA chaperone which is thought to regulate pigment production [27]. This finding may indicate that *L. aquaemixtae* CECT 8399^T is able to produce this blue pigment in an appropriate medium.

Quorum sensing seems to play an important role in indigoidine production [29] and is usually related with antibiotic compound production [30]. AntiSMASH Server detected two homoserine lactone clusters, an ectoine cluster, a Type I PolyKetide Synthase cluster and a bacteriocin cluster. Two *luxR* genes and a *luxI* gene (coding for acyl-homoserine lactone synthase) were annotated in this draft genome; thus, this bacterium may use quorum sensing to modulate pigmentation and antibacterial compound synthesis.

Ectoine is one of the so-called compatible solutes that counteract dehydration caused by water efflux in hypersaline environments. Ectoine synthase gene *ectC* and three ectoine utilization protein coding genes *eutA*,*C*,*E* are encoded in this genome. This finding is in agreement to the slightly moderate salinity tolerance found in this strain (up to 8% NaCl) and could be of biotechnological interest.

Surprisingly, although this strain was reported to be non-motile [6], this genome harbor *fli*, *flh*, *flg* genes coding for complete flagella machinery so we suggest this observation must be revised.

This genome encodes for a complete Glycolysis pathway including, 6-phosphofructokinase, which is missing in the majority of *Roseobacter* genomes [23], but also present in *L. aquimarina* DSM 24565^T genome.

Genes for methyl halide metabolism and phenylacetate catabolism were searched. These pathways have been found in other *Leisingera* genomes [28,31] and play important roles in environment conservation. All genes for the methyl halide oxidation pathway are present with the exception of *purU* coding for a formyltetrahydrofolate deformilase, one of the essential enzymes [30]. However, all genes for phenylacetate utilization pathway are present suggesting a possible ability of aromatic compound degradation.

Different siderophore genes have been found in *L. aquaemixtae* CECT 8399^T. Four putative siderophore transport system proteins (YusV, YfhA, YfiZ, YfiY) are encoded forming an operon. Three ferric siderophore transport system, two biopolymer transport proteins ExbB and a periplasmic protein TonB are also encoded contiguously with that operon and a ferric hydroxamate uptake *fhuA* gene. Besides, two enterobactin transport regulator genes, an enterobactin exporter EntS and a phosphopantetheinyltransferase component of enterobactin

Table 1
Genome statistics (annotation by Prokka).

Attribute	Value	% of Total
Genome size (bp)	4,614,060	100
DNA coding (bp)	4,136,999	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