AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Complete Genome Sequence of the *Arcobacter canalis* Type Strain LMG 29148

William G. Miller,[a] Emma Yee,[a] Mary H. Chapman[a]

[a]Produce Safety and Microbiology Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Albany, California, USA

ABSTRACT *Arcobacter canalis* was originally recovered from shellfish and from a sewage-contaminated canal. *Arcobacter canalis* is closely related to the marine bacterium *Arcobacter marinus*. This study describes the complete whole-genome sequence of the *A. canalis* type strain LMG 29148 (=F138-33^T; =CECT 8984^T), which was recovered from oysters.

Arcobacter canalis was originally recovered from shellfish and sewage-contaminated canal water in Catalonia, Spain (1). Based on phylogenetic and genomic analyses (1), *A. canalis* was determined to be highly related to *Arcobacter marinus* (2). The original description of *A. canalis* also identified several phenotypic discriminatory markers to distinguish it from *A. marinus* and ostensibly from other *Arcobacter* spp. (1). These included, for example, nitrate reduction, catalase activity, and growth on medium containing 2% NaCl. As part of a project to obtain complete genomes for all *Arcobacter* type strains, we sequenced an *A. canalis* type strain. In this study, we report the first closed genome sequence of the *A. canalis* type strain LMG 29148 (=F138-33^T; =CECT 8984^T), which was recovered from oysters.

A. canalis strain LMG 29148^T was obtained from the BCCM/LMG culture collection and grown, both initially and in one subculture, aerobically at 30°C for 48 h on brain heart infusion agar (Thermo Fisher Scientific, Waltham, MA) amended with 5% horse blood and 2% (wt/vol) NaCl. Approximately 5 μl of cells, representing multiple individual colonies, were removed from the plate using a sterile inoculating loop. Genomic DNA was prepared from these cells using the Promega Wizard genomic DNA purification kit (Madison, WI); a single preparation of genomic DNA was used to construct the Illumina and PacBio libraries. The Illumina MiSeq library was constructed using the Illumina Nextera DNA Flex kit, and the 20-kb PacBio SMRTbell library was prepared using the SMRTbell template prep kit 1.0, following the manufacturer's instructions. Illumina sequencing was performed on a MiSeq instrument at 8.0 pM, with dual-index paired-end reads, using the MiSeq reagent kit v2 (300 cycle). PacBio sequencing was performed on an RS II sequencer, with reads assembled using the Hierarchical Genome Assembly Process (HGAP) v. 3.0 in the SMRT Analysis software v. 2.3.0. Default parameters were used for all software unless otherwise specified. Sequencing metrics are presented in Table 1. A single PacBio contig was obtained and processed using Geneious Prime v. 2019.1.3 (Biomatters Ltd., Auckland, New Zealand), as follows. The PacBio contig was circularized manually within Geneious, thus removing >99.9% of the bases with a Q score of <40 (generally those within ~8 kb of the contig ends). Then, the Illumina MiSeq reads were quality trimmed and assembled onto the circularized PacBio contig. Using the "Find Variations/SNPs" module with a default minimum variation of 0.3, 25 variations (1-bp indels) were identified and corrected to the MiSeq consensus sequence. The final coverage was 357×.

Genomic data for strain LMG 29148^T are presented in Table 1. The genome size is 2,829,476 bp, with a G+C content of 27.5%. Protein-, rRNA-, and tRNA-encoding genes

Citation Miller WG, Yee E, Chapman MH. 2019. Complete genome sequence of the *Arcobacter canalis* type strain LMG 29148. Microbiol Resour Announc 8:e01156-19. https://doi.org/10.1128/MRA.01156-19.

Editor David A. Baltrus, University of Arizona

Address correspondence to William G. Miller, william.miller@ars.usda.gov.

Received 17 September 2019
Accepted 14 October 2019
Published 31 October 2019

TABLE 1 Sequencing metrics and genomic data for *A. canalis* strain LMG 29148^T

Feature[a]	Value(s)[b]
Sequencing metrics	
Illumina MiSeq platform	
No. of reads	1,764,438
No. of bases	260,767,753
Avg length (bases)	148
Coverage (×)	92.2
PacBio platform	
No. of reads	96,414
No. of bases	749,500,177
Avg length (bases)	7,773.8
Coverage (×)	264.9
Genomic data	
Chromosome	
Size (bp)	2,829,476
G+C content (%)	27.49
No. of CDS[c]	2,679
Assigned function (% CDS)	1,017 (38.0)
General function annotation (% CDS)	1,025 (38.3)
Domain/family annotation only (% CDS)	182 (6.8)
Hypothetical (% CDS)	455 (17.0)
No. of pseudogenes	23
Genomic islands/CRISPR	
No. of genetic islands	5
No. of CDS in genetic islands	175
CRISPR/Cas locus type	I-B
Gene content/pathways	
Signal transduction	
Che proteins	*che*$(A)_2(B)_2CD(R)_2V(W)_2(Y)_4$
No. of methyl-accepting chemotaxis proteins	33
No. of response regulators	57
No. of histidine kinases	67
No. of response regulator/histidine kinase fusions	4
No. of diguanylate cyclases	23
No. of diguanylate phosphodiesterases (HD-GYP, EAL)	9 [1]
No. of diguanylate cyclase/phosphodiesterases	11
No. of other signal transduction genes	12 [1]
Motility	
Flagellin genes	*fla1–fla7*
Restriction/modification	
No. of type I systems (*hsd*)	1
No. of type II systems	0
No. of type III systems	2
Transcription/translation	
No. of transcriptional regulatory proteins	60 [1]
Non-ECF σ factors	σ^{70}
No. of ECF σ factors	0
No. of tRNAs	63
No. of ribosomal loci	6
Nitrate/nitrite reduction	*napABDFGH*, *nirA*, *nrfAH*
Nitrogen fixation (*nif*)	No
Osmoprotection	$(BCCT)_5$,[d] *ectABC*
Pyruvate → acetyl-CoA	
Pyruvate dehydrogenase (E1/E2/E3)	Yes
Pyruvate/ferredoxin oxidoreductase	*por*
Urease	No
Vitamin B_{12} biosynthesis	Yes

[a] CDS, coding sequences; ECF, extracytoplasmic function; CoA, coenzyme A.
[b] Numbers in square brackets indicate pseudogenes or fragments.
[c] Numbers do not include pseudogenes.
[d] BCCT, betaine/carnitine/choline transporter.

were identified using GeneMark, RNAmmer, and ARAGORN (3–5), respectively, and annotated as described previously (6). The data presented here support the taxonomic relatedness of *A. canalis* and *A. marinus*; the genomes of these species are highly syntenic, with an average nucleotide identity of 95.2%. However, we identified errors in

the original phenotypic discriminatory markers for both species (1, 2). Although *A. canalis* was originally described as being unable to reduce nitrate (1), the *A. canalis* genome contains a full set of nitrate reductase genes (Table 1); controlled tests performed in our laboratory (7) demonstrate that strain LMG 29148^T is able to reduce nitrate, thus confirming the genomic data. Similarly, we have demonstrated that *A. marinus* strain JCM 15502^T is catalase positive, contradicting previous descriptions (1, 2), and its genome encodes the catalase gene. Additionally, we routinely grow *A. canalis* on medium without amended NaCl (0.5% [wt/vol final]), thus disputing the original description (1). These discrepancies suggest that a complete reassessment of the phenotypic markers used to discriminate *A. canalis* and *A. marinus* should be performed.

Data availability. The complete genome sequence of *A. canalis* strain LMG 29148^T has been deposited in GenBank under the accession number CP042812. All MiSeq and PacBio sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under accession number SRP217059.

ACKNOWLEDGMENT

This work was funded by the United States Department of Agriculture, Agricultural Research Service, under CRIS project 2030-42000-230-051.

REFERENCES

1. Pérez-Cataluña A, Salas-Massó N, Figueras MJ. 2018. *Arcobacter canalis* sp. nov., isolated from a water canal contaminated with urban sewage. Int J Syst Evol Microbiol 68:1258–1264. https://doi.org/10.1099/ijsem.0.002662.
2. Kim HM, Hwang CY, Cho BC. 2010. *Arcobacter marinus* sp. nov. Int J Syst Evol Microbiol 60:531–536. https://doi.org/10.1099/ijs.0.007740-0.
3. Besemer J, Borodovsky M. 2005. GeneMark: Web software for gene finding in prokaryotes, eukaryotes and viruses. Nucleic Acids Res 33: W451–W454. https://doi.org/10.1093/nar/gki487.
4. Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
5. Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res 32:11–16. https://doi.org/10.1093/nar/gkh152.
6. Miller WG, Yee E, Bono JL. 2018. Complete genome sequence of the *Arcobacter molluscorum* type strain LMG 25693. Microbiol Resour Announc 7:e01293-18. https://doi.org/10.1128/MRA.01293-18.
7. Cook GT. 1950. A plate test for nitrate reduction. J Clin Pathol 3:359–362. https://doi.org/10.1136/jcp.3.4.359.