Complete Genome Sequence of the *Arcobacter suis* Type Strain LMG 26152

William G. Miller,[a] Emma Yee,[a] James L. Bono[b]

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

[b]Meat Safety and Quality Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Clay Center, Nebraska, USA

ABSTRACT *Arcobacter* species are prevalent in pigs, and strains have been isolated from pig feces and pork meat; some *Arcobacter* strains may be porcine abortifaciens. *Arcobacter suis* was recovered from pork meat in Spain. This study describes the whole-genome sequence of the *A. suis* type strain LMG 26152 (=F41^T =CECT 7833^T).

*A*rcobacter species are often isolated from swine, pig manure, and pork meat (1–6), and some species or strains are possible porcine abortifaciens (7, 8). *Arcobacter suis*, represented by a single strain (the type strain), was originally isolated from retail market pork (9). Subsequent to the initial description, *A. suis* was also recovered from water buffalo milk (10), operational taxonomic units similar to *A. suis* were identified in samples from a spinach-processing plant (11), and *A. suis* was potentially identified in sewage (12). In this study, we report the first closed genome sequence of the *A. suis* type strain LMG 26152 (=F41^T =CECT 7833^T), isolated in 2008 from pork meat in Catalonia, Spain (9).

Arcobacter suis strain LMG 26152^T was grown at 30°C for 48 h aerobically on anaerobe basal agar (Oxoid) plus 5% horse blood. A loop of cells (~5 μl) was taken from the plate, and genomic DNA was isolated with the Wizard genomic DNA purification kit (Promega, Madison, WI). Shotgun and paired-end Roche 454 libraries were constructed following the manufacturer's protocols and with standard methods. PacBio SMRTbell libraries were prepared from 10 μg of genomic DNA with the standard 20-kb PacBio protocol (13). Shotgun and paired-end 454 libraries were sequenced on a GS-FLX+ instrument with Titanium chemistry and standard protocols. The resulting reads were assembled into 79 total contigs and a chromosomal scaffold of 33 contigs with Newbler v. 2.6 (Roche); Roche standard flowgram format (SFF) files were not processed before assembly, and 454 reads were quality controlled within the Newbler assembler. Low-quality contigs were deleted, and the remaining 25 contigs were positioned at one or more locations within the scaffold gaps with the Perl script contig_extender3 (14). PacBio sequencing was performed as described previously (14); the single chromosomal contig was assembled along with the 454 contigs with SeqMan v. 8.0 (DNASTAR, Madison, WI). Chromosomal assembly was also validated with an optical restriction map (restriction enzyme XbaI; OpGen, Gaithersburg, MD). Illumina HiSeq reads were obtained from SeqWright (Houston, TX) and assembled with Newbler v. 2.6. HiSeq reads were not processed before assembly, and any quality trimming of the reads was performed within Newbler. The HiSeq reads and contigs were used to verify and error correct the 454 and PacBio base calls, as described previously (15). The final coverage across the genome was 917×.

Sequencing metrics and genomic data for *A. suis* strain LMG 26152^T are presented in Table 1. *A. suis* strain LMG 26152^T has a circular genome of 2,639,269 bp, with an average G+C content of 27.4%. Putative coding sequences (CDSs), tRNA/transfer-messenger RNA

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

TABLE 1 Sequencing metrics and genomic data for *Arcobacter suis* strain LMG 26152[T]

Feature	Value(s)[a]
Sequencing metrics	
454 (shotgun) platform	
No. of reads	185,325
No. of bases	103,818,564
Average length (bases)	560.2
Coverage ($\times$)	39.3
454 (paired-end) platform	
No. of reads	76,546
No. of bases	24,718,237
Average length (bases)	322.9
Coverage ($\times$)	9.4
Illumina HiSeq 2000 platform	
No. of reads	17,658,830
No. of bases	1,765,883,000
Average length (bases)	100
Coverage ($\times$)	669.1
PacBio platform	
No. of reads	174,492
No. of bases	524,997,072
Average length (bases)	3,008.7[b]
Coverage ($\times$)	198.9
Newbler metrics[c]	
N50ContigSize (454) (bases)	142,251
Q40PlusBases (454) (%)	99.84
N50ContigSize (HiSeq pool 1) (bases)	142,383
Q40PlusBases (HiSeq pool 1) (%)	99.99
N50ContigSize (HiSeq pool 2) (bases)	142,376
Q40PlusBases (HiSeq pool 2) (%)	99.99
N50ContigSize (HiSeq pool 3) (bases)	142,376
Q40PlusBases (HiSeq pool 3) (%)	99.99
Genomic data	
Chromosome	
Size (bp)	2,639,269
G+C content (%)	27.39
No. of CDSs[d]	2,523
Assigned function (% CDSs)	976 (38.7)
General function annotation (% CDSs)	923 (36.6)
Domain/family annotation only (% CDSs)	173 (6.9)
Hypothetical (% CDSs)	451 (17.9)
Pseudogenes	34
Genomic islands/CRISPR	
No. of genetic islands	9
No. of CDSs in genetic islands	157, [7]
No. of CRISPR-Cas loci	0
Gene content/pathways	
IS elements, mobile elements, or tranposases	3, [2] (IS3)
Signal transduction	
Che proteins	*cheABCDRVW(Y)$_2$*
No. of methyl-accepting chemotaxis proteins	25
No. of response regulators	42, [1]
No. of histidine kinases	53, [1]
No. of response regulator/histidine kinase fusions	6
No. of diguanylate cyclases	17
No. of diguanylate phosphodiesterases (HD-GYP, EAL)	7, 22
No. of diguanylate cyclase/phosphodiesterases	13
No. of other	8, [1]
Motility	
Flagellin genes	*fla*
Restriction/modification	
No. of type I systems (*hsd*)	0
No. of type II systems	1
No. of type III systems	1

TABLE 1 (Continued)

Feature	Value(s)[a]
Transcription/translation	
No. of transcriptional regulatory proteins	54
Non-ECF[e] σ factors	σ^{54}, σ^{70}
No. of ECF σ factors	3
No. of tRNAs	59
No. of ribosomal loci	5
CO dehydrogenase (*coxSLF*)	Yes
Ethanolamine utilization (*eutBCH*)	Yes
Nitrogen fixation (*nif*)	Yes
Osmoprotection	*betA*
Pyruvate → acetyl-CoA[f]	
Pyruvate dehydrogenase (E1/E2/E3)	Yes
Pyruvate:ferredoxin oxidoreductase	*porABDG*
Urease	*ureAB*
Vitamin B_{12} biosynthesis	Yes

[a]Numbers in square brackets indicate pseudogenes or fragments.
[b]Maximum length, 24,428 bases.
[c]Features and values taken from largeContigMetrics within 454NewblerMetrics.txt for each assembly. Large contigs were defined as those ≥500 bases. Due to the large number of HiSeq reads, the total reads were split into 3 pools and assembled independently.
[d]Numbers do not include pseudogenes. CDSs, coding sequences.
[e]ECF, extracytoplasmic function.
[f]CoA, coenzyme A.

(tmRNA) genes, and rRNA loci were identified with GeneMark, ARAGORN, and RNAmmer, respectively (16–18). A preliminary GenBank-formatted file was created with the genome sequence and the GeneMark-derived CDS coordinates. Identification of putative pseudogenes and genes missed in the original GeneMark analysis and manual curation of each putative CDS were performed with the GenBank-formatted file and Artemis v. 16 (19). Annotation was accomplished with BLASTP to compare the proteome of strain LMG 26152[T] to proteins in both the NCBI nonredundant (nr) database and a custom protein database constructed from the proteomes of all current completed *Arcobacter* and *Campylobacter* genomes. Annotation was further refined with an analysis of Pfam motifs (20).

The LMG 26152[T] genome is predicted to encode 2,523 putative protein-coding genes, 34 pseudogenes, 5 rRNA operons, and 59 tRNA-encoding genes, and it contains 9 genomic islands ranging from 5.5 to 34.3 kb in size. The LMG 26152[T] genome contains a *nif/rpoN* nitrogen fixation gene cluster (21; GenBank accession number CP001999) and the same set of adenosylcobalamin biosynthesis genes identified previously in *A. bivalviorum* (22). The *A. suis* genome also encodes the B_{12}-dependent EutBC ethanolamine ammonia-lyase and the EutH ethanolamine permease. The acetaldehyde produced by EutBC would presumably be converted to ethanol and acetyl-coenzyme A by a putative AdhE dehydrogenase (Asuis2568). Two large genes encoding T1SS repeat domain-containing proteins were identified in the *A. suis* genome: *asuis0242* (9,252 bp) and *asuis0243* (16,326 bp). Similarly to *A. mytili* (23), these genes contain tandemly repeated internal motifs (5 × 639 bp, *asuis0242*; 39 × 647 bp, *asuis0243*). A 34,282-bp gene with no internal repeats encoding a putative repeats-in-toxin (RTX) family calcium-binding protein was also identified.

Data availability. The complete genome sequence of *A. suis* strain LMG 26152[T] has been deposited in GenBank under the accession number CP032100. 454, HiSeq, and PacBio sequencing reads have been deposited in the NCBI Sequence Read Archive (SRA) under the accession number SRP155204.

REFERENCES

1. De Smet S, Vandamme P, De Zutter L, On SL, Douidah L, Houf K. 2011. *Arcobacter trophiarum* sp. nov., isolated from fattening pigs. Int J Syst Evol Microbiol 61:356–361. https://doi.org/10.1099/ijs.0.022665-0.

2. Forsythe SJ. 2006. *Arcobacter*, p 181–221. *In* Motarjemi Y, Adams M (ed), Emerging foodborne pathogens. Woodhead Publishing Ltd, Cambridge, England.

3. Miller WG, Wesley IV, On SL, Houf K, Megraud F, Wang G, Yee E, Srijan A, Mason CJ. 2009. First multi-locus sequence typing scheme for *Arcobacter* spp. BMC Microbiol 9:196. https://doi.org/10.1186/1471-2180-9-196.

4. Wesley IV, Miller WG. 2010. *Arcobacter*: an opportunistic human foodborne pathogen? p 185–212. *In* Scheld WM, Grayson ML, Hughes JM (ed), Emerging infections 9. ASM Press, Washington, DC.

5. Whiteduck-Léveillée K, Whiteduck-Léveillée J, Cloutier M, Tambong JT, Xu R, Topp E, Arts MT, Chao J, Adam Z, André Lévesque C, Lapen DR, Villemur R, Talbot G, Khan IUH. 2015. *Arcobacter lanthieri* sp. nov., isolated from pig and dairy cattle manure. Int J Syst Evol Microbiol 65:2709–2716. https://doi.org/10.1099/ijs.0.000318.

6. Ramees TP, Dhama K, Karthik K, Rathore RS, Kumar A, Saminathan M, Tiwari R, Malik YS, Singh RK. 2017. *Arcobacter*: an emerging food-borne zoonotic pathogen, its public health concerns and advances in diagnosis and control—a comprehensive review. Vet Q 37:136–161. https://doi.org/10.1080/01652176.2017.1323355.

7. Houf K, On SL, Coenye T, Debruyne L, De Smet S, Vandamme P. 2009. *Arcobacter thereius* sp. nov., isolated from pigs and ducks. Int J Syst Evol Microbiol 59:2599–2604. https://doi.org/10.1099/ijs.0.006650-0.

8. On SL, Jensen TK, Bille-Hansen V, Jorsal SE, Vandamme P. 2002. Prevalence and diversity of *Arcobacter* spp. isolated from the internal organs of spontaneous porcine abortions in Denmark. Vet Microbiol 85:159–167. https://doi.org/10.1016/S0378-1135(01)00503-X.

9. Levican A, Collado L, Figueras MJ. 2013. *Arcobacter cloacae* sp. nov. and *Arcobacter suis* sp. nov., two new species isolated from food and sewage. Syst Appl Microbiol 36:22–27. https://doi.org/10.1016/j.syapm.2012.11.003.

10. Giacometti F, Salas-Massó N, Serraino A, Figueras MJ. 2015. Characterization of *Arcobacter suis* isolated from water buffalo (*Bubalus bubalis*) milk. Food Microbiol 51:186–191. https://doi.org/10.1016/j.fm.2015.06.004.

11. Hausdorf L, Neumann M, Bergmann I, Sobiella K, Mundt K, Frohling A, Schlüter O, Klocke M. 2013. Occurrence and genetic diversity of *Arcobacter* spp. in a spinach-processing plant and evaluation of two *Arcobacter*-specific quantitative PCR assays. Syst Appl Microbiol 36:235–243. https://doi.org/10.1016/j.syapm.2013.02.003.

12. Fisher JC, Levican A, Figueras MJ, McLellan SL. 2014. Population dynamics and ecology of *Arcobacter* in sewage. Front Microbiol 5:525. https://doi.org/10.3389/fmicb.2014.00525.

13. PacBio. 2015. Procedure and checklist: 20 kb template preparation using BluePippin size-selection system. https://www.pacb.com/wp-content/uploads/2015/09/Procedure-Checklist-20-kb-Template-Preparation-Using-BluePippin-Size-Selection.pdf.

14. Miller WG, Yee E, Lopes BS, Chapman MH, Huynh S, Bono JL, Parker CT, Strachan NJC, Forbes KJ. 2017. Comparative genomic analysis identifies a *Campylobacter* clade deficient in selenium metabolism. Genome Biol Evol 9:1843–1858. https://doi.org/10.1093/gbe/evx093.

15. Miller WG, Yee E. 2018. Complete genome sequence of the *Arcobacter trophiarum* type strain LMG 25534. Microbiol Resour Announc 7:e01110–e01118. https://doi.org/10.1128/MRA.01110-18.

16. 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.

17. Lagesen K, Hallin P, Rødland EA, Stærfeldt 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.

18. 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.

19. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944–945. https://doi.org/10.1093/bioinformatics/16.10.944.

20. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer EL, Eddy SR, Bateman A, Finn RD. 2012. The Pfam protein families database. Nucleic Acids Res 40:D290–D301. https://doi.org/10.1093/nar/gkr1065.

21. Mizutani Y, Tanaka R. 2017. Genome sequence of *Arcobacter* sp. strain LA11, isolated from the abalone *Haliotis discus*. Genome Announc 5:e00032-17. https://doi.org/10.1128/genomeA.00032-17.

22. Miller WG, Yee E, Bono JL. 2018. Complete genome sequence of the *Arcobacter bivalviorum* type strain LMG 26154. Microbiol Resour Announc 7:e01076. https://doi.org/10.1128/MRA.01076-18.

23. Miller WG, Yee E, Bono JL. 2018. Complete genome sequence of the *Arcobacter mytili* type strain LMG 24559. Microbiol Resour Announc 7:e01078. https://doi.org/10.1128/MRA.01078-18.