



Complete Genome Sequence of *Campylobacter armoricus* CA639, Which Carries Two Plasmids, Compiled Using Oxford Nanopore and Illumina Sequencing Technologies

Amine M. Boukerb,^{a*} Julien Schaeffer,^a Joëlle Serghine,^a Gregory Carrier,^b Françoise S. Le Guyader,^a Michèle Gourmelon^a

^aIFREMER RBE-SGMM-LSEM, Laboratoire de Santé Environnement et Microbiologie, Plouzané, Institut Français de Recherche pour l'Exploitation de la Mer, France

^bIFREMER RBE-BRM-LPBA, Laboratoire de Physiologie et Biotechnologie des Algues, Institut Français de Recherche pour l'Exploitation de la Mer, Nantes, France

ABSTRACT As determined by a hybrid approach combining Oxford Nanopore MinION and Illumina MiniSeq sequence data, *Campylobacter armoricus* strain CA639 harbored a circular chromosome of 1,688,169 bp with a G+C content of 28.47% and two plasmids named pCA639-1 and pCA639-2, with lengths of 51,123 and 28,139 bp, and G+C contents of 26.5% and 28.45%, respectively.

Campylobacter armoricus is a novel urease-positive bacterial species phylogenetically classified within the *Campylobacter lari* group (1, 2). This group forms a distinct clade within the epsilon subdivision of the *Proteobacteria* and its members are among the thermotolerant *Campylobacter* spp. (3). We report here the complete sequence of the river water isolate *Campylobacter armoricus* CA639 and its native plasmids pCA639-1 and pCA639-2. This strain was isolated from the river Le Rat (La Fresnaye catchment, Brittany, France) on 4 March 2014 using the ISO-10272:2016 method (1, 2). Bacterial DNA was extracted from an overnight culture in trypto-casein-soy agar (bioMérieux, Marcy-l'Étoile, France) supplemented with 5% (vol/vol) sheep blood (Oxoid, Thermo Scientific, Inc.) at 42°C in a microaerobic atmosphere, using the DNA QIAamp minikit 250 (Qiagen, Venlo, The Netherlands) and used for Illumina and Nanopore sequencing. Genomic libraries were prepared using the Nextera DNA Flex library prep kit (Illumina, San Diego, CA, USA), and sequencing was performed on an Illumina MiniSeq platform with a 2 × 150 paired-end protocol (1). Default parameters were used for all software except where otherwise noted. Raw reads were quality filtered and adapter trimmed with Trimmomatic v.0.36 (4). An Oxford Nanopore Technologies (ONT) sequencing library was prepared using the manufacturer's 1D genomic DNA by ligation kit (SQK-LSK 108), and sequencing was carried out on a MinION device using flow cell type R9.4.1 (FLO-MIN106D). Porechop v.0.2.1 (5) was used for adaptor trimming, and NanoFilt v.2.2.0 (6) was used to remove reads of <500 bp or with average quality scores of <10. Thus, we used a robust pipeline relying on a combination of Oxford Nanopore long-read (681,890; N_{50} value, 15,543 bp; 9.7 Gb of data) and Illumina short-read (1,309,028; 2 × 150-bp reads) technologies to scaffold and polish sequencing data.

Several approaches were used to construct *de novo* assemblies using default parameters (Fig. 1). Based on the obtained statistics (Fig. 1A), the Unicycler hybrid assembly was selected for downstream analyses. This reported one circular chromosome of 1,688,169 bp (28.47% G+C content) and two plasmids named pCA639-1 and pCA639-2 with lengths of 51,123 and 28,139 bp and G+C contents of 26.5 and 28.45%, respectively (Fig. 1B). BMap v.38.71 (<https://sourceforge.net/projects/bbmap/>) was used to calculate the average coverages for the chromosome (101.7× for short reads and 1,301.2× for long reads), pCA639-1 (87.9× and 328.9×, respectively), and pCA639-2 (178.8× and 280.4×, respectively).

Citation Boukerb AM, Schaeffer J, Serghine J, Carrier G, Le Guyader FS, Gourmelon M. 2020. Complete genome sequence of *Campylobacter armoricus* CA639, which carries two plasmids, compiled using Oxford Nanopore and Illumina sequencing technologies. Microbiol Resour Announc 9:e01309-19. <https://doi.org/10.1128/MRA.01309-19>.

Editor Simon Roux, DOE Joint Genome Institute

Copyright © 2020 Boukerb et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Amine M. Boukerb, amine.boukerb@univ-rouen.fr, or Michèle Gourmelon, michele.gourmelon@ifremer.fr.

* Present address: Amine M. Boukerb, Laboratory of Microbiology Signals and Microenvironment (LMSM EA4312), University of Rouen Normandy, Normandy University, Evreux, France.

Received 17 October 2019

Accepted 14 November 2019

Published 2 January 2020

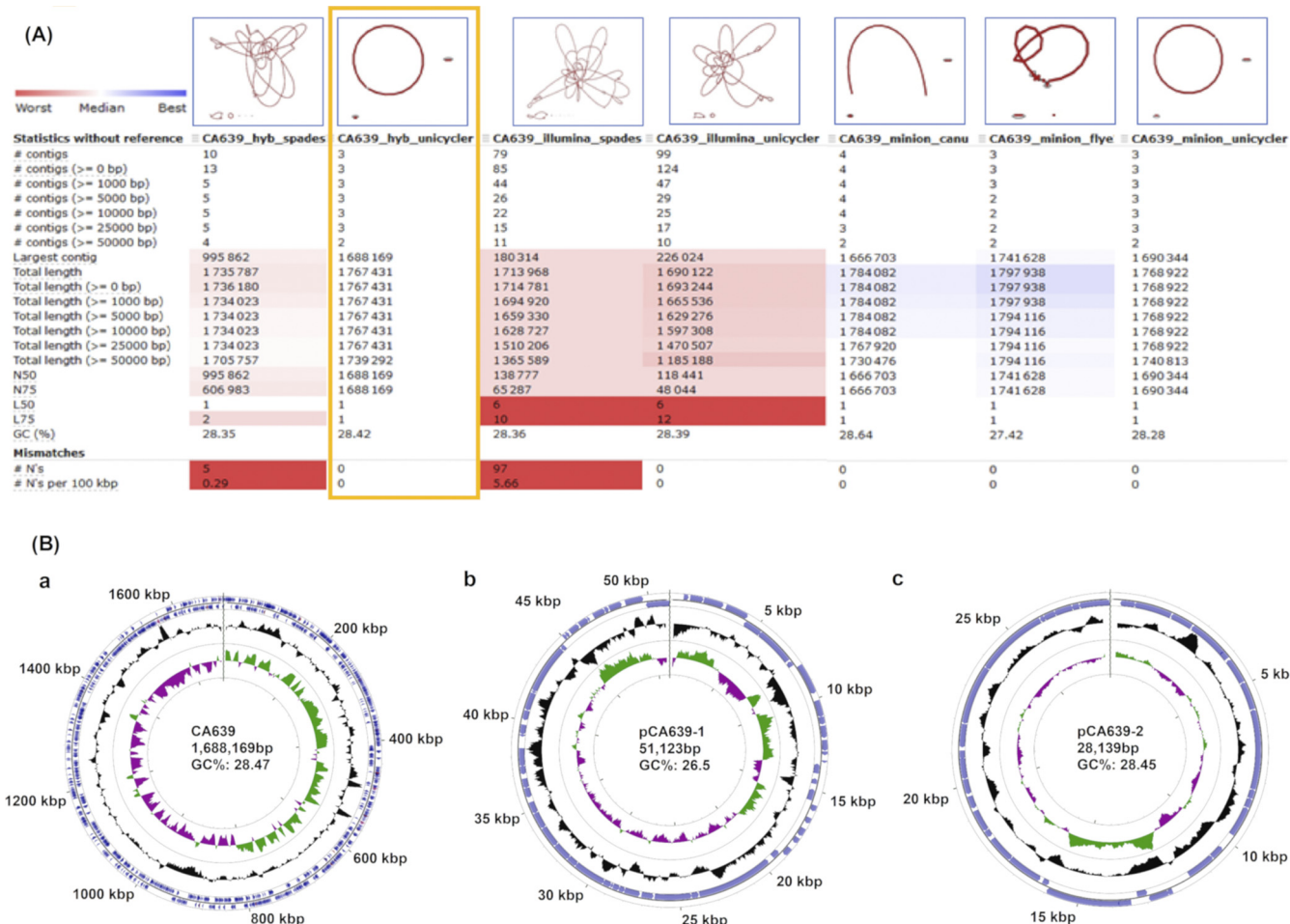


FIG 1 (A) Visualization of assembly graphs and statistics for each strategy was produced with Bandage v.0.8.1 (9) and QUAST v.5.0.0 (10), respectively. First, we constructed MiniSeq assemblies (illumina) using SPAdes v.3.12.0 (11) or Unicycler v.0.4.7 (12). Second, MinION assemblies (minion) were achieved using Canu v.1.5 (13), Flye v.2.4 (14), or Unicycler. These three assemblies were aligned to MinION reads using Minimap2 v.2.17 (15) and SAMtools v.1.9 (16) and then polished using Nanopolish v.0.11.0 (17). An additional round of Nanopolish did not improve their accuracy. Moreover, the Canu assembly was polished using Pilon v.1.23 (18) with the flags “-fix bases” and then “-fix all” by aligning MiniSeq reads using Bowtie 2 v.2.3.4.3 (19) and SAMtools. Third, we added MinION reads to the obtained MiniSeq-based assemblies to resolve ambiguous regions in the sequencing graph, creating SPAdes hybrid and Unicycler hybrid assemblies (Hyb). (B) Circular maps of the *C. armoricus* CA639 replicons (a, chromosome; b and c, plasmids) from the hybrid assembly using Unicycler were drawn using the online CGView server (http://stothard.afns.ualberta.ca/cgview_server/). Counting from the outside toward the center, circle 1 (outermost circle) shows distances from the putative origin of replication in kilobase pairs. Circle 2 shows annotated CDS (blue) encoded on the forward and reverse strands. The *rrs* operons and tRNA genes in the chromosome are indicated in pink and gray, respectively. Circle 3 shows G+C contents higher and lower than the average G+C content (black). Circle 4 shows G+C skew, with positive values in green and negative values in purple.

Prokka v.1.14 (7) predicted 1,640 putative coding sequences (CDS), with 862 (52.6%) having assigned functions, including 3 rRNA operons and 43 tRNAs for the chromosome and 59 and 35 CDS for the two respective plasmids. The chromosome contains one prophage integrase and an ISHp1 transposase (IS1595 family). In addition to the results that were obtained for virulence (i.e., the *cdtABC* operon, *ciaB*, *flaC*, *porA*, and *cadF*) and antibiotic resistance (i.e., *cmeABC*, *cmeR*, *cosR*, *macAB*, *oxa-184*, and *oxa-493*) coding gene screening (1) using ABRicate v.0.8.7 (8), we detected a chloramphenicol acetyltransferase type III gene (*cat3*), a bicyclomycin resistance gene (*bcr*), and other multidrug efflux pump-coding genes that may be involved in antibiotic resistance. pCA639-1 harbored genes coding for the Tra/Vir type IV secretion system (T4SS) and a Cag pathogenicity island protein. A blastn search of the sequence of this plasmid against the NCBI database showed 79% query coverage and 94.56% identity with that of *Campylobacter lari* pCL2100 (GenBank accession number CP000933). pCA639-2 carried several conjugative transfer genes and shared 82% query coverage and 95.06%

identity with pGMI16-001 (GenBank accession number [CP028188](#)) carried by *Campylobacter coli* strain CFSAN054106, suggesting an intraspecies dissemination.

This study highlights the value of combining short- and long-read sequencing data for high-quality genome assemblies and annotation of repetitive genomic regions. The complete genome sequence of *C. armoricus* CA639 comprises essential data for taxonomic and comparative genomic studies within a One Health approach, a concept which recognizes that the health of people is connected to the health of animals and the environment.

Data availability. The sequencing data have been deposited in the DDBJ/EMBL/GenBank databases under accession numbers [CP044262](#) for the chromosome and [CP044261](#) and [CP044263](#) for plasmids pCA639-1 and pCA639-2, respectively. The Illumina paired-end fastq and ONT base-called fastq files are available in the Sequence Read Archive under accession numbers [SRR10390899](#) and [SRR10162491](#), respectively.

ACKNOWLEDGMENTS

We acknowledge the Pôle de Calcul et de Données Marines (PCDM; <http://www.ifremer.fr/pcdm>) for providing DATARMOR (storage, data access, computational resources, visualization, Web services, consultation, and support services).

This work was supported by funding from the European Commission's Horizon 2020 Research and Innovation program under the COMPARE project (grant agreement 643476).

REFERENCES

- Boukerb AM, Penny C, Serghine J, Walczak C, Cauchie HM, Miller WG, Losch S, Ragimbeau C, Mossong J, Mégraud F, Lehours P, Bénédjat L, Gourmelon M. 2019. *Campylobacter armoricus* sp. nov., a novel member of the *Campylobacter lari* group isolated from surface water and stools from humans with enteric infection. *Int J Syst Evol Microbiol*, in press. <https://doi.org/10.1099/ijsem.0.003836>.
- Rincé A, Baliere C, Hervio-Heath D, Cozien J, Lozach S, Parnaudeau S, Le Guyader FS, Le Hello S, Giard JC, Sauvageot N, Benachour A, Strubbia S, Gourmelon M. 2018. Occurrence of bacterial pathogens and human noroviruses in shellfish-harvesting areas and their catchments in France. *Front Microbiol* 9:2443. <https://doi.org/10.3389/fmicb.2018.02443>.
- Miller WG, Yee E, Chapman MH, Smith TP, Bono JL, Huynh S, Parker CT, Vandamme P, Luong K, Korch J. 2014. Comparative genomics of the *Campylobacter lari* group. *Genome Biol Evol* 6:3252–3266. <https://doi.org/10.1093/gbe/evu249>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Micromol Lett* 3:e000132. <https://doi.org/10.1099/mgen.0.000132>.
- De Coster W, D'Hert S, Schultz DT, Cruets M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- Seemann T. 2018. ABRicate: mass screening of contigs for antimicrobial and virulence genes. <https://github.com/tseemann/abricate>.
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* 31:3350–3352. <https://doi.org/10.1093/bioinformatics/btv383>.
- Mikheenko A, Prijibelski A, Saveliev V, Antipov D, Gurevich A. 2018. Versatile genome assembly evaluation with QUAST-LG. *Bioinformatics* 34:i142–i150. <https://doi.org/10.1093/bioinformatics/bty266>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prijibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>.
- Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. *Nat Methods* 12:733–735. <https://doi.org/10.1038/nmeth.3444>.
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
- Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.