Draft Genome Sequence of *Helicobacter* sp. Strain CaF467b, Isolated from a Pig Manure Storage Tank

Galen Guo,ª Wen Chen,ª Michel Cloutier,ª John Chmara,ª Suzanne Gerdis,ª Julie Chapados,ª Jeremy Dettman,ª Izhar U. H. Khanª

ªOttawa Research & Development Centre, Science & Technology Branch, Agriculture and Agri-Food Canada, Ottawa, Ontario, Canada

ABSTRACT In this report, we present the draft genome sequence of an unclassified *Helicobacter* strain, CaF467b. This bacterial isolate was recovered from a pig manure storage tank. The draft genome sequence is 1,655,514 bp in length with 1,709 predicted genes and a G+C content of 34.07%.

*H*elicobacter species are Gram-negative, curved bacilli commonly found in the gastrointestinal environment of humans (1, 2) and animals (3, 4). Recently, non-*Helicobacter pylori* species have been considered emerging human pathogens since they have been reported in clinical cases (5–7) that are often acquired through direct contact with infected animals (8). Therefore, there is an urgent need to understand the full nature of their pathogenesis.

Strain CaF467b was recovered from a pig manure storage tank located in Ayer's Cliff, Quebec, Canada. The sample was aseptically collected in a sterile bottle containing was processed $1\times$ phosphate-buffered saline (PBS). The suspension was 10-fold serially diluted and isolated on modified Karmali agar (Thermo Fisher Scientific, KS, USA) containing antimicrobial supplements under microaerophilic (5% O_2, 85% N_2, 10% CO_2) conditions for 48 h. DNA was extracted from a purified colony using a DNeasy UltraClean microbial kit (Qiagen, MD, USA) as per the manufacturer's instructions. Initial 16S rRNA amplicon sequencing, using the primer pair pA-F (5′-AGAGTTTGATCCTGGCTCAG-3′) and pH-R (5′-AAGGAGGTG ATCCAGCCGC-3′), showed that strain CaF467b had the highest sequence identity to *Helicobacter canadensis* strain MIT 98-5491 (Accession number GCF_000162575.1).

Libraries were prepared and pooled using the DNA prep and NextSeq 500 mid output reagent kits (Illumina, CA, USA). Whole-genome sequencing was performed at the AAFC Ottawa Research and Development Centre on the Illumina NextSeq 500 platform (2×150 bp). We obtained a total of 1,826,288 paired-end raw reads; adapters and low-quality bases were trimmed using fastp v.0.20.1 (9), and the cleaned paired-end reads were assembled using Megahit v.1.2.9 (10), where sequences of $>5,000$ bp were excluded. The taxonomy was estimated using the *anvi-run-scg-taxonomy* function from Anvi'o v.7.1 (11) against the GTDB release 202 (12). Assemblies were refined such that 20 contigs were identified as *Helicobacter* using *anvi-refine* from Anvi'o (11). CheckM v.1.1.3 (13) was used to assess the quality of the final draft genome of strain CaF467b, which showed a completeness of 99.71% with 0.35% contamination. The whole-genome alignment and assembly were improved using RagTag (14, 15), with *H. canadensis* (GCF_000162575.1) as the reference genome as determined by average nucleotide identity (ANI) analysis (1).

The draft genome is 1,655,514 bp long, with an average coverage of $119.98\times$, an N_{50} value of 116,880 bp, and a G+C content of 34.07%. The ANI was calculated using MinHash (16) through GTDB-tk (17). ANI analysis showed that this draft genome has a 90.3% similarity to the reference genome of *H. canadensis* (GCF_000162575.1) and 98.8% similarity to the draft genome of *Helicobacter* sp. strain 11-8110 (GCF_003288905.1). Annotation using the Prokaryotic Genome Annotation Pipeline (PGAP) yielded a total of 1,653 coding genes, 18 pseudogenes, and 35 tRNA genes (18). Using ABRicate v.1.0.1 (19), we screened for potential antibiotic resistance genes and detected *tetW*, with a coverage of 85.5% and

identity of 99.82% against the NCBI AMRFinderPlus database (20) and a coverage of 85.52% and identity of 98.72% against the CARD database (21).

The draft genome sequence reported here could serve as a reference genome for comparative genomics, expand the taxonomic branch of other non-*H. pylori Helicobacter* spp., and be mined for zoonotic pathogenic potential.

Data availability. The whole-genome shotgun project for strain CaF467b has been deposited at DDBJ/ENA/GenBank under the accession number JALAIY000000000.1 and under the BioProject accession number PRJNA809382. The version described in this paper is version JALAIY010000000. The raw sequencing files can be found under the SRA accession number SRR18128650.

REFERENCES

1. Fox JG, Chien CC, Dewhirst FE, Paster BJ, Shen Z, Melito PL, Woodward DL, Rodgers FG. 2000. *Helicobacter canadensis* sp. nov. isolated from humans with diarrhea as an example of an emerging pathogen. J Clin Microbiol 38: 2546–2549. https://doi.org/10.1128/JCM.38.7.2546-2549.2000.
2. Marshall BJ, Warren JR. 1984. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. Lancet 323:1311–1315. https://doi.org/10.1016/S0140-6736(84)91816-6.
3. Van den Bulck K, Decostere A, Baele M, Driessen A, Debongnie J-C, Burette A, Stolte M, Ducatelle R, Haesebrouck F. 2005. Identification of non-*Helicobacter pylori* spiral organisms in gastric samples from humans, dogs, and cats. J Clin Microbiol 43:2256–2260. https://doi.org/10.1128/JCM.43.5.2256-2260.2005.
4. Péré-Védrenne C, Flahou B, Loke MF, Ménard A, Vadivelu J. 2017. Other Helicobacters, gastric and gut microbiota. Helicobacter 22:e12407. https://doi.org/10.1111/hel.12407.
5. Da Cunha T, Dharan M. 2021. Acute peptic ulcer disease caused by non-*Helicobacter pylori*-helicobacter. Cureus 13:e14993. https://doi.org/10.7759/cureus.14993.
6. Hayashi T, Tomida J, Kawamura Y, Yoshida M, Yokozawa I, Kaneko S. 2017. Unusual manifestation of *Helicobacter cinaedi* infection: a case report of intracranial subdural empyema and bacteremia. BMC Infect Dis 17:40. https://doi.org/10.1186/s12879-016-2129-3.
7. Katsuma A, Yamamoto I, Tsuchiya Y, Kawabe M, Yamakawa T, Katsumata H, Mafune A, Nakada Y, Kobayashi A, Koike K, Shimizu A, Tanno Y, Ohkido I, Tsuboi N, Hori S, Yamamoto H, Yokoo T. 2017. Helicobacter cinaedi bacteremia with cellulitis in a living-donor kidney transplant recipient identified by matrix-assisted laser desorption ionization time-of-flight mass spectrometry: a case report. BMC Res Notes 10:87. https://doi.org/10.1186/s13104-017-2403-5.
8. Haesebrouck F, Pasmans F, Flahou B, Chiers K, Baele M, Meyns T, Decostere A, Ducatelle R. 2009. Gastric helicobacters in domestic animals and nonhuman primates and their significance for human health. Clin Microbiol Rev 22:202–223. https://doi.org/10.1128/CMR.00041-08.
9. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
10. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics 31:1674–1676. https://doi.org/10.1093/bioinformatics/btv033.
11. Eren AM, Kiefl E, Shaiber A, Veseli I, Miller SE, Schechter MS, Fink I, Pan JN, Yousef M, Fogarty EC, Trigodet F, Watson AR, Esen ÖC, Moore RM, Clayssen Q, Lee MD, Kivenson V, Graham ED, Merrill BD, Karkman A, Blankenberg D, Eppley JM, Sjödin A, Scott JJ, Vázquez-Campos X, McKay LJ, McDaniel EA, Stevens SLR, Anderson RE, Fuessel J, Fernandez-Guerra A, Maignien L, Delmont TO, Willis AD. 2021. Community-led, integrated, reproducible multi-omics with Anvi'o. Nat Microbiol 6:3–6. https://doi.org/10.1038/s41564-020-00834-3.
12. Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil P-A, Hugenholtz P. 2021. GTDB: an ongoing census of bacterial and archaeal diversity through a phylogenetically consistent, rank normalized and complete genome-based taxonomy. Nucleic Acids Res 50:D785–D794. https://doi.org/10.1093/nar/gkab776.
13. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res 25:1043–1055. https://doi.org/10.1101/gr.186072.114.
14. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20:224. https://doi.org/10.1186/s13059-019-1829-6.
15. Alonge M, Lebeigle L, Kirsche M, Aganezov S, Wang X, Lippman ZB, Schatz MC, Soyk S. 2021. Automated assembly scaffolding elevates a new tomato system for high-throughput genome editing. bioRxiv. https://doi.org/10.1101/2021.11.18.469135.
16. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 17:132. https://doi.org/10.1186/s13059-016-0997-x.
17. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics 36:1925–1927. https://doi.org/10.1093/bioinformatics/btz848.
18. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, Coulouris G, Chitsaz F, Derbyshire MK, Durkin AS, Gonzales NR, Gwadz M, Lanczycki CJ, Song JS, Thanki N, Wang J, Yamashita RA, Yang M, Zheng C, Marchler-Bauer A, Thibaud-Nissen F. 2021. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. Nucleic Acids Res 49: D1020–D1028. https://doi.org/10.1093/nar/gkaa1105.
19. Seemann T. 2020. ABRicate (1.1.0). Github. https://github.com/tseemann/abricate.
20. Feldgarden M, Brover V, Gonzalez-Escalona N, Frye JG, Haendiges J, Haft DH, Hoffmann M, Pettengill JB, Prasad AB, Tillman GE, Tyson GH, Klimke W. 2021. AMRFinderPlus and the Reference Gene Catalog facilitate examination of the genomic links among antimicrobial resistance, stress response, and virulence. Sci Rep 11:12728. https://doi.org/10.1038/s41598-021-91456-0.
21. Alcock BP, Raphenya AR, Lau TTY, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen A-LV, Cheng AA, Liu S, Min SY, Miroshnichenko A, Tran H-K, Werfalli RE, Nasir JA, Oloni M, Speicher DJ, Florescu A, Singh B, Faltyn M, Hernandez-Koutoucheva A, Sharma AN, Bordeleau E, Pawlowski AC, Zubyk HL, Dooley D, Griffiths E, Maguire F, Winsor GL, Beiko RG, Brinkman FSL, Hsiao WWL, Domselaar GV, McArthur AG. 2020. CARD 2020: antibiotic resistome surveillance with the Comprehensive Antibiotic Resistance Database. Nucleic Acids Res 48:D517–D525. https://doi.org/10.1093/nar/gkz935.