Genome Sequence of *Zymomonas mobilis* subsp. *mobilis* NRRL B-1960

Katherine Chacon-Vargas, Alexandra A. Chirino, Meghan M. Davis,
Sophia A. Debler, Wynn Raphael Haimer, Justin J. Wilbur, Xiaoli Mo,
Baxter W. Worthing, Ethan G. Wainblat, Shu Zhao, John G. Gibbons

Biology Department, Clark University, Worcester, Massachusetts, USA

ABSTRACT *Zymomonas mobilis* subsp. *mobilis* is an efficient ethanol producer with application for industrial production of biofuel. To supplement existing *Z. mobilis* genomic resources and to facilitate genomic research, we used Oxford Nanopore and Illumina sequencing to assemble the complete genome of the beer spoilage isolate *Z. mobilis* subsp. *mobilis* strain NRRL B-1960.

Zymomonas mobilis subsp. *mobilis* is a Gram-negative alphaproteobacterium capable of highly efficient ethanol production through the Entner-Doudroff pathway (1). *Z. mobilis* has garnered interest for its prospective biofuel application, and it is utilized in industrial-scale fermentations to produce bionic acid, levan, and sorbitol (2, 3). The genomes of several natural isolates of *Z. mobilis* from North America, South America, Europe, and Africa have been previously sequenced (4–8). Here, we sequenced the genome of *Zymomonas mobilis* subsp. *mobilis* NRRL B-1960 (NCIMB 8227), originally isolated from "bad beer" in the United Kingdom (1), to facilitate population and comparative genomics research of this industrially important species.

Z. mobilis NRRL B-1960 was grown in 0.5% yeast extract and 2% glucose liquid medium at 30°C for 24 h. DNA was extracted using the PureLink microbiome DNA purification kit (Thermo Fisher Scientific), and DNA integrity was evaluated via gel electrophoresis. We constructed an Oxford Nanopore (ONT) rapid 1D library from 500 ng of DNA, according to the manufacturer's instructions (kit version RAD002) and performed sequencing on two flow cells (R9 version) of the MinION device. Poretools version 0.6.0 (9) was used to convert ONT data from fast5 to fastq format. We performed error correction of ONT reads using Canu version 1.4 (10) and generated a total of 12,302 corrected reads, with an average read length of 4,152 bp, representing ~25× coverage. A paired-end 101-bp Illumina library was constructed and sequenced at Macrogen (Rockville, MD) from 1 μg of DNA. Illumina sequence data were deduplicated using Tally (11), adapter and quality trimmed using Trim Galore version 0.4.3 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/), and errors corrected using SPAdes version 3.10.0 (12), resulting in 14,838,803 paired-end reads, representing ~1,400× coverage. We incorporated both sequence data types in SPAdes to produce an assembly using k-mer sizes of 27, 37, 47, 57, 67, 77, 87, and 97 (12). Last, preassembly-improved Illumina data was used for genome polishing via Pilon version 1.21 (13).

We assembled the genome into a single scaffold consisting of 2,045,798 bp, with a GC content of 46.09%. Two additional scaffolds were assembled which correspond to plasmid sequences. The larger plasmid sequence, pZMO1960_1, is 34,463 bp, with a GC content of 45.77%, and has an ~5.5-kb region nearly identical to a *Z. mobilis* CP4 plasmid (GenBank accession no. CP006894) (6). The other plasmid sequence, pZMO1960_1A, is 1,743 bp, with a GC content of 38.15%, and it is identical to the *Z. mobilis* NCIMB 11163 plasmid pZMO1A (GenBank accession no. GQ293074) (5). The RAST server (14, 15) was used for genome annotation, and clustered regularly inter-

spaced short palindromic repeat (CRISPR) arrays were predicted using CRISPRFinder (16). The genome contains 1,971 protein-coding genes, 3 rRNA gene loci, 51 tRNA-encoding genes, and 3 CRISPR arrays.

Phylogenetic analysis of 74,852 polymorphic sites across the sequenced natural isolates of *Z. mobilis* indicates that the NRRL B-1960 genome is most closely related to the NCIMB 11163 strain, which was isolated from spoiled beer in England (5). We conservatively identified 11 genes unique to the NRRL B-1960 genome, including 2 TonB-dependent receptor-encoding genes, via BLAST (17) searches of the NRRL B-1960 genes against the 5 sequenced *Z. mobilis* genomes (4–8).

Accession number(s). The *Zymomonas mobilis* subsp. *mobilis* NRRL B-1960 genome was assigned GenBank accession number CP021053 for the chromosome and accession numbers CP021791 and CP021792 for the plasmids.

REFERENCES

1. Swings J, De Ley J. 1977. The biology of *Zymomonas*. Bacteriol Rev 41:1–46.
2. He MX, Wu B, Qin H, Ruan ZY, Tan FR, Wang JL, Shui ZX, Dai LC, Zhu QL, Pan K, Tang XY, Wang WG, Hu QC. 2014. *Zymomonas mobilis*: a novel platform for future biorefineries. Biotechnol Biofuels 7:101. https://doi.org/10.1186/1754-6834-7-101.
3. Yang S, Fei Q, Zhang Y, Contreras LM, Utturkar SM, Brown SD, Himmel ME, Zhang M. 2016. *Zymomonas mobilis* as a model system for production of biofuels and biochemicals. Microb Biotechnol 9:699–717. https://doi.org/10.1111/1751-7915.12408.
4. Desiniotis A, Kouvelis VN, Davenport K, Bruce D, Detter C, Tapia R, Han C, Goodwin LA, Woyke T, Kyrpides NC, Typas MA, Pappas KM. 2012. Complete genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* centrotype ATCC 29191. J Bacteriol 194:5966–5967. https://doi.org/10.1128/JB.01398-12.
5. Kouvelis VN, Saunders E, Brettin TS, Bruce D, Detter C, Han C, Typas MA, Pappas KM. 2009. Complete genome sequence of the ethanol producer *Zymomonas mobilis* NCIMB 11163. J Bacteriol 191:7140–7141. https://doi.org/10.1128/JB.01084-09.
6. Kouvelis VN, Teshima H, Bruce D, Detter C, Tapia R, Han C, Tampako-poulou VO, Goodwin L, Woyke T, Kyrpides NC, Typas MA, Pappas KM. 2014. Finished genome of *Zymomonas mobilis* subsp. *mobilis* strain CP4, an applied ethanol producer. Genome Announc 2(1):e00845-13. https://doi.org/10.1128/genomeA.00845-13.
7. Pappas KM, Kouvelis VN, Saunders E, Brettin TS, Bruce D, Detter C, Balakireva M, Han CS, Savvakis G, Kyrpides NC, Typas MA. 2011. Genome sequence of the ethanol-producing *Zymomonas mobilis* subsp. *mobilis* lectotype strain ATCC 10988. J Bacteriol 193:5051–5052. https://doi.org/10.1128/JB.05395-11.
8. Seo JS, Chong H, Park HS, Yoon KO, Jung C, Kim JJ, Hong JH, Kim H, Kim JH, Kil JI, Park CJ, Oh HM, Lee JS, Jin SJ, Um HW, Lee HJ, Oh SJ, Kim JY, Kang HL, Lee SY, Lee KJ, Kang HS. 2005. The genome sequence of the ethanologenic bacterium *Zymomonas mobilis* ZM4. Nat Biotechnol 23:63–68. https://doi.org/10.1038/nbt1045.
9. Loman NJ, Quinlan AR. 2014. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics 30:3399–3401. https://doi.org/10.1093/bioinformatics/btu555.
10. 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.
11. Davis MP, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods 63:41–49. https://doi.org/10.1016/j.ymeth.2013.06.027.
12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski 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.
13. 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.
14. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: rapid annotations using subsystems technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
15. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.
16. Grissa I, Vergnaud G, Pourcel C. 2007. CRISPRFinder: a Web tool to identify clustered regularly interspaced short palindromic repeats. Nucleic Acids Res 35:W52–W57. https://doi.org/10.1093/nar/gkm360.
17. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.