Draft Genome Sequence of *Bacillus* sp. Strain EB106-08-02-XG196, Isolated from High-Nitrate-Contaminated Sediment

Xiaoxuan Ge,[a] Michael P. Thorgersen,[a] Farris L. Poole II,[a] Adam M. Deutschbauer,[b] John-Marc Chandonia,[b] Pavel S. Novichkov,[b] Paul D. Adams,[c] Adam P. Arkin,[b,d] Terry C. Hazen,[e] Michael W. W. Adams[a]

ᵃDepartment of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA
ᵇEnvironmental Genomics and Systems Biology Division, Lawrence Berkeley National Laboratory, Berkeley, California, USA
ᶜMolecular Biosciences and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, California, USA
ᵈDepartment of Bioengineering, University of California, Berkeley, Berkeley, California, USA
ᵉDepartment of Civil and Environmental Engineering, University of Tennessee, Knoxville, Tennessee, USA

ABSTRACT *Bacillus* sp. strain EB106-08-02-XG196 was isolated from a high-nitrate- and heavy metal-contaminated site at the Oak Ridge Reservation in Tennessee. We report the draft genome sequence of this strain to provide insights into the genomic basis for surviving in this unique environment.

The Oak Ridge Reservation (ORR) in Tennessee contains a site termed the S-3 ponds, which has been used for the disposal of waste liquids produced by the Y-12 nuclear plant and is contaminated with nitrate and heavy metals (1). The groundwater at this site has a pH value as low as 3.0 and contains up to 230 mM nitrate and high concentrations of over 20 different metals (2–4). In contrast, the concentration of the metal molybdenum, which is essential for nitrate reduction by microorganisms, is lower in highly contaminated groundwater than it is in the uncontaminated pristine groundwater (5, 6).

Strain EB106-08-02-XG196 (referred to as XG196) was isolated from highly contaminated ORR sediment core EB106, located 21 m downstream of the S-3 ponds (5). It was identified as a *Bacillus* sp. strain by 16S rRNA gene sequencing. The ORR sediment sample was collected and processed under anaerobic conditions (5). Sediment samples (1 g) were incubated anaerobically in 5 ml of a defined medium described previously (7). Mixed carbon sources (formate, acetate, ethanol, lactate, succinate, and glucose, 2 mM each) and 0.1 g/liter yeast extract were used for the enrichment. Cultures were incubated at room temperature for 2 to 7 days and then streaked onto agar plates of defined medium. A single colony of strain XG196 was isolated and cultivated for genomic DNA preparation.

XG196 genomic DNA was prepared using a ZymoBead genomic DNA kit (Zymo Research, Irvine, CA, USA). The Illumina sequencing library was made by shearing DNA to ~800 bp (Covaris, Woburn, MA, USA), and the KAPA HyperPrep kit (Roche, Basel, Switzerland) was then used to add Illumina-compatible adapters to the library. The genome was sequenced using the Illumina HiSeq 2500 platform. This generated 10,798,254 paired-end reads of 151 bp each, totaling 1,630,536,354 bp, with a genome coverage of $215\times$ and an N_{50} value of 283,772 bp. The sequencing reads were trimmed using Trimmomatic v0.36 (8) and assembled *de novo* using SPAdes v3.12.0 (9). An initial genome annotation was performed using Prokka v1.12 (10), executed using KBase (11). The KBase applications kb_trimmomatic v1.2.13, kb_SPAdes v1.1.3, and ProkkaAnnotation v2.0.3 were used for the corresponding steps in the pipeline. Default parameters were used for all software unless otherwise specified. Species trees were built for the XG196 genome using two KBase software packages, GTDB-Tk Classify v1.3.0 and Insert Genome Into

Species Tree v2.1.10. Both rely on phylogenetic marker genes other than 16S rRNA, and both identified *Bacillus niacini* as having the genome most similar to that of XG196 (11). Finally, the KBase-derived assembly was deposited in GenBank. Due to compatibility issues, the GenBank PGAP was rerun on the assembly to produce the final draft genome. We ran the RAST annotation pipeline in KBase v0.1.1 on the GenBank-derived genome to classify features. The final draft genome of strain XG196 contained 6,010,169 bp in 55 contigs longer than 500 bp, with a G+C content of 38.35%. A total of 5,721 coding genes and 234 noncoding features were annotated, including 78 noncoding RNA genes, 42 noncoding regulatory features, and 114 other noncoding features.

Data availability. This whole-genome shotgun project has been deposited in GenBank under the accession number JABWSY000000000. The version described in this paper is the first version (JABWSY000000000.1). The raw sequence reads have been deposited in the SRA under the accession number PRJNA633127. All KBase analyses and data objects are available in the KBase static narrative (https://kbase.us/n/60201/86).

REFERENCES

1. Brooks SC. 2001. Waste characteristics of the former S-3 ponds and outline of uranium chemistry relevant to NABIR Field Research Center studies. Oak Ridge National Laboratory, Oak Ridge, TN.

2. Smith MB, Rocha AM, Smillie CS, Olesen SW, Paradis C, Wu L, Campbell JH, Fortney JL, Mehlhorn TL, Lowe KA, Earles JE, Phillips J, Techtmann SM, Joyner DC, Elias DA, Bailey KL, Hurt RA, Preheim SP, Sanders MC, Yang J, Mueller MA, Brooks S, Watson DB, Zhang P, He Z, Dubinsky EA, Adams PD, Arkin AP, Fields MW, Zhou J, Alm EJ, Hazen TC. 2015. Natural bacterial communities serve as quantitative geochemical biosensors. mBio 6: e00326-15. https://doi.org/10.1128/mBio.00326-15.

3. Thorgersen MP, Lancaster WA, Vaccaro BJ, Poole FL, Rocha AM, Mehlhorn T, Pettenato A, Ray J, Waters RJ, Melnyk RA, Chakraborty R, Hazen TC, Deutschbauer AM, Arkin AP, Adams MWW. 2015. Molybdenum availability is key to nitrate removal in contaminated groundwater environments. Appl Environ Microbiol 81:4976–4983. https://doi.org/10.1128/AEM.00917-15.

4. Revil A, Skold M, Karaoulis M, Schmutz M, Hubbard SS, Mehlhorn TL, Watson DB. 2013. Hydrogeophysical investigations of the former S-3 ponds contaminant plumes, Oak Ridge Integrated Field Research Challenge site, Tennessee. Geophysics 78:EN29–EN41. https://doi.org/10.1190/geo2012-0177.1.

5. Ge X, Vaccaro BJ, Thorgersen MP, Poole FL, Majumder EL, Zane GM, De León KB, Lancaster WA, Moon JW, Paradis CJ, von Netzer F, Stahl DA, Adams PD, Arkin AP, Wall JD, Hazen TC, Adams MWW. 2019. Iron- and aluminium-induced depletion of molybdenum in acidic environments impedes the nitrogen cycle. Environ Microbiol 21:152–163. https://doi.org/10.1111/1462-2920.14435.

6. Moura JJ, Brondino CD, Trincão J, Romao MJ. 2004. Mo and W bis-MGD enzymes: nitrate reductases and formate dehydrogenases. J Biol Inorg Chem 9:791–799. https://doi.org/10.1007/s00775-004-0573-9.

7. Widdel F, Bak F. 1992. Gram-negative mesophilic sulfate-reducing bacteria, p 3352–3378. *In* Balows A, Trüper HG, Dworkin M, Harder W, Schleifer KH (ed), The prokaryotes. Springer, New York, NY.

8. 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.

9. 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.

10. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

11. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, Joachimiak MP, Keegan KP, Kondo S, Kumar V, Land ML, Meyer F, Mills M, Novichkov PS, Oh T, Olsen GJ, Olson R, Parrello B, Pasternak S, Pearson E, Poon SS, Price GA, Ramakrishnan S, Ranjan P, Ronald PC, Schatz MC, Seaver SMD, Shukla M, Sutormin RA, Syed MH, Thomason J, Tintle NL, Wang D, Xia F, Yoo H, Yoo S, Yu D. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.