AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Complete Genome Sequences of Two Interactive Moderate Thermophiles, *Paenibacillus napthalenovorans* 32O-Y and *Paenibacillus* sp. 32O-W

Robert R. Butler III, Jia Wang, Benjamin C. Stark, Jean-François Pombert

Department of Biology, Illinois Institute of Technology, Chicago, Illinois, USA

Microorganisms with the capability to desulfurize petroleum are in high demand with escalating restrictions currently placed on fuel purity. Thermophilic desulfurizers are particularly valuable in high-temperature industrial applications. We report the whole-genome sequences of *Paenibacillus napthalenovorans* 32O-Y and *Paenibacillus* sp. 32O-W, which can and cannot, respectively, metabolize dibenzothiophene.

Received 15 December 2015 **Accepted** 18 December 2015 **Published** 11 February 2016

Citation Butler RR, III, Wang J, Stark BC, Pombert J-F. 2016. Complete genome sequences of two interactive moderate thermophiles, *Paenibacillus napthalenovorans* 32O-Y and *Paenibacillus* sp. 32O-W. Genome Announc 4(1):e01717-15. doi:10.1128/genomeA.01717-15.

Address correspondence to Jean-François Pombert, jpombert@iit.edu.

Biodesulfurization of petroleum is critical in the face of escalating restrictions on fuel purity, and alkylated derivatives of dibenzothiophene (DBT) are among the most common contaminants in petroleum fuel precursors (1). Thermophilic microorganisms capable of desulfurization are especially valuable, as they can be used in high temperature industrial settings (2–5). Previous research isolated two bacterial strains, *Paenibacillus napthalenovorans* 32O-Y, which can metabolize DBT, and *Paenibacillus* sp. 32O-W, which cannot. Curiously, when the two strains are in a mixed culture, 32O-Y metabolizes DBT at a higher rate than in isolation (6). Here, we present the whole-genome sequences of both 32O-Y and 32O-W.

Isolation, purification, and culture of the two species were previously described (6). Genomic DNA was extracted using the PowerSoil (MoBio, Carlsbad, CA, USA) and MasterPure Gram-positive (Epicentre, Madison, WI, USA) DNA purification kits. Long-read sequencing was conducted on a PacBio RS II sequencer (Pacific Biosciences, Menlo Park, CA, USA) using two single-molecule real-time cells and P4-C2 chemistry at the University of Michigan DNA Sequencing Core (Ann Arbor, MI, USA). DNA was purified using AMPure XP beads (Beckman Coulter, Brea, CA, USA) and sheared to 10 kbp (N_{50}s: 32O-W, 6,562 bp; 32O-Y, 7,096 bp) with g-TUBEs (Covaris, Woburn, MA, USA). Additional short-read sequencing was conducted on a MiSeq sequencer (Illumina, San Diego, CA, USA) at Université Laval's Plate-forme d'Analyses Génomiques (Québec, QC, Canada). A 300-bp paired-end protocol was run on TruSeq (Illumina) DNA libraries (insert sizes: 32O-W, 392 ± 106 bp; 32O-Y, 433 ± 123 bp).

Assembly of 32O-W (5,200,139 bp, 56.34% G+C, 93× coverage) was completed using the HGAP2 protocol in SMRT-Analysis version 2.2.0 (7). The assembly was trimmed and then corrected via sequential runs of the RS_Resequencing.1 protocol. Basecalling was further polished by mapping Illumina reads (169× coverage) to the assembly using Geneious version 7.1.9 (8). HGAP2 assembly of 32O-Y produced six contigs. Illumina reads were assembled *de novo* using SPAdes version 3.5.0 (9) and Geneious and compared to the HGAP2 assembly. 32O-Y HGAP2 contigs were joined using overlaps from the other assemblies. This assembly (5,375,858 bp, 49.69% G+C; PacBio 109×, Illumina 122×) was then trimmed and corrected as above.

Genomes were annotated using Prokka version 1.11 (10), with a database containing six reference genomes (*Paenibacillus curdlanolyticus* YK9, NZ_AEDD01000040; *Paenibacillus darwinianus*, NZ_KK082115; *Paenibacillus harenae* DSM 16969, NZ_KE383842; *Paenibacillus pinihumi* DSM 23905, NZ_KE383864; *Paenibacillus* sp. G1, NZ_CBVJ010000001; *Paenibacillus* sp. JDR-2, NC_012914). The RNAmmer version 1.2 (11) switch was used for rRNA annotation, and the tRNA annotations were verified with tRNAscan-SE version 1.3.1 (12). Hypothetical protein annotations were checked with InterProScan versions 5.14-53.0 (13), and conserved domain hits above 1×10^{-10} were amended. An NCBI genome submission check (https://www.ncbi.nlm.nih.gov/genomes/frameshifts/frameshifts.cgi) identified potential pseudogenes (three in 32O-W, seven in 32O-Y). A total of 5,054/4,752 coding sequences, 94/69 tRNAs, 34/19 rRNAs, 1/7 repeat regions, 1/1 tmRNA, and 382/439 signal peptides were annotated in the 32O-Y/32O-W genomes.

Nucleotide sequence accession numbers. Whole-genome sequences were deposited in GenBank under the accession numbers CP013652 (32O-Y) and CP013653 (32O-W). Both species are in the ARS Culture Collection, Peoria, IL, USA, under NRRL catalog numbers B-65351 and B-65352.

ACKNOWLEDGMENTS

We would like to thank the following people: Joelle Salazar and Batzaya Davaadelger (IIT) for the original isolation of *Paenibacillus* sp. from soil samples; Christina McHenry (University of Michigan) for help with PacBio sequencing; Brian Boyle (Université Laval) for help with Illumina

sequencing; and Yukun Sun and Shadi Beidas (IIT) for help with software and scriptwriting.

FUNDING INFORMATION

This work was supported by start-up funds from the Illinois Institute of Technology to Jean-François Pombert.

REFERENCES

1. **Kilbane JJ**. 2006. Microbial biocatalyst developments to upgrade fossil fuels. Curr Opin Biotechnol **17:**305–314. http://dx.doi.org/10.1016/j.copbio.2006.04.005.
2. **Konishi J, Ishii Y, Onaka T, Okumura K, Suzuki M**. 1997. Thermophilic carbon-sulfur-bond-targeted biodesulfurization. Appl Environ Microbiol **63:**3164–3169.
3. **Ishii Y, Konishi J, Okada H, Hirasawa K, Onaka T, Suzuki M**. 2000. Operon structure and functional analysis of the genes encoding thermophilic desulfurizing enzymes of *Paenibacillus* sp. A11-2. Biochem Biophys Res Commun **270:**81–88. http://dx.doi.org/10.1006/bbrc.2000.2370.
4. **Kirimura K, Furuya T, Nishii Y, Ishii Y, Kino K, Usami S**. 2001. Biodesulfurization of dibenzothiophene and its derivatives through the selective cleavage of carbon-sulfur bonds by a moderately thermophilic bacterium *Bacillus subtilis* WU-S2B. J Biosci Bioeng **91:**262–266. http://dx.doi.org/10.1016/S1389-1723(01)80131-6.
5. **Kayser KJ, Cleveland L, Park H-S, Kwak J-H, Kolhatkar A, Kilbane JJ**. 2002. Isolation and characterization of a moderate thermophile, *Mycobacterium phlei* GTIS10, capable of dibenzothiophene desulfurization. Appl Microbiol Biotechnol **59:**737–746. http://dx.doi.org/10.1007/s00253-002-1030-8.
6. **Wang J, Davaadelger B, Salazar JK, Butler RR, III, Pombert J-F, Kilbane JJ, Stark BC**. 2015. Isolation and characterization of an interactive culture of two *Paenibacillus* species with moderately thermophilic desulfurization ability. Biotechnol Lett **37:**2201–2211. http://dx.doi.org/10.1007/s10529-015-1918-x.
7. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.
8. **Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A**. 2012. Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics **28:**1647–1649. http://dx.doi.org/10.1093/bioinformatics/bts199.
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. http://dx.doi.org/10.1089/cmb.2012.0021.
10. **Seemann T**. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.
11. **Lagesen K, Hallin P, Rødland EA, Staerfeldt H-H, Rognes T, Ussery DW**. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res **35:**3100–3108. http://dx.doi.org/10.1093/nar/gkm160.
12. **Lowe TM, Eddy SR**. 1997. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res **25:**955–964. http://dx.doi.org/10.1093/nar/25.5.0955.
13. **Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S**. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics **30:**1236–1240. http://dx.doi.org/10.1093/bioinformatics/btu031.