



Draft Genome Sequences of Thermophiles Isolated from Yates Shaft, a Deep-Subsurface Environment

Nitin K. Singh,^a Courtney Carlson,^{a,b} Rajesh K. Sani,^b  Kasthuri Venkateswaran^a

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA^a; South Dakota School of Mines and Technology, Rapid City, South Dakota, USA^b

ABSTRACT The whole-genome sequences of seven thermophiles that could grow at >55°C, but not at 37°C, were generated. These thermophilic bacteria will play a useful role as model microorganisms, and analyzing their genomes will help to understand the observed production of novel bioactive compounds, including thermostable enzymes and macromolecules.

Several thermophilic spore-forming bacteria were collected from a site known to harbor vast numbers of prokaryotic diversity (1)—the former Homestake Gold Mine in Lead, South Dakota, USA, now known as the Sanford Underground Research Facility (SURF) (2). The microbiome of this environment is relatively little studied and is host to significant evolutionary pressures exerted by its naturally inhospitable conditions (hot, dark, and deep subsurface), rendering it a probable location to find hardy novel microbes. Within SURF, 123 thermophilic isolates were collected from the Yates Shaft (44.35° N, 103.76° W) at a depth of 1,480 m, where the temperature was 55°C and pH was 7.5. Initially, the 16S rRNA gene was sequenced to phylogenetically identify all 123 thermophiles, and a subset of these spore-forming thermophiles was tested for resistance to UVC₂₅₄ irradiation. Based on the results of these two tests, seven strains were chosen for whole-genome sequencing, and phylogenetic analyses showed they belonged to the genus *Geobacillus*.

Uncovering the molecular mechanisms of extremophiles, such as *Geobacillus* spp., will advance knowledge about the resistance of microorganisms to extreme extraterrestrial conditions (UV, etc.) and will play a crucial role in the field of astrobiology. Previously, UV resistance has been tested among large assortments of *Bacillus* spp. (3), but underlying genetic mechanisms of UV resistance have been not been reported in *Geobacillus* spp. In addition, thermophiles and their associated biomolecules are also germane to the development of new biotechnology. For example, the ingenious manipulation of thermostable enzymes (e.g., *Taq* polymerase) has led to the development of widely used molecular biology techniques (4). Furthermore, the industrial applications of thermostable enzymes are widespread in diverse fields (5, 6). They are convenient biocatalysts for these applications due to their ability to maintain activity at extreme temperatures (60 to 70°C) and to resist destruction by organic solvents, detergents, low or high pH, and other denaturing agents (7–9). Operating industrial processes at elevated temperatures with the aid of thermostable enzymes is desirable because it facilitates the use of higher substrate concentrations, lower viscosity, reduced risk of mesophilic microbial contamination, and high reaction rates (10, 11).

In this study, we determined the draft genome sequences of seven strains isolated from SURF. Whole-genome sequencing was done on an Illumina HiSeq2500 instrument with a paired-end module. The NGS QC Toolkit version 2.3 (12) was used to filter the data for high-quality vector- and adaptor-free reads for genome assembly (cutoff read length for high quality, 80%; cutoff quality score, 20). High-quality vector-filtered reads

Received 4 April 2017 Accepted 7 April 2017 Published 1 June 2017

Citation Singh NK, Carlson C, Sani RK, Venkateswaran K. 2017. Draft genome sequences of thermophiles isolated from Yates Shaft, a deep-subsurface environment. *Genome Announc* 5:e00405-17. <https://doi.org/10.1128/genomeA.00405-17>.

Copyright © 2017 Singh 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 Kasthuri Venkateswaran, kjvenkat@jpl.nasa.gov.

TABLE 1 Statistics summary for the seven draft Yates Shaft thermophilic bacterial genome sequences

Strain	Strain designation	NCBI accession no.	Isolation location	No. of contigs	Genome size (bp)	N_{50} (bp)	Final coverage (%)	% GC	No. of error-corrected reads	Total no. of genes
<i>Geobacillus</i> sp. 44B	SURF-44B	NADP00000000	Yates Shaft, SURF	284	3,543,185	35,604	100	44.62	54,051,938	3,676
<i>Geobacillus</i> sp. 44C	SURF-44C	NADQ00000000	Yates Shaft, SURF	121	3,220,397	85,620	100	42.32	66,691,843	3,365
<i>Geobacillus</i> sp. 46C-IIa	SURF-46C-IIa	NADR00000000	Yates Shaft, SURF	51	3,474,909	191,187	100	52.06	67,323,860	3,552
<i>Geobacillus</i> sp. 47C-IIb	SURF-47C-IIb	NADS00000000	Yates Shaft, SURF	167	3,347,490	59,962	100	49.60	58,967,837	3,453
<i>Geobacillus thermocatenuatus</i>	SURF-48B	NAGD00000000	Yates Shaft, SURF	230	3,386,398	43,530	100	52.63	62,078,032	3,626
<i>Geobacillus thermocatenuatus</i>	SURF-114	NAGE00000000	Yates Shaft, SURF	153	3,431,480	52,431	100	52.36	60,187,546	3,591
<i>Geobacillus thermocatenuatus</i>	SURF-189	NAGF00000000	Yates Shaft, SURF	363	3,553,583	32,887	100	52.37	51,162,127	3,854

were assembled with MaSuRCA (13), under the default parameter predicted by the assembler for each genome, and annotated with the help of the Rapid Annotations using Subsystems Technology (RAST) (14). Table 1 summarizes assembly statistics (number of contigs, total genome size, N_{50} size, final coverage, G+C percentage, number of error-corrected reads used for assembly, and number of coding sequences). The genome sizes were in the range of 3.2 to 3.5 Mb per genome. The G+C content was in the range of 42.3 to 52.6% for all of the strains under study.

Accession number(s). The whole-genome sequence data were deposited at DDBJ/EMBL/GenBank under the accession numbers listed in Table 1. The versions described in this paper are the first versions.

ACKNOWLEDGMENTS

We thank several members of the South Dakota School of Mines and Technology and the Biotechnology and Planetary Protection Group for isolating and for help analyzing the thermophiles.

Part of the research reported here was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with NASA and at the South Dakota School of Mines and Technology, under a contract with NASA (grant no. NNX16AQ98A).

REFERENCES

- Rastogi G, Osman S, Kukkadapu R, Engelhard M, Vaishampayan PA, Andersen GL, Sani RK. 2010. Microbial and mineralogical characterizations of soils collected from the deep biosphere of the former Homestake gold mine, South Dakota. *Microb Ecol* 60:539–550. <https://doi.org/10.1007/s00248-010-9657-y>.
- Heise J. 2015. The Sanford Underground Research Facility at Homestake. *AIP Conf Proc* 1672:120003. <https://doi.org/10.1063/1.4928009>.
- Newcombe DA, Schuerger AC, Benardini JN, Dickinson D, Tanner R, Venkateswaran K. 2005. Survival of spacecraft-associated microorganisms under simulated Martian UV irradiation. *Appl Environ Microbiol* 71:8147–8156. <https://doi.org/10.1128/AEM.71.12.8147-8156.2005>.
- Brock TD, Freeze H. 1969. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol* 98:289–297.
- Synowiecki J. 2010. Some applications of thermophiles and their enzymes for protein processing. *Afr J Biotechnol* 9:7020–7025.
- Bhalla A, Bansal N, Kumar S, Bischoff KM, Sani RK. 2013. Improved lignocellulose conversion to biofuels with thermophilic bacteria and thermostable enzymes. *Bioresour Technol* 128:751–759. <https://doi.org/10.1016/j.biortech.2012.10.145>.
- Cowan D, Daniel R, Morgan H. 1985. Thermophilic proteases: properties and applications. *Trends Biotechnol* 3:68–72. [https://doi.org/10.1016/0167-7799\(85\)90080-0](https://doi.org/10.1016/0167-7799(85)90080-0).
- Cowan DA. 1997. Thermophilic proteins: stability and function in aqueous and organic solvents. *Comp Biochem Physiol A Physiol* 118:429–438. [https://doi.org/10.1016/S0300-9629\(97\)00004-2](https://doi.org/10.1016/S0300-9629(97)00004-2).
- Gupta A, Khare SK. 2006. A protease stable in organic solvents from solvent tolerant strain of *Pseudomonas aeruginosa*. *Bioresour Technol* 97:1788–1793. <https://doi.org/10.1016/j.biortech.2005.09.006>.
- Bruins ME, Janssen AE, Boom RM. 2001. Thermozyms and their applications: a review of recent literature and patents. *Appl Biochem Biotechnol* 90:155–186.
- Eichler J. 2001. Biotechnological uses of archaeal enzymes. *Biotechnol Adv* 19:261–278. [https://doi.org/10.1016/S0734-9750\(01\)00061-1](https://doi.org/10.1016/S0734-9750(01)00061-1).
- Patel RK, Jain M. 2012. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. *PLoS One* 7:e30619. <https://doi.org/10.1371/journal.pone.0030619>.
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>.
- 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>.