AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

PROKARYOTES

Draft Genome Sequences from a Novel Clade of *Bacillus cereus Sensu Lato* Strains, Isolated from the International Space Station

Kasthuri Venkateswaran,[a] Aleksandra Checinska Sielaff,[a] Shashikala Ratnayake,[b] Robert K. Pope,[b] Thomas E. Blank,[b] Victor G. Stepanov,[c] George E. Fox,[c] Sandra P. van Tongeren,[d] Clinton Torres,[e] Jonathan Allen,[e] Crystal Jaing,[e] Duane Pierson,[f] Jay Perry,[g] Sergey Koren,[b]* Adam Phillippy,[b]* Joy Klubnik,[b] Todd J. Treangen,[b] M. J. Rosovitz,[b] Nicholas H. Bergman[b]

Biotechnology and Planetary Protection Group, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA[a]; National Biodefense Analysis and Countermeasures Center, Ft. Detrick, Maryland, USA[b]; Department of Biology and Biochemistry, University of Houston, Houston, Texas, USA[c]; Department of Medical Microbiology, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands[d]; Lawrence Livermore National Laboratory, Berkeley, California, USA[e]; Johnson Space Center, Houston, Texas, USA[f]; Marshall Space Flight Center, Huntsville, Alabama, USA[g]

ABSTRACT The draft genome sequences of six *Bacillus* strains, isolated from the International Space Station and belonging to the *Bacillus anthracis-B. cereus-B. thuringiensis* group, are presented here. These strains were isolated from the Japanese Experiment Module (one strain), U.S. Harmony Node 2 (three strains), and Russian Segment Zvezda Module (two strains).

Among the six *Bacillus cereus sensu lato* strains reported here, three U.S. Harmony Node 2 isolates (ISSFR-3F, ISSFR-9F, and ISSFT-23F) and one Japanese Experiment Module isolate (JEM-2) were sequenced and assembled with both Illumina MiSeq and PacBio RSII sequence data. The remaining assemblies, including two Russian isolates, were sequenced and assembled with only MiSeq data. The MiSeq runs yielded on average 24 to 54 million 300-bp reads (from 1,402× to 3,093× average coverage), while PacBio yielded 4,000 to 116,000 reads (from 7× to 202× average coverage) (Table 1). Due to the extremely high coverage (>1,000×), Illumina MiSeq reads were randomly down-sampled to 100× using an estimated genome size of 5.3 Mb, resulting in an average of 1.2 to 1.5 million paired-end reads per isolate. Next, the down-sampled reads were assembled using iMetAMOS (1) with IDBA_UD and SPAdes (2). IDBA_UD was selected as the best assembly for all six isolates. Low confidence bases within the selected IDBA_UD (3) assemblies were masked out by mapping all reads to the assembled contigs and detecting conflicting variants with FreeBayes (4). The PacBio reads were assembled following the methods described by Berlin et al. (5) with Celera Assembler version 8.3rc1 and polished with Quiver (6). A second round of polishing was performed post-Quiver using the available MiSeq data as input to Pilon (7).

The six International Space Station (ISS) isolates were aligned (NUCmer [8], Parsnp [9]) against members of the *B. cereus sensu lato* group genomes (10). The PacBio assemblies were used for all isolates with sufficient read coverage, and Illumina assemblies were used for the remaining isolates. Based on genome size estimates (5.2 to 5.3 Mb), NUCmer pairwise alignments (>99.9% average pairwise nucleotide identity) and maximum-likelihood phylogenetic placement, all six isolates were found to exhibit a very high degree of similarity.

Received 31 May 2017 **Accepted** 5 June 2017 **Published** 10 August 2017

Citation Venkateswaran K, Checinska Sielaff A, Ratnayake S, Pope RK, Blank TE, Stepanov VG, Fox GE, van Tongeren SP, Torres C, Allen J, Jaing C, Pierson D, Perry J, Koren S, Phillippy A, Klubnik J, Treangen TJ, Rosovitz MJ, Bergman NH. 2017. Draft genome sequences from a novel clade of *Bacillus cereus sensu lato* strains, isolated from the International Space Station. Genome Announc 5:e00680-17. https://doi.org/10.1128/genomeA.00680-17.

Address correspondence to Kasthuri Venkateswaran, kjvenkat@jpl.nasa.gov, or Nicholas H. Bergman, nicholas.bergman@nbacc.dhs.gov.

* Present address: Sergey Koren and Adam Phillippy, Genome Informatics Section, Computational and Statistical Genomics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland, USA.

T.J.T., M.J.R., and N.H.B. contributed equally to this article.

TABLE 1 *De novo* assembly statistics of the *Bacillus* spp. isolated from the ISS

Isolate no.	ISS module	GenBank accession no.	Sequencing platform	No. of reads	Avg. read coverage (×)	N_{50} (Mb)	No. of contigs
ISSFR-3F	U.S.[a]	CP018931	PacBio	116,418	202	5.2	2
ISSFR-9F	U.S.	CP018933	PacBio	15,939	42	5.2	2
ISSFR-23F	U.S.	MSMO00000000	PacBio	8,145	17	0.6	16
JEM-2	Japan[b]	CP018935	PacBio	40,305	80	5.2	2
S1-R4H1-FB	Russian Federation[c]	NBNT00000000	Illumina	4,275,902	242	0.1	329
S2-R3J1-FB-BA1	Russian Federation	NBNR00000000	Illumina	11,266,760	638	0.4	388

[a]U.S., U.S. Segment Harmony Node 2.
[b]Japan, Kibo Japanese Experiment Module (JEM).
[c]Russian Federation, Russian Segment Zvezda Module.

Finally, due to their high genomic similarity to the *B. anthracis* type strain (>98% average nucleotide identity), all six genomes were examined for evidence of pathogenicity. However, none of the commonly known *B. anthracis* signature elements were identified. Specifically, all six ISS isolates (i) contain the *plcR* (11) ancestral "C" allele, which has been used in large-scale phylogenetic analyses to distinguish *B. anthracis* strains from the rest of the *B. cereus* group; (ii) lack significant hits to pXO1 and pXO2 plasmids; and (iii) are phylogenetically placed outside of the *B. anthracis* clade. Results were consistent with a comparative genomic analysis performed using the Lawrence Livermore National Laboratory Microbial Threat Characterization Pipeline. Altogether, the collective genomic evidence supports the conclusion that the six ISS isolates represent a novel *Bacillus* sp. located within the *B. cereus sensu lato* group.

Accession number(s). The complete genome sequences were deposited in NCBI under the accession numbers listed in Table 1 and can be accessed from the NASA GeneLab system (GLDS-64; https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-64).

ACKNOWLEDGMENTS

Part of the research described in this publication was carried out at the Jet Propulsion Laboratory (JPL), California Institute of Technology, under a contract with NASA. The contributions of S.R., R.K.P., T.E.B., S.K., A.P., J.K., T.J.T., M.J.R., and N.H.B. were funded under contract no. HSHQDC-07-C-00020 awarded by the Department of Homeland Security (DHS) Science and Technology Directorate (S&T) for the management and operation of the National Biodefense Analysis and Countermeasures Center (NBACC), a Federally Funded Research and Development Center. This research was also funded by 2012 Space Biology (NNH12ZTT001N) grant no. 19-12829-26 under Task Order NNN13D111T awarded to K.V., which also funded the postdoctoral fellowship for A.C.S., and by JPL subcontract 1506453 to G.E.F. The research carried out at Lawrence Livermore National Laboratory was funded by 2014 Space Biology (NNH14ZTT002N) grant no. NNX15AJ29G awarded to C.J., which also funded J.A. and C.T. The contribution of S.P.V.T. was supported by the European Space Agency (MAP Project no. AO-LS-99-MAP-LSS-018 "biofilms") and SRON (MG-064/MG-068).

We thank Kazuyuki Tasaki, Director of the JEM Utilization Center, and Julie Robinson, Chief Scientist of the ISS (U.S.) for arranging Japanese Aerospace Exploration Agency (JAXA) approval to use information on strains collected during the routine operation and maintenance of the JEM-Kibo Module. We also thank Hermie J. M. Harmsen and Hendrik I. J. Roest for providing the Russian isolates and logistical support, as well as Jonathan Hnath, Kathy Fronda, Gregory Horn, and Henry Lupari from NBACC for technical assistance.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the DHS or S&T. In no event, shall DHS, NBACC, S&T, or Battelle National Biodefense Institute have any responsibility or liability for any use, misuse, inability to use, or reliance upon the information contained herein. DHS does not endorse any products or commercial services mentioned in this publication.

REFERENCES

1. Koren S, Treangen TJ, Hill CM, Pop M, Phillippy AM. 2014. Automated ensemble assembly and validation of microbial genomes. BMC Bioinformatics 15:126. https://doi.org/10.1186/1471-2105-15-126.
2. 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.
3. Peng Y, Leung HC, Yiu SM, Chin FY. 2012. IDBA-UD: a de novo assembler for single-cell and metagenomic sequencing data with highly uneven depth. Bioinformatics 28:1420–1428. https://doi.org/10.1093/bioinformatics/bts174.
4. Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. arXiv:1207.3907. https://arxiv.org/abs/1207.3907.
5. Berlin K, Koren S, Chin CS, Drake JP, Landolin JM, Phillippy AM. 2015. Assembling large genomes with single-molecule sequencing and locality-sensitive hashing. Nat Biotechnol 33:623–630. https://doi.org/10.1038/nbt.3238.
6. Chin CS, 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. https://doi.org/10.1038/nmeth.2474.
7. 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.
8. Delcher AL, Salzberg SL, Phillippy AM. 2003. Using MUMmer to identify similar regions in large sequence sets. Curr Protoc Bioinformatics Chapter 10:Unit 10.3. https://doi.org/10.1002/0471250953.bi1003s00.
9. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The Harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. Genome Biol 15:524. https://doi.org/10.1186/PREACCEPT-2573980311437212.
10. Zwick ME, Joseph SJ, Didelot X, Chen PE, Bishop-Lilly KA, Stewart AC, Willner K, Nolan N, Lentz S, Thomason MK, Sozhamannan S, Mateczun AJ, Du L, Read TD. 2012. Genomic characterization of the *Bacillus cereus* sensu lato species: backdrop to the evolution of Bacillus anthracis. Genome Res 22:1512–1524. https://doi.org/10.1101/gr.134437.111.
11. Easterday WR, Van Ert MN, Simonson TS, Wagner DM, Kenefic LJ, Allender CJ, Keim P. 2005. Use of single nucleotide polymorphisms in the *plcR* gene for specific identification of *Bacillus anthracis*. J Clin Microbiol 43:1995–1997. https://doi.org/10.1128/JCM.43.4.1995-1997.2005.