AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequences of *Bacillaceae* Strains Isolated from the International Space Station

Robert Daudu,[a] Nitin K. Singh,[a] Jason M. Wood,[a] Marilyne Debieu,[b] Niamh B. O'Hara,[b,c] Christopher E. Mason,[d,e] Kasthuri Venkateswaran[a]

[a]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, USA
[b]Biotia, New York, New York, USA
[c]Department of Cell Biology, College of Medicine, SUNY Downstate Health Sciences University, Brooklyn, New York, USA
[d]Department of Physiology and Biophysics, Weill Cornell Medicine, New York, New York, USA
[e]The WorldQuant Initiative for Quantitative Prediction, Weill Cornell Medicine, New York, New York, USA

ABSTRACT The draft genome sequences of 29 bacterial isolates belonging to the family *Bacillaceae* were collected from the International Space Station, assembled, and identified. Further analysis of these sequences will enable us to understand their roles for space and biotechnological applications.

The family *Bacillaceae* is composed of mostly aerobic or facultatively anaerobic rods with a Gram-positive cell wall. Members of this group are ubiquitous in nature, found in soil, sediments, air, freshwater, marine ecosystems, and foods and in extreme environments with high temperatures (1). *Bacillaceae* groups are reported to produce a wide range of valuable and thermostable extracellular enzymes (2). The majority of *Bacillaceae* can form endospores that are crucial for survival in harsh environments (3, 4). This spore-forming property makes it important for NASA planetary protection purposes because spore-forming bacteria are more likely to survive sterilization procedures and the harsh environments of outer space, making them a potential threat for forward contamination (5).

This report presents the draft genome sequences of three species (*Bacillus amyloliquefaciens*, *Bacillus cereus*, and *Paenibacillus polymyxa*) belonging to the family *Bacillaceae*, isolated from predefined surfaces at various locations on the International Space Station. The samples were collected using premoistened polyester wipes (6), brought back to Earth, aseptically transferred to a 500-ml bottle containing 200 ml of phosphate-buffered saline, mixed by vigorously shaking, and concentrated via a filtration system. An appropriate aliquot of each sample was inoculated onto either Reasoner's 2A (R2A) (25°C; 7 days) or blood (37°C; 2 days) agar. A single colony from each incubated plate was used for genomic DNA extraction using the ZymoBIOMICS MagBead DNA kit according to the manufacturer's instructions (Zymo Research, USA). All isolates were identified by 16S rRNA sequencing based on similarity to their type strain 16S rRNA, as previously reported (6). To create the whole-genome sequences (WGS) of these strains, shotgun libraries were prepared using the Illumina Nextera Flex protocol (7), using NovaSeq 6000 S4 flow cell 2×150 paired-end (PE) sequencing. Verification of the quality of the raw sequencing data was carried out using FastQC v0.11.7 (8). Quality control for adapter trimming and quality filtering were performed using fastp v0.20.0 (9), and then SPAdes v3.11.1 (10) was used to assemble all the cleaned sequences. Fastp quality control was based on the following three parameters: (i) correction of mismatches in overlapped regions of paired-end reads, (ii) trimming of autodetected adapter sequences, and (iii) quality trimming at the 5′ and 3′ ends. To determine the quality of the assembled sequences, the number of contigs, the N_{50} value, and the total

Citation Daudu R, Singh NK, Wood JM, Debieu M, O'Hara NB, Mason CE, Venkateswaran K. 2020. Draft genome sequences of *Bacillaceae* strains isolated from the International Space Station. Microbiol Resour Announc 9:e00701-20. https://doi.org/10.1128/MRA.00701-20.

Editor Steven R. Gill, University of Rochester School of Medicine and Dentistry

Address correspondence to Kasthuri Venkateswaran, kjvenkat@jpl.nasa.gov.

Received 6 July 2020
Accepted 2 October 2020
Published 29 October 2020

TABLE 1 Summary of draft whole-genome sequences of *Bacillaceae* strains isolated from various International Space Station environmental surfaces

Sample name	Bacterial species	WGS accession no.	SRA accession no.	Isolation location[a]	No. of contigs	Genome size (bp)	N_{50} (bp)	Depth of coverage (×)	G+C content (%)	No. of filtered reads
IIF7SW-P1	*Bacillus amyloliquefaciens*	JABVDW000000000	SRR11948426	Overhead 3	33	3,908,309	234,631	745.98	46.42	26,708,590
IIF7SW-P2	*Bacillus amyloliquefaciens*	JABVDS000000000	SRR11948402	Overhead 3	34	3,906,939	268,709	866.52	46.43	33,191,426
IIF7SW-P3	*Bacillus amyloliquefaciens*	JABVDR000000000	SRR11948401	Overhead 3	33	3,908,243	234,631	743.30	46.42	27,127,172
IIF7SW-P4	*Bacillus amyloliquefaciens*	JABVDP000000000	SRR11948399	Overhead 3	31	3,906,699	268,709	571.88	46.43	21,235,408
IIF7SW-P5	*Bacillus amyloliquefaciens*	JABVDV000000000	SRR11948415	Overhead 3	36	3,907,462	268,709	495.54	46.43	18,131,724
IIF7SW-B1	*Bacillus amyloliquefaciens*	JABVDX000000000	SRR11948427	Overhead 3	32	3,907,264	268,709	542.41	46.42	20,063,358
IIF7SW-B4	*Bacillus amyloliquefaciens*	JABVDQ000000000	SRR11948400	Overhead 3	22	4,214,765	1,035,357	358.93	45.76	16,420,634
IIF7SC-B1	*Bacillus amyloliquefaciens*	JABVDT000000000	SRR11948403	Field control	36	3,907,040	268,709	577.23	46.42	21,558,310
IIF7SC-B4	*Bacillus amyloliquefaciens*	JABVDU000000000	SRR11948404	Field control	35	3,907,001	268,702	480.80	46.42	18,065,320
IIF4SW-P2	*Bacillus thuringiensis*	JABVDN000000000	SRR11948425	Dining table	18	5,303,558	2,681,506	381.70	35.08	24,097,822
IIF4SW-P3	*Bacillus thuringiensis*	JABVDI000000000	SRR11948420	Dining table	18	5,302,927	2,862,962	354.91	35.07	21,342,744
IIF4SW-P4	*Bacillus thuringiensis*	JABVDH000000000	SRR11948419	Dining table	17	5,303,162	3,967,155	416.52	35.07	24,652,908
IIF4SW-P5	*Bacillus thuringiensis*	JABVDM000000000	SRR11948424	Dining table	18	5,303,374	2,863,400	377.68	35.08	24,892,178
IIF4SW-B1	*Bacillus thuringiensis*	JABVDJ000000000	SRR11948421	Dining table	17	5,302,884	3,967,588	452.68	35.07	25,835,874
IIF4SW-B2	*Bacillus thuringiensis*	JABVDG000000000	SRR11948418	Dining table	18	5,303,293	2,863,295	450	35.08	25,806,770
IIF4SW-B3	*Bacillus thuringiensis*	JABVDK000000000	SRR11948422	Dining table	18	5,303,485	2,863,488	399.11	35.08	22,970,622
IIF2SG-B1	*Bacillus thuringiensis*	JABVDL000000000	SRR11948423	CRV2-L2	19	5,303,842	2,682,377	388.39	35.08	20,471,002
IIF3SG-B3	*Bacillus thuringiensis*	JABVDF000000000	SRR11948417	CRV2-L3	18	5,303,031	2,863,399	291.96	35.07	18,136,072
IIF7SG-B4	*Bacillus thuringiensis*	JABVDO000000000	SRR11948398	CRV2-L7	17	5,302,496	3,967,299	349.55	35.08	20,338,930
IIF1SG-B5	*Bacillus thuringiensis*	JABVDE000000000	SRR11948416	CRV2-L1	18	5,303,272	2,864,997	162.05	35.07	11,404,858
IIF2*SW-P2	*Paenibacillus polymyxa*	JABVCY000000000	SRR11948409	WHC	34	5,788,525	1,496,248	298.66	45.51	18,337,554
IIF2*SW-P4	*Paenibacillus polymyxa*	JABVCV000000000	SRR11948406	WHC	30	5,790,156	1,573,833	416.52	45.51	25,137,632
IIF2SW-B2	*Paenibacillus polymyxa*	JABVDB000000000	SRR11948412	WHC	35	5,788,707	1,496,248	412.5	45.51	24,747,812
IIF5SW-B3	*Paenibacillus polymyxa*	JABVDC000000000	SRR11948413	Overhead 4	34	5,789,544	1,497,677	443.30	45.51	26,964,124
IIF5SW-B4	*Paenibacillus polymyxa*	JABVDA000000000	SRR11948411	Overhead 4	36	5,789,252	1,496,248	310.71	45.51	18,843,168
IIF8SW-P3	*Paenibacillus polymyxa*	JABVCW000000000	SRR11948407	Crew quarters	34	5,790,736	670,850	424.55	45.51	26,480,032
IIF8SW-P4	*Paenibacillus polymyxa*	JABVCZ000000000	SRR11948410	Crew quarters	32	5,790,548	1,573,833	364.29	45.51	22,062,790
IIF8SW-P5	*Paenibacillus polymyxa*	JABVCU000000000	SRR11948405	Crew quarters	33	5,789,239	1,496,161	354.91	45.51	21,821,506
IIF8SW-B4	*Paenibacillus polymyxa*	JABVCX000000000	SRR11948408	Crew quarters	37	5,789,049	670,850	362.95	45.51	22,771,816

[a] WHC, waste and hygiene compartment; CRV, commercial resupply vehicle. Hyphenated designations indicate the CRV number followed by the location.

length were calculated using QUAST v5.0.2 (11). Default parameters were used for all software. The average nucleotide identity (ANI) (12) was calculated using OrthoANIu by comparing each of the 29 *Bacillaceae* scaffolds to the WGS of the respective type strains. The ANI range was 93.98% to 94.17% for *B. amyloliquefaciens*, and the ANI was 98% for both *B. cereus* and *P. polymyxa*. All other genomic statistics are given in Table 1.

Data availability. The WGS and raw data have been deposited in GenBank under the BioProject accession number PRJNA637984 and also in the NASA GeneLab system (https://genelab-data.ndc.nasa.gov/genelab/accession/GLDS-303/). The version described in this paper is the final version.

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. Government sponsorship is acknowledged. This research was funded by a 2012 Space Biology NNH12ZTT001N grant (no. 19-12829-26) under task order NNN13D111T award to K.V., which also funded a postdoctoral fellowship for J.M.W., a JPL graduate fellowship to R.D., and a subcontract to Biotia, Inc.

We thank astronaut Terry Virts for collecting samples aboard the International Space Station (ISS) and the Implementation Team (Fathi Karouia) at NASA Ames Research Center for coordinating this effort. We thank Ryan Kemp (Zymo Corporation) for DNA extraction and Dan Butler (Cornell Medicine) for generating the shotgun sequencing using NovaSeq. R.D. thanks Wei-Jen Lin for providing guidance and direction regarding undertaking research at JPL.

REFERENCES

1. Beladjal L, Gheysens T, Clegg JS, Amar M, Mertens J. 2018. Life from the ashes: survival of dry bacterial spores after very high temperature exposure. Extremophiles 22:751–759. https://doi.org/10.1007/s00792-018-1035-6.

2. Mandic-Mulec I, Stefanic P, van Elsas JD. 2015. Ecology of *Bacillaceae*. Microbiol Spectr 3:TBS-0017-2013. https://doi.org/10.1128/microbiolspec.TBS-0017-2013.
3. Gopal N, Hill C, Ross PR, Beresford TP, Fenelon MA, Cotter PD. 2015. The prevalence and control of Bacillus and related spore-forming bacteria in the dairy industry. Front Microbiol 6:1418. https://doi.org/10.3389/fmicb.2015.01418.
4. Ulrich N, Nagler K, Laue M, Cockell CS, Setlow P, Moeller R. 2018. Experimental studies addressing the longevity of *Bacillus subtilis* spores—the first data from a 500-year experiment. PLoS One 13:e0208425. https://doi.org/10.1371/journal.pone.0208425.
5. Horneck G, Moeller R, Cadet J, Douki T, Mancinelli RL, Nicholson WL, Panitz C, Rabbow E, Rettberg P, Spry A, Stackebrandt E, Vaishampayan P, Venkateswaran KJ. 2012. Resistance of bacterial endospores to outer space for planetary protection purposes—experiment PROTECT of the EXPOSE-E mission. Astrobiology 12:445–456. https://doi.org/10.1089/ast.2011.0737.
6. Checinska Sielaff A, Urbaniak C, Mohan GBM, Stepanov VG, Tran Q, Wood JM, Minich J, McDonald D, Mayer T, Knight R, Karouia F, Fox GE, Venkateswaran K. 2019. Characterization of the total and viable bacterial and fungal communities associated with the International Space Station surfaces. Microbiome 7:50. https://doi.org/10.1186/s40168-019-0666-x.
7. Singh NK, Bezdan D, Checinska Sielaff A, Wheeler K, Mason CE, Venkateswaran K. 2018. Multi-drug resistant *Enterobacter bugandensis* species isolated from the International Space Station and comparative genomic analyses with human pathogenic strains. BMC Microbiol 18:175. https://doi.org/10.1186/s12866-018-1325-2.
8. Andrews S. 2015. FastQC: a quality tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
9. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
10. 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.
11. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
12. Yoon S-H, Ha S-M, Lim J, Kwon S, Chun J. 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. Antonie Van Leeuwenhoek 110:1281–1286. https://doi.org/10.1007/s10482-017-0844-4.