




Draft Genome Sequences of Two Potentially Novel *Bacillus* Isolates from Backyard and Commercial Chicken Gastrointestinal Tracts

 Jitendra Keshri,^a Rocio Ramirez,^{a,b} Mark E. Berrang,^c Brian B. Oakley^a

^aCollege of Veterinary Medicine, Western University of Health Sciences, Pomona, California, USA

^bGraduate College of Biomedical Sciences, Western University of Health Sciences, Pomona, California, USA

^cUSDA Agricultural Research Service, U.S. National Poultry Research Center, Athens, Georgia, USA

ABSTRACT Here, we present the draft genome sequences of two *Bacillus* strains, HF117_J1_D and USDA818B3_A, isolated in Pomona, California, from the gastrointestinal (GI) tract of backyard and commercial broiler chickens, respectively. The draft genomes of both strains appear to represent novel species.

A total of 379 *Bacillus* species have been recognized as of 2019 (<http://www.bacterio.net/bacillus.html>). We isolated *Bacillus* sp. strain HF117_J1_D from fecal samples of a backyard chicken and strain USDA818B3_A from the cecal contents of a commercial broiler chicken in Pomona, California. Both strains were isolated aerobically at 37°C after 24 to 48 h of growth. Strain HF117_J1_D was grown on tryptic soy agar (TSA), and strain USDA818B3_A was grown on brain heart infusion (BHI) agar (BD, Franklin Lakes, NJ, USA). Genomic DNA was extracted with the TaKaRa NucleoSpin microbial DNA isolation kit. Isolates were initially identified as *Bacillus* spp. using 16S rRNA sequencing. The genomes of both strains were sequenced using the Illumina NovaSeq platform through the IIGB Genomics Core at the University of California, Riverside.

Dual-indexed libraries suitable for 150-bp paired-end reads on the NovaSeq platform were constructed using a plexWell 96 library preparation kit for Illumina sequencing platforms (seqWell, Inc., Beverly, MA, USA), and 1,667,450 and 4,753,763 paired-end 150-bp reads were generated for strains HF117_J1_D and USDA818B3_A, respectively. Default parameters for all software were used during analyses unless otherwise specified. Quality filtering of the reads was performed using Trimmomatic v0.39 (1). High-quality reads were used for *de novo* genome assembly with SPAdes v3.12.0 using the careful assembly method (2). Scaffolds were filtered for a minimum 200-bp read length and 30× coverage. The quality of the subsequent assemblies was assessed using QUAST (3). Genome completeness rates were found to be 92.7% and 99.38% for strains HF117_J1_D and USDA818B3_A, respectively, using CheckM lineage (4) and the Microbial Genomes Atlas (MiGA; <http://microbial-genomes.org/>) workflow (5). A search for similar genomes using the Mash/MinHash algorithm (6) showed the nearest reference representative genome to strains HF117_J1_D and USDA818B3_A to be *Bacillus fordii* DSM 16014 (GenBank accession no. [ARGB000000000](https://genbank.ncbi.nlm.nih.gov/GenBank/FASTA/ARGB000000000)) (Mash distance, 16%) and *Bacillus bataviensis* 2482 (v2) (GenBank accession no. [VIVN010000000](https://genbank.ncbi.nlm.nih.gov/GenBank/FASTA/VIVN010000000)) (Mash distance, 17%). The species identification tool Specl (7), based on 40 universal single-copy marker genes, confirmed that both genomes could not be assigned to any species cluster. Taxonomic novelty at the species level for both genomes was verified against reference genomes of NCBI's RefSeq database using MiGA analysis. The assembled genomes were annotated using the IMG/M Pipeline (8), and average coverage was calculated using the BMap tool. Results of the assembly quality analysis conducted using the MiGA workflow based on 106 essential genes were excellent (85.1%) and high (78.2%) for

Citation Keshri J, Ramirez R, Berrang ME, Oakley BB. 2020. Draft genome sequences of two potentially novel *Bacillus* isolates from backyard and commercial chicken gastrointestinal tracts. *Microbiol Resour Announc* 9:e00492-20. <https://doi.org/10.1128/MRA.00492-20>.

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

Copyright © 2020 Keshri 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 Brian B. Oakley, boakley@westernu.edu.

Received 1 May 2020

Accepted 2 May 2020

Published 28 May 2020

TABLE 1 Assembly and annotation metrics

Statistic	Data for strain:	
	HF117_J1_D	USDA818B3_A
No. of quality-filtered paired-end reads	1,667,450	3,939,604
No. of scaffolds	210	90
N_{50} (bp)	48,679	270,360
Largest scaffold length (bp)	242,713	717,443
Assembly length (bp)	3,817,981	5,714,796
Avg genome coverage (×)	82	176
G+C content (%)	40.89	39.09
No. of rRNA genes	17	16
No. of tRNA genes	55	44
No. of functional protein coding genes	2,981	4,299
No. of hypothetical protein coding genes	808	1,208
No. of genes associated with KEGG pathways	1,068	1,505
No. of genes associated with KEGG orthology	1,829	2,705
GenBank accession no.	JAABNQ000000000	JAABNN000000000
SRA accession no.	SRR10903101	SRR10903100

HF117J1_D and USDA818B3_A, respectively. The assembly and annotation metrics are presented in Table 1.

Based on the “compare genome analysis” function in IMG, which uses genome-wide average nucleotide identity (ANI), the maximum pairwise ANI of strain HF117_J1_D with known genomes of *Bacillus* is 82.3% with *Bacillus fordii* DSM 16014 (GenBank accession no. ARGB000000000) (aligned fraction, 71.9%). The highest ANI of USDA818B3_A with known genomes of *Bacillus* is 83.1% with *Bacillus bataviensis* 2482 (v2) (GenBank accession no. VIVN010000000) (aligned fraction, 71.5%). The ANI between the two isolates is only 69.1%, suggesting that strains HF117_J1_D and USDA818B3_A are distantly related to each other. The availability of these genomes will be useful in further studies.

Data availability. The draft genome sequence and raw read sequences for both strains have been deposited in DDBJ/EMBL/GenBank under the accession numbers listed in Table 1.

ACKNOWLEDGMENT

Funding was provided by USDA NIFA grants (no. 1015210 and 1011327) to the Western University of Health Sciences College of Veterinary Medicine and B.B.O.

REFERENCES

- 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>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>.
- 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>.
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
- Rodríguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. *Nucleic Acids Res* 46:W282–W288. <https://doi.org/10.1093/nar/gky467>.
- Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, Phillippy AM. 2016. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biol* 17:132. <https://doi.org/10.1186/s13059-016-0997-x>.
- Mende DR, Sunagawa S, Zeller G, Bork P. 2013. Accurate and universal delineation of prokaryotic species. *Nat Methods* 10:881–884. <https://doi.org/10.1038/nmeth.2575>.
- Chen IA, Chu K, Palaniappan K, Pillay M, Ratner A, Huang J, Huntemann M, Varghese N, White JR, Seshadri R, Smirnova T, Kirton E, Jungbluth SP, Woyke T, Eloe-Fadrosh EA, Ivanova NN, Kyrpides NC. 2019. IMG/M v.5.0: an integrated data management and comparative analysis system for microbial genomes and microbiomes. *Nucleic Acids Res* 47:D666–D677. <https://doi.org/10.1093/nar/gky901>.