Draft Genome Sequences of *Bacillus subtilis* Strains TNC1(2019), TNC3(2019), and TNW1(2019), as Well as *Bacillus velezensis* Strains TNC2(2019) and TNW2(2019), Isolated from Cabbage Kimchee

Gabriel G. Perron,[a,b,c] Tejaswee Neupane,[a,b] Rachael A. Mendoza[a]

[a]Department of Biology, Reem-Kayden Center for Science and Computation, Bard College, Annandale-on-Hudson, New York, USA
[b]Bard Food Lab, Center for Experimental Humanities, Bard College, Annandale-on-Hudson, New York, USA
[c]Center for the Study of Land, Air, and Water, Bard College, Annandale-on-Hudson, New York, USA

ABSTRACT We report the draft genome sequences of five novel *Bacillus* strains isolated from five different batches of fermented Napa cabbage kimchee. Strains TNC1(2019), TNC3(2019), and TNW1(2019) were identified as *Bacillus subtilis*, while TNC2(2019) and TNW2(2019) were identified as *Bacillus velezensis*.

Bacillus is a genetically and metabolically diverse genus of bacteria that have been used for the development of environmental and medicinal tools (1). The genus also plays an important role in the making of many traditional fermented foods, such as natto and miso, two products made from fermented soybeans (2, 3). For this reason, the characterization of novel *Bacillus* strains is of interest for the development of novel agricultural and biotechnological applications such as probiotics (4).

To investigate the genetic diversity of *Bacillus* strains in fermented food, we isolated the most frequent bacterial strain found in five different batches of Napa cabbage kimchee. In addition to Napa cabbage, the five kimchee preparations included the same quantities of Asian pear, garlic, chili powder, dried shrimp, and ginger and were fermented for 14 days. For each ferment, bacterial count and diversity were estimated from a 10-fold serial dilution initiated with 1 ml of brine. We then plated 100 μl of each dilution onto LB plates and incubated the plates at 37°C for 48 h. For each sample, we picked one colony from the most frequent colony type, which was then grown in LB broth at 37°C for 48 h. The bacterial cultures were then streaked onto LB plates to confirm the unique colony morphology, and we used 1.8 ml of overnight cultures to extract genomic DNA using the DNeasy UltraClean microbial kit (Qiagen, Hilden, Germany). Whole-genome sequencing was conducted at the Microbial Genome Sequencing Center, LLC (Pittsburgh, PA), using the 151-bp paired-end read libraries designed for the Illumina MiSeq platform. We obtained a total of 3,653,401 raw reads for TNC1(2019), 3,551,954 raw reads for TNC2(2019), 3,581,451 raw reads for TNC3(2019), 3,863,404 raw reads for TNW1(2019), and 3,466,478 raw reads for TNW2(2019).

We assembled the genomes using the bioinformatic pipeline described by Shrestha et al. (5), using default parameters unless specified otherwise. Briefly, raw reads were trimmed using Trimmomatic v0.36 with the following parameters: slidingwindow, 4:15; leading, 3; trailing, 3; and minlen, 50 (6). Trimmed reads were then assembled using the SPAdes *de novo* assembler v3.11, testing assemblies with kmer values of 21, 33, 55, 77, 99, and 127 (7); contigs smaller than 500 bp or with low coverage (lower than or equal to 2×) were filtered out. Assembly statistics were estimated using QUAST v4.5 (8) and BBMap v35.82 (9). We also estimated genome completeness and confirmed that no

TABLE 1 Summary of the draft genome sequences for the *Bacillus* sp. strains described in this study

Isolate[a]	Coordinates	Closest species match (DDH [%])	No. of contigs	Genome size (bp)	Completeness (%)	G+C content (%)	N_{50} (bp)	Median read depth (×)	Genome accession no.	SRA accession no.
TNC1(2019)	42°1′12″N, 73°54′28″W	*B. subtilis* ATCC 6051 (91.7)	15	4,080,193	99.59	43.7	2,103,272	105	JAAEBK000000000	SRR10969390
TNC2(2019)	42°1′49″N, 73°54′17″W	*B. velezensis* NRRL B-41580 (93.1)	24	4,032,556	99.59	46.4	414,810	101	JAAEBJ000000000	SRR10969389
TNC3(2019)	42°1′47″N, 73°54′16″W	*B. subtilis* ATCC 6051 (87.7)	18	4,070,250	99.59	43.7	1,024,243	102	JAAEBI000000000	SRR10969388
TNW1(2019)	42°1′12″N, 73°54′28″W	*B. subtilis* ATCC 6051 (86.1)	23	4,044,726	99.59	43.5	1,057,433	113	JAAEBH000000000	SRR10969387
TNW2(2019)	42°1′26″N, 73°54′23″W	*B. velezensis* FZB42 (91.0)	40	4,050,273	99.59	46.1	295,280	98	JAAEBG000000000	SRR10969386

[a] All isolates were obtained from kimchee in Annandale, New York.

contamination was detected using CheckM v1.0.18 (10). Finally, species identity, based on DNA-DNA hybridization (DDH) values calculated with the closest relative, was estimated for each strain using the genome BLAST distance phylogeny approach (11), as implemented on the Type (Strain) Genome Server (TYGS) using default settings (12).

We found that the draft genomes of strains TNC1(2019), TNC3(2019), and TNW1(2019) were most closely related to *Bacillus subtilis* subsp. *subtilis* (ATCC 6051) (NCBI accession number ASM608879v1), with DDH values of 91.7%, 90.0%, and 86.8%, respectively. The three draft genomes ranged in size from 4,044,726 bp to 4,080,193 bp, with an average G+C content of 43.6% and median coverage of 106.7×. We found that the draft genome of strain TNC2(2019) was most closely related to *Bacillus velezensis* NRRL B-41580 (NCBI accession number ASM146182v1), a type strain isolated from river sediments in Spain (13), and that of strain TNW2(2019) was most closely related to *Bacillus velezensis* FZB42 (NCBI accession number ASM1578v2), a strain formerly known as *Bacillus amyloliquefaciens* subsp. *plantarum* FZB42 and isolated from beet rhizosphere (14). The two strains had an average G+C content of 46.3%, median coverage of 100×, and DDH values of 93.1% and 91.0%, respectively.

Data availability. The draft genomes and raw reads for each strain have been deposited in DDBJ/EMBL/GenBank under the accession numbers listed in Table 1.

REFERENCES

1. Liu Y, Lai Q, Göker M, Meier-Kolthoff JP, Wang M, Sun Y, Wang L, Shao Z. 2015. Genomic insights into the taxonomic status of the *Bacillus cereus* group. Sci Rep 5:14082. https://doi.org/10.1038/srep14082.
2. Onda T, Yanagida F, Tsuji M, Shinohara T, Yokotsuka K. 2003. Time series analysis of aerobic bacterial flora during miso fermentation. Lett Appl Microbiol 37:162–168. https://doi.org/10.1046/j.1472-765x.2003.01371.x.
3. Lin NN, Lee YF, Chi YJ, Wang MF, Chan YC, Chan KC, Chen YJ, Chiu YT. 2017. *Bacillus subtilis*-fermented red bean (red bean natto) reduces hyperlipidemia levels in hamsters fed an atherogenic diet. J Food Biochem 41:e12264. https://doi.org/10.1111/jfbc.12264.
4. Jeżewska-Frąckowiak J, Seroczyńska K, Banaszczyk J, Jedrzejczak G, Żylicz-Stachula A, Skowron PM. 2018. The promises and risks of probiotic *Bacillus* species. Acta Biochim Pol 65:509–519. https://doi.org/10.18388/abp.2018_2652.
5. Shrestha SD, Guttman DS, Perron GG. 2017. Draft genome sequences of 10 environmental *Pseudomonas aeruginosa* strains isolated from soils, sediments, and waters. Genome Announc 5:e00804-17. https://doi.org/10.1128/genomeA.00804-17.
6. 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.
7. 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.

8. Mikheenko A, Valin G, Prjibelski A, Saveliev V, Gurevich A. 2016. Icarus: visualizer for de novo assembly evaluation. Bioinformatics 32:3321–3323. https://doi.org/10.1093/bioinformatics/btw379.
9. Marić J. 2015. Long read RNA-seq mapper. Master thesis. University of Zagreb, Zagreb, Croatia.
10. 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.
11. Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. BMC Bioinformatics 14:60. https://doi.org/10.1186/1471-2105-14-60.
12. Meier-Kolthoff JP, Göker M. 2019. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. Nat Commun 10:2182. https://doi.org/10.1038/s41467-019-10210-3.
13. Dunlap CA, Kim S-J, Kwon S-W, Rooney AP. 2016. *Bacillus velezensis* is not a later heterotypic synonym of *Bacillus amyloliquefaciens*; *Bacillus methylotrophicus*, *Bacillus amyloliquefaciens* subsp. *plantarum* and "*Bacillus oryzicola*" are later heterotypic synonyms of *Bacillus velezensis* based on phylogenomics. Int J Syst Evol Microbiol 66:1212–1217. https://doi.org/10.1099/ijsem.0.000858.
14. Fan B, Wang C, Ding X, Zhu B, Song X, Borriss R. 2019. AmyloWiki: an integrated database for *Bacillus velezensis* FZB42, the model strain for plant growth-promoting bacilli. Database (Oxford) 2019:baz071. https://doi.org/10.1093/database/baz071.