Draft Genome Sequences of Three *Bacillus* Species Isolated from Biofouled Reverse-Osmosis Membranes

Zahid ur Rehman, Muhammad Ali Iftikhar, TorOve Leiknes

Water Desalination and Reuse Center, King Abdullah University of Science and Technology, Thuwal, Kingdom of Saudi Arabia
Department of Ophthalmology, Luton and Dunstable University Hospital NHS Foundation Trust, Luton, United Kingdom

**ABSTRACT** Here, we present the draft genome sequences of three bacteria belonging to the genus *Bacillus* which were isolated from biofouled reverse-osmosis (RO) membranes harvested from a full-scale desalination plant. The sizes of the assembled genomes for RO1, RO2, and RO3 were 4.22 Mb, 4.15 Mb, and 4.23 Mb, respectively.

The bacteria belonging to the genus *Bacillus* have expansive physiology that allows them to colonize diverse environments such as soil, water, air, and lake sediments, as well as extreme environments, including salt marshes, hot springs, and acid water (1). The bacilli in seawater can initiate biofouling of reverse-osmosis (RO) membranes (2), which leads to increased maintenance and water treatment costs. In addition, bacilli have been shown to produce xanthine oxidase, a free radical-generating enzyme that can be implemented in biological control of fouling (3). Therefore, these bacteria provide a valuable resource for studies on the formation of biofouling and its control.

A 7-year-old biofouled RO membrane module was obtained from a seawater desalination plant located on the coast of the Red Sea in Saudi Arabia (22.299815N, 39.116812E). A 1-cm² membrane piece was cut, mixed with 10 ml of 1/100 phosphate-buffered saline (PBS), and vortex mixed. After mixing, 100 µl of PBS was plated onto freshly prepared marine agar plates and incubated at 30°C for 72 h. Three phenotypically distinct (based on color) colonies were selected and streaked onto fresh marine agar plates. This procedure was repeated three times to obtain pure cultures of the isolates.

For DNA extraction, all of the isolates were grown in 20 ml of marine broth for 48 h, with shaking at 120 rpm. The cell culture was centrifuged at 6,000 × g for 10 min, and the resulting cell pellet was used for DNA extraction using the DNeasy PowerWater kit (Qiagen, Germany). The NEBNext Ultra II DNA library preparation kit (New England Biolabs, USA) was used to prepare the sequencing library according to the manufacturer’s instructions. The library preparation and sequencing were performed by DNA-Sense (Aalborg, Denmark). Paired-end sequencing (2 × 301 bp) of the samples was performed on the MiSeq platform (Illumina, USA) using the MiSeq reagent kit v. 3 (600 cycles).

The bioinformatic processing of the sequence reads was performed using default parameters for all software unless otherwise specified. The sequencing reads were trimmed using Cutadapt v. 1.16 (4) and assembled using MEGAHIT v. 1.1.3 (5). The genomic features of all of the strains are given in Table 1. The genomes were annotated using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v. 4.11 and Prokka v. 1.14-dev (6, 7). The numbers of protein-coding genes, rRNAs, and tRNAs detected for RO1, RO2, and RO3 are given in Table 1. The taxonomic assignment of the genome by GTDB-Tk v. 1.0.2 (8) showed that RO1 and RO2 are related to *Bacillus* sp. CHD6a (average nucleotide identity [ANI] values of...
91% and 92.2%, respectively), while RO3 is related to *Bacillus aquimaris* (ANI value of 83.1%). The ANI values were below the species demarcation limit (ANI values of ≥95%) (9, 10), which suggests that these isolates may represent novel species of the genus *Bacillus* and thus may require new species names.

**Data availability.** The genome sequences reported in this article were deposited in DDBJ/ENA/GenBank under BioProject number PRJNA616073 and accession numbers JAAZWB000000000 (RO1), JAAXCU000000000 (RO2), and JAAXCV000000000 (RO3). Raw reads were deposited in the Sequence Read Archive (SRA) under accession numbers SRR11849547 (RO1), SRR11849546 (RO2), and SRR11849545 (RO3). The genomic versions described in this paper are versions JAAZWB010000000, JAAXCU010000000, and JAAXCV010000000.

**ACKNOWLEDGMENTS**

This research was funded by baseline finding funding (grant BAS/1/1061-01-01) awarded to T.L. by King Abdullah University of Science and Technology (KAUST).

We thank the staff at the KAUST desalination plant for help with sample collection.

**REFERENCES**

1. 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.

2. Lee J, Ren X, Yu H-W, Kim S-J, Kim IS. 2010. Membrane biofouling of seawater reverse osmosis initiated by sporogenic *Bacillus* strain. Environ Eng Res 15:141–147. https://doi.org/10.4491/eer.2010.15.3.141.

3. Nagaraj V, Skillman L, Li D, Xie Z, Ho G. 2017. Control of biofouling by xanthine oxidase on seawater reverse osmosis membranes from a desalination plant: enzyme production and screening of bacterial isolates from the full-scale plant. Lett Appl Microbiol 65:73–81. https://doi.org/10.1111/lam.12747.

4. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBOt 1 J 17:10–12. https://doi.org/10.14806/ej.17.1.200.

5. Li DH, Luo RB, Liu CM, Leung CM, Ting HF, Sadakane K, Yamashita H, Lam TW. 2016. MEGAHIT v1.0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

6. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.

7. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt K, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.

8. Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a tool to classify genomes with the Genome Taxonomy Database. Bioinformatics 2019:btz848. https://doi.org/10.1093/bioinformatics/btz848.

9. Rashid MHO, Young JPW, Everall I, Clercx P, Willems A, Braun MS, Wink M. 2015. Average nucleotide identity of genome sequences supports the description of *Rhizobium lentis* sp. nov., *Rhizobium bangladeshense* sp. nov. and *Rhizobium binae* sp. nov. from lentil (*Lens culinaris*) nodules. Int J Syst Evol Microbiol 65:3037–3045. https://doi.org/10.1099/ijsem.0.003373.

10. Jain C, Rodriguez-R LM, Phillips AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun 9:5114. https://doi.org/10.1038/s41467-018-07641-9.

**TABLE 1 Genomic features of RO isolates**

| Genomic feature | Data for strain: |
|----------------|-----------------|
|                | RO1             | RO2             | RO3             |
| Assembly length (bp) | 4,224,738       | 4,154,828       | 4,223,915       |
| No. of contigs     | 92              | 57              | 34              |
| No. of reads       | 1,610,647       | 2,107,415       | 1,860,114       |
| GC content (%)     | 40              | 40              | 43              |
| Contig N50 (bp)    | 92,363          | 118,722         | 226,369         |
| Completeness (%)   | 98.85           | 98.56           | 98.56           |
| Coverage (%)       | 217             | 268             | 238             |
| No. of protein-coding genes | 4,049         | 4,059           | 4,250           |
| No. of tRNAs       | 65              | 38              | 79              |
| No. of complete rRNAs | 14             | 10              | 6               |

**Table 1**

Genomic features of RO isolates

| Genomic feature | Data for strain: |
|----------------|-----------------|
|                | RO1             | RO2             | RO3             |
| Assembly length (bp) | 4,224,738       | 4,154,828       | 4,223,915       |
| No. of contigs     | 92              | 57              | 34              |
| No. of reads       | 1,610,647       | 2,107,415       | 1,860,114       |
| GC content (%)     | 40              | 40              | 43              |
| Contig N50 (bp)    | 92,363          | 118,722         | 226,369         |
| Completeness (%)   | 98.85           | 98.56           | 98.56           |
| Coverage (%)       | 217             | 268             | 238             |
| No. of protein-coding genes | 4,049         | 4,059           | 4,250           |
| No. of tRNAs       | 65              | 38              | 79              |
| No. of complete rRNAs | 14             | 10              | 6               |