Draft Genome Sequence of *Bacillus cereus* Strain UAEU-H3K6M1, a Bacterium with Potential Bioremediation Abilities, Isolated from Petroleum Sludge

Manal A. Al Hefeiti,ª Joseph Mafofo,ᵇ Gess Thoms Xavier,ᵇ Sathishkumar Ramaswamy,ᵇ Bincy Baby,ᶜ Divinlal Harilal,ᵇ Ranjit Vijayan,ᶜ 🅿 S. Salman Ashrafª

ªDepartment of Chemistry, College of Science, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

ᵇAgiomix FZ-LLC, Dubai Science Park, Dubai, United Arab Emirates

ᶜDepartment of Biology, College of Science, United Arab Emirates University, Al Ain, Abu Dhabi, United Arab Emirates

ABSTRACT   Here, we report the draft genome sequence of *Bacillus cereus* strain UAEU-H3K6M1, which was isolated from petroleum sludge in the desert. It is composed of around 5.4 Mbp and has a GC content of 35%. Functional annotation and pathway modeling showed multiple genes with potential bioremediation abilities.

*B*acillus cereus strain UAEU-H3K6M1 was isolated from a petroleum sludge sample and shows interesting bioremediation potential. *In vitro* studies show potential uses with various classes of organic pollutants, including aromatic dyes and contaminants of emerging concern. Here, we report the annotated draft genome sequence (~5.4 Mbp) of this strain.

Due to ever-increasing pollution of our environment, there is a pressing need to find novel and efficient approaches for remediation of our natural resources, especially water bodies. A number of microbial species that can degrade a range of pollutants have been previously reported (1). We were interested in screening the microbial population in a sample of petroleum sludge from the United Arab Emirates (UAE) to isolate bacteria capable of degrading various classes of aromatic pollutants. A total of 12 different bacterial isolates that showed very interesting bioremediation capacities were purified through conventional culture techniques. The crucial criteria for selection included the efficiency of the isolate in degrading azo dyes under both aerobic and anaerobic conditions. The most promising of these isolates were further screened for their capacity to degrade various classes of contaminants of emerging concern, such as sulfamethoxazole, prometryn, and fluometuron, under aerobic conditions.

Strain UAEU-H3K6M1 was then selected for genome analysis based on the criteria described above. An overnight culture (37°C with gentle shaking) of this strain in nutrient broth was used to isolate genomic DNA using an ABIOPure DNA isolation kit (Alliance Bio, USA). Genomic DNA (1 ng) was used to generate Illumina sequencing libraries using the Nextera XT DNA library prep kit (Illumina) and subsequently sequenced using an Illumina MiSeq instrument. A total of 35,124,272 paired-end reads were generated, which represented a coverage of 242× (total bases/reference genome size = 1,314,502,756/5,427,083 = 242.11). These data were preprocessed for quality using Trimmomatic (version 0.36) (2) and FASTX-Toolkit (version 0.0.13) (3) to remove Illumina adapter sequences and filter out low-quality bases, respectively. Preprocessing reduced the total number of sequence reads to 18,225,662 paired-end reads.

These sequence data were then assembled *de novo* using ABySS (4), IDBA-UD (5), SPAdes (6), and Velvet (7), and an integrated assembly was produced with CISA (8) by

using the Tychus pipeline (https://github.com/Abdo-Lab/Tychus). The integrated assembly produced a total of 16 contigs with an N_{50} length of 573,109 bp and an L_{50} of 4. The largest contig was 1,120,625 bp long. The total assembly sequence was shown to be 5,408,131 bp long with a GC content of 35%. Gene prediction and annotation performed using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (https://www.ncbi.nlm.nih.gov/genome/annotation_prok/) identified 5,493 protein-coding genes, 246 pseudogenes, and 40 RNA genes. The most common subsystems identified by rapid genome annotations using RAST (9) were those related to amino acids and derivatives (n = 495 coding sequences [CDSs]), carbohydrates (n = 421), cofactors, vitamins, prosthetic groups, and pigments (n = 223), cell wall and capsule (n = 169), RNA metabolism (n = 149), membrane transport (n = 147), fatty acids, lipids, and isoprenoids (n = 140), protein metabolism (n = 109), nucleosides and nucleotides (n = 107), and defense (n = 105). We identified several protein-coding genes, including copper oxidase and azoreductase 2 and 4, which have potential bioremediation abilities (10).

Data availability. The *Bacillus cereus* strain UAEU-H3K6M1 genome sequence was deposited at DDBJ/ENA/GenBank under the accession number PHQW00000000. The version described in this paper is version PHQW02000000.

REFERENCES

1. Ye S, Zeng G, Wu H, Zhang C, Dai J, Liang J, Yu J, Ren X, Yi H, Cheng M, Zhang C. 2017. Biological technologies for the remediation of co-contaminated soil. Crit Rev Biotechnol 20:1–5. https://doi.org/10.1080/07388551.2017.1304357.
2. 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.
3. Hannon Lab. 2014. FASTX-Toolkit. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
4. Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJM, Birol İ. 2009. ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123. https://doi.org/10.1101/gr.089532.108.
5. Peng Y, Leung HCM, Yiu SM, Chin FYL. 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.
6. 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.
7. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
8. Lin S-H, Liao Y-C. 2013. CISA: contig integrator for sequence assembly of bacterial genomes. PLoS One 8:e60843. https://doi.org/10.1371/journal.pone.0060843.
9. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrello B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. BMC Genomics 9:75. https://doi.org/10.1186/1471-2164-9-75.
10. Mahmood S, Khalid A, Arshad M, Mahmood T, Crowley DE. 2016. Detoxification of azo dyes by bacterial oxidoreductase enzymes. Crit Rev Biotechnol 36:639–651. https://doi.org/10.3109/07388551.2015.1004518.