

Complete genome sequence of *Bacillus subtilis* strain MARUCo01 isolated from marine sediments of the Indian Ocean in Bagamoyo, Tanzania

Reuben Maghembe,^{1,2,3} Edward Moto,⁴ Abdalah Makaranga,¹ France Mdoe,¹ James Mpemba,⁵ Maximilian A. K. Magulye,⁶ Deogratus Mark,⁷ Andrew Mtewa⁸

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Bacillus subtilis* has emerged as a species with potential for versatile nonribosomal peptides and polyketides of therapeutic importance, including antibiotics. From our molecular bioprospecting project, we report a full genome of *Bacillus subtilis* strain MARUCo01 locally isolated from sediments of the Indian Ocean along the coast of Bagamoyo in Tanzania.

KEYWORDS genomics, *Bacillus subtilis*, strain MARUCo01, genome analysis, average nucleotide identity (ANI)

Bacillus subtilis is species of widely distributed bacteria associated with multiple benefits to human, including the production of antimicrobials and the potential for probiotics (1). These, among other benefits, are a trigger for increased search for and manipulation of novel *Bacillus subtilis* strains in the expression of various secondary metabolites as drug scaffolds. In our ongoing bioprospecting project, as described in our recently published work (2), surface sediment samples were collected from a mangrove forest area in the Indian Ocean shore along the Bagamoyo Coast in Tanzania (6°25'32.5"S 38°54'08.1"E, <https://www.google.com/maps/search/?api=1&query=-6.4256828,38.9022509>). The sediment samples were collected in sterile plastic bottles and kept there at 4°C. A piece of the sediment (about 5 g) was suspended in 200 mL of 0.80% NaCl, followed by serial dilutions (10⁻¹ to 10⁻⁶) with phosphate-buffered saline (PBS) (pH 7.2), and the subsamples of these dilutions were plated on nutrient agar (NA). The individual colonies were then picked and restreaked on NA and incubated at 28°C for 48 h before DNA extraction.

Cells were picked directly from the plated colonies, resuspended in PBS, and the genomic DNA was extracted using a ZymoBIOMICS DNA cvMiniprep Kit (ZR D4300) based on the manufacturer's guide. Then the TruSeq DNA PCR-Free Kit and TruSeq Nano DNA Kit were used to build libraries, and the Illumina Novaseq 6000 platform (2.5 G bp) was used to sequence the entire genome, producing small paired-end reads with an average length of 151 bp. Raw reads were quality controlled using FastQC and Trimmomatic (v 0.38) under the following criteria: minimal sequence length for both reads prior to a sequence pair removal is 20 bp, and the minimal adapter overlap (stringency) is 1 bp. With the use of the Unicycler pipeline (v. 0.4.8) (3), the reads were *de novo* assembled into contigs, which were then checked for species contamination with the ContEst16S pipeline (4). A chromosome was then generated using CONTIGuator (v.2.7.40) (5) and then annotated using the best-placed reference protein approach (set; GeneMarkS-2+) of the Prokaryotic Genome Annotation Pipeline (PGAP v.6.3; https://www.ncbi.nlm.nih.gov/genome/annotation_prok/). The taxonomic

Editor Julia A. Maresca, University of Delaware College of Engineering, Newark, Delaware, USA

Address correspondence to Reuben Maghembe, rmaghembe@gmail.com.

The authors declare no conflict of interest.

Received 2 June 2023

Accepted 19 July 2023

Published 22 August 2023

Copyright © 2023 Maghembe 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/).

placement was predicted using the Microbial Genomes Atlas (MiGA) (6) and then confirmed with average nucleotide identity (ANI) calculation with FastANI (7).

From whole-genome Illumina sequencing, 19,041,058 reads were generated, which, upon quality control, were reduced to 12,490,734. *De novo* assembly resulted in 29 contigs (a total of 4,058,030 bp) with N50 of 995,325 and GC content of 43.7%. PGAP annotation revealed 4,206 genes, with 4,037 coding sequences. Combining MiGA taxonomic approach with FastANI estimation, the MARUCo01 strain was unambiguously placed into the species *Bacillus subtilis*, and the closest relative identified was *Bacillus subtilis* subsp. *subtilis* str. 168 (NCBI accession number [AL009126.3](https://.ncbi.nlm.nih.gov/nucl/AL009126.3)) with an ANI value of 98.74%.

ACKNOWLEDGMENTS

This project was privately funded. However, the authors acknowledge the contribution from individuals such as Clement Mlay from Pwani University (Kenya), Michael Silayo from Marian University College, and Laurent Wambura from the Government Chemist Laboratory Authority (GCLA, Tanzania) for their support sample handling from the field to the laboratory.

AUTHOR AFFILIATIONS

¹Biological and Marine Sciences Unit, Faculty of Natural and Applied Sciences, Marian University College, Bagamoyo, Tanzania

²Department of Biological Sciences, Faculty of Science, University of Botswana, Gaborone, Botswana

³Department of Microbiology and Immunology, Faculty of Biomedical Sciences, Kampala International University Western Campus, Kampala, Uganda

⁴Department of Biology, College of Natural and Mathematics Sciences, University of Dodoma, Dodoma, Tanzania

⁵Department of Biochemistry and Physiology, St. Francis University College of Health and Allied Sciences, Ifakara, Tanzania

⁶Department of Immunology and Molecular Biology, School of Biomedical Sciences, Makerere University, Kampala, Uganda

⁷Department of Disease Control, Tanzania Agricultural Research Institute (TARI), Dar es Salaam, Tanzania

⁸Chemistry Section, Department of Applied Sciences, Malawi Institute of Technology, Malawi University of Science and Technology, Limbe, Malawi

AUTHOR ORCID*s*

Reuben Maghembe  <http://orcid.org/0000-0003-2453-5993>

AUTHOR CONTRIBUTIONS

Reuben Maghembe, Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Software, Writing – original draft | Edward Moto, Funding acquisition, Project administration, Resources, Validation, Writing – review and editing | Abdalah Makaranga, Conceptualization, Data curation, Formal analysis, Visualization, Writing – review and editing | France Mdoe, Formal analysis, Investigation, Methodology, Resources | James Mpemba, Formal analysis, Investigation, Methodology, Resources | Maximilian A. K. Magulye, Data curation, Formal analysis, Methodology, Software, Visualization, Writing – review and editing | Deogratius Mark, Formal analysis, Investigation, Methodology, Resources | Andrew Mtewa, Funding acquisition, Investigation, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome sequence was submitted to GenBank and can be accessed via the accession numbers [CP107539.1](https://ncbi.nlm.nih.gov/nucl/CP107539.1) (chromosome), [CP107540.1](https://ncbi.nlm.nih.gov/nucl/CP107540.1) (plasmid), [OP586436.1](https://ncbi.nlm.nih.gov/nucl/OP586436.1) (16S rRNA),

and BioSample and BioProject accession numbers [SAMN31159204](#) and [PRJNA887360](#), respectively. The assembly and raw read sequence data sets can be accessed through [GCA_025837035](#) and [SRS15977960](#), respectively.

REFERENCES

1. Arai EN, Yoneda S, Yoneda N, Ito M, Tsuda S, Shiozaki A, Nohira T, Hyodo H, Kumazawa K, Suzuki T, Nagasaki S, Makino S, Saito S. 2022. Probiotics including *Clostridium butyricum*, *Enterococcus faecium*, and *Bacillus subtilis* may prevent recurrent spontaneous preterm delivery. *J Obstet Gynaecol Res* 48:688–693. <https://doi.org/10.1111/jog.15166>
2. Maghembe RS, Mdoe FP, Makaranga A, Mpemba JA, Mark D, Mlay C, Moto EA, Mtewa AG. 2023. Complete genome sequence data of *Priestia megaterium* strain MARUCo02 isolated from marine mangrove-inhabited sediments of the Indian Ocean in the Bagamoyo coast. *Data Brief* 48:109119. <https://doi.org/10.1016/j.dib.2023.109119>
3. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
4. Yoon S-H, Ha S-M, Kwon S, Lim J, Kim Y, Seo H, Chun J. 2017. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 67:1613–1617. <https://doi.org/10.1099/ijsem.0.001755>
5. Galardini M, Mengoni A, Bazzicalupo M. 2015. Mapping contigs using CONTIGuator, p. 163–176. In Mengoni, A, Galardini, M, Fondi, M (eds.), *Bacterial pangenomics: methods and protocols*. Springer New York, New York, NY.
6. 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>
7. Jain C, Rodríguez-R LM, Phillippy 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>