AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Complete Genome Sequence of *Alkalihalobacillus* sp. Strain LMS39, a Haloalkaliphilic Bacterium Isolated from a Hypersaline Lake

**Alex Kipnyargis,[a,b] Romano Mwirichia,[a] Birgit Pfeiffer,[b] Rolf Daniel[b]**

[a]Department of Biological Sciences, University of Embu, Embu, Kenya

[b]Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Göttingen, Germany

**ABSTRACT** Here, we report the complete genome sequence of a haloalkaliphilic bacterium (*Alkalihalobacillus* sp. strain LMS39) isolated from Lake Magadi, a hypersaline lake in Kenya. The genome comprised 4,850,562 bp with a GC content of 37%.

Haloalkaliphilic microorganisms are ideal models for studying adaptation to extreme ecosystems (1) and encode biocatalysts that are utilized in industrial applications (2). Strain LMS39 was isolated from dry sediments of Lake Magadi, Kenya (1°43′–2°00′S, 36°13′–36°18′E). Sediment (0.1 g) was serially diluted, and $10^{-9}$ and $10^{-10}$ dilutions were spread onto basal agar medium containing peptone (2 g/L), yeast extract (0.5 g/L), $K_2HPO_4$ (1 g/L), $CaCl_2{\cdot}2H_2O$ (0.05 g/L), $MgSO_4{\cdot}7H_2O$ (0.1 g/L), and agar (14 g/L) (3). The medium was prepared with sterile lake water and supplemented with 1% (wt/vol) cellulose. Singularization was conducted by restreaking the strains at least three times. Pure cultures were grown on Trypticase soy broth (TSB) supplemented with $NaCO_3$ (1% [wt/vol]) and NaCl (4% [wt/vol]) in a rotary shaker (180 rpm) at 37°C for 12 h.

Chromosomal DNA was extracted using the MasterPure complete DNA and RNA purification kit as recommended by the manufacturer (Epicentre, Madison, WI, USA). Sanger sequencing and analysis of the LMS39 16S rRNA gene sequence revealed 97% identity to that of *Alkalihalobacillus bogoriensis* sp. strain LBB3 (NR042894). An Illumina paired-end sequencing library was generated using the Nextera XT DNA sample preparation kit. Sequencing was conducted using the MiSeq system and reagent kit v3 (600 cycles) according to the manufacturer's protocol (Illumina, San Diego, CA, USA). Libraries for Nanopore sequencing were generated using 1.5 $\mu$g of isolated DNA without size selection. The 1D genomic DNA sequencing protocol for the MinION device was conducted using the ligation sequencing 1D kit (SQK-LSK109) as recommended by the manufacturer (Oxford Nanopore Technologies, Oxford, UK). After end repair using the NEBNext FFPE repair mix (New England Biolabs, Ipswich, MA, USA), sequencing was performed using a SpotON flow cell Mk I (R9.4) and MinKNOW software v18.12.6 (Oxford Nanopore Technologies). Default parameters were used for all software unless otherwise specified. Trimming and adapter removal was performed using fastp v0.20.1 (4) and Porechop v0.2.4 (https://github.com/rrwick/Porechop) for the Illumina and Nanopore reads, respectively. Potential phiX contamination was removed from the Illumina reads using Bowtie2 v2.3.5.1 (5). After quality assessment, 1,884,551 Nanopore reads ($N_{50}$, 5,594 bp) and 2,524,128 Illumina paired-end reads were obtained.

*De novo* genome assembly was conducted using Unicycler v0.5.0 (6) and validated with Bandage v0.8.1 (7), resulting in 1 circular genome (4,850,562 bp) with a GC content of 37.0%. Annotation using PGAP v6.1 (8) yielded 4,749 protein-encoding genes, of which 4,571 had functional assignments. Additionally, genes encoding 89 tRNAs, 30 rRNAs, and 4 noncoding RNAs (ncRNAs) were identified. To assess the relationship between the LMS39 genome and publicly available ones, the type strain genomes available from the Type Strain Genome

**Editor** Julia A. Maresca, University of Delaware



Address correspondence to Rolf Daniel, rdaniel@gwdg.de.

The authors declare no conflict of interest.

**Received** 5 April 2022
**Accepted** 3 June 2022
**Published** 23 June 2022

Server (TYGS) (9) were employed. The average nucleotide identity (ANI) was assessed using the ANIm method in pyani v0.2.11 (10). LMS39 is most closely related to *Alkalihalobacillus bogoriensis* ATCC BAA-922 (GenBank accession number GCA_000621445) with an average nucleotide identity of 92.8%. Accordingly, the strain was designated *Alkalihalobacillus* sp. strain LMS39. Analysis of the carbohydrate-active enzymes (CAZymes) (11) showed that the genome of LMS39 possessed genes encoding cellulase, $\beta$-glucosidases, $\beta$-xylanase, $\beta$-xylosidase $\alpha$-amylase, neopullulanases, and pectate lyases.

**Data availability.** The annotated genome sequence of *Alkalihalobacillus* sp. strain LMS39 has been submitted to GenBank under the accession number CP093300.1. The raw reads were deposited at the NCBI Sequence Read Archive (SRA) under the accession number SRR18516662 for the Illumina reads and SRR18516661 for the Nanopore reads.

## ACKNOWLEDGMENTS

We acknowledge the German Academic Exchange Service (DAAD) for supporting A. Kipnyargis. We appreciate the support of Open Access Publication Funds of the University of Göttingen. The funders had no role in the study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

1. Pontefract A, Zhu TF, Walker VK, Hepburn H, Lui C, Zuber MT, Ruvkun G, Carr CE. 2017. Microbial diversity in a hypersaline sulfate lake: a terrestrial analog of ancient Mars. Front Microbiol 8:1819. https://doi.org/10.3389/fmicb.2017.01819.
2. Sorokin DY, Gumerov VM, Rakitin AL, Beletsky AV, Damsté JSS, Muyzer G, Mardanov AV, Ravin NV. 2014. Genome analysis of *Chitinivibrio alkaliphilus* gen. nov., sp. nov., a novel extremely haloalkaliphilic anaerobic chitinolytic bacterium from the candidate phylum Termite Group 3. Environ Microbiol 16:1549–1565. https://doi.org/10.1111/1462-2920.12284.
3. Kiplimo D, Mugweru J, Kituyi S, Kipnyargis A, Mwirichia R. 2019. Diversity of esterase and lipase producing haloalkaliphilic bacteria from Lake Magadi in Kenya. J Basic Microbiol 59:1173–1184. https://doi.org/10.1002/jobm.201900353.
4. Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
5. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
6. 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.
7. Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of de novo genome assemblies. Bioinformatics 31:3350–3352. https://doi.org/10.1093/bioinformatics/btv383.
8. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. Nucleic Acids Res 44:6614–6624. https://doi.org/10.1093/nar/gkw569.
9. 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.
10. Arahal DR. 2014. Whole-genome analyses: average nucleotide identity, p 103–122. *In* Goodfellow M, Sutcliffe I, Chun J (ed), Methods in microbiology: new approaches to prokaryotic systematics, vol 41. Academic Press, Oxford, UK.
11. Cantarel BI, Coutinho PM, Rancurel C, Bernard T, Lombard V, Henrissat B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for glycogenomics. Nucleic Acids Res 37:D233–D238. https://doi.org/10.1093/nar/gkn663.