AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Draft Genome Sequence of a Novel Marine Anaerobic Ammonium-Oxidizing Bacterium, "*Candidatus* Scalindua sp."

Muhammad Ali,[a] Dario Rangel Shaw,[a] Pascal E. Saikaly[a]

[a]Biological and Environmental Science and Engineering Division, Water Desalination and Reuse Research Center, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

ABSTRACT A novel anaerobic ammonium-oxidizing (anammox) bacterium was detected in an upflow column reactor treating synthetic nitrogen-rich saline solution. Here, we assembled a 4.59-Mb draft genome sequence of this bacterium, identified as a member of the genus "*Candidatus* Scalindua," that has 84% nucleotide-level genomic similarity with the closest related anammox bacterium ("*Candidatus* Scalindua rubra").

Anaerobic ammonium-oxidizing (anammox) bacteria play a vital role in the global nitrogen cycle and in energy-efficient treatment of N-rich wastewaters (1). There are more than 100 full-scale anammox installations worldwide for the treatment of N-rich wastewaters (2), and there is now growing interest in applying anammox bacteria for the treatment of N-rich saline wastewaters (3). Thus, we started a 1-liter upflow column reactor (XK 50/60 column; GE Healthcare, UK) for the treatment of synthetic N-rich saline solution prepared using fresh Red Sea water (~3.5% salinity) containing NH_4^+ and NO_2^- (~5 mM each). The reactor was operated at ambient temperature (35°C) with a constant feeding rate of 2.2 liters · day^{-1}. The column reactor was inoculated with anammox biofilm attached to a nonwoven fabric sheet harvested from another reactor (4).

The biomass was extracted from the upflow column reactor under steady-state conditions, and genomic DNA was extracted using a DNA extraction kit (FastDNA Spin kit for soil; MP Biomedicals) according to the manufacturer's instructions. The DNA was quantified using a Qubit fluorometer (Thermo Fisher Scientific, USA), and 50 ng was used to prepare Nextera libraries following the manufacturer's instructions (Illumina, USA). The prepared DNA libraries were paired-end sequenced (2 × 250 bp) on a HiSeq 2500 instrument (Illumina) and generated approximately 44 million reads in total. The sequence reads were trimmed for Nextera adaptors using cutadapt v. 1.10 (5) with a minimum Phred score of 20 and a minimum length of 150 bp. The trimmed reads were assembled using SPAdes v. 3.7.1 (6). The reads were mapped back to the assembly using Burrows-Wheeler Aligner (BWA) v. 0.7.15-r1142-dirty (7) to generate coverage files for metagenomic binning. Open reading frames (ORFs) were predicted in the assembled scaffolds using Prodigal (8). A set of 117 hidden Markov models (HMMs) of essential single-copy genes were searched against the ORFs using HMMER3 (http://hmmer.janelia.org/) with default settings, with the exception that the option "-cut_tc" was used (9). Identified proteins were taxonomically classified using a BLASTP search against the RefSeq v.52 protein database with a maximum E value cutoff of 10^{-5}. MEtaGenome ANalyzer (MEGAN) was used to extract class-level taxonomic assignments from the BLASTP output (10). The script network.pl (http://madsalbertsen.github.io/mmgenome/) was used to obtain paired-end read connections between scaffolds. The 16S rRNA genes were identified using BLAST v. 2.2.28+ (11) and were classified using SINA v. 1.2.11 (12) with the minimum identity adjusted to 0.80. The required data set

Citation Ali M, Shaw DR, Saikaly PE. 2019. Draft genome sequence of a novel marine anaerobic ammonium-oxidizing bacterium, "*Candidatus* Scalindua sp." Microbiol Resour Announc 8:e00297-19. https://doi.org/10.1128/MRA.00297-19.

Editor Irene L. G. Newton, Indiana University, Bloomington

Address correspondence to Pascal E. Saikaly, pascal.saikaly@kaust.edu.sa.

Received 15 March 2019
Accepted 22 April 2019
Published 16 May 2019

for binning was generated according to the description in the mmgenome package v. 0.6.3 (13). The genome was extracted by using the mmgenome package in R v. 3.3.1 (14), and the extracted genome was annotated using PROKKA v. 1.12-beta (15).

A 4.59-Mb genome sequence comprising 121 contigs (GC content of 41% and N_{50} value of 92,628 bp) was obtained, and 2,565 gene-coding regions, 41 tRNAs, and a single rRNA (*rrn*) operon were annotated. The genome is 90% complete based on CheckM v1.0.5 (16). The calculated average nucleotide identity with the closest related anammox bacterium ("*Ca*. Scalindua rubra" [GenBank accession number MAYW00000000]) was 83.72%, and the genome therefore is considered novel.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number RBMW00000000. The Sequence Read Archive (SRA) accession number is SRR7904086. The BioProject accession number is PRJNA492998.

ACKNOWLEDGMENT

This work was funded by the Center Competitive Funding Program (grant FCC/1/1971-33-01) to P.E.S. from the King Abdullah University of Science and Technology (KAUST).

REFERENCES

1. Ali M, Okabe S. 2015. Anammox-based technologies for nitrogen removal: advances in process start-up and remaining issues. Chemosphere 141:144–153. https://doi.org/10.1016/j.chemosphere.2015.06.094.
2. Lackner S, Gilbert EM, Vlaeminck SE, Joss A, Horn H, Van Loosdrecht MCM. 2014. Full-scale partial nitritation/anammox experiences—an application survey. Water Res 55:292–303. https://doi.org/10.1016/j.watres.2014.02.032.
3. Li J, Qi P, Qiang Z, Dong H, Gao D, Wang D. 2018. Is anammox a promising treatment process for nitrogen removal from nitrogen-rich saline wastewater? Bioresour Technol 270:722–731. https://doi.org/10.1016/j.biortech.2018.08.115.
4. Kindaichi T, Awata T, Suzuki Y, Tanabe K, Hatamoto M, Ozaki N, Ohashi A. 2011. Enrichment using an up-flow column reactor and community structure of marine anammox bacteria from coastal sediment. Microbes Environ 26:67–73. https://doi.org/10.1264/jsme2.ME10158.
5. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
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. Li H, Durbin R. 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 26:589–595. https://doi.org/10.1093/bioinformatics/btp698.
8. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.
9. Dupont CL, Rusch DB, Yooseph S, Lombardo M-J, Alexander Richter R, Valas R, Novotny M, Yee-Greenbaum J, Selengut JD, Haft DH, Halpern AL, Lasken RS, Nealson K, Friedman R, Craig Venter J. 2012. Genomic insights to SAR86, an abundant and uncultivated marine bacterial lineage. ISME J 6:1186–1199. https://doi.org/10.1038/ismej.2011.189.
10. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. https://doi.org/10.1093/molbev/msr121.
11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
12. Pruesse E, Peplies J, Glöckner FO. 2012. SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28:1823–1829. https://doi.org/10.1093/bioinformatics/bts252.
13. Karst S, Kirkegaard RH, Albertsen M. 2016. mmgenome: a toolbox for reproducible genome extraction from metagenomes. bioRxiv. https://doi.org/10.1101/059121.
14. R Core Team. 2013. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
15. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.
16. 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.