AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Three Draft Single-Cell Genome Sequences of Novel SAR324 Strains Isolated from the Abyssopelagic Southern Ocean

Diego J. Castillo,[a] Marc W. Van Goethem,[a] Thulani P. Makhalanyane[a]

[a]Centre for Microbial Ecology and Genomics (CMEG), Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

ABSTRACT SAR324 is a ubiquitous and phylogenetically distinct clade of *Deltaproteobacteria* in marine environments. Here, we present three single-cell amplified genome sequences from the SAR324 lineage, obtained from the abyssopelagic zone of the Indian sector of the Southern Ocean.

Members of SAR324 mediate important biogeochemical processes in the oceans (1–6). However, the lack of sequence representatives of this clade limits efforts to gain a mechanistic understanding of their precise functional roles (2, 3). Members of SAR324 have been found throughout the water column (4, 5) and are metabolically versatile (2, 6). Current physiological insights regarding this clade are derived primarily from environmental samples, based on 16S rRNA gene surveys (5), metagenomics and metatranscriptomics (2, 3, 7–9), and single-cell genomics (6, 10). Here, we present three SAR324 genome sequences, obtained from the abyssopelagic zone of the Southern Ocean using single-cell genomics.

A water sample was collected at a depth of 4,154 m in the Indian sector of the Southern Ocean (47.994°S, 37.034°E) and preserved as detailed previously (11). Fluorescence-activated cell sorting and multiple displacement amplification were performed at Bigelow Laboratory for Ocean Sciences (ME, USA), as previously described (11). Single-cell amplified genomes (SAGs; $n = 41$) were selected for library preparation using the Nextera XT DNA kit, according to the manufacturer's instructions. The libraries were sequenced at Admera Health, LLC (NJ, USA) using an Illumina HiSeq X sequencer (150-bp paired-end reads). Bioinformatics analysis was conducted using KBase (12). The raw reads were processed using Trimmomatic v0.36 (13) and assembled using SPAdes v3.13.0 (14), with "single-cell" entered as the DNA source. The assemblies were evaluated using QUAST (15), while the SAG completeness and contamination were estimated using CheckM v1.018 (16). The assembly quality was determined using minimum information about a single amplified genome (MISAG) standards (17). The genomic coverage was calculated using BBTools (18). The genome statistics are provided in Table 1. Genome Taxonomy Database Toolkit (GTDB-Tk) v1.1.0 release 89 (19) was used to assign taxonomy, and protein-encoding regions were identified using Prokka v1.14.5 (20). Average nucleotide identities (ANI) of reciprocal hits were calculated between our three SAR324 SAGs and against two SAR324 draft genome sequences: SAR324 bacterium lautmerah1 (3) and SAR324 Arctic96AD-7 (genome 046) (9) (http://enve-omics.ce.gatech.edu/ani/). Finally, we compared the 16S rRNA gene of each SAG to all 16S rRNA gene sequences available in the NCBI nonredundant (nr) database. Default parameters were used for all software unless otherwise noted.

Based on current standards (17), our three SAR324 genome sequences were classified as medium-quality draft genome sequences and were taxonomically assigned as deltaproteobacterial group SAR324 (strain Arctic96AD-7). SAR324_K2 and SAR324_N8 had the highest genome similarity, with 96.75% ANI. Both genomes were distinct from SAR324_I22 (ca. 81.9% ANI), suggesting that they may have different identities at the species or genus level, within the same family. ANI percentages between our three

Citation Castillo DJ, Van Goethem MW, Makhalanyane TP. 2021. Three draft single-cell genome sequences of novel SAR324 strains isolated from the abyssopelagic Southern Ocean. Microbiol Resour Announc 10:e00759-21. https://doi.org/10.1128/MRA.00759-21.

Editor Frank J. Stewart, Montana State University

Copyright © 2021 Castillo et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Thulani P. Makhalanyane, thulani.makhalanyane@up.ac.za.

Received 29 July 2021
Accepted 8 September 2021
Published 30 September 2021

TABLE 1 Genome statistics and comparisons of three potentially novel SAR324 SAGs

Characteristic	Data for strain:		
	SAR324_I22	SAR324_K2	SAR324_N8
No. of raw paired-end reads	11,549,492	8,050,736	10,928,754
No. of quality-filtered paired-end reads	10,024,104	6,997,348	9,413,936
Assembly size (bp)	2,453,850	2,467,220	2,070,041
Coverage (×)	601	422	549
G+C content (%)	44.72	41.42	42.43
Estimated genome completeness (%)	56.15	50.71	50.14
Predicted genome size (bp)	4,370,169	4,865,352	4,128,522
Estimated contamination (%)	3.2	3.36	3.03
Genome quality	Medium	Medium	Medium
No. of contigs	707	505	646
Largest contig (bp)	87,181	87,899	52,735
N_{50} (bp)	13,771	23,927	11,591
No. of protein-coding genes	2,346	2,337	1,952
No. of tRNA genes	31	32	19
No. of rRNA genes	8	3	3
ENA raw read accession no.	ERR6548281	ERR6548282	ERR6548383
ENA assembly accession no.	CAJYYN010000000.1	CAJYYM010000000.1	CAJYYL010000000.1
ANI with bacterium lautmerah1 (%)	74.02	80.96	73.11
ANI with Arctic96AD-7 (%)	79.37	97.01	97.54
16S rRNA gene highest similarity isolate name	*Pseudomonadaceae* SI-3	Uncultured delta proteobacterium	Uncultured bacterium clone
GenBank accession no.	CP026511.1	GU474888.1	HQ674466.1
Similarity (%)	100	99.87	99.79

SAR324 SAGs and two SAR324 reference genomes (3, 9) and comparisons of the 16S rRNA gene sequences available in the NCBI (Table 1) revealed substantial differences within this versatile group. Metabolic pathway reconstructions showed evidence for carbon fixation (phosphoribulokinase; *prk*) in SAR324_N8. Indicator genes for sulfur oxidation (*soxAB*) were found in both SAR324_I22 and SAR324_K2 (3 copies of *soxB*) but not in SAR324_N8.

Data availability. The genome assemblies for the three SAGs have been deposited at the ENA under accession number PRJEB47084, and the accession numbers are given in Table 1.

ACKNOWLEDGMENTS

We thank the scientists aboard the May 2018 trip of the *SA Agulhas II*, especially Mancha Mabaso, Sandra Boitumelo Phoma, and Jarishma Gokul, for sample acquisition.

We gratefully acknowledge the National Research Foundation of South Africa (NRF) (UID 110717). We thank the Centre for High Performance Computing (Cape Town, South Africa) and the Centre for Bioinformatics and Computational Biology, University of Pretoria, for providing computational resources.

REFERENCES

1. Waite DW, Chuvochina M, Pelikan C, Parks DH, Yilmaz P, Wagner M, Loy A, Naganuma T, Nakai R, Whitman WB, Hahn MW, Kuever J, Hugenholtz P. 2020. Proposal to reclassify the proteobacterial classes Deltaproteobacteria and Oligoflexia, and the phylum Thermodesulfobacteria into four phyla reflecting major functional capabilities. Int J Syst Evol Microbiol 70:5972–6016. https://doi.org/10.1099/ijsem.0.004213.
2. Cao H, Dong C, Bougouffa S, Li J, Zhang W, Shao Z, Bajic VB, Qian P-Y. 2016. Delta-proteobacterial SAR324 group in hydrothermal plumes on the South Mid-Atlantic Ridge. Sci Rep 6:22842. https://doi.org/10.1038/srep22842.
3. Haroon MF, Thompson LR, Stingl U. 2016. Draft genome sequence of uncultured SAR324 bacterium lautmerah10, binned from a Red Sea metagenome. Genome Announc 4:e01711-15. https://doi.org/10.1128/genomeA.01711-15.
4. Wright TD, Vergin KL, Boyd PW, Giovannoni SJ. 1997. A novel delta-subdivision proteobacterial lineage from the lower ocean surface layer. Appl Environ Microbiol 63:1441–1448. https://doi.org/10.1128/aem.63.4.1441-1448.1997.
5. Pham VD, Konstantinidis KT, Palden T, DeLong EF. 2008. Phylogenetic analyses of ribosomal DNA-containing bacterioplankton genome fragments from a 4000 m vertical profile in the North Pacific Subtropical Gyre. Environ Microbiol 10:2313–2330. https://doi.org/10.1111/j.1462-2920.2008.01657.x.
6. Swan BK, Martinez-Garcia M, Preston CM, Sczyrba A, Woyke T, Lamy D, Reinthaler T, Poulton NJ, Masland EDP, Gomez ML, Sieracki ME, DeLong EF, Herndl GJ, Stepanauskas R. 2011. Potential for chemolithoautotrophy among ubiquitous bacteria lineages in the dark ocean. Science 333:1296–1300. https://doi.org/10.1126/science.1203690.
7. Sheik CS, Jain S, Dick GJ. 2014. Metabolic flexibility of enigmatic SAR324 revealed through metagenomics and metatranscriptomics. Environ Microbiol 16:304–317. https://doi.org/10.1111/1462-2920.12165.

8. Li M, Jain S, Baker BJ, Taylor C, Dick GJ. 2014. Novel hydrocarbon monooxygenase genes in the metatranscriptome of a natural deep-sea hydrocarbon plume. Environ Microbiol 16:60–71. https://doi.org/10.1111/1462-2920.12182.
9. Cao S, Zhang W, Ding W, Wang M, Fan S, Yang B, Mcminn A, Wang M, Xie B-B, Qin Q-L, Chen X-L, He J, Zhang Y-Z. 2020. Structure and function of the Arctic and Antarctic marine microbiota as revealed by metagenomics. Microbiome 8:47. https://doi.org/10.1186/s40168-020-00826-9.
10. Chitsaz H, Yee-Greenbaum JL, Tesler G, Lombardo M-J, Dupont CL, Badger JH, Novotny M, Rusch DB, Fraser LJ, Gormley NA, Schulz-Trieglaff O, Smith GP, Evers DJ, Pevzner PA, Lasken RS. 2011. Efficient de novo assembly of single-cell bacterial genomes from short-read data sets. Nat Biotechnol 29:915–921. https://doi.org/10.1038/nbt.1966.
11. Stepanauskas R, Fergusson EA, Brown J, Poulton NJ, Tupper B, Labonté JM, Becraft ED, Brown JM, Pachiadaki MG, Povilaitis T, Thompson BP, Mascena CJ, Bellows WK, Lubys A. 2017. Improved genome recovery and integrated cell-size analyses of individual uncultured microbial cells and viral particles. Nat Commun 8:84. https://doi.org/10.1038/s41467-017-00128-z.
12. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. Nat Biotechnol 36:566–569. https://doi.org/10.1038/nbt.4163.
13. 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.
14. 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.
15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086.
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.
17. Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Ettema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Eren AM, Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol 35:725–731. https://doi.org/10.1038/nbt.3893.
18. Bushnell B. 2018. BBTools: a suite of fast, multithreaded bioinformatics tools designed for analysis of DNA and RNA sequence data. Joint Genome Institute.
19. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Oxford University Press.
20. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics 30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153.