



Metagenome-Assembled Genomes of Four Southern Ocean Archaea Harbor Multiple Genes Linked to Polyethylene Terephthalate and Polyhydroxybutyrate Plastic Degradation

Yvonne Marvellous Akpudo,^{a,b} Oliver K. Bezuiddt,^{a,b}  Thulani P. Makhwanyane^{a,b}

^aMicrobiome@UP, Department of Biochemistry, Genetics, and Microbiology, University of Pretoria, Pretoria, South Africa

^bSouth African Research Chair in Marine Microbiomics, Department of Biochemistry, Genetics, and Microbiology, University of Pretoria, Pretoria, South Africa

ABSTRACT Here, we present four archaeal metagenome-assembled genomes (MAGs) (three *Thaumarchaeota* MAGs and one *Thermoplasmatota* MAG) from a polar upwelling zone in the Southern Ocean. These archaea harbor putative genes encoding enzymes such as polyethylene terephthalate (PET) hydrolases (PETases) and polyhydroxybutyrate (PHB) depolymerases, which are associated with microbial degradation of PET and PHB plastics.

Nitrosopumilus and Marine Group (MG) II archaea affiliated with the phyla *Thaumarchaeota* and *Thermoplasmatota*, respectively, are widely distributed in shallow depths of the oceans (1–5). However, except for their capacity to sequester nitrogen and carbon in the oceans (1, 6, 7), their metabolic functions, including those linked to the biodegradation of plastic polymers, remain largely unclear. To reduce this knowledge deficit, we present four draft genomes of *Nitrosopumilus* and MG II archaea from the Southern Ocean. We further explore their metabolic potential linked to the degradation of polyethylene terephthalate (PET) and polyhydroxybutyrate (PHB).

Two water samples (10 L each) were collected from depths of approximately 5 m and 110 m at a polar upwelling zone (54.007°S, 0.011°E) during the SCALE (Southern Ocean seAsonal Experiment) expedition in October and November 2019. After sequential filtration with 0.5- μ m glass-fiber and 0.2- μ m polycarbonate filters, DNA was extracted from the latter size fraction using the DNeasy PowerWater kit (Qiagen, GmbH, Germany) as detailed in the manufacturer's protocol. Sequencing libraries were generated using the Nextera XT kit (Illumina, Inc., San Diego, CA, USA) and sequenced using an Illumina NovaSeq S4 platform (2 × 150 bp) by Admera Biopharma Service Department (South Plainfield, NJ, USA). The Metagenome-ATLAS v2.4.4 workflow (8) was used to generate metagenome-assembled genomes (MAGs). ATLAS incorporates BBTtools v38.87 for quality control (9) and metaSPAdes v3.14.1 for assembly of contigs (10). These assemblies were binned into MAGs using MetaBAT2 (11) and MaxBin2 (12). The resultant MAGs were refined and dereplicated with DASTool v1.1.2 (13) and dRep v2.5.4 (14), respectively. Completeness and contamination levels were computed using CheckM v1.1.3 (15). MAGs were uploaded onto KBase (16) for taxonomic assignments using GTDB-Tk v1.0.2 and the Genome Taxonomy Database (GTDB) reference release 89 (17). Also, the average nucleotide identity (ANI) values for these MAGs were calculated against each other using FastANI v1.32 (18). For this study, we selected four nonredundant archaeal MAGs, i.e., *Nitrosopumilus* and MG II, MG IIa-L1 (family *Poseidoniceae*), MG IIb-O2, and MG IIb-O3 (family *Thalassoarchaeaceae*). The NCBI Prokaryotic Genome Annotation Pipeline (PGAP) v6.1 (19) was used for annotation of these MAGs, and the resultant data are presented in Table 1. To screen for plastic-degrading genes in these MAGs, protein-coding sequences were predicted with Prodigal v2.6.3 (20), and then HMMScan v3.1b2 (21) was implemented against profiles generated from cluster analysis of reference

Editor Kenneth M. Stedman, Portland State University

Copyright © 2023 Akpudo 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/).

Address correspondence to Thulani P. Makhwanyane, thulani.makhwanyane@up.ac.za.

The authors declare no conflict of interest.

Received 15 October 2022

Accepted 19 January 2023

Published 16 February 2023

TABLE 1 General features of archaeal MAGs from a polar upwelling zone in the Southern Ocean

Genomic feature	Data for MAG:			
	S9-metabat_32 (<i>Nitrosopumilus</i>)	S10-maxbin_48 (MG IIa-L1)	S10-metabat_04 (MG IIb-O2)	S10-metabat_06 (MG IIb-O3)
No. of raw paired-end reads	63,532,776	53,101,891	53,101,891	53,101,891
Genome size (bp)	988,157	1,829,283	1,426,462	1,466,140
GC content (%)	30.98	46.22	38.01	35.85
No. of contigs	181	554	63	50
Minimum contig length (bp)	1,500	1,000	1,500	1,500
N_{50} (bp)	6,719	4,282	47,072	46,931
Completeness (%) ^a	91.13	66.67	82.31	75.73
Contamination (%) ^a	0.97	1.2	6.06	0
No. of predicted genes ^b	1,320	1,767	1,280	1,313
No. of coding sequences (with protein) ^b	1,278	1,719	1,241	1,272
No. of tRNAs ^b	36	36	29	35
No. of rRNAs ^b	1	3	2	3
No. of noncoding RNAs ^b	2	2	2	2
GenBank assembly accession no. for GTDB closest reference strain ^c	GCA_001437625.1	GCA_009937025.1	GCA_003672125.1	GCA_002496735.1
ANI (%) ^d	93.49	79.83	79.76	93.34
BioSample accession no.	SAMN23575727	SAMN23575728	SAMN23575729	SAMN23575730
GenBank accession no.	JAJPRB000000000	JAJPRC000000000	JAJPRD000000000	JAJPRE000000000
SRA accession no.	SRR17224336	SRR17224335	SRR17224335	SRR17224335

^a Estimated with CheckM v1.1.3 (15).

^b Numbers obtained from annotation with NCBI PGAP v6.1 (19).

^c GTDB release 89 nearest reference strain determined with GTDB-Tk v1.0.2 (17).

^d ANI between MAG and nearest GTDB reference strain estimated with FastANI v1.32 (18).

sequences obtained from the Plastics Microbial Biodegradation Database (PMBD) (22) using MCL v14-137 (23). Putative hits homologous to plastic-degrading genes were further analyzed using the HHPred interactive server with HHSuite v3.3.0 (24–26) at high similarity specification (>95% probability), which compared the secondary structures and domains of these predicted genes with references in the Protein Data Bank (PDB) mmCIF70 (27) and Pfam-A v34 (28) databases, respectively. Unless otherwise stated, default parameters were used for all software.

As shown in Fig. 1, this analysis resulted in the prediction of multiple genes encoding enzymes that catalyze PET (29, 30) and PHB (31) plastic degradation in MG II genomes. The *Nitrosopumilus* genome harbored only putative phthalate dioxygenase (PDO) and reductase (*ophA1*) genes, which are associated with the degradation of the plasticizer phthalate (7, 32).

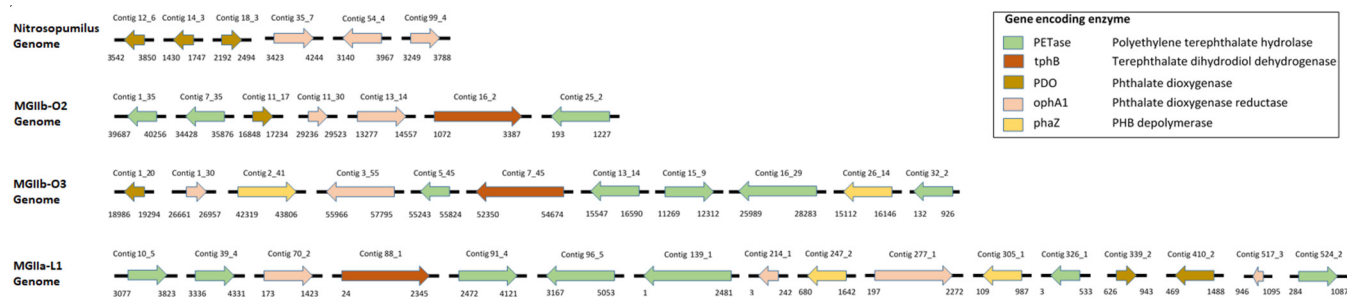


FIG 1 Predicted plastic-polymer-degrading genes, with their orientations indicated, for contiguous sequences of archaeal MAGs from a polar upwelling zone in the Southern Ocean. The predicted genes are shown as arrows, indicating contigs in ascending order. The secondary structure and domains were compared to references in the PDB mmCIF70 and Pfam-A databases, respectively, using HHPred, which yielded high similarity of >95% probability. Labels above the arrows represent contig identification numbers, as specified in the protein-coding sequence from Prodigal. The numbers before and after the underscore represent the sequence position in the genome and the gene position within the contig, respectively. Labels below the arrows show the distance between the annotated start and end, equated as the gene size and represented by the relative length of each arrow.

Data availability. Raw sequence reads (SRA accession numbers [SRR17224336](https://www.ncbi.nlm.nih.gov/sra/SRR17224336) and [SRR17224335](https://www.ncbi.nlm.nih.gov/sra/SRR17224335)) are available in the NCBI GenBank database under BioProject accession number [PRJNA785751](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA785751). The archaeal MAGs were assigned BioSample numbers [SAMN23575727](https://www.ncbi.nlm.nih.gov/biosample/SAMN23575727) to [SAMN23575730](https://www.ncbi.nlm.nih.gov/biosample/SAMN23575730), as well as GenBank accession numbers [JAJPRB000000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRB000000000), [JAJPRC000000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRC000000000), [JAJPRD000000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRD000000000), and [JAJPRE000000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRE000000000). The versions described in this paper are the first versions ([JAJPRB010000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRB010000000), [JAJPRC010000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRC010000000), [JAJPRD010000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRD010000000), and [JAJPRE010000000](https://www.ncbi.nlm.nih.gov/genbank/JAJPRE010000000)).

ACKNOWLEDGMENTS

We thank the crew aboard the *SA Agulhas II* during the October–November 2019 SCALE Expedition.

We acknowledge financial support from the National Research Foundation of South Africa (NRF) (grants UID 110717, UID 129225, UID 136491, and UID 148867). We acknowledge funding received from the European Union Horizon 2020 research and innovation program as part of the AtlantECO project (grant agreement 862923). We also 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. We thank the Technology Innovation Agency (TIA) for supporting the project.

REFERENCES

- Zhang CL, Xie W, Martin-Cuadrado A-B, Rodriguez-Valera F. 2015. Marine Group II archaea, potentially important players in the global ocean carbon cycle. *Front Microbiol* 6:1108. <https://doi.org/10.3389/fmicb.2015.01108>.
- Rinke C, Rubino F, Messer LF, Youssef N, Parks DH, Chuvpochina M, Brown M, Jeffries T, Tyson GW, Seymour JR, Hugenholtz P. 2019. A phylogenomic and ecological analysis of the globally abundant Marine Group II archaea (*Ca. Poseidoniales* ord. nov.). *ISME J* 13:663–675. <https://doi.org/10.1038/s41396-018-0282-y>.
- Martin-Cuadrado A-B, Garcia-Heredia I, Molto AG, Lopez-Ubeda R, Kimes N, López-García P, Moreira D, Rodriguez-Valera F. 2015. A new class of marine Euryarchaeota group II from the Mediterranean deep chlorophyll maximum. *ISME J* 9:1619–1634. <https://doi.org/10.1038/ismej.2014.249>.
- Castillo DJ, Dithugoe CD, Bezuidt OK, Makhalanyane TP. 2022. Microbial ecology of the Southern Ocean. *FEMS Microbiol Ecol* 98:fiac123. <https://doi.org/10.1093/femsec/fiac123>.
- Phoma S, Vikram S, Jansson JK, Ansoorge IJ, Cowan DA, Van de Peer Y, Makhalanyane TP. 2018. Agulhas Current properties shape microbial community diversity and potential functionality. *Sci Rep* 8:10542. <https://doi.org/10.1038/s41598-018-28939-0>.
- Graham ED, Tully BJ. 2021. Marine dadabacteria exhibit genome streamlining and phototrophy-driven niche partitioning. *ISME J* 15:1248–1256. <https://doi.org/10.1038/s41396-020-00834-5>.
- Correll CC, Batié CJ, Ballou DP, Ludwig ML. 1992. Phthalate dioxygenase reductase: a modular structure for electron transfer from pyridine nucleotides to [2Fe-2S]. *Science* 258:1604–1610. <https://doi.org/10.1126/science.1280857>.
- Kieser S, Brown J, Zdobnov EM, Trajkovski M, McCue LA. 2020. ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data. *BMC Bioinformatics* 21:257. <https://doi.org/10.1186/s12859-020-03585-4>.
- Bushnell B, Rood J, Singer E. 2017. BBMerge: accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056. <https://doi.org/10.1371/journal.pone.0185056>.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
- Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>.
- Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>.
- Sieber CM, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>.
- Olm MR, Brown CT, Brooks B, Banfield JF. 2017. dRep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. *ISME J* 11:2864–2868. <https://doi.org/10.1038/ismej.2017.126>.
- 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>.
- 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>.
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
- Jain C, Rodriguez-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>.
- 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>.
- 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>.
- Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>.
- Gan Z, Zhang H. 2019. PMBD: a comprehensive plastics microbial biodegradation database. *Database (Oxford)* 2019:baz119. <https://doi.org/10.1093/database/baz119>.
- Enright AJ, Van Dongen S, Ouzounis CA. 2002. An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Res* 30:1575–1584. <https://doi.org/10.1093/nar/30.7.1575>.
- Zimmermann L, Stephens A, Nam S-Z, Rau D, Kübler J, Lozajic M, Gabler F, Söding J, Lupas AN, Alva V. 2018. A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core. *J Mol Biol* 430:2237–2243. <https://doi.org/10.1016/j.jmb.2017.12.007>.
- Gabler F, Nam SZ, Till S, Mirdita M, Steinegger M, Söding J, Lupas AN, Alva

- V. 2020. Protein sequence analysis using the MPI Bioinformatics Toolkit. *Curr Protoc Bioinformatics* 72:e108. <https://doi.org/10.1002/cpbi.108>.
26. Meier A, Söding J. 2015. Automatic prediction of protein 3D structures by probabilistic multi-template homology modeling. *PLoS Comput Biol* 11: e1004343. <https://doi.org/10.1371/journal.pcbi.1004343>.
27. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow GV, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranović V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao Y-P, Voigt M, Westbrook JD, Young JY, Zardecki C, Zhuravleva M. 2021. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res* 49:D437–D451. <https://doi.org/10.1093/nar/gkaa1038>.
28. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, Heger A, Hetherington K, Holm L, Mistry J, Sonnhammer EL, Tate J, Punta M. 2014. Pfam: the protein families database. *Nucleic Acids Res* 42: D222–D230. <https://doi.org/10.1093/nar/gkt1223>.
29. Austin HP, Allen MD, Donohoe BS, Rorrer NA, Kearns FL, Silveira RL, Pollard BC, Dominick G, Duman R, El Omari K, Mykhaylyk V, Wagner A, Michener WE, Amore A, Skaf MS, Crowley MF, Thorne AW, Johnson CW, Woodcock HL, McGeehan JE, Beckham GT. 2018. Characterization and engineering of a plastic-degrading aromatic polyesterase. *Proc Natl Acad Sci U S A* 115:E4350–E4357. <https://doi.org/10.1073/pnas.1718804115>.
30. Dimarogona M, Nikolaivits E, Kanelli M, Christakopoulos P, Sandgren M, Topakas E. 2015. Structural and functional studies of a *Fusarium oxysporum* cutinase with polyethylene terephthalate modification potential. *Biochim Biophys Acta* 1850:2308–2317. <https://doi.org/10.1016/j.bbagen.2015.08.009>.
31. Hisano T, Kasuya K-I, Tezuka Y, Ishii N, Kobayashi T, Shiraki M, Oroudjev E, Hansma H, Iwata T, Doi Y, Saito T, Miki K. 2006. The crystal structure of polyhydroxybutyrate depolymerase from *Penicillium funiculosum* provides insights into the recognition and degradation of biopolyesters. *J Mol Biol* 356:993–1004. <https://doi.org/10.1016/j.jmb.2005.12.028>.
32. Mahto JK, Neetu N, Waghmode B, Kuatsjah E, Sharma M, Sircar D, Sharma AK, Tomar S, Eltis LD, Kumar P. 2021. Molecular insights into substrate recognition and catalysis by phthalate dioxygenase from *Comamonas testoteroni*. *J Biol Chem* 297:101416. <https://doi.org/10.1016/j.jbc.2021.101416>.