Four *Lentisphaerae* Family Metagenome-Assembled Genomes from the South Atlantic Ocean

Oluwatayo A. Makinde,[a] Percy Mutseka Lunga,[a] Oliver K. I. Bezuidt,[a] ⓘ Thulani P. Makhalanyane[a]

[a]Department of Biochemistry, Genetics and Microbiology, University of Pretoria, Pretoria, South Africa

ABSTRACT   We present four *Lentisphaerae* metagenome-assembled genomes (MAGs) from the South Atlantic Ocean. The medium-quality genomes, affiliated with the family of *Lentisphaeraceae,* ranged from 4.86 to 5.46 Mbp and harbored the genetic capacity to produce secondary metabolites. This resource provides a basis for investigating the functional attributes of this phylum.

*L*entisphaerae is a member of the superphylum PVC (*Planctomycetes, Verrucomicrobia*, and *Chlamydia*) (1), known for their biotechnological importance and capacity to produce transparent exopolymers (2, 3). These constituents mediate the cycling of organic carbon particles between the euphotic zone and the deep ocean (4–7). Currently, there are only two validly described orders of this phylum, the *Victivallales* and the *Lentisphaerales*, due to difficulties associated with reproducing the ideal laboratory conditions required to isolate members of this species (7–10). Metagenome-assembled genomes (MAGs) have provided substantial insights regarding *Lentisphaerae* from several ecosystems (11–13). Nevertheless, we lack taxonomic and functional insights regarding *Lentisphaerae* from diverse ecosystems, such as the South Atlantic Ocean.

South Atlantic water samples (45 liters) were retrieved from three stations at a 5-m depth (Table 1), filtered through 0.2-μm polyethersulfone (PES) filter membranes (Merck, Republic of South Africa), and stored at -80°C until further processing. Metagenomic extractions were performed as described previously (14, 15) using the DNeasy PowerSoil kit (Qiagen, Hilden, Germany), and the resultant high-quality DNA was used to construct libraries using the KAPA HyperPrep kit (KAPA Biosystems, Massachusetts, USA) as detailed by the manufacturer. Sequencing was performed using an Illumina HiSeq 2000 platform (2×250 bp). Trimmomatic version 0.36 and PRINSEQ lite version 0.20.4 (-lc_method dust) were used to remove adaptors and overrepresented and low-quality reads from raw reads, respectively (16, 17). The resultant reads were assembled into contiguous segments using MegaHIT version 1.2.3 (18) and binned using MetaBAT 2 (Table 1) (19). CheckM version 1.0.18 was used to evaluate the completeness and contamination value of each MAG (20). The Genome Taxonomy Database Toolkit (GTDB-Tk) version 1.6.0 release 89 (P.-A. Chaumeil, A. J. Mussig, P. Hugenholtz, and D. H. Parks) was used to assign taxonomy, and the relative abundance of each MAG was determined using CoverM version 0.6.1 (21). Gene annotation was done using PGAP version 6.1 (22). The protein-encoding regions were identified using Prodigal version 2.6.3 (23), and the amino acid identities (AAI) of the four *Lentisphaerae* MAGs were calculated using AAI calculator (24) against the *Lentisphaera araneosa* genome sequence (GCF_0000170755.1), which had the highest similarly as determined using GTDB-Tk classification. The biosynthetic potentials of the four *Lentisphaerae* were determined using antiSMASH version 5.2 (25), using the strict detection mode option. The default parameters were used for all software unless otherwise indicated.

The four *Lentisphaerae* genomes were classified as medium-quality draft genomes, consistent with criteria established by the Genomic Standards Consortium (26). These genomes were assigned to the *Lentisphaeraceae* family based on GTDB-Tk classification. The AAI ranged from 47.95 to 49.38%, based on the reference genome with the closest placement (*L. araneosa*).

TABLE 1 Summary of genome statistics and comparisons of four novel *Lentisphaerae* MAGs, as well as sampling coordinates

Characteristic	Value or description for[a]:			
	Lentisphaeraceae_S22.bin.5	Lentisphaeraceae_S24.bin.22	Lentisphaeraceae_S14.bin.16	Lentisphaeraceae_S10.bin.24
Sampling coordinates	47°28.906′S, 09°59.970′W	47°28.906′S, 09°59.970′W	38°22.947′S, 11°14.942′W	36°35.591′S, 10°06.384′W
Depth of sequencing per metagenome (Gbp)	6	6	6	6
Bin coverage/relative abundance (%)	0.95	0.77	0.59	1.42
Genome size (bp)	5,465,546	5,228,557	4,867,689	5,264,327
G+C content (%)	39.34	39.03	40.84	39.19
Estimated genome completeness (%)	94.8	89.02	78.63	95.14
Estimated genome contamination (%)	9.42	4.17	8.16	6.93
Genome quality	Medium	Medium	Medium	Medium
No. of contigs	655	731	738	177
Largest contig (bp)	42,385	33,909	34,929	1,5864,44
N_{50} (bp)	10,158	8,194	7,627	44,916
No. of genes associated with:				
Protein coding	4,943	4,780	4,646	4,337
tRNA	53	35	36	38
rRNA	2	0	2	3
Secondary metabolic pathway for:				
Aryl polyene	1	2	2	0
Ripp-like	1	1	1	0
Ectoin	1	0	1	1
NRPS	0	0	1	0
Terpene	1	1	1	1
NCBI accession no. for:				
Raw reads	SRR15221251	SRR15221249	SRR15221260	SRR15221242
Assembly	JAJPOT000000000	JAJPOS000000000	JAJPOR000000000	JAJPOQ000000000
AAI with GCA_000170755.1_ASM17075v1 (%)[b]	48.5	48.58	47.95	49.38
Name of isolate with highest-similarity 16S rRNA gene	Uncultured bacterium clone	NA	NA	NA
GenBank accession no.	KF771565.1	NA	NA	NA
% similarity	98.98	NA	NA	NA

[a] NA, not available.
[b] AAI, amino acid identity.

This result suggests that the four genomes belong to the same family as *L. araneosa* (GCF_0000170755.1) (Table 1). Interestingly, all the genomes harbored genes for unique secondary metabolic pathways, including terpene, aryl polyene, Ripp-like, ectoin, and nonribosomal peptide synthetase (NRPS) (Table 1).

Data availability. The NCBI BioProject accession number is PRJNA748242, while the Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank and the accession numbers are given in Table 1.

REFERENCES

1. Rivas-Marín E, Devos DP. 2018. The paradigms they are a-changin': past, present and future of PVC bacteria research. Antonie Van Leeuwenhoek 111:785–799. https://doi.org/10.1007/s10482-017-0962-z.

2. Wagner M, Horn M. 2006. The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr Opin Biotechnol 17:241–249. https://doi.org/10.1016/j.copbio.2006.05.005.

3. Van Niftrik L, Devos DP. 2017. *Planctomycetes-Verrucomicrobia-Chlamydiae* bacterial superphylum: new model organisms for evolutionary cell biology. Front Microbiol 8:1458. https://doi.org/10.3389/fmicb.2017.01458.

4. Alldredge AL, Passow U, Logan BE. 1993. The abundance and significance of a class of large, transparent organic particles in the ocean. Deep Sea Res 1 Oceanogr Res Papers 40:1131–1140. https://doi.org/10.1016/0967-0637(93)90129-Q.

5. Passow U, Alldredge AL, Logan BE. 1994. The role of particulate carbohydrate exudates in the flocculation of diatom blooms. Deep Sea Res 1 Oceanogr Res Papers 41:335–357. https://doi.org/10.1016/0967-0637(94)90007-8.

6. Logan BE, Passow U, Alldredge AL, Grossartt H-P, Simont M. 1995. Rapid formation and sedimentation of large aggregates is predictable from coagulation rates (half-lives) of transparent exopolymer particles (TEP). Deep Sea Res 2 Top Stud Oceanogr 42:203–214. https://doi.org/10.1016/0967-0645(95)00012-F.

7. Cho J-C, Vergin KL, Morris RM, Giovannoni SJ. 2004. *Lentisphaera araneosa* gen. nov., sp. nov, a transparent exopolymer producing marine bacterium, and the description of a novel bacterial phylum, *Lentisphaerae*. Environ Microbiol 6:611–621. https://doi.org/10.1111/j.1462-2920.2004.00614.x.

8. Thrash JC, Cho J-C, Vergin KL, Morris RM, Giovannoni SJ. 2010. Genome sequence of *Lentisphaera araneosa* HTCC2155T, the type species of the order *Lentisphaerales* in the phylum *Lentisphaerae*. J Bacteriol 192:2938–2939. https://doi.org/10.1128/JB.00208-10.

9. Hedlund BP, Cho J-C, Derrien M, Costa KC. 2010. Phylum XXII. Lentisphaerae Cho, Vergin, Morris and Giovannoni 2004a, 1005[VP] (effective publication: Cho, Vergin, Morris and Giovannoni 2004b, 617.), p 785–793. *In* Krieg NR, Staley JT, Brown DR, Hedlund BP, Paster BJ, Ward NL, Ludwig W, Whitman WB (ed), Bergey's manual of systematic bacteriology, vol 4. Springer, New York, NY.

10. Van Passel MW, Kant R, Airi P, Lucas S, Copeland A, Lapidus A, del Rio TG, Dalin E, Tice H, Bruce D, Goodwin L, Pitluck S, Walston Davenport K, Sims D, Brettin TS, Detter JC, Han S, Larimer FW, Land ML, Hauser L, Kyrpides N, Ovchinnikova G, Richardson PP, de Vos WM, Smidt H, Zoetendal EG. 2011. Genome sequence of *Victivallis vadensis* ATCC BAA-548, an anaerobic bacterium from the phylum *Lentisphaerae*, isolated from the human gastrointestinal tract. J Bacteriol 193:2373–2374. https://doi.org/10.1128/JB.00271-11.

11. Gharechahi J, Salekdeh GH. 2018. A metagenomic analysis of the camel rumen's microbiome identifies the major microbes responsible for lignocellulose degradation and fermentation. Biotechnol Biofuels 11:1–19. https://doi.org/10.1186/s13068-018-1214-9.

12. Peterson BD, Peterson BD, McDaniel EA, Schmidt AG, Lepak RF, Janssen SE, Tran PQ, Marick RA, Ogorek JM, DeWild JF, Krabbenhoft DP, McMahon KD. 2020. Mercury methylation trait dispersed across diverse anaerobic microbial guilds in a eutrophic sulfate-enriched lake. Environ Sci Technol 54:15840–15851. https://doi.org/10.1021/acs.est.0c05435.

13. Wilkinson T, Korir D, Ogugo M, Stewart RD, Watson M, Paxton E, Goopy J, Robert C. 2020. 1200 high-quality metagenome-assembled genomes from the rumen of African cattle and their relevance in the context of sub-optimal feeding. Genome Biol 21:1–25. https://doi.org/10.1186/s13059-020-02144-7.

14. Hirai M, Nishi S, Tsuda M, Sunamura M, Takaki Y, Nunoura T. 2017. Library construction from subnanogram DNA for pelagic sea water and deep-sea sediments. Microbes Environ 32:336–343. https://doi.org/10.1264/jsme2.ME17132.

15. Phoma BS, Makhalanyane TP. 2021. Depth-dependent variables shape community structure and functionality in the Prince Edward Islands. Microb Ecol 81:396–409. https://doi.org/10.1007/s00248-020-01589-4.

16. 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.

17. Schmieder R, Edwards R. 2012. Insights into antibiotic resistance through metagenomic approaches. Future Microbiol 7:73–89. https://doi.org/10.2217/fmb.11.135.

18. Li D, Luo R, Liu C-M, Leung C-M, Ting H-F, Sadakane K, Yamashita H, Lam T-W. 2016. MEGAHIT v1. 0: a fast and scalable metagenome assembler driven by advanced methodologies and community practices. Methods 102:3–11. https://doi.org/10.1016/j.ymeth.2016.02.020.

19. 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.

20. 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.

21. Woodcroft BJ, Newell R. 2017. CoverM: read coverage calculator for metagenomics. https://github.com/wwood/CoverM.

22. 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.

23. 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.

24. Rodriguez-R LM, Konstantinidis KT. 2014. Bypassing cultivation to identify bacterial species. Microbe 9:111–118. https://doi.org/10.1128/microbe.9.111.1.

25. Blin K, Shaw S, Kautsar SA, Medema MH, Weber T. 2021. The antiSMASH database version 3: increased taxonomic coverage and new query features for modular enzymes. Nucleic Acids Res 49:D639–D643. https://doi.org/10.1093/nar/gkaa978.

26. 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.