Check for updates

Metagenome-Assembled Genome Sequences of Three Uncultured *Planktomarina* sp. Strains from the Northeast Atlantic Ocean

Matilde Marques,[a] Nuno Borges,[a] 🆔 Sandra Godinho Silva,[a] Ulisses Nunes da Rocha,[b] Asunción Lago-Lestón,[c] Tina Keller-Costa,[a] 🆔 Rodrigo Costa[a,d,e,f]

[a]iBB–Institute for Bioengineering and Biosciences, Department of Bioengineering, Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal
[b]Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany
[c]Department of Medical Innovation, Centro de Investigación Científica y de Educación Superior de Ensenada (CICESE), Ensenada, Mexico
[d]Centro de Ciências do Mar (CCMAR), Universidade do Algarve, Faro, Portugal
[e]Department of Energy Joint Genome Institute, Berkeley, California, USA
[f]Lawrence Berkeley National Laboratory, Berkeley, California, USA

ABSTRACT We report three metagenome-assembled genomes (MAGs) of *Planktomarina* strains from coastal seawater (Portugal) to help illuminate the functions of understudied *Rhodobacteraceae* bacteria in marine environments. The MAGs encode proteins involved in aerobic anoxygenic photosynthesis and a versatile carbohydrate metabolism, strengthening the role of *Planktomarina* species in oceanic carbon cycling.

Bacterioplankton communities are key players in nutrient cycling in the world's oceans (1). The *Roseobacter* clade-affiliated (RCA) cluster (*Rhodobacteraceae, Alphaproteobacteria*) has been found to constitute up to 35% of coastal marine bacterioplankton, to be widely distributed from temperate to polar waters, and to play an important role in the degradation of phytoplankton-derived organic matter (1–5). The genus *Planktomarina* is a dominant member of the RCA community (3, 6), with the photoheterotrophic type strain *Planktomarina temperata* RCA23 isolated in 2013 from the Wadden Sea (2). We report three metagenome-assembled *Planktomarina* sp. genomes from the Northeast Atlantic Ocean that represent a novel candidate species within the RCA cluster.

Three seawater samples were obtained at ca. 18-m depth off the coast of Faro, Algarve, Portugal (latitude 36.979778, longitude –7.989111) in June 2014. Samples were filtered through sterile 0.22-μm nitrocellulose membrane filters, and the total community DNA was extracted using the UltraClean soil DNA isolation kit (Mo Bio Laboratories) (7; T. Keller-Costa, A. Lago-Leston, J. P. Saraiva, R. Toscan, S. G. Silva, J. Gonçalves, C. J. Cox, N. C. Kyrpides, U. Nunes da Rocha, and R. Costa, submitted for publication). DNA libraries were prepared using the Nextera DNA sample preparation kit from Illumina and subjected to paired-end metagenome sequencing (Illumina HiSeq 2500; depth, ~20 million 101-bp reads per sample) with 200 cycles. The raw sequence reads were trimmed and quality checked with the Trim Galore module with MetaWRAP v1.0.5 (8). Assemblies were performed with the quality-filtered, unmerged reads using metaSPAdes (9). Eukaryotic contigs, assigned with EukRep (10), were thereafter filtered out. The metagenome-assembled genomes (MAGs) were then binned from the obtained prokaryotic-enriched assemblies with MetaBAT2 (11) and subjected to taxonomic classification with GTDB-Tk v0.3.2 (12) on the KBase server (13) and with the Microbial Genomes Atlas (MiGA) (14). Three MAGs identified as a *Planktomarina* sp. (strains SW01_Bin08, SW02_Bin14, and SW04_Bin04) showing standard completeness

Address correspondence to Rodrigo Costa, rodrigoscosta@tecnico.ulisboa.pt.

TABLE 1 General features of metagenome-assembled *Planktomarina* genomes from the Northeast Atlantic Ocean

Isolate	Original sample[a] (BioSample no.)	Compl/cont[b]	Size (bp)	No. of contigs	No. of CDSs[c]	N_{50} (bp)	GC content (%)	ANI (%)[d]	AAI (%)[d]	Genome accession no.
SW04_Bin04	SW04 (SAMEA3913377)	78.4/1.8	2,419,173	144	2,489	27,162	54.8	87.43	91.94	GCA_902732755
SW01_Bin08	SW01 (SAMEA3913374)	72.1/1.8	2,208,500	136	2,255	22,913	55.0	87.56	92.19	GCA_902732765
SW02_Bin14	SW02 (SAMEA3913375)	80.2/2.7	2,269,649	125	2,326	27,294	55.0	87.46	91.95	GCA_902732775

[a] The original metagenome BioSample from which the MAG was obtained.
[b] Compl, completeness score; Cont, contamination score.
[c] CDSs, coding DNA sequences.
[d] ANI and AAI were obtained with MiGA against the genome sequence of type strain *P. temperata* RCA23 (accession number GCA_000738435).

and contamination scores were chosen for this study (Table 1). The average nucleotide identity blast (ANIb) and average amino acid identity (AAI) values were calculated between MAGs and *P. temperata* RCA23 using JSpeciesWS (15) and aai.rb (16), respectively. Functional annotations were performed with the RAST server (17), and clusters of orthologous groups (COGs) of proteins and protein families (Pfams) were predicted with WebMGA (18). antiSMASH v5.0 (19) was used to identify biosynthetic gene clusters. Default parameters were used for all software unless otherwise noted.

The three MAGs shared 99% ANI and AAI among themselves but presented only 87.5% ANI with *P. temperata* RCA23 (Table 1), suggesting that they may represent a novel *Planktomarina* species. Nevertheless, the MAGs shared 1,217 "core" COGs with *P. temperata* RCA23 out of 1,575 "pangenome" COGs detected across all four genomes. Among the common features between the MAGs and the *P. temperata* RCA23 genome, we highlight genes encoding proteins involved in aerobic anoxygenic photosynthesis (e.g., chlorophyll *a* synthase, EC 2.5.1.62) and several metallo-beta-lactamases associated with antibiotic resistance, along with a terpenoid gene cluster showing 100% similarity to a carotenoid cluster (BGC0000647) from *Rhodobacter sphaeroides*. Strengthening the role of *Planktomarina* spp. in carbon cycling, multiple glycoside hydrolase genes (e.g., alpha-amylase [PF00128], alpha-galactosidase [EC 3.2.1.22], and beta-galactosidase [EC 3.2.1.23]) were also common to all genomes, with chitinase-encoding genes (EC 3.2.1.14) being furthermore detected in the *P. temperata* RCA23 and SW01_Bin08 genomes.

Data availability. The metagenome-assembled genomes as well as the raw metagenome data have been deposited at ENA/EMBL/GenBank under the study accession number PRJEB13222. The BioSample accession numbers of the MAGs are SAMEA6497282 through SAMEA6497284, and their GenBank accession numbers are GCA_902732755 (SW04_Bin04), GCA_902732765 (SW01_Bin08), and GCA_902732775 (SW02_Bin14). The MAGs were obtained from the marine metagenome BioSamples SW01 (SAMEA3913374), SW02 (SAMEA3913375), and SW04 (SAMEA3913377).

REFERENCES

1. Selje N, Simon M, Brinkhoff T. 2004. A newly discovered *Roseobacter* cluster in temperate and polar oceans. Nature 427:445–448. https://doi.org/10.1038/nature02272.

2. Giebel H-A, Kalhoefer D, Gahl-Janssen R, Choo Y-J, Lee K, Cho J-C, Tindall BJ, Rhiel E, Beardsley C, Aydogmus ÖO, Voget S, Daniel R, Simon M, Brinkhoff T. 2013. *Planktomarina temperata* gen. nov., sp. nov., belonging to the globally distributed RCA cluster of the marine *Roseobacter* clade, isolated from the German Wadden Sea. Int J Syst Evol Microbiol 63:4207–4217. https://doi.org/10.1099/ijs.0.053249-0.

3. Giebel H-A, Kalhoefer D, Lemke A, Thole S, Gahl-Janssen R, Simon M, Brinkhoff T. 2011. Distribution of *Roseobacter* RCA and SAR11 lineages in the North Sea and characteristics of an abundant RCA isolate. ISME J 5:8–19. https://doi.org/10.1038/ismej.2010.87.

4. Bakenhus I, Dlugosch L, Billerbeck S, Giebel H-A, Milke F, Simon M. 2017. Composition of total and cell-proliferating bacterioplankton community in early summer in the North Sea—*Roseobacters* are the most active component. Front Microbiol 8:1771. https://doi.org/10.3389/fmicb.2017.01771.

5. Zhou J, Richlen ML, Sehein TR, Kulis DM, Anderson DM, Cai Z. 2018. Microbial community structure and associations during a marine dinoflagellate bloom. Front Microbiol 9:1201. https://doi.org/10.3389/fmicb.2018.01201.

6. Mayali X, Franks PJS, Azam F. 2008. Cultivation and ecosystem role of a marine *Roseobacter* clade-affiliated cluster bacterium. Appl Environ Microbiol 74:2595–2603. https://doi.org/10.1128/AEM.02191-07.

7. Keller-Costa T, Eriksson D, Gonçalves JMS, Gomes NCM, Lago-Leston A, Costa R. 2017. The gorgonian coral Eunicella labiata hosts a distinct prokaryotic consortium amenable to cultivation. FEMS Microbiol Ecol 93:1–19. https://doi.org/10.1093/femsec/fix143.

8. Uritskiy GV, DiRuggiero J, Taylor J. 2018. MetaWRAP—a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome 6:158. https://doi.org/10.1186/s40168-018-0541-1.

9. 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.

10. West PT, Probst AJ, Grigoriev IV, Thomas BC, Banfield JF. 2018. Genome-reconstruction for eukaryotes from complex natural microbial communities. Genome Res 28:569–580. https://doi.org/10.1101/gr.228429.117.

11. 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.

12. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. Bioinformatics:btz848. https://doi.org/10.1093/bioinformatics/btz848.

13. 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.

14. Rodriguez-R LM, Gunturu S, Harvey WT, Rosselló-Mora R, Tiedje JM, Cole JR, Konstantinidis KT. 2018. The Microbial Genomes Atlas (MiGA) webserver: taxonomic and gene diversity analysis of Archaea and Bacteria at the whole genome level. Nucleic Acids Res 46:W282–W288. https://doi.org/10.1093/nar/gky467.

15. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a Web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinformatics 32:929–931. https://doi.org/10.1093/bioinformatics/btv681.

16. Rodriguez-R LM, Konstantinidis KT. 2016. The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. PeerJ Prepr 4:e1900v1. https://doi.org/10.7287/peerj.preprints.1900v1.

17. Overbeek R, Olson R, Pusch GD, Olsen GJ, Davis JJ, Disz T, Edwards RA, Gerdes S, Parrello B, Shukla M, Vonstein V, Wattam AR, Xia F, Stevens R. 2014. The SEED and the Rapid Annotation of microbial genomes using Subsystems Technology (RAST). Nucleic Acids Res 42:D206–D214. https://doi.org/10.1093/nar/gkt1226.

18. Wu S, Zhu Z, Fu L, Niu B, Li W. 2011. WebMGA: a customizable Web server for fast metagenomic sequence analysis. BMC Genomics 12:444. https://doi.org/10.1186/1471-2164-12-444.

19. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. Nucleic Acids Res 47:W81–W87. https://doi.org/10.1093/nar/gkz310.