



Draft Genome Sequence of “*Candidatus Moanabacter tarae*,” Representing a Novel Marine Verrucomicrobial Lineage

Julian Vosseberg,^{a,b} Joran Martijn,^a Thijs J. G. Ettema^a

^aDepartment of Cell and Molecular Biology, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

^bTheoretical Biology and Bioinformatics, Department of Biology, Utrecht University, Utrecht, The Netherlands

ABSTRACT The Tara Oceans Consortium has published various metagenomes of marine environmental samples. Here, we report a contig of 2.6 Mbp from the assembly of a sample collected near the Marquesas Islands in the Pacific Ocean, covering a nearly complete novel verrucomicrobial genome. We propose the name “*Candidatus Moanabacter tarae*” for the corresponding bacterium.

Verrucomicrobia is a bacterial phylum within the PVC superphylum, which additionally includes, *inter alia*, *Planctomycetes* and *Chlamydiae* (1). Although the phylum comprises a relatively small fraction of the marine bacterioplankton (2), its widespread occurrence tentatively points to significant ecological roles (3). From the vast majority of prokaryotic diversity, including verrucomicrobial lineages, genomic sequences are lacking. Recent advances in cultivation-independent approaches have led to the elucidation of many novel lineages and have provided various new insights (4). A promising source for the discovery and retrieval of new genomes is the metagenomic data of ocean samples generated by the Tara Oceans Consortium, which have been made publically available (5).

The sample of interest had been collected at Tara sampling site 125 (8°54'14.35"S 142°33'58.54"W), marine epipelagic mixed layer (140 m), size fraction 0.22 to 3 μm (GenBank BioSample number [ERS492926](https://www.ncbi.nlm.nih.gov/biosample/ERS492926)), and had subsequently been sequenced using Illumina HiSeq technology (SRA accession number [ERR599156](https://www.ncbi.nlm.nih.gov/sra/ERR599156)) (5). The preprocessing and assembly were performed as described previously (6). The raw sequence data were downloaded from the European Bioinformatics Institute (EMBL-EBI) (ENA accession number [ERR599156](https://www.ebi.ac.uk/ena/record/ERR599156)). From the resulting sequencing reads, we removed adapter sequences and low-quality ends and reads using Trimmomatic v0.35 (ILLUMINACLIP, TruSeq-3-PE-2.fa:2:30:10:1:true; LEADING, 3; TRAILING, 3; MINLEN, 60; AVGQUAL, 29) (7). Assembly was performed with the meta option in SPAdes v3.7.0 (8), using *k*-mers of 21, 33, 55, and 77.

The longest contig in this assembly was 2,633,965 bp and had a GC content of 46%. Based on the presence of 131 out of 139 bacterial single-copy marker genes (9) and the genome completeness estimation method miComplete v0.2.0 (<https://bitbucket.org/evolegiolab/micomplete>), which takes the general distances between these marker genes into account, the contig was estimated to cover 92% of the complete genome. This genome was thereby calculated to have a length of approximately 2.85 Mbp.

The contig was annotated with Prokka v1.13 (default parameters) (10). A total of 2,237 coding sequences, 48 tRNA genes, and 3 rRNA genes were predicted to be present. A BLAST search of the 16S rRNA gene recovered several uncultured marine sediment bacteria (GenBank accession numbers [FN553635](https://www.ncbi.nlm.nih.gov/nuccore/FN553635), [JF809735](https://www.ncbi.nlm.nih.gov/nuccore/JF809735), [AB694308](https://www.ncbi.nlm.nih.gov/nuccore/AB694308), and [AB694307](https://www.ncbi.nlm.nih.gov/nuccore/AB694307); 91 to 92% identity), epixenosomes of *Euplotidium arenarium* (87% identity), and *Coralimargarita akajimensis* (86% identity) as the top hits, demonstrating the verrucomicrobial origin and novel nature of the contig. The verrucomicrobial affiliation of this new genome was confirmed by its position in a reference genome tree based on

Received 9 July 2018 Accepted 11 September 2018 Published 18 October 2018

Citation Vosseberg J, Martijn J, Ettema TJG. 2018. Draft genome sequence of “*Candidatus Moanabacter tarae*,” representing a novel marine verrucomicrobial lineage. *Microbiol Resour Announc* 7:e00951-18. <https://doi.org/10.1128/MRA.00951-18>.

Editor Frank J. Stewart, Georgia Institute of Technology

Copyright © 2018 Vosseberg 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 Thijs J. G. Ettema, thijs.ettema@icm.uu.se.

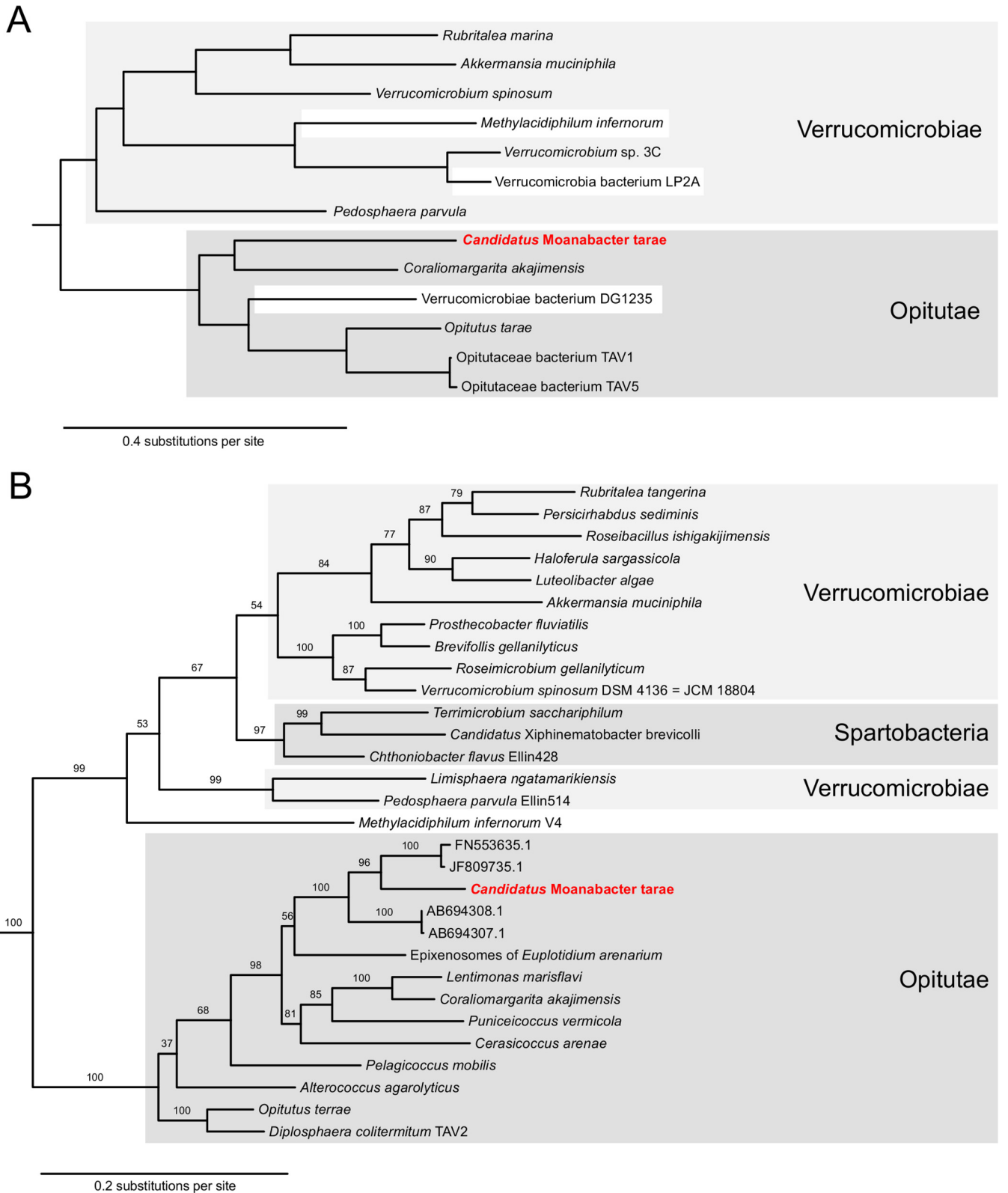


FIG 1 Phylogenetic position of the newly identified “*Candidatus Moanabacter tarae*” within the verrucomicrobial class *Opitutae*. (A) Placement of “*Candidatus Moanabacter tarae*” (red) within the *Verrucomicrobia* in the CheckM reference tree, which is based on 43 universal single-copy marker genes. (B) Verrucomicrobial phylogeny based on the 16S rRNA gene indicating the evolutionary relationship of “*Candidatus Moanabacter tarae*” (red) with the distinct verrucomicrobial classes. Branch labels correspond to bootstrap values. The tree was rooted with *Lentisphaerae*, *Chlamydiae*, and *Planctomycetes* (data not shown).

Downloaded from <http://mra.asm.org/> on November 27, 2018 by guest

43 marker genes (Fig. 1A), as inferred by CheckM v1.0.11 (tree command, with the predicted proteins provided as input) (11).

To more precisely infer the phylogenetic position of this novel verrucomicrobial lineage, we aligned the 16S rRNA gene with the top BLAST hits and other 16S rRNA sequences from a diverse set of *Verrucomicrobia*, *Lentisphaerae*, *Chlamydiae*, and *Planctomycetes* species, with MAFFT E-INS-i v7.050b (12) and inferred a phylogenetic tree with IQ-TREE v1.5.0a (GTR+R4 selected by the $-m$ TESTNEW option, with 100 nonparametric bootstraps) (13). The resulting tree was in line with the BLAST and CheckM results and indicated that the organism represented by the contig was a member of a clade further consisting of marine sediment bacteria, within the verrucomicrobial class *Opitutae* (Fig. 1B).

For the newly identified marine bacterium reported here, we propose the name “*Candidatus* Moanabacter tarae,” named after the Marquesan word for ocean, “moana,” and the Tara Oceans project, with Tara aptly also being the name of a sea goddess in Polynesian mythology.

Data availability. The genome sequence reported here has been deposited in GenBank under the accession number [CP029803](https://doi.org/10.1093/bioinformatics/btu170).

ACKNOWLEDGMENTS

We are grateful to the Tara Oceans Consortium for collecting and publishing the sequencing data and for their permission to use their data. The assembly was performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) through the Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX).

This work was supported by grants from the European Research Council (ERC starting grant 310039-PUZZLE_CELL), the Swedish Foundation for Strategic Research (SSF-FFL5), and the Swedish Research Council (VR grant 2015-04959) to T.J.G.E.

REFERENCES

- 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>.
- Sunagawa S, Coelho LP, Chaffron S, Kultima JR, Labadie K, Salazar G, Djahanschiri B, Zeller G, Mende DR, Alberti A, Cornejo-Castillo FM, Costea PI, Cruaud C, d'Ovidio F, Engelen S, Ferrera I, Gasol JM, Guidi L, Hildebrand F, Kokoszka F, Lepoivre C, Lima-Mendez G, Poulain J, Poulos BT, Royo-Llonch M, Sarmiento H, Vieira-Silva S, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Bowler C, de Vargas C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Not F, Ogata H, Pesant S, Speich S, Stemmann L, Sullivan MB, Weissenbach J, Wincker P, Karsenti E, Raes J, Acinas SG, Bork P, Boss E, Bowler C, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sieracki M, Velayoudon D. 2015. Structure and function of the global ocean microbiome. *Science* 348:1261359. <https://doi.org/10.1126/science.1261359>.
- Cardman Z, Arnosti C, Durbin A, Ziervogel K, Cox C, Steen AD, Teske A. 2014. *Verrucomicrobia* are candidates for polysaccharide-degrading bacterioplankton in an arctic fjord of Svalbard. *Appl Environ Microbiol* 80:3749–3756. <https://doi.org/10.1128/AEM.00899-14>.
- Solden L, Lloyd K, Wrighton K. 2016. The bright side of microbial dark matter: lessons learned from the uncultivated majority. *Curr Opin Microbiol* 31:217–226. <https://doi.org/10.1016/j.mib.2016.04.020>.
- Pesant S, Not F, Picheral M, Kandels-Lewis S, Le Bescot N, Gorsky G, Iudicone D, Karsenti E, Speich S, Troublé R, Dimier C, Searson S, Acinas SG, Bork P, Boss E, Bowler C, De Vargas C, Follows M, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Jaillon O, Kandels-Lewis S, Karp-Boss L, Karsenti E, Krzic U, Not F, Ogata H, Pesant S, Raes J, Reynaud EG, Sardet C, Sieracki M, Speich S, Stemmann L, Sullivan MB, Sunagawa S, Velayoudon D, Weissenbach J, Wincker P. 2015. Open science resources for the discovery and analysis of Tara Oceans data. *Sci Data* 2:150023. <https://doi.org/10.1038/sdata.2015.23>.
- Martijn J, Vosseberg J, Guy L, Offre P, Ettema TJG. 2018. Deep mitochondrial origin outside the sampled alphaproteobacteria. *Nature* 557:101–105. <https://doi.org/10.1038/s41586-018-0059-5>.
- 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>.
- Nurk S, Meleshko D, Korobeynikov A, Pevzner P. 2016. metaSPAdes: a new versatile metagenomics assembler. *Genome Res.* 27:824–834. <https://doi.org/10.1101/gr.213959.116>.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, Darling A, Malfatti S, Swan BK, Gies EA, Dodsworth JA, Hedlund BP, Tsiamis G, Sievert SM, Liu W-T, Eisen JA, Hallam SJ, Kyrpides NC, Stephanoukas R, Rubin EM, Hugenholtz P, Woyke T. 2013. Insights into the phylogeny and coding potential of microbial dark matter. *Nature* 499:431–437. <https://doi.org/10.1038/nature12352>.
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
- 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>.
- Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* 30:772–780. <https://doi.org/10.1093/molbev/mst010>.
- Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 32:268–274. <https://doi.org/10.1093/molbev/msu300>.