



Draft Whole-Genome Sequence of *Triparma laevis* f. *inornata* (Parmales, Bolidophyceae), Isolated from the Oyashio Region, Western North Pacific Ocean

 Akira Kuwata,^a Kenji Saitoh,^a Yoji Nakamura,^b Mutsuo Ichinomiya,^c  Naoki Sato^d

^aTohoku National Fisheries Research Institute, Fisheries Research and Education Agency, Shiogama, Miyagi, Japan

^bNational Research Institute of Fisheries Science, Fisheries Research and Education Agency, Yokohama, Kanagawa, Japan

^cPrefectural University of Kumamoto, Kumamoto, Kumamoto, Japan

^dUniversity of Tokyo, Tokyo, Japan

Author order was determined on the basis of affiliation and seniority.

ABSTRACT We present the first draft whole-genome sequence for the Parmales (Bolidophyceae, Heterokonta), a picoplanktonic sister group of diatoms, using a *Triparma laevis* f. *inornata* strain that was isolated from the Oyashio region in the western North Pacific Ocean.

The Parmales (Bolidophyceae) are a group of pico-sized marine phytoplankton with a silicified cell wall that are distributed widely in the world's oceans, from polar to subtropical regions (1). Analysis of small-subunit ribosomal DNA and *rbcL* genes and the organellar genomes of *Triparma laevis* revealed that the Parmales are members of the Heterokonta and share a common ancestor with diatoms, which are the most successful phytoplankton in modern oceans (2, 3). The origin of the silica cell wall and the early evolution of diatoms are still unclear, and genomic analysis of Parmales may help address these issues. We present the first draft whole-genome sequence for the Parmales, that of a *Triparma laevis* f. *inornata* strain (NIES-2565) that was isolated from the Oyashio region of the western North Pacific Ocean.

The algal strain was cultured in f/2 medium (4) at 5°C under ~30 $\mu\text{mol photons m}^{-2} \text{ s}^{-1}$ (14-h light/10-h dark photoperiod) (2). Cells in the exponential growth phase, ~3 days after inoculation, were collected on polycarbonate filters (0.6- μm pore size; Millipore). DNA was extracted by treatment with sodium *N*-dodecanoylsarcosinate and proteinase K, extracted with phenol-chloroform, and purified by CsCl ultracentrifugation (3).

Preparation of sequence templates for 454 FLX+ (shotgun and paired-end) sequencing and Illumina GAll (paired-end) sequencing followed the manufacturers' instructions using the Roche 454 GS FLX Titanium rapid library preparation kit and the Illumina TruSeq DNA PCR-free sample preparation kit, respectively. To obtain short (~800-bp) fragments for 454 shotgun and Illumina paired-end sequencing, an S220 ultrasonicator (Covaris, Woburn, MA, USA) was used to shear genomic DNA. The HydroShear device (Digilab, Marlborough, MA, USA) was used to obtain longer (~3-kb) fragments for 454 paired-end sequencing. We first assembled 454 genomic shotgun reads (~2.75 million reads) and paired-end reads (~2.27 million pairs of reads) (31.1 \times coverage overall) with the Newbler assembler (version 2.9; Roche Diagnostics). Newbler excluded about 15% of reads by filtering low-quality or highly repetitive reads. The paired-end insert length given by Newbler (~1.6 kb) may be underestimated because of skew to shorter inserts with both reads mapped. For Illumina reads, low-quality end nucleotides were trimmed using a Perl script (34.4 \times coverage subsequently) before

Citation Kuwata A, Saitoh K, Nakamura Y, Ichinomiya M, Sato N. 2020. Draft whole-genome sequence of *Triparma laevis* f. *inornata* (Parmales, Bolidophyceae), isolated from the Oyashio region, western North Pacific Ocean. Microbiol Resour Announc 9:e00367-20. <https://doi.org/10.1128/MRA.00367-20>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2020 Kuwata 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 Akira Kuwata, akuwata@affrc.go.jp.

Received 17 April 2020

Accepted 14 July 2020

Published 13 August 2020

further analysis. Because 454 pyrosequencing reads inherently have frequent miscalls of the number of mononucleotide tracts (5), length disagreements were overridden by Illumina reads upon $>8\times$ coverage (1,602 sites) using an available Perl script (<http://cse.fra.affrc.go.jp/ksaitoh/script/OverrideSeqs.pl>). These length disagreements were detected by mapping Illumina paired-end reads (2×71 to 73 bp, with insert lengths of ~ 800 bp) onto the 454 scaffolds by Bowtie 2 (version 2.1.0) (6). The improved 454 scaffolds and contigs were then bridged by Illumina paired-end reads with SSPACE-basic (version 2.0) (7) with the options `-k 5 -a 0.7 -x 1 -m 50 -o 10 -t 12 -r 0.8`. Gaps were filled with the Illumina reads by GapFiller (version 1.10) (8) with the options `-m 30 -o 2 -r 0.7 -n 10 -d 50 -t 10 -g 0 -T 4 -i 1`.

As a result, 13,538 contigs (N_{50} , 15,437 bp) with 62,124,971 bp in total were obtained. Bridging these contigs yielded 2,930 scaffolds (≥ 500 bp) totaling 62,842,229 bp in length (N_{50} , 89,610 bp), but some of those were derived from the plastid and mitochondrion of *T. laevis* (3) or bacteria in the culture. To clean these, the scaffolds were further classified based on 4-bp nucleotide frequency using principal-component analysis and t-SNE (9, 10). Finally, a total of 902 scaffolds were collected as those of the *T. laevis* nuclear genome, totaling 42,573,571 bp with an N_{50} value of 83,218 bp and a G+C content of 51.3%. The genome sequence quality was assessed by Benchmarking Universal Single-Copy Orthologs (BUSCO) (version 4.0.5) (11) with a stramenopile set of 100 single-copy orthologs (stramenopiles_odb10), and 83% of those orthologs were completely captured.

Data availability. The genome sequence of *T. laevis* has been deposited at the DDBJ under the accession numbers BLQM01000001 to BLQM01000902. Raw data are available in the DDBJ Read Archive with the accession numbers DRR213786 to DRR213791, DRR213793 to DRR213797, and DRR214684 to DRR214685.

ACKNOWLEDGMENTS

Financial support for this work was provided by Grants-in-Aid for Scientific Research 22657027, 23370046, 26291085, and 17H03724 from the Japan Society for the Promotion of Science and the Canon Foundation.

We thank D. Jewson for editing the manuscript.

REFERENCES

1. Kuwata A, Yamada K, Ichinomiya M, Yoshikawa S, Tragin M, Vault D, Lopes dos Santos A. 2018. Bolidophyceae, a sister picoplanktonic group of diatoms: a review. *Front Mar Sci* 5:370. <https://doi.org/10.3389/fmars.2018.00370>.
2. Ichinomiya M, Yoshikawa S, Kamiya M, Ohki K, Takaichi S, Kuwata A. 2011. Isolation and characterization of Parmales (Heterokonta/Heterokontophyta/Stramenopiles) from the Oyashio region, western North Pacific. *J Phycol* 47:144–151. <https://doi.org/10.1111/j.1529-8817.2010.00926.x>.
3. Tajima N, Saitoh K, Sato S, Maruyama F, Ichinomiya M, Yoshikawa S, Kurokawa K, Ohta H, Tabata S, Kuwata A, Sato N. 2016. Sequencing and analysis of the complete organellar genomes of Parmales, a closely related group to Bacillariophyta (diatoms). *Curr Genet* 62:887–896. <https://doi.org/10.1007/s00294-016-0598-y>.
4. Guillard RRL, Ryther JH. 1962. Studies of marine planktonic diatoms. I. *Cyclotella nana* Hustedt and *Detonula confervacea* (Cleve) Gran. *Can J Microbiol* 8:229–239. <https://doi.org/10.1139/m62-029>.
5. Robasky K, Lewis NE, Church GM. 2014. The role of replicates for error mitigation in next-generation sequencing. *Nat Rev Genet* 15:56–62. <https://doi.org/10.1038/nrg3655>.
6. Langmead B, Salzberg S. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
7. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27:578–579. <https://doi.org/10.1093/bioinformatics/btq683>.
8. Boetzer M, Pirovano W. 2012. Towards (almost) closed genomes with GapFiller. *Genome Biol* 13:R56. <https://doi.org/10.1186/gb-2012-13-6-r56>.
9. van der Maaten L, Hinton G. 2008. Visualizing data using t-SNE. *J Mach Learn Res* 9:85.
10. Nakamura Y. 2019. Horizontal gene transfer in marine environment: a technical perspective on metagenomics. *In* Gojobori T, Wada T, Kobayashi T, Mineta K (ed). *Marine metagenomics*, p 65–84. Springer Nature, Singapore.
11. Seppey M, Manni M, Zdobnov EM. 2019. BUSCO: assessing genome assembly and annotation completeness. *Methods Mol Biol* 1962: 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.