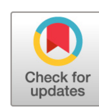
GENOME SEQUENCES



Genome Sequence of the Edible Green Alga *Ulva prolifera*, Originating from the Yoshinogawa River in Japan

Keita Tamura,[a,b] Hidemasa Bono[a,b]

[a]Laboratory of Genome Informatics, Graduate School of Integrated Sciences for Life, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan
[b]Laboratory of BioDX, Genome Editing Innovation Center, Hiroshima University, Higashi-Hiroshima, Hiroshima, Japan

ABSTRACT We report the genome sequence of *Ulva prolifera*, which originated from the Yoshinogawa River in Japan, using Oxford Nanopore Technologies MinION and Illumina sequencing reads. The genome assembly size is 103.8 Mbp, consisting of 142 scaffolds with an N_{50} value of 4.11 Mbp.

Ulva prolifera (Sujiaonori in Japanese) is one of the most important edible green alga in Japan. It grows in brackish river mouths; however, natural populations of *U. prolifera* in Japan have been decreasing, and a tank cultivation system has been developed for sustainable production (1). *U. prolifera* is also one of the sources of troublesome green tide (2). The genome sequence of *U. prolifera* collected from the Yellow Sea was recently reported (3); however, it is known that brackish strains and bloom-forming marine habitat strains show different characteristics, and new subspecies for the latter strains have been proposed (4, 5).

An *U. prolifera* sample for genome sequencing was collected from a land culture at Hashirijima Island (Hiroshima, Japan) that was originally derived from the Yoshinogawa River (Tokushima, Japan). The thallus, with a length of >20 cm, was cut into small pieces, frozen in liquid nitrogen, and finely powdered using a Micro Smash MS-100 cell disruptor (Tomy Seiko). Genomic DNA (gDNA) was extracted using the NucleoBond high-molecular-weight (HMW) DNA kit (Macherey-Nagel), and we did not perform either DNA shearing or size selection. A Nanopore sequencing library was prepared from 1 μg of gDNA using a ligation sequencing kit (SQK-LSK110; Oxford Nanopore Technologies [ONT]). Nanopore sequencing was performed on an ONT MinION sequencer using a R9.4.1 flow cell (FLO-MIN106D). One-half volume of the library was sequenced for 24 h, and the other one-half volume of the library was sequenced for 22 h after a flow cell wash step. The raw FAST5 files obtained for the two sequencing runs were base called using Guppy v6.0.1 (ONT) with a high-accuracy model (Table 1). An Illumina sequencing library was prepared using the NEBNext Ultra DNA library preparation kit for Illumina (New England Biolabs) and sequenced on an Illumina NovaSeq 6000 system in the 150-bp paired-end mode (Table 1).

The base-called Nanopore reads were trimmed by fastp v0.23.2 (6) to remove low-quality head reads, and reads longer than 1,000 bp were collected (15,851 Mbp after trimming) for subsequent analysis. The Nanopore reads were assembled using NECAT v0.0.1_update20200803 (7) following a modified method (8). First the draft assembly by NECAT was polished with Medaka v1.6.0 (https://github.com/nanoporetech/medaka) using Nanopore reads, and then redundant haplotypes were removed with Purge Haplotigs v1.1.2 (9) using the Nanopore reads mapped with minimap2 v2.23 (10). Finally, a NECAT bridging step was performed using the primary contigs generated by Purge Haplotigs as a replacement for the draft assembly by NECAT. The bridged contigs from NECAT were polished three times with HyPo v1.0.3 (11) using the Illumina reads (after trimming with fastp v0.23.2) mapped with BWA-MEM v0.7.17 (12). Contigs that showed >90% identity and coverage against mitochondrial or chloroplast RefSeq sequences of *Ulva* species by BLASTn v2.12.0 were removed as organellar contigs. BESST v2.2.8 (13, 14) was used for scaffolding of the contigs using the Illumina

Editor Jason E. Stajich, University of California, Riverside
Copyright © 2022 Tamura and Bono. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.
Address correspondence to Keita Tamura, tamurak@hiroshima-u.ac.jp.
The authors declare no conflict of interest.
Received 23 May 2022
Accepted 16 August 2022
Published 29 August 2022

TABLE 1 Summary of the sequence read and genome assembly data

Parameter	Finding
Nanopore sequencing	
No. of reads	5,280,010
No. of bases	17,282,510,862
Avg read length (bp)	3,273.2
Read N_{50} (bp)	9,388
Illumina sequencing	
Total no. of reads	74,904,442
Total no. of bases	11,235,666,300
Genome assembly	
Total length (bp)	103,843,212
No. of scaffolds	142
Scaffold N_{50} (bp)	4,106,524
Scaffold L_{50}	8
Largest scaffold (bp)	9,543,602
No. of contigs	143
Contig N_{50} (bp)	4,106,524
Contig L_{50}	9
Largest contig (bp)	9,355,488
GC content (%)	55.57

reads mapped with BWA-MEM v0.7.17, and the final assembly was obtained (Table 1). Only two contigs were scaffolded with a 1-bp gap by BESST. The largest sequence (9,543,602 bp) (Table 1) was generated by this step.

Genome completeness analysis with BUSCO v5.2.2 (15) using the chlorophyta_odb10 data set (5 August 2020) and default settings showed 80.5% completeness (71.8% single-copy and 8.7% duplicated). Our BUSCO analysis of previously published *U. prolifera* (3) and *Ulva mutabilis* (16) genomes with the same settings showed relatively lower completeness scores (77.2% and 79.2%, respectively). Future studies will annotate genes to analyze this genome in more detail.

Data availability. This whole-genome shotgun project has been deposited in DDBJ under the accession numbers BRCE01000001 to BRCE01000142. The raw Nanopore and Illumina reads have been deposited in DDBJ under the accession numbers DRR361634 and DRR361635, respectively. Additional information on the parameter settings and draft assembly results with tested assemblers can be found at figshare (https://doi.org/10.6084/m9.figshare.19728862).

ACKNOWLEDGMENTS

We thank Mishima Foods Co., Ltd., for providing a sample of *U. prolifera* for genome sequencing.

This work was supported by the Center of Innovation for Bio-Digital Transformation (BioDX), a Japan Science and Technology Agency (JST) open innovation platform for industry-academia cocreation (COI-NEXT) (grant JPMJPF2010). Computations were partially performed on the supercomputer at the ROIS National Institute of Genetics.

REFERENCES

1. Hiraoka M, Oka N. 2008. Tank cultivation of *Ulva prolifera* in deep seawater using a new "germling cluster" method. J Appl Phycol 20:97–102. https://doi.org/10.1007/s10811-007-9186-3.
2. Cai C, Liu F, Jiang T, Wang L, Jia R, Zhou L, Gu K, Ren J, He P. 2018. Comparative study on mitogenomes of green tide algae. Genetica 146:529–540. https://doi.org/10.1007/s10709-018-0046-7.
3. He Y, Shen S, Yu D, Wang Y, Yin J, Wang Z, Ye Y. 2021. The *Ulva prolifera* genome reveals the mechanism of green tides. J Ocean Limnol 39:1458–1470. https://doi.org/10.1007/s00343-020-0212-5.
4. Hiraoka M. 2021. Massive *Ulva* green tides caused by inhibition of biomass allocation to sporulation. Plants 10:2482. https://doi.org/10.3390/plants10112482.
5. Cui J, Monotilla AP, Zhu W, Takano Y, Shimada S, Ichihara K, Matsui T, He P, Hiraoka M. 2018. Taxonomic reassessment of *Ulva prolifera* (Ulvophyceae, Chlorophyta) based on specimens from the type locality and Yellow Sea green tides. Phycologia 57:692–704. https://doi.org/10.2216/17-139.1.
6. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
7. Chen Y, Nie F, Xie S-Q, Zheng Y-F, Dai Q, Bray T, Wang Y-X, Xing J-F, Huang Z-J, Wang D-P, He L-J, Luo F, Wang J-X, Liu Y-Z, Xiao C-L. 2021. Efficient assembly of Nanopore reads via highly accurate and intact error correction. Nat Commun 12:60. https://doi.org/10.1038/s41467-020-20236-7.

8. Naito K. 2021. Nanopore de yomu shokubutsu genome, p 177–199. *In* Long read WET & DRY kaiseki guide, sequence wo motto jiyu ni! Repeat hairetsu kara kozo heni, direct RNA, de novo assembly made, kenkyu no kanosei wo gun to hirogeru oyo jizaina saishin gijutsu, doko demo, dare demo, kyo kara tsukaeru! (Saikyo no step up series). Yodosha Company, Ltd., Tokyo, Japan.
9. Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. BMC Bioinformatics 19:460. https://doi.org/10.1186/s12859-018-2485-7.
10. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
11. Kundu R, Casey J, Sung W-K. 2019. HyPo: super fast & accurate polisher for long read genome assemblies. bioRxiv. https://doi.org/10.1101/2019.12.19.882506.
12. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997. https://doi.org/10.48550/arXiv.1303.3997.
13. Sahlin K, Chikhi R, Arvestad L. 2016. Assembly scaffolding with PE-contaminated mate-pair libraries. Bioinformatics 32:1925–1932. https://doi.org/10.1093/bioinformatics/btw064.
14. Sahlin K, Vezzi F, Nystedt B, Lundeberg J, Arvestad L. 2014. BESST: efficient scaffolding of large fragmented assemblies. BMC Bioinformatics 15:281. https://doi.org/10.1186/1471-2105-15-281.
15. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. Mol Biol Evol 38:4647–4654. https://doi.org/10.1093/molbev/msab199.
16. De Clerck O, Kao S-M, Bogaert KA, Blomme J, Foflonker F, Kwantes M, Vancaester E, Vanderstraeten L, Aydogdu E, Boesger J, Califano G, Charrier B, Clewes R, Del Cortona A, D'Hondt S, Fernandez-Pozo N, Gachon CM, Hanikenne M, Lattermann L, Leliaert F, Liu X, Maggs CA, Popper ZA, Raven JA, Van Bel M, Wilhelmsson PKI, Bhattacharya D, Coates JC, Rensing SA, Van Der Straeten D, Vardi A, Sterck L, Vandepoele K, Van de Peer Y, Wichard T, Bothwell JH. 2018. Insights into the evolution of multicellularity from the sea lettuce genome. Curr Biol 28:2921–2933.e5. https://doi.org/10.1016/j.cub.2018.08.015.