



De Novo Genome Assembly of Stinkhorn Mushroom *Clathrus columnatus* (Basidiomycota, Fungi) Using Illumina and Nanopore Sequencing Data

✉ Eri Ogiso-Tanaka,^a Hiyori Itagaki,^b Muneyuki Ohmae,^c Tsuyoshi Hosoya,^d Kentaro Hosaka^{a,d}

^aCenter for Molecular Biodiversity Research, National Museum of Nature and Science, Tsukuba, Ibaraki, Japan

^bDepartment of Biological Science, Graduate School of Science, The University of Tokyo, Bunkyo-ku, Tokyo, Japan

^cHokken Co. Ltd., Mibu-machi, Tochigi, Japan

^dDepartment of Botany, National Museum of Nature and Science, Tsukuba, Ibaraki, Japan

ABSTRACT We report the reference genome of *Clathrus columnatus* isolate MO-923, which was isolated from Chichijima Island, the Ogasawara (Bonin) Islands, Japan. Oxford Nanopore Technologies MinION and Illumina sequence reads were assembled using NECAT and polished using Pilon to yield a 36.51-Mb genome with 10,625 predicted protein-coding genes.

The basidiomycete *Clathrus columnatus* is a stinkhorn fungus in the family Clathraceae (subclass Phallomycetidae, order Phallales) (1). The species was originally described from South Carolina, USA, but was later reported from Africa, Australasia, Central America, and South America (2). In Japan, it was first collected from Hahajima, the Ogasawara (Bonin) Islands (3), but was later reported from wider areas, including both Hahajima and Chichijima Islands (4, 5).

The material for this study was collected at Mount Chuo, Chichijima Island (27°04'28.9"N, 142°13'07.1"E), on 7 November 2018. In the laboratory, immature fresh fruit bodies were dissected in half using a clean, sterilized razor blade, and a small piece of internal tissue was placed on potato dextrose agar (PDA) (Nissui, Japan) to obtain an isolate. After confirmation of mycelial growth without obvious contamination, the isolate was transferred to PDA slants and was maintained as isolate MO-923 until further experiments. The remaining material was photographed and dried for 24 h with low heat (~45°C) and good air circulation using a food dehydrator. A dried voucher specimen was deposited at the fungal herbarium of the National Museum of Nature and Science (TNS) under registration number TNS-F-82131 for future studies.

Prior to DNA extraction, the isolate of *Clathrus columnatus* was transferred again to PDA plates prepared in 9-cm petri dishes. Ten inoculated PDA plates were incubated at 20°C in the dark. After 2 weeks, fragments of mycelial colonies were harvested by peeling them from the agar using sterilized tweezers. DNA was extracted by the method described at protocols.io (6).

The library preparation for the Illumina platform was performed using the NEBNext Ultra II FS DNA library preparation kit for Illumina (New England Biolabs, USA) following a modified manufacturer's protocol (7). The library was then sequenced on an Illumina MiSeq instrument to obtain 300-bp paired-end reads using a 600-cycle v3 kit (Illumina, USA) (Table 1).

The genomic DNA (gDNA) (1 μL) was diluted in EB buffer to a final volume 100 μL and sheared in g-TUBEs (Covaris, USA) with two passages at 4,000 × g or 1,300 × g for 1 min in a centrifuge (TOMY, Japan). The library preparation was performed using the ligation sequencing kit SQK-LSK109 (Oxford Nanopore Technologies, United Kingdom) following the manufacturer's protocol. The library was then sequenced on a MinION system using

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2022 Ogiso-Tanaka 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 Eri Ogiso-Tanaka, ogiso@kahaku.go.jp, or Kentaro Hosaka, khosaka@kahaku.go.jp.

The authors declare no conflict of interest.

Received 4 November 2021

Accepted 19 January 2022

Published 10 February 2022

TABLE 1 Genome assembly, Nanopore sequencing, and Illumina sequencing data

Parameter	Finding
Sample	<i>Clathrus columnatus</i> isolate MO-923
No. of contigs	19
Total length (bp)	36,509,728
N_{50} (bp)	3,174,166
GC content (%)	41.3
Coverage (×)	164
Nanopore sequencing	
No. of reads	1,108,254
No. of bases	8,298,334,373
Avg length (bp)	7,488
Illumina sequencing	
Total no. of reads	4,585,114
Total size (Mb)	2,348

R9.4.1 flow cell sequencing chemistry with MinKNOW v4.2.5 software. Guppy GPU v4.5.2+bcc53d3 (Oxford Nanopore Technologies) was used for base calling (Table 1). For the long-read genome assembly, Filtlong v0.2.0 (<https://github.com/rwick/Filtlong>) was used to extract the top 278,079 longer reads (total of 5 Gbp) and high-quality reads from the 1.1 million reads obtained (8.30 Gbp). The filtered Nanopore reads were assembled using NECAT (8). For Illumina short reads, quality control was performed with Trimmomatic v0.30 (9). Then, the processed reads were used for polishing of the sequences assembled from Nanopore reads. Six rounds of assembly polishing were performed with Pilon v1.13 (10) using Illumina reads aligned with BWA v0.7.9a (11) for short-read mapping. As a result, we obtained a complete genome assembly with 19 contigs containing 36,509,728 bp (GC content, 41.27%). The N_{50} value for the contigs was 3,174,166 bp, and the Benchmarking Universal Single-Copy Orthologs (BUSCO) completeness value was 92.07% (BUSCO v3 fungi_odb9 [12] and QAST v5.0.2 [13]). Annotation of the assembled genome was carried out using the Funannotate pipeline v1.5.2 (14), which included masking repetitive elements with the RepeatMasker tool (<https://www.repeatmasker.org>) and annotating genes with AUGUSTUS (15), GlimmerHMM (16), SNAP (17), and EVM (18). Gene function was predicted using the Pfam (19), MEROPS (20), CAZy (21), InterProScan (22), and Swiss-Prot (23) databases. A total of 10,625 protein-coding genes were predicted.

Data availability. The annotated complete genome assembly of *Clathrus columnatus* isolate MO-923 is available in GenBank and DDBJ under the accession numbers [GCA_020884735.1](https://ncbi.nlm.nih.gov/GenBank/record/GCA_020884735.1) and [BPWL01000001](https://ncbi.nlm.nih.gov/GenBank/record/BPWL01000001) to [BPWL01000019](https://ncbi.nlm.nih.gov/GenBank/record/BPWL01000019). A table with all Web addresses was uploaded at figshare (https://figshare.com/articles/online_resource/Data_availability_Clathrus_columnatus_genome/17246096). The Illumina MiSeq and Oxford Nanopore Technologies sequencing raw data were deposited in DDBJ under the accession numbers [DRA012831](https://ncbi.nlm.nih.gov/BioProject/record/DRA012831) and [DRA012832](https://ncbi.nlm.nih.gov/BioProject/record/DRA012832) (BioProject accession number [PRJDB12304](https://ncbi.nlm.nih.gov/BioProject/record/PRJDB12304)). The voucher specimen (dried specimen) from which the isolate was obtained is housed at the fungal herbarium of the Department of Botany, TNS, under the registration number TNS-F-82131.

ACKNOWLEDGMENT

This work was supported by the Director's Support Fund of the TNS, Japan.

REFERENCES

- Hosaka K, Bates ST, Beever RE, Castellano MA, Colgan W, Domínguez LS, Nouhra ER, Geml J, Giachini AJ, Kenney SR, Simpson NB, Spatafora JW, Trappe JM. 2006. Molecular phylogenetics of the gomphoid-phalloid fungi with an establishment of the new subclass Phallomycetidae and two new orders. *Mycologia* 98:949–959. <https://doi.org/10.1080/15572536.2006.11832624>.
- Dring DM. 1980. Contributions towards a rational arrangement of the *Clathraceae*. *Kew Bull* 35:1–96. <https://doi.org/10.2307/4117008>.
- Ito S, Imai S. 1937. Fungi of the Bonin Islands I. *Trans Sapporo Nat Hist Soc* 40:1–12.
- Hongo T. 1978. Higher fungi of the Bonin Islands II. *Rep Tottori Mycol Inst* 16: 59–65.
- Hosaka K. 2018. Distribution data of some mushroom species distributed in and around the Ogasawara Islands. *Jpn Mem Natl Mus Nat Sci Tokyo* 52:17–37.
- Ogiso-Tanaka E, Itagaki H, Ohmae M, Hosoya T, Hosaka K. 2021. Fungal DNA extraction for Nanopore sequencing. *protocolsio* <https://doi.org/10.17504/protocols.io.b2vfqe3n>.
- Li H, Wu K, Ruan C, Pan J, Wang Y, Long H. 2019. Cost-reduction strategies in massive genomics experiments. *Mar Life Sci Technol* 1:15–21. <https://doi.org/10.1007/s42995-019-00013-2>.

8. Chen Y, Nie F, Xie SQ, Zheng YF, Dai Q, Bray T, Wang YX, Xing JF, Huang ZJ, Wang DP, He LJ, Luo F, Wang JX, Liu YZ, Xiao CL. 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>.
9. 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>.
10. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
11. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997.
12. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
13. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
14. Palmer JM, Stajich J. 2020. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo <https://doi.org/10.5281/zenodo.1134477>.
15. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* 34:W435–W439. <https://doi.org/10.1093/nar/gkl200>.
16. Majoros WH, Pertea M, Salzberg SL. 2004. TigrScan and GlimmerHMM: two open-source ab initio eukaryotic gene-finders. *Bioinformatics* 20:2878–2879. <https://doi.org/10.1093/bioinformatics/bth315>.
17. Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
18. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. 2008. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol* 9:R7. <https://doi.org/10.1186/gb-2008-9-1-r7>.
19. Mistry J, Chuguransky S, Williams L, Qureshi M, Salazar GA, Sonnhammer ELL, Tosatto SCE, Paladin L, Raj S, Richardson LJ, Finn RD, Bateman A. 2021. Pfam: the protein families database in 2021. *Nucleic Acids Res* 49: D412–D419. <https://doi.org/10.1093/nar/gkaa913>.
20. Rawlings ND, Barrett AJ, Bateman A. 2010. MEROPS: the peptidase database. *Nucleic Acids Res* 38:D227–D233. <https://doi.org/10.1093/nar/gkp971>.
21. Terrapon N, Lombard V, Drula E, Coutinho PM, Henrissat B. 2017. The CAZy database/the Carbohydrate-Active Enzyme (CAZy) database: principles and usage guidelines, p 117–131. *In* Aoki-Kinoshita K (eds), *A practical guide to using glycomics databases*. Springer, Tokyo, Japan. https://doi.org/10.1007/978-4-431-56454-6_6.
22. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* 30:1236–1240. <https://doi.org/10.1093/bioinformatics/btu031>.
23. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Methods Mol Biol* 1374:23–54. https://doi.org/10.1007/978-1-4939-3167-5_2.