AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Chromosome-Level Genome Sequence of *Aspergillus chevalieri* M1, Isolated from Katsuobushi

Chihiro Kadooka,[a,b]* Kazuki Mori,[c] Kayu Okutsu,[a] Yumiko Yoshizaki,[a,b] Kazunori Takamine,[a,b] Kosuke Tashiro,[d] Hisanori Tamaki,[a,b] Taiki Futagami[a,b]

[a]Education and Research Centre for Fermentation Studies, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan
[b]United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Japan
[c]Cell Innovator Co., Ltd., Fukuoka, Japan
[d]Laboratory of Molecular Gene Technology, Faculty of Agriculture, Kyushu University, Fukuoka, Japan

ABSTRACT In this study, we report the chromosome-level genome sequence of the osmophilic filamentous fungus *Aspergillus chevalieri* M1, which was isolated from a dried bonito, katsuobushi. This fungus plays a significant role in the fermentation and ripening process. Thus, elucidating the sequence data for this fungus will aid in subsequent genomic research on the fungi involved in katsuobushi production.

Naturally occurring fungi have been used in the traditional production of dried bonito, katsuobushi, in Japan. The fungi play a significant role in the degradation of proteins and lipids, as well as the formation of flavor during the production of katsuobushi. In a previous study, we analyzed the fungal flora transition during katsuobushi production and isolated *Aspergillus chevalieri* (1). *A. chevalieri* isolates were further classified into teleomorphic and anamorphic strains, based on morphological analysis. The teleomorphic strain was dominant throughout the katsuobushi production (1). *A. chevalieri* was also identified in different katsuobushi samples in Japan (2). In addition, *A. chevalieri* was isolated from fermented foods such as bagoong, a traditional fermented fish prepared with salted anchovy from Luzon Island in the Philippines (3), and meju, a brick of dried fermented soybeans in Korea (4).

To understand the fungi used for the production of katsuobushi, we sequenced the genome of the teleomorphic *A. chevalieri* strain M1 (1). Strain M1 was cultivated in M40Y medium (40% [wt/vol] sucrose, 2% [wt/vol] malt extract, 0.5% [wt/vol] yeast extract, and 2% [wt/vol] agar) (DSMZ medium 187). Then, the mycelia retrieved were subjected to DNA extraction using DNAs-ici!-F (Rizo Inc., Tsukuba, Japan) and RNA extraction using RNAiso Plus (TaKaRa Bio Inc., Shiga, Japan) and the SV total RNA isolation system (Promega, Madison, WI). The genomic DNA of strain M1 was sequenced to 178-fold and 179-fold coverage using Oxford Nanopore Technologies (ONT) MinION and Illumina NovaSeq 6000 platforms, respectively, with PCR-free workflows. ONT and Illumina sequencing libraries were prepared using the ligation 1D kit (SQK-LSK109; ONT) and the NEBNext Ultra II DNA library preparation kit (E7645; New England BioLabs [NEB]), respectively. The ONT reads and Illumina reads were used for *de novo* assembly and error correction, respectively. The ONT reads were assembled by Canu v2.0 (5), Flye v2.8-b1674 (6), and Raven v1.5.0 (7). The Flye and Raven assemblies were used to bridge separate contigs generated by Canu. We choose the better assembly metrics based on telomere-to-telomere genome assembly. Consequently, the genome of strain M1 was assembled into 9 contigs, consisting of 8 chromosomes and 1 mitochondrial DNA, thus indicating that we successfully sequenced the nearly complete genome sequences of strain M1. Five chromosomes were generated only by Canu, while chromosomes 3, 7, and 8 were generated by an assembly in which 2 Canu contigs

Citation Kadooka C, Mori K, Okutsu K, Yoshizaki Y, Takamine K, Tashiro K, Tamaki H, Futagami T. 2021. Chromosome-level genome sequence of *Aspergillus chevalieri* M1, isolated from katsuobushi. Microbiol Resour Announc 10:e00385-21. https://doi.org/10.1128/MRA.00385-21.

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Taiki Futagami, futagami@agri.kagoshima-u.ac.jp.

* Present address: Chihiro Kadooka, Department of Applied Microbial Technology, Faculty of Biotechnology and Life Science, Sojo University, Kumamoto, Japan.

Received 14 April 2021
Accepted 14 August 2021
Published 16 September 2021

TABLE 1 Replicons of *Aspergillus chevalieri* M1

Location[a]	GenBank accession no.	Size (Mb)	GC content (%)	No. of coding sequences	No. of rRNAs	No. of tRNAs
Chr. 1	AP024416.1	5.59	49.2	1,944	0	28
Chr. 2	AP024417.1	4.31	49.2	1,471	0	36
Chr. 3	AP024418.1	4.1	49.3	1,423	0	21
Chr. 4	AP024419.1	3.79	49.4	1,366	0	18
Chr. 5	AP024420.1	3.72	49.4	1,312	0	24
Chr. 6	AP024421.1	2.96	49.2	1,043	0	11
Chr. 7	AP024422.1	2.8	49.3	943	27 (82)[b]	11
Chr. 8	AP024423.1	2.43	49.1	839	0	20
MT	AP024424.1	0.05	28.1	15	1	27

[a] Chr., chromosome; MT, mitochondria.
[b] The number of rRNA genes is not clear due to their highly repetitive structure. The number in parentheses indicates the estimated copy number based on the median per-base coverage.

were bridged by Flye and Raven contigs. The final assembly was polished using medaka v1.0.3 (8) with ONT reads, Pilon v1.23 (9) with ONT reads, and Pilon v1.23 (9) with Illumina reads. The genome annotation of the chromosomal contigs and mitochondrial contig obtained was performed based on the Funannotate v1.8.1 pipeline (10) and MFannot v1.1 (11), respectively. Gene prediction was performed by using SNAP v2006-07-28 (12), AUGUSTUS v3.3.3 (13), GlimmerHMM v3.0.4 (14), and GeneMark-ES v4.61_lic (15) via the Funannotate v1.8.1 pipeline (10). For the analysis, RNA-seq reads for strain M1 were obtained with the NovaSeq 6000 system and used for gene prediction. The sequencing library was prepared using the NEBNext Ultra directional RNA library preparation kit for Illumina (E7420; NEB). The Illumina reads were *de novo* assembled by Trinity v2.8.5 (16) and analyzed with the sequence alignment tool HISAT v2.2.0 (17). The proteins were annotated with MEROPS v12.0 (18), UniProt v2020_05 (19), MIBiG v1.4 (20), Pfam v33.1 (21), and dbCAN2 v9.0 (22) (based on the CAZy database v7/30/2020 [23]) using sequence alignment tools such as DIAMOND v2.0.6 (24) and HMMER v3.3.2 (25). Annotation was also performed using InterProScan v5.47-82.0 (26), eggNOG-mapper v1.0.3 (27) (for the EggNOG v4.5 database [28]), antiSMASH v5.1.2 (29), SignalP v4.1 (30), Phobius v1.01 (31), tRNAscan-SE v2.0.7 (32), and Barrnap v0.9 (33). The nearly complete genome of strain M1 includes 29,748,498 bp, with a GC content of 49.2%, and is composed of 10,356 predicted coding sequences and 196 tRNAs. Genome completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.2 with the ascomycota_odb10 data set (34), which resulted in 97.9% complete and single-copy BUSCOs, 0.4% complete and duplicate-copy BUSCOs, 0.5% fragmented-copy BUSCOs, and 1.2% missing BUSCOs. The details regarding the replicons present are summarized in Table 1. The chromosome-level genome sequence data for strain M1 will allow detailed genomic characterization to determine the possible applications and importance of *A. chevalieri* in fermentation industries.

Data availability. The nucleotide sequences of *A. chevalieri* M1 chromosomes and mitochondria have been deposited in DDBJ/ENA/GenBank under accession numbers AP024416 to AP024424. Raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession numbers DRX256206, DRX251720, and DRX251721.

ACKNOWLEDGMENTS

This work was funded by the Sankei Chemical Foundation and the Towa Foundation for Food Science and Research. C.K. was supported by a Grant-in-Aid for JSPS Research Fellows (grant 17J02753).

REFERENCES

1. Kadooka C, Nakamura E, Kubo S, Okutsu K, Yoshizaki Y, Takamine K, Tamaki H, Futagami T. 2020. Analysis of the fungal population involved in katsuobushi production. J Gen Appl Microbiol 66:239–243. https://doi.org/10.2323/jgam.2019.09.003.
2. Takenaka S, Nakabayashi R, Ogawa C, Kimura Y, Yokota S, Doi M. 2020. Characterization of surface *Aspergillus* community involved in traditional fermentation and ripening of katsuobushi. Int J Food Microbiol 327:108654. https://doi.org/10.1016/j.ijfoodmicro.2020.108654.
3. Murao M, Fujita T, Yamagata K. 1982. Identification of the species of genus *Aspergillus* isolated from *Bagoong* (fermented fish). Mem Fac Agr Kinki Univ 15:9–32. (In Japanese.)
4. Hong SB, Kim DH, Lee M, Baek SY, Kwon SW, Samson RA. 2011. Taxonomy

of *Eurotium* species isolated from meju. J Microbiol 49:669–674. https://doi.org/10.1007/s12275-011-0376-y.
5. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res 27:722–736. https://doi.org/10.1101/gr.215087.116.
6. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol 37:540–546. https://doi.org/10.1038/s41587-019-0072-8.
7. Vaser R, Šikić M. 2021. Time- and memory-efficient genome assembly with Raven. Nat Comput Sci 1:332–336. https://doi.org/10.1038/s43588-021-00073-4.
8. Oxford Nanopore Technologies. 2018. Medaka. https://nanoporetech.github.io/medaka.
9. 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.
10. Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo https://doi.org/10.5281/zenodo.4054262.
11. Beck N, Lang BF. 2010. MFannot. http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl.
12. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.
13. Stanke M, Steinkamp R, Waack S, Morgenstern B. 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Res 32:W309–W312. https://doi.org/10.1093/nar/gkh379.
14. 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.
15. Lukashin AV, Borodovsky M. 1998. GeneMark.hmm: new solutions for gene finding. Nucleic Acids Res 26:1107–1115. https://doi.org/10.1093/nar/26.4.1107.
16. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.
17. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol 37:907–915. https://doi.org/10.1038/s41587-019-0201-4.
18. Rawlings ND, Barrett AJ, Finn R. 2016. Twenty years of the MEROPS database of proteolytic enzymes, their substrates and inhibitors. Nucleic Acids Res 44:D343–D350. https://doi.org/10.1093/nar/gkv1118.
19. Wu CH, Apweiler R, Bairoch A, Natale DA, Barker WC, Boeckmann B, Ferro S, Gasteiger E, Huang H, Lopez R, Magrane M, Martin MJ, Mazumder R, O'Donovan C, Redaschi N, Suzek B. 2006. The Universal Protein Resource (UniProt): an expanding universe of protein information. Nucleic Acids Res 34:D187–D191. https://doi.org/10.1093/nar/gkj161.
20. Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hooft JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Linington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. Nucleic Acids Res 48:D454–D458. https://doi.org/10.1093/nar/gkz882.
21. El-Gebali S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, Qureshi M, Richardson LJ, Salazar GA, Smart A, Sonnhammer ELL, Hirsh L, Paladin L, Piovesan D, Tosatto SCE, Finn RD. 2019. The Pfam protein families database in 2019. Nucleic Acids Res 47:D427–D432. https://doi.org/10.1093/nar/gky995.
22. Zhang H, Yohe T, Huang L, Entwistle S, Wu P, Yang Z, Busk PK, Xu Y, Yin Y. 2018. dbCAN2: a meta server for automated carbohydrate-active enzyme annotation. Nucleic Acids Res 46:W95–W101. https://doi.org/10.1093/nar/gky418.
23. Lombard V, Golaconda Ramulu H, Drula E, Coutinho PM, Henrissat B. 2014. The Carbohydrate-Active enZymes database (CAZy) in 2013. Nucleic Acids Res 42:D490–D495. https://doi.org/10.1093/nar/gkt1178.
24. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. Nat Methods 12:59–60. https://doi.org/10.1038/nmeth.3176.
25. Potter SC, Luciani A, Eddy SR, Park Y, Lopez R, Finn RD. 2018. HMMER Web server: 2018 update. Nucleic Acids Res 46:W200–W204. https://doi.org/10.1093/nar/gky448.
26. Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong S-Y, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031.
27. Huerta-Cepas J, Forslund K, Coelho LP, Szklarczyk D, Jensen LJ, von Mering C, Bork P. 2017. Fast genome-wide functional annotation through orthology assignment by eggNOG-Mapper. Mol Biol Evol 34:2115–2122. https://doi.org/10.1093/molbev/msx148.
28. Ferrés I, Iraola G. 2018. Phylen: automatic phylogenetic reconstruction using the EggNOG database. J Open Sci Softw 3:593. https://doi.org/10.21105/joss.00593.
29. 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.
30. Almagro Armenteros JJ, Tsirigos KD, Sonderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. 2019. SignalP 5.0 improves signal peptide predictions using deep neural networks. Nat Biotechnol 37:420–423. https://doi.org/10.1038/s41587-019-0036-z.
31. Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined transmembrane topology and signal peptide prediction: the Phobius web server. Nucleic Acids Res 35:W429–W432. https://doi.org/10.1093/nar/gkm256.
32. Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. Methods Mol Biol 1962:1–14. https://doi.org/10.1007/978-1-4939-9173-0_1.
33. Seemann T. 2018. Barrnap. https://github.com/tseemann/barrnap.
34. 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.