Chromosome-Level Genome Sequence of the Black Koji Fungus *Aspergillus luchuensis* RIB2601

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

[a]Cell Innovator Co., Ltd., Fukuoka, Japan
[b]Education and Research Centre for Fermentation Studies, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan
[c]United Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Japan
[d]Research Support Center, Institute for Research Promotion, Kagoshima University, Kagoshima, Japan
[e]Laboratory of Molecular Gene Technology, Faculty of Agriculture, Kyushu University, Nishi-ku, Fukuoka, Japan
[f]Department of Applied Biochemistry and Food Science, Faculty of Agriculture, Saga University, Saga, Japan

ABSTRACT *Aspergillus luchuensis* is used for the production of awamori and shochu, which are traditional Japanese distilled alcoholic beverages. Here, we determined the chromosome-level genome sequence of *A. luchuensis* RIB2601.

The black koji fungus *Aspergillus luchuensis* is used for the production of awamori and shochu, which are traditional distilled alcoholic beverages indigenous to Japan (1–3). *A. luchuensis* was originally used to produce awamori in Okinawa, Japan's southernmost prefecture; then, it was used to produce shochu (4). The black koji fungus plays an important role in supplying glycoside hydrolases for decomposing starch contained in the ingredients of awamori and shochu during the fermentation process (1–4). In addition, it excretes a large amount of citric acid, which can prevent the growth of contaminating microbes during fermentation (1–4).

Currently, the genome sequence of *A. luchuensis* NBRC 4314 (RIB2604) is available (5); however, there are a variety of strains of the black koji fungus (1–3). We previously studied an amylolytic enzyme-overproducing mutant of *A. luchuensis* RIB2601 (6); therefore, we sequenced the genome of RIB2601. Strain RIB2601 was cultivated in yeast extract-peptone-dextrose medium (2% [wt/vol] glucose, 1% [wt/vol] yeast extract, and 2% [wt/vol] peptone) overnight. Then, the retrieved mycelia were subjected to DNA extraction using DNAs-ici!-F (Rizo, Inc., Tsukuba, Japan). The genomic DNA of RIB2601 was sequenced with coverages of 37- and 336-fold using Oxford Nanopore Technologies (ONT) MinION and Illumina HiSeq 2000 instruments, respectively. ONT and Illumina sequencing libraries were prepared using the ligation 1D (SQK-LSK109) and Illumina TruSeq DNA sample prep kits, respectively. The ONT reads and Illumina reads were used for *de novo* assembly and error correction, respectively. The ONT reads were assembled using Canu v2.0 (7); then, the initial assembly and trimmed corrected ONT reads were reassembled using Flye v2.8-b1674 (8). The final assembly was polished using medaka v1.0.3 (9) with ONT reads, Pilon v1.23 (10) with ONT reads, and Pilon v1.23 (10) with Illumina reads. The genome sequence of RIB2601 was assembled into nine contigs, which consist of eight chromosomes and one mitochondrial DNA. In addition, we found telomeres on both ends of each chromosome sequence, thus indicating that we successfully sequenced the nearly complete genome sequence of RIB2601. Genome annotation of the obtained chromosomal contigs and mitochondrial contig was performed using the Funannotate v1.8.1 pipeline (11) and MFannot v1.1 (12), respectively. Gene prediction was performed using SNAP v2006-07-28 (13), AUGUSTUS v3.3.3 (14), GlimmerHMM v3.0.4 (15), and GeneMark-ES v4.61_lic (16) via the Funannotate v1.8.1 pipeline (11). For the analysis, transcriptome sequencing (RNA-seq) reads of strain RIB2601 (6) (Sequence Read

TABLE 1 Replicons of *Aspergillus luchuensis* strain RIB2601

Location[a]	GenBank accession no.	Size (Mb)	% GC	No. of CDS[b]	No. of rRNAs[c]	No. of tRNAs
Chr 1	AP024434.1	6.19	49.5	2,011	NA	53
Chr 2	AP024435.1	4.96	48.9	1,553	NA	46
Chr 3	AP024436.1	4.48	49.4	1,477	NA	33
Chr 4	AP024437.1	3.78	49.2	1,294	9 (70)[d]	18
Chr 5	AP024438.1	4.82	47.9	1,490	NA	30
Chr 6	AP024439.1	3.97	48.7	1,323	NA	40
Chr 7	AP024440.1	3.19	48.2	1,026	NA	14
Chr 8	AP024441.1	4.07	48.6	1,364	NA	39
MT	AP024442.1	0.03	26.4	15	1	26

[a] Chr, chromosome; MT, mitochondria.
[b] CDS, coding DNA sequences.
[c] NA, not applicable.
[d] 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.

Archive [SRA] accession numbers SRX2414184 to SRX2414186) were *de novo* assembled using Trinity v2.8.5 (17) and used for gene prediction with the sequence alignment tool HISAT v2.2.0 (18). The proteins were annotated using MEROPS v12.0 (19), UniProt v2020_05 (20), MIBiG v1.4 (21), Pfam v33.1 (22), and dbCAN2 v9.0 (23) (based on CAZy database v7/30/2020 [24]) with sequence alignment tools such as DIAMOND v2.0.6 (25) and HMMER v3.3.2 (26). The annotation was also performed using InterProScan v5.47-82.0 (27), eggNOG-mapper v1.0.3 (28) (for the EggNOG v4.5 database [29]), antiSMASH v5.1.2 (30), SignalP v4.1 (31), Phobius v1.01 (32), tRNAscan-SE v2.0.7 (33), and Barrnap v0.9 (34). The RIB2601 genome includes 35,508,746 bp with a GC content of 48.8% and is comprised of 11,553 predicted coding sequences and 287 tRNAs. The genome completeness was assessed using BUSCO v5.1.2 with the ascomycota_odb10 data set (35), resulting in 98.9% complete and single-copy, 0.2% complete and duplicate-copy, 0.3% fragmented-copy, and 0.6% missing benchmarking universal single-copy orthologs (BUSCOs). The details for each replicon are summarized in Table 1. The chromosome-level genome sequence will aid in subsequent genomics research of black koji fungi.

Data availability. The nucleotide sequences of the *A. luchuensis* RIB2601 chromosomes and mitochondria have been deposited at DDBJ/ENA/GenBank under accession numbers AP024434 to AP024442. The raw sequence reads were deposited in the SRA under accession numbers DRX251232 and DRX251231.

REFERENCES

1. Yamada O, Takara R, Hamada R, Hayashi R, Tsukahara M, Mikami S. 2011. Molecular biological researches of Kuro-Koji molds, their classification and safety. J Biosci Bioeng 112:233–237. https://doi.org/10.1016/j.jbiosc.2011.05.005.
2. Hong S-B, Lee M, Kim D-H, Varga J, Frisvad JC, Perrone G, Gomi K, Yamada O, Machida M, Houbraken J, Samson RA. 2013. *Aspergillus luchuensis*, an industrially important black *Aspergillus* in East Asia. PLoS One 8:e63769. https://doi.org/10.1371/journal.pone.0063769.
3. Hong S-B, Yamada O, Samson RA. 2014. Taxonomic re-evaluation of black koji molds. Appl Microbiol Biotechnol 98:555–561. https://doi.org/10.1007/s00253-013-5332-9.
4. Sameshima Y. 2004. Technical changes of shochu, and the subject in the 21th century. J Brew Soc Japan 99:495–500. https://doi.org/10.6013/jbrewsocjapan1988.99.495.
5. Yamada O, Machida M, Hosoyama A, Goto M, Takahashi T, Futagami T, Yamagata Y, Takeuchi M, Kobayashi T, Koike H, Abe K, Asai K, Arita M, Fujita N, Fukuda K, Higa K-I, Horikawa H, Ishikawa T, Jinno K, Kato Y, Kirimura K, Mizutani O, Nakasone K, Sano M, Shiraishi Y, Tsukahara M, Gomi K. 2016. Genome sequence of *Aspergillus luchuensis* NBRC 4314. DNA Res 23:507–515. https://doi.org/10.1093/dnares/dsw032.
6. Kojo T, Kadooka C, Komohara M, Onitsuka S, Tanimura M, Muroi Y, Kurazono S, Shiraishi Y, Oda K, Iwashita K, Onoue M, Okutsu K, Yoshizaki Y, Takamine K, Futagami T, Mori K, Tamaki H. 2018. Characterization of amylolytic enzyme overproducing mutant of *Aspergillus luchuensis* obtained by ion beam mutagenesis. J Gen Appl Microbiol 63:339–346. https://doi.org/10.2323/jgam.2017.02.004.
7. 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.
8. 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.
9. Oxford Nanopore Technologies. 2018. Medaka. https://nanoporetech.github.io/medaka/.
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. Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo. https://doi.org/10.5281/zenodo.4054262.

12. Beck N, Lang BF. 2010. MFannot. http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl.

13. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59.

14. 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.

15. 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.

16. 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.

17. 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. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol 29:644–652. https://doi.org/10.1038/nbt.1883.

18. 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.

19. 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.

20. 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.

21. 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.

22. 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.

23. 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.

24. 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.

25. 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.

26. 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.

27. 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.

28. 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.

29. Ferrés I, Iraola G. 2018. Phylen: automatic phylogenetic reconstruction using the EggNOG database. J Open Source Softw 3:593. https://doi.org/10.21105/joss.00593.

30. 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.

31. 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.

32. Käll L, Krogh A, Sonnhammer ELL. 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.

33. 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.

34. Seemann T. 2018. Barrnap. https://github.com/tseemann/barrnap.

35. 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.