



Draft Genome Sequence of *Aspergillus oryzae* BP2-1, Isolated from Traditional Malted Rice in South Korea

Jongbum Jeon,^a Jung A Kim,^b Sook-Young Park,^c Gye-Won Kim,^d Cheon-Seok Park,^{e,f} Changmu Kim,^g Hye Yoon Park,^b Joo-Hong Yeo,^b Yong-Hwan Lee,^a Soonok Kim^b

^aDepartment of Agricultural Biotechnology, Interdisciplinary Program in Agricultural Genomics, Center for Fungal Genetic Resources, and Center for Fungal Pathogenesis, Seoul National University, Seoul, Republic of Korea

^bGenetic Resources Assessment Division, National Institute of Biological Resources, Incheon, Republic of Korea

^cDepartment of Plant Medicine, Suncheon National University, Suncheon, Republic of Korea

^dBrewing Research Center, Academic Industry Cooperation, Hankyong National University, Anseong, Republic of Korea

^eGraduate School of Biotechnology, Kyung Hee University, Yongin, Republic of Korea

^fInstitute of Life Science and Resources, Kyung Hee University, Yongin, Republic of Korea

^gMicroorganism Resources Division, National Institute of Biological Resources, Incheon, Republic of Korea

Jongbum Jeon, Jung A Kim, and Sook-Young Park contributed equally to this work. Author order was determined both alphabetically and in order of increasing seniority.

ABSTRACT The fungus *Aspergillus oryzae* strain BP2-1 was isolated from the traditional malted starter culture nuruk. We report here the draft whole-genome sequence of *A. oryzae* BP2-1, which is comprised of 14 scaffolds with a total length of 39,455,382 bp and a GC content of 47.13%.

Aspergillus oryzae is extensively used in industry, especially in the production of fermented foods and alcoholic beverages. Discovery of novel *A. oryzae* strains may assist in the development of novel enzymes and chemicals. An *A. oryzae* strain, BP2-1, which confers a unique flavor to makgeolli, was isolated during a survey of fungal strains from traditional nuruk in 2014. Moreover, extracts of malted rice with this strain exhibited lipid metabolism improvement activity by increasing peroxisome proliferator-activated receptor α (PPAR α) activity in monkey kidney cells (CV-1) and a skin-lightening effect by inhibiting tyrosinase activity. Here, we report the draft genome sequence of *A. oryzae* BP2-1.

This strain was isolated from a nuruk sample collected from Donghae (Gangwon, South Korea) according to the protocols described by Yang et al. (1) and was deposited in the National Institute of Biological Resources (NIBR) culture collection under the accession no. NIBRFGC000134774. High-quality genomic DNA was extracted from *A. oryzae* BP2-1 germlings grown overnight at 23°C in potato dextrose broth with shaking at 120 rpm using a DNeasy minikit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. Genome sequencing of *A. oryzae* BP2-1 was conducted at Theragen Etx Bio Institute (Suwon, South Korea). One short-insert paired-end (PE) library (fragment size, 550 bp) and two long-insert mate pair (MP) libraries (insert sizes, 5 kb and 10 kb) were generated using a TruSeq DNA sample prep kit and a Nextera MP library prep kit (Illumina, CA, USA), respectively. The single-molecule real-time (SMRT) sequencing library was prepared according to the PacBio standard library preparation protocol using a PacBio DNA template prep kit (Pacific Biosciences, CA, USA), with a fragment size of 20 kb. A total of 996,216 reads with an average read size of 7,484 bp were generated on 8 cells of a PacBio RS II sequencer. Reads were assembled using an overlap-layout-consensus (OLC) algorithm and FALCON v. 0.2JASM (2), generating 41 contigs with a total length of 38,361,741 bp and an N_{50} value of 2,463,276 bp. After filtering with NextClip v. 1.3 (3), we obtained 46,143,916 reads with a Q30 of 88.70% from the PE

Citation Jeon J, Kim JA, Park S-Y, Kim G-W, Park C-S, Kim C, Park HY, Yeo J-H, Lee Y-H, Kim S. 2020. Draft genome sequence of *Aspergillus oryzae* BP2-1, isolated from traditional malted rice in South Korea. Microbiol Resour Announc 9:e01405-19. <https://doi.org/10.1128/MRA.01405-19>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2020 Jeon 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 Yong-Hwan Lee, yonglee@snu.ac.kr, or Soonok Kim, sokim90@korea.kr.

Received 8 November 2019

Accepted 26 November 2019

Published 2 January 2020

library, 31,740,314 high-quality reads from the 5-kb inserted MP library, and 22,346,808 high-quality reads from the 10-kb inserted MP library from 101 cycles on an Illumina HiSeq 2000 platform. Using SOAPdenovo v. 2.04 (4), these short-read data sets were assembled into 78 contigs with a total length of 39,948,698 bp and an N_{50} value of 2,254,781 bp.

These two assemblies were merged into 23 contigs using HaploMerger2, followed by cleaning twice using the faDnaPolishing.pl script provided by HaploMerger2 (removeShortSeq=500) (5). Scaffolding and gap filling were performed using SSPACE-standard v3.0 (6), SSPACE-LongRead v1.1 (7), and GapFiller v1.10 (8), using default parameters. The final assembly consisted of 14 scaffolds with a total length of 39,455,382 bp and an N_{50} value of 4,366,015 bp. The GC content of the assembled genome was 47.13%. Genome assembly was validated using BUSCO v. 3.0.2b with the fungal ortholog data set (fungi_odb9) (9). In the assembly, 98.6% of the orthologs, including one that was duplicated, were found to be complete. This draft genome sequence will provide valuable information to identify genes for various bioactivities and will also facilitate comparative genomics with other publicly available *Aspergillus* fungal sequences for evolutionary studies.

Data availability. The genome sequence of *Aspergillus oryzae* BP2-1 obtained in this study has been deposited at GenBank under the accession no. [NGZNO0000000](https://ncbi.nlm.nih.gov/GenBank/NGZNO0000000). The version described in this article is the first version, NGZNO1000000. SRA data of PacBio and Illumina sequences were also deposited at GenBank under accession no. [SRR9306672](https://ncbi.nlm.nih.gov/SRA/SRR9306672) to [SRR9306682](https://ncbi.nlm.nih.gov/SRA/SRR9306682) (BioProject no. [PRJNA368788](https://ncbi.nlm.nih.gov/BioProject/PRJNA368788)).

ACKNOWLEDGMENT

This work was supported by the National Institute of Biological Resources, funded by the Ministry of Environment of the Republic of Korea (projects NIBR201830101 and NIBR201921101).

REFERENCES

1. Yang S, Lee J, Kwak J, Kim K, Seo M, Lee Y-W. 2011. Fungi associated with the traditional starter cultures used for rice wine in Korea. *J Korean Soc Appl Biol Chem* 54:933–943. <https://doi.org/10.1007/BF03253183>.
2. Wang A, Wang Z, Li Z, Li LM. 2018. BAUM: improving genome assembly by adaptive unique mapping and local overlap-layout-consensus approach. *Bioinformatics* 34:2019–2028. <https://doi.org/10.1093/bioinformatics/bty020>.
3. Leggett RM, Clavijo BJ, Clissold L, Clark MD, Caccamo M. 2014. NextClip: an analysis and read preparation tool for Nextera long mate pair libraries. *Bioinformatics* 30:566–568. <https://doi.org/10.1093/bioinformatics/btt702>.
4. Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaoqian Z, Liu G, Liao X, Li Y, Yang H, Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *Gigascience* 1:18. <https://doi.org/10.1186/2047-217X-1-18>.
5. Huang S, Chen Z, Huang G, Yu T, Yang P, Li J, Fu Y, Yuan S, Chen S, Xu A. 2012. HaploMerger: reconstructing allelic relationships for polymorphic diploid genome assemblies. *Genome Res* 22:1581–1588. <https://doi.org/10.1101/gr.133652.111>.
6. 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>.
7. Boetzer M, Pirovano W. 2014. SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics* 15:211. <https://doi.org/10.1186/1471-2105-15-211>.
8. Nadalin F, Vezzi F, Policriti A. 2012. GapFiller: a *de novo* assembly approach to fill the gap within paired reads. *BMC Bioinformatics* 13 Suppl 14:S8. <https://doi.org/10.1186/1471-2105-13-S14-S8>.
9. Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017. BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol* 35: 543–548. <https://doi.org/10.1093/molbev/msx319>.