




Chromosome-Level Genome Sequence of *Aspergillus puulaauensis* MK2, a Fungus Isolated from a Dead Hard Tick

 Taiki Futagami,^{a,b} Kazuki Mori,^c Chihiro Kadooka,^{a,b*} Hiroko Niihara,^d Kosuke Tashiro,^e Hisanori Tamaki,^{a,b} Tetsuya Tanaka^d

^aUnited Graduate School of Agricultural Sciences, Kagoshima University, Kagoshima, Japan

^bEducation and Research Centre for Fermentation Studies, Faculty of Agriculture, Kagoshima University, Kagoshima, Japan

^cCell Innovator Co., Ltd., Fukuoka, Japan

^dLaboratory of Infectious Diseases, Joint Faculty of Veterinary Medicine, Kagoshima University, Kagoshima, Japan

^eLaboratory of Molecular Gene Technology, Faculty of Agriculture, Kyushu University, Fukuoka, Japan

ABSTRACT *Aspergillus puulaauensis* strain MK2 was isolated from a dead hard tick (*Haemaphysalis longicornis*). Here, we determined the chromosome-level genome sequence of *A. puulaauensis* MK2.

During a *Haemaphysalis longicornis* hard tick study, we observed that the ticks were covered with fungi, resulting in subsequent tick death. *H. longicornis* is known as a vector of several pathogens causing illnesses, including *Theileria* spp., *Babesia* spp., *Rickettsia* spp., and severe fever with thrombocytopenia syndrome virus (1). We isolated *Aspergillus puulaauensis* MK2 from the dead tick. The species was identified based on the DNA sequence of the β -tubulin-encoding *benA* gene with direct colony PCR using a primer set of Bt₂a and Bt₂b (2). The *benA* gene exhibits 100% identity to those of *A. puulaauensis* NRRL 58602, which was isolated from indoor air in West Virginia, USA, and is known to produce sterigmatocystin (3, 4).

To further investigate the MK2 strain, we sequenced the genome of the MK2 strain. The MK2 strain was cultured in yeast extract-peptone-dextrose medium (2% [wt/vol] glucose, 1% [wt/vol] yeast extract, and 2% [wt/vol] peptone), and then the retrieved mycelia 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 MK2 was sequenced to 170- and 164-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) and Flye v2.8-b1674 (6). The Flye assembly was used to bridge separate contigs generated by Canu. We chose the better assembly metrics based on telomere-to-telomere genome assembly. Consequently, the genome of strain MK2 was assembled into 9 contigs, consisting of 8 chromosomes and 1 mitochondrial DNA, indicating that we successfully sequenced the nearly complete genome sequence of strain MK2. Seven chromosomes were generated only by Canu, while chromosome 1 was generated by an assembly in which 2 Canu contigs were bridged by a Flye contig. The final assembly was polished using medaka v1.0.3 (7) with ONT reads, Pilon v1.23 (8) with ONT reads, and Pilon v1.23 (8) with Illumina reads. Genome annotation of the chromosomal and mitochondrial contigs obtained was performed based on the Funannotate v1.8.1 pipeline (9) and MFannot v1.1 (10), respectively. Gene prediction was performed by using SNAP v2006-07-28 (11), AUGUSTUS v3.3.3 (12), GlimmerHMM v3.0.4 (13), and GeneMark-ES v4.61_lic (14) via the Funannotate v1.8.1 pipeline (9). For the analysis, RNA-seq reads for strain MK2 were obtained with the

Citation Futagami T, Mori K, Kadooka C, Niihara H, Tashiro K, Tamaki H, Tanaka T. 2021. Chromosome-level genome sequence of *Aspergillus puulaauensis* MK2, a fungus isolated from a dead hard tick. *Microbiol Resour Announc* 10:e00372-21. <https://doi.org/10.1128/MRA.00372-21>.

Editor Jason E. Stajich, University of California, Riverside

Copyright © 2021 Futagami 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 Taiki Futagami, k2130730@kadai.jp, or Tetsuya Tanaka, k6199431@kadai.jp.

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

Received 10 April 2021

Accepted 9 August 2021

Published 9 September 2021

TABLE 1 Replicons of *Aspergillus puulaauensis* strain MK2

Location ^a	GenBank accession no.	Size (Mb)	GC content (%)	No. of coding sequences	No. of rRNAs	No. of tRNAs
Chr. 1	AP024443.1	6.03	49.9	2,396	0	19
Chr. 2	AP024444.1	5.43	50.2	2,207	0	18
Chr. 3	AP024445.1	4.51	50.1	1,781	0	16
Chr. 4	AP024446.1	4.39	49.8	1,733	0	29
Chr. 5	AP024447.1	4.23	49.6	1,663	0	31
Chr. 6	AP024448.1	3.59	49.5	1,478	12 (69) ^b	7
Chr. 7	AP024449.1	3.32	49.3	1,267	0	8
Chr. 8	AP024450.1	2.82	49.3	1,089	0	7
MT	AP024451.1	0.03	24.2	14	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.

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 (15) and analyzed with the sequence alignment tool HISAT v2.2.0 (16). The proteins were annotated with MEROPS v12.0 (17), UniProt v2020_05 (18), MIBiG v1.4 (19), Pfam v33.1 (20), and dbCAN2 v9.0 (21) (based on the CAZy database v7/30/2020 [22]) using sequence alignment tools such as DIAMOND v2.0.6 (23) and HMMER v3.3.2 (24). Annotation was also performed using InterProScan v5.47-82.0 (25), eggNOG-mapper v1.0.3 (26) (for the eggNOG v4.5 database [27]), antiSMASH v5.1.2 (28), SignalP v4.1 (29), Phobius v1.01 (30), tRNAscan-SE v2.0.7 (31), and Barrnap v0.9 (32). The MK2 genome consists of 34,346,679 bp, with a GC content of 49.8%, and is composed of 13,627 predicted coding sequences and 162 tRNAs. Genome completeness was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.1.2 with the ascomycota_odb10 data set (33), which resulted in 99.2% complete and single-copy BUSCOs,

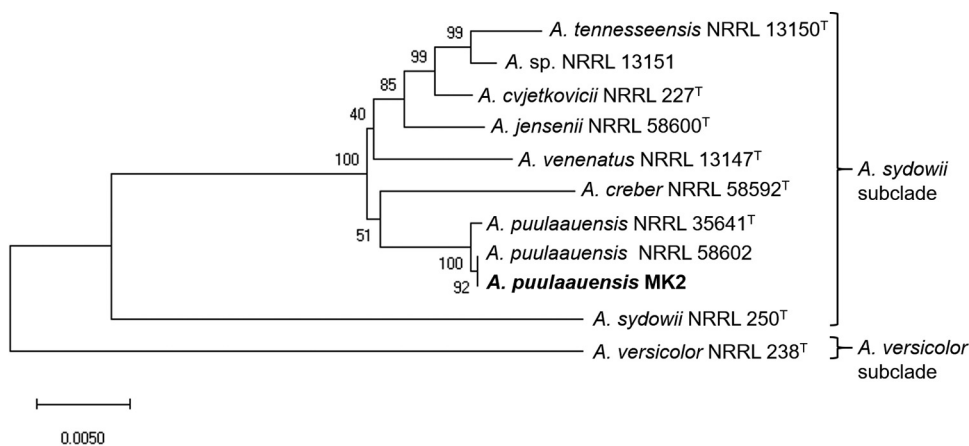


FIG 1 Phylogenetic analysis of *Aspergillus versicolor* clade strains, showing the phylogenetic position of strain MK2. The position of strain MK2 is indicated by bold type. The tree was constructed by a neighbor-joining method based on an alignment of concatenated *benA*, *caM*, *tsr1*, and *rpb2* sequences with complete gap deletion with MEGA X (34). Bootstrap values (1,000 replicates) are indicated at the branches. All of the other analytical options were set to defaults. Sequence data were obtained from the DDBJ using the following accession numbers: NRRL 13150 *benA*, JN853976.1; NRRL 13150 *caM*, JN854017.1; NRRL 13150 *tsr1*, JN853872.1; NRRL 13150 *rpb2*, JN853806.1; NRRL 13151 *benA*, JN853977.1; NRRL 13151 *caM*, JN854018.1; NRRL 13151 *tsr1*, JN853874.1; NRRL 13151 *rpb2*, JN853807.1; NRRL 227 *benA*, JN853971.1; NRRL 227 *caM*, EF652352.1; NRRL 227 *tsr1*, JN853866.1; NRRL 227 *rpb2*, EF652176.1; NRRL 58600 *benA*, JN854007.1; NRRL 58600 *caM*, JN854046.1; NRRL 58600 *tsr1*, JN853863.1; NRRL 58600 *rpb2*, JN853835.1; NRRL 13147 *benA*, JN854003.1; NRRL 13147 *caM*, JN854014.1; NRRL 13147 *tsr1*, JN853876.1; NRRL 13147 *rpb2*, JN853803.1; NRRL 58592 *benA*, JN853980.1; NRRL 58592 *caM*, JN854043.1; NRRL 58592 *tsr1*, JN853887.1; NRRL 58592 *rpb2*, JN853832.1; NRRL 35641 *benA*, JN853979.1; NRRL 35641 *caM*, JN854034.1; NRRL 35641 *tsr1*, JN853895.1; NRRL 35641 *rpb2*, JN853823.1; NRRL 58602 *benA*, JN853999.1; NRRL 58602 *caM*, JN854048.1; NRRL 58602 *tsr1*, JN853896.1; NRRL 58602 *rpb2*, JN853837.1; NRRL 250 *benA*, JN853933.1; NRRL 250 *caM*, EF652362.1; NRRL 250 *tsr1*, JN853901.1; NRRL 250 *rpb2*, EF652186.1; NRRL 238 *benA*, JN853941.1; NRRL 238 *caM*, EF652354.1; NRRL 238 *tsr1*, JN853911.1; NRRL 238 *rpb2*, EF652178.1.

0.4% complete and duplicate-copy BUSCOs, 0.1% fragmented-copy BUSCOs, and 0.3% missing BUSCOs. Details regarding the replicons present are summarized in Table 1. To confirm the species-level identification of strain MK2, we performed phylogenetic analysis using *benA* and the genes encoding calmodulin (*caM*), pre-rRNA processing protein (*tsr1*), and RNA polymerase II subunit (*rpb2*) for strain MK2 and *Aspergillus versicolor* clade strains (2) with MEGA X software (34). The position of strain MK2 was closest to *A. puulaauensis* NRRL 58602 (2, 3), confirming that strain MK2 is classified as *A. puulaauensis* (Fig. 1).

Data availability. The nucleotide sequences of *A. puulaauensis* MK2 chromosomes and mitochondria have been deposited in DDBJ/ENA/GenBank under accession numbers AP024443 to AP024451. Raw sequence reads have been deposited in the Sequence Read Archive (SRA) under accession numbers DRX256207, DRX251723, and DRX251722.

ACKNOWLEDGMENTS

This work was funded by a supporting program for interdisciplinary projects in Kagoshima University and a JSPS KAKENHI grant (grant 20KK0154). C.K. was supported by a Grant-in-Aid for JSPS Research Fellows (grant 17J02753).

REFERENCES

- Rochlin I, Toledo A. 2020. Emerging tick-borne pathogens of public health importance: a mini-review. *J Med Microbiol* 69:781–791. <https://doi.org/10.1099/jmm.0.001206>.
- Glass NL, Donaldson GC. 1995. Development of primer sets designed for use with the PCR to amplify conserved genes from filamentous ascomycetes. *Appl Environ Microbiol* 61:1323–1330. <https://doi.org/10.1128/aem.61.4.1323-1330.1995>.
- Jurjević Z, Peterson SW, Horn BW. 2012. *Aspergillus* section *Versicolores*: nine new species and multilocus DNA sequence based phylogeny. *IMA Fungus* 3:59–79. <https://doi.org/10.5598/imafungus.2012.03.01.07>.
- Jurjević Z, Peterson SW, Solfrizzo M, Peraica M. 2013. Sterigmatocystin production by nine newly described *Aspergillus* species in section *Versicolores* grown on two different media. *Mycotoxin Res* 29:141–145. <https://doi.org/10.1007/s12550-013-0160-4>.
- 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>.
- 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>.
- Oxford Nanopore Technologies. 2018. Medaka. <https://nanoporetech.github.io/medaka>.
- 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>.
- Palmer JM, Stajich JE. 2020. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo <https://doi.org/10.5281/zenodo.4054262>.
- Beck N, Lang BF. 2010. MFannot. <http://megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
- 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>.
- 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>.
- 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>.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Muceli 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>.
- 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>.
- 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>.
- 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>.
- Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hoof 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>.
- 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, Paladini 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>.
- 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>.
- 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>.
- 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>.
- 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>.
- 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, Scheremetjov 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>.
- 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>.
- 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>.
- 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>.
29. 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>.
 30. Käll L, Krogh A, Sonnhammer EL. 2007. Advantages of combined trans-membrane topology and signal peptide prediction—the Phobius web server. *Nucleic Acids Res* 35:W429–W432. <https://doi.org/10.1093/nar/gkm256>.
 31. 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.
 32. Seemann T. 2018. Barrnap. <https://github.com/tseemann/barrnap>.
 33. 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.
 34. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 35:1547–1549. <https://doi.org/10.1093/molbev/msy096>.