

# Draft genome sequence of an entomopathogenic fungus *Amphichorda felina*

Ngoc-Hung Nguyen,<sup>1,2</sup> Phuong-Thao Nguyen,<sup>2</sup> Toshimitsu Fukihar, <sup>3</sup> Haruka Kameda,<sup>3,4</sup> Kiminori Shimizu<sup>2,5</sup>

**AUTHOR AFFILIATIONS** See affiliation list on p. 2.

**ABSTRACT** An entomopathogenic fungus, *Amphichorda felina* (formerly classified as *Beauveria felina*), was sequenced through Illumina MiSeq short reads and Oxford Nanopore long reads. The hybrid assembly resulted in a genome of approximately 29.6 Mb with 9 scaffolds and an N50 of 3.28 Mb.

**KEYWORDS** entomopathogen

*Amphichorda felina* (formerly *Beauveria felina*) is an entomopathogenic fungus with biocontrol potential. Recent multi-locus phylogenetic analyses revealed that this species is distinct from other *Beauveria* species, leading to its reclassification into *Amphichorda* (Bionectriaceae, Sordariomycetes, Hypocreales) (1). To support future biocontrol applications, we sequenced and assembled the genome of *A. felina* strain CBM-FB-45539.

The wild strain CBM-FB-45539 was isolated from deer feces collected on 4 December 2021 on Yakushima Island, Japan. Dung was incubated in a moist chamber (1,000 mL bottle with sterilized, moistened moss) at 22°C under a 12 h light/dark cycle (2). After two weeks, basidiomata developed; a fragment was aseptically transferred to MYC medium to establish CBM-FB-45539. The strain was deposited in the Natural History Museum and Institute, Chiba, Japan (CBM-FB-45539) and maintained on MYC agar at 4°C until DNA extraction. Mycelia from 7-day cultures were harvested, and DNA was extracted separately for MGI and Nanopore sequencing using the DNeasy Plant Mini Kit (Qiagen) following the manufacturer's protocol. Taxonomic status was confirmed by ITS rDNA sequencing (primers ITS1/ITS4 [3]; GenBank [PV413099.1](#)). BLASTn against NCBI nt (accessed 8 December 2022) showed 98.93% sequence identity to *A. felina* strain CBS 250.34 (GenBank [AY261369.1](#)), and monophyly was confirmed by phylogenetic analysis (4).

Illumina libraries were prepared using the MGIEasy PCR-Free DNA Library Prep Set and sequenced on DNBSEQ-G400RS (paired-end 150 bp). Nanopore DNA was prepared without shearing or size selection using the Native Barcoding Kit 24 V14 (SQK-NBD114.24) and sequenced on PromethION (FLO-PRO114M flow cells). Base calling was performed using Dorado v0.7.3 with the SUP model. We obtained 36,771,966 MiSeq read pairs (~9.79 Gb) and 361,240 long reads (2.86 Gb). Long reads were adapter-trimmed (Porechop v0.2.4), contamination-filtered (NanoLyse v1.1.0 [5]), and size-filtered >1,000 bp (SeqKit v2.2.0 [6]). Illumina reads were quality-filtered and adapter-trimmed (cutadapt v4.2 [7]), retaining reads with Phred quality >20 and minimum length of 20 bp.

The mitochondrial genome was assembled and circularized from Illumina reads using GetOrganelle v1.7.7.1 (8), which automatically detects and trims overlapping terminal regions. Reads from both platforms were mapped to the mitochondrial genome using Minimap2 v2.22 (9), and mitochondrial reads were removed (samtools v1.16.1 [10]). The nuclear genome was assembled from remaining reads using MaSuRCA v4.1.2 (11), polished by mapping MiSeq reads (BWA-MEM v0.7.17 [12]) followed by four rounds of

**Editor** Leighton Pritchard, University of Strathclyde, Glasgow, United Kingdom

Address correspondence to Kiminori Shimizu, [shimizuk@rs.tus.ac.jp](mailto:shimizuk@rs.tus.ac.jp).

The authors declare no conflict of interest.

See the funding table on p. 3.

**Received** 8 October 2025

**Accepted** 3 February 2026

**Published** 24 February 2026

Copyright © 2026 Nguyen et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

**TABLE 1** Genome features of the *Amphichorda felina* assembly and annotation

| Characteristic                              | Value         |
|---------------------------------------------|---------------|
| Chromosome contigs                          |               |
| Total assembled length (bp)                 | 29,588,262    |
| No. of contigs                              | 9             |
| N50 (bp)                                    | 3,278,244     |
| L50                                         | 4             |
| Max length (bp)                             | 4,728,797     |
| No. of complete BUSCOs (%)                  | 3,747 (98.1%) |
| No. of incomplete BUSCOs (%)                | 6 (0.2%)      |
| No. of missing BUSCOs (%)                   | 64 (1.4%)     |
| Repeats                                     | 1.74%         |
| GC                                          | 54.68%        |
| No. of protein-coding genes                 | 9,270         |
| No. of tRNAs                                | 157           |
| No. of Phobius secretome genes              | 1,109         |
| No. of Phobius transmembrane proteins       | 2,018         |
| No. of antiSMASH biosynthetic gene clusters | 49            |
| Mitochondrial genome                        |               |
| Length (bp)                                 | 27,301        |
| No. of protein-coding genes                 | 17            |
| No. of rRNAs                                | 2             |
| No. of tRNAs                                | 24            |

Pilon v1.24 (13). The mitochondrial genome was annotated using MFannot (14) and MITOS2 v2.1.3+galaxy0 (15) on Galaxy (16) (Genetic Code 4); both resulting annotations were inspected in IGV (17) and reconciled by manual curation to produce a single final mitochondrial genome annotation. Genome completeness was assessed using BUSCO v5.4.3 with 3,817 Sordariomycetes orthologs (18). Repeats were predicted (RepeatModeler v2.0.1 [19]) and masked (RepeatMasker v4.1.1 [20]). The Funannotate v1.8.1 pipeline (21) predicted 9,427 protein-coding genes across the nine assembled contigs (Table 1). The complete mitochondrial genome is a circular molecule of 27,301 bp (single closed contig) containing 43 genes: 2 rRNA, 17 protein-coding, and 24 tRNA genes. Default parameters were used unless otherwise specified.

## ACKNOWLEDGMENTS

This work was partly supported by the Japan Science and Technology Agency (JPMJCR21N6 to N.-H.N.).

## AUTHOR AFFILIATIONS

<sup>1</sup>Department of Computational Biology and Medical Sciences, Graduate School of Frontier Sciences, University of Tokyo, Kashiwa, Chiba, Japan

<sup>2</sup>Department of Biological Science and Technology, Tokyo University of Science, Tokyo, Japan

<sup>3</sup>Natural History Museum and Institute, Chiba, Japan

<sup>4</sup>Graduate School of Human and Environmental Studies, Kyoto University, Kyoto, Japan

<sup>5</sup>Medical Mycology Research Center, Chiba University, Chiba, Japan

## AUTHOR ORCIDs

Ngoc-Hung Nguyen  <http://orcid.org/0000-0001-7008-8800>

Kiminori Shimizu  <http://orcid.org/0000-0001-5966-976X>

## FUNDING

| Funder                              | Grant(s)   | Author(s)        |
|-------------------------------------|------------|------------------|
| Japan Science and Technology Agency | JPMJCR21N6 | Ngoc-Hung Nguyen |

## AUTHOR CONTRIBUTIONS

Ngoc-Hung Nguyen, Conceptualization, Data curation, Formal analysis, Investigation, Writing – original draft | Phuong-Thao Nguyen, Data curation, Investigation | Toshimitsu Fukiharu, Resources | Haruka Kameda, Resources | Kiminori Shimizu, Conceptualization, Writing – review and editing

## DATA AVAILABILITY

The Amphicorda felina genome project, including raw Illumina and Nanopore data, has been deposited at DDBJ/EMBL/GenBank under BioProject accession number [PRJNA1174440](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1174440). The raw MiSeq and Nanopore sequencing reads have been deposited at the Sequence Read Archive under accession numbers [SRR31037621](https://www.ncbi.nlm.nih.gov/sra/SRR31037621) and [SRR31037620](https://www.ncbi.nlm.nih.gov/sra/SRR31037620), respectively. The assembled nuclear genome and assembled mitochondrial sequences are available under GenBank accession numbers [JBIMJW000000000](https://www.ncbi.nlm.nih.gov/nuccore/JBIMJW000000000) (the version described in this paper is version [JBIMJW010000000.1](https://www.ncbi.nlm.nih.gov/nuccore/JBIMJW010000000.1)) and [PQ516710.1](https://www.ncbi.nlm.nih.gov/nuccore/PQ516710.1), respectively.

## REFERENCES

- Guerra-Mateo D, Gené J, Baulin V, Cano-Lira JF. 2023. Phylogeny and taxonomy of the genus *Amphichorda* (Bionectriaceae): an update on Beauveria-like strains and description of a novel species from marine sediments. *Diversity (Basel)* 15:795. <https://doi.org/10.3390/d15070795>
- Fukiharu T, Takarada K, Hosoya T, Kinjo N. 2005. Three *Coprinus* species occurred on the animal dung collected at Yatsugatake range, central Honshu, Japan. *Bull Natl Sci Mus Ser B Bot* 31:117–126.
- White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. *PCR Protocols*:315–322. <https://doi.org/10.1016/B978-0-12-372180-8.50042-1>
- Nguyen HN. 2025. Supplement for "Draft genome sequence of an entomopathogenic fungus *Amphichorda felina*". Zenodo. <https://doi.org/10.5281/zenodo.17284441>.
- De Coster W, D'Hert S, Schultz DT, Cruts M, Van Broeckhoven C. 2018. NanoPack: visualizing and processing long-read sequencing data. *Bioinformatics* 34:2666–2669. <https://doi.org/10.1093/bioinformatics/bty149>
- Shen W, Le S, Li Y, Hu F. 2016. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLoS One* 11:e0163962. <https://doi.org/10.1371/journal.pone.0163962>
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet j* 17:10. <https://doi.org/10.14806/ej.17.1.200>
- Jin J-J, Yu W-B, Yang J-B, Song Y, dePamphilis CW, Yi T-S, Li D-Z. 2020. GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes. *Genome Biol* 21:241. <https://doi.org/10.1186/s13059-020-02154-5>
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Li Heng, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. 2013. The MaSuRCA genome assembler. *Bioinformatics* 29:2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. <https://doi.org/10.48550/arXiv.1303.3997>
- 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>
- Valach M, Burger G, Gray MW, Lang BF. 2014. Widespread occurrence of organelle genome-encoded 5S rRNAs including permuted molecules. *Nucleic Acids Res* 42:13764–13777. <https://doi.org/10.1093/nar/gku1266>
- Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsche G, Pütz J, Middendorf M, Stadler PF. 2013. MITOS: improved de novo metazoan mitochondrial genome annotation. *Mol Phylogenet Evol* 69:313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>
- Abueg LAL, Afgan E, Allart O, Awan AH, Bacon WA, Baker D, Bassetti M, Batut B, Bernt M, Blankenberg D, et al. 2024. The galaxy platform for accessible, reproducible, and collaborative data analyses: 2024 update. *Nucleic Acids Res* 52:W83–W94. <https://doi.org/10.1093/nar/gkae410>
- Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>
- 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>
- Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* 117:9451–9457. <https://doi.org/10.1073/pnas.1921046117>
- Smit A, Hubley R, Green P. 2020. RepeatMasker open-4.0. Available from: <http://www.repeatmasker.org>
- Palmer J. 2020. Funannotate: fungal genome annotation scripts. Available from: <https://funannotate.readthedocs.io/en/latest>