

Draft genome assemblies of two *Beauveria bassiana* strains from Taiwan using the hybrid assembly approach

Kusum Mushyakhwo,¹ Yu-Shin Nai²

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT We report the draft genome sequences of two *Beauveria bassiana* strains, Bb-NCHU-143 and Bb-NCHU-153, isolated from mycosed coleopteran (beetles) in Taiwan. Genome assemblies were generated using a hybrid genome assembly approach. The draft genomes of Bb-NCHU-143 and Bb-NCHU-153 are 34.21 Mb (7,994 genes) and 34.20 Mb (7,151 genes), respectively.

KEYWORDS *Beauveria bassiana*, hybrid assembly, whole-genome sequencing

Beauveria bassiana is a widely studied entomopathogenic fungus (EPF), well known for its effectiveness as a biocontrol agent against a variety of insect pests. In this study, we report the hybrid genome assemblies of two indigenous *B. bassiana* strains, Bb-NCHU-143 and Bb-NCHU-153, which exhibit high pathogenicity against the agricultural pest, fall army worm (*Spodoptera frugiperda*) (1).

The EPF strains, Bb-NCHU-143 and Bb-NCHU-153, were isolated from cadavers of coffee berry borer (*Hypothenemus hampei*) from Chiayi, Taiwan and red palm weevil (*Rhynchophorus ferrugineus*) from Changhua, Taiwan, respectively. Isolation, purification, and identification of the strains have been previously reported (1–3). The strains were stored at –80°C in 20% glycerol at the Insect Pathology and Genomics Laboratory, National Chung Hsing University (Taichung, Taiwan). The EPF strains were recovered from –80°C and cultured on quarter-strength Sabouraud dextrose agar (HiMedia Co.) at 25°C for 10 days in darkness for genomic DNA extraction.

Fungal genomic DNA was extracted using a Yeast Genomic DNA Kit (Geneaid, Taiwan), following the manufacturer's protocol. The extracted DNA was subjected to both Illumina and Oxford Nanopore Technologies (ONT) sequencing. For Illumina sequencing, a library was constructed using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, USA) and sequenced on an Illumina HiSeq platform with paired-end (PE) reads of 2 × 150 bp (Macrogen, Korea). Quality trimming and adapter removal of the NGS reads were performed using Trimmomatic v0.39 (4). For ONT sequencing, the library was prepared using the Ligation Sequencing Kit SQK-LSK109 (ONT, Cambridge, UK) and sequenced on an R9.4.1 SpotON flow cell using a MinION device operated with ONT MinKNOW software version 22.05.5. ONT sequencing reads were trimmed using Porechop v0.2.4 (5) (<https://github.com/rrwick/Porechop>) to remove adapters. The raw Illumina sequencing data for strain Bb-NCHU-143 consisted of 42,561,883 reads, with 35,927,457 reads retained after quality filtering. For strain Bb-NCHU-153, a total of 39,237,588 reads were generated, of which 32,371,398 high-quality reads remained after filtering. ONT sequencing followed by Guppy basecalling v6.1.7 (ONT community site: <https://community.nanoporetech.com>) (6), produced 468,422 reads (2,989,362,299 bases) and 81,298 reads (637,332,762 bases) for Bb-NCHU-143 and Bb-NCHU-153, respectively. After adapter trimming, 468,140 reads (2,937,027,379 bases) and 81,304 reads (633,983,475 bases) for Bb-NCHU-143 and Bb-NCHU-153, respectively, were used for hybrid genome assembly.

Editor André O. Hudson, Rochester Institute of Technology, Rochester, New York, USA

Address correspondence to Yu-Shin Nai, ysnai@nchu.edu.tw.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 7 October 2025

Accepted 8 November 2025

Published 26 November 2025

Copyright © 2025 Mushyakhwo and Nai. 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/).

TABLE 1 Genome features of *Beauveria bassiana* strains assembly and annotation

Assembled genome	Bb-NCHU-143	Bb-NCHU-153
Number of contigs	11	13
Largest contig (bp)	7,146,430	7,946,388
Total length (bp)	34,214,673	34,200,021
Coverage	387.59×	290.12×
GC (%)	50.69	50.52
<i>N</i> ₅₀ (bp)	5,547,362	5,644,419
Complete BUSCO marker genes (%)	99.60%	99.70%
Fragmented BUSCO marker genes (%)	0.10%	0.00%
Missing BUSCO marker genes (%)	0.30%	0.30%
Total protein-coding genes	7,994	7,151
Gene density (gene/Mb)	233.64	209.09
Number of exons	20,901	19,009
Number of exons per gene	2.62	2.66
Number of tRNA	127	200
Number of secondary metabolites	38	41
Repeat sequence density (%)	5.94	6.36
TE density	6.49%	6.83%

Genome assembly was performed using MaSuRCA v4.1.0 (7), integrating both Illumina and ONT sequencing reads. Default parameters were used, except: LHE_COVERAGE = 15, JF_SIZE = 1000000000, CA_PARAMETERS = cgwErrorRate = 0.15, ovlMinLen = 50. Assembly quality was evaluated using QUAST v5.2.0 (8), and completeness was evaluated using BUSCO v5.8.2 (9) against the fungal_odb10 database. The fungal genome was annotated using BRAKER (10–12), incorporating RNA-seq-based gene prediction. Public RNA-seq data (SRR3105334; PRJNA309019) from NCBI were aligned to the assembly with HISAT2 v2.2.1 (13). Transfer RNA (tRNA) genes were predicted using tRNAscan-SE v2.0 (14). Secondary metabolite gene clusters were identified using the antiSMASH web server (15). Transposable element (TE) regions were predicted with RepeatModeler2 (16) and annotated using RepeatMasker (17). The information on genome assembly and analysis is summarized in Table 1. Default parameters were used except where otherwise noted.

ACKNOWLEDGMENTS

We thank Ms. Ju-Chun Chang for gDNA sample preparation, Mr. Tzu-Han Chen for high-throughput sequencing library preparation, and Dr. Yu-Feng Huang for the data consultation.

This research was supported by NSTC 113-2313-B-005-024-MY3 from the National Science and Technology Council.

AUTHOR AFFILIATIONS

¹International Doctoral Program in Agriculture, National Chung Hsing University, Taichung, Taiwan

²Department of Entomology, National Chung Hsing University, Taichung, Taiwan

AUTHOR ORCIDs

Kusum Mushyakhwo  <http://orcid.org/0009-0007-5950-2055>

Yu-Shin Nai  <http://orcid.org/0000-0003-0104-6839>

FUNDING

Funder	Grant(s)	Author(s)
National Science and Technology Council	NSTC 113-2313-B-005-024-MY3	Yu-Shin Nai

AUTHOR CONTRIBUTIONS

Kusum Mushyakhwo, Conceptualization, Data curation, Formal analysis, Methodology, Validation, Writing – original draft, Writing – review and editing | Yu-Shin Nai, Conceptualization, Data curation, Writing – review and editing

DATA AVAILABILITY

The *Beauveria bassiana* genome project, including the raw Illumina and Nanopore data, has been deposited in NCBI under BioProject accession number [PRJNA1237937](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1237937). This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession numbers [JBPXMI000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBPXMI000000000) (*Beauveria bassiana*-NCHU-143) and [JBPYGC000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBPYGC000000000) (*Beauveria bassiana*-NCHU-153). The versions described in this paper are version [JBPXMI010000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBPXMI010000000) (Bb-NCHU-143) and [JBPYGC010000000](https://www.ncbi.nlm.nih.gov/nuclseq/JBPYGC010000000) (Bb-NCHU-153). The raw Illumina sequencing reads have been deposited under accession numbers [SRR32764875](https://www.ncbi.nlm.nih.gov/nuclseq/SRR32764875) (Bb-NCHU-143) and [SRR32764876](https://www.ncbi.nlm.nih.gov/nuclseq/SRR32764876) (Bb-NCHU-153). The Nanopore sequencing data were deposited under accession numbers [SRR32764873](https://www.ncbi.nlm.nih.gov/nuclseq/SRR32764873) (Bb-NCHU-143) and [SRR32764874](https://www.ncbi.nlm.nih.gov/nuclseq/SRR32764874) (Bb-NCHU-153).

REFERENCES

- Wu S-S, Tseng C-T, Yang Y-H, Liu Y-C, Chang J-C, Gyawali P, Li Y-H, Yang T-H, Tsai Y-F, Tang L-C, Nai Y-S. 2022. Potential of a combination of entomopathogenic fungal strains and a non-ionic surfactant to control the fall armyworm (*Spodoptera frugiperda*). *J Asia Pac Entomol* 25:102001. <https://doi.org/10.1016/j.aspen.2022.102001>
- Chang J-C, Wu S-S, Liu Y-C, Yang Y-H, Tsai Y-F, Li Y-H, Tseng C-T, Tang L-C, Nai Y-S. 2021. Construction and selection of an entomopathogenic fungal library from soil samples for controlling *Spodoptera litura*. *Front Sustain Food Syst* 5:596316. <https://doi.org/10.3389/fsufs.2021.596316>
- Liu Y-C, Ni N-T, Chang J-C, Li Y-H, Lee MR, Kim JS, Nai Y-S. 2021. Isolation and selection of entomopathogenic fungi from soil samples and evaluation of fungal virulence against insect pests. *J Vis Exp* 175:e62882. <https://doi.org/10.3791/62882-v>
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Completing bacterial genome assemblies with multiplex MinION sequencing. *Microb Genom* 3:e000132. <https://doi.org/10.1099/mgen.0.000132>
- Technologies O. N. 2020. Oxford nanopore technologies. GitHub, United States. Available from: <https://community.nanoporetech.com>
- Zimin AV, Puiu D, Luo M-C, Zhu T, Koren S, Marçais G, Yorke JA, Dvořák J, Salzberg SL. 2017. Hybrid assembly of the large and highly repetitive genome of *Aegilops tauschii*, a progenitor of bread wheat, with the MaSuRCA mega-reads algorithm. *Genome Res* 27:787–792. <https://doi.org/10.1101/gr.213405.116>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Manni M, Berkeley MR, Seppely M, Simão FA, Zdobnov EM. 2021. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol* 38:4647–4654. <https://doi.org/10.1093/molbev/msab199>
- Brůna T, Hoff KJ, Lomsadze A, Stanke M, Borodovsky M. 2021. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR Genom Bioinform* 3:lqaa108. <https://doi.org/10.1093/nargab/lqaa108>
- Gabriel L, Hoff KJ, Brůna T, Borodovsky M, Stanke M. 2021. TSEBRA: transcript selector for BRAKER. *BMC Bioinformatics* 22:1–12. <https://doi.org/10.1186/s12859-021-04482-0>
- Hoff KJ, Lange S, Lomsadze A, Borodovsky M, Stanke M. 2016. BRAKER1: unsupervised RNA-Seq-based genome annotation with genemark-ET and AUGUSTUS. *Bioinformatics* 32:767–769. <https://doi.org/10.1093/bioinformatics/btv661>
- 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>
- Chan PP, Lowe TM. 2019. tRNAscan-SE: searching for tRNA genes in genomic sequences. In Kollmar M (ed), *Gene prediction: Methods in Molecular Biology*. Humana, New York, NY.
- 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>
- 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>
- Chen N. 2004. Using repeat masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform* 5. <https://doi.org/10.1002/0471250953.bi0410s05>