





Draft Genome Sequence of the Lichenized Fungus *Bacidia giganteis*

 Jessica L. Allen,^a Steven J. M. Jones,^b  R. Troy McMullin^c

^aBiology Department, Eastern Washington University, Cheney, Washington, USA

^bCanada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency, Vancouver, British Columbia, Canada

^cResearch and Collections, Canadian Museum of Nature, Ottawa, Ontario, Canada

ABSTRACT The draft genome sequence of *Bacidia giganteis*, a lichenized fungus in the order Lecanorales, was sequenced directly from a herbarium specimen collected from the type locality at Sleeping Giant Provincial Park in Ontario, Canada. Using long-read sequencing on the Oxford Nanopore PromethION platform, we assembled a nearly complete genome sequence.

We sequenced the genome of *Bacidia giganteis*, a crustose, epiphytic, lichenized fungus in the family Ramalinaceae. The species was recently described as new to science from Sleeping Giant Provincial Park along the north shore of Lake Superior in Ontario, Canada (1). A whole lichen specimen was collected at the type locality (1), and DNA was extracted from whole thallus material. The tissue was disrupted using metal bead-based homogenization; proteins were pelleted via centrifugation and the DNA-containing supernatant drawn off. Then, multiple ethanol washes were conducted over a silica column. All steps followed a protocol similar to those for commercially available kits (2). Bead cleanup at a 0.6:1 bead-to-sample ratio was first performed to remove small DNA fragments using PCR Clean DX C-1003-450 (Aline Biosciences, Woburn, MA, USA). The quality of the genomic DNA (gDNA) was assessed using pulse-field gel electrophoresis (PFGE) with 1% agarose gel by running with 0.5× Tris-borate-EDTA (TBE) on a CHEF pulse field gel apparatus (Bio-Rad, Hercules, CA, USA) for 22 h. The library construction and sequencing were conducted following Oxford Nanopore's genomic DNA by ligation (SQK-LSK109) protocol. Briefly, 1.14 μg of genomic DNA was formalin-fixed, paraffin-embedded (FFPE) repaired, end-prepped, and dA-tailed using the NEBNext FFPE DNA repair mix (M6630) and the NEBNext Ultra II end prep/dA-tailing module (E7546). NEBNext quick ligase (E6056S) was used to ligate the Oxford Nanopore adapters. A final size selection using a 0.4:1 bead-to-library ratio was performed to select against smaller molecules. Then, 380 ng of the final library was loaded onto a PromethION 24 instrument using the R9.4.1 pore flow cell and v19.06.9 software (MinKNOW graphical user interface [GUI] v4.0.23). Sequencing was run for 72 h. Base calling was performed using Fast-Bonito v0.2.2 (<https://github.com/nanoporetech/bonito>). The N_{50} value of the reads was 26 kb. A total of 14,958,030 reads containing 32 Gbp were generated.

The reads were corrected using Canu v2.0 (3); they were then assembled using Flye v2.8.2 with "--pacbio-hifi" mode, the option "--hifi-error 0.006," and using the purge_haplotigs tool to remove duplicated disjointigs between the "consensus" and "repeat" stages of the Flye assembly pipeline (4, 5). The assembly was polished by mapping all of the reads to the assembly using Minimap2 v2.17, then implementing one round of Racon v1.4.17 polishing and one round of Medaka v1.2 polishing (6) (<https://github.com/nanoporetech/medaka>). To remove nontarget organism sequences from the assembly, we used a Diamond v0.9.32 blastx search against the NCBI nonredundant protein database to assign the taxa to each scaffold, then built an annotated GC depth

Citation Allen JL, Jones SJM, McMullin RT. 2021. Draft genome sequence of the lichenized fungus *Bacidia giganteis*. Microbiol Resour Announc 10:e00686-21. <https://doi.org/10.1128/MRA.00686-21>.

Editor Antonis Rokas, Vanderbilt University

Copyright © 2021 Allen 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 R. Troy McMullin, tmcmullin@nature.ca.

Received 22 July 2021

Accepted 12 October 2021

Published 4 November 2021

TABLE 1 Summary statistics for the *Bacidia gigantensis* genome assembly and annotation

Characteristic ^a	Value
Total assembly length (Mb)	33.12
No. of scaffolds	24
N_{50} (Mb)	18.1
L_{50}	8
No. of complete single-copy BUSCOs (%)	1,186 (92.7)
No. of duplicate single-copy BUSCOs (%)	1 (0.5)
No. of incomplete BUSCOs (%)	35 (2.6)
No. of missing BUSCOs (%)	35 (4.6)
% repeat	7.69
%GC	44.67
No. of tRNAs	51
No. of Phobius secretome genes	931
No. of Phobius transmembrane proteins	1,798
No. of antiSMASH biosynthetic gene clusters	35
No. of antiSMASH biosynthetic enzymes	78
No. of antiSMASH smCOGs	103
No. of CAZYmes	205
Total no. of annotations	14,923
NCBI BioProject accession no.	PRJNA748063
Sequence Read Archive accession no.	SRX11510989
GenBank assembly accession no.	GCA_019456465.1

^aBUSCOs, benchmarking universal single-copy orthologs; smCOGs, secondary metabolism clusters of orthologous groups; CAZYmes, carbohydrate-active enzymes.

plot using the BITAT python script from McKenzie et al. (7) (<https://github.com/biorover/LethariaGenomes>). We then kept only scaffolds with 70× to 80× coverage and a GC content below 0.6 that were annotated as Ascomycota, as was the *Bacidia gigantensis* genome. We used QAST v5.0.2 to evaluate the final metrics for the assembly. The assembly was highly contiguous and comprised 24 scaffolds, with a total length of 33.12 Mb, with an N_{50} value of 18.1 Mb and an L_{50} value of 8 Mb (8). The genome sequence was also highly complete, with a BUSCO v3.1 homology search against the Pezizomycotina database, resulting in 92.7% complete single-copy genes, 0.5% duplicate single-copy genes, 2.6% incomplete genes, and 4.6% missing genes (9).

We used Funannotate v1.8.7 for gene prediction in the nuclear genome using all default evidence-guided and *ab initio* predictors (<https://github.com/nextgenusfs/funannotate>). We first masked the repeat regions, which comprised 7.69% of the genome. Protein evidence was derived from the Joint Genome Institute database for the following accessions: *Cladonia grayi* Cgr/DA2myc/ss v2.0, *Lobaria pulmonaria* Scotland reference genome v1.0, *Usnea florida* ATCC 18376 v1.0, and *Xanthoria parietina* 46-1-SA22 v1.1. *Aspergillus fumigatus* was selected as the seed species for BUSCO searches. A total of 14,923 valid annotations were compiled, and all genome statistics are summarized in Table 1.

Data availability. All data are available under the National Center for Biotechnology Information BioProject accession number [PRJNA748063](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA748063). The raw reads are available under the Sequence Read Archive accession number [SRX11510989](https://www.ncbi.nlm.nih.gov/sra/SRX11510989); the assembled nuclear genome sequence is available under the GenBank assembly accession number [GCA_019456465.1](https://www.ncbi.nlm.nih.gov/genbank/GCA_019456465.1).

ACKNOWLEDGMENTS

Funding for the sequencing was supported, in part, through the CanSeq150 program of Canada's Genomics Enterprise (<https://www.cgen.ca/>). Partial support for J.L.A. came from the U.S. National Science Foundation (Division of Environmental Biology award number 2115191).

We acknowledge the technical expertise of the scientists at Canada's Michael Smith Genome Sciences Centre in the generation of the DNA sequence data. We also thank Roger Bull and Anna Ginter for extracting the DNA.

REFERENCES

1. McMullin RT, McCune M, Lendemer JC. 2020. *Bacidia gigantensis* (Ramalinaceae), a new species with homosekikaic acid from the north shore of Lake Superior in Ontario, Canada. *Bryologist* 123:215–224. <https://doi.org/10.1639/0007-2745-123.2.215>.
2. Alexander PJ, Rajanikanth G, Bacon CD, Bailey CD. 2006. Recovery of plant DNA using a reciprocating saw and silica-based columns. *Mol Ecol Notes* 7:5–9. <https://doi.org/10.1111/j.1471-8286.2006.01549.x>.
3. 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>.
4. Roach MJ, Schmidt SA, Borneman AR. 2018. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* 19:460. <https://doi.org/10.1186/s12859-018-2485-7>.
5. Kolmogorov M, Yuan J, Lin Y, Pevzner P. 2019. Assembly of long error-prone reads using repeat graphs. *Nat Biotechnol* 37:540–546. <https://doi.org/10.1038/s41587-019-0072-8>.
6. Vaser R, Sovic I, Nagarajan N, Šikic M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
7. McKenzie SK, Walston RF, Allen JL. 2020. Complete, high-quality genomes from long-read metagenomic sequencing of two wolf lichen thalli reveals enigmatic genome architecture. *Genomics* 112:3150–3156. <https://doi.org/10.1016/j.ygeno.2020.06.006>.
8. 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>.
9. Seppy 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.