



Draft Genome Sequence of *Mortierella alpina* Strain LL118, Isolated from an Aspen (*Populus tremuloides*) Leaf Litter Sample

 Shu Yang,^a  Boris A. Vinatzer^a

^aSchool of Plant and Environmental Sciences, Virginia Tech, Blacksburg, Virginia, USA

ABSTRACT *Mortierella alpina* is a filamentous fungus commonly associated with soil and is one of very few fungal species known to include strains with ice nucleation activity. Here, we report the draft genome sequence of the ice nucleation-active *M. alpina* strain LL118, isolated from aspen leaf litter collected in Alberta, Canada.

Mortierella species are filamentous fungi in the subphylum Mortierellomycotina (1). *Mortierella alpina* is the only *Mortierella* species that includes strains with ice nucleation activity (INA) (2, 3). Since INA plays a key role in atmospheric processes (4, 5), sequencing the genome of a *M. alpina* strain with INA is an important step toward understanding the possible contribution of *M. alpina* to atmospheric processes. This paper reports the genome assembly and transcriptome sequencing (RNA-seq)-based annotation of *M. alpina* strain LL118, a strain with INA that was isolated from an aspen (*Populus tremuloides*) leaf sample collected in Alberta, Canada (6).

Genomic DNA was extracted from mycelium grown in potato dextrose broth for 14 days at 28°C using the ZymoBIOMICS DNA miniprep kit (Zymo Research). Whole-genome sequencing was performed on an Illumina HiSeq 3000 platform at the Iowa State University DNA Facility with the Nextera DNA Flex library prep kit (Illumina), and the quality of reads was checked using FastQC v0.11.9 (7). All tools were run with default parameters unless otherwise specified. A total of 9.8 million 2×100 -bp reads were generated. The reads were trimmed using Trimmomatic v0.39 (8) to remove adapters. *De novo* assembly was performed using SPAdes v3.15.2 (9) with `-k 21,33,55,77 --careful --cov-cutoff auto`. The completeness of the assembly was assessed using Benchmarking Universal Single-Copy Orthologs (BUSCO) v5.0.0 (10) with `-m genome -l mucoromycota_odb10`.

The genome assembly was found to have $54\times$ coverage using BMAP v38.90 (11). The genome was determined to comprise 35,609,760 bp with a G+C content of 51.44%. The BUSCO assessment revealed the presence of 1,575 genes (97.6%) based on the lineage-specific profile library mucoromycota_odb10 (1,614 genes). The assembly statistics are shown in Table 1.

Total RNA was extracted from mycelium grown on potato dextrose agar for 30 days at 6°C, room temperature, and 28°C (three replicates for each temperature) using the RNeasy plant minikit (Qiagen). RNA-seq was performed on an Illumina Nova Seq 6000 platform at Novogene Corporation, Inc. (Sacramento, CA), using the TruSeq stranded mRNA library prep kit (Illumina), and the quality of reads was again checked using FastQC v0.11.9 (7). A total of 197.4 million 2×150 -bp reads were generated. The RNA-seq reads were aligned to the genome assembly using STAR v2.7.8a (12) and assembled using Trinity v2.12.0 (13) and StringTie v2.1.5 (14).

The MAKER v3.01.03 pipeline (15) with SNAP v2013-02-16 (16) and AUGUSTUS v3.4.0 *ab initio* gene predictions (17) were employed for genome annotation using the assembled transcriptomes obtained above and the proteome of *Linnemannia elongata* (ID: UP000078512) from the UniProt database (18, 19). Functional annotations were performed using InterProScan v5.46-81.0 (20) and BLASTP from BLAST v2.10.0+ (21). The completeness of the genome annotation was assessed using BUSCO v5.0.0 (10) with `-m proteins -l mucoromycota_odb10`.

Editor Vincent Bruno, University of Maryland School of Medicine

Copyright © 2021 Yang and Vinatzer. 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 Shu Yang, shuyang@vt.edu.

Received 29 August 2021

Accepted 3 November 2021

Published 24 November 2021

TABLE 1 Assembly summary and annotation features of *Mortierella alpina* strain LL118

Characteristic	Value ^a
Strain	F156N33
Assembly size (bp)	35,609,760
No. of contigs	2,222
Maximum contig length (bp)	504,467
Minimum contig length (bp)	200
Avg contig length (bp)	16,169
N_{50} contig length (bp)	96,689
GC content (%)	51.44
Assembly BUSCO coverage (%)	C: 97.6; F: 0.7; M: 1.7
Annotation BUSCO coverage (%)	C: 91.7; F: 2.1; M: 6.2
No. of predicted coding genes	8,852
Mean gene length (bp)	3,205

^aBUSCO coverage: C, complete BUSCOs; F, fragmented BUSCOs; M, missing BUSCOs.

The BUSCO assessment revealed the presence of 1,480 complete genes (91.7%) and 34 fragmented genes (2.1%) (Table 1), indicating a good-quality genome annotation. A total of 8,852 protein-coding genes were predicted. Of the predicted proteins, 7,288 (82.3%) received a functional annotation.

Data availability. The draft genome sequence has been deposited in DDBJ/ENA/GenBank under accession number [JAIFTL000000000.1](https://accession.gtrdb.org/acc/acc/JAIFTL000000000.1), BioSample accession number [SAMN20056918](https://www.ncbi.nlm.nih.gov/biosample/SAMN20056918), and BioProject accession number [PRJNA743604](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA743604). The raw DNA sequences have been deposited in the Sequence Read Archive (SRA) under accession number [SRR15443011](https://www.ncbi.nlm.nih.gov/sra/SRR15443011), and the raw RNA sequences can be found under accession numbers [SRR16290171](https://www.ncbi.nlm.nih.gov/sra/SRR16290171), [SRR16290170](https://www.ncbi.nlm.nih.gov/sra/SRR16290170), and [SRR16290169](https://www.ncbi.nlm.nih.gov/sra/SRR16290169).

ACKNOWLEDGMENTS

Funding was provided by the National Science Foundation (grant IOS-1754721). Funding to B. A. Vinatzer was also provided in part by the Virginia Agricultural Experiment Station and the Hatch Program of the National Institute of Food and Agriculture, U.S. Department of Agriculture.

REFERENCES

- Spatafora JW, Chang Y, Benny GL, Lazarus K, Smith ME, Berbee ML, Bonito G, Corradi N, Grigoriev I, Gryganskyi A, James TY, O'Donnell K, Roberson RW, Taylor TN, Uehling J, Vilgalys R, White MM, Stajich JE. 2016. A phylum-level phylogenetic classification of zygomycete fungi based on genome-scale data. *Mycologia* 108:1028–1046. <https://doi.org/10.3852/16-042>.
- Pummer BG, Budke C, Augustin-Bauditz S, Niedermeier D, Felgitsch L, Kampf CJ, Huber RG, Liedl KR, Loerting T, Moschen T, Schauperl M, Tollinger M, Morris CE, Wex H, Grothe H, Pöschl U, Koop T, Fröhlich-Nowoisky J. 2015. Ice nucleation by water-soluble macromolecules. *Atmos Chem Phys* 15: 4077–4091. <https://doi.org/10.5194/acp-15-4077-2015>.
- Fröhlich-Nowoisky J, Hill TCJ, Pummer BG, Yordanova P, Franc GD, Pöschl U. 2015. Ice nucleation activity in the widespread soil fungus *Mortierella alpina*. *Biogeosciences* 12:1057–1071. <https://doi.org/10.5194/bg-12-1057-2015>.
- Murray BJ, O'Sullivan D, Atkinson JD, Webb ME. 2012. Ice nucleation by particles immersed in supercooled cloud droplets. *Chem Soc Rev* 41: 6519–6554. <https://doi.org/10.1039/c2cs35200a>.
- Tan I, Storelvmo T, Zelinka MD. 2016. Observational constraints on mixed-phase clouds imply higher climate sensitivity. *Science* 352:224–227. <https://doi.org/10.1126/science.aad5300>.
- Vasebi Y, Mechan Llontop ME, Hanlon R, Schmale DG, III, Schnell R, Vinatzer BA. 2019. Comprehensive characterization of an aspen (*Populus tremuloides*) leaf litter sample that maintained ice nucleation activity for 48 years. *Biogeosciences* 16:1675–1683. <https://doi.org/10.5194/bg-16-1675-2019>.
- Andrews S, Krueger F, Segonds-Pichon A, Biggins L, Krueger C, Wingett S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>. Accessed 6 April 2021.
- 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>.
- Prijbelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. 2020. Using SPAdes *de novo* assembler. *Curr Protoc Bioinformatics* 70:e102. <https://doi.org/10.1002/cpbi.102>.
- 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.
- Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Lawrence Berkeley National Lab (LBNL), Berkeley, CA.
- Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29:15–21. <https://doi.org/10.1093/bioinformatics/bts635>.
- Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, MacManes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, LeDuc RD, Friedman N, Regev A. 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat Protoc* 8:1494–1512. <https://doi.org/10.1038/nprot.2013.084>.
- Perteau M, Perteau GM, Antonescu CM, Chang T-C, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol* 33:290–295. <https://doi.org/10.1038/nbt.3122>.
- Campbell MS, Holt C, Moore B, Yandell M. 2014. Genome annotation and curation using MAKER and MAKER-P. *Curr Protoc Bioinformatics* 48: 4.11.1–11.39. <https://doi.org/10.1002/0471250953.bi0411s48>.
- Korf I. 2004. Gene finding in novel genomes. *BMC Bioinformatics* 5:59. <https://doi.org/10.1186/1471-2105-5-59>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntetically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <https://doi.org/10.1093/bioinformatics/btn013>.
- Uehling J, Gryganskyi A, Hameed K, Tschaplinski T, Misztal PK, Wu S, Desirò A, Vande Pol N, Du Z, Zienkiewicz A, Zienkiewicz K, Morin E, Tisserant E, Splivallo

- R, Hainaut M, Henrissat B, Ohm R, Kuo A, Yan J, Lipzen A, Nolan M, LaButti K, Barry K, Goldstein AH, Labbé J, Schadt C, Tuskan G, Grigoriev I, Martin F, Vilgalys R, Bonito G. 2017. Comparative genomics of *Mortierella elongata* and its bacterial endosymbiont *Mycobacterium cysteinexigens*. *Environ Microbiol* 19: 2964–2983. <https://doi.org/10.1111/1462-2920.13669>.
19. UniProt Consortium. 2021. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
 20. 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, Scheremetjew 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>.
 21. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.