Draft Genome Sequence of *Mortierella alpina* Isolate CDC-B6842

Kizee A. Etienne,[a] Marcus C. Chibucos,[b,c] Qi Su,[b] Joshua Orvis,[b] Sean Daugherty,[b] Sandra Ott,[b] Naomi A. Sengamalay,[b] Claire M. Fraser,[b] Shawn R. Lockhart,[a] Vincent M. Bruno[b,c]

Centers for Disease Control and Prevention, Atlanta, Georgia, USA[a]; Institute for Genome Sciences[b] and Department of Microbiology and Immunology,[c] University of Maryland School of Medicine, Baltimore, Maryland, USA

We report the draft genome sequence of *Mortierella alpina* isolate CDC-B6842. *M. alpina* is a nonpathogenic member of the *Mucoromycotina* subphylum of fungi that is an important model for understanding the molecular mechanisms of lipid production and metabolism.

Address correspondence to Vincent M. Bruno, vbruno@umaryland.edu.

Fungi belonging to the subphylum *Mucoromycotina* consist of a diverse group of organisms with industrial, biotechnological, and clinical relevance. They have been studied for their use in commercial applications, such as with biodiesel and other lipid-based products (1, 2). *Mortierella alpina*, a saprophytic member of *Mucoromycotina*, has been considered an important oleaginous model organism for understanding lipid metabolism and production. Unlike other genera of *Mucoromycotina*, *Mortierella* spp. are not known to cause disease in plants, animals, or humans. Sequencing the genomes of several isolates from this diverse group of fungi will allow us to understand the evolution of pathogenicity among the fungi within this subphylum.

M. alpina strain CDC-B6842 was isolated from the skin of a human leg lesion in Minnesota in 2004 but was demonstrated to be nonpathogenic. Genomic DNA was extracted using the Gene-Rite kit (Carlsberg, CA), according to the manufacturer's instructions. The genomic DNA was sequenced at the University of Maryland School of Medicine, Institute for Genome Sciences, Genomics Resource Center (http://www.igs.umaryland.edu) using a combination of paired-end libraries (average insert of 483 bp) and mate-pair (3-kb) libraries on the Illumina HiSeq 2000. A total of 44.5 million 100-bp reads were generated. The draft genome data were assembled using the MaSuRCA genome assembler (3). The resulting genome assembly contained 1,185 contigs with an average size of 33,358 bp. This resulted in a predicted genome size of 39.53 Mb with a G+C content of 50.4%. Both the estimated genome size and G+C content are consistent with those calculated for *M. alpina* strain ATCC 32222 (2).

Structural and functional predictions were done using the Institute for Genome Sciences (IGS) eukaryotic annotation pipeline protocol 1.0 at the Institute for Genome Sciences, Informatics Resource Center (http://www.igs.umaryland.edu). Briefly, repeat annotation was performed using RepeatModeler (http://www.repeatmasker.org) and RepeatMasker (4). Genes were predicted *ab initio* using four gene prediction programs: CEGMA (5), GeneMark-ES (6), Augustus (7), and SNAP (8). Augustus and SNAP used CEGMA predictions for parameter training. Spliced alignments of Swiss-Prot protein models against the *M. alpina*

genome were generated with AAT (9), using cutoffs of 60% identity and 80% similarity. All structural evidence was combined using EVidenceModeler (10). A total of 9,666 protein-coding genes, 233 tRNA genes, and 10 rRNA genes were predicted from this pipeline. While our estimation of the number of tRNA genes is consistent with what was reported for strain ATCC 32222, we estimate fewer protein-coding genes (2).

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. AZCI00000000. The version described in this paper is the first version, AZCI01000000.

REFERENCES

1. **Vongsangnak W, Ruenwai R, Tang X, Hu X, Zhang H, Shen B, Song Y, Laoteng K.** 2013. Genome-scale analysis of the metabolic networks of oleaginous zygomycete fungi. Gene **521:**180–190. http://dx.doi.org/10.1016/j.gene.2013.03.012.
2. **Wang L, Chen W, Feng Y, Ren Y, Gu Z, Chen H, Wang H, Thomas MJ, Zhang B, Berquin IM, Li Y, Wu J, Zhang H, Song Y, Liu X, Norris JS, Wang S, Du P, Shen J, Wang N, Yang Y, Wang W, Feng L, Ratledge C, Zhang H, Chen YQ.** 2011. Genome characterization of the oleaginous fungus *Mortierella alpina*. PLoS One **6:**e28319. http://dx.doi.org/10.1371/journal.pone.0028319.
3. **Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA.** 2013. The MaSuRCA genome assembler. Bioinformatics **29:**2669–2677. http://dx.doi.org/10.1093/bioinformatics/btt476.
4. **Tarailo-Graovac M, Chen N.** 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics **25:**4.10.1–4.10.14. http://dx.doi.org/10.1002/0471250953.bi0410s25.
5. **Parra G, Bradnam K, Korf I.** 2007. CEGMA: a pipeline to accurately

annotate core genes in eukaryotic genomes. Bioinformatics **23:** 1061–1067. http://dx.doi.org/10.1093/bioinformatics/btm071.

6. **Ter-Hovhannisyan V, Lomsadze A, Chernoff YO, Borodovsky M.** 2008. Gene prediction in novel fungal genomes using an *ab initio* algorithm with unsupervised training. Genome Res. **18:**1979–1990. http://dx.doi.org/10.1101/gr.081612.108.

7. **Stanke M, Morgenstern B.** 2005. Augustus: a web server for gene prediction in eukaryotes that allows user-defined constraints. Nucleic Acids Res. **33:**W465–W467. http://dx.doi.org/10.1093/nar/gki458.

8. **Korf I.** 2004. Gene finding in novel genomes. BMC Bioinformatics **5:**59. http://dx.doi.org/10.1186/1471-2105-5-59.

9. **Huang X, Adams MD, Zhou H, Kerlavage AR.** 1997. A tool for analyzing and annotating genomic sequences. Genomics **46:**37–45. http://dx.doi.org/10.1006/geno.1997.4984.

10. **Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR.** 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. **9:**R7. http://dx.doi.org/10.1186/gb-2008-9-1-r7.