

Draft Genome Sequence of the Cellulolytic Fungus *Chaetomium globosum*

Christina A. Cuomo,^a Wendy A. Untereiner,^b Li-Jun Ma,^{a,c} Manfred Grabherr,^{a,d} Bruce W. Birren^a

Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA^a; Brandon University, Brandon, Manitoba, Canada^b; University of Massachusetts, Amherst, Massachusetts, USA^c; Science for Life Laboratory, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden^d

***Chaetomium globosum* is a filamentous fungus typically isolated from cellulosic substrates. This species also causes superficial infections of humans and, more rarely, can cause cerebral infections. Here, we report the genome sequence of *C. globosum* isolate CBS 148.51, which will facilitate the study and comparative analysis of this fungus.**

Received 7 January 2015 Accepted 13 January 2015 Published 26 February 2015

Citation Cuomo CA, Untereiner WA, Ma L-J, Grabherr M, Birren BW. 2015. Draft genome sequence of the cellulolytic fungus *Chaetomium globosum*. *Genome Announc* 3(1): e00021-15. doi:10.1128/genomeA.00021-15.

Copyright © 2015 Cuomo et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](http://creativecommons.org/licenses/by/3.0/).

Address correspondence to Christina A. Cuomo, cuomo@broadinstitute.org, or Wendy Untereiner, untereiner@brandonu.ca.

Chaetomium (family *Chaetomiaceae*, class *Sordariomycetes*, phylum *Ascomycota*) is a genus of filamentous fungi encompassing species important in the decomposition of plant and other cellulose-rich materials that can be isolated easily from dung, plant debris, and soil. *Chaetomium* includes thermophilic species capable of growth at elevated temperatures as well as a few species that cause infections in vertebrates. *Chaetomium globosum*, the type species of the genus, is commonly isolated from decaying plant material, seeds, and other cellulosic substrates. Of the more than 150 species of *Chaetomium* described to date, *C. globosum* is the most frequently isolated and inhabits the widest range of environments (1).

Chaetomium globosum is a medically important fungus and is encountered typically as an agent of skin and nail infections in humans. This species more rarely causes cerebral and systemic infections, but such mycoses have high mortality rates, particularly in immunocompromised patients (2). *Chaetomium globosum* is important to human health as a contaminant in indoor environments since it is known to produce mycotoxins (3, 4) and act as an allergen.

The sequenced strain CBS 148.51 was isolated from stored cotton in Washington, DC. This strain is commonly used in testing paper and polymers for fungal resistance (5). Three size-selected libraries were constructed from genomic DNA. These included a 4-kb plasmid, a 10-kb plasmid, and a 40-kb fosmid library; each library was paired-end sequenced using Sanger technology. The resulting 568,566 reads were assembled using Arachne version 3.0, with an average of 8.9× sequence depth in the final assembly.

Based on the assembly, the genome size was estimated to be 34.3 Mb with a GC content of 55.6%. The assembly was organized in 1,245 contigs, which are linked by paired-end reads into 37 scaffolds. The average base is found in a scaffold of N_{50} size 4.72 Mb and a contig of N_{50} size 50.76 kb. The assembly is highly contiguous; the largest 8 scaffolds account for 98% of the assembly bases. For gene prediction, Fgenesh, Fgenesh+, and Genewise, which were previously trained for the related species *Neurospora crassa* (6), were used. A total of 11,124 protein-coding genes were

predicted by aligning publicly available expressed sequence tags to the genome and combining those alignments with the output from Fgenesh, Fgenesh+, Genewise, and GeneId. A total of 5.57% of the assembly was identified as interspersed repeats using RepeatMasker (<http://www.repeatmasker.org>) to identify copies of *de novo* repeats identified by RepeatModeler (<http://www.repeatmasker.org/RepeatModeler.html>) and fungal sequences from RepBase (7).

This genome sequence of *C. globosum* will serve as an important reference for further studies of the basis of its cellulose specificity, for genes that enable human infection and for further comparative studies with other fungi.

Nucleotide sequence accession number. The whole-genome sequence and annotation of *C. globosum* isolate CBS 148.51 (ATCC 6205) have been deposited at DDBJ/EMBL/GenBank under the accession number [AAFU00000000](https://www.ncbi.nlm.nih.gov/nuccore/AAFU00000000).

ACKNOWLEDGMENTS

We thank the Broad Institute Genomics Platform for generating all DNA sequence described here and David DeCaprio for running gene prediction methods.

This work was supported by the NHGRI.

REFERENCES

- Domsch KH, Gams W, Anderson TH. 2007. *Compendium of soil fungi*, 2nd ed. Academic Press, London, United Kingdom.
- Serena C, Ortoneda M, Capilla J, Pastor FJ, Sutton DA, Rinaldi MG, Guarro J. 2003. *In vitro* activities of new antifungal agents against *Chaetomium* spp. and inoculum standardization. *Antimicrob Agents Chemother* 47:3161–3164. <http://dx.doi.org/10.1128/AAC.47.10.3161-3164.2003>.
- Nielsen KF, Gravesen S, Nielsen PA, Andersen B, Thrane U, Frisvad JC. 1999. Production of mycotoxins on artificially and naturally infested building materials. *Mycopathologia* 145:43–56. <http://dx.doi.org/10.1023/A:1007038211176>.
- Pieckova E. 2003. *In vitro* toxicity of indoor *Chaetomium* Kunze ex Fr. *Ann Agric Environ Med* 10:9–14.
- Gu J-G, Gu J-D. 2005. Methods currently used in testing microbiological degradation and deterioration of a wide range of polymeric materials with various degree of degradability: a review. *J Polym Environ* 13:65–74. <http://dx.doi.org/10.1007/s10924-004-1230-7>.

6. Galagan JE, Calvo SE, Borkovich KA, Selker EU, Read ND, Jaffe D, FitzHugh W, Ma LJ, Smirnov S, Purcell S, Rehman B, Elkins T, Engels R, Wang S, Nielsen CB, Butler J, Endrizzi M, Qui D, Ianakiev P, Bell-Pedersen D, Nelson MA, Werner-Washburne M, Selitrennikoff CP, Kinsey JA, Braun EL, Zelter A, Schulte U, Kothe GO, Jedd G, Mewes W, Staben C, Marcotte E, Greenberg D, Roy A, Foley K, Naylor J, Stange-Thomann N, Barrett R, Gnerre S, Kamal M, Kamvyselis M, Mauceli E, Bielke C, Rudd S, Frishman D, Krystofova S, Rasmussen C, Metzenberg RL, Perkins DD, Kroken S, Cogoni C, Macino G, Catchside D, Li W, Pratt RJ, Osmani SA, DeSouza CP, Glass L, Orbach MJ, Berglund JA, Voelker R, Yarden O, Plamann M, Seiler S, Dunlap J, Radford A, Aramayo R, Natvig DO, Alex LA, Mannhaupt G, Ebbole DJ, Freitag M, Paulsen I, Sachs MS, Lander ES, Nusbaum C, Birren B. 2003. The genome sequence of the filamentous fungus *Neurospora crassa*. *Nature* 422: 859–868. <http://dx.doi.org/10.1038/nature01554>.
7. Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* 110:462–467. <http://dx.doi.org/10.1159/000084979>.