

Draft Genome Sequence of *Geotrichum candidum* Strain 3C

 Dmitrii E. Polev,^a Kirill S. Bobrov,^b Elena V. Eneyskaya,^b Anna A. Kulminskaya^{b,c}

Research Resource Center for Molecular and Cell Technologies, St. Petersburg State University, St. Petersburg, Russia^a; National Research Centre, Kurchatov Institute, B.P. Konstantinov Petersburg Nuclear Physics Institute, Gatchina, Orlova Roscha, Russia^b; St. Petersburg State Polytechnical University, St. Petersburg, Russia^c

We report here the draft genome sequence of *Geotrichum candidum* strain 3C, which is a filamentous yeast-like fungus that holds great promise for biotechnology. The genome was sequenced using Ion Torrent and 454 platforms. The estimated genome size was 41.4 Mb, and 14,579 protein-coding genes were predicted *ab initio*.

Received 22 August 2014 Accepted 2 September 2014 Published 2 October 2014

Citation Polev DE, Bobrov KS, Eneyskaya EV, Kulminskaya AA. 2014. Draft genome sequence of *Geotrichum candidum* strain 3C. *Genome Announc.* 2(5):e00956-14. doi:10.1128/genomeA.00956-14.

Copyright © 2014 Polev 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 Dmitrii E. Polev, dmitry.polev@gmail.com.

Geotrichum candidum is a filamentous yeast-like fungus. The cellulolytic enzymes from *G. candidum* attract the most interest due to their potential applications in many fields such as protoplast generation and the textile, paper, and detergent industries (1, 2). In recent years, growing attention has been drawn to fuel ethanol production by using cellulases (3). Conversely, *G. candidum* and related species may degrade various natural and artificial materials. A species close to *G. candidum* was found to be responsible for biodeterioration in compact discs (CDs), destroying the information pits (4). The *G. candidum* strain 3C itself was found and isolated from a rotting rope (5). Previous studies demonstrated that this strain possesses high cellulolytic and xylanolytic activities (6, 7). So, this fungus holds great promise for biotechnology; however, there are no genome sequences of *G. candidum* strains available at the present time.

We sequenced the whole genome of *G. candidum* strain C3 using Roche 454 and Ion Torrent (Life Technologies) platforms. The DNA library for Roche 454 GS Junior was prepared using GS Rapid Library Prep Kit and sequenced with the GS Junior Titanium sequencing kit. The amount of raw data was more than 84 Mb in 182,317 reads with an average read length of 460 bp. This appeared to be insufficient for the *de novo* genome assembly, and we sequenced the genomic DNA using Ion Torrent PGM with an Ion Plus fragment library kit, an Ion PGM sequencing 400 kit, and Ion 318 Chip v2 to get additional information. With the Ion PGM instrument we obtained 1.3-G total bases in 4.39 million reads with a median read length of 343 bp. Both datasets were used for the *de novo* genome assembly using MIRA 4.0 (8). Genome assembly resulted in 560 large contigs (500 bp or more) with a total consensus length of 41,384,521 bp, largest contig size of 1,363,582 bp, and N_{50} of 437,602 bp. The average total coverage was 31.25, and the G+C content was 46.96%. We masked the repetitive sequences using RepeatMasker version open-4.0.5 (<http://www.repeatmasker.org>), and used the self-training gene prediction software GeneMark-ES 2.0 (9) to create a training data set for Augustus. We used Augustus 3.0.2 (10) trained with the output data from GeneMark-ES for *ab initio* gene prediction. In this way we could identify 14,579 protein-coding genes.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank un-

der the accession no. [JMRO00000000](https://doi.org/10.1128/genomeA.00956-14). The version described in this paper is version JMRO01000000.

ACKNOWLEDGMENTS

The research was supported by Research Resource Center for Molecular and Cell Technologies, St. Petersburg State University, and the Russian Foundation for Basic Research (project 14-08-01041-a).

REFERENCES

- Kuhad RC, Gupta R, Singh A. 2011. Microbial cellulases and their industrial applications. *Enzyme Res.* 2011:280696. <http://dx.doi.org/10.4061/2011/280696>.
- Phitsuwan P, Laohakunjit N, Kerdchoechuen O, Kyu KL, Ratanakhanokchai K. 2013. Present and potential applications of cellulases in agriculture, biotechnology, and bioenergy. *Folia Microbiol. (Praha)* 58:163–176. <http://dx.doi.org/10.1007/s12223-012-0184-8>.
- Singh A, Pant D, Korres NE, Nizami A-S, Prasad S, Murphy JD. 2010. Key issues in life cycle assessment of ethanol production from lignocellulosic biomass: challenges and perspectives. *Bioresour. Technol.* 101:5003–5012. <http://dx.doi.org/10.1016/j.biortech.2009.11.062>.
- Garcia-Guinea J, Cárdenes V, Martínez AT, Martínez MJ. 2001. Fungal bioturbation paths in a compact disk. *Naturwissenschaften* 88:351–354. <http://dx.doi.org/10.1007/s001140100249>.
- Rodionova NA, Tiunova NA, Feniksova RV, Kudriashova TI, Martynovich LI. 1974. Cellulolytic enzymes of *geotrichum candidum*. *Dokl. Akad. Nauk SSSR* 214:1206–1209. (In Russian.)
- Tiunova NA, Rodionova NA, Martinovich LI, Gogolev MN. 1980. Preparation of cellulolytic enzymes from *geotrichum candidum*. *Prikl. Biokhim. Mikrobiol.* 16:185–190. (In Russian.)
- Rodionova NA, Dubovaia NV, Eneiskaia EV, Martinovich LI, Gracheva IM, Bezborodov AM. 2000. Purification and characteristic of endo-(1–4)-beta-xylanase from *geotrichum candidum* 3C. *Prikl. Biokhim. Mikrobiol.* 36:535–540. (In Russian.)
- Chevreur B, Wetter T, Suhai S. 1999. Genome sequence assembly using trace signals and additional sequence information, p 45–56. *In* Computer science and biology: Proceedings of the German Conference on Bioinformatics (GCB) 99. Hannover, Germany.
- 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>.
- Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntentically mapped cDNA alignments to improve *de novo* gene finding. *Bioinformatics* 24:637–644. <http://dx.doi.org/10.1093/bioinformatics/btn013>.