Draft Genome Sequences of *Candida glabrata* Isolates 1A, 1B, 2A, 2B, 3A, and 3B

Othilde Elise Håvelsrud, Peter Gaustad

Department of Microbiology, Oslo University Hospital, Oslo, Norway

ABSTRACT Here, we report the draft genome sequences of six *Candida glabrata* isolates. The isolates were taken from blood samples from patients after recurrent *C. glabrata* infection. Two isolates were taken from each of three patients a minimum 3 months apart.

Address correspondence to Peter Gaustad, peter.gaustad@medisin.uio.no.

*C*andida glabrata is an important causal agent of candidemia in humans (1). A recent surveillance study from Norway, where our isolates were collected, identified *C. glabrata* as the most common causative agent of candidemia after *Candida albicans* (2). Here, we present the draft genome sequences of six isolates, two each from three patients with recurring candidemia. These draft genome sequences may help ascertain if the recurrent infections were caused by the same strain and whether the strain genomes evolve between infections. They also present a good opportunity for further investigation into the virulence of *C. glabrata*.

DNA was extracted using the YeaStar genomic DNA kit (Zymo Research), according to the manufacturer's protocol I. The DNA was purified and concentrated using the Amicon Ultra-0.5 ml 30K centrifugal filters for DNA purification and concentration (Merck Millipore). The DNA concentration was determined by the Qubit double-stranded DNA (dsDNA) high-sensitivity (HS) assay kit (Life Technologies). The DNA quality was assessed by agarose gel electrophoresis and 260/280 ratios measured on a NanoDrop 2000 spectrophotometer (Thermo Scientific). Sequencing was performed at the Genomics Core Facility at Radiumhospitalet (http://oslo.genomics.no) using Illumina MiSeq technology. Approximately 15.4, 15.7, 16.3, 24.1, 15.6, and 14.2 million paired-end reads of 101 nucleotides (nt) were generated for *C. glabrata* isolates 1A, 1B, 2A, 2B, 3A, and 3B, respectively. After removal of read pairs containing ambiguous bases, the reads were assembled with Velvet version 1.2.10 and AMOScmp-shortReads version 3.1.0 (3, 4). VelvetOptimiser was used to set optimal parameter values for the Velvet assemblies. Only contigs >1,000 nt were retained in the Velvet assemblies. *C. glabrata* CBS138 (5, 6) was used as a reference for the AMOScmp-shortReads assemblies. Feature response curves were created in AMOS version 3.1.10 in order to evaluate assembly quality (4). Contigs from the two assemblies were joined with GAM-NGS version 1.1b using the AMOScmp-shortReads assembly as a master (7). SSPACE version 3.0 was used to scaffold the contigs (8). Both the intermediate contigs and the final scaffolds were validated with REAPR version 1.0.17 (9). mothur 1.36.1 was used to filter out all scaffolds <1,000 nt (10). The final assemblies consist of 12.6 Mbp (119 scaffolds), 12.6 Mbp (144 scaffolds), 14.6 Mbp (196 scaffolds), 12.8 Mbp (132 scaffolds), 12.8 Mbp (189 scaffolds), and 12.9 Mbp (183 scaffolds) for 1A, 1B, 2A, 2B, 3A, and 3B, respectively. The completeness of the assemblies was assessed with CEGMA version 2.4.010312 (11). The draft genome sequences were estimated to be 96 to 97% complete. The average G+C content was 38.6% for all draft genome sequences. The final scaffolds were annotated with MAKER version 2.31 (12). A total of 5,214 (1A), 5,191 (1B), 5,985 (2A), 5,293 (2B), 5,324 (3A), and 5,331 (3B) predicted protein-coding genes were detected.

The predicted protein-coding genes cover approximately 65% of the draft genome sequences.

Accession number(s). These whole-genome shotgun projects have been deposited at DDBJ/EMBL/GenBank under the accession numbers LLWO00000000, LMAA00000000, LLZZ00000000, LLZY00000000, LMAY00000000, and LMAB00000000 for 1A, 1B, 2A, 2B, 3A, and 3B, respectively. The versions described in this paper are the first versions LLWO01000000, LMAA01000000, LLZZ01000000, LLZY01000000, LMAY01000000, and LMAB01000000, respectively.

REFERENCES

1. Guinea J. 2014. Global trends in the distribution of *Candida* species causing candidemia. Clin Microbiol Infect 20(Suppl 6):5–10. https://doi.org/10.1111/1469-0691.12539.
2. Hesstvedt L, Gaustad P, Andersen CT, Haarr E, Hannula R, Haukland HH, Hermansen NO, Larssen KW, Mylvaganam H, Ranheim TE, Sandven P, Nordøy I, Group tNYS, Kanestrøm A, Grub C, Onken A, Thielsen C, Skaare D, Tofteland S, Sønsteby L-J, Hjetland R, Hide R, Vik E, Kümmel A, Åsheim S. 2015. Twenty-two years of candidaemia surveillance: results from a Norwegian national study. Clin Microbiol Infect 21:938–945.
3. Schatz MC, Phillippy AM, Sommer DD, Delcher AL, Puiu D, Narzisi G, Salzberg SL, Pop M. 2013. Hawkeye and AMOS: visualizing and assessing the quality of genome assemblies. Brief Bioinform 14:213–224.
4. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
5. Dujon B, Sherman D, Fischer G, Durrens P, Casaregola S, Lafontaine I, De Montigny J, Marck C, Neuvéglise C, Talla E, Goffard N, Frangeul L, Aigle M, Anthouard V, Babour A, Barbe V, Barnay S, Blanchin S, Beckerich JM, Beyne E, Bleykasten C, Boisramé A, Boyer J, Cattolico L, Confanioleri F, De Daruvar A, Despons L, Fabre E, Fairhead C, Ferry-Dumazet H, Groppi A, Hantraye F, Hennequin C, Jauniaux N, Joyet P, Kachouri R, Kerrest A, Koszul R, Lemaire M, Lesur I, Ma L, Muller H, Nicaud JM, Nikolski M, Oztas S, Ozier-Kalogeropoulos O, Pellenz S, Potier S, Richard GF, Straub ML, et al. 2004. Genome evolution in yeasts. Nature 430:35–44. https://doi.org/10.1038/nature02579.
6. Koszul R, Malpertuy A, Frangeul L, Bouchier C, Wincker P, Thierry A, Duthoy S, Ferris S, Hennequin C, Dujon B. 2003. The complete mitochondrial genome sequence of the pathogenic yeast *Candida* (*Torulopsis*) *glabrata*. FEBS Lett 534:39–48. https://doi.org/10.1016/S0014-5793(02)03749-3.
7. Vicedomini R, Vezzi F, Scalabrin S, Arvestad L, Policriti A. 2013. GAM-NGS: genomic assemblies merger for next generation sequencing. BMC Bioinformatics 14(Suppl 7):S6. https://doi.org/10.1186/1471-2105-14-S7-S6.
8. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.
9. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol 14:R47. https://doi.org/10.1186/gb-2013-14-5-r47.
10. Schloss PD, Westcott SL, Ryabin T, Hall JR, Hartmann M, Hollister EB, Lesniewski RA, Oakley BB, Parks DH, Robinson CJ, Sahl JW, Stres B, Thallinger GG, Van Horn DJ, Weber CF. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. Appl Environ Microbiol 75:7537–7541. https://doi.org/10.1128/AEM.01541-09.
11. Parra G, Bradnam K, Korf I. 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071.
12. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Alvarado AS, Yandell M. 2008. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res 18:188–196. https://doi.org/10.1101/gr.6743907.