



De novo Nanopore Genome Sequencing of the Clinical *Diutina catenulata* Type-strain CBS565

Received: 14 January 2022 / Accepted: 2 April 2022 / Published online: 10 May 2022
© The Author(s) 2022

Abstract *Diutina catenulata* is an ascomycetous yeast, that is regularly fluconazole-resistant and increasingly reported as the cause of invasive infection in humans. Here, we describe the *de novo* genome assembly of the clinical *D. catenulata* type-strain CBS565 and provide insights into the genome and compared it to an Illumina-sequenced environmental strain.

Keywords *Diutina catenulata* · Antifungal resistant · Emerging pathogen · Long-read sequencing · Nanopore sequencing

Diutina catenulata is an ascomycetous yeast belonging to the *Saccharomycetales* and a member of the *Debaryomycetaceae/Metschnikowiaceae* clade [1, 2]. It was originally described in 1942 as *Candida catenulata* but recently assigned to the novel genus *Diutina* that presently accommodates six other species *D. neorugosa*, *D. pseudorugosa*, *D. ranogensis*, *D. rugosa*, *D. scorzettiae* and *D. siamensis* [1]. *Diutina* yeasts have been recognized as emerging pathogens [3, 4]. Molecular characterization and antifungal susceptibility testing of the clinically relevant *Diutina* species showed that *D. mesorugosa* was in fact indistinguishable from *D. rugosa*, thus the former was found to be a synonym for the latter [3]. *Diutina catenulata* has been mainly reported from birds and

contaminated cheeses but this species, which frequently display high fluconazole MIC-values is also increasingly reported as a cause of invasive infection in humans [4–6]. Here, we provide the *de novo* genome assembly and annotation of the clinical *D. catenulata* type-strain CBS565.

The *D. catenulata* type-strain CBS565 was isolated in 1926 from faeces of a patient with dysentery who at that time lived in Puerto Rico [1]. Genomic DNA extraction was performed as previously described [7]. The nanopore sequencing library preparation was performed with the ligation sequencing kit (SQK-LSK109; ONT, Oxford, UK) and the native barcoding kit (EXP-NBD114; ONT). The two libraries were run onto a MinION flow cell (FLO-MIN106; ONT) following the manufacturer's protocol.

The raw nanopore reads were basecalled using Guppy v5.0.16 + b9fcd7b5b (ONT) using the settings—flowcell FLO-MIN106—kit SQK-LSK109—barcode_kits EXP-NBD114—device cuda:0, followed by demultiplexing and barcode trimming with the same software. *De novo* genome assembly was performed with Flye v2.9 (<https://github.com/fenderglass/Flye>; [8]) using the parameters—nano-raw < fastq > —out-dir < directory > —genome-size 13 m.

The assembled genome quality was assessed using GenomeQC [9]. The total genome size was 14,464,696 bp with an N50 of 2,438,920 bp, distributed over 9 contigs (range 3,918,888–370,337 bp;

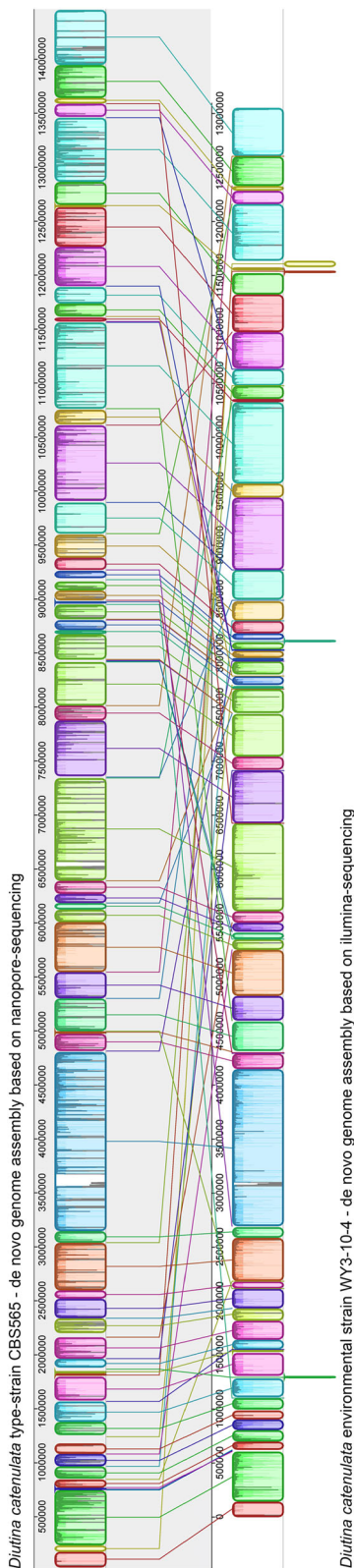


Fig. 1 Genome alignment of *Diutina catenulata* strains CBS565 and WY3-10-4. Mauve [18] was used with default LCB weight settings to visualize the de novo genome assembly of the nanopore-sequenced strain CBS565 (upper lane) versus that of the recently published de novo assembled illumina-sequenced strain WY3-10-4 [19] (lower lane). Homologues sequence-segments (coloured boxes) are connected by a line, and box that is downward indicates an inverted segment compared to the other genome. Numbers indicate sequence-length in base pairs

coverage of 107–138X) and a circular mitochondrial genome of 20,926 bp (3,528X coverage). Percentages of adenine, thymine, cytosine and guanine were 23.36, 23.37, 26.69 and 26.58%, respectively, resulting in a GC-content of 53.27%, which is close to 54.5% as previously determined by DNA–DNA-reassociation analyses [1, 10].

Annotation was performed using the Funannotate pipeline v1.8.9 (<https://github.com/nextgenusf/funannotate>). Ab initio gene prediction was done using GeneMark-ES v4 [11], Augustus v3.3.3 [12], SNAP [13] and GlimmerHMM v3.0.4 [14]. The latter three were trained with validated protein predictions found by BUSCO v2.0 [15]. Functional annotation was performed using InterProScan v5.52-86.0 [16] and SignalP v5.0 [17] as part of the Funannotate pipeline. The final annotated genome contained 5,209 predicted genes, consisting of 4918 mRNAs and 291 tRNAs. Among the 106 CAZymes predicted, 65 were observed by SignalP to be extracellular, for the 177 predicted proteases 65 were found to be extracellular.

We used Mauve [18] to compare the de novo assembly of CBS565 to that of the published genome of the environmental strain WY3-10-4 [19], which showed various structural variations (Fig. 1). The Average Nucleotide Identity (ANI) was calculated using Oat [20], this yielded an ANI of 98.06%.

With Orthofinder [21] the CBS565 and WY3-10-4 proteomes were clustered using the default parameters. In total 4577 orthologous clusters were built. A total of 129 unique putative proteins in CBS565 and 550 in WY3-10-4 were found. Some well-known virulence factors of *Candida albicans* (e.g., proteases, hyphal wall protein, agglutinin-like sequence protein) were used to identify virulence factors in CBS565 by BLAST [22]. At least one copy of these virulence factors were found in the genome of CBS565 (Supplementary Data).

Author contributions SB: performed research, analyzed the data. FM: conceived the study, wrote the manuscript. MZ: performed research, analyzed the data. BGE: performed research. FH: conceived the study, performed research, analyzed the data, wrote the manuscript.

Funding No funding was received for conducting this study.

Data transparency The data has been deposited in NCBI Genomes. BioProject accession number PRJNA770350,

BioSample accession number SAMN22215821, and sequencing reads with SRA accession numbers SRR16292071 and SRR16311897.


Declarations


Conflict of interest The authors declare that there are no other competing interests.


Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>


Author name and affiliation

^{1,2}Sander Boden 

³Florent Morio 

²Miaomiao Zhou 

¹Bert Gerrits van den Ende 

^{1,4}Ferry Hagen 

¹Westerdijk Fungal Biodiversity Institute, Uppsalalaan 8, 3584 CT Utrecht, The Netherlands
e-mail: f.hagen@wi.knaw.nl
e-mail: f.hagen@gmail.com

²Avans University of Applied Sciences, Breda, The Netherlands

³Nantes Université, CHU de Nantes, Cibles et Médicaments des Infections et de l'immunité, IICiMed, UR1155, 44000 Nantes, France

⁴Department of Medical Microbiology, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands

References

1. Khunnamwong P, Lertwattanasakul N, Jindamorakot S, Limtong S, Lachance MA. Description of *Diutina* gen. nov., *Diutina siamensis*, f.a. sp. nov., and reassignment of *Candida catenulata*, *Candida mesorugosa*, *Candida neorugosa*, *Candida pseudorugosa*, *Candida ranongensis*, *Candida rugosa* and *Candida scorzettiae* to the genus *Diutina*. Int J

- Syst Evol Microbiol. 2015;65(12):4701–4709. <https://doi.org/10.1099/ijsem.0.000634>.
2. Chapter 90. *Candida* (Berkhout). In: Kurtzman CP, Fell JW, Boekhout T, editors. The Yeasts: A Taxonomic Study. 5th edition. Amsterdam: Elsevier Publishers; 2011.
3. Ming C, Huang J, Wang Y, Lv Q, Zhou B, Liu T, Cao Y, Gerrits van den Ende B, Al-Hatmi AMS, Ahmed SA, Huang G, Bai F, de Hoog S, Kang Y. Revision of the medically relevant species of the yeast genus *Diutina*. Med Mycol. 2019;57(2):226–233. <https://doi.org/10.1093/mmy/myy001>.
4. Stavrou AA, Lackner M, Lass-Flörl C, Boekhout T. The changing spectrum of *Saccharomycotina* yeasts causing candidemia: phylogeny mirrors antifungal susceptibility patterns for azole drugs and amphotericin B. FEMS Yeast Res. 2019;19(4):foz037. <https://doi.org/10.1093/femsyr/foz037>.
5. Borman AM, Muller J, Walsh-Quantick J, Szekely A, Patterson Z, Palmer MD, Fraser M, Johnson EM. Fluconazole resistance in isolates of uncommon pathogenic yeast species from the United Kingdom. Antimicrob Agents Chemother. 2019;63(8):e00211–e219. <https://doi.org/10.1128/AAC.00211-19>.
6. Chen XF, Zhang W, Fan X, Hou X, Liu XY, Huang JJ, Kang W, Zhang G, Zhang H, Yang WH, Li YX, Wang JW, Guo DW, Sun ZY, Chen ZJ, Zou LG, Du XF, Pan YH, Li B, He H, Xu YC. Antifungal susceptibility profiles and resistance mechanisms of clinical *Diutina catenulata* isolates with high MIC values. Front Cell Infect Microbiol. 2021;11:739496. <https://doi.org/10.3389/fcimb.2021.739496>.
7. de Jong AW, Francisco EC, de Almeida JN, Brandão IB Jr, Pereira FM, Dias PHP, de Miranda Costa MM, de Souza Jordão RT, Vu D, Colombo AL, Hagen F. Nanopore genome sequencing and variant analysis of the susceptible *Candida auris* strain L1537/2020, Salvador, Brazil Mycopathologia. 2021;186(6):883–7. <https://doi.org/10.1007/s11046-021-00593-7>.
8. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019;37(5):540–6. <https://doi.org/10.1038/s41587-019-0072-8>.
9. Manchanda N, Portwood JL 2nd, Woodhouse MR, Seetharam AS, Lawrence-Dill CJ, Andorf CM, Hufford MB. GenomeQC: a quality assessment tool for genome assemblies and gene structure annotations. BMC Genomics. 2020;21(1):193. <https://doi.org/10.1186/s12864-020-6568-2>.
10. Stenderup A, Bak AL. Deoxyribonucleic acid base composition of some species within the genus *Candida*. J Gen Microbiol. 1968;52:231–6. <https://doi.org/10.1099/00221287-52-2-231>.
11. Lomsadze A, Ter-Hovhannisyan V, Chernoff YO, Borodovsky M. Gene identification in novel eukaryotic genomes by self-training algorithm. Nucleic Acids Res. 2005;33(20):6494–506. <https://doi.org/10.1093/nar/gki937>.
12. Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics. 2011;27(6):757–63. <https://doi.org/10.1093/bioinformatics/btr010>.
13. Korf I. Gene finding in novel genomes. BMC Bioinformatics. 2004;5:59. <https://doi.org/10.1186/1471-2105-5-59>.

14. Delcher AL, Bratke KA, Powers EC, Salzberg SL. Identifying bacterial genes and endosymbiont DNA with Glimmer. *Bioinformatics*. 2007;23(6):673–9. <https://doi.org/10.1093/bioinformatics/btm009>.
15. Manni M, Berkeley MR, Seppey M, Simão FA, Zdobnov EM. BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol*. 2021;38(10):4647–54. <https://doi.org/10.1093/molbev/msab199>.
16. Blum M, Chang H, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-Lafosse T, Qureshi M, Raj S, Richardson L, Salazar GA, Williams L, Bork P, Bridge A, Gough J, Haft DH, Letunic I, Marchler-Bauer A, Mi H, Natale DA, Necci M, Orengo CA, Pandurangan AP, Rivoire C, Sigrist CJA, Sillitoe I, Thanki N, Thomas PD, Tosatto SCE, Wu CH, Bateman A, Finn RD. The InterPro protein families and domains database: 20 years on. *Nucleic Acids Res*. 2021;49(D1):D344–54. <https://doi.org/10.1093/nar/gkaa977>.
17. Almagro Armenteros JJ, Tsirigos KD, Sønderby CK, Petersen TN, Winther O, Brunak S, von Heijne G, Nielsen H. SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat Biotechnol*. 2019;37:420–3. <https://doi.org/10.1038/s41587-019-0036-z>.
18. Darling AE, Mau B, Perna NT. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. *PLoS ONE*. 2010;5(6): e11147. <https://doi.org/10.1371/journal.pone.0011147>.
19. O'Brien CE, McCarthy CGP, Walshe AE, Shaw DR, Sumski DA, Krassowski T, Fitzpatrick DA, Butler G. Genome analysis of the yeast *Diutina catenulata*, a member of the Debaryomycetaceae/Metschnikowiaceae (CTG-Ser) clade. *PLoS ONE*. 2018;13(6): e0198957. <https://doi.org/10.1371/journal.pone.0198957>.
20. Lee I, Kim YO, Park SC, Chun J. OrthoANI: An improved algorithm and software for calculating average nucleotide identity. *Int J Syst Evol Microbiol*. 2015;66:1100–3. <https://doi.org/10.1099/ijsem.0.000760>.
21. Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;20:238. <https://doi.org/10.1186/s13059-019-1832-y>.
22. Mba IE, Nweze EI. Mechanism of *Candida* pathogenesis: revisiting the vital drivers. *Eur J Clin Microbiol Infect Dis*. 2020;39:1797–819. <https://doi.org/10.1007/s10096-020-03912-w>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.