AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

| Mycology | Announcement

Genome sequences of six clinical isolates of *Candida parapsilosis* exhibiting different degrees and temporal regulation of biofilm formation

Sulman Shafeeq,[1] Srisuda Pannanusorn,[1,2] Jacques Dainat,[3] Ali Dadvar,[1] Christian Tellgren-Roth,[4] Bengt Sennblad,[5] Björn Nystedt,[5] Ute Römling[1]

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Candida parapsilosis* is a major pathogen causing central venous catheter–associated bloodstream infections with biofilm formation as virulence factor. We sequenced the genomes of six *C. parapsilosis* isolates from bloodstream infections displaying no, low, and high biofilm under conditions mimicking the clinical setting.

KEYWORDS *Candida parapsilosis*, PacBio RSII system, biofilm formation, clinical isolates

Bloodstream infection caused by *Candida* species is a primary cause of morbidity and mortality in intensive care units worldwide (1–9). Thereby, the environmental species *C. parapsilosis* is the second/third most frequently isolated species (1, 10) causing superficial to invasive infections (1, 4). The ability to grow in parenteral nutrition with high glucose to form biofilms on catheters and high transmissibility contribute to the infection potential (11–13).

Here, we report whole-genome sequencing of six Swedish *C. parapsilosis* bloodstream isolates (14–16) with no (SMI 416, SMI 798), low (SMI 706 with high initial adherence), and high (SMI 588, SMI 596, SMI 828 [hospital outbreak strain (12)]) biofilm-forming ability (Fig. 1A).

Strains maintained at −80°C were plated on YPD (1% yeast extract, 2% peptone, and 2% glucose) agar plates and incubated at 37°C for 48 h. Subsequently, isolates were grown overnight in 50 mL YPD medium at 37°C with shaking, and genomic DNA was isolated using the Genomic-tip 500/G columns (QIAGEN) following the manufacturer's instructions. RNA contamination and DNA integrity were investigated on a 0.5% agarose gel (1×TAE buffer), and quantity and purity were assessed by 260/280 and 260/230 ratios measured by NanoDrop 2000c (Thermo Scientific). Genomic DNA was sheared with Megaruptor 2, removing fragments below 3 kbp with SPRI beads. SMRTbell libraries prepared at the National Genomics Infrastructure (Science for Life Laboratory, Uppsala) according to the manufacturer's instructions were sequenced in SMRT cells on the PacBio RS II system.

The quality of reads was evaluated with FastQC (17), InSeqt (http://grabherr.github.io/InSeqt/), and Kraken ((18), contamination). *De novo* genome assembly was performed by HGAP3 (Hierarchical Genome-Assembly Process)/HGAP4 (for SMI 706) algorithms from the PacBio SMRT tools (19). The assemblies were further evaluated using QUAST (20) to assess basic statistics and BlobTools (21) to again identify contaminations.

Annotation of each individual genome with three ab-initio tools used transcriptional data from RNA-sequencing experiments (differential gene expression in early versus late biofilms by the high spider-like biofilm-forming yeast *Candida parapsilosis* SMI828, by S.S. Manfred Grabherr, B.S., B.N., U.R., in preparation) and reviewed proteins from UniProt/Swiss-Prot (downloaded 2016-08 (22, 23)). Reads from RNA-sequencing were mapped

Editor Jason E. Stajich, University of California Riverside, Riverside, California, USA

Address correspondence to Ute Römling, ute.romling@ki.se.

The authors declare no conflict of interest.

See the funding table on p. 3.

Received 27 December 2024
Accepted 22 July 2025
Published 11 August 2025

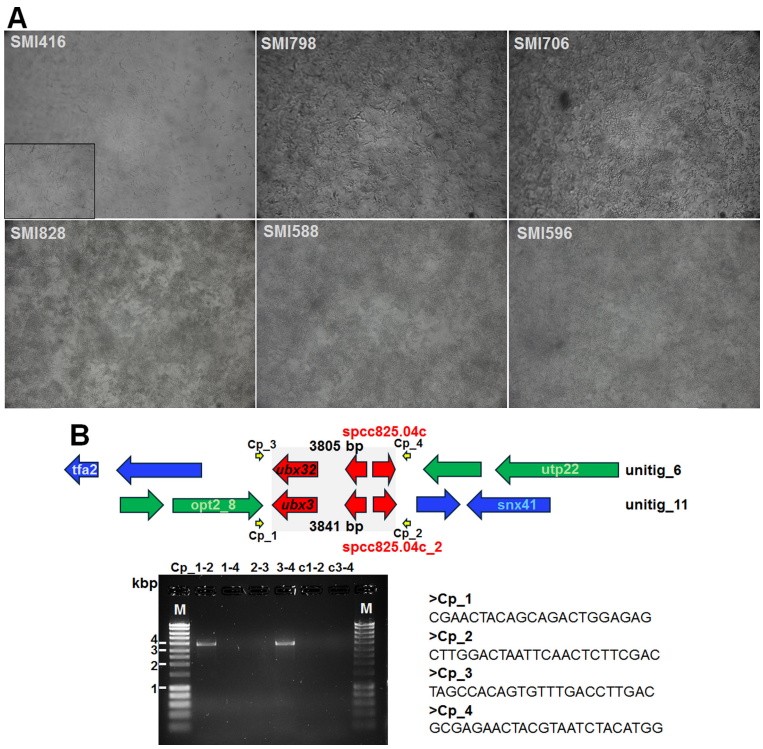

FIG 1 Biofilm formation of six *C. parapsilosis* isolates at 48 h on silicon pads developed in YNB medium with10% glucose mimicking parenteral solution as documented by light microscopy (A) with no, SMI416 (inset, silicon pad incubated with uninoculated medium), SMI798; low, SMI706; and high, SMI588, SMI596, SMI828, biofilm-forming ability (15); (B) translocation observed between *C. parapsilosis* SMI828 and CDC317 encompassing a 3,654 bp sequence duplication with 98.91% identity (gray box) in the two contigs in both strains. The translocation has been experimentally verified by PCR in SMI828 using the indicated primers located outside the repetitive sequences. Green and blue arrows indicate the location of genes on CDC317 contigs 005806 and 006372, respectively. Empy arrows/no text indicate gene of unknown function.

using *tophat2* (v.2.0.9) (24) and assembled with the *stringtie* package (v.1.2.2) (25). The three ab initio tools, GeneMark-ET v.4.3 (26), Augustus (integrates transcriptional and protein data [Augustus (27)]), and SNAP (pure ab initio (28)) were run within the *Maker* v.3.0 package (29) to combine annotations for a comprehensive gene build. Translated coding sequences were searched with BlastP against the UniProt/Swiss-Prot reference data set (22, 23) and run against InterProScan v.5.7-48 (30) to retrieve InterPro, PFAM,

TABLE 1 Basic characteristics of *de novo* assembled and annotated genomes of the six Swedish *C. parapsilosis* isolates

Strains	GC-content %	Length [bp]	Coverage	#contigs	Largest contig [bp]	N50 [bp]	Number of raw reads	L50	Predicted genes
	High biofilm-forming strains								
SMI828	38.7	13,575,533	206.70	38	1,981,830	1,105,301	229,353	5	6,128
SMI588	38.7	13,478,561	144.05	30	2,092,877	1,400,889	180,203	4	6,160
SMI596	38.7	13,244,759	131.82	21	3,027,088	2,087,741	174,490	3	6,052
	No biofilm-forming strains								
SMI798	38.7	13,395,816	131.28	28	3,025,112	2,006,230	169,821	3	6,070
SMI416	38.7	13,414,442	133.33	33	3,031,530	1,402,579	168,108	4	6,137
	Low biofilm-forming strains								
SMI706	38.5	13,322,571	327	29	1,990,489	1,465,361	203,964	4	6,038
	Reference strain								
CDC317	38.5	13,030,174	9.2	8	n.r.[a]	2,100,000	3,023,470	3	5,827

[a] n.r., not relevant.

Gene Ontology (GO), MetaCyc, UniPathway, KEGG, and Reactome assignments. The data output was parsed using the AnnIE annotation tool (31). Mitochondrial genomes were also annotated using the MFannot server (32).

The six *de novo* assembled and annotated CTG clade *C. parapsilosis* genomes are larger than the *C. parapsilosis* CDC317 reference genome (eight chromosomes of 12,998,393 bp in total and a mitochondrial genome of 31,781 bp) (Table 1). Comparative analysis indicated the presence of a large chromosomal translocation in SMI 828 (and all other sequenced genomes) compared to CDC317 (Fig. 1B).

ACKNOWLEDGMENTS

Srisuda Pannanusorn received a personal fellowship from Thammasat University, Pathum Thani, Thailand, and Karolinska Institutet/MTC. The authors would like to express our sincere appreciation to Victor Fernandez for initiation of the *Candida parapsilosis* project. This work was supported by a grant from the EU (European Commission) under grant agreement no. 634588 through the NOMORFILM project (Horizon 2020). Support by NBIS (National Bioinformatic Infrastructure Sweden) is gratefully acknowledged. BS is financially supported by the Knut and Alice Wallenberg Foundation (KAW 2017.0003). Computations were partially enabled by resources provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partially funded by the Swedish Research Council through grant agreement no. 2022-06725.

AUTHOR AFFILIATIONS

[1]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[2]Department of Biotechnology, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand
[3]Department of Medical Biochemistry Microbiology and Genomics, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden
[4]Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
[5]Department of Cell and Molecular Biology, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden

PRESENT ADDRESS

Sulman Shafeeq, Mölnlycke Health Care AB, Gothenburg, Sweden
Jacques Dainat, Joint Research Unit for Infectious Diseases and Vectors Ecology Genetics Evolution and Control, University of Montpellier, Montpellier, France

AUTHOR ORCIDs

Sulman Shafeeq http://orcid.org/0000-0003-1152-526X
Ali Dadvar https://orcid.org/0009-0001-0938-5125
Ute Römling http://orcid.org/0000-0003-3812-6621

FUNDING

Funder	Grant(s)	Author(s)
European Commission	634588	Ute Römling
Thammasat University	personal grant	Srisuda Pannanusorn

AUTHOR CONTRIBUTIONS

Sulman Shafeeq, Data curation, Formal analysis, Investigation, Methodology, Supervision, Validation, Visualization, Writing – review and editing | Srisuda Pannanusorn, Data curation, Formal analysis, Investigation, Methodology | Jacques Dainat, Formal analysis, Investigation, Methodology, Writing – review and editing | Ali Dadvar, Formal analysis, Methodology, Visualization | Christian Tellgren-Roth, Formal analysis, Investigation, Methodology | Bengt Sennblad, Formal analysis, Investigation, Methodology | Björn Nystedt, Formal analysis, Investigation, Methodology, Validation, Writing – review and editing | Ute Römling, Conceptualization, Data curation, Formal analysis, Funding acquisition, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in DDBJ/ENA/GenBank under the project ID PRJEB73879/ERP158620 (https://www.ebi.ac.uk/ena/browser/view/PRJEB73879). The version described in this paper is the first version of CAXIDQ010000000, CAXIDR010000000, CAXIDS010000000, CAXIDT010000000, CAXIDU010000000 and CAXIDV010000000. RNA sequencing data have been deposited in DDBJ/ENA/GenBank under the project ID ERP173917.

REFERENCES

1. Govrins M, Lass-Flörl C. 2024. *Candida parapsilosis* complex in the clinical setting. Nat Rev Microbiol 22:46–59. https://doi.org/10.1038/s41579-023-00961-8
2. Boattini M, Pinto MF, Christaki E, Fasciana T, Falces-Romero I, Tofarides A, Bianco G, Cendejas-Bueno E, Tricoli MR, Tsiolakkis G, García-Rodríguez J, Matzaras R, Comini S, Giammanco A, Kasapi D, Almeida A, Gartzonika K, Cavallo R, Costa C. 2023. Multicentre surveillance of *Candida* species from bood cultures during the SARS-CoV-2 pandemic in Southern Europe (CANCoVEU Project). Microorganisms 11:560. https://doi.org/10.3390/microorganisms11030560
3. Prigitano A, Blasi E, Calabrò M, Cavanna C, Cornetta M, Farina C, Grancini A, Innocenti P, Lo Cascio G, Nicola L, Trovato L, Cogliati M, Esposto MC, Tortorano AM, Romanò L, On Behalf Of The FiCoV Study Group. 2023. Yeast bloodstream infections in the COVID-19 patient: a multicenter Italian Study (FiCoV Study). J Fungi (Basel) 9:277. https://doi.org/10.3390/jof9020277
4. Koulenti D, Karvouniaris M, Paramythiotou E, Koliakos N, Markou N, Paranos P, Meletiadis J, Blot S. 2023. Severe *Candida* infections in critically ill patients with COVID-19. J Intensive Med 3:291–297. https://doi.org/10.1016/j.jointm.2023.07.005
5. Agnelli C, Valerio M, Bouza E, Guinea J, Sukiennik T, Guimarães T, Queiroz-Telles F, Muñoz P, Colombo AL. 2022. Prognostic factors of *Candida* spp. bloodstream infection in adults: a nine-year retrospective cohort study across tertiary hospitals in Brazil and Spain. Lancet Reg Health Am 6:100117. https://doi.org/10.1016/j.lana.2021.100117
6. Horn DL, Neofytos D, Anaissie EJ, Fishman JA, Steinbach WJ, Olyaei AJ, Marr KA, Pfaller MA, Chang CH, Webster KM. 2009. Epidemiology and outcomes of candidemia in 2019 patients: data from the prospective antifungal therapy alliance registry. Clin Infect Dis 48:1695–1703. https://doi.org/10.1086/599039
7. Andes DR, Safdar N, Baddley JW, Playford G, Reboli AC, Rex JH, Sobel JD, Pappas PG, Kullberg BJ, Mycoses Study G. 2012. Impact of treatment strategy on outcomes in patients with candidemia and other forms of invasive candidiasis: a patient-level quantitative review of randomized trials. Clin Infect Dis 54:1110–1122. https://doi.org/10.1093/cid/cis021
8. Moran C, Grussemeyer CA, Spalding JR, Benjamin DK Jr, Reed SD. 2009. *Candida albicans* and non-*albicans* bloodstream infections in adult and pediatric patients: comparison of mortality and costs. Pediatr Infect Dis J 28:433–435. https://doi.org/10.1097/INF.0b013e3181920ffd
9. Koehler P, Stecher M, Cornely OA, Koehler D, Vehreschild MJGT, Bohlius J, Wisplinghoff H, Vehreschild JJ. 2019. Morbidity and mortality of candidaemia in Europe: an epidemiologic meta-analysis. Clin Microbiol Infect 25:1200–1212. https://doi.org/10.1016/j.cmi.2019.04.024
10. Branco J, Miranda IM, Rodrigues AG. 2023. *Candida parapsilosis* virulence and antifungal resistance mechanisms: a comprehensive review of key determinants. J Fungi (Basel) 9:80. https://doi.org/10.3390/jof9010080
11. Montagna MT, Lovero G, Borghi E, Amato G, Andreoni S, Campion L, Lo Cascio G, Lombardi G, Luzzaro F, Manso E, Mussap M, Pecile P, Perin S, Tangorra E, Tronci M, Iatta R, Morace G. 2014. Candidemia in intensive care unit: a nationwide prospective observational survey (GISIA-3 study) and review of the European literature from 2000 through 2013. Eur Rev Med Pharmacol Sci 18:661–674.
12. Brillowska-Dabrowska A, Schön T, Pannanusorn S, Lönnbro N, Bernhoff L, Bonnedal J, Haggstrom J, Wistedt A, Fernandez V, Arendrup MC. 2009. A nosocomial outbreak of *Candida parapsilosis* in southern Sweden verified by genotyping. Scand J Infect Dis 41:135–142. https://doi.org/10.1080/00365540802585301
13. Chandra J, Mukherjee PK. 2015. *Candida biofilms*: development, architecture, and resistance. Microbiol Spectr 3. https://doi.org/10.1128/microbiolspec.MB-0020-2015
14. Pannanusorn S, Fernandez V, Römling U. 2013. Prevalence of biofilm formation in clinical isolates of *Candida species* causing bloodstream infection. Mycoses 56:264–272. https://doi.org/10.1111/myc.12014
15. Pannanusorn S, Ramírez-Zavala B, Lünsdorf H, Agerberth B, Morschhäuser J, Römling U. 2014. Characterization of biofilm formation and the role of BCR1 in clinical isolates of *Candida parapsilosis*. Eukaryot Cell 13:438–451. https://doi.org/10.1128/EC.00181-13
16. Pannanusorn S, Elings MA, Römling U, Fernandez V. 2012. Pyrosequencing of a hypervariable region in the internal transcribed spacer 2 to identify clinical yeast isolates. Mycoses 55:172–180. https://doi.org/10.1111/j.1439-0507.2011.02064.x
17. Wingett SW, Andrews S. 2018. FastQ screen: a tool for multi-genome mapping and quality control. F1000Res 7:1338. https://doi.org/10.12688/f1000research.15931.2
18. Davis MPA, van Dongen S, Abreu-Goodger C, Bartonicek N, Enright AJ. 2013. Kraken: a set of tools for quality control and analysis of high-throughput sequence data. Methods 63:41–49. https://doi.org/10.1016/j.ymeth.2013.06.027
19. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474
20. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

21. Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. F1000Res 6:1287. https://doi.org/10.12688/f1000research.12232.1
22. UniProt Consortium T. 2018. UniProt: the universal protein knowledgebase. Nucleic Acids Res 46:2699. https://doi.org/10.1093/nar/gky092
23. Boutet E, Lieberherr D, Tognolli M, Schneider M, Bansal P, Bridge AJ, Poux S, Bougueleret L, Xenarios I. 2016. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt knowledgebase: how to use the entry view. Methods Mol Biol 1374:23–54. https://doi.org/10.1007/978-1-4939-3167-5_2
24. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36
25. Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. Nat Biotechnol 33:290–295. https://doi.org/10.1038/nbt.3122
26. Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. Nucleic Acids Res 42:e119. https://doi.org/10.1093/nar/gku557
27. Keller O, Kollmar M, Stanke M, Waack S. 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27:757–763. https://doi.org/10.1093/bioinformatics/btr010
28. Korf I. 2004. Gene finding in novel genomes. BMC Bioinformatics 5:59. https://doi.org/10.1186/1471-2105-5-59
29. Cantarel BL, Korf I, Robb SMC, Parra G, Ross E, Moore B, Holt C, Sánchez Alvarado A, 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
30. Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S. 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics 30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031
31. Tate R, Hall B, DeRego T, Geib S. 2014. AnnIE: the ANNotation information extractor. Association for Computational Linguistics. https://aclanthology.org/2022.acl-demo.5/.
32. Lang BF, Beck N, Prince S, Sarrasin M, Rioux P, Burger G. 2023. Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. Front Plant Sci 14:1222186. https://doi.org/10.3389/fpls.2023.1222186