UCLA
UCLA Previously Published Works

Title

Genome Sequence for Candida albicans Clinical Oral Isolate 529L

Permalink

https://escholarship.org/uc/item/1q95g9n1

Journal

Microbiology Resource Announcements, 8(25)

ISSN

2576-098X

Authors

Cuomo, Christina A

Fanning, Saranna
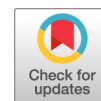
Gujja, Sharvari

et al.

Publication Date

2019-06-20

DOI

10.1128/mra.00554-19

Copyright Information

Peer reviewed

Genome Sequence for *Candida albicans* Clinical Oral Isolate 529L

Christina A. Cuomo,a Saranna Fanning,b Sharvari Gujja,a Qiandong Zeng,a Julian R. Naglik,c Scott G. Filler,d Aaron P. Mitchellb

aBroad Institute of MIT and Harvard, Cambridge, Massachusetts, USA
bDepartment of Biological Sciences, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA
cCentre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom
dDivision of Infectious Diseases, Los Angeles Biomedical Research Institute at Harbor-UCLA Medical Center, Torrance, California, USA

ABSTRACT The diploid heterozygous yeast *Candida albicans* is the most common cause of fungal infection. Here, we report the genome sequence assembly of the clinical oral isolate 529L. As this isolate grows as a commensal, this genome will serve as a reference for experimental and genetic studies of mucosal colonization.

*C*andida albicans is a major cause of bloodstream infections in immunocompromised individuals but is more commonly found in the human mycoflora. Here, we sequenced and assembled the genome of the 529L isolate, which was collected from a patient with oral candidiasis at Guy's Hospital, United Kingdom (1). In a model of persistent oral and vaginal colonization, the 529L isolate colonized these sites for over 5 weeks with higher fungal burden than with other clinical strains, including SC5314 (1). The 529L isolate has been used in mouse models of oral and vaginal colonization (1–4).

Cells were grown overnight at 30°C in four 4-ml cultures of YPD broth (1% [wt/vol] Difco yeast extract, 2% [wt/vol] Bacto peptone, 2% [wt/vol] dextrose) with shaking. DNA was prepared with the Qiagen Genomic-tip 100/G (catalog number 10243) using the Qiagen genomic buffer set (catalog number 19060), following the manufacturer's yeast protocol. Cell wall digestion was accomplished with lyticase (catalog number L2524; Sigma). DNA preparations from the four cultures were pooled.

For genome sequencing, two libraries were constructed from genomic DNA. For an ~180-base-insert library, 100 ng of genomic DNA was sheared to a median size of ~250 bp using a Covaris LE instrument; the resulting fragments were cleaned using SPRI AMPure XP beads, followed by end repair, A-base addition, and adapter ligation (New England BioLabs) (5). A ~3-kb insert library was prepared using the 2- to-5-kb-insert mate pair library prep kit (V2; Illumina). Libraries were sequenced on the Illumina HiSeq 2000 platform to generate paired 101-base reads totaling 35,388,222 reads of average quality 34.4 for the 180-base library and 52,230,472 reads of average quality 36.7 for the 3-kb library. Based on an evaluation of assemblies of *C. albicans* genomes at different coverage levels (6), a subset of approximately 100× sequence coverage of both libraries (28,316,832 total reads) was error corrected, filtered, and assembled using ALLPATHS (7) version R45597 with parameters HAPLOIDIFY=True and ASSISTED_PATCHING = 2.1. The quality of the assembly was evaluated using GAEMR v0.1.0 (https://github.com/broadinstitute/GAEMR); sequencing coverage appeared to be even across the assembly, with no large regions of aneuploidy noted. Three contigs that GAEMR identified as having sequence similarity to mitochondrial sequences were removed from the assembly. The final assembly with 176× read depth includes 87 scaffolds consisting of 608 contigs, with a scaffold N_{50} value of 1.2 Mb and a contig N_{50} value of 65.5 kb. The total scaffold length is 14.7 Mb, and the average GC content is 33.5%; this is similar in size and GC content to other *C. albicans* genomes (6, 8).

The genome was annotated by transferring gene coordinates from the SC5314 reference version A21-s02-m01-r01 curated by the *Candida* Genome Database (9) using unique NUCmer (MUMmer v3 [10]) alignments. In cases where the alignment-based mapping resulted in a gene with internal frameshifts, a transcript predicted by Prodigal v2.5 (11) or the longest overlapping open reading frame was substituted if it was longer than the mapped gene. For reference genes not mapped by this method, BLAST v2.2.25 alignments were used to identify missing loci, and gene structures were added using Prodigal v2.5 and GeneWise v2.2.0 (12). A total of 6,211 protein-coding genes were predicted using this approach, similar to the gene content of other *C. albicans* genomes (6, 8).

Data availability. The *C. albicans* 529L assembly and annotation reported here are available in GenBank under accession number ASHC00000000. Raw sequence reads have been deposited in the NCBI Sequence Read Archive under accession numbers SRX276261 and SRX276262.

REFERENCES

1. Rahman D, Mistry M, Thavaraj S, Challacombe SJ, Naglik JR. 2007. Murine model of concurrent oral and vaginal *Candida albicans* colonization to study epithelial host-pathogen interactions. Microbes Infect 9:615–622. https://doi.org/10.1016/j.micinf.2007.01.012.

2. Moyes DL, Runglall M, Murciano C, Shen C, Nayar D, Thavaraj S, Kohli A, Islam A, Mora-Montes H, Challacombe SJ, Naglik JR. 2010. A biphasic innate immune MAPK response discriminates between the yeast and hyphal forms of *Candida albicans* in epithelial cells. Cell Host Microbe 8:225–235. https://doi.org/10.1016/j.chom.2010.08.002.

3. Ibrahim AS, Luo G, Gebremariam T, Lee H, Schmidt CS, Hennessey JP, Jr, French SW, Yeaman MR, Filler SG, Edwards JE, Jr. 2013. NDV-3 protects mice from vulvovaginal candidiasis through T- and B-cell immune response. Vaccine 31:5549–5556. https://doi.org/10.1016/j.vaccine.2013.09.016.

4. Break TJ, Jaeger M, Solis NV, Filler SG, Rodriguez CA, Lim JK, Lee C-CR, Sobel JD, Netea MG, Lionakis MS. 2015. CX$_3$CR1 is dispensable for control of mucosal *Candida albicans* infections in mice and humans. Infect Immun 83:958–965. https://doi.org/10.1128/IAI.02604-14.

5. Fisher S, Barry A, Abreu J, Minie B, Nolan J, Delorey TM, Young G, Fennell TJ, Allen A, Ambrogio L, Berlin AM, Blumenstiel B, Cibulskis K, Friedrich D, Johnson R, Juhn F, Reilly B, Shammas R, Stalker J, Sykes SM, Thompson J, Walsh J, Zimmer A, Zwirko Z, Gabriel S, Nicol R, Nusbaum C. 2011. A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. Genome Biol 12:R1. https://doi.org/10.1186/gb-2011-12-1-r1.

6. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, Zeng Q, Zisson E, Wang JM, Greenberg JM, Berman J, Bennett RJ, Cuomo CA. 2015. Genetic and phenotypic intra-species variation in *Candida albicans*. Genome Res 25:413–425. https://doi.org/10.1101/gr.174623.114.

7. Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518. https://doi.org/10.1073/pnas.1017351108.

8. van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, Cuomo C, Berriman M, Scherer S, Magee BB, Whiteway M, Chibana H, Nantel A, Magee PT. 2007. Assembly of the *Candida albicans* genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol 8:R52. https://doi.org/10.1186/gb-2007-8-4-r52.

9. Costanzo MC, Arnaud MB, Skrzypek MS, Binkley G, Lane C, Miyasato SR, Sherlock G. 2006. The *Candida* Genome Database: facilitating research on *Candida albicans* molecular biology. FEMS Yeast Res 6:671–684. https://doi.org/10.1111/j.1567-1364.2006.00074.x.

10. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.

11. Hyatt D, Chen G-L, LoCascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics 11:119. https://doi.org/10.1186/1471-2105-11-119.

12. Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. Genome Res 14:988–995. https://doi.org/10.1101/gr.1865504.