

High-Quality Draft Genome Sequence of *Candida apicola* NRRL Y-50540

Leticia Vega-Alvarado,^b Jorge Gómez-Angulo,^a Zazil Escalante-García,^a Ricardo Grande,^b Anne Gschaedler-Mathis,^a Lorena Amaya-Delgado,^a Alejandro Sanchez-Flores,^b Javier Arrizon^a

Unidad de Biotecnología Industrial, Centro de Investigación y Asistencia en Tecnología y Diseño del Estado de Jalisco A.C., Jalisco, Mexico^a; Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología de la UNAM, Morelos, México^b

***Candida apicola*, a highly osmotolerant ascomycetes yeast, produces sophorolipids (biosurfactants), membrane fatty acids, and enzymes of biotechnological interest. The genome obtained has a high-quality draft for this species and can be used as a reference to perform further analyses, such as differential gene expression in yeast from *Candida* genera.**

Received 27 March 2015 Accepted 13 May 2015 Published 11 June 2015

Citation Vega-Alvarado L, Gómez-Angulo J, Escalante-García Z, Grande R, Gschaedler-Mathis A, Amaya-Delgado L, Sanchez-Flores A, Arrizon J. 2015. High-quality draft genome sequence of *Candida apicola* NRRL Y-50540. *Genome Announc* 3(3):e00437-15. doi:10.1128/genomeA.00437-15.

Copyright © 2015 Vega-Alvarado et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Alejandro Sánchez-Flores, alexsf@ibt.unam.mx, or Javier Arrizon, jparrizon@ciatej.mx.

Candida apicola, a highly osmotolerant ascomycetes yeast, produces sophorolipids (biosurfactants), membrane fatty acids, and enzymes, such as reductases and proteases (1–3). Naturally, this yeast has been found in wine and cachaça fermentation processes. In cachaça fermentation, it exhibits the capacity to produce volatile compounds (4–6). Recently, it was found that *C. apicola* yeasts isolated from the mezcal fermentation process secreted β -fructofuranosidases with fructosyltransferase activity, useful for prebiotic synthesis. Therefore, this fermentation process could reveal certain interesting features that can be discovered by performing whole-genome sequencing of *C. apicola*. It could lead us to a specific gene catalogue for biotechnological applications in fermentation bioprocesses.

Genomic DNA from *C. apicola* NRRL Y-50540 (YPD culture) was isolated and prepared as Illumina sequencing libraries to generate a total of 13,207,584 paired-end reads (estimated coverage $\sim 211\times$) with a length of 72 bases, using the Illumina GAIIx platform. The assembly was performed with Velvet v1.2.10 using a k-mer size of 51. An assembly of 9,769,876 bp in 40 contigs with lengths greater than or equal to 1,000 bp, was obtained with N_{50}/N_{90} values of 773,945/186,965 bp, respectively (7). The average contig length was 107,585 bp, giving a considerable space to search for genes. The average G+C content was 41.6%, which is similar to that of other *Candida* species. Gene prediction was performed using Augustus v2.7 (8) using three different *Candida* species profiles (*C. albicans*, *C. guilliermondii*, and *C. tropicalis*). We predicted 3,818 protein-coding genes by intersecting all three predictions. Using CEGMA v2.5 (9), we obtained 92% genome completeness. A group of enzymes from other yeast species, with functions related to carbohydrate polymer synthesis and degradation, such as secretory aspartyl protease (SAP2p), exoinulinases, and invertases, are present in the genome. However, the best hits for these proteins have only $\sim 35\%$ identity, implying a high divergence at the sequence level and probably several modifications in terms of substrate recognition and activity compared to those present in other yeasts found in fermentation processes.

Therefore, the *C. apicola* genome presented here is, to our knowledge, the first high-quality draft genome for this species and can be used as a reference to perform further analyses, such as differential gene expression of enzymes related to the synthesis and degradation of biotechnological molecules of interest, for example, in different fermentation conditions, which is one of our main interests. The availability of these genomes can also contribute to the understanding of the *Candida* genus, which is arguably biased by the amount of information related to the opportunistic human pathogen *Candida albicans*.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at GenBank under the accession numbers [LBNK01000001](https://www.ncbi.nlm.nih.gov/nuclink/LBNK01000001) to [LBNK01000040](https://www.ncbi.nlm.nih.gov/nuclink/LBNK01000040).

ACKNOWLEDGMENTS

We thank the “Unidad de Secuenciación Masiva y Bioinformática, Instituto de Biotecnología (USMB), UNAM,” for DNA sequencing advice and bioinformatics analysis. The USMB is part of the “Laboratorio Nacional de Respuesta a Enfermedades Emergentes,” which has been created and funded by the “CONACYT—Programa de Laboratorios Nacionales.” We are also thankful for the financial support of CONACYT project CB-2012-01-181766.

We thank Veronica Jimenez-Jacinto for preparing and submitting the sequencing data to the NCBI-SRA repository.

REFERENCES

1. Bednarski W, Adamczak M, Tomasik J, Płaszczyk M. 2004. Application of oil refinery waste in the biosynthesis of glycolipids by yeast. *Bioresour Technol* 95:15–18. [http://dx.doi.org/10.1016/j.biortech.2004.01.009](https://doi.org/10.1016/j.biortech.2004.01.009).
2. Girhard M, TF, Weber E, Smit MS, Urlacher VB. 2012. Cytochrome P450 reductase from *Candida apicola*: versatile redox partner for bacterial P450s. *Appl Microb Biotechnol* 97:1625–1635. [http://dx.doi.org/10.1007/s00253-012-4026-z](https://doi.org/10.1007/s00253-012-4026-z).
3. Reid VJ, Theron LW, du Toit M, Divol B. 2012. Identification and partial characterization of extracellular aspartic protease genes from *Metschnikowia pulcherrima* IWBT Y1123 and *Candida apicola* IWBT Y1384. *Appl Environ Microbiol* 78:6838–6849. [http://dx.doi.org/10.1128/AEM.00505-12](https://doi.org/10.1128/AEM.00505-12).
4. Arrizon J, Morel S, Gschaedler A, Monsan P. 2012. Fructanase and fructosyltransferase activity of non-*Saccharomyces* yeasts isolated from fer-

- menting musts of mezcal. *Bioresour Technol* 110:560–565. <http://dx.doi.org/10.1016/j.biortech.2012.01.112>.
5. Oliveira ES, Rosa CA, Morgano MA, Serra GE. 2005. The production of volatile compounds by yeasts isolated from small Brazilian cachaça distilleries. *World J Microbiol Biotechnol* 21:1569–1576. <http://dx.doi.org/10.1007/s11274-005-7894-x>.
 6. Tofalo R, Chaves-López C, Di Fabio F, Schirone M, Felis GE, Torriani S, Paparella A, Suzzi G. 2009. Molecular identification and osmotolerant profile of wine yeasts that ferment a high sugar grape must. *Int J Food Microbiol* 130:179–187. <http://dx.doi.org/10.1016/j.ijfoodmicro.2009.01.024>.
 7. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. *Genome Res* 18:821–829. <http://dx.doi.org/10.1101/gr.074492.107>.
 8. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. *BMC Bioinformatics* 7:62. <http://dx.doi.org/10.1186/1471-2105-7-62>.
 9. Parra G, Bradnam K, Ning Z, Keane T, Korf I. 2009. Assessing the gene space in draft genomes. *Nucleic Acids Res* 37:289–297. <http://dx.doi.org/10.1093/nar/gkn916>.