



Draft Genome Sequence of *Metschnikowia australis* Strain UFMG-CM-Y6158, an Extremophile Marine Yeast Endemic to Antarctica

Thiago M. Batista,^a Heron O. Hilário,^a Rennan G. Moreira,^c Carolina Furtado,^d Valéria M. Godinho,^b Luiz H. Rosa,^b Glória R. Franco,^a Carlos A. Rosa^b

Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil^a; Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil^b; Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil^c; Instituto Nacional do Câncer, Rio de Janeiro, Brazil^d

ABSTRACT Here we report the draft genome sequence of *Metschnikowia australis* strain UFMG-CM-Y6158, a yeast endemic to Antarctica. We isolated the strain from the marine seaweed *Acrosiphonia arcta* (*Chlorophyta*). The genome is 14.3 Mb long and contains 4,442 predicted protein-coding genes.

The genus *Metschnikowia* comprises a clade consisting of approximately 81 species. The sexual life cycles of the members of this clade involve the formation of elongated asci containing two, often needle-shaped, spores (1). *M. australis* is a species endemic to Antarctica, and has been isolated from seawater, marine invertebrates, sponges, and macroalgae (2–6). Owing to the extremely cold environment of Antarctica, *M. australis* may have unique metabolic traits enabling it to survive under such stressful conditions; exploring these can help identify potential antifreeze compounds for biotechnological use.

We isolated *M. australis* strain UFMG-CM-Y6158 from a marine macroalgae, *Acrosiphonia arcta* (*Chlorophyta*), collected in Admiralty Bay of King George Island in Keller Peninsula, Antarctica (5). We cultivated the strain on marine agar (Himedia, India) at 10°C for 15 days, and the genomic DNA was isolated by phenol:chloroform (1:1) extraction. We assessed DNA quality by gel electrophoresis and determined its purity and quantity using both the NanoDrop 1000 UV-Vis spectrophotometer and the Qubit version 2.0 fluorometer with the Qubit dsDNA HS assay kit (Thermo Fisher Scientific). We used the Nextera XT DNA kit (Illumina) to construct paired-end libraries and assessed their quality using Bioanalyzer HS Assay (Agilent Technologies). Generated fragments with a mean length of 1,167 bp were sequenced using the Illumina MiSeq sequencer, whereas those with a mean length of 550 bp were sequenced using the Illumina HiSeq 2500 sequencer. The former generated 1,585,122 reads (2 × 301) with 35× coverage, while the latter generated 103,312,458 reads (2 × 101) with 745× coverage. We assembled the genome using SPAdes version 3.9.1 (7). The assembled draft genome consisted of 14,356,710 bp over 160 contigs (>505 bp) with a G+C content of 47.2%. The longest contig was 1,116,518 bp long, and the N_{50} contig length was 542,232 bp. CEGMA (8) analysis showed that the assembly was 95.9% complete, whereas analysis with BUSCO version 2 (9) using the *Saccharomycetales* lineage data set indicated 90.2% completeness based on the presence of conserved orthologous genes among species of the genus. We identified 4,442 protein-coding genes using MAKER2 (10). A sequence similarity search using the BLASTx tool in BLAST version 2.2.31+ (11) returned 4,348 protein matches (97.8%), with $e\text{-value} \leq 1e^{-6}$, against NCBI's nonredundant database. We identified 249 tRNAs using tRNAscan-SE (12).

Received 17 March 2017 Accepted 30 March 2017 Published 18 May 2017

Citation Batista TM, Hilário HO, Moreira RG, Furtado C, Godinho VM, Rosa LH, Franco GR, Rosa CA. 2017. Draft genome sequence of *Metschnikowia australis* strain UFMG-CM-Y6158, an extremophile marine yeast endemic to Antarctica. *Genome Announc* 5:e00328-17. <https://doi.org/10.1128/genomeA.00328-17>.

Copyright © 2017 Batista et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Carlos A. Rosa, carlrosa@icb.ufmg.br.

Using the OrthoVenn web platform (13), we compared *M. australis* protein-coding genes with those of two previously sequenced *Metschnikowia* genomes—*M. fructicola* and *M. bicuspidata*. The analysis showed that *M. australis* has a much shorter predicted proteome than that of *M. fructicola* (5,851 protein-coding genes) and *M. bicuspidata* (6,028 protein-coding genes). Additionally, we found six exclusive clusters of paralogous genes, of which four did not match any protein in the NCBI and UniProt-Swissprot databases. These results highlight the importance of investigating yeast endemic to Antarctica, such as *M. australis*, not only to identify novel genes associated with adaptation to extreme environments, but also for potential application in biotechnology.

Accession number(s). Data related to this whole-genome shotgun project have been deposited at DDBJ/ENA/GenBank under the accession number [MVNQ00000000](#). The version described in this paper is the first version, MVNQ01000000.

ACKNOWLEDGMENTS

Laboratório Multusuário de Bioinformática—EMBRAPA Informática Agropecuária, Campinas, Brazil, provided access to genome annotation. This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq-PROANTAR 407230/2013-0, INCT Criosfera, 457499/2014-1); Fundação de Apoio à Pesquisa do Estado de Minas Gerais (FAPEMIG-0050-13, APQ-01525-14); and CAPES (23038.003478/2013-92).

REFERENCES

- Lachance MA, Hurtado E, Hsiang T. 2016. A stable phylogeny of the large-spored *Metschnikowia* clade. *Yeast* 33:261–275. <https://doi.org/10.1002/yea.3163>.
- Loque CP, Medeiros AO, Pellizzari FM, Oliveira EC, Rosa CA, Rosa LH. 2010. Fungal community associated with marine macroalgae from Antarctica. *Polar Biol* 33:641–648. <https://doi.org/10.1007/s00300-009-0740-0>.
- Lachance M-A. 2011. *Metschnikowia* Kamienski (1899), p 575–620. In Kurtzman CP, Fell JW, Boekhout T (ed), *The yeasts: a taxonomic study*, 5th ed, vol 1. Elsevier, London.
- Vaca I, Faúndez C, Maza F, Paillavil B, Hernández V, Acosta F, Levicán G, Martínez C, Chávez R. 2013. Cultivable psychrotolerant yeasts associated with Antarctic marine sponges. *World J Microbiol Biotechnol* 29: 183–189. <https://doi.org/10.1007/s11274-012-1159-2>.
- Godinho VM, Furbino LE, Santiago IF, Pellizzari FM, Yokoya NS, Pupo D, Alves TM, Junior PA, Romanha AJ, Zani CL, Cantrell CL, Rosa CA, Rosa LH. 2013. Diversity and bioprospecting of fungal communities associated with endemic and cold-adapted macroalgae in Antarctica. *ISME J* 7:1434–1451. <https://doi.org/10.1038/ismej.2013.77>.
- Furbino LE, Godinho VM, Santiago IF, Pellizzari FM, Alves TMA, Zani CL, Junior PA, Romanha AJ, Carvalho AG, Gil LH, Rosa CA, Minnis AM, Rosa LH. 2014. Diversity patterns, ecology and biological activities of fungal communities associated with the endemic macroalgae across the Antarctic peninsula. *Microb Ecol* 67:775–787. <https://doi.org/10.1007/s00248-014-0374-9>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- 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>.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210–3212. <https://doi.org/10.1093/bioinformatics/btv351>.
- Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* 12:491. <https://doi.org/10.1186/1471-2105-12-491>.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. Blast+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Lowe TM, Eddy SR. 1997. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* 25:955–964. <https://doi.org/10.1093/nar/25.5.0955>.
- Wang Y, Coleman-Derr D, Chen G, Gu YQ. 2015. OrthoVenn: a web server for genome wide comparison and annotation of orthologous clusters across multiple species. *Nucleic Acids Res* 43:W78–W84. <https://doi.org/10.1093/nar/gkv487>.