

Draft genome sequence of the extremophile endemic marine antarctic yeast *Metchnikowia australis*

Heron Oliveira Hilario¹, Thiago Mafra Batista¹, Rennan Garcia Moreira¹, Valéria Martins Godinho², Carlos Augusto Rosa², Luiz Henrique Rosa², Glória Regina Franco¹

*Departamento de Bioquímica & Imunologia, Universidade Federal de Minas Gerais¹,
Departamento de Microbiologia, Universidade Federal de Minas Gerais²*

Antarctica, the earth's fifth biggest continent, is an harsh environment. Almost all of its 14.000.000 km² surface is always under ice, originated from the snow falling constantly since past ages. In the winter this ice sheet grows over the Antarctic Circle, till the the northernmost tip of the continent, where situates the Antarctic Peninsula. These geocycles selected diverse survival strategies for life in the region, and only the organisms adapted to deal with extreme situations are capable of flourishing, in special, island microorganisms which are poorly characterized due to their difficult cultivation conditions. As one of the less explored places on earth, Antarctica potentially harbors indigenous microorganisms which may present features that could be exploited by the scientific and medical communities, such as novel candidates for antibiotics, anti-freezing agents and other pharmacological composites. The Mycoantar project integrates the annual expeditions of PROANTAR (Brazilian Antarctic Program), investigating this biodiversity and has already isolated many fungi from the antarctic region. The yeast endemic *Metschnikowia australis* was isolated from marine macroalgae from Antarctica and after diverse microbiological characterization, it was further sequenced for deeper bioinformatic investigation. Paired-end libraries were constructed using Nextera XT DNA Kit, producing fragments with mean of 1,167 bp that were sequenced in the Illumina Miseq and fragments of 550 bp that were sequenced in the Illumina HiSeq 2500. A total of 1,585,122 reads (2x301) was generated by MiSeq with an estimated genome coverage of 35x, and 103,312,458 reads (2x101) were generated by Hiseq with an estimated genome coverage of 745x. The genome was assembled using SPADIS 3.9.1, with default parameters. The estimated genome size is 14,356,710 bp comprising 160 contigs with mean of 89,729 bp, and the longest contig length of 1,116,518 bp, N₅₀ value of 542,232 and GC content of 47.2%. The predictor Maker2 was able to find 4,442 Open Reading Frames. The search for sequence similarity against the non-redundant database from NCBI revealed hits of these ORFs with 4,348 distinct proteins. Comparative analysis of orthologous proteins present in the genome of *M. australis* and other *Metschnikowia* species (*M. bicuspidata* and *M. fructicola*) revealed several shared clusters and 163 singletons, however, six clusters were composed solely of *M. australis* proteins. Downstream analysis will be carried out to deep investigate genes from *M. australis* and their involvement in the cold adaptation.