Genome sequence of Torulaspora microellipsoides CLIB 830T

Virginie Galeote, Frederic Bigey, Hugo Devillers, Sylvie Dequin, Kenneth Wolfe, Cécile Neuvéglise

To cite this version:

Virginie Galeote, Frederic Bigey, Hugo Devillers, Sylvie Dequin, Kenneth Wolfe, et al.. Genome sequence of Torulaspora microellipsoides CLIB 830T. Genome Announcements, American Society for Microbiology, 2018, 6 (26), 10.1128/genomeA.00615-18. hal-02623200

HAL Id: hal-02623200
https://hal.inrae.fr/hal-02623200
Submitted on 26 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.
Genome Sequence of *Torulaspora microellipsoides* CLIB 830<sup>T</sup>

Virginie Galeote,a Frédéric Bigey,a Hugo Devillers,b Raúl A. Ortiz-Merino,c Sylvie Dequin,a Kenneth H. Wolfe,c Cécile Neuvégliseb

aSPO, INRA, SupAgro, Université Montpellier, Montpellier, France
bMicalis Institute, INRA, AgroParisTech, Université Paris-Saclay, Jouy-en-Josas, France
cUCD Conway Institute and School of Medicine, University College Dublin, Dublin, Ireland

**ABSTRACT** We report here the genome sequence of the ascomycetous yeast *Torulaspora microellipsoides* CLIB 830<sup>T</sup>. A reference genome for this species, which has been found as a donor of genetic material in wine strains of *Saccharomyces cerevisiae*, will undoubtedly give clues to our understanding of horizontal transfer mechanisms between species in the wine environment.

The genus *Torulaspora* belongs to the Saccharomycotina subphylum and is now composed of eight species (*T. delbrueckii*, *T. globosa*, *T. franciscae*, *T. pretoriensis*, *T. microellipsoides*, *T. maleae*, *T. quercuum*, and *T. incana*); *T. maleae*, *T. quercuum*, and *T. incana* have been described in the past decade (1–4). *Torulaspora delbrueckii* is probably the most widely distributed species, found in nature and in anthropic environments requiring fermentation, such as wine or bread, whereas other species, such as *T. microellipsoides*, appear more sporadically. However, their presence in fermentation vats may play an important role. Indeed, *T. microellipsoides* has been shown to be the donor of a DNA region of at least 158 kb to *Saccharomyces cerevisiae* wine strains, which confers to them an adaptive advantage during wine fermentation (5, 6). Yet, only the *T. delbrueckii* genome sequence is available (7). Sequencing the genome of *T. microellipsoides* strain CLIB 830<sup>T</sup> (=CBS 427) may therefore contribute to an increase in the genomics knowledge of this clade.

Total genomic DNA was used to construct a paired-end (PE) 500-bp insert library and a mate pair (MP) 6-kb insert library. Both libraries were sequenced using the Illumina HiSeq 2000 platform, resulting in raw sequencing coverage depths of 197× (PE) and 177× (MP). Sequencing reads were cleaned using Trimmomatic version 0.32 (8), resulting in a sequencing depth of 319×. A preliminary assembly of all reads was obtained using SOAPdenovo2 version 2.04 (9). MP reads were mapped with BWA version 0.6.2 (10), and inward read pairs were removed using BowTie version 2.2.3 (11). A second assembly of PE and cleaned MP reads was performed using SOAPdenovo2, with a k-mer of 61. Gap closing was performed using GapCloser version 1.12 (9). The final assembly was made of 46 scaffolds (N<sub>50</sub>, 1.2 Mb) larger than 1 kb; 8 of them corresponded to mitochondrial DNA. The remaining 38 scaffolds were suitable for automatic annotation.

The structural annotation of protein-coding genes was performed using the Amadea Annotation transfer tool (Isoft, France), with the *Lachancea kluyveri* CBS3082<sup>T</sup> genome as a reference (revised version available at http://gryc.inra.fr [12]). Missing genes were investigated through a BLASTX search against the NCBI RefSeq database, with a comparison to an annotation from the YGAP pipeline (13), and manual curation. In total, 5,239 protein-coding genes were predicted, including 145 pseudogenes. tRNA genes were identified using tRNAscan-SE version 1.3.1 (14). Transposable elements were identified by BLAST with known yeast elements from different families, such as Ty1-copia, Ty3-gypsy, and hAT. A family of 38 Rover elements was identified, including intact and degenerate copies and miniature inverted-repeat transposable elements.
The accession numbers of the 46 scaffolds are FYBL01000001 to FYBL01000046. The deposited at the European Nucleotide Archive (ENA) under BioProject no. PRJEB7632. version

ACKNOWLEDGMENTS

The development of the annotation transfer tool based on Amadea Biopack was financially supported by the ANR project GB-3G (2010 BLAN 1606). We thank the FP7 Marie Curie Initial Training Networks (ITN) (FP7-PEOPLE-2013-ITN, YEASTCELL).

We thank Jean-Luc Legras for his helpful advice and discussions.

REFERENCES


