

Draft Genome Sequence of *Lachancea lanzarotensis* CBS 12615^T, an Ascomycetous Yeast Isolated from Grapes

Véronique Sarilar,^{a,b} Hugo Devillers,^{a,b} Kelle C. Freel,^c Joseph Schacherer,^c Cécile Neuveglise^{a,b}

INRA, UMR1319 Micalis, Jouy-en-Josas, France^a; AgroParisTech, UMR1319 Micalis, Jouy-en-Josas, France^b; CNRS, Department of Genetics, Genomics and Microbiology, University of Strasbourg, UMR7156, Strasbourg, France^c

We report the genome sequencing of the yeast *Lachancea lanzarotensis* CBS 12615^T. The assembly comprises 24 scaffolds, for a total size of 11.46 Mbp. The annotation revealed 5,058 putative protein-coding genes. Detection of seven centromeres supports a chromosome fusion, which occurred after divergence from *Lachancea thermotolerans* and *Lachancea kluyveri*.

Received 26 February 2015 Accepted 3 March 2015 Published 16 April 2015

Citation Sarilar V, Devillers H, Freel KC, Schacherer J, Neuveglise C. 2015. Draft genome sequence of *Lachancea lanzarotensis* CBS 12615^T, an ascomycetous yeast isolated from grapes. *Genome Announc* 3(2):e00292-15. doi:10.1128/genomeA.00292-15.

Copyright © 2015 Sarilar 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/4.0/).

Address correspondence to Cécile Neuveglise, ncecile@grignon.inra.fr.

Lachancea lanzarotensis is a newly described species isolated from grapes and wine fermentation during a study of yeast communities in vineyards and wineries in the Canary Islands (1). This species, naturally present in grape must, contributes to spontaneous alcoholic fermentation during the early phases of wine fermentation, before *Saccharomyces cerevisiae* becomes dominant and completes the process. Knowledge of its gene repertoire, especially concerning enzymes involved in fermentation process, will provide clues to understand the wine microbial ecosystem and its functioning, as well as enhance aroma and wine quality.

The CBS 12615^T genome was sequenced using Illumina-Solexa technology, from a 500-bp library, on paired-end reads. An ~70-fold coverage was generated. Reads were assembled using SOAPdenovo2 v2.04 (2) with a k-mer size of 75, as recommended by KmerGenie v1.6741 (3). Gaps were filled in using GapCloser v1.12 (2). The rDNA unit was assembled through iterative runs of Newbler v2.7 (<http://454.com/products/analysis-software/index.asp>) and manually integrated between two scaffolds harboring a partial rDNA unit at one of their extremities. The MAT locus was localized by synteny with that of *Lachancea thermotolerans* (4) and represented as a stretch of five “N” in scaffold 11. Based on the reference genomes of the closely related species *L. thermotolerans* and *Lachancea kluyveri* (4), putative protein-coding genes were annotated using the Amadea annotation transfer tool (Isoft, France). Transposable elements were identified by BLAST with known *Ty1*, *Ty3*, and *hAT* sequences from the *Lachancea* clade as queries (4, 5). tRNA genes were identified using tRNAscan-SE v1.3.1 (6). Additionally, snRNAs were identified by sequence homology with *L. thermotolerans* (4) snRNA sequences. Centromeres were localized by searching for characteristic motifs in syntenic regions of *L. thermotolerans* centromeres (4) using MEME v4.9.1 (7).

The current draft comprises 24 scaffolds interrupted by 52 gaps, for a total size of 11,461,889 bp and a G+C content of around 44.3%. A total of 5,058 putative protein-coding genes have been identified, 321 of which harbor spliceosomal introns within the coding sequences. A total of 59 additional genes have been

annotated as dubious models or pseudogenes, with frameshifts, stops in translation, or dubious starts or stops. The genome contains 182 tRNAs. Whenever possible, functional annotation was transferred from *S. cerevisiae* or refseq protein sequences, or experimentally validated proteins from other *Lachancea* species. During this process, 3,863 proteins showed at least 50% sequence similarity with *S. cerevisiae*. Finally, only 48 putative proteins have no known homologs. Class I and class II transposable elements were detected: LTR-retrotransposons with five *Ty3*-like and two *Ty1*-like degenerate copies, as well as 25 copies of *hAT* elements including one putative autonomous copy. Only seven centromeres have been identified, suggesting that *L. lanzarotensis* has one less chromosome than *L. thermotolerans* and *L. kluyveri*. Dot-plot comparison showed that no large DNA deletion occurred, which is compatible with the fact that CBS 12615 has undergone a chromosomal fusion.

Further comparison of the genome of CBS 12615 against other *Lachancea* species will provide additional insights concerning structural genome evolution and its functional impact.

Nucleotide sequence accession numbers. This whole-genome shotgun project (PRJEB7950) has been deposited at the European Nucleotide Archive under the accession numbers [LN736360](https://www.ebi.ac.uk/ena/record/LN736360) to [LN736383](https://www.ebi.ac.uk/ena/record/LN736383).

ACKNOWLEDGMENTS

This work was supported by an ANR grant (GB-3G project, 2010-BLAN-1606).

We thank Jean-Philippe Meyniel for technical assistance on Amadea Bio-pack and all members of the GB-3G consortium for fruitful discussions.

REFERENCES

- González SS, Alcoba-Flórez J, Laich F. 2013. *Lachancea lanzarotensis* sp. nov., an ascomycetous yeast isolated from grapes and wine fermentation in Lanzarote, Canary Islands. *Int J Syst Evol Microbiol* 63:358–363. <http://dx.doi.org/10.1099/ijms.0.046029-0>.
- Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, Tang J, Wu G, Zhang H, Shi Y, Liu Y, Yu C, Wang B, Lu Y, Han C, Cheung DW, Yiu S-M, Peng S, Xiaojian Z, Liu G, Liao X, Li Y, Yang H,

- Wang J, Lam T-W, Wang J. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read *de novo* assembler. *GigaScience* 1:18. <http://dx.doi.org/10.1186/2047-217X-1-18>.
3. Chikhi R, Medvedev P. 2014. Informed and automated k-mer size selection for genome assembly. *Bioinformatics* 30:31–37. <http://dx.doi.org/10.1093/bioinformatics/btt310>.
4. The Génolevures Consortium, Souciet J-L, Dujon B, Gaillardin C, Johnston M, Baret PV, Cliften PV, Sherman DJ, Weissenbach J, Westhof E, Wincker P, Jubin C, Poulain J, Barbe V, Ségurens B, Artiguenave F, Anthouard V, Vacherie B, Val M-E, Fulton RS, Minx P, Wilson R, Durrens P, Jean G, Marck C, Martin T, Nikolski M, Rolland T, Seret M-L, Casarégola S, Despons L, Fairhead C, Fischer G, Lafontaine I, Leh V, Lemaire M, de Montigny J, Neuvéglise C, Thierry A, Blanc-Lenfle I, Bleykasten C, Diffels J, Fritsch E, Frangeul L, Goëffon A, Jauniaux N, Kachouri-Lafond R, Payen C, Potier S, Pribylova L, Ozanne C, Richard G-F, Sacerdot C, Straub M-L, Talla E. 2009. Comparative genomics of protoploid *Saccharomycetaceae*. *Genome Res* 19:1696–1709. <http://dx.doi.org/10.1101/gr.091546.109>.
5. Sarilar V, Bleykasten-Grosshans C, Neuvéglise C. 2015. Evolutionary dynamics of *hAT* DNA transposon families in *Saccharomycetaceae*. *Genome Biol Evol* 7:172–190. <http://dx.doi.org/10.1093/gbe/evu273>.
6. 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. <http://dx.doi.org/10.1093/nar/25.5.0955>.
7. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. 2009. MEME suite: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208. <http://dx.doi.org/10.1093/nar/gkp335>.