



# Genome Sequence of the Diploid Yeast *Debaryomyces hansenii* TMW 3.1188

 Tobias Link,<sup>a</sup>  Rebekka H. Lülff,<sup>a</sup>  Marina Parr,<sup>b</sup>  Maik Hilgarth,<sup>a</sup>  Matthias A. Ehrmann<sup>a</sup>

<sup>a</sup>Lehrstuhl für Mikrobiologie, Technische Universität München, Freising, Germany

<sup>b</sup>Department of Bioinformatics, Wissenschaftszentrum Weißenstephan, Technische Universität München, Freising, Germany

Tobias Link and Rebekka H. Lülff contributed equally to the manuscript.

**ABSTRACT** *Debaryomyces hansenii* TMW 3.1188 is a halotolerant diploid yeast that was isolated from lupine moromi fermentation. Here, we report on the 24.77-Mbp genome of a diploid strain of the species *D. hansenii*.

*Debaryomyces hansenii* is a yeast that is commonly found in marine environments or in food fermentations such as cheese or doenjang (1–5). Strain TMW 3.1188 was isolated from a spontaneous lupine moromi fermentation with 10% NaCl (wt/vol) after 12 weeks (6). The identity of the isolate was initially verified via 28S rDNA sequencing using the primers V9G (7) and LR5 (8) (GenBank accession number [OP179623](#)). However, because the *ACT1* gene was reported to have greater variability within *Debaryomyces* species, both alleles from TMW 3.1188 were used to verify the affiliation with the *D. hansenii* species (Fig. 1) (9, 10). Isolation of the genomic DNA of TMW 3.1188 was carried out with harvested cells that had been grown for 24 h at 30°C in DSMZ 90 medium with 5% NaCl (wt/vol). DNA was isolated using R-zymolyase (Zymo Research), followed by phenol-chloroform extraction and enzymatic digestion of proteins with proteinase K.

Quantification and fragment distribution analysis of the DNA were performed with a Qubit fluorometer (Life Technologies) and a Femto Pulse electrophoresis system (Agilent). The library was created with the SMRTbell Express template preparation kit v2.0 (Pacific Biosciences [PacBio]). The library was size selected to >17.5 kbp using a BluePippin device (Sage Science) and then sequenced using the Sequel sequencing kit v3.0 (PacBio) according to the manufacturer's instructions. The read mode was set to continuous long reads (CLR) on a PacBio Sequel instrument.

A total of 10,724,419,855 bases were sequenced. The mean subread length was 14,003 bases, with an  $N_{50}$  value of 19,076 bp. The mean of the longest subreads was 14,126 bases, with an  $N_{50}$  value of 19,076 bp. After sequencing, data quality was checked via the PacBio single-molecule real-time (SMRT) Link software. The genome was assembled with HGAP4 in SMRT Link v10.0.0.108728 with default parameters, except that the estimated genome size was set to 12 Mbp. The assembly generated 26 contigs, with a total sequence length of 24,773,645 bp, an  $N_{50}$  value of 1,604,673 bp, and a GC content of 36.23%.

Genomic comparisons using FastANI v1.3 (11) and genomdiff in GenomeTools v1.6.2 (12) revealed that TMW 3.1188 has the greatest similarity to strain CBS767<sup>T</sup> (GenBank assembly accession number [GCF\\_000006445.2](#)), with an average nucleotide identity (ANI) value of 95.72%, while the  $K_r$  distance, which estimates the number of substitutions per site between two unaligned DNA sequences, has a value of 0.055 (13). The ANI value with respect to the type strain of the species *Debaryomyces fabryi*, CBS789<sup>T</sup> (GenBank assembly accession number [GCF\\_001447935.2](#)), is 85.33%, and the  $K_r$  distance is 0.102. The ANI values with respect to strains J6 (GenBank assembly accession number [GCA\\_001682995.1](#)) and MTCC 234 (GenBank assembly accession number [GCA\\_000239015.2](#)) are only 84.5%. The  $K_r$  distance from TMW 3.1188 to J6 is 0.119 and that to MTCC234 is 0.129.

**Editor** Antonis Rokas, Vanderbilt University

**Copyright** © 2022 Link et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

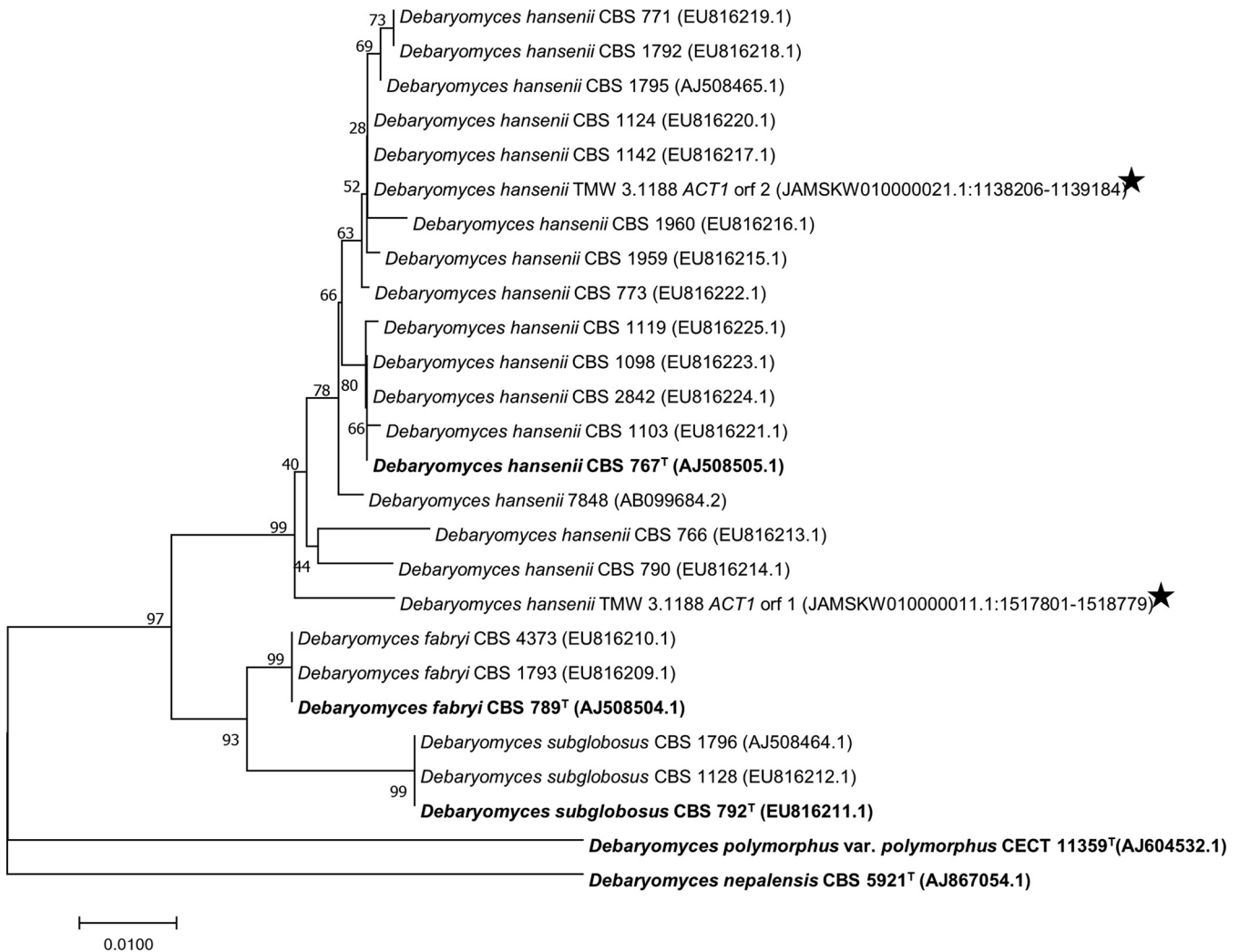
Address correspondence to Matthias A. Ehrmann, [matthias.ehrmann@tum.de](mailto:matthias.ehrmann@tum.de).

The authors declare no conflict of interest.

**Received** 18 July 2022

**Accepted** 2 September 2022

**Published** 26 October 2022



**FIG 1** Phylogenetic tree of *Debaryomyces* species. The evolutionary history was inferred using the neighbor-joining method (15). The optimal tree, with the sum of branch lengths of 0.23095609, is shown. The percentages of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches (16). The tree is drawn to scale, with branch lengths in the same units as the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the maximum composite likelihood method (17) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated. There were a total of 752 positions in the final data set. Evolutionary analyses were conducted in MEGA7 (18). Type strains were indicated by bold letters with superscript T. Black stars indicate the positions of the sequences from TMW 3.1188. The accession numbers and positions in the chromosome are indicated in parentheses.

Furthermore, genomic comparison using the D-GENIES (14) online tool revealed that each chromosome of CBS767<sup>T</sup> had two contigs of TMW 3.1188 aligning to it, one with a higher ANI value (>57%) and one with a lower ANI value (34 to 40%), proving the existence of two chromosomal sets. Contigs 10, 12, 13, and 26 of TMW 3.1188 align to the mitochondrial genome of *Debaryomyces hansenii* (GenBank assembly accession number [GCF\\_000006445.2](https://www.ncbi.nlm.nih.gov/assembly/GCF_000006445.2)).

**Data availability.** The assembled genome is available under assembly accession number [GCA\\_024256405.1](https://www.ncbi.nlm.nih.gov/assembly/GCA_024256405.1), with BioProject accession number [PRJNA841823](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA841823). The raw reads are available under SRA accession number [SRR19753073](https://www.ncbi.nlm.nih.gov/sra/SRR19753073). The 28S rDNA sequence is available under GenBank accession number [OP179623](https://www.ncbi.nlm.nih.gov/nuccore/OP179623).

## ACKNOWLEDGMENTS

The German Ministry of Food and Agriculture (BMEL) supported part of this work (project 28-1-A4.001-17).

We acknowledge technical support from the Functional Genomics Center Zurich, the University of Zurich, and ETH Zurich.

## REFERENCES

1. Shim JM, Lee KW, Yao Z, Kim HJ, Kim JH. 2016. Properties of doenjang (soybean paste) prepared with different types of salts. *J Microbiol Biotechnol* 26: 1533–1541. <https://doi.org/10.4014/jmb.1605.05019>.
2. Petersen KM, Jespersen L. 2004. Genetic diversity of the species *Debaryomyces hansenii* and the use of chromosome polymorphism for typing of strains isolated from surface-ripened cheeses. *J Appl Microbiol* 97: 205–213. <https://doi.org/10.1111/j.1365-2672.2004.02293.x>.
3. Berrocal CA, Rivera-Vicens RE, Nadathur GS. 2016. Draft genome sequence of the heavy-metal-tolerant marine yeast *Debaryomyces hansenii* J6. *Genome Announc* 4:6–7. <https://doi.org/10.1128/genomeA.00983-16>.
4. Núñez F, Lara MS, Peromingo B, Delgado J, Sánchez-Montero L, Andrade MJ. 2015. Selection and evaluation of *Debaryomyces hansenii* isolates as potential bioprotective agents against toxigenic penicillia in dry-fermented sausages. *Food Microbiol* 46:114–120. <https://doi.org/10.1016/j.fm.2014.07.019>.
5. Van Wyk N, Pretorius IS, Von Wallbrunn C. 2020. Assessing the oenological potential of *Nakazawaea ishiwadae*, *Candida railenensis* and *Debaryomyces hansenii* strains in mixed-culture grape must fermentation with *Saccharomyces cerevisiae*. *Fermentation* 6:49. <https://doi.org/10.3390/fermentation6020049>.
6. Lülfi RH, Vogel RF, Ehrmann MA. 2021. Microbiota dynamics and volatile compounds in lupine based Moromi fermented at different salt concentrations. *Int J Food Microbiol* 354:109316. <https://doi.org/10.1016/j.ijfoodmicro.2021.109316>.
7. Hoog GS, Ende AHGG. 1998. Molecular diagnostics of clinical strains of filamentous Basidiomycetes. *Mycoses* 41:183–189. <https://doi.org/10.1111/j.1439-0507.1998.tb00321.x>.
8. Wagner T, Fischer M. 2001. Natural groups and a revised system for the European poroid *Hymenochaetales* (Basidiomycota) supported by nLSU rDNA sequence data. *Mycol Res* 105:773–782. <https://doi.org/10.1017/S0953756201004257>.
9. Groenewald M, Daniel H-M, Robert V, Poot GA, Smith MT. 2008. Polyphasic re-examination of *Debaryomyces hansenii* strains and reinstatement of *D. hansenii*, *D. fabryi* and *D. subglobosus*. *Persoonia* 21:17–27. <https://doi.org/10.3767/003158508X336576>.
10. Corredor M, Davila A-M, Casarégola S, Gaillardin C. 2003. Chromosomal polymorphism in the yeast species *Debaryomyces hansenii*. *Antonie Van Leeuwenhoek* 84:81–88. <https://doi.org/10.1023/a:1025432721866>.
11. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. 2018. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun* 9:5114. <https://doi.org/10.1038/s41467-018-07641-9>.
12. Gremme G, Steinbiss S, Kurtz S. 2013. GenomeTools: a comprehensive software library for efficient processing of structured genome annotations. *IEEE/ACM Trans Comput Biol Bioinform* 10:645–656. <https://doi.org/10.1109/TCBB.2013.68>.
13. Haubold B, Pfaffelhuber P, Domazet-Lošo M, Wiehe T. 2009. Estimating mutation distances from unaligned genomes. *J Comput Biol* 16:1487–1500. <https://doi.org/10.1089/cmb.2009.0106>.
14. Cabanettes F, Klopp C. 2018. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. *PeerJ* 6:e4958. <https://doi.org/10.7717/peerj.4958>.
15. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4:406–425. <https://doi.org/10.1093/oxfordjournals.molbev.a040454>.
16. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* (N Y) 39:783–791. <https://doi.org/10.2307/2408678>.
17. Tamura K, Nei M, Kumar S. 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. *Proc Natl Acad Sci U S A* 101:11030–11035. <https://doi.org/10.1073/pnas.0404206101>.
18. Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874. <https://doi.org/10.1093/molbev/msw054>.