

Chromosome level genome assembly and annotation of *Hanseniaspora mollemarum* CBS 18055 strain

Bennis, Nicole Xanthe; Hanji, Harshitha Indudhar; van den Broek, Marcel; Daran, Jean Marc

DOI

[10.1128/mra.01005-24](https://doi.org/10.1128/mra.01005-24)

Publication date

2025

Document Version

Final published version

Published in

Microbiology Resource Announcements

Citation (APA)

Bennis, N. X., Hanji, H. I., van den Broek, M., & Daran, J. M. (2025). Chromosome level genome assembly and annotation of *Hanseniaspora mollemarum* CBS 18055 strain. *Microbiology Resource Announcements*, 14(3). <https://doi.org/10.1128/mra.01005-24>

Important note

To cite this publication, please use the final published version (if applicable). Please check the document version above.

Copyright

Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy

Please contact us and provide details if you believe this document breaches copyrights. We will remove access to the work immediately and investigate your claim.

Chromosome level genome assembly and annotation of *Hanseniaspora mollemarum* CBS 18055 strain

Nicole Xanthe Bennis,¹ Harshitha Indudhar Hanji,¹ Marcel van den Broek,¹ Jean-Marc Daran¹

AUTHOR AFFILIATION See affiliation list on p. 2.

ABSTRACT *Hanseniaspora* species gained attention due to the ability of these species to ferment simple sugars and to actively contribute to the development of bouquet aromas in wine and cider fermentations. We present a chromosome-level assembly of an isolate of *Hanseniaspora mollemarum* that would enhance its potential applications.

KEYWORDS yeasts, saccharomycotina, *Hanseniaspora*, genomes

The *Hanseniaspora* genus comprises over 20 species including the recently described *Hanseniaspora gamundiae* (1), *Hanseniaspora smithiae* (2), *Hanseniaspora terricola* (3), and *Hanseniaspora mollemarum*. The latter species was isolated in a soil sample from a Dutch garden (CBS 15034) (4). Subsequently, new strains from Great-Britain and Hungary have been deposited at the Westerdijk Fungal Biodiversity Institute (<https://wi.knaw.nl/>). However, the sole *H. mollemarum* genome currently available derives from the CBS 15034 strain. To match the interest of the beverage industry for *Hanseniaspora* strains, it is critical to gain both physiological and genomic knowledge. Here, we report the genome sequence of a new strain of *H. mollemarum*.

The *H. mollemarum* CBS 18055 was isolated from an apple tree (*Malus domestica*) located in the south of France (43°01'24.7"N, 0°58'30.4"E). The sample was grown at 4°C on YPD medium (10 g L⁻¹ yeast extract, 20 g L⁻¹ Bacto-peptone, and 20 g L⁻¹ glucose) supplemented with 3% ethanol. Single colonies were obtained by re-streaking the culture on a solid YPD medium supplemented with 25 µg mL⁻¹ chloramphenicol and 50 µg mL⁻¹ ampicillin incubated at room temperature. This process was repeated thrice. Final isolates were stored in 30% (vol/vol) glycerol at -80°C.

Sanger sequencing of the ITS and the D1/D2 region of the rDNA of CBS 18055 amplified with primers ITS1/ITS4 (5) and NL1/NL4 (6), respectively, revealed a 99.98 and 100% identity to that of *H. mollemarum* CBS 15034, respectively.

For genome sequencing, total DNA was extracted using the Genomic-tip 100 /G kit protocol (Qiagen, Hilden, Germany). Shotgun library was prepared using Oxford Nanopore Technologies' SQK-LSK109 kit (ONT, Oxford, United Kingdom) and sequenced on a MinION MK1B device with R10 flow cell. Raw FAST5 signal files were base-called using GPU Guppy (ONT, version 4.5.4) in high-accuracy mode. After filtering for length (>1 kb), 94,060 reads were obtained, yielding 1.04 Gb (N50: 19,089 bp), representing ~112-fold coverage of a *H. mollemarum* genome. Canu version 2.0 with settings genomeSize = 12 m, useGrid = 0, and nanopore-raw was used for *de novo* assembly yielding eight contigs (7). To correct errors in the assembly, a 150 bp read length TruSeq PCR-Free Illumina library (Illumina, San Diego, CA) with a 350 bp insert-size was sequenced on a NovaSeq 6000 (Illumina) by Novogene (Cambridge, United Kingdom) yielding 18,811,684 reads for a total of 2.82 Gigabases (307-fold coverage). Reads quality was assessed with FastQC v0.11.5. The genome was polished by mapping the untrimmed Illumina reads with Burrows-Wheeler Aligner (BWA version 0.7.15-r1142-dirty; default parameters used)

Editor Jason E. Stajich, University of California Riverside, Riverside, California, USA

Address correspondence to Jean-Marc Daran, j.g.daran@tudelft.nl.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 4 October 2024

Accepted 15 January 2025

Published 29 January 2025

Copyright © 2025 Bennis 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/).

TABLE 1 Statistics of the *H. mollemarum* assembly GCA_042466605.1

			Canu (ONT)	UniCycler hybrid
Assembly size (Mb)			9.18 ($n = 105$)	
N50 (Mb)			1.25 ($n = 4$)	
N90 (Kb)			105.20 ($n = 9$)	
	GenBank	Chr	Size (bp)	Size (bp)
contig1 (bp)	CP157652.1	Chr 8	1,960,557	
contig2 (bp)	CP157651.1	Chr 7	1,481,168	
contig3 (bp)	CP157650.1	Chr 6	1,331,897	
contig4 (bp)	CP157649.1	Chr 5	1,256,963	
contig5 (bp)	CP157648.1	Chr 4	1,027,777	
contig6 (bp)	CP157647.1	Chr 3	994,132	
contig7 (bp)	CP157646.1	Chr 2	662,091	
contig8 (bp)	CP157645.1	Chr 1	469,443	
contig9 – circular (bp)	CP157653.1	mtDNA		32,902
		Nuclear DNA (bp)	9,184,028	
		Total (bp)		9,216,930

(8, 9) to the assembly and further processed with SAMtools (version 1.3.1) (10, 11) and polished once by Pilon (version 1.18) with settings fix all (12).

This yielded a genome assembly of 9.18 Mb (N50 = 1.25 Mb) with a GC content of 34.96%, comprising eight contigs. Contig 1–7 (CHR2-to-8) showed telomeric sequences at both ends, whereas contig 8 (CHR1) had telomere sequences on one chromosome arm indicating that this contig also represented (part of) a linear chromosome (Table 1). Due to Canu's being notoriously known to have difficulty in assembling circular contigs, the fastq files were assembled using UniCycler Hybrid assembler (13). A circular contig of 32.9 kbp was identified as mitochondrial chromosome (Table 1). Annotation of the polished assembly was performed using Funannotate v1.8.15 (<https://github.com/nextgenusfu/funannotate>) (14) using ab initio gene predictors SNAP (v2013_11_29), Augustus (3.5.0), Genemark.HMM (ES Suite v4), and glimmerhmm (v3.0.4) in combination with default Funannotate databases, for example, pfam, interpro, and Uniprot as previously described in (15).

ACKNOWLEDGMENTS

Nicole X. Bennis was supported by funding of the Top consortia for Knowledge and Innovation (TKIs) AgriFood which was granted a PPP allowance from the Ministry of Economic Affairs and Climate Policy (Project Habitats #PPPS1701).

AUTHOR AFFILIATION

¹Department of Biotechnology, Delft University of Technology, Delft, the Netherlands

AUTHOR ORCIDs

Nicole Xanthe Bennis  <http://orcid.org/0000-0002-3403-0844>

Marcel van den Broek  <http://orcid.org/0000-0002-7734-582X>

Jean-Marc Daran  <http://orcid.org/0000-0003-3136-8193>

FUNDING

Funder	Grant(s)	Author(s)
EZK Topconsortium voor Kennis en Innovatie (TKI)	PPPS1701	Marcel van den Broek

AUTHOR CONTRIBUTIONS

Harshitha Indudhar Hanji, Investigation, Methodology | Marcel van den Broek, Data curation, Formal analysis, Methodology, Resources, Software, Validation | Jean-Marc Daran, Conceptualization, Funding acquisition, Investigation, Supervision, Writing – original draft, Writing – review and editing.

DATA AVAILABILITY

The genome reported in this publication was deposited at NCBI (<https://www.ncbi.nlm.nih.gov/>) under the BioProject accession number [PRJNA977567](https://doi.org/10.1093/bioinformatics/btp698). The BioProject includes the assembly [GCA_042466605.1](https://doi.org/10.1093/gigascience/giab008) and the associated Illumina and Nanopore reads [SRX23808396](https://doi.org/10.1093/gigascience/giab008) and [SRX23808395](https://doi.org/10.1093/gigascience/giab008), respectively. The annotation was deposited at the 4TU repository (<https://data.4tu.nl/>) under URL <https://doi.org/10.4121/06bb6616-2da3-4211-8694-ff5ae37935dc.v1>.

REFERENCES

- Čadež N, Bellora N, Ulloa R, Hittinger CT, Libkind D. 2019. Genomic content of a novel yeast species *Hanseniaspora gamundiae* sp. nov. from fungal stromata (Cyttaria) associated with a unique fermented beverage in Andean Patagonia, Argentina. *PLoS ONE* 14:e0210792. <https://doi.org/10.1371/journal.pone.0210792>
- Čadež N, Bellora N, Ulloa R, Tome M, Petković H, Groenewald M, Hittinger CT, Libkind D. 2021. *Hanseniaspora smithiae* sp. nov., a novel apiculate yeast species from patagonian forests that lacks the typical genomic domestication signatures for fermentative environments. *Front Microbiol* 12:679894. <https://doi.org/10.3389/fmicb.2021.679894>
- Liu Z, Wang M-M, Wang G-S, Li A-H, Wang Q-M. 2019. *Hanseniaspora terricola* sp. nov., an ascomycetous yeast isolated from Tibet. *Int J Syst Evol Microbiol* 71. <https://doi.org/10.1099/ijsem.0.004741>
- Groenewald M, Lombard L, de Vries M, Lopez AG, Smith M, Crous PW. 2018. Diversity of yeast species from Dutch garden soil and the description of six novel Ascomycetes. *FEMS Yeast Res* 18. <https://doi.org/10.1093/femsyr/foy076>
- White TJ, Bruns T, Lee S, Taylor J. 1990. Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics, p 315–322. In Innis MA, Gelfand DH, Sninsky JJ, White TJ (ed), *PCR Protocols*. Academic Press, San Diego.
- O'Donnell K. 1993. *Fusarium* and its near relatives, p 225–233. In Reynolds DR, Taylor JW (ed), *The fungal holomorph: mitotic, meiotic and pleomorphic speciation in fungal systematic*. CAB International, Wallingford.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv Genomics*. <https://doi.org/10.48550/arXiv.1303.3997>
- Li H, Durbin R. 2010. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics* 26:589–595. <https://doi.org/10.1093/bioinformatics/btp698>
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. *Gigascience* 10:giab008. <https://doi.org/10.1093/gigascience/giab008>
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLOS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- Palmer JM, Stajich J. 2020. Funannotate v1.8.1: eukaryotic genome annotation (v1.8.1). Zenodo. <https://doi.org/10.5281/zenodo.1134477>
- van den Broek M, Ortiz-Merino RA, Bennis NX, Wronska AK, Hassing E-J, Daran-Lapujade P, Daran J-M. 2024. Draft genome sequence of the *Saccharomyces cerevisiae* SpyCas9 expressing strain IMX2600, a laboratory and platform strain from the CEN.PK lineage for cell-factory research. *Microbiol Resour Announc* 13:e0055023. <https://doi.org/10.1128/mra.00550-23>