AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

# Genome Sequence of the Yeast *Saprochaete ingens* CBS 517.90

**Viktória Hodorová,[a] Hana Lichancová,[a] Stanislav Zubenko,[b] Karolina Sienkiewicz,[c] Sarah Mae U. Penir,[d] Philipp Afanasyev,[e] Dominic Boceck,[f] Sarah Bonnin,[g] Siras Hakobyan,[h] Urszula Smyczynska,[i] Erik Zhivkoplias,[j] Maryna Zlatohurska,[k] Eugeniusz Tralle,[l] Alina Frolova,[b,m] Leszek P. Pryszcz,[g,l] Broňa Brejová,[n] Tomáš Vinař,[o] Jozef Nosek[a]**

[a]Department of Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Bratislava, Slovak Republic
[b]Kyiv Academic University, Kiev, Ukraine
[c]Faculty of Mathematics, Informatics and Mechanics, University of Warsaw, Warsaw, Poland
[d]Department of Meiosis, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany
[e]Laboratory of Evolutionary Genomics, Vavilov Institute of General Genetics, Moscow, Russia
[f]Algorithms in Bioinformatics, ZBIT Center for Bioinformatics, University of Tübingen, Tübingen, Germany
[g]Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology, Barcelona, Spain
[h]Institute of Molecular Biology NAS RA, Yerevan, Armenia
[i]Department of Biostatistics and Translational Medicine, Medical University of Lodz, Lodz, Poland
[j]Biology Education Centre, Uppsala University, Uppsala, Sweden
[k]Institute of Microbiology and Virology, National Academy of Science of Ukraine, Kiev, Ukraine
[l]International Institute of Molecular and Cell Biology in Warsaw, Warsaw, Poland
[m]Institute of Molecular Biology and Genetics, National Academy of Sciences of Ukraine, Kiev, Ukraine
[n]Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovak Republic
[o]Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Bratislava, Slovak Republic

**ABSTRACT** Chromosome-scale genome assembly of the yeast *Saprochaete ingens* CBS 517.90 was determined by a combination of technologies producing short (HiSeq X; Illumina) and long (MinION; Oxford Nanopore Technologies) reads. The 21.2-Mbp genome sequence has a GC content of 36.9% and codes for 6,475 predicted proteins.

The yeast *Saprochaete ingens* was originally described as *Candida ingens* (1) and later classified into the *Magnusiomyces/Saprochaete* clade (*Dipodascaceae*, Saccharomycotina, Ascomycota). In this clade, teleomorphic and anamorphic stages were named *Magnusiomyces* and *Saprochaete*, respectively. To investigate claims that *Saprochaete ingens* and *Magnusiomyces ingens* do not represent different reproductive stages of the same species but rather distinct taxa (2–4), we sequenced the genome of *S. ingens* ex-holotype strain CBS 517.90, isolated from a wine cellar in Western Cape Province, South Africa (1), and compared it to the recently determined *M. ingens* genome (5).

The yeasts were grown overnight in yeast extract-peptone-dextrose (YPD) medium (1% [wt/vol] yeast extract, 2% [wt/vol] peptone, and 1% [wt/vol] glucose) at 28°C, and the genomic DNA was purified using a Genomic-tip 100/G (Qiagen) (6). A total of 111,042 long reads (mean, 13,586.5 nucleotides [nt]; median, 5,776 nt; longest read, 192,848 nt) totaling 1.5 Gbp (~71× coverage) were obtained with a MinION Mk-1B device on an R9.4.1 flow cell, using ligation kit SQK-LSK109, and base called by ONT Albacore (v. 2.3.1). A paired-end (2 × 151-nt) TruSeq PCR-free DNA library was sequenced on a HiSeq X Ten platform by Macrogen Korea, yielding 172,059,934 reads (25.98 Gbp; ~1,226× coverage). No additional read trimming or filtering was performed. Unless otherwise noted, all tools were used with default parameters.

Eleven contigs of the initial long-read assembly (miniasm v. 0.3-r179 [7]; minimap2 v. 2.13-r852 [option -x ava-ont] [8]; polished by Racon v. 1.3.1 [option –include-

**Citation** Hodorová V, Lichancová H, Zubenko S, Sienkiewicz K, Penir SMU, Afanasyev P, Boceck D, Bonnin S, Hakobyan S, Smyczynska U, Zhivkoplias E, Zlatohurska M, Tralle E, Frolova A, Pryszcz LP, Brejová B, Vinař T, Nosek J. 2019. Genome sequence of the yeast *Saprochaete ingens* CBS 517.90. Microbiol Resour Announc 8:e01366-19. https://doi.org/10.1128/MRA.01366-19.

**Editor** Christina A. Cuomo, Broad Institute



Address correspondence to Broňa Brejová, brejova@dcs.fmph.uniba.sk, or Jozef Nosek, jozef.nosek@uniba.sk.

**Received** 30 October 2019
**Accepted** 6 November 2019
**Published** 12 December 2019

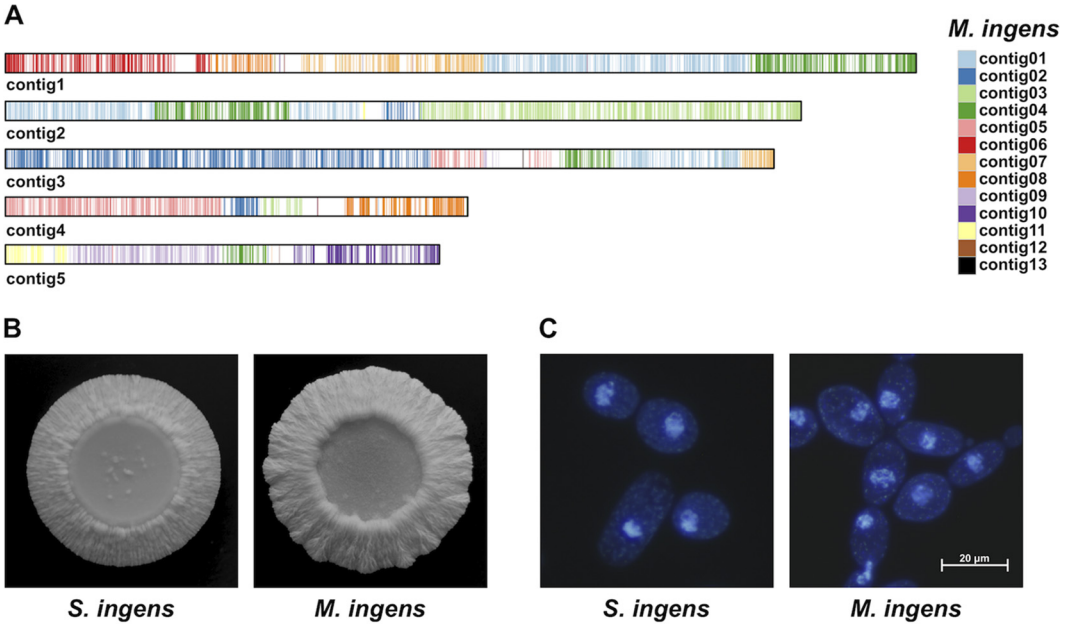


**FIG 1** (A) Nuclear contigs of *S. ingens* CBS 517.90 colored based on alignments to contigs of *M. ingens* NRRL Y-17630 (CBS 521.90) (5). The comparison was performed using Last Aligner v. 830 (option -E1e-10) (15), postprocessed by last-split to keep only the best match at each *M. ingens* locus, and visualized using ggplot2 (16). (B) Differentiated colonies of *S. ingens* CBS 517.90 and *M. ingens* NRRL Y-17630 grown on yeast extract-malt extract-peptone (YM) plates (0.3% [wt/vol] yeast extract, 0.3% [wt/vol] malt extract, and 0.5% [wt/vol] peptone) containing 1% (wt/vol) glucose at 28°C for about 2 weeks. (C) Nuclear and mitochondrial DNA of *S. ingens* CBS 517.90 and *M. ingens* NRRL Y-17630 cells stained with 4′,6-diamidino-2-phenylindole (DAPI) and visualized using an Olympus BX50 microscope.

unpolished] [9]) were compared with long-read assemblies by wtdgb2 v. 2.3 (options -g 20 m -x ont) (10) and Canu v. 1.7 (options genomeSize = 25m overlapper=mhap utgReAlign=true) (11). Based on the comparison, four pairs of contigs were connected, two contigs were extended to telomeres, and seven local misassemblies were corrected. A short contig containing only ribosomal DNA (rDNA) repeats was discarded, with and additional eight copies of rDNA present in contig 4. The resulting assembly was polished with short reads (four iterations of pilon v. 1.21 [12]; BWA-MEM v. 0.7.17-r1188 [option -M] [13]). The rDNA repeat and the mitochondrial genome were polished separately from the rest of the genome to avoid ambiguous alignments.

The assembly is 21.2 Mbp long and consists of five nuclear contigs (between 2.7 and 5.7 Mbp) and a mitochondrial genome (35.5 kbp). Nine nuclear contig ends are terminated by telomeric repeats $(CA_3G_{5-8})_n$, indicating five chromosomes with one telomeric region missing from the assembly. Genes were annotated using Augustus v. 3.2.3 (option –uniqueGeneId=true) (14), with initial parameters estimated from *Magnusiomyces capitatus* (5) and then trained on the 3,341 predicted *S. ingens* genes with at least 80% protein-level identity to their closest *M. ingens* ortholog. A total of 14 predictions were discarded due to in-frame stop codons, resulting in 6,475 nuclear protein-coding genes.

The nuclear genome comparison of *S. ingens* and *M. ingens* (Fig. 1A) shows that, although the genomes exhibit a long-range synteny, the alignments are fragmented and have only about 77% identity (median among alignments with at least 1,000 matches). The comparison thus demonstrates that, despite these two yeasts exhibiting many common features, such as similar assimilation profiles (3, 4) and colony and cell morphologies (Fig. 1B and C), they represent different species.

**Data availability.** The assembly has been deposited in ENA (accession no. CABVLU010000000). Illumina and MinION reads have been deposited under accession no. ERR3510534 and ERR3509916, respectively. The assembly and its annotation can also be viewed interactively in a genome browser available at http://genome.compbio.fmph.uniba.sk/.

## ACKNOWLEDGMENTS

*S. ingens* strain CBS 517.90 was purchased from the Westerdijk Fungal Biodiversity Institute (The Netherlands). This work was initiated at the #NGSchool2018: Nanopore sequencing & personalised medicine bioinformatics school organized in Lublin, Poland (16 to 23 September 2018; https://ngschool.eu/2018), supported by International Visegrad Fund project 21810033. The computations were done with the help of cloud services and resources from national e-infrastructure providers through the Training Infrastructure of the EGI Federation.

This project was supported by grants from the Slovak Research and Development Agency (APVV-14-0253 and APVV-18-0239 to J.N.) and the Scientific Grant Agency (VEGA 1/0684/16 to B.B., VEGA 1/0458/18 to T.V., and VEGA 1/0027/19 to J.N.), and from the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement 665778 (to L.P.P.).

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

1. van der Walt JP, van Kerken AE. 1961. *Candida ingens* nov. spec. Antonie Van Leeuwenhoek 27:284–286. https://doi.org/10.1007/bf02538457.
2. Smith MT, Poot GA. 2003. Genome comparisons in the genus *Dipodascus* de Lagerheim. FEMS Yeast Res 3:301–311. https://doi.org/10.1016/S1567-1356(03)00013-8.
3. de Hoog GS, Smith MT. 2011. *Magnusiomyces* Zender (1977), p 565–574. *In* Kurtzman CP, Fell JW, Boekhout T (ed), The yeasts: a taxonomic study, 5th ed. Elsevier, London, United Kingdom.
4. de Hoog GS, Smith MT. 2011. *Saprochaete* Coker & Shanor ex D.T.S. Wagner & Dawes (1970), p 1317–1327. *In* Kurtzman CP, Fell JW, Boekhout T (ed), The yeasts: a taxonomic study, 5th ed. Elsevier, London, United Kingdom.
5. Brejová B, Lichancová H, Brázdovič F, Hegedűsová E, Forgáčová Jakúbková M, Hodorová V, Džugasová V, Baláž A, Zeiselová L, Cillingová A, Neboháčová M, Raclavský V, Tomáška Ľ, Lang BF, Vinař T, Nosek J. 2019. Genome sequence of the opportunistic human pathogen *Magnusiomyces capitatus*. Curr Genet 65:539–560. https://doi.org/10.1007/s00294-018-0904-y.
6. Hodorova V, Lichancova H, Bujna D, Nebohacova M, Tomaska L, Brejova B, Vinar T, Nosek J. 2018. *De novo* sequencing and high-quality assembly of yeast genomes using a MinION device. London Calling, 24 to 25 May 2018, London, United Kingdom. https://nanoporetech.com/resource-centre/de-novo-sequencing-and-high-quality-assembly-yeast-genomes-using-minion-device.
7. Li H. 2016. Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. Bioinformatics 32:2103–2110. https://doi.org/10.1093/bioinformatics/btw152.
8. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
9. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. Genome Res 27:737–746. https://doi.org/10.1101/gr.214270.116.
10. Ruan J, Li H. 2019. Fast and accurate long-read assembly with wtdbg2. bioRxiv. https://doi.org/10.1101/530972.
11. 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.
12. 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.
13. 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.
14. Stanke M, Schöffmann O, Morgenstern B, Waack S. 2006. Gene prediction in eukaryotes with a generalized hidden Markov model that uses hints from external sources. BMC Bioinformatics 7:62. https://doi.org/10.1186/1471-2105-7-62.
15. Frith MC, Kawaguchi R. 2015. Split-alignment of genomes finds orthologies more accurately. Genome Biol 16:106. https://doi.org/10.1186/s13059-015-0670-9.
16. Wickham H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York, NY.