



# Genome Sequence of an Arthroconidial Yeast, *Saprochaete fungicola* CBS 625.85

 Broňa Brejová,<sup>a</sup> Hana Lichancová,<sup>b</sup> Viktória Hodorová,<sup>b</sup> Martina Neboháčová,<sup>b</sup>  Lubomír Tomáška,<sup>c</sup>  Tomáš Vinař,<sup>d</sup>  
 Jozef Nosek<sup>b</sup>

<sup>a</sup>Department of Computer Science, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, Bratislava, Slovak Republic

<sup>b</sup>Department of Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Ilkovičova, Bratislava, Slovak Republic

<sup>c</sup>Department of Genetics, Faculty of Natural Sciences, Comenius University in Bratislava, Ilkovičova, Bratislava, Slovak Republic

<sup>d</sup>Department of Applied Informatics, Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava, Mlynská dolina, Bratislava, Slovak Republic

**ABSTRACT** *Saprochaete fungicola* is an arthroconidial yeast classified in the *Magnusiomyces/Saprochaete* clade of the subphylum Saccharomycotina. Here, we report the genome sequence of holotype strain CBS 625.85, assembled to five putative chromosomes. The genome sequence is 20.2 Mbp long and codes for 6,138 predicted proteins.

*Saprochaete fungicola* is an anamorphic yeast reproducing by fragmentation of hyphae into asexual spores dubbed arthroconidia (1). Arthroconidia facilitate dissemination, and in some pathogenic fungi, their formation contributes to virulence (2). To provide a resource to study genetic control of arthroconidiogenesis, we determined the genome sequence of the strain CBS 625.85, originally isolated from ascocarps of *Nectria cinnabarina*, a plant pathogenic fungus causing coral spots (1). Genomic DNA sequencing was performed using a combination of HiSeq 2000 (Illumina) and MinION (Oxford Nanopore Technologies) platforms. DNA was isolated using a standard protocol (3), and total cellular RNA was prepared from a culture grown in yeast extract-peptone-galactose (YPGal) medium (1% [wt/vol] yeast extract, 2% [wt/vol] peptone, and 2% [wt/vol] galactose) at 28°C using hot phenol extraction (4) and an RNeasy minikit (Qiagen).

In total, 3.0 Gbp (~148× genome coverage) were sequenced in 309,350 long reads (mean, 9.7 kbp; longest read, 251 kbp) using a MinION Mk-1B device with an R9.4.1 flow cell and SQK-LSK109 kit. The paired-end (2 × 101 nucleotides [nt]) TruSeq PCR-free DNA library was sequenced on a HiSeq 2000 instrument at Macrogen Korea, yielding 40,313,550 short reads (4.1 Gbp; coverage, ~203×). Finally, 38,086,316 transcriptome sequencing (RNA-Seq) reads were generated from a TruSeq nonstranded mRNA paired-end (2 × 101 nt) library using a NovaSeq 6000 instrument at Macrogen Korea.

Assembly with SPAdes v. 3.12.0 (5) resulted in 367 contigs ( $N_{50}$  value, 0.4 Mbp), and long-read assembly with miniasm v. 0.3-r179 (6) and minimap2 v. 2.13-r852 (7) polished by racon v. 1.3.1 (8) yielded 10 contigs ( $N_{50}$  value, 5.1 Mbp). Two contigs contained mitochondrial DNA (mtDNA), of which one complete copy (circular 33-kbp contig) was retained for the final assembly. Two contigs not supported by Illumina reads were discarded as contamination. One ribosomal DNA (rDNA) contig was discarded, and eight copies of the rDNA cluster were retained elsewhere. To further correct the long-read assembly, the rDNA cluster was polished separately by two iterations of pilon v. 1.21 (9) with BWA-MEM v. 0.7.17-r1188 (10), and four contig ends were extended using SPAdes contigs. Identified by a significant decrease in long-read coverage, five local misassemblies were corrected, also using SPAdes assembly. All changes were based on reliable long-range overlaps and supported by multiple long reads. The final

**Citation** Brejová B, Lichancová H, Hodorová V, Neboháčová M, Tomáška L, Vinař T, Nosek J. 2019. Genome sequence of an arthroconidial yeast, *Saprochaete fungicola* CBS 625.85. *Microbiol Resour Announc* 8:e00092-19. <https://doi.org/10.1128/MRA.00092-19>.

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

**Copyright** © 2019 Brejová 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/).

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

**Received** 25 January 2019

**Accepted** 11 March 2019

**Published** 11 April 2019

assembly contains five nuclear contigs of lengths 5.4, 5.1, 4.4, 3.2, and 2.1 Mbp, with an overall G+C content of 40.6%. These likely correspond to full-length chromosomes since they terminate on both ends by putative telomeric repeats ( $[(AACAG)_{0-1}A_{2-6}G_{0-1}A_{0-1}G_{4-7}]$ ) with the predominating motif  $A_2GAG_6$ .

RNA-Seq reads, processed by Trimmomatic v. 0.36 (11), were assembled into transcripts by Trinity v. 2.8.4 (12) and aligned to the genome by blat v. 36 × 2 (13). Augustus v. 3.2.3 (14) trained on the related *Magnusiomyces capitatus* genome (15) with RNA-Seq evidence was used for initial gene predictions. Of these genes, 4,785 best supported by RNA-Seq transcripts (99% identity on 99% of length as identified by blat) were used to retrain Augustus parameters for *S. fungicola* and, together with the RNA-Seq evidence, to predict the final set of 6,138 protein-coding genes. The high-contiguity genome sequence of *S. fungicola* will be instrumental in comparative and functional studies focused on biology and evolution of arthroconidial yeasts.

**Data availability.** The genome assembly has been deposited in ENA under the accession number [CAACAHO10000000](https://ena.ebi.ac.uk/ena/record/CAACAHO10000000), and the Illumina, MinION, and RNA-Seq reads were deposited in SRA under the accession numbers [ERR3046939](https://www.ncbi.nlm.nih.gov/sra/ERR3046939), [ERR3046967](https://www.ncbi.nlm.nih.gov/sra/ERR3046967), and [ERR3046965](https://www.ncbi.nlm.nih.gov/sra/ERR3046965), respectively. The genome annotations are available through a genome browser at <http://genome.compbio.fmph.uniba.sk/> and are also archived through Zenodo (16).

## ACKNOWLEDGMENTS

This project was supported by grants from the Slovak Research and Development Agency (APVV-14-0253 to J.N. and 15-0022 to L.T.) and VEGA (1/0052/16 to L.T., 1/0684/16 to B.B., and 1/0458/18 to T.V.). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

## REFERENCES

- 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 (eds), *The yeasts: a taxonomic study*, 5th ed, vol 2. Elsevier, London, United Kingdom. <https://doi.org/10.1016/B978-0-444-52149-1.00097-5>.
- Barrera CR, Szanislo PJ. 1985. Formation and germination of fungal arthroconidia. *Crit Rev Microbiol* 12:271–292. <https://doi.org/10.3109/10408418509104431>.
- 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, 24th–25th May 2018, London, United Kingdom. <https://nanoporetech.com/resource-centre/de-novo-sequencing-and-high-quality-assembly-yeast-genomes-using-minion-device>.
- Collart MA, Oliviero S. 1993. Preparation of RNA. *Curr Protoc Mol Biol* 23:13.12.1–13.12.5. <https://doi.org/10.1002/0471142727.mb1312s23>.
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyskin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 19:455–477. <https://doi.org/10.1089/cmb.2012.0021>.
- Li H. 2016. Minimap and minimap: fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* 32:2103–2110. <https://doi.org/10.1093/bioinformatics/btw152>.
- Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>.
- 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>.
- 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>.
- 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>.
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>.
- Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* 29:644–652. <https://doi.org/10.1038/nbt.1883>.
- Kent WJ. 2002. BLAT—the BLAST-like alignment tool. *Genome Res* 12:656–664. <https://doi.org/10.1101/gr.229202>.
- 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>.
- Brejová B, Lichancová H, Brázdovič F, Hegedúsová E, Forgáčová Jakúbková M, Hodorová V, Džugasová V, Baláž A, Zeiselová L, Cillingová A, Neboháčová M, Raclavský V, Tomáška L, 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>.
- Brejová B, Lichancová H, Hodorová V, Neboháčová M, Tomáška L, Vinař T, Nosek J. 2019. Annotations of sapFunA1 genome assembly (version v2) [data set]. Zenodo. <https://doi.org/10.5281/zenodo.2566797>.