AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Long-Read Genome Assembly of *Saccharomyces uvarum* Strain CBS 7001

Jingxuan Chen,[a] **David J. Garfinkel,**[b,c] **Casey M. Bergman**[a,b]

[a]Institute of Bioinformatics, University of Georgia, Athens, Georgia, USA
[b]Department of Genetics, University of Georgia, Athens, Georgia, USA
[c]Department of Biochemistry & Molecular Biology, University of Georgia, Athens, Georgia, USA

ABSTRACT Here, we report a long-read genome assembly for *Saccharomyces uvarum* strain CBS 7001 based on PacBio whole-genome shotgun sequence data. Our assembly provides an improved reference genome for an important yeast in the *Saccharomyces sensu stricto* clade.

Saccharomyces uvarum is a globally distributed yeast species that can be commonly extracted from low-temperature fermentations and other natural substrates (1–6). Low-coverage Sanger sequence assemblies were initially reported for an *S. uvarum* reference strain called CBS 7001 (also called MCYC 623) (7, 8). Later, Scannell et al. (9) generated a chromosome-level scaffolded assembly for CBS 7001 that combined Sanger and Illumina short-read sequences. All previous CBS 7001 assemblies contain a large number of gaps (Table 1), limiting the comprehensive analysis of transposable elements (TEs) in this species (9, 10). Here, we generated a long-read genome assembly for strain CBS 7001 to improve the reference genome and support analysis of TEs in *S. uvarum*.

To prepare DNA for PacBio sequencing, a single colony of strain CBS 7001 (NCYC 2909) was inoculated in 7 mL yeast extract-peptone-dextrose (YPD) liquid broth and cultured for ~24 h at 30°C. DNA was isolated using the Wizard genomic DNA purification kit (Promega), and a PacBio library was prepared using the SMRTbell Express template prep kit (Pacific Biosciences), following the >15-kb size-selection protocol that includes Covaris g-TUBE shearing. PacBio sequencing was performed using the Sequel II instrument (sequencing kit v2.1), yielding 655,409 reads (average length, 8,490 bp). HGAP4 (smrtlink-release_5.1.0.26412) (11) was used for *de novo* assembly, and the resulting PacBio contigs were polished five times using Pilon v1.24 (12) with 33,377,822 51-bp Illumina Genome Analyzer (GA) II reads (SRA accession number SRR173084). The polished contigs were scaffolded using RagTag v2.0.1 (13) with CBS 7001 ultrascaffolds from reference 9 that corrected the name swap for ChrX and ChrXII noted in reference 5. The assembly statistics were calculated using stats.sh from BBMap v38.90 (14) and BUSCO v5.0.0 (saccharomycetes_odb10) (15). TEs were annotated using a RepeatMasker-based pipeline described in reference 16. Genome alignment was performed using MUMmer v3.23 (17). The PacBio reads were aligned to the genome assemblies using minimap2 v2.17 with the parameters "-ax map-pb" (18). Default software parameters were used except where otherwise noted.

Our CBS 7001 assembly is 11,968,998 bp long (contig N_{50}, 836,618 bp) with an overall GC content of 40.18% (Table 1). Of the chromosomes, 15/16 are assembled in scaffolds with one contig each; ChrXII is assembled in one scaffold with one gap at the rDNA locus. In total, 97.3% of *Saccharomycetes* benchmarking universal single-copy orthologs (BUSCOs) are present and single copy. Our assembly was colinear with all scaffolds in the Scannell et al. (9) assembly except for one discrepancy at a sequencing gap in the previous assembly (ChrIII: 219401 to 219500). PacBio reads aligned to the

Editor Antonis Rokas, Vanderbilt University

Address correspondence to Casey M. Bergman, cbergman@uga.edu.

The authors declare no conflict of interest.

Received 29 September 2021
Accepted 4 December 2021
Published 6 January 2022

TABLE 1 Statistics for *de novo* assemblies of CBS 7001[a]

Assembly	NCBI accession no.	No. of contigs	No. of scaffolds	Total length (bp)	Contig N_{50} (bp)	Complete BUSCOs (%)	No. of full-length *Tsu4* copies
Kellis et al. (7)	GCA_000166995.1	1,098	1,098	11,475,890	4,795	88.3	0
Cliften et al. (8)	GCA_000167035.1	588	586	11,862,889	19,032	94.2	3
Scannell et al. (9)	NA[b]	363	16	11,502,913	18,172	97.1	0
This study	CM034478 to CM034493	17	16	11,968,998	836,618	97.3	9

[a]The contigs within a scaffold were defined by 10 consecutive Ns.

[b]NA, not applicable. The CBS 7001 genome assembly from Scannell et al. (9) is available at http://www.saccharomycessensustricto.org/.

previous assembly did not span this gap but did align continuously through the corresponding region in our assembly. Our assembly corrects this misassembly, the ChrX ↔ ChrXII name swap (5), and closes over 340 sequencing gaps in the previous assembly. Our assembly provides confirmation that *Tsu4* is the only TE with full-length copies in the CBS 7001 genome (19–22): we found nine full-length *Tsu4* copies in our assembly, tripling previous estimates (8, 23) (Table 1).

Data availability. The assembly produced here was deposited at NCBI under accession numbers CM034478 to CM034493. The PacBio data used to generate the assembly are available under SRA accession number PRJNA753102.

ACKNOWLEDGMENTS

We thank Marissa Howard, Noah Workman, and Magdy Alabady at the University of Georgia Genomics and Bioinformatics Core for assistance with the PacBio library preparation and sequencing, the Georgia Advanced Computing Resource Center for computational resources, and Caiti Smukowski Heil (North Carolina State University) for comments on the manuscript.

This work was funded by the University of Georgia Research Foundation (C.M.B.) and NIH grant R01GM124216 (D.J.G. and C.M.B.).

REFERENCES

1. Rainieri S, Zambonelli C, Hallsworth JE, Pulvirenti A, Giudici P. 1999. Saccharomyces uvarum, a distinct group within Saccharomyces sensu stricto. FEMS Microbiol Lett 177:177–185. https://doi.org/10.1111/j.1574-6968.1999.tb13729.x.
2. Pulvirenti A, Nguyen H, Caggia C, Giudici P, Rainieri S, Zambonelli C. 2000. Saccharomyces uvarum, a proper species within Saccharomyces sensu stricto. FEMS Microbiol Lett 192:191–196. https://doi.org/10.1111/j.1574-6968.2000.tb09381.x.
3. Nguyen HV, Gaillardin C. 2005. Evolutionary relationships between the former species Saccharomyces uvarum and the hybrids Saccharomyces bayanus and Saccharomyces pastorianus; reinstatement of Saccharomyces uvarum (Beijerinck) as a distinct species. FEMS Yeast Res 5:471–483. https://doi.org/10.1016/j.femsyr.2004.12.004.
4. Libkind D, Hittinger CT, Valerio E, Goncalves C, Dover J, Johnston M, Goncalves P, Sampaio JP. 2011. Microbe domestication and the identification of the wild genetic stock of lager-brewing yeast. Proc Natl Acad Sci U S A 108:14539–14544. https://doi.org/10.1073/pnas.1105430108.
5. Almeida P, Goncalves C, Teixeira S, Libkind D, Bontrager M, Masneuf-Pomarede I, Albertin W, Durrens P, Sherman DJ, Marullo P, Todd Hittinger C, Goncalves P, Sampaio JP. 2014. A Gondwanan imprint on global diversity and domestication of wine and cider yeast Saccharomyces uvarum. Nat Commun 5:4044. https://doi.org/10.1038/ncomms5044.
6. Rodríguez ME, Pérez-Través L, Sangorrín MP, Barrio E, Lopes CA. 2014. Saccharomyces eubayanus and Saccharomyces uvarum associated with the fermentation of Araucaria araucana seeds in Patagonia. FEMS Yeast Res 14:948–965. https://doi.org/10.1111/1567-1364.12183.
7. Kellis M, Patterson N, Endrizzi M, Birren B, Lander ES. 2003. Sequencing and comparison of yeast species to identify genes and regulatory elements. Nature 423:241–254. https://doi.org/10.1038/nature01644.
8. Cliften P, Sudarsanam P, Desikan A, Fulton L, Fulton B, Majors J, Waterston R, Cohen BA, Johnston M. 2003. Finding functional features in Saccharomyces genomes by phylogenetic footprinting. Science 301:71–76. https://doi.org/10.1126/science.1084337.
9. Scannell DR, Zill OA, Rokas A, Payen C, Dunham MJ, Eisen MB, Rine J, Johnston M, Hittinger CT. 2011. The awesome power of yeast evolutionary genetics: new genome sequences and strain resources for the Saccharomyces sensu stricto genus. G3 (Bethesda) 1:11–25. https://doi.org/10.1534/g3.111.000273.
10. Peona V, Blom MPK, Xu L, Burri R, Sullivan S, Bunikis I, Liachko I, Haryoko T, Jønsson KA, Zhou Q, Irestedt M, Suh A. 2021. Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. Mol Ecol Resour 21:263–286. https://doi.org/10.1111/1755-0998.13252.
11. Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods 10:563–569. https://doi.org/10.1038/nmeth.2474.
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. Alonge M, Soyk S, Ramakrishnan S, Wang X, Goodwin S, Sedlazeck FJ, Lippman ZB, Schatz MC. 2019. RaGOO: fast and accurate reference-guided scaffolding of draft genomes. Genome Biol 20:224. https://doi.org/10.1186/s13059-019-1829-6.
14. Bushnell B. 2014. BBMap: a fast, accurate, splice-aware aligner. Technical report LBNL-7065E. Lawrence Berkeley National Lab, Berkeley, CA.
15. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212. https://doi.org/10.1093/bioinformatics/btv351.
16. Czaja W, Bensasson D, Ahn HW, Garfinkel DJ, Bergman CM. 2020. Evolution of Ty1 copy number control in yeast by horizontal transfer and recombination. PLoS Genet 16:e1008632. https://doi.org/10.1371/journal.pgen.1008632.

17. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL. 2004. Versatile and open software for comparing large genomes. Genome Biol 5:R12. https://doi.org/10.1186/gb-2004-5-2-r12.
18. Li H. 2018. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
19. Bon E, Neuvéglise C, Casaregola S, Artiguenave F, Wincker P, Aigle M, Durrens P. 2000. Genomic exploration of the hemiascomycetous yeasts: 5. Saccharomyces bayanus var. uvarum. FEBS Lett 487:37–41. https://doi.org/10.1016/S0014-5793(00)02276-6.
20. Neuveglise C, Feldmann H, Bon E, Gaillardin C, Casaregola S. 2002. Genomic evolution of the long terminal repeat retrotransposons in hemiascomycetous yeasts. Genome Res 12:930–943. https://doi.org/10.1101/gr.219202.
21. Liti G, Peruffo A, James SA, Roberts IN, Louis EJ. 2005. Inferences of evolutionary relationships from a population survey of LTR-retrotransposons and telomeric associated sequences in the Saccharomyces sensu stricto complex. Yeast 22:177–192. https://doi.org/10.1002/yea.1200.
22. Smukowski HC, Patterson K, Shang M, Hickey A, Alcantara E, Dunham MJ. 2021. Transposable element mobilization in interspecific yeast hybrids. Genome Biol Evol 13:evab033. https://doi.org/10.1093/gbe/evab033.
23. Bergman CM. 2018. Horizontal transfer and proliferation of Tsu4 in Saccharomyces paradoxus. Mob DNA 9:18. https://doi.org/10.1186/s13100-018-0122-7.