AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements™

EUKARYOTES

# Complete Sequence of the Intronless Mitochondrial Genome of the *Saccharomyces cerevisiae* Strain CW252

**Delphine Naquin,[a] Cristina Panozzo,[b] Geneviève Dujardin,[b] Erwin van Dijk,[a] Yves d'Aubenton-Carafa,[a] Claude Thermes[a]**

[a]I2BC Next-Generation Sequencing Facility, Institut de Biologie Intégrative de la Cellule, UMR9198, CNRS CEA Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

[b]Biogenesis and Functioning of Mitochondrial Oxphos Complexes, Institut de Biologie Intégrative de la Cellule, UMR9198, CNRS CEA Université Paris-Sud, Université Paris-Saclay, Gif-sur-Yvette, France

**ABSTRACT** The mitochondrial genomes of *Saccharomyces cerevisiae* strains contain up to 13 introns. An intronless recombinant genome introduced into the nuclear background of *S. cerevisiae* strain W303 gave the *S. cerevisiae* CW252 strain, which is used to model mitochondrial respiratory pathologies. The complete sequence of this mitochondrial genome was obtained using a hybrid assembling methodology.

The mitochondrial genome of the *Saccharomyces cerevisiae* CW252 strain (1, 2) was sequenced using both Illumina short-read and MinION long-read sequencing technologies. In the first step, reads issued from 2 runs of the Illumina MiSeq platform and GAIIx system with a quality score greater than 30 were purified by mapping with TopHat2 (3) on the mitochondrial genome of the *S. cerevisiae* S288c parent strain; they were then sampled by Trinity (4) to obtain an estimated coverage of 30×. An assembly of the genome obtained by Velvet (5) with a k-mer size of 55 produced 34 contigs ($N_{50}$, 3.7 kb), which were subsequently scaffolded by SSPACE (6) to obtain 10 contigs ($N_{50}$, 57.5 kb). Some of these scaffolds seemed to be chimeric, since a mapping of the original paired-end reads on this assembly by BWA (7) displayed only 85% success. A correction step was then performed by REAPR (8) to break the chimeric scaffolds, producing a coherent assembly of 38 scaffolds of good quality ($N_{50}$, 2 kb). Examination of the mapping of these scaffolds on the S288c mitochondrial genome showed that many of the gaps between scaffolds contained GC-rich clusters which could hamper Illumina sequencing (9). In a second step, two-dimensional (2D) base calling of a single run on a MinION R9.4 flow cell, using Metrichor, produced 29,276 reads with a mean length of 4 kb and a total length of 120,259 kb. The correction, trimming, and assembly of these reads by Canu version 1.4 (10) produced 2 contigs, one of 84 kb corresponding to the mitochondrial genome, and another of 34 kb corresponding to contamination from the nuclear genome. The mitochondrial contig was first trimmed at both ends to remove overlapping extremities. To improve the quality of the MinION sequencing, the assembly was polished using two strategies, (i) the Pilon version 1.21 hybrid method (11) to directly correct the assembly using the Illumina reads by five iterative correction steps and (ii) a prior polishing of the sequence by Nanopolish (12), followed by four correction steps by Pilon. The second method gave the best results, with approximately 2-fold fewer nucleotide variations than the Illumina scaffolds, and the assembly was retained as a bona fide backbone. To improve the final assembly, the scaffolds issued from the Illumina assembly were replaced on this backbone, and the resulting mitochondrial genome was then annotated. The unique contig sequence was 70,523 bp long, and no nucleotide variation was shown in several regions spanning 8,500 nucleotides (nt) sequenced by the Sanger method. This genome displays the mosaic structure of its construction. Particularly, the 3′ sequence of the *COX3* gene is different

**Received** 21 February 2018 **Accepted** 24 March 2018 **Published** 26 April 2018

**Citation** Naquin D, Panozzo C, Dujardin G, van Dijk E, d'Aubenton-Carafa Y, Thermes C. 2018. Complete sequence of the intronless mitochondrial genome of the *Saccharomyces cerevisiae* strain CW252. Genome Announc 6: e00219-18. https://doi.org/10.1128/genomeA.00219-18.



Address correspondence to Delphine Naquin, delphine.naquin@i2bc.paris-saclay.fr.

D.N. and C.P. contributed equally to this work.

from that of its S288C parent strain and corresponds to the sequence of the *S. cerevisiae* D273-10B strain (13).

**Accession number(s).** The genome sequence is available in GenBank with accession number MG916964.

## ACKNOWLEDGMENTS

Illumina and MinION library preparation, sequencing, and bioinformatics analyses were performed at the I2BC Next-Generation Sequencing facility.

We thank Christopher Herbert for critical reading of the manuscript.

## REFERENCES

1. Seraphin B, Boulet A, Simon M, Faye G. 1987. Construction of a yeast strain devoid of mitochondrial introns and its use to screen nuclear genes involved in mitochondrial splicing. Proc Natl Acad Sci U S A 84:6810–6814.
2. Saint-Georges Y, Bonnefoy N, di Rago J-P, Chiron S, Dujardin G. 2002. A pathogenic cytochrome *b* mutation reveals new interactions between subunits of the mitochondrial $bc_1$ complex. J Biol Chem 277:49397–49402. https://doi.org/10.1074/jbc.M207219200.
3. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg S. 2013. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol 14:R36. https://doi.org/10.1186/gb-2013-14-4-r36.
4. 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.
5. Zerbino DR, Birney E. 2008. Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
6. Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27:578–579. https://doi.org/10.1093/bioinformatics/btq683.
7. Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
8. Hunt M, Kikuchi T, Sanders M, Newbold C, Berriman M, Otto TD. 2013. REAPR: a universal tool for genome assembly evaluation. Genome Biol 14:R47. https://doi.org/10.1186/gb-2013-14-5-r47.
9. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. 2013. Characterizing and measuring bias in sequence data. Genome Biol 14:R51. https://doi.org/10.1186/gb-2013-14-5-r51.
10. 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.
11. 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.
12. Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled *de novo* using only nanopore sequencing data. Nat Methods 12:733–735. https://doi.org/10.1038/nmeth.3444.
13. Thalenfeld BE, Tzagoloff A. 1980. Assembly of the mitochondrial membrane system. Sequence of the *oxi 2* gene of yeast mitochondrial DNA. J Biol Chem 255:6173–6180.