Draft Genome Sequence of *Saccharomyces cerevisiae* LW2591Y, a Laboratory Strain for *In Vivo* Multigene Assemblies

Jennifer Gerke,[a] Dominik Schneider,[b] Anja Poehlein,[b] Virginia W. Cornish,[c] Gerhard H. Braus[a]

[a]Molecular Microbiology and Genetics, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Göttingen, Germany
[b]Genomic and Applied Microbiology & Göttingen Genomics Laboratory, Institute of Microbiology and Genetics, Georg-August-Universität Göttingen, Göttingen, Germany
[c]Department of Chemistry, Columbia University, New York, New York, USA

ABSTRACT *Saccharomyces cerevisiae* is an industrially preferred cell factory for the heterologous production of proteins and chemicals. Here, we present the draft genome sequence of the laboratory strain *Saccharomyces cerevisiae* LW2591Y, which has been designed for robust and efficient assembly of multigene pathways.

*S*accharomyces cerevisiae LW2591Y has been developed for the *in vivo* chromosomal assembly of multigene pathways for synthetic biology (1). The reiterative recombination assembly method is a simple, efficient, and robust way to assemble an indefinite number of DNA constructs. Recently, LW2591Y was used for the heterologous production of the fragrance geraniol (2) and for the development of a colorimetric assay to detect pathogen-derived peptides (3).

LW2591Y was cultivated in yeast-extract-peptone-dextrose medium overnight, and total genomic DNA was isolated by potassium acetate extraction (2). Illumina shotgun libraries were prepared using the NEBNext Ultra II DNA library prep kit with beads (New England Biolabs). Sequencing was performed on a MiSeq system with the v3 reagent kit (Illumina) with 600 cycles, resulting in 6,729,204 paired-end reads. For long-read sequencing, libraries were prepared from high-molecular-weight DNA using the ligation sequencing kit 1D (SQK-LSK109) and the native barcode expansion kit (barcode 7, EXP-NBD104; barcode 14, EXP-NBD114) (Oxford Nanopore Technologies). Sequencing was performed twice on a MinION Mk1B device and a SpotON R9.4.1 flow cell (Oxford Nanopore Technologies) for 96 h using MinKNOW v19.12.5 software for sequencing and Guppy v4.0.15 for demultiplexing and base calling, resulting in 786,167 reads.

Default parameters were used for all software unless otherwise specified. Short-reads were quality filtered by fastp v0.20.0 (4) (base Phred score, ≥Q20; correction by overlap, read clipping with a sliding window of 4 with a Phred score of Q ≥ 20; required minimum length, ≥50 bp). Reads were adapter trimmed by Cutadapt v2.5 (5). Potential phiX contamination was removed by mapping the quality-filtered short reads against another sequence (GenBank accession number NC_001422) using Bowtie 2 v2.3.5.1 (6). The Nanopore reads were processed with fastp v0.20.0 (4) (base Phred score, ≥Q10; clipping by sliding window of 10 with a Phred score of ≥Q10; required minimum length, ≥5,000 bp), followed by Porechop v0.2.4 (https://github.com/rrwick/Porechop). After processing, 6,525,536 short reads with an average length of 257 bp and 387,790 long reads with an average length of 9,314 bp (N_{50}, 9,753 bp [7]) were obtained.

De novo hybrid assembly was performed with Unicycler v0.4.8 (8) in normal mode. Contigs with less than 200 bp were removed. The assembly resulted in 29 contigs, which were aligned using BLASTn (9) with the *Saccharomyces cerevisiae* S288C

reference chromosomes and mitochondrial DNA (GenBank accession number GCA _000146045.2), as well as a 2-μm plasmid (J01347.1). The 27 contigs assigned to the S288C chromosomes were subsequently mapped with Mauve v20150226 (build 10) (10) and Lasergene v17.1 (DNASTAR, Madison, WI) and scaffolded manually by adding 100 Ns. Short contigs mapping to repetitive elements of S288C chromosome 12 were inserted only once.

The draft genome sequence comprises 11,853,285 bp with an N_{50} value of 799,432 bp (7). Coverages were estimated with QualiMap v2.2.2 (11) using Bowtie 2 v2.3.5.1 (6) and minimap2 v2.17-r941 (12). The short-read coverages were 113$\times$ for the chromosomal DNA, 2,654$\times$ for the mitochondrial DNA, and 424$\times$ for the plasmid DNA. The long-read coverages were 278$\times$ for the chromosomal DNA, 2,148$\times$ for the mitochondrial DNA, and 259$\times$ for the plasmid DNA. The total GC content is 38.1%.

A breseq v0.35.0 analysis (2, 13) of the short reads against the reference strain S288C revealed 235 single nucleotide polymorphisms and insertions/deletions with 100% frequency (see Table S1 at https://figshare.com/s/acf5207e3169bc9e8fc1). In 56 genes, the mutations induce changes in the corresponding amino acid sequences.

Data availability. The data for *Saccharomyces cerevisiae* LW2591Y have been deposited at GenBank under accession numbers CP059522 through CP059539, with BioProject accession number PRJNA611915 and SRA accession numbers SRX9268844 (Oxford Nanopore) and SRX7891670 (MiSeq).

REFERENCES

1. Wingler LM, Cornish VW. 2011. Reiterative recombination for the in vivo assembly of libraries of multigene pathways. Proc Natl Acad Sci U S A 108:15135–15140. https://doi.org/10.1073/pnas.1100507108.
2. Gerke J, Frauendorf H, Schneider D, Wintergoller M, Hofmeister T, Poehlein A, Zebec Z, Takano E, Scrutton NS, Braus GH. 2020. Production of the fragrance geraniol in peroxisomes of a product-tolerant baker's yeast. Front Bioeng Biotechnol 8:582052. https://doi.org/10.3389/fbioe.2020.582052.
3. Ostrov N, Jimenez M, Billerbeck S, Brisbois J, Matragrano J, Ager A, Cornish VW. 2017. A modular yeast biosensor for low-cost point-of-care pathogen detection. Sci Adv 3:e1603221. https://doi.org/10.1126/sciadv.1603221.
4. Chen S, Zhou Y, Chen Y, Gu J. 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34:i884–i890. https://doi.org/10.1093/bioinformatics/bty560.
5. Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J 17:10–12. https://doi.org/10.14806/ej.17.1.200.
6. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
7. Tan G, Polychronopoulos D, Lenhard B. 2019. CNEr: a toolkit for exploring extreme noncoding conservation. PLoS Comput Biol 15:e1006940. https://doi.org/10.1371/journal.pcbi.1006940.
8. 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.
9. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
10. Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res 14:1394–1403. https://doi.org/10.1101/gr.2289704.
11. Okonechnikov K, Conesa A, García-Alcalde F. 2016. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics 32:292–294. https://doi.org/10.1093/bioinformatics/btv566.
12. Li H. 2018. minimap2: pairwise alignment for nucleotide sequences. Bioinformatics 34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.
13. Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol 1151:165–188. https://doi.org/10.1007/978-1-4939-0554-6_12.