Draft genome sequence of the Saccharomyces cerevisiae SpyCas9 expressing strain IMX2600, a laboratory and platform strain from the CEN.PK lineage for cell-factory research

van den Broek, Marcel; Ortiz-Merino, Raul A.; Bennis, Nicole X.; Wronska, Anna K.; Hassing, Else Jasmijn; Daran-Lapujade, Pascale; Daran, Jean Marc G.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Draft genome sequence of the *Saccharomyces cerevisiae Spy*Cas9 expressing strain IMX2600, a laboratory and platform strain from the CEN.PK lineage for cell-factory research

Marcel van den Broek,[1] Raul A. Ortiz-Merino,[1] Nicole X. Bennis,[1] Anna K. Wronska,[1] Else-Jasmijn Hassing,[1] Pascale Daran-Lapujade,[1] Jean-Marc G. Daran[1]

AUTHOR AFFILIATION See affiliation list on p. 2.

ABSTRACT The biobased-economy aims to create a circular biotechnology ecosystem to transition from a fossil fuel-based to a sustainable industry based on biomass. For this, new microbial cell-factories are essential. We present the draft genome of the CEN.PK-derived *Saccharomyces cerevisiae Spy*Cas9 expressing strain (IMX2600), that serve as chassis of new cell-factories.

KEYWORDS *Saccharomyces cerevisiae*, CRISPR, biotechnology, cell factory, metabolic engineering

Strains belonging to the CEN.PK family were created as part of a multidisciplinary project funded by the Volkswagen Foundation in 1993–1994 (1). This lineage was chosen as a model for physiology and metabolic engineering research. With advent of CRISPR editing technology (2, 3) and increasing complexity of genetic engineering strategies that involve iteration of transformation rounds, chromosomal integration of *Spycas9* can accelerate strain engineering (4). To this end, the *Saccharomyces cerevisiae* strain IMX2600 (MATa *MAL2-8c SUC2 can1Δ::Spy*cas9-natNT2) was developed from the prototrophic strain CEN.PK113-7D (MATa *MAL2-8c SUC2*) (5). IMX2600 have been used in constructing microbial cell factories (6–8). To ensure accurate sequencing analysis during construction and/or subsequent adaptive laboratory evolution strategy, a reference genome sequence of this platform strain is essential.

A −80°C stocked glycerol vial of the strain IMX2600 was inoculated and grown in 500 mL shake flasks with 100 mL YPD complex medium (10 g L^{-1} Yeast extract, 20 g L^{-1} BactoPeptone, 20 g L^{-1} glucose). Total DNA was extracted using the YeaStar genomic DNA kit from Zymo Research (Zymo Research, Irvine, CA). Shotgun library preparation using Oxford_Nanopore Technologies' SQK-LSK109 kit (ONT, Oxford, UK) was sequenced on a MinION MK1B device with R10 flow cell. Raw FAST5 signal files were base-called using GPU Guppy 4.0.11 in high-accuracy mode. After filtering for quality and length, 974,343 reads were obtained, yielding 8.69 Gb (N50: 14,169 bp), representing ~724× coverage of a *S. cerevisiae* genome (Table 1). Canu version 2.0 with settings genomeSize = 12 m, useGrid = 0 and nanopore-raw (9) was used for *de novo* assembly yielding full-length chromosomes from telomere to telomere. Nucmer from the MUMmer package (Version 3.1) (10) was used to align and trim the 2μ plasmid and the mitochondrial DNA [nucmer (-maxmatch-nosimplify)]. For error correcting the assembled genome sequence, a 300 bp read length TruSeq PCR-Free Illumina library (Illumina, San Diego, CA) with a 550 bp insert-size yielding 2,793,193 paired reads for a total of 1.6 Gigabases (133-fold coverage) was sequenced on a MiSeq (Illumina). Reads quality was assessed with FastQC v0.11.5. The genome was polished by mapping the untrimmed Illumina reads with Burrows-Wheeler Aligner (BWA version 0.7.15-r1142-dirty; default parameters

TABLE 1 IMX2600 assembly data and accession numbers at NCBI repositories (https://www.ncbi.nlm.nih.gov/)

Bioproject	PRJNA976676	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA976676
Biosample	SAMN35437089	https://www.ncbi.nlm.nih.gov/biosample/?term=SAMN35437089
	Assembly size (Mb)	12.26
	GC content (%)	38.26
	N50 (bp)	914,509
Experiment	Oxford nanopore	SRX20564515
	Illumina	SRX20564514
SRA	Oxford nanopore	https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR24792116&display=metadata
	Illumina	https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR24792117&display=metadata
Assembly (annotated)	CP127195– CP127212	https://www.ncbi.nlm.nih.gov/nuccore?term=976676%5BBioProject%5D

used) (11, 12) to the assembly and further processed with SAMtools (version 1.3.1) (13, 14) and polished once by Pilon (version 1.18) with settings fix all (15). This yielded a genome assembly of 12.26 Megabase (N_{50} = 915,509 bp) composed of 16 chromosomes, a 86,554 bases mitochondrial genome and a 6,318 bases 2µ plasmid (Table 1) with a GC content of 38.26%. CHRXII that harbors the rDNA locus was assembled in one contig of 1,099,799 bp long and four smaller contigs ($2x$~8 kb +$1x$ ~55 kb +$1x$ ~100 kb) consisting exclusively of rDNA repeats that were not added to the assembly. The annotation of the polished assembly was performed using Funannotate v1.8.15 (https://github.com/nextgenusfs/funannotate) (16). The "predict" (Gene prediction) step was performed using the command (--species "Saccharomyces cerevisiae" --strain IMX2600—augustus_species saccharomyces_cerevisiae_S288C). Interproscan (version 5.25–64, with "-goterms") was applied in the "predict" step and used as input in the (Functional) "annotate" step. A total of 5,533 coding sequences and 287 tRNA's were identified.

AUTHOR AFFILIATION

[1]Department of Biotechnology, Delft University of Technology, Van der Maasweg, Delft, the Netherlands

AUTHOR ORCIDs

Marcel van den Broek http://orcid.org/0000-0002-7734-582X
Raul A. Ortiz-Merino http://orcid.org/0000-0003-4186-8941
Nicole X. Bennis http://orcid.org/0000-0002-3403-0844
Anna K. Wronska http://orcid.org/0000-0002-6103-3039
Pascale Daran-Lapujade http://orcid.org/0000-0002-4097-7831
Jean-Marc G. Daran http://orcid.org/0000-0003-3136-8193

AUTHOR CONTRIBUTIONS

Marcel van den Broek, Data curation, Formal analysis, Methodology, Software | Raul A. Ortiz-Merino, Formal analysis, Methodology, Project administration, Software | Nicole X. Bennis, Investigation, Validation | Anna K. Wronska, Investigation, Validation | Else-Jasmijn Hassing, Investigation, Validation | Pascale Daran-Lapujade, Conceptualization, Project administration, Writing – original draft | Jean-Marc G. Daran, Conceptualization, Funding acquisition, Investigation, Methodology, Supervision, Writing – original draft

DATA AVAILABILITY

The genome sequencing and assembly of *Saccharomyces cerevisiae* strain IMX2600 have been deposited as Bioproject at NCBI under accession number PRJNA976676 (Table 1).

REFERENCES

1. Entian KD, Kotter P. 2007. Yeast genetic strain and plasmid collections. Meth. Enzymol 36:629–666. https://doi.org/10.1016/S0580-9517(06)36025-4
2. DiCarlo JE, Norville JE, Mali P, Rios X, Aach J, Church GM. 2013. Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res 41:4336–4343. https://doi.org/10.1093/nar/gkt135
3. Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. eLife 2:e00471. https://doi.org/10.7554/eLife.00471
4. Mans R, van Rossum HM, Wijsman M, Backx A, Kuijpers NGA, van den Broek M, Daran-Lapujade P, Pronk JT, van Maris AJA, Daran JM. 2015. CRISPR/Cas9: a molecular swiss army knife for simultaneous introduction of multiple genetic modifications in *Saccharomyces cerevisiae*. FEMS Yeast Res 15:fov004. https://doi.org/10.1093/femsyr/fov004
5. Wronska AK, van den Broek M, Perli T, de Hulster E, Pronk JT, Daran JM. 2021. Engineering oxygen-independent biotin biosynthesis in *Saccharomyces cerevisiae*. Metab Eng 67:88–103. https://doi.org/10.1016/j.ymben.2021.05.006
6. Bouwknegt J, Wiersma SJ, Ortiz-Merino RA, Doornenbal ESR, Buitenhuis P, Giera M, Müller C, Pronk JT. 2021. A squalene-hopene cyclase in *Schizosaccharomyces japonicus* represents a eukaryotic adaptation to sterol-limited anaerobic environments. Proc Natl Acad Sci U S A 118:e2105225118. https://doi.org/10.1073/pnas.2105225118
7. Perli T, Vos AM, Bouwknegt J, Dekker WJC, Wiersma SJ, Mooiman C, Ortiz-Merino RA, Daran JM, Pronk JT. 2021. Identification of oxygen-independent pathways for pyridine nucleotide and coenzyme A synthesis in anaerobic fungi by expression of candidate genes in yeast. mBio 12:e0096721. https://doi.org/10.1128/mBio.00967-21
8. Postma ED, Hassing EJ, Mangkusaputra V, Geelhoed J, de la Torre P, van den Broek M, Mooiman C, Pabst M, Daran JM, Daran-Lapujade P. 2022. Modular, synthetic chromosomes as new tools for large scale engineering of metabolism. Metab Eng 72:1–13. https://doi.org/10.1016/j.ymben.2021.12.013
9. 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
10. 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
11. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.arXiv. Genomics. https://doi.org/10.48550/arXiv.1303.3997
12. 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
13. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of SAMtools and BCFtools. Gigascience 10:giab008. https://doi.org/10.1093/gigascience/giab008
14. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352
15. 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
16. Palmer JM, Stajich J. 2020. Funannotate V1.8.1: eukaryotic genome annotation. Zenodo 16. https://doi.org/10.5281/zenodo.1134477