

# Whole genome sequencing and assembly of *Saccharomyces cerevisiae* genomes using Oxford Nanopore data

Andrew G. Matveenko  
Dpt. of Genetics and Biotechnology, St.  
Petersburg State University, St. Petersburg,  
Russia  
a.matveenko@spbu.ru

Anton B. Matiiv  
Dpt. of Genetics and Biotechnology, St.  
Petersburg State University, St. Petersburg,  
Russia  
antonmatiiv@yandex.ru

Yury A. Barbitoff  
Dpt. of Genetics and Biotechnology, St.  
Petersburg State University, St. Petersburg,  
Russia; Bioinformatics Institute, St.  
Petersburg, Russia  
barbitoff@bk.ru

Evgenia M. Maksiukenko  
Vavilov Institute of General Genetics, St.  
Petersburg Branch; Dpt. of Genetics and  
Biotechnology, St. Petersburg State  
University, St. Petersburg, Russia  
jmrose@yandex.ru

Svetlana E. Moskalenko  
Vavilov Institute of General Genetics, St.  
Petersburg Branch; Dpt. of Genetics and  
Biotechnology, St. Petersburg State  
University, St. Petersburg, Russia  
s.moskalenko@spbu.ru

Alexandra V. Beliavskaia  
University of Liverpool, Liverpool, UK  
alex.beliavskaia@gmail.com

Alexander V. Predeus  
University of Liverpool, Liverpool, UK;  
Bioinformatics Institute, St. Petersburg,  
Russia  
predeus@bioinf.me

Galina A. Zhouravleva  
Dpt. of Genetics and Biotechnology, St.  
Petersburg State University, St. Petersburg,  
Russia  
g.zhouravleva@spbu.ru

**Abstract** — The Peterhof genetic collection (PGC) of yeast *Saccharomyces cerevisiae* is one of the rare examples of a large genetic collection established independently of reference S288C strain. We assembled genomes of two widely used PGC strains, 1A-D1628 and 74-D694, using Oxford Nanopore MinION sequencing data. Subsequent analysis of structural variations showed a number of differences between PGC strains and S288C.

**Keywords** — *Saccharomyces cerevisiae*, NGS, Oxford Nanopore, structural variations

## Motivation and aim

*Saccharomyces cerevisiae* is one of the most thoroughly studied eukaryotic model organism. Hundreds of yeast strains have been described in wild environment, industrial lineages, and laboratories throughout the world. Most of the common laboratory yeast strains, and, specifically, the reference strain S288C ascend to the so-called “Berkeley yeast”. The Peterhof genetic collection (PGC) of *S. cerevisiae* was founded 50 years ago and it is one of the rare examples of a large genetic collection established independently. First PGC strains ascend to a distillery lineage, which is unrelated to the Berkeley yeast. For many years, studies of translation, prion biology, and other fields have benefited from PGC, and several laboratory strains are now widely used throughout the world.

Recently, next-generation sequencing (NGS) was applied extensively to study the diversity of yeast strains. To place the PGC into known interrelations of yeast lineages we previously attempted whole genome sequencing of several strains using Ion Torrent NGS technology [1]. This allowed the identification of genomic variations that caused several phenotypic traits in these strains, e.g., clumping phenotype, phenylalanine auxotrophy, nonsense suppression caused by defective *SUP35* transcription [2], etc. The closest strain to the PGC progenitor, 15V-P4, was shown to differ greatly from other laboratory stocks. Despite the progress achieved, the obtained assemblies of the PGC genomes were incomplete and required substantial improvement.

## Methods

We performed genome sequencing of two strains, 1A-D1628, and 74-D694, using Oxford Nanopore MinION. Using the obtained reads we made a draft assembly that comprises all yeast chromosomes and mtDNA. We compared several long-read assemblers, with Canu showing best results. The obtained assemblies were polished by raw Nanopore signal and, subsequently, by short reads produced by Illumina HiSeq.

## Results

The final assemblies were highly contiguous: 15 and 14 chromosomes (out of 16) were fully assembled for 1A-D1628 and 74-D694, respectively. 97.3% and 97.8% complete BUSCO were discovered in 1A-D1628 and 74-D694, respectively, (compared to 98.0% for reference assembly) demonstrating high reliability of the obtained sequences. We next performed structural variants analysis and discovered several gross rearrangements, as well as multiple intergenic deletions, insertions and duplications in PGC genomes. All structural variations tested have been confirmed by PCR or Sanger sequencing. We also compared the number of *CUP1* gene clusters and copper resistance of the strains. PGC strains were less resistant to copper, and contained less copies of the *CUP1* gene, as shown by qPCR. Comparative analysis and annotation of the whole genome sequences are continuing.

## ACKNOWLEDGMENTS

The work is supported by RCs “MCT” and “Biobank” of SPbSU, by the State research program 0112-2016-0015, by the RFBR grant 20-34-70156, and the RSF grant 18-14-00050.

## REFERENCES

- [1] Drozdova, P.B. et al/ (2016). Genome sequencing and comparative analysis of *Saccharomyces cerevisiae* strains of the Peterhof genetic collection. *PLoS One*, 11(5).
- [2] Matveenko, A.G., Drozdova, P.B., Moskalenko, S.E., Tarasov, O.V., & Zhouravleva, G.A. (2019). Whole genome sequencing data and analyses of the underlying *SUP35* transcriptional regulation for a *Saccharomyces cerevisiae* nonsense suppressor mutant. *Data in brief*, 23, 103694.