Chromosome-level genome assembly using both long-read and short-read sequencing and structural variant analysis of two yeast strains from the Peterhof genetic collection

P-01.1-39

A. Matveenko *[I], Y. Barbitoff *[I,II], A. Matiiv[I], E. Maksiutenko[I,III], P. Drozdova[IV], S. Moskalenko[I,III], D. Polev[V], A. Beliavskaia[VI,VII], L. Danilov[I], A. Predeus[II,VII], G. Zhouravleva[I]

[I]Department of Genetics and Biotechnology, St. Petersburg State University, St. Petersburg, Russia, [II]Bioinformatics Institute, St. Petersburg, Russia, [III]Vavilov Institute of General Genetics, Russian Academy of Sciences, St Petersburg Branch, St. Petersburg, Russia, [IV]Irkutsk State University, Irkutsk, Russia, [V]CerbaLab Ltd., St. Petersburg, Russia, [VI]St. Petersburg State University, St. Petersburg, Russia, [VII]University of Liverpool, Liverpool, United Kingdom

The yeast *Saccharomyces cerevisiae* is a model eukaryotic organism, which has been thoroughly studied for decades. Thousands of yeast strains have been described both on phenotypic and genomic level; however, the majority of investigations utilizes limited number of laboratory strains, which are closely related to the reference S288C strain. The Peterhof genetic collection (PGC) was established independently from S288C, though the lineages have been crossed several times during strain evolution. Several PGC strains are extensively used for studies in the fields of prion biology and translation termination; however, the genomic data is scarce and limited to short-read technologies (Drozdova PB et al. (2016) PLoS ONE, 11, e0154722). We analysed the genomes of two widely-used PGC strains, 74-D694 and U-1A-D1628, using both long-read sequencing with Oxford Nanopore (ONT) and short-read Illumina techniques. Reference-quality assemblies were obtained by constructing draft assemblies from ONT reads using canu, followed by polishing with Nanopore raw signal and Illumina short reads. Hybrid assembly also allowed us to reconstruct sequences of circular molecules, *i.e.*, mitochondrial DNA and 2-micron yeast plasmid. Structural variant (SV) analysis showed multiple mid-length insertions and deletions within coding sequences, *e.g.*, in the *NUP100* and *SCH9* genes. High contiguity of the assemblies allowed us to deduce possible routes of reciprocal unbalanced translocations between chromosomes I, VIII, IX, XI, and XVI of the PGC strains. We also showed that SV-driven formation of hybrid flocculin alleles is likely responsible for the lack of invasive growth of the strains studied (Barbitoff YA et al. (2021) G3, in press). This work was supported by the RSF grant 18-14-00050 "Genetic and epigenetic regulation of translation termination", RFBR grant 20-34-70156, State Research Program 0112-2016-0015, and by RCs “Development of Molecular and Cellular Technologies” and “Biobank” of SPbSU.

* The authors marked with an asterisk equally contributed to the work.