De novo sequencing and high-quality assembly of yeast genomes using a MinION device

Viktoria Hodorova[1], Hana Lichancova[1], Dominik Bujna[3], Martina Nebohacova[1], Lubomir Tomaska[2], Brona Brejova[3], Tomas Vinar[4], and Jozef Nosek[1]

Departments of [1]Genetics and [2]Biochemistry, Faculty of Natural Sciences, Comenius University in Bratislava, Slovakia; Departments of [3]Computer Science and [4]Applied Informatics, Faculty of Mathematics, Physics, and Informatics, Comenius University in Bratislava, Slovakia

Yeasts are excellent model organisms for investigation of a wide repertoire of biological phenomena. Relatively small genomes and high degree of genetic as well as phenotypic heterogeneity resulting from diverse evolutionary trajectories make these unicellular fungi perfectly suited for comparative genome analyses. Although the number of yeast genome sequences available in public databases is constantly increasing, only a few genomes have been completed into compact high-quality assemblies consisting of chromosome-sized contigs. In this work, we used several non-conventional yeast species as examples to demonstrate that a single run on a MinION device can generate the sequence data in gigabase amounts in less than 24 hours. Resulting data are suitable for high-quality assemblies of nearly complete genomes with the sequencing depth highly exceeding 100X.

Experimental protocol:

DNA isolation:

- yeast cultures were grown O/N in 100 mL of complex medium (YPD)
- cells were converted to sphaeroplasts using zymolyase treatment and lysed in SDS/TE buffer
- nucleic acids were extracted by phenol / chloroform / isoamyl alcohol, precipitated with ethanol, and resuspended in TE buffer
- RNA was removed using RNase A
- DNA was purified on a Genomic 100/G column (Qiagen), the eluate was then precipitated with ethanol and resuspended in TE buffer; typical yield was 10-30 µg

Sample QC (example of *J. angkorensis*):

- **Nanodrop conc.:** 293 ng/µl
  A260/280 = 1.86
  A260/230 = 2.12
- **Qubit conc.:** 144 ng/µl

DNA sequencing:

- 2-3 µg of genomic DNA (based on Qubit)
- sequencing libraries were prepared using SQK-LSK108 kit; each library was loaded into a FLO-MIN106 (R9.4 or R9.4.1) flow cell in a MinION Mk-1B device

Data analysis:

- **Basecalling:** Albacore 2.0.2
- **Assembly:** minimap / miniasm (Li, 2016)
- **Assembly statistics:** Quast (Gurevich *et al.* 2013)

Read lengths:

Read qualities:

Sequence data:

	Raw data [Mbp]	No. of reads	Mean length [bp]	Median lenght [bp]	Max length [bp]
Jaminaea angkorensis	6290	556420	11305.5	8083	118582
Magnusiomyces magnusii	4607	742765	6203.1	2881	186804
Magnusiomyces tetrasperma	7230	1455622	4967.0	1495	184269
Saprochaete clavata	5390	481871	11187.4	4803	198579
Yarrowia lipolytica	2920	642717	4551.5	3331	90529
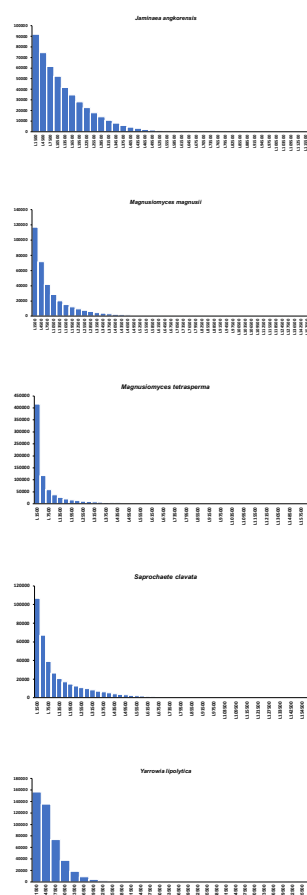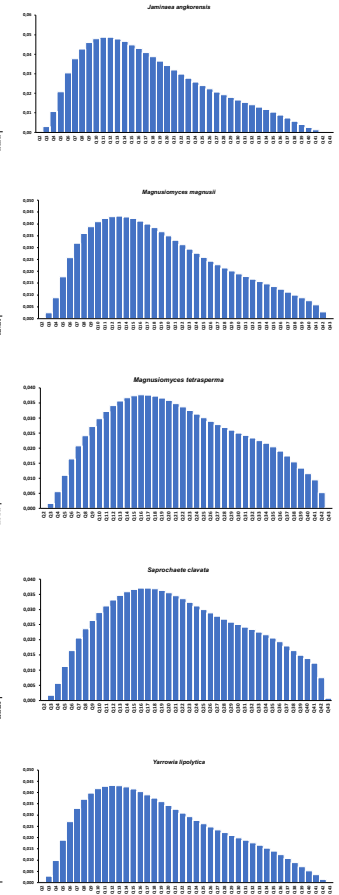
Genome assemblies:

	Genome size [Mbp]	No. of contigs	Longest contig [Mbp]	N50 [Mbp]	Genome coverage
Jaminaea angkorensis	20.1	37	2.8	0.84	299
Magnusiomyces magnusii	22.9	34	3.67	1.8	150
Magnusiomyces tetrasperma	29.7	68	3.28	1.12	192
Saprochaete clavata	23.4	75	2.15	0.55	217
Yarrowia lipolytica	20.3	19	4.46	3.42	116

Read lengths:
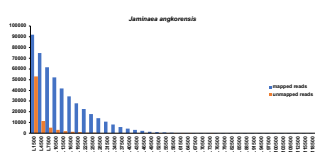
Read qualities:
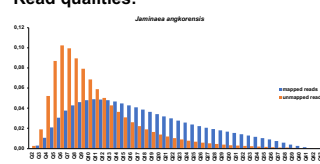
References:

1. Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G. (2013) QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29(8): 1072-1075.
2. Li, H. (2016) Minimap and miniasm: fast mapping and *de novo* assembly for noisy long sequences. *Bioinformatics* 32(14): 2103-2110.