High-quality genome assemblies of male and female *Populus* x *sibirica* plants

Pushkova E.N.[1*], Krasnov G.S.[1], Dvorianinova E.M.[1, 2], Novakovskiy R.O.[1], Povkhova L.V.[1, 2], Melnikova N.V.[1], Dmitriev A.A.[1]
[1] Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia;
[2] Moscow Institute of Physics and Technology, Dolgoprudny, Russia * e-mail: pushkova18@gmail.com

The genus *Populus* is represented by dioecious species with XY or ZW sex-determination system and is characterized by polymorphic mechanisms of sex determination

The *Populus* mechanisms of sex determination are actively studied and recent works showed its genetic nature, however, there is still no consensus on this issue

Populus × *sibirica* is one of the most common poplars in the Moscow region, but to date, there was no genome assembly for it

This work was aimed to obtain high-quality genome assemblies of male and female trees of *Populus* × *sibirica* using a combination of the Oxford Nanopore platform with long reads and Illumina with high-accuracy reads for further comparative analysis

A female poplar plant is on the left and male one is on the right

Materials and Methods

Plant material

Young leaves from male and female trees of *Populus* x *sibirica*

Extraction and purification of high-molecular-weight genomic DNA

Cell lysis:
• Buffer: 100 mM Tris-HCl pH 9.5; 2% CTAB; 1.4 M NaCl; 1% PEG 8000; 20 mM EDTA with 12 µl β-mercaptoethanol and 0.04 g PVP K30
• Purification: chloroform

DNA precipitation:
• Buffer: 1% CTAB, 50 mM Tris-HCl pH 8.0, 10 mM EDTA

DNA purification and elution:
• Set Blood & Cell Culture DNA Mini Kit (Qiagen, Germany)

Elimination of short DNA fragments (up to 10 kb):
• Short Read Eliminator Kit (Circulomics, USA)

DNA purification on beads:
• AMPure XP beads (Beckman Coulter, USA) in the ratio 1: 0.7 (sample:beads)

Genome sequencing on the Illumina platform

HiSeq 2500 (Illumina, USA) with a read length of 250+250 bp

Genome sequencing on the Oxford Nanopore platform

MinION (Oxford Nanopore Technologies, UK) with a FLO-MIN-106 R9.4 and FLO-MIN-110 R10 flow-cells (Oxford Nanopore Technologies)

Bioinformatics analysis

Basecalling of Nanopore reads
- guppy 3.2.2, flip-flop algorithm

Preparation of reads
- Trimmomatic
- Porechop

Initial genome assembly
- Flye
- Shasta
- Raven
- Wtdbg2

Assembly polishing algorithms
- Racon (Nanopore reads)
- POLCA (Illumina reads)

Assembly completeness assessment
- BUSCO

Assembly statistics evaluation
- QUAST

Results

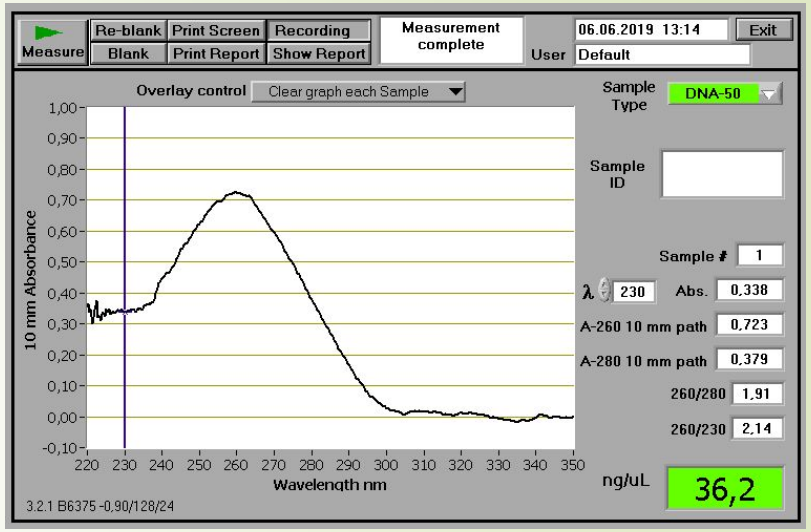
We developed and optimized the method of obtaining the pure high-molecular-weight genomic DNA from young poplar leaves needed for Nanopore sequencing. The extracted according to this protocol DNA showed A260/280 values equal to 1.7-2.0 and A260/230 values equal to 2.0-2.2

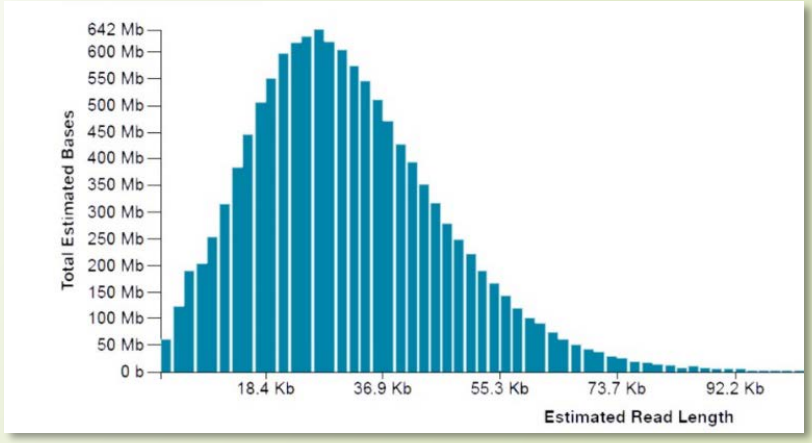
The obtained proximity of concentration values measured using a Qubit fluorometer and a Nanodrop spectrophotometer served as an additional criterion of DNA purity. Short Read Eliminator Kit (Circulomics) enabled the removal of short DNA fragments, and the average length of DNA was about 50 kb

For each *Populus* x *sibirica* tree, we received about 22 Gb by Nanopore reads with N50 equal to 20 kb and 36-52 million paired-end 2x125 bp Illumina reads

We compared several genome assemblers. The best result was obtained for the Raven assembler: genome length was 474 Mb with N50 of 0.33 Mb for the male poplar and 495 Mb with N50 of 0.32 Mb for the female one. According to BUSCO, the genome completeness was 96.8% for the male and 97.9% for the female.

The comparative analysis of high-quality genomes of male and female *Populus* x *sibirica* plants provides the opportunity to determine the genes associated with poplar sex and clarify the sex determination mechanism of this species





Acknowledgments: This work was funded by RFBR according to the research project 17-29-08036