

# Extraction of high-molecular-weight DNA from poplar plants for Nanopore sequencing

E.N. Pushkova\*, A.D. Beniaminov, G.S. Krasnov, R.O. Novakovskiy, L.V. Povkhova, N.V. Melnikova, A.A. Dmitriev

Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia

DOI 10.18699/ICG-PlantGen2019-51

© Autors, 2019

\* e-mail: pushkova18@gmail.com

**Abstract:** Poplar (*Populus trichocarpa*) is a model object for genetic studies in trees. The extraction of pure high-molecular-weight DNA from plant cells is difficult because of a high level of storage substances that adversely affect Nanopore sequencing quality. We developed and optimized a methodology for isolation of pure genomic DNA from poplar plants using modified CTAB buffer, QIAGEN Genomic-tip 20/G, and AMPure XP beads. The developed method allowed us to isolate pure DNA samples with the  $A_{260/280}$  ratio in the range of 1.8–1.9, the  $A_{260/230}$  ratio in the range of 2.2–2.6, and a length of about 50 kb. The obtained DNA can be used for sequencing on a Nanopore platform as well as for a wide range of molecular studies.

**Key words:** *Populus*; Nanopore sequencing; high-molecular-weight DNA.

## 1. Introduction

Genomic studies can be used to address problems both in applied and basic research. Plant genomes compared to the human genome and other animals' genomes are much more complex, which makes their analysis more problematic. Plants have the largest genomes compared to other living organisms – from 61 Mb (*Genlisea tuberosa*) (Fleischmann et al., 2014) to 152 Gb (*Paris japonica*) (Pellicer et al., 2010) – which is almost 50 times the size of the human genome. Moreover, plant genomes exhibit a wide range of ploidy (the presence of more than two genomes (genome duplication) per cell), contain regions with a high GC content and regions with transposable elements. These structural features of plant genomes represent a severe problem of a whole genome assembly using short reads produced by sequencing platforms generating reads with an average read length of 300–1000 bp. However, improvements in sequencing technology with the Nanopore platform have made it possible to obtain reads with a length of about 100 thousand nucleotides (van Dijk et al., 2014). This allows one not only to simplify and accelerate the assembly of previously unpublished genomes, but also to improve the already existing assemblies. Nanopore sequencing technology requires high-quality DNA. However, extracting pure high-molecular-weight DNA from plant cells is difficult because of a high level of storage substances, such as polysaccharides and secondary metabolites, for example, alkaloids, phenolics, terpenes, which not only interfere with DNA isolation, but also adversely affect its quality, as well as Nanopore sequencing quality. Moreover, these metabolites vary considerably, which makes it impossible to use a universal DNA isolation protocol. Our study was aimed to optimize the isolation protocol of pure high-molecular-weight genomic DNA from poplar plants for Nanopore sequencing. Poplar (*Populus trichocarpa*) is a model object for genetic studies in trees (Ellis et al., 2010). That was the first tree, whose genome was sequenced entirely in 2006 (Tuskan et al., 2006). At present, there is a reference genome assembly for *P. trichocarpa*; however, this assembly continues to be refined and changed. The complete genome

assemblies of poplar species will help geneticists and breeders in developing tree cultivars with desired properties.

## 2. Materials and methods

### 2.1 Plant material

The material was obtained in the flowering period from the leaves of *Populus × sibirica*, the most common poplar in Moscow. We chose young, partially expanded leaves, which provide the highest quantity of DNA per unit of tissue volume.

### 2.2 DNA extraction

The leaf material was lysed in a modified Carlson buffer taken from the Nanopore protocol (100 mM Tris-HCl, pH 9.5; 2 % CTAB; 1.4 M NaCl; 1 % PEG 8000; 20 mM EDTA) with the addition of 0.04 M  $\beta$ -mercaptoethanol (BioRad, USA), 1 % PVP (BioChemica PanReac AppliChem, Germany). The homogenate was incubated in the Dwarf thermostat (DNA Technology, Russia).

For DNA purification, chloroform (Acros Organics, USA) was added to the homogenate and centrifuged in IEC Micro-max (Thermo Fisher Scientific, USA); and the aqueous phase was collected. For further purification from polysaccharides, 5× CTAB buffer (5 % CTAB, 350 mM EDTA) was added to the aqueous phase. For the precipitation of DNA, we used a precipitation buffer (1 % CTAB, 50 mM Tris-HCl, pH 8.0, 10 mM EDTA). Purification of the obtained DNA was performed using QIAGEN Genomic-tip 20/G (Germany). DNA was dissolved in TE buffer (Evrogen, Russia) and further purified using AMPure XP beads (Beckman Coulter, UK). The purity and quality of the DNA were assessed using a NanoDrop 2000C spectrophotometer (Thermo Fisher Scientific, USA). The electrophoretic analysis was performed in an SE.1 electrophoretic tank (Helicon, Russia) with the Elf-4 power source (DNA Technology, Russia) using a 0.8 % agarose (Helicon, Russia). The DNA concentration was measured on the Qubit 2.0 fluorometer (Invitrogen, USA) using a dsDNA high-sensitivity (HS) assay kit (Invitrogen, USA).

### 3. Results and discussion

#### 3.1 Extraction of high-molecular-weight DNA from poplars

The first step to obtain a high-quality pure DNA is to destroy the cell wall, which we performed by physical tissue destruction – grinding with a pestle in liquid nitrogen, which enables cell destruction thus inhibiting any biochemical and physical processes that could damage the DNA. We poured liquid nitrogen into the mortar and 0.2 g of leaf material was pounded with a pestle to a flour-like consistency. We concluded that plant material used for DNA extraction should be less than 200–300 mg in weight; otherwise, we would fail to avoid the adverse effects of polyphenols, polysaccharides, and others. Next, we chemically acted on the plant cells using one of the most widely used methods based on the CTAB buffer (Murray and Thompson, 1980). However, the standard method did not meet the requirements for the quality and purity of the DNA, so we modified and changed the composition of the Carlson buffer in order to obtain high-quality DNA. Cell lysis was performed in 5 ml of Carlson buffer with 12.5  $\mu$ l  $\beta$ -mercaptoethanol and 1 % PVP. The homogenate was incubated at 65 °C for 1 hour mixing the sample by inversion halfway through. The cationic surfactant, cetyltrimethylammonium bromide (CTAB) included in the buffer lyses the cell wall. Mercaptoethanol destroys disulfide bridges, including those in proteins, breaking down their tertiary and quaternary structures, and also acts as a natural antioxidant inhibiting the oxidation processes that directly or indirectly damage the DNA. The chaotropic agents present in the extraction buffer, such as salts, denature macromolecules breaking hydrogen bonds, hydrophobic interactions, and van der Waals forces. High salt concentration, for example, precipitates polysaccharides, which otherwise may form a jelly complex with DNA. The presence of EDTA, a chelating agent binding metal ions ( $Mg^{2+}$ ,  $Ca^{2+}$ ,  $Fe^{3+}$ , etc.) enables deactivation of metal-dependent enzymes located in plant extracts, including nucleases that destroy DNA. Polyvinylpyrrolidone (PVP), which is also present in the extraction buffer, binds with phenols preventing the formation of their complexes with DNA. We continued our DNA extraction with chloroform, which helps remove proteins, polysaccharides, and phenol residues. After incubation, we let the samples cool to room temperature and added 1 volume of chloroform. The sample was vortexed for two pulses of 5 s each and centrifuged at 5500 g for 10 min at room temperature. Then, we carefully transferred the top (aqueous) phase to a new tube without disturbing the interphase. However, even after modifying the stage of chemical cell lysis, we were unable to get rid of polysaccharides completely. To remove polysaccharide residues during the extraction step with chloroform, we collected the aqueous phase and added 0.2 volumes of 5 $\times$  CTAB buffer and incubated it at 65 °C for 10 minutes. Then, we repeated the chloroform extraction. If the solution becomes distinctively cloudy, indicating the presence of polysaccharides, we recommend to repeat the chloroform extraction with the addition of 0.2 volumes of 5 $\times$  CTAB buffer. Next, we added 2 volumes of precipitation buffer to the top phase and mixed it thoroughly by inverting the tube 10 times. Then, the samples were incubated for 1 hour at room temperature and centrifuged at 10000 g for 15 minutes

at room temperature. We carefully discarded the supernatant without disturbing the pellet.

At this stage, the amount of isolated DNA was sufficient, approximately 5  $\mu$ g; however, the DNA purity and the presence of proteins and RNA in the sample did not meet the requirements of sequencing on the Nanopore platform. We decided to carry out the post-treatment stage on the QIAGEN DNeasy Plant Mini Kit columns. Unfortunately, after testing this kit, we were faced with large losses of material and low-quality DNA. The length of high-molecular DNA after the columns was about 40 kb, which was not that we wanted. We decided to try cleaning on QIAGEN Genomic-tip 20/G columns.

#### 3.2 Purification of high-molecular-weight DNA from poplar plants with QIAGEN Genomic-tip 20/G

The precipitate of the high-molecular-weight DNA obtained at the previous step was dissolved in 2 ml of G lysis buffer supplemented with 4  $\mu$ l of RNase A (100 mg/ml), which contributed to RNA hydrolysis. RNA removal is a crucial stage in DNA isolation for sequencing with the Nanopore platform, since hairpin RNA structures significantly reduce the quality of sequencing and clog the protein pores in the flow cell. We incubated the sample at 37 °C for 30 minutes and then added 100  $\mu$ l of alkaline protease K and incubated the sample at 50 °C for 30 minutes. Protease hydrolyzes proteins, including histones associated with the DNA, and effectively inactivates nucleases. DNA samples were then deposited to a column using the kit protocol. The method is based on ion-exchange chromatography and is implemented by binding a positively charged resin in a column with negatively charged DNA. Elution from the matrix is carried out by decreasing salt concentration. Proteins and RNA at low salt concentrations are easily separated from the matrix, while DNA is more tightly bound to the support and is eluted at lower concentrations. After appropriate purification from RNA and proteins, the parameter  $A_{260/280}$  of dissolved DNA samples was about 1.8, and the concentration was about 400 ng/ $\mu$ l. However, the concentration of the DNA assessed by Qubit 2.0 was about 250 ng/ $\mu$ l. This discrepancy in values may indicate remaining contamination in the sample. We decided to add an additional stage of purification on magnetic beads.

#### 3.3 DNA purification after columns with AMPure XP beads

The DNA sample obtained was diluted to a concentration of 60 ng/ $\mu$ l in a final volume of 50  $\mu$ l with TE buffer of pH 8. Then, DNA was purified again using AMPure XP beads in a ratio of 1:0.7. We chose 1:0.7, for its provides the best selection for DNA fragments > 2 kb. We incubated the sample for 10 minutes at room temperature, placed the tube on a magnetic stand, and discarded the supernatant. Keeping the tubes on the magnetic stand, we washed beads 2 times with 200  $\mu$ l of freshly prepared 70 % ethanol without disturbing the pellet. Then, we air dried the sample for 30 seconds, removed the tubes from the magnetic stand and resuspended beads in 40  $\mu$ l of TE buffer. The sample was incubated for 3 min at 50 °C and then for 5 min at room temperature. At this stage, we increased the incubation time at 50 °C from 1 minute to 3 minutes. This contributed to a better DNA yield. The principle

of this method is based on a reversible binding of DNA on the surface of magnetic particles, which are easily precipitated from the suspension using a magnetic stand. Thus, during the cleaning with beads, we remove various impurities, such as salts etc.

### 3.4 Quality control of isolated genomic DNA

The quantity and quality of the isolated DNA were evaluated by the  $A_{260/280}$  and  $A_{260/230}$  ratios used to assess the amount of protein and metabolite contamination left after the nucleic acid isolation process. The recommended  $A_{260/280}$  and  $A_{260/230}$  ratios before Nanopore sequencing are 1.80 and 2.0–2.2, respectively. Before measuring the absorbance of samples at 280, 260, and 230 nm, it is necessary to calibrate the spectrophotometer using a solution (not containing DNA) in which the DNA is resuspended. Also, we do not recommend using samples with concentrations below 50 ng/μl, as it may lead to incorrect values. The  $A_{260/280}$  ratio for the poplar high-molecular-weight genomic DNA samples was about 1.8, which corresponds to the recommended value. Its reduction indicates protein and phenol contamination, since these compounds have an absorption maximum at 280 nm. The increased  $A_{260/280}$  ratio indicates the contamination of the DNA sample with RNA molecules. The  $A_{260/230}$  ratio for the DNA samples was about 2.4. Absorption at a wavelength of 230 nm can be caused by contamination with phenolates, thiocyanates, and other organic compounds. The concentration of the DNA samples was also assessed during the evaluation of the  $A_{260/280}$  and  $A_{260/230}$  ratios using a spectrophotometer. Next, we measured the DNA concentration using a fluorometer, which confirmed the concentration value obtained by the spectrophotometer. This suggests that the obtained high-molecular-weight DNA is of good quality and purity. This is an important step of DNA purity control because if the values from the two instruments were to diverge by more than 2-fold, it would indicate the presence of various contaminants (mentioned above). Therefore, we would recommend adding a purification step. To further assess the quality and size of the isolated DNA, an electrophoretic analysis was performed in a 0.8 % agarose gel. As a control, pure high-molecular-weight DNA from phage λ at different concentrations was used. The electropherogram showed a clear band whose length corresponded to the length of the control DNA, approximately 50 kb, indicating that the isolated high-molecular-weight DNA was not degraded. In addition, we compared the isolated DNA with phage λ DNA at varying concentrations regarding luminescence intensity determining the amount of the DNA in an aliquot. As a result

of the work, we would like to especially emphasize the importance of the sample preparation stage for sequencing on the Nanopore platform. The key to success lies in optimizing the DNA isolation and purification protocol for each object individually, as well as a correct and accurate assessment of the purity, amount, and size of DNA. Using the DNA obtained with the protocol developed, we obtained 9 Gb of sequencing data per run of Nanopore MinION with R9.4.1 flow cell.

## 4. Conclusion

During this study, we developed and optimized a methodology of isolation of pure high-molecular-weight genomic DNA from poplar plants that can be sequenced using the Nanopore platform. We selected conditions for cell lysis and DNA purification using QIAGEN Genomic-tip 20/G and AMPure XP beads, which allowed us to dispose of main plant storage substances such as polysaccharides and secondary metabolites, for instance, alkaloids, phenolics, terpenes. The developed method of isolation and purification of DNA allowed us to isolate pure DNA samples with the  $A_{260/280}$  ratio in the range of 1.8–1.9, the  $A_{260/230}$  ratio in the range of 2.2–2.6, a concentration of about 300 ng/μl, and a length of about 50 kbp. The obtained pure DNA can be used for sequencing on a Nanopore platform as well as for a wide range of molecular research.

## References

- Ellis B., Jansson S., Strauss S.H. et al. Why and How Populus Became a “Model Tree”. New York, NY: Springer New York, 2010.
- Fleischmann A., Michael T.P., Rivadavia F. et al. Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. *Ann Bot.* 2014;114(8):1651–1663. DOI 10.1093/aob/mcu189.
- Murray M.G., Thompson W.F. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 1980;8(19):4321–4325. DOI 10.1093/nar/8.19.4321.
- Pellicer J., Fay M.F., Leitch I.J. The largest eukaryotic genome of them all? *Botanical J. Linnean Society.* 2010;164(1):10–15. DOI 10.1111/j.1095-8339.2010.01072.x.
- Tuskan G.A., Difazio S., Jansson S. et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Sci.* 2006;313(5793):1596–1604. DOI 10.1126/science.1128691.
- van Dijk E.L., Auger H., Jaszczyszyn Y., et al. Ten years of next-generation sequencing technology. *Trends Genet.* 2014;30(9):418–426. DOI 10.1016/j.tig.2014.07.001.

**Acknowledgements.** The reported study was funded by RFBR according to the research project 18-34-20113 mol\_a\_ved.

**Conflict of interest.** The authors declare no conflict of interest.