DOI 10.18699/SBB-2022-007

Sequencing *Linum usitatissimum* L. genomic DNA extracted from nuclei

Dvorianinova E.M.[1, 2]*, Bolsheva N.L.[1], Rozhmina T.A.[1, 3], Sigova E.A.[1, 2], Pushkova E.N.[1], Melnikova N.V.[1], Dmitriev A.A.[1]

[1] *Engelhardt Institute of Molecular Biology, RAS, Moscow, Russia*

[2] *Moscow Institute of Physics and Technology, Moscow, Russia*

[3] *Federal Research Center for Bast Fiber Crops, Torzhok, Russia*

** dvorianinova.em@phystech.edu*

Key words: flax, high-molecular-weight DNA, whole genome sequencing, Nanopore

Motivation and Aim: The genetics of agricultural plants has shown the importance of assembling accurate and contiguous genomes [1]. These sequences form the basis for studying crucial agricultural traits at the molecular level. However, poor assembly quality undermines the validity of further genetic analysis [2]. Thus, obtaining high-quality plant genomes with available techniques is a relevant challenge. The current work aimed at producing a high-quality genome sequence of a flax (*Linum usitatissimum* L.) variety using the Oxford Nanopore Technologies (ONT) platform, which requires extremely pure DNA.

Methods and Algorithms: We adapted a plant nuclei isolation protocol for flax. The organelles were obtained from the *L. usitatissimum* line #3896 (provided by the Institute for Flax) leaves. Pure high-molecular-weight DNA was extracted from the nuclei with the Nanobind Plant Nuclei Big DNA Kit (Circulomics, USA). The SQK-LSK109 kit was used to prepare 1D genomic DNA libraries for sequencing on the ONT platform (MinION, R9.4.1 flow-cells). The received fast5 reads were basecalled with Guppy 5.0.11 (dna_r9.4.1_450bps_sup.cfg); the adapters were removed with Porechop. *L. usitatissimum* line #3896 genome was assembled with Canu 2.2 (genomeSize = 400m) and polished with Racon (two iterations) and Medaka [3–5]. Genome statistics were evaluated using QUAST and BUSCO (eudicots_odb10 dataset) [6, 7]. Cytosine methylation levels and methylation frequencies were quantified with Megalodon and the megalodon.sh script from METEORE respectively [8, 9].

Results: The developed nuclei isolation protocol allowed us to extract pure high-molecular-weight DNA and perform two successful sequencing runs on the ONT platform. We received 14.5 and 16.5 Gb of raw fast5-reads with an N50 of 12.4 and 15.7 kb respectively. Having merged the two datasets, we assembled and polished the genome sequence of *L. usitatissimum* line #3896. The assembly demonstrated 93.8 % BUSCO completeness and consisted of 1697 contigs with a total length of 447.0 Mb and an N50 of 6.2 Mb. In the assembled genome, cytosine nucleotides were modified in the CG, CHG, and CHH contexts (Table 1). Nearly a half of the CG cytosines were highly methylated (≥ 50 %). In contrast, only 0.02 % of the CHH cytosines had high methylation frequency.

Table 1. Cytosine methylation in the genome assembly of *L. usitatissimum* line #3896

Methylation context	CG	CHG	CHH
Context abundance, % of the called CN sites	17.0	11.9	71.1
Percentage of sites with high methylation levels (≥ 50 %)	53.8	2.8	0.02

Conclusion: The employed method of flax nuclei isolation enabled us to extract pure high-molecular-weight DNA of *L. usitatissimum*. The received amount of ONT data was adequate to produce an assembly with acceptable QUAST and BUSCO statistics. Our dataset can be used in molecular genetic research on flax, including such genome regulation studies as methylation analysis.

Acknowledgements: The work was financially supported by the Ministry of Science and Higher Education of the Russian Federation, grant No. 075-15-2021-1064.

References

1. Li G. et al. A high-quality genome assembly highlights rye genomic characteristics and agronomically important genes. *Nat Genet*. 2021;53(4):574-584.
2. Van Bel M., Bucchini F., Vandepoele K. Gene space completeness in complex plant genomes. *Curr Opin Plant Biol*. 2019;48:9-17.
3. Koren S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722-736.
4. Vaser R. et al. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res*. 2017;27(5):737-746.
5. https://github.com/nanoporetech/medaka
6. Gurevich A. et al. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075.
7. Simão F.A. et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210-3212.
8. Yuen Z.W.-S. et al. Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nat Commun*. 2021;12(1):1-12.
9. https://github.com/nanoporetech/megalodon