Comparative analysis of flax (*Linum usitatissimum* L.) genomes and transcriptomes

Elena Pushkova[I,*], George Krasnov[I], Roman Novakovskiy[I], Liubov Povkhova[I,II], Artemy Beniaminov[I], Nadezhda Bolsheva[I], Tatiana Rozhmina[I,III], Alexey Dmitriev[I] , Nataliya Melnikova[I]

[I]*Engelhardt Institute of Molecular Biology, Russian Academy of Sciences, Moscow, Russia*
[II]*Moscow Institute of Physics and Technology, Dolgoprudny, Russia*
[III]*Federal Research Center for Bast Fiber Crops, Torzhok, Russia*
**e-mail: pushkova18@gmail.com*

Flax (*Linum usitatissimum* L.) is widely used to produce fiber and seeds

Flax seeds are the richest source of omega-3 fatty acids, which reduce the risk of cancer and cardiovascular diseases, and lignans, which have antibacterial, antifungicide, and antioxidant activities

Flax fiber is used in the manufacturing of high-quality textiles, medical products, and composite materials

The aim of the present work was to obtain high-quality flax genomes and transcriptomes for genetically diverse cultivars and lines, which have breeding value and different direction of use, for further comparative analysis

Materials and Methods

Plant material

Seeds of 6 flax cultivars/lines that differ in economically significant traits and are valuable for breeding (LM98, #3896, Diplomat, Atlant, Universal, Alizee) were obtained from the Institute for Flax (Torzhok, Russia). Leaves of 2-week-old plants were used for genome sequencing, while roots and shoots of 7-day-old seedlings and leaves, flowers, and stems of 6-week-old plants – for transcriptome sequencing

RNA extraction

Quick-RNA Miniprep Kit (Zymo Research, USA)

High-molecular-weight genomic DNA extraction and purification

Cell lysis:
- DNA-EXTRAN-3 kit (Synthol)

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, USA)

Transcriptome sequencing on the Illumina platform

NextSeq (Illumina, USA) with a read length of 86 bp

Genome sequencing on the Illumina platform

HiSeq 2500 (Illumina) 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 flow-cell (Oxford Nanopore Technologies)

Bioinformatics analysis

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

Preparation of reads
- trimmomatic

Initial genome assembly
- Shasta
- Flye
- Wtdbg2

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

Assembly completeness assessment
- BUSCO

Assembly statistics evaluation
- QUAST

Mapping of transcriptome reads to the genome
- STAR

Generating of plots
- MultiQC

Evaluation of gene expression
- BEDTools

Genome annotation
- funannotate

Results

We developed and optimized the method of obtaining the pure high-molecular-weight genomic DNA from flax leaves

The extracted DNA was about 50 kb long, A260/280 values were in the range 1.75-1.87, and A260/230 values – 2.0-2.7

From 20 to 25 million paired-end 2×250 bp reads were obtained on the Illumina platform for each of 6 flax cultivars/lines that, after trimming, corresponded to 20-25-fold coverage of the flax genome

From 6 to 10 Gb were obtained on the Oxford Nanopore platform that corresponded to 15-25-fold genome coverage

We compared several genome assemblers, and Flye provided the best results: N50 values ranged from 200 kb to 1 Mb depending on the genotype, while the completeness of the assemblies was over 90% according to BUSCO

From 6 to 16 million 86 bp reads were received as a result of transcriptome sequencing for 5 tissues of 6 flax cultivars/lines. After bioinformatics analysis, for the majority of samples, more than 87% of reads were uniquely mapped to the *L. usitatissimum* genome, and about 8-10% of reads were mapped to several loci

The obtained genome assemblies are the basis for molecular genetic studies in flax, allows the assessment of the differences in *L. usitatissimum* cultivars/lines at the genome-wide level, and lay the foundation for the development and introduction of marker-assisted and genomic selection of flax and its genome editing. The obtained transcriptome sequencing data enables determining genes that control the key processes occurring in different flax tissues, revealing genes with expression differences between flax genotypes with diverse characteristics, and establishing associations between gene expression and phenotype

ACKNOWLEDGMENTS: This work was financially supported by the Russian Science Foundation, grant 16-16-00114