DOI: 10.25205/978-5-4437-1843-9-5

COMPARATIVE ANALYSIS OF CHROMOSOME-LEVEL FLAX GENOME ASSEMBLIES FROM R9 NANOPORE DATA

E. M. Dvorianinova[1], A. A. Arkhipov[1], N. L. Bolsheva[1], E. N. Pushkova[1], T. A. Rozhmina[2], E. V. Borkhert[1], E. A. Sigova[1], D. A. Zhernova[1], N. V. Melnikova[1], A. A. Dmitriev[1]

[1] *Engelhardt Institute of Molecular Biology RAS, Moscow*
[2] *Federal Research Center for Bast Fiber Crops, Torzhok*

✉ dvorianinova.em@phystech.edu

Abstract

Flax is a perspective agriculture of a vast application. In this work, we assembled the genomes of six genetically distinct flax varieties valuable for breeding from R9 Nanopore reads. The constructed assemblies had a chromosome level and possessed a quite similar structure. However, a range of chromosomal rearrangements were observed, including long inversions and deletions.

Flax (*Linum usitatissimum* L.) is a multipurpose crop cultivated to obtain oil and fiber. This produce is demanded by numerous industries, e.g., manufacturing functional nutrition, bioactive supplements, paints, medical products, fibers, and composite materials. Therefore, each industry requires growing flax varieties with a target complex of characteristics tailored for their main use. Modern breeding methods can effectively solve this task. The techniques employed are based on the knowledge of the key plant characteristics organization at the genome level. Meanwhile, interspecific genetic heterogeneity is common for plants. To receive complete information on genetic determinants of the valuable plant traits, the power of a comparative genomic analysis can be harnessed. This study aimed to obtain and compare chromosome-level genome assemblies of genetically diverse flax varieties of high breeding value.

The genomes of six varieties (Alizee, K-470, K-1570, K-1230, K-432, and K-5543) were sequenced on the Oxford Nanopore Technologies (ONT) platform on the PromethION instrument (R9.4.1 flow-cells). The received genomic data was basecalled with Dorado 0.9.6 and filtered by the quality threshold of 15 or 16 and the length of 10-15 kb. The assembly was performed using Hifiasm 0.25.0. The resulting genome sequences were analyzed with QUAST 5.0.2, BUSCO 5.8.3, and tidk 0.2.65. The final assemblies were aligned to the genome of the K-470 variety using LAST 2.39.3.

The updates in sequencing technologies and data processing instruments allowed us to assemble flax genomes of high contiguity and completeness even from R9 Nanopore data. For the six sequenced varieties, we received an average volume of raw reads of 131 Gb with an average N50 of 25 kb. The obtained genome assemblies had a BUSCO completeness of 95.4 % and a size of 477.8–492.9 Mb. The L50 parameters of these assemblies were close to half of a haploid set of flax chromosomes (2n = 30) with the median of 8. These assemblies had an L90 of 17–26 and an N50 of 17.1–29.7 Mb. The obtained genomes were checked for the presence of a telomere repeat across the contigs. The majority of the analyzed contigs had telomeres at least at one end. The assemblies of Alizee, K-470, and K-432 possessed the highest number of contigs (9) ending and beginning with telomeres. The conducted analyses demonstrated that the constructed genomes had a near-chromosome level.

The alignment of the obtained genomes of different flax genotypes demonstrated high similarity between them. Nevertheless, a range of structural variations was present in flax chromosomes — mainly long deletions and inversions. For instance, the K-1230 genome had a 4-Mb deletion in chromosome 8, two deletions of 2- and 3-Mb lengths in chromosome 10, a 2-Mb deletion in chromosome 6, and a 3-Mb inversion in chromosome 8. In the assembly of the variety K-1570, we observed a 3.5-Mb deletion followed by a 4.5-Mb inversion in chromosome 10. The identified structural variations can have an impact on genome regulation of the studied varieties. For instance, the deleted sequences in the genome of K-1230 were similar to microsatellites, which can play an important role in gene expression.

The results of the study contribute to the understanding of the genomic diversity of flax. The obtained genomes are a multifunctional bioinformatics instrument, which can simplify the marker-assisted and genomic selection of flax varieties with a target complex of traits.