

Structural variation in flax *Linum usitatissimum* L. genome

Duk M.A.^{1,2*}, Kanapin A.A.¹, Samsonova A.A.¹, Rozhmina T.A.³, Samsonova M.G.¹

¹ Centre for Computational Biology, Peter the Great St. Petersburg Polytechnic University, St. Petersburg, Russia

² The Ioffe Institute, St. Petersburg, Russia

³ Federal Research Center for Bast Fiber Crops, Thorzhok, Russia

* duk@mail.ioffe.ru

Key words: flax, genomics, sequencing, structural variation

Motivation and Aim: Flax is one of the significant agricultural crops for Russia cultivated for fiber and oil. The characterization of the genetic diversity is of great importance for the sustainability and diversification of the production of flax, as well as for successful breeding programs. The collection of the Federal Research Center for Bast Fiber Crops covers the most genetic diversity of this crop. It includes flax varieties from Eurasia with a large proportion of inherited Russian local forms, which distinguishes it from the collections used in previous genetic studies. Here we analyzed two types of structural variation, namely, presence-absence variations (PAV) and copy-number variations (CNV) in individual genomes from this collection.

Methods and Algorithms: Two datasets were analyzed, one of 100 accessions sequenced with a mean coverage of 20.6x, and the other of 306 accessions with mean coverage 9.3x. All plants have been grown in the experimental field of Federal Research Center for Bast Fiber Crops in Torzhok, Russia. For new insertion analysis we considered reads which were more than 1000 bp in length, do not align to the reference genome and have less than 25 % identity with it. CNVnator was used to analyze deletions and copy number increase [1]. Functional content of regions with structural variation was analyzed using flax genome annotation, Pfam database [2] and Gene Ontology (GO).

Results: Insertions of new regions in individual genomes (i.e. regions absent in the reference genome) were enriched for genes encoding proteins with antimicrobial functions (*MatE*, *Antimicrobial10*), phytopathogen cleavage (*TAXi_C*, *TAXi_N*), viral activity and controlling DNA-processes, energy exchange and cell membrane formation. Deletions in individual genomes happen in regions with genes involved in response to oxidative and salt stress, immune response (*TPR*, *PPR*, *ARM*), response to drought and temperature fluctuations (*WD40*, *TPR*, *LRR*), and energy exchange. Copy number increase happens in regions enriched for genes controlling response to abiotic stress (*RING/U-box*, *PPR*, *TTR*, *LRR*), seed formation and response to light. Both insertions and deletions contain plenty of domains related to retrotransposon activity, reflecting transposone relocation in individual genomes.

Conclusion: We found large structural variation in individual genomes of flax accessions that predominately happens in regions controlling stress and immune responses, membrane remodelling and energy metabolism. We conclude that individual genomes adapt to regional growing conditions through structural variation. Our results provide a valuable resource for future pangenomic studies.

Acknowledgements: The research is funded by the Ministry of Science and Higher Education of the Russian Federation under the strategic academic leadership program "Priority 2030" (Agreement 075-15-2021-1333 dated 30.09.2021).

References

1. Abyzov A., Urban A.E., Snyder M., Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res.* 2011;21(6):974-84. doi: 10.1101/gr.114876.110.
2. Mistry J., Chuguransky S., Williams L. et al. Pfam: The protein families database in 2021. *Nucleic Acids Res.* 2021;49(D1):D412-D419. doi: 10.1093/nar/gkaa913.