

Reconstruction and analysis of potato *Solanum tuberosum* pangenome of Siberian cultivars

Karetnikov D.I.^{1,2,3*}, Genaev M.A.^{1,2}, Nesterov M.A.^{1,2}, Ibragimova S.M.^{1,2}, Vasiliev G.V.^{1,2}, Toschakov S.V.⁴, Gavrilenko T.A.⁵, Afonnikov D.A.^{1,2}, Salina E.A.^{1,2}, Patrushev M.V.⁴, Kochetov A.V.^{1,2}

¹ Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

² Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia

³ Novosibirsk State University, Novosibirsk, Russia

⁴ Kurchatov Institute, Moscow, Russia

⁵ N.I. Vavilov All-Russian Institute of Plant Genetic Resources, St. Petersburg, Russia

* karetnikovmit@bionet.nsc.ru

Key words: potato, *Solanum tuberosum*, pangenome, genome structural variation, DNA sequencing, genome annotation

Motivation and Aim: Genome sequence of the crop opens the way for studying the structure of its chromosomes, distribution of repetitive and coding sequences, identification and annotation of genes. It makes possible to study gene functions and develop markers to find associations with phenotypic traits. However, there may be significant structural variations in the genomes of different organisms of the same species that cannot be represented based on the reference sequence alone and need to be described within the concept of pangenome, the complete set of genes of organisms of the same species. Sets of both common and variable genes can be distinguished in the pangenome. The variable part is particularly important for the study because it is formed by genes associated with plant adaptation to environmental conditions, resistance to diseases and abiotic and biotic stresses. *Solanum tuberosum* L. (potato) is the most important crop produced almost all over the world. Total production is over 370 million tons per year and the number of varieties is over 4000. A feature of potato genetics is the autoteraploid genome and high heterozygosity due to the predominantly vegetative method of propagation. In this paper, a study of pangenome of potato varieties of Siberian breeding was performed.

Methods and Algorithms: DNA from plants of 7 varieties was sequenced by paired-end short reads using the Illumina platform, allowing to obtain DNA assembly at the exome level. This assembly was further improved by using the DM1-3516 R44 reference genome. Structural variation was assessed for genotypes and were compared between Siberian cultivars and additionally four foreign varieties sequenced earlier [1].

Results: Based on the obtained assemblies, protein-coding genes were identified and the pangenome was analyzed. We determined its variable and conservative parts. It was shown that the pangenome structure of Siberian varieties is open, i.e., the pangenome size grows indefinitely as the number of individual genomes increases.

Conclusion: The results obtained provide an opportunity to further study the features of the structure and functions of protein-coding genes for potato varieties of Siberian cultivars.

Acknowledgements: The work was funded by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Kyriakidou M., Anglin N.L., Ellis D. et al. *Scientific Data*. 2020;7(1):1-6.