

Analysis of genome structure and its variations in potato cultivars grown in Russia

Karetnikov D.I.^{1, 2*}, Vasiliev G.V.^{1, 2}, Toshchakov S.V.³, Shmakov N.A.^{1, 2}, Genaev M.A.^{1, 2}, Nesterov M.A.^{1, 2}, Ibragimova S.M.^{1, 2}, Rybakov D.A.⁴, Gavrilenko T.A.⁴, Salina E.A.^{1, 2}, Patrushev M.V.³, Kochetov A.V.^{1, 2}, Afonnikov D.A.^{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 Kurchatov Institute, Moscow, Russia

⁴ N.I. Vavilov All-Russian Institute of Plant Genetic Resources (VIR), St. Petersburg, Russia

* karetnikovmit@bionet.nsc.ru

Key words: potato; *Solanum tuberosum* L.; genome structural variation; copy number variation; single nucleotide polymorphism; potato diversification

Motivation and Aim: *Solanum tuberosum* L. (potato) is the most important crop grown almost all over the world. The total production is more than 370 million tons per year and the number of cultivars is more than 4000. A peculiarity of potato genetics is the autotetraploid genome and high heterozygosity due to the predominantly vegetative mode of reproduction.

Whole-genome sequencing opens the way to study genome structure, identify single-nucleotide polymorphisms associated with different phenotypic traits, and reconstruct the pangenome at the level of a single species. One of the important features in genome analysis is copy number variation (CNV), defined as regions of the genome ranging in size from 1 kilobase to several megabases that exhibit different copy numbers in a population. At the functional level, CNV is associated with genes involved in stress responses and resistance to various factors.

Previously, 15 potato cultivars of *S. tuberosum* grown in Russia were sequenced [1]. Based on the sequenced genomes, the pangenome was reconstructed and potential CNVs associated with the genes were identified and analyzed. In this work, the sample of potatoes grown in Russia was expanded to 30.

Methods and Algorithms: DNA from 30 *S. tuberosum* cultivars was sequenced by paired-end short reads using the Illumina platform, which allowed DNA assembly at the exome level. Based on the short reads and DM1-3 v4.04 reference genome, CNVs in the genomes of the 30 cultivars were identified using the CNVpytor. In addition, we identified single nucleotide polymorphisms in all chromosomes, including extra chromosomes by GATK software package.

Results: Based on the collected genomes, we reconstructed the potato pangenome of *S. tuberosum*. CNV analysis revealed quantitative differences between cultivars grown in Russia and South American cultivars. Also, short reads made it possible to identify single nucleotide polymorphisms in all 30 genomes and conduct their further quantitative analysis.

Conclusion: Our findings lay a foundation for subsequent genetic research on Russian potato varieties, specifically focusing on how the structure of the genome and single nucleotide polymorphisms correlates with the plants' agronomic traits.

Funding: 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. This research was supported in part through computational resources of HPC facilities at collaborative center “Bioinformatics” ICG SB RAS.

References

1. Karetnikov D.I., Vasiliev G.V., Toshchakov S.V. et al. Analysis of Genome Structure and Its Variations in Potato Cultivars Grown in Russia. *Int J Mol Sci.* 2023;24:5713. doi 10.3390/ijms24065713