DOI 10.18699/PlantGen2021-071

De novo guar genome assembly and application of "omics" technologies to speed up guar (*Cyamopsis tetragonoloba* (L.) Taub.) breeding in Russia

Grigoreva E.A.[1, 2, 3]*, Tkachenko A.A.[2], Ulianich P.S.[1], Karzhaev D.S.[1], Barbitob Y.A.[2, 4], Changalidi A.I.[2, 4], Volkov V.A.[1, 3], Potokina E.K.[1, 3]
[1] *Sirius University of Science and Technology, Sochi, Russia*
[2] *ITMO University, St. Petersburg, Russia*
[3] *Saint-Petersburg State Forestry University, St. Petersburg, Russia*
[4] *Bioinformatics Institute, St. Petersburg, Russia*
* *email: L.Grigoreva@gmail.com*

Guar (*Cyamopsis tetragonoloba* (L.) Taub.) is a widely used plant mostly cultivated in India and Pakistan. Seeds of the guar is served as a rich source of galactomannan polysaccharide (guar gum) actively used in many fields of industry. The main interest in guar cultivation in Russia is supported by the high commercial value of guar gum for the oil industry. However, due to the long photoperiod typical for the most regions of Russia, guar cannot initiate flowering processes in time, which leads to a significant loss of yield. Thus, the deep understanding of genetic mechanisms underlying guar flowering processes is essential. The main problem for guar breeding is the lack of information about the guar genome. Omics technologies, combining different experimental approaches, could help to investigate important agricultural traits and, in particular, flowering processes. Here, for the first time, we present *de novo* guar genome assembly, constructed based on the results of short reads and long reads sequencing approaches. Genome assembly was performed by MaSuRCa assembler and contained 90.66% scaffolds with the length greater than 300 Kb with N50 = 2518853 b.p. This genome assembly was employed to generate genome-guided transcriptome assembly for guar and performing differential expression study between guar genotypes of different photoperiod sensitivity. For those guar genotypes results of metabolite profiling were also generated. We linked information about 1067 differentially expressed genes with concentrations of 64 metabolites and revealed that some biomarkers of early flowering in guar are linked to the specific transcripts. Besides, the created reference guar genome was used to generate a large set of SNP markers, useful for GWAS and marker-assisted breeding of the agriculturally important legume crop.
Acknowledgments: The reported study was funded by RFBR, project number 19-316-51007.