doi 10.18699/bgrs2024-06-43

Testing CAPS-markers designed on partially phased assembly of autotetraploid *Solanum tuberosum* genome

Larichev K.*, Sergeeva E., Karetnikov D., Afonnikov D.**, Salina E., Kochetov A.
Kurchatov Genomic Center of the Institute of Cytology and Genetics, SB RAS, Novosibirsk, Russia
** klarichev@bionet.nsc.ru; ** ada@bionet.nsc.ru*

Key words: *Solanum tuberosum*; genetic polymorphism; phased assembly; Illumina; CAPS markers

Motivation and Aim: As one of the key food crops, potato (*Solanum tuberosum* L.) is an attractive target for breeding. However, potato breeding is significantly complicated by its autotetraploid genome (2n = 4x = 48), as well as high heterozygosity and allelic diversity caused by clonal methods of reproduction. While marker assisted selection (MAS) can help alleviate some of the complications associated with tetraploid genome, taking the full advantage of MAS requires extensive knowledge of genetic polymorphisms that can be used as genetic markers. One of the most efficient ways to discover new genetic polymorphisms is by whole-genome sequencing. Recently, with the advent of long read sequencing and Hi-C, phased genome assembly, which provides the most comprehensive information regarding genetic polymorphisms, became viable even for polyploid species such as potato [1, 2]. However, such tools are quite expensive and not always available.

While the information about genomic variability acquired by short read sequencing may be incomplete (reads from different homologous chromosomes can be erroneously assembled into one contig), it can still be used to discover new genetic polymorphisms. Here we demonstrate how partially phased sequences assembled with Illumina short reads can be used to successfully discover novel genetic polymorphisms, even in organisms with complex genomes such as potato.

Methods and Algorithms: For our search of novel genetic polymorphisms we focused on 9 genes encoding the key enzymes of plant carbohydrate metabolism (ADP-glucose pyrophosphorylases *AGPase S3*, *AGPase S1*, beta-amylase *BAM1*, sucrose synthase *SuSy4*, starch phosphorylases *Pho1*, starch synthases *GBSS1*, *SS1*, *SS4*, Rubisco activase). Previously, using Illumina short reads, our colleagues assembled genomes of five Russian potato cultivars: Fritella, Krasavchik, Kolobok, Krepysh, Krasa Meschery. The coverage of assemblies on *S. tuberosum* DM1-3 R44 SolTub_3.0 genome assembly [3] ranged from 51 to 108 [4]. Reads belonging to selected genes were *de novo* assembled by SPAdes with default parameters and phased with freebayes and whatshap programmes. As a reference, we used gene sequences from DM1-3. The resulting partially phased gene sequences, as well as reference sequences and transcripts were aligned using MUSCLE algorithm in MEGA X. Conserved (no indels or substitutions) and polymorphic regions were annotated manually.

To verify SNPs predicted by assemblies, we used cleaved amplified polymorphic sequence (CAPS) assay. PCR primers for the assay were designed in Primer3 with the following considerations: (1) primers should anneal to conserved regions (preferably exons); (2) the length of PCR products should be 350-1100 bp; (3) each pair amplifies different parts of a gene; (4) amplification products should contain as least one SNP detectable using CAPS assay, i. e., one of the SNP alleles is a part of a restriction site.

Designed primers were used for CAPS assay of five Russian potato varieties (Fritella, Krasavchik, Kolobok, Krepysh, Krasa Meschery) in two replicas. Briefly, PCR products, amplified using cultivar genomic DNA as a template, were digested by an appropriate restriction enzyme (SibEnzyme, Russia) according to manufacturer's instructions. Digestion products were fractionated by agarose gel electrophoresis. Digestion patterns, corresponding to different SNP alleles, and the resulting cultivar allele distribution were discerned manually. PCR amplification was performed according to the following protocol: initial denaturation: 95 ℃ 5 min., denaturation 95 °C 15 sec., annealing 60 °C 15 sec., elongation 72 °C 30–90 sec. depending on the product length, final elongation 72 °C 5 min., 25 cycles in total.

Results: After *de novo* assembly and phasing, only two of the four haplotypes were resolved for each gene. For each cultivar, the resolved haplotypes demonstrated different SNP distribution. Using the assembled sequences, and taking into account SNP distribution in different cultivars, we designed 27 primer pairs amplifying various SNP containing regions of chosen genes. PCR products contained up to six SNPs, with three being the average. Observed amplicon length generally correlated to the predicted length in assembly. In some cases, deviations of 30-50 bp were observed, presumably caused by indels.

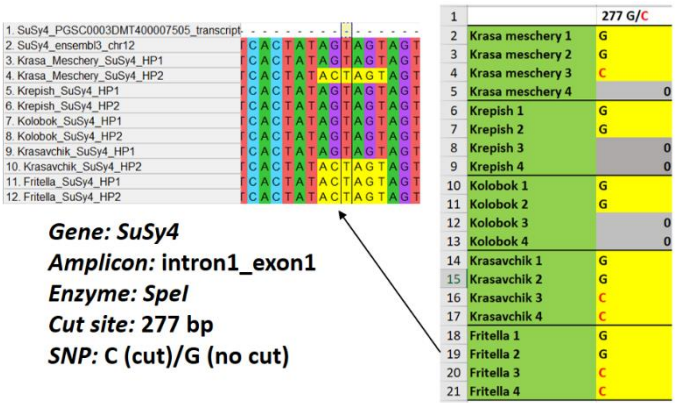

Fig. 1. Predicted haplotypes alignment in MEGA X and observed data obtained by CAPS assay

By using CAPS assay, we were able to detect more allelic variants (3–4) than was predicted by individual assemblies for most of the genes. For example, in the Fritella cultivar assembly, one of the SNPs in gene *SuSy4* was represented as 'C' allele in both haplotypes, however, with CAPS assay we additionally detected allele 'G' (Fig. 1). Thus, because of partial phasing, some SNPs were hidden in our assembly. It is worth noting that while the Fritella cultivar assembly did not show polymorphism in this site, it did in other cultivar's assemblies, which is why we designated this site as a SNP in the first place. In total, only *BAM1* gene had the same amount of detected and predicted alleles.

Conclusion: We demonstrated that partially phased assemblies can be used to discover novel CAPS markers for potato cultivars genotyping. Our results indicate that while some SNPs may be hidden due to partial phasing, and thus designated as 'non-polymorphic', using assemblies of several cultivars can help alleviate this problem by

providing additional information about these sites. In other words, combined, these assemblies provide more information about SNPs than individually, which should be taken into account when using partially phased assemblies as a means to find new genetic polymorphisms.

Funding: The study is supported by the Kurchatov Genomic Center of the Institute of Cytology and Genetics of Siberian Branch of the Russian Academy of Sciences, in agreement with the Ministry of Education and Science of the Russian Federation, No. 075-15-2019-1662.

References

1. Sun H., Jiao W.-B., Krause K. et al. Chromosome-Scale and Haplotype-Resolved Genome Assembly of a Tetraploid Potato Cultivar. *Nat Genet*. 2022;54:342-348. doi 10.1038/s41588-022-01015-0
2. Hoopes G., Meng X., Hamilton J.P. et al. Phased, chromosome-scale genome assemblies of tetraploid potato reveal a complex genome, transcriptome, and predicted proteome landscape underpinning genetic diversity. *Mol Plant*. 2022;15(3):520-536. doi 10.1016/j.molp.2022.01.003
3. The Potato Genome Sequencing Consortium. Genome Sequence and Analysis of the Tuber Crop Potato. *Nature*. 2011;475:189-195. doi 10.1038/nature10158
4. 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(6):5713. doi 10.3390/ijms24065713