Genetic foundations of breeding

UDC 633.49:631.522/524:577.21
doi: 10.15389/agrobiology.2018.5.907eng
doi: 10.15389/agrobiology.2018.5.907rus

Pho1a GENE FRAGMENT VARIABILITY IN TUBER-BEARING AND NONTUBER-BEARING POTATO SPECIES (Solanum subgenus Potatoe) AND S. tuberosum L. CULTIVARS

E.O. SHMELKOVA1, M.A. SLUGINA1, A.A. MELESHIN2, E.Z. KOCHIEVA1

1Research Center of Biotechnology RAS, Federal Agency for Scientific Organizations, 33/2, Leninskii prospect, Moscow, 119071 Russia, e-mail shmelkovfa@gmail.com (corresponding author), ekochieva@yandex.ru, mash-inmail@mail.ru.
2Lorkh All-Russian Research Institute of Potato Farming, Federal Agency for Scientific Organizations, 23, ul. Lorkha, pos. Korenevo-1, Lyubertsy Region, Moscow Province, 140051 Russia, e-mail a-mela@mail.ru

The authors declare no conflict of interests

Acknowledgements:
Received July 25, 2018

Abstract

Starch is the main metabolite in potato tubers. Therefore structure and functional analysis of starch metabolism genes are of fundamental and applied interest. The final starch amount in sink organs (fruits, seeds and tubers) depends not only on the amylose and amylopectin synthesis, but also on the catabolic enzymes activity. Proteins that participate in starch biosynthesis are rather well studied, while the starch degradation reactions are not fully understood. To date, more data on the crucial role of starch-phosphorylases in this process have been reported. Starch phosphorylases are widespread among plant species, but the coding genes structure and genetic diversity remain unclear. In potato tubers starch is cleaved by L-form of starch phosphorylase encoded by the Pho1a (STP23) gene. In the current work Pho1a gene fragment (exon II—exon IV) variability was analyzed for the first time in 15 wild and 81 cultivated potato accessions. The chosen gene fragment corresponds to the regulatory part of the glycosyltransferase domain and comprises glucose-6-P binding site, pyridoxal phosphate cofactor binding site and active site (glucose binding). The nucleotide and amino acid polymorphism is determined. A total of 96 potato accessions were used for allelic diversity analysis: 15 wild species from Potatooe and Estolonifera subsections (where S. etuberosum is a nontuberspecies), 67 cultivated potato varieties and 14 breeding lines of S. tuberosum. Nuclear DNA was isolated from young leaves using potassium-acetate method with phenol-chloroform additional purification. Primer combination Pho2F (5’-CTGAAATGGAAGCAAGCTTA-3’)—Pho4R (5’-GGCTATGGACTTAGGTACA-3’) was designed for chosen fragment amplification. The sequences of all varieties and breeding lines of S. tuberosum had a length of 670 bp. The length of the obtained Pho1a sequences in species ranged from 666 bp (S. vema, S. lignicaule) up to 672 bp (S. pinnatisectum). Totally 59 SNPs were detected, 15 of them localized in exons. It allowed us to identify 11 allelic variants, moreover 9 of them were found in wild species. Cultivated potato S. tuberosum has two allelic variants. The Pho1a_A2 allelic variant was observed in the majority of analyzed potato cultivars and all the breeding lines. Interestingly, the same variant was detected in some wild potato species, belonging not only to superseries Rotata, but also to Stellata that is considered to be more ancient. The Pho1a_A10 allelic variant was found in 9 cultivars (Bintje, Red Scarlett, Ushkonir, Karasaiskii, Aurora, Aladin, Chernskii, Plamy, Udacha). The Pho1a gene fragment translation revealed that 3 out of 15 exonic SNPs led to amino acid substitutions. In potato cultivars with the Pho1a_A10 allelic variant neutral M139I substitution was detected. The other neutral substitutions M139L and T157S were identified in S. ciriurfolium and S. vema, correspondingly. The only radical substitution R212S was detected in nontuber-bearing S. etuberosum. The potential role of the found amino acid substitutions in the functional protein domain requires the further investigation. Further search for the allelic variants associated with starch content in tubers can be used in potato breeding programs.

Keywords: starch phosphorylase, Pho1a, nucleotide and amino acid variability, wild spe-
Starch can be considered the main source of carbon and energy in plant cells. The final starch amount in heterotrophic organs (fruits, seeds, and tubers) depends not only on the synthesis of starch components, the amylose and amylopectin, but also on the activity of catabolic enzymes. Starch biosynthesis enzymes are rather well studied, while the starch degradation reactions are not fully understood [1]. The catabolism in the leaves is described in more details; some data proves catabolism in the endosperm of cereals, and less information is obtained for potato tubers [2].

In general, the starch-destroying enzymes can be divided into two categories: hydrolytic (ES 3.2.1.1, ES 3.2.1.2 amylases, 4-α-glucanotransferase ES 2.4.1.25, maltase ES 3.2.1.20, isoamylase ES 3.2.1.68) and phosphorylolytic (α-glycan phosphorylase ES 2.4.1.1, maltose phosphorylase ES 2.4.1.8, α-glucan-H₂O-dikinase ES 2.7.9.4, phosphoglucon-H₂O-dikinase ES 2.7.9.5) [3-5]. Their comparative activity may vary depending on the plant stage of development or environmental conditions. Which of enzymes groups has bigger importance is a rather controversial issue. It is supposed that the process itself has been initiated due to glycans phosphorylation that makes the starch grains surface more hydrophilic and, thus, more accessible to hydrolytic enzymes, creating selective protein–carbohydrate and protein–protein interactions additionally [5-8].

Starch phosphorylases, the plant analogs of α-glycan phosphorylases (ES 2.4.1.1), need more attention among other phosphorylolytic enzymes [9]. The fundamental role of starch phosphorylase is in the catalysis of starch decomposition due to the replacement of carbon with phosphorus in the glycoside bond with the formation of glucose-1-monophosphate [10, 11]. However, in the case of phosphate lack, starch phosphorylases can carry out the reverse reaction of starch synthesis [12].

Starch phosphorylases are widespread among plant species [13]. The presence of many isoforms is a characteristic feature of starch phosphorylases, which differ in kinetic properties and localization in the cell [13]. Most higher plants with known genomes and transcriptomes have two types of starch phosphorylases, the plastidic (Pho1/L-form/L-SP) and cytosolic (Pho2/H-form/r H-SP) [14]. The plastid protein of about 105 kDa has low affinity to branched glycans, while the cytosolic form of about 90 kDa is characterized by high affinity to linear and branched glycans and even to heteroglycans [15-17]. For the first time, various forms of plant starch phosphorylases have been found in potato tubers (Solanum tuberosum L.) and pea seeds (Pisum sativum L.) [18, 19]. Structure and genes polymorphism of starch phosphorylases are widely studied among the representatives of the Monocotyledons class: barley (Hordeum vulgare L.) [20], rice (Oryza sp.) [21], corn (Zea mays L.) [6, 22, 23]. At the same time, very few similar data are obtained for the Dicotyledons: the full-size sequence of the yam gene [24] and the cDNA of four potato varieties [25] are known.

Starch is the main metabolite in potato tubers; therefore, the structural and functional analysis of the genes of starch phosphorylases and evaluation of their variability is of not only fundamental but also applied interest because it can shed light on the function of these proteins, as well as be used in the selection of new varieties with high starch content, resistant to cold-induced sweetening.

Starch decomposition in potato tubers is carried out due to the L-form of starch phosphorylase, which is encoded by the Phola gene (STP23). The full-size sequence of this gene of potato is unavailable now. The NCBI GenBank contains only the corresponding S. tuberosum mRNA (NM 001288286.1); however, it is known that the gene is located on chromosome 3 [26], has a length of
16.4 kbp and consists of 15 exons and 14 introns [27].

In the present paper, the allelic polymorphism of the fragment of the *Pho1a* gene from exon II to exon IV on the broad samples, including both wild species of potatoes and the varieties and lines of *S. tuberosum* was analyzed for the first time. As a result, we identified 11 allelic variants that can be used further in breeding programs.

The work objective of the present investigation was to assess the variability of the starch phosphorylase *Pho1a* gene in the area of exons II-IV in tuber-bearing and non-tuber-bearing potato species (*Solanum*, the *Potatoe* subgenus), as well as in domestic and foreign varieties and lines of cultivated species of *S. tuberosum*.

**Techniques.** To investigate allelic polymorphism, 96 samples of potato were selected, including 15 wild species of the *Potatoe* and *Estolonifera* subsections, one of which was non-tuber-bearing (*S. etuberosum*), 67 varieties and 14 lines of cultivated potato of *S. tuberosum*. The sequence of exons of rice *Oryza sativa* (AK063766.1), corn *Zea mays* (NM 001309854.1), barley *Hordeum vulgare* (JQ277327.1), and yam *Ipomoea batatas* (L.) Lam. (M64362.1) available in the GenBank NCBI database were taken for additional evaluation of polymorphism. Seeds of wild species were obtained from the collections of Vavilov Institute of Plant Industry (Saint Petersburg, Russia) and CGN (Centre for Genetic Resources, Wageningen, the Netherlands). Varieties and lines are provided by Lorch Potato Research Institute (the Moscow Province, Russia).

Nuclear DNA was isolated from young leaves with the potassium acetate method with additional purification with the phenol-chloroform mixture [28]. Primer combination Pho2F (5´-CTGAA CATGAA GCAACGTA-3´)–Pho4R (5´-GGCTA-TGGACTTAGGTACA-3´) was designed for chosen fragment amplification. The reaction mixture for PCR contained 1× buffer (50 mM Tris-HCl, pH 8.0, 50 mM KCl, 1.5 M MgCl₂, 20 mM dNTPs), 10 µM of the respective primer, 0.25 U of Taq DNA polymerase (Dialat Ltd, Russia) and 100 ng of potato genomic DNA. The temperature-time profile of PCR was as follows: the first cycle — 5 min at 95 °C; 30 s at 94 °C, 40 s at 55 °C, 1 min at 72 °C (35 cycles); the final elongation — 1 min at 72 °C. PCR was carried out with commercial reagents (Dialat Ltd, Russia) in an amplifier Bio-Rad C1000 (Bio-Rad Laboratories, Inc., USA).

PCR products were visualized by electrophoresis in 1% agarose gel LE 2 Agarose (Helicon, Russia) with 1× TBE buffer stained with ethidium bromide, and documented in a BioDocII system (Biometra GmbH, Germany). Amplification products were sequenced with the same primers in the Bioengineering CCU on the platform Applied Biosystems 3730 DNA Analyzer (Applied Biosystems, USA).

For alignment and analysis of nucleotide and amino acid sequences polymorphism, MEGA 7.0 software was used (https://www.megassoftware.net/) [29]. The sequence of *S. tuberosum* mRNA from the GenBank NCBI database (NM 001288286.1) was the reference. The functional effect of amino acid substitutions was calculated with Provean program (http://provean.jcvi.org/index.php).

**Results.** To analyze the polymorphism of the *Pho1a gene*, a fragment from exons II to IV encoding the N-terminal regulatory part of the functional glycosyltransferase domain (ID 10136827) was selected, which includes glucose-6-phosphate binding site, pyrophosphate binding site and the active binding site for glucose [20]. The selection of sites for primers annealing was based on the possibility of the *Pho1a* amplification in phylogenetically distant species, as well as the ability to discriminate the target isoform of the gene from paralogs of *Pho1b* and *Pho2*. The gene fragment was amplified and sequenced with the
help of the developed Pho2f-Pho4R primer pair.

Sequences of all varieties and breeding lines of *S. tuberosum* had the same length, 670 bp. The length of the obtained Pho1a sequences in species ranged from 666 bp (*S. vernei, S. lignicaule*) up to 672 bp (*S. pinnatisectum*). The sequence lengths varied due to the presence of insertions and deletions localized in introns only.

In total, 59 point nucleotide substitutions (single nucleotide polymorphisms, SNPs) were detected among 96 potato samples under investigation; 58 of these SNPs were found in wild species sequences. The joint degree of polymorphism for the entire selection was 8.75%, and for a set of tuber-bearing and non-tuber-bearing potato varieties was 8.6%. Only two polymorphic sites (0.29%) were found in the analyzed varieties of the cultivated potato *S. tuberosum*, the breeding lines had no substitutions at all (Table 1). The great majority of substitutions were localized in the intron sequences. The variability of the studied exon sequences was 5.95% due to the presence of 15 SNPs, 14 of which were found in the species and only one replacement was found in the varieties of *S. tuberosum*.

### 1. The number and proportion of variable sites in the composition of the studied Pho1a sequences (II-IV exons) of wild species as well as varieties and lines of *Solanum tuberosum*

<table>
<thead>
<tr>
<th>Analyzed sequence</th>
<th>Wild species and the Prior variety</th>
<th>Varieties</th>
<th>Breeding lines</th>
<th>All samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full-length gene</td>
<td>58 (8.60 %)</td>
<td>2 (0.29 %)</td>
<td>0</td>
<td>59 (8.75 %)</td>
</tr>
<tr>
<td>cDNA</td>
<td>14 (5.55 %)</td>
<td>1 (0.39 %)</td>
<td>0</td>
<td>15 (5.95 %)</td>
</tr>
<tr>
<td>Protein</td>
<td>3 (3.57 %)</td>
<td>1 (1.19 %)</td>
<td>0</td>
<td>3 (3.57 %)</td>
</tr>
</tbody>
</table>

Based on the presence of the above-mentioned substitutions in the exon sequences, it is possible to distinguish 11 allelic variants (Table 2 submitted online on the journal website http://www.agrobiology.ru). Nine of them are typical for wild species, and seven are unique to the selection under investigation. The Pho1a_A1 allelic variant was typical only for the non-tuber-bearing variety *S. etuberosum*; however, despite its evolutionary distance from tuber-bearing species [30], the differences were of only one SNP. The multiple SNPs that distinguish the species of the *Rotata* superseries from the evolutionary more ancient species of the *Stellata* superseries were not identified as well (see Table 2 on the journal website http://www.agrobiology.ru). Moreover, some *Stellata* species (*S. polyadenium, S. chacoense, S. lignicaule*) had the same allelic variant Pho1a_A2 that the majority of the analyzed cultivars of *S. tuberosum*.

The analyzed sequences of the cultivated potato *S. tuberosum* were represented by two allelic variants. The Pho1a_A10 allelic variant was found in nine varieties of Russian and foreign selection (Bintje, Red Scarlet, Ushkonir, Karasaiskii, Aurora, Aladdin/Aladin, Chernskii, Plamya, Udacha). Varieties differed from wild species with the presence of nucleotide substitution G150A (the index corresponds to the number in the sequence of cDNA).

All 14 potato breeding lines were monomorphic in the Pho1a fragment and had one common Pho1a_A2 allele with the majority of varieties, which was also found in a number of wild species (see Table 2). The same allelic variant of Pho1a_A2 in wild species and most varieties can most likely be explained by the use of wild species samples in crossbreeding. For example, Charodei, Kholmogorski, Sudarinya, Ocharovanie, and Sirenevyi tuman varieties – hybrids obtained with the use of *S. demissum, S. vernei, S. stoloniferum*.

The attempt to associate the presence of a certain allelic variant with the qualitative characteristics of tubers (starch content, cold resistance, cold-induced sweetening) was not successful.
It is quite difficult to carry out a comparative analysis of the polymorphism of this part of the starch phosphorylase gene of potato and other plants since the Pho1a sequence is studied in a small number of monocotyledonous species [6, 20-23], and for dicotyledons, such information is extremely limited [24]. The polymorphism of the investigated plot of wild and cultivated tomato species was studied previously (Solanum, the Lycopersicon section). In total, 24 SNPs were detected in the composition of such sequences in 11 tomato species, 7 of them in the sequences of exons, 17 in the sequences of introns, the total polymorphism was 3.6% (M.A. Slugina, unpublished data). It is interesting that three species of wild potato (S. limbaniense, S. bulbocastanum, S. pinnatisectum) had indels common for all types of tomato in their introns sequences, which again reveal the phylogenetic relationships of tomato and potato.

Comparison of the Pho1a fragment (exons II-IV) in the studied potato samples with sequences in other species taken from the NCBI database showed a rather conservative structure of this gene area in the analyzed monocotyledonous and dicotyledonous plant species. At the same time, the homology of primary exon sequences was rather high, and that of intron sequences was very low. In the investigated area, the localization of introns coincided, but their lengths differed. The analysis revealed 87 SNPs (33.98%) (Fig., A, posted on the journal website http://www.agrobiology.ru). The greatest homology (89.45%) was in potato with yam. The most distant were the sequences of the Pho1a fragment of potato and corn (23.05%).

The Pho1 protein has a size of 966 a.a.r. and consists of four sections: N-terminal transit peptide, which delivers protein to plastids, a regulatory domain, central insertion L78 (distinguishes this isoform from Pho2), and a catalytic domain. The regulatory domain L78 and the catalytic domain form a functional glycosyltransferase domain, the function of which is to transfer sugar fragments from the activated donor molecule to the acceptor molecule, followed by the formation of a glycoside bond [20, 25]. When aligning the sequences of known starch phosphorylases of mono- and dicotyledonous plants, it is shown that the regulatory and catalytic domains are the most conservative [20].

Encoding sequences of the studied fragment (exon II—exon IV) of potato starch phosphorylase Pho1a gene were translated, which allowed estimating
their amino acid variability. The sequence available for analysis was 84 a.a.r. (positions from 129 to the 212 a.a.r.); it was a part of the regulatory domain and included glucose-6-phosphate and pyridoxal phosphate-binding sites, as well as the active binding site with glucose [20].

Only four of the 15 exon SNPs in the analyzed fragment led to the substitutions of amino acid residues. At the same time, A/T31 and G/T33 were parts of the same codon and led to the substitution of the same amino acid position: at position 139, methionine was replaced by isoleucine (M139I) or leucine (M139L). Substitution of M139I was typical for a group of varieties of Russian and foreign selection. Among them are Ushkonir, Karasaiskii, Aurora, Bintje, Chernskii, Red Scarlett, Aladin, Plamya, Udacha. Amino acid polymorphism was not found in other studied varieties and breeding lines. It is interesting that for the previously cloned and sequenced full-size cDNA of potato cultivars Dianna, Theresa, Saturna, and Satina, amino acid polymorphism on the plot of the second and fourth exons has not been identified [25]. That is, the presence of variants of the sequence of starch phosphorylase is shown on a wider sample of varieties and breeding lines of S. tuberosum presented in this article.

Significant variability of the Pho1a fragment has been identified in the wild species. The substitution M139L typical for wild tuber-bearing S. circacaeolium was detected. The amino acid substitution T157S unique for the selection was found in the S. vernei sample. However, according to the calculations in the Provean program [31], both substitutions are neutral and do not lead to a change in protein charge. R212S substitution in non-tuber-bearing S. etuberosum, in contrast to the above-mentioned, is the radical that may cause changes in the Pho1a protein structure and thus affect its functionality.

Alignment of the obtained amino acid sequences with translated sequences (exons II-IV) in rice, corn, barley, and yam revealed 14 amino acid substitutions, among which three were radical (see Fig. B, http://www.agrobiology.ru).

In general, the extremely low amino acid variability of the analyzed Pho1a regulatory domain in wild potato species and varieties is consistent with the data obtained for starch phosphorylases of other plant species [20], which indicates the important role of this domain for binding to substrates and cofactors, and thus explains the preservation of its conservatism during evolution. Amino acid substitutions identified in the regulatory domain can potentially affect the activity of the enzyme. In the future, it is necessary to continue the search for associations of detected substitutions with starch content in tubers, taking into account the prospects of their use in breeding programs.

Thus, the analysis of the area of Pho1a gene from exons II to IV in 15 wild species, 67 samples of cultivated potatoes and 14 breeding lines is carried out and description of its nucleotide and amino acid polymorphism is created on the basis of these data. Despite the conservatism of the studied exons, 59 nucleotide substitutions were described on this fragment, 15 of which were localized in exons, and three led to amino acid substitution. The unique replacement SNP636, leading to the substitution of arginine with serine R/S, found in non-tuber-bearing S. etuberosum, is radical and can potentially cause a change in protein conformation. In total, 11 allelic variants are determined. Further search for their associations with the starch content in potato tubers can be used in breeding programs. The influence of the identified amino acid substitutions in a functionally significant area and their potential impact on the activity of the starch phosphorylase enzyme requires further study.

REFERENCES


23. Nader Nieto C. Structural and functional characterization of natural alleles of potato (*Solanum tuberosum L.*) starch phosphorylases associated with tuber quality traits. University of Cologne,
Cologne, 2011.


