agronomy

MDPI

Article

Meta-Analysis of Genetic Factors for Potato Starch Phosphorylation

Vadim Khlestkin [1,2], Tatyana Erst [1], Alexander Igoshin [1], Irina Rozanova [3] and Elena Khlestkina [1,3,*]

[1] The Federal Research Center Institute of Cytology and Genetics SB RAS, 630090 Novosibirsk, Russia; dir2645@yandex.ru (V.K.); erst@bionet.nsc.ru (T.E.); igoshin@bionet.nsc.ru (A.I.)
[2] Russian Research Institute of Farm Animal Genetics and Breeding—Branch of the L.K. Ernst Federal Research Center for Animal Husbandry, 196601 Saint-Petersburg, Russia
[3] The N.I. Vavilov Federal Research Center All-Russian Institute of Plant Genetic Resources (VIR), 190000 Saint-Petersburg, Russia; fermoza@gmail.com
* Correspondence: director@vir.nw.ru

Abstract: Starch is one of the most demanded renewable feedstock in the world. The degree of phosphorylation of native potato (*Solanum tuberosum* L.) starch is a practically important quantitative trait, significantly influencing its physical and chemical properties. In this study, we evaluated the genetic diversity of the population of potato varieties and quantified phosphorus content in potato tuber starch harvested in 2017, 2018, and 2019. With the statistical methods, the most promising varieties for the next generation of breeding were identified for the first time. Genotyping and chemotyping data were utilized for genome-wide associations study (GWAS) in order to reveal genetic factors underlying the trait. GWAS based on a general linear model (GLM) with principal component analysis (PCA) was performed. The approach allowed us to identify two new, and confirm two previously found, significant SNPs on chromosome 5 associated with phosphorus content in starch. A search for the protein products coded in the genome regions carrying the significant SNPs revealed a cluster of genes that code glycoside and protein kinases, thus forming an operon-like structure. The genetic markers can be used for marker-assisted selection or to be considered as potential targets for genome editing to improve the industrially important properties of potato native starch via "intravital modification".

Keywords: potato; starch; meta-analysis; genetics; GWAS; phosphorylation; SNP; markers; gene cluster

Citation: Khlestkin, V.; Erst, T.; Igoshin, A.; Rozanova, I.; Khlestkina, E. Meta-Analysis of Genetic Factors for Potato Starch Phosphorylation. *Agronomy* **2022**, *12*, 1343. https://doi.org/10.3390/agronomy12061343

Academic Editor: Richard G. F. Visser

Received: 7 May 2022
Accepted: 29 May 2022
Published: 31 May 2022

1. Introduction

People's concerns about the environmentally friendly production of important industrial products, as well as the development of approaches to economically modify them, are pushing engineers to reduce the number of individual production steps. For chemical products, "one-pot" syntheses, continuous processes in flow reactors, and "atom economy" calculations have been developed. One of the promising approaches to economical and environmentally friendly methods for the synthesis of complex molecules is the "intravital modification" of the product. By "intravital modification" we mean the optimization of the chemical or physical structure and/or composition of biochemical products during their natural biosynthesis in the body by means of selection techniques, genome modification, or growing conditions. By affecting the genome of living producents with modern methods of genome editing or traditional breeding, as well as supplying with special components in the nutrient medium, it is possible to obtain economically valuable primary or secondary metabolites with improved structure and composition, baring practically significant properties.

Starch is one of the most important natural renewable resources [1]. In a recent article [2], we described the results of a genome-wide association study (GWAS) of potato *Solanum tuberosum* for phosphorus content (i.e., phosphate groups) in native potato starch. It

has been reasonably suggested that the phosphorus content of potato starch is an inherited trait. Several SNPs were identified on chromosomes 1, 4, 5, 7, 8, 10, and 11, associated with this trait. However, the *p*-values of most of the SNPs found did not overcome the Bonferroni level, probably due to the relatively small population size (90 varieties). So the SNPs needed further validation. In addition, a number of detected markers had no chromosome assignment at that time. Furthermore, it remained unclear whether the level of phosphorus content in potato starch can stably reproduce from year to year in different environmental conditions.

So, the main aim of the current study is to identify potato varieties and their special genetic features for "intravital modification" of potato tubers starch, namely, for an increase of phosphorus content in starch. For that, the following objectives have to be achieved: (1) meta-data on the content of phosphorus in potato starch for three years (90 varieties in 2017, 79 of them in 2018, and 74 in 2019) obtained; (2) a population structure evaluation to show proper population diversity for application in GWAS; (3) annotation of previously found and novel SNPs, associated with the trait; (4) identification of the most promising varieties with contrasting trait values.

Meta-analysis is a set of approaches for combining the results of several studies and it can increase the statistical power [3] of a scientific study. Overall, an increase in the reliability of the found markers, proof of the quality of the population, and the choice of promising varieties would make possible an effective next-generation breeding program for further "intravital modification" of starch to obtain its hyperphosphorylated modification, comparable in properties to the chemically produced analogs (for example, nutrition additive E1410).

2. Materials and Methods

2.1. Plant Material

The set of 90 potato (*Solanum tuberosum* L.) varieties from the ICG "GenAgro" collection (Novosibirsk, Russia) was grown during the period May to October 2017 in the same field in the Novosibirsk region (Michurinsky settlement, 54°52′ N and 83°00′ E). The growing of potato plants was performed according to the standard procedure. Briefly, seed tubers of all cultivars were planted in two rows with 0.75 m spacing and 0.3 m distance between the plants on the rows. In total, 10 plants were planted in the row, so, the length of each row was 10 m. Each cultivar was planted in three replicates, and the distances between the replicates' plots were 2 m. Sowing was performed in the first decade of May and harvesting in the 3rd decade of September. After harvesting, tubers were stored for 3 weeks at +4 °C. Only healthy tubers were collected for further analyses. From the healthy tubers, 25% of the smallest and largest tubers were removed. Among the rest, only five morphologically typical for a given variety of tubers were selected for starch isolation.

The procedure was repeated in 2018 and 2019 with 79 and 74 potato varieties, respectively. The names and varieties grown each year are shown in Supplementary Materials as Table S1.

2.2. Starch Isolation

Potato starch was isolated from the tubers according to the typical procedure, described elsewhere (for example, see [4]).

2.3. DNA Isolation and Genotyping

DNA was isolated from the tuber's skin using DNeasyPlant Mini Kit (Qiagen) according to the standard procedure. A set of 15, 214 (71.7%) scorable SNPs [2] was used for GWAS analysis. Genotyping information is available in [5].

2.4. Phosphorus Content Analysis

Analysis of the starch samples was performed using the phosphomolibdate method with spectrophotometric detection according to the standard procedure, described in GOST

7698–93 "Starch. Rules for acceptance and methods of analysis" (Russian and Belarus standard, correlates with ISO 3946-82). In brief, a starch probe was decomposed in a mixture of sulfuric and nitric acids, neutralized, and reacted with ammonium molybdate. Spectrophotometric detection at 825 nm allowed to determine the phosphorus content in the starch probe using a calibration curve.

2.5. PCA and Population Structure Analysis

The principal component analysis (PCA) is used to reveal correlations in the data and reduce the input data dimension with the least loss of information. For phenotypic data, PCA analysis was calculated by the standard method using the PAST 2.17 package [6]. PCA for genotyping data was conducted using the classical Gower algorithm through the distance matrix [7] with the JACOBI4 package [8]. Preliminarily, the tetraploid potato genome was encoded in digital format, where the effector allele was taken as 0, the non-effector allele was taken as 1, and the intermediate forms were coded as 0.25, 0.5, and 0.75. For example: AAAA allele is designated as 1, AAAG—as 0.75, AAGG—0.5, AGGG—0.25, and GGGG—as 0.

In order to calculate the population structure matrix (Q-matrix) containing membership coefficients for each individual of the potato mapping panel, the genotyping data was analyzed by Bayesian cluster analysis in STRUCTURE v.2.3.4 [9]. For passing the quality control for the analysis of the population structure, 4707 markers were selected, the data for which were known for all varieties. For data analysis, a model was chosen that assumes intraspecific hybridization and mixing of populations (admixture model). Estimation of k was carried out using hoc quantity (Δk) [10]. The length of the burn-in and Markov chain Monte Carlo (MCMC) was 5000.

2.6. Partial Least Squares (PLS) Analysis

PLS analysis was chosen as a method to treat phenotype and genotype data in order to identify varieties with genotypic and phenotypic variability for the trait of interest. PLS analysis was performed on the basis of the phenotypic trait measured for three years' harvests and for genotypic data using the PAST 2.17 software package [6]. Phenotypic (phosphorus content in the starch of three-year harvests) and genotypic sets of principal components were taken as blocks to perform the PLS analysis. As a result, phenotypic and genotypic bicomponents were obtained that corresponded to each other as much as possible.

2.7. Association Analysis

A general linear model (GLM) taking into account population membership estimates derived from principal components analysis was performed for phosphorus content measured every year with the help of TASSEL 5 package [11] to detect significant marker associations. Since TASSEL has especially been developed for diploid genome analysis, we re-coded tetraploid potato genome from four-letter code to numerical, taking into account the dose of a certain allele. For the details, refer to [2]. To identify significant SNPs the Bonferroni correction was applied. The Benjamin-Hochberg criterion [12] or false discovery rate (FDR) was also considered, with a threshold value of 0.05. To determine which markers crossed the threshold value according to the Benjamin-Hochberg test, the *p*-values obtained in the TASSEL 5 program were preliminarily ranked in ascending order. Further, the values were multiplied by the total number of tests (in our case, the number of studied markers was 15,214), and divided by the rank corresponding to the marker. The markers that passed the threshold value showed *p*-values less than 0.05 according to the Benjamin-Hochberg test. The purpose of the FDR criterion was to set the threshold below the specified Bonferroni value, which could facilitate the search for markers that are significant enough but did not pass the strict Bonferroni threshold. Inclusion criteria for SNPs were genotyping rate >90%, and minor allele frequency >5%. The relation of identified SNPs to genes and their

association with certain proteins was confirmed on the site https://plants.ensembl.org (accessed on 25 April 2022).

2.8. Meta-Analysis

The results of the genome-wide association analysis (*p*-values of the SNPs) for three years of phosphorus content were pooled using the Fisher's method modified for correlated data [13]. The use of an effective rather than a nominal number of tests in this method made it possible to avoid overestimation of SNP significance due to the correlation between the results. Before pooling, the *p*-values for each year were rescaled using the standard genomic control method which implies genomic inflation factor λ being brought to one [14].

2.9. Basic Local Alignment Tool (BLAST)

For identification of qualitative trait loci (QTL), genes, related proteins, and their statistical parameters BLAST supported by platforms https://plants.ensembl.org (accessed on 25 April 2022) and https://www.ncbi.nlm.nih.gov/ (accessed on 25 April 2022) were applied.

3. Results

The library of potato tuber starch of the three harvests of 2017, 2018, and 2019 was obtained through the starch isolation procedure described above. Chemotyping (phosphorus content) was performed in three replicates for each variety of each crop (2017–2019). The phosphorus content table is available in Supplementary Materials as Table S1. Phosphorus content in starch varied in 0.0456 to 0.1075% interval.

Pairwise correlations of phosphorus content in the starch of crops harvested in different years are depicted in Figure 1A–C.

Of 15,214 scorable SNPs, 4707 SNPs data was obtained for each potato variety and used as an input for evaluation of the quality of the population. Parameter Δk was calculated for the different number of subpopulations k. The magnitude of Δk is shown in Figure 2.

The population structure of potato varieties studied was calculated for potential 2 to 7 subpopulations and presented in Figure 3.

Correlations between the first genotype and phenotype bicomponents (see Discussion section) were calculated by the PLS method and depicted in Figure 4. As a result of PLS analysis, a clear correlation between first bicomponents was demonstrated, which implies a strong genetic basis of the trait studied. The opposite sides of the correlation cloud are promising for further investigation as their distinct genetics clearly result in different phenotypes.

We evaluated whether GLM + PCA analysis possesses proper accuracy in finding significant SNPs in the quantile-quantile plots. Difference in expected and calculated $-\log_{10}(p)$ did not exceed 5% (Figure 5).

Table 1. *p*-Values for SNPs with the highest $-\log_{10}(p)$ according to meta-analysis.

SNP	*p*-Value (2017)	*p*-Value (2018)	*p*-Value (2019)	*p*-Value Meta-Analysis
solcap_snp_c2_55899	6.14×10^{-5}	7.97×10^{-5}	7.00×10^{-4}	4.23×10^{-7}
PotVar0091465	2.46×10^{-4}	4.86×10^{-4}	3.83×10^{-4}	2.16×10^{-6}
PotVar0001436	7.86×10^{-4}	5.98×10^{-5}	1.20×10^{-3}	2.46×10^{-6}
solcap_snp_c2_50231	3.49×10^{-4}	9.93×10^{-5}	1.87×10^{-3}	2.69×10^{-6}

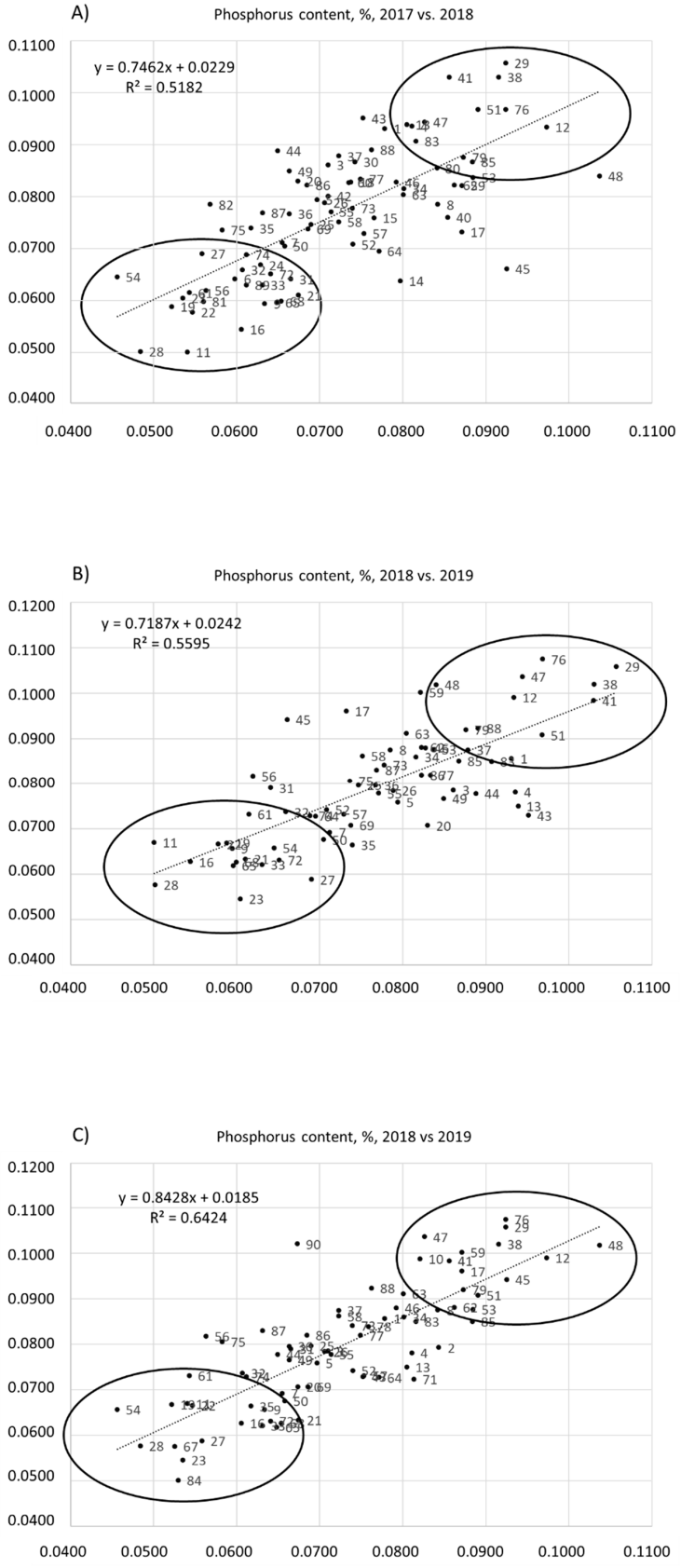

Figure 1. Pairwise correlations of phosphorus content in the starch of crops harvested in three years: (**A**) 2017 vs. 2018; (**B**) 2018 vs. 2019; (**C**) 2018 vs. 2019.

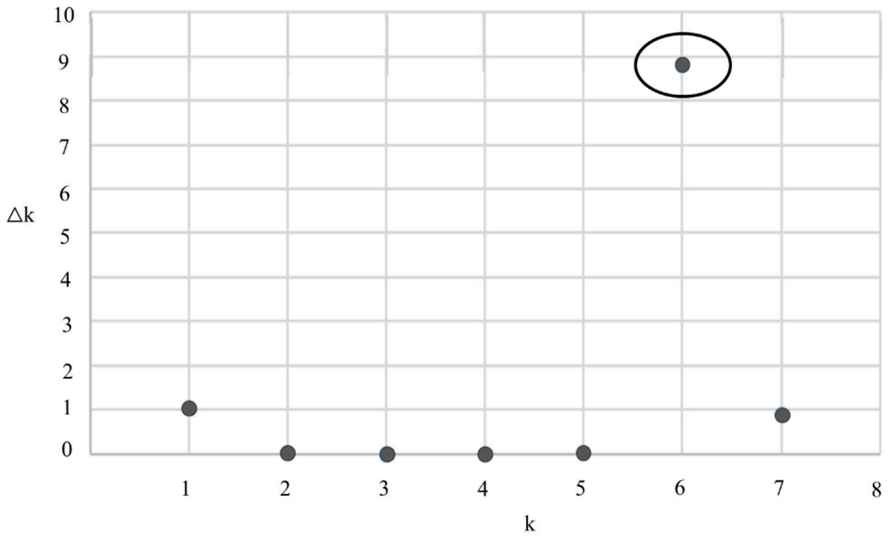

Figure 2. Plot Δk values vs. k. Oval encircles point for the optimal number of subpopulations k.

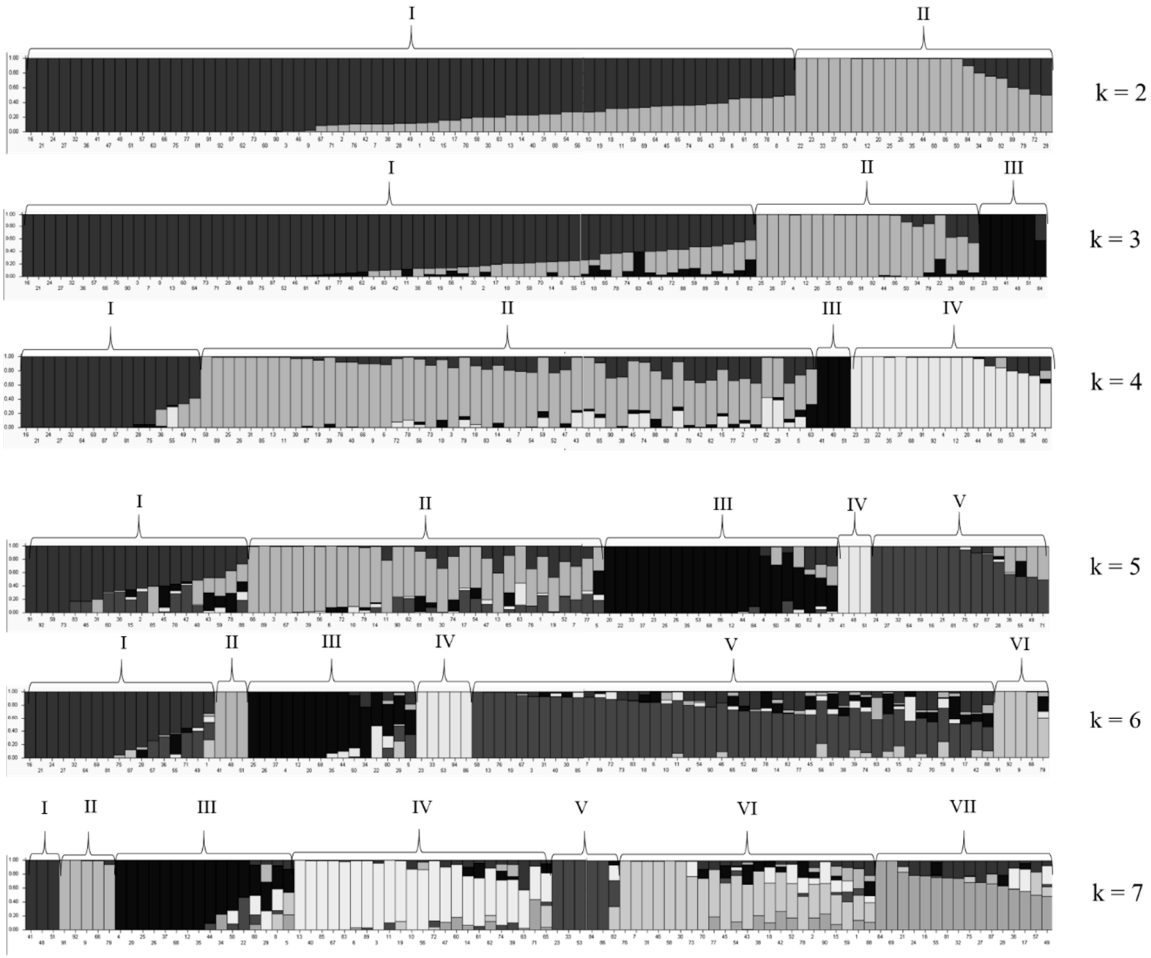

Figure 3. Structure of subpopulations of potato varieties in this study for k from 2 to 7.

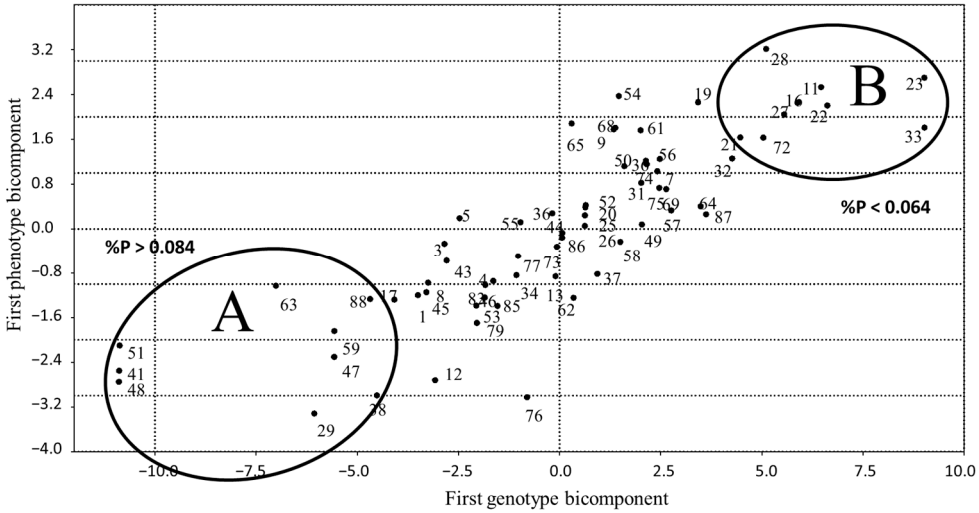

Figure 4. Correlation of phenotype and genotype bicomponents. Ovals encircle 10% quantiles from both sides of the cloud, related to the most promising varieties for breeding. Oval A includes nine varieties (29, 38, 41, 47, 48, 51, 59, 63, 88) with phosphorus content in starch >0.084%. Oval B includes nine varieties (11, 16, 21, 22, 23, 27, 28, 33, 72) with phosphorus content in starch <0.064%.

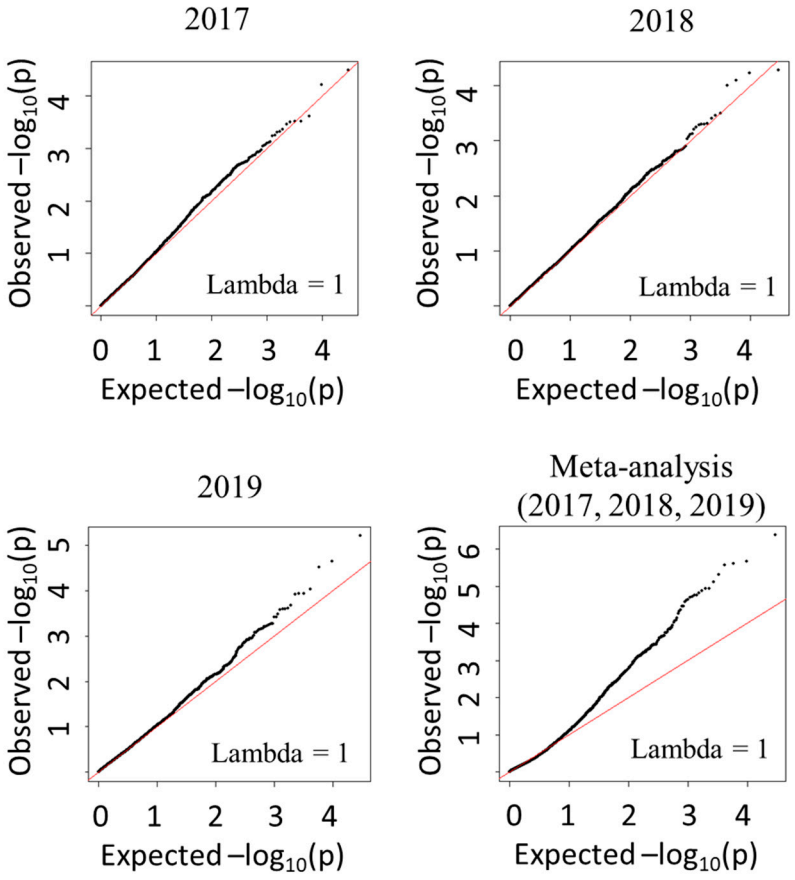

Figure 5. Quantile-quantile plots for $-\log_{10}(p)$ for 2017, 2018, 2019, and meta-analysis data.

Standard GWAS analysis was performed with phosphorus content data for 2017, 2018, and 2019 crops, followed by an analysis of meta-data, united all three years' data (Figure 6). The *p*-values of the most significant SNPs obtained in the meta-analysis are presented in Table 1.

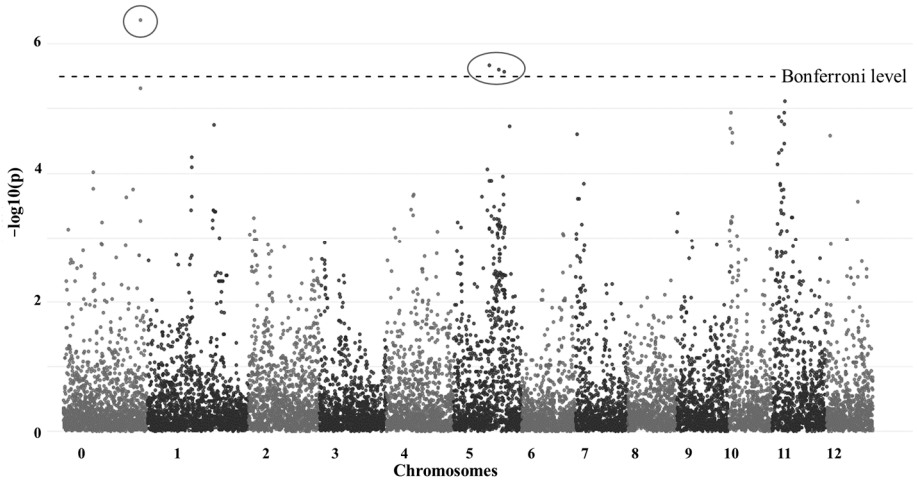

Figure 6. Manhattan plot for $-\log_{10}(p)$ according to meta-analysis data. Ovals encircle four SNPs with $-\log_{10}(p)$ overcoming Bonferroni level.

4. Discussion

In our paper [2], 17 SNPs and 8 new QTLs were identified on chromosomes 1, 4, 5, 7, 8, 10, and 11, with a high probability associated with the phosphorus content in starch. However, most of the *p*-values of the found markers laid within the false discovery rate but did not overcome the Bonferroni level. Chemotypic data (phosphorus content) for GWAS were obtained from the analysis of potato starch from a single crop in 2017, which means that there are still some issues to be addressed: whether the phosphorus content remains stable from year to year, and how much it depends on the agronomic background of potato cultivation and weather factor variations between years.

Thus, the results obtained in [2] may be validated in order to subsequently use the obtained markers in industrial potato breeding to obtain hyperphosphorylated starch by "intravital modification". There are known examples when the problem of a small number of available varieties was solved using a meta-analysis of data obtained under different agronomic conditions [15] or crops harvested at different times [16]. We used that approach in our study to confirm SNPs associated with phosphorus content in potato tuber starch.

The studied trait, phosphorus content in starch, showed a large positive correlation between the yields of different years (Pierson's coefficients > 0.7, Table 2). Moreover, the same varieties from year to year demonstrate the maximum or minimum values of the chemotypic trait (Figure 1), despite variable environmental conditions during 2017–2019. This indicates a significant contribution of genetic factors to the phosphorus content in the potato tuber starch trait and a minimal environmental effect.

Table 2. Pierson's correlation coefficients for pairwise comparison of phosphorus content in the starch of crops harvested in different years.

	2017	**2018**	**2019**
2017	1		
2018	0.72 *	1	
2019	0.80 *	0.75 *	1

* $p < 0.001$.

Once genetic contribution in the trait is suggested, proper population diversity for associations study should be provided. An independent analysis of 4707 SNP markers to identify the population structure yielded a dataset for estimating the number of subpopulations, which is Δk values plotted vs. k (Figure 2). A significant increase of Δk value at k = 6 indicates not less than 6 subpopulations in the varieties set studied. Figure 3 demonstrates the clusterisation of the set of varieties for k from 2 to 7. It shows that no

significant isolation of the varieties in any number of separate subpopulations is observed. Also, phosphorus content in starch does not depend on belonging to the subpopulations. Overall, that means, that the studied population of potato varieties is diverse enough to be efficiently used for GWAS analysis.

PCA analysis of the data on phosphorus content in starch showed that 84.8% and 9.5% phenotypic variance are explained by the first and second components, respectively. PCA for genotypic data showed the first three components explained 31.8% of the variance.

To calculate the relationship between phenotypic and genotypic bicomponents, a phenotype-genotype covariance analysis by the PLS method was performed and gave a set of linear bicomponents. The first bicomponent defines the most (91.5%) of the total covariance (Table 3).

Table 3. Linear bi-components for "phenotype-genotype" co-variation.

Axis	Sing.Val.	% Covar
1	5.5948	91.503
2	1.3609	5.4137
3	1.027	3.0835

Table 4 presents the correlation coefficients between the three bicomponents (both phenotypic and genotypic) and values of the chemotypic trait over three years. It can be seen that the content of phosphorus in potato starch strongly correlates only with the first bicomponent. Accordingly, the first bicomponent most fully reflects the total phosphorus content in potato starch, and thus only the first components (phenotypic and genotypic) are of interest for identification of the promising varieties for breeding and genetic modifications.

Table 4. Correlation coefficients for bi-components with phosphorus content in potato tuber starch for the crops of 2017, 2018, and 2019.

Year	First Bicomponent		Second Bicomponent		Third Bicomponent	
	Phenotype	Genotype	Phenotype	Genotype	Phenotype	Genotype
2017	0.93 *	0.78 *	0.14	0.09	0.36 *	0.27 *
2018	0.89 *	0.74 *	0.47 *	0.32 *	0.11	0.06
2019	0.94 *	0.79 *	0.22	0.21	0.20	0.21

* Significant values ($p < 0.05$).

The correlation between the vectors of the first phenotypic bicomponent and the first genotypic one is shown in Figure 6. When examining varieties based on the phosphorus content in starch, variety samples showing the maximum values of the trait in the sample, concentrated at one end of the correlation cloud (block A). Varieties with a lower percentage of phosphorus in starch are on the opposite side (block B). As a result, two groups of varieties were identified as genetically different with low and high phosphorus content in starch varieties. Accordingly, the selected groups of varieties with low and high values of the trait (percentage of phosphorus in starch) in the sample are potentially a significant genetic resource for further breeding.

None of the markers found in the current study showed a significant degree of reliability in the previous work [2]. The negative logarithm *p*-values of two of them overcame FDR, the other two have not been identified as significant markers of increased phosphorus content in starch at all. These markers and GWAS were not detected based on separately taken data on starch from the 2018 and 2019 crops either. However, as a result of the GWAS meta-analysis, the negative logarithm of the *p*-value of the SNPs found exceeded the Bonferroni level, making them validated and considered reliable biomarkers of phosphorus-enriched starch (Table 5). All found significant SNPs are located on

chromosome 5. The solcap_snp_c2_55899 marker is annotated according to [17]. Two polymorphisms (solcap_snp_c2_55899 and PotVar0091465) are located within the same QTL (positions 9,639,637 and 9,834,091), and the other two are quite far from each other, and from other markers: solcap_snp_c2_50231 is situated at position 36,834,630 and PotVar0001436 at position 20,274,370, thus marking three QTLs associated with the content of phosphorus in starch on the same chromosome.

Table 5. Significant SNPs discovered.

N	SNP	Ch	Position	Threshold		*p*-Value		Polymorphysm	Minor Allele	Minor Allele Frequency
				Current Study	Article [2]	Current Study, Meta-Analysis	Article [2]			
1	solcap_snp_c2_55899	5	9,639,637	Bonferroni	FDR	4.23×10^{-7}	9.01×10^{-6}	T/G	G	0.30
2	PotVar0091465	5	9,834,091	Bonferroni	Below FDR	2.16×10^{-6}	–	A/G	G	0.34
3	PotVar0001436	5	20,274,370	Bonferroni	Below FDR	2.46×10^{-6}	–	T/C	C	0.38
4	solcap_snp_c2_50231	5	36,834,630	Bonferroni	FDR	2.69×10^{-6}	2.70×10^{-5}	T/C	C	0.33

SNPs solcap_snp_c2_50231 and PotVar0001436 do not fall directly into any mapped gene. Known genes are not detected at a distance closer than ±15,000 bp from them. The PotVar0091465 SNP falls into the structure of the gene encoding the GWD protein (or Starch-granule-bound R1 protein), a known starch biosynthesis enzyme responsible for glycoside phosphorylation. Curiously, the second SNP of the same QTL, solcap_snp_c2_55899, is located next (±15,000 bp) to four annotated genes that also encode phosphorylating enzymes of the kinase class: LescPth4, two proteins designated as LescPth2, and an unknown kinase. However, all four of these proteins are not glycoside, but protein kinases, thus making up the family of proteins with similar functions. Due to the fact that their substrates are serine and threonine amino acids, which, like glycosides, bare hydroxyl groups for phosphorylation, and the average residual weight of the proteins GWD, LescPth4, LescPth2, and an unknown kinase are very close (111, 113, and 115 g/mol), all of the four genes are related in composition and origin, forming a superfamily of proteins with a similar function (Table 6). Their location within the same QTL allows for the identification of the detected genomic cluster [18], which probably is an operon-like structure [19] on chromosome 5 of the *S.tuberosum* genome.

Indeed, BLAST analysis of 1448 bp partial coding sequence for one of these proteins (Gene code PGSC0003DMG400016992, Transcript ID PGSC0003DMT400043783) allows a finding of 30 paralogous sequences ranging from 363 to 48 bp. All of them are located in positions from 9,700,478 to 9,402,861 on chromosome 5. That confirms the structural relationship of these genes, most likely formed as a result of segmental duplication.

Quite recently [20] allelic status of the *GWD* gene and thus the corresponding haplotypes were shown to be responsible for the variation of phosphorus content in potato tuber starch. However, in the current study, we showed that there are at least two more QTLs at chromosome 5, reliably associated with the trait. Additionally, the length of *GWD* does not exceed 20,000 bp (15,519 bp in [2], 19,188 bp in [20]) and solcap_snp_c2_55899 (position 5: 9,639,637) is situated far (194,454 bp) away from PotVar0091465 (position 5: 9,834,091, situated in *GWD*), which means that the *GWD*-containing QTL also contains a few more kinases-coding related genes.

Table 6. Description of genes and their products (according to Plant.Essemble.com), associated with the protein-coding SNPand QTLs identified in the current study.

N	SNP	Gene Code	Gene Statistics	Protein Name	Protein Statistics
1	solcap_snp_c2_50231	–	–	–	–
2	solcap_snp_c2_55899	PGSC0003DMG401016989	Exons: 2, Coding exons: 1, Transcript length: 2961 bps, Translation length: 212 residues	LescPth4, or protein serine/threonine kinase	Ave. residue weight: 113.766 g/mol Charge: 0.0 Isoelectric point: 6.4888 Molecular weight: 24,118.36 g/mol Number of residues: 212 aa
		PGSC0003DMG402016989	Exons: 1, Coding exons: 1, Transcript length: 1372 bps, Translation length: 68 residues	Protein kinase family protein	Ave. residue weight: 115.902 g/mol Charge: 2.5 Isoelectric point: 8.0567 Molecular weight: 7,881.32 g/mol Number of residues: 68 aa
		PGSC0003DMG400016991	Exons: 1, Coding exons: 1, Transcript length: 1101 bps, Translation length: 205 residues	LescPth2, protein kinase	Ave. residue weight: 113.319 g/mol Charge: −5.0 Isoelectric point: 5.4433 Molecular weight: 36,375.41 g/mol Number of residues: 321 aa
		PGSC0003DMG400016992	Exons: 1, Coding exons: 1, Transcript length: 1448 bps, Translation length: 120 residues	LescPth2, protein kinase	Ave. residue weight: 113.414 g/mol Charge: 7.0 Isoelectric point: 10.3705 Molecular weight: 15,991.43 g/mol Number of residues: 141 aa
3	PotVar0091465	PGSC0003DMG400007677	Exons: 34, Coding exons: 33, Transcript length: 5008 bps, Translation length: 1464 residues	GWD, or Starch-granule-bound R1 protein	Ave. residue weight: 111.535 g/mol Charge: −10.0 Isoelectric point: 5.8464 Molecular weight: 163,287.07 g/molNumber of residues: 1464 aa
4	PotVar0001436	–	–	–	–

5. Conclusions

The study demonstrated the possibility of GWAS analysis with a limited number (74–90 in this case) of samples to identify significant SNPs. At the same time, a meta-analysis of data from harvests of different years significantly increases the level of reliability of calculations. Meta-analysis made it possible to annotate two SNP markers on chromosome 5 and validate two previously identified markers to the level of practical significance with the potential to be used in potato breeding. An operon-like cluster of genes encoding a family of kinases and associated with phosphorus content in potato starch on chromosome 5 is identified. PLS analysis allowed us to identify the most promising potato varieties for the next generation of breeding techniques.

In the future, investigation of the discovered gene cluster can be aimed at using them as targets for breeding or genome editing in order to obtain potato and starch with a high content of phosphate groups.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/agronomy12061343/s1, Table S1: Phosphorus content in potato starch, %.

Author Contributions: Conceptualization, V.K. and E.K.; methodology, V.K.; software, A.I.; validation, T.E., I.R.; formal analysis, I.R., T.E.; investigation, T.E.; resources, E.K.; data curation, I.R.; writing—original draft preparation, V.K.; writing—review and editing, E.K.; visualization, I.R.; supervision, E.K.; project administration, V.K.; funding acquisition, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Ministry of Science and Higher Education of the Russian Federation, grant number FWNR-2022-0017.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Genotyping data is published elsewhere [Khlestkin VK, Erst TV, Rozanova IV, Efimov VM, Khlestkina EK. 2020. Genetic loci determining potato starch yield and granule morphology revealed by genome-wide association study (GWAS). PeerJ 8:e10286 DOI 10.7717/peerj.10286].

Acknowledgments: Authors thank Vadim M. Efimov (The Federal Research Center Institute of Cytology and Genetics SB RAS, Novosibirsk, Russia) for his consultation and advice on biological statistics.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Khlestkin, V.K.; Peltek, S.E.; Kolchanov, N.A. Review of direct chemical and biochemical transformations of starch. *Carbohydr. Polym.* **2018**, *181*, 460–476. [CrossRef] [PubMed]
2. Khlestkin, V.K.; Rozanova, I.V.; Efimov, V.M.; Khlestkina, E.K. Starch phosphorylation associated SNPs found by genome-wide association studies in the potato (*Solanum tuberosum* L.). *BMC Genet.* **2019**, *20* (Suppl. S1), 29. [CrossRef]
3. Law, M.; Jackson, D.; Turner, R.; Rhodes, K.; Viechtbauer, W. Two new methods to fit models for network meta-analysis with random inconsistency effects. *BMC Med. Res. Methodol.* **2016**, *16*, 87. [CrossRef] [PubMed]
4. Chung, H.-J.; Li, X.-Q.; Kalinga, D.; Lim, S.-T.; Yada, R.; Liu, Q. Physicochemical properties of dry matter and isolated starch from potatoes grown in different locations in Canada. *Food Res. Int.* **2014**, *57*, 89–94. [CrossRef]
5. Khlestkin, V.K.; Erst, T.V.; Rozanova, I.V.; Efimov, V.M.; Khlestkina, E.K. Genetic loci determining potato starch yield and granule morphology revealed by genome-wide association study (GWAS). *PeerJ* **2020**, *8*, e10286. [CrossRef] [PubMed]
6. Hammer, Ø.; Harper, D.A.T.; Ryan, P.D. PAST: Palentological statistics software package for education and data analysis. *Paleontol. Electron.* **2001**, *4*, 9.
7. Gower, J.C. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* **1966**, *53*, 325–338. [CrossRef]
8. Polunin, D.A.; Shtaiger, I.A.; Efimov, V.M. Development of the software package jacobi 4 for multidementional analysis of microchip data. *Bull. Novosib. State Univ. Ser. Inf. Technol.* **2014**, *2*, 90–98. (In Russian)
9. Pritchard, J.K.; Stephens, M.; Donnelly, P. Inference of population structure using multilocus genotype data. *Genet* **2000**, *155*, 945–959. [CrossRef] [PubMed]
10. Evanno, G.; Regnaut, S.; Goudet, J. Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Mol. Ecol.* **2005**, *14*, 2611–2620. [CrossRef] [PubMed]
11. Bradbury, P.J.; Zhang, Z.; Kroon, D.E.; Casstevens, T.M.; Ramdoss, Y.; Buckler, E.S. TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics* **2007**, *23*, 2633–2635. [CrossRef] [PubMed]
12. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J. R. Stat. Soc. Ser. B (Methodol.)* **1995**, *57*, 289–300. [CrossRef]
13. Cinar, O.; Viechtbauer, W. The poolr Package for Combining Independent and Dependent p Values. *J. Stat. Softw.* **2022**, *101*, 1–42. [CrossRef]
14. van den Berg, S.; Vandenplas, J.; van Eeuwijk, F.A.; Lopes, M.S.; Veerkamp, R.F. Significance testing and genomic inflation factor using high-density genotypes or whole-genome sequence data. *J. Anim. Breed. Genet.* **2019**, *136*, 418–429. [CrossRef] [PubMed]
15. Lo, S.; Muñoz-Amatriaín, M.; Hokin, S.A.; Cisse, N.; Roberts, P.A.; Farmer, A.D.; Xu, S.; Close, T.J. A genome-wide association and meta-analysis reveal regions associated with seed size in cowpea [*Vigna unguiculata* (L.) Walp]. *Theor. Appl. Genet.* **2019**, *132*, 3079–3087. [CrossRef] [PubMed]
16. Leonova, I.N.; Skolotneva, E.S.; Salina, E.A. Genome-wide association study of leaf rust resistance in Russian spring wheat varieties. *BMC Plant Biol.* **2020**, *20*, 135. [CrossRef] [PubMed]
17. Mengist, M.F. *Investigating the Genetics and Physiological Basis of Differences in Cadmium and Zinc Concentrations in Tubers of Potato (*Solanum tuberosum *L.): Implications for Food Safety and Biofortification*; University College Cork: Cork, Ireland, 2018.
18. Hourcadec, D.; Garcia, A.D.; Post, T.W.; Taillon-Miller, P.; Holers, V.M.; Wagner, L.M.; Bora, N.S.; Atkinson, J.P. Analysis of the human regulators of complement activation (RCA) gene cluster with yeast artificial chromosomes (YACs). *Genomics* **1992**, *12*, 289–300. [CrossRef]
19. Osbourn, A.E.; Field, B. Operons. *Cell. Mol. Life Sci.* **2009**, *66*, 3755–3775. [CrossRef] [PubMed]
20. Uitdewilligen, J.G.A.M.L.; Wolters, A.M.A.; van Eck, H.J.; Visser, R.G.F. Allelic variation for alpha-Glucan Water Dikinase is associated with starch phosphate content in tetraploid potato. *Plant Mol. Biol.* **2022**, *108*, 469–480. [CrossRef] [PubMed]