Identification of DNA markers of high protein content in pea seeds as a promising raw material for food production

Irina Mironova[1,2*], *Karina* Gainullina[3,4], *Artem* Slinkin[1], *Ilmir* Khabibullin[1], and *Ruzel* Khabibullin[1]

[1]FSBEI of HE "Bashkir State Agrarian University", Ufa, 450001, Russia
[2]FSBEI of HE "Ufa State Petroleum Technological University", Ufa, 450064, Russia
[3]Institute of Biochemistry and Genetics – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, 450054, Russia
[4]Experimental Station "Ufimskaya" – Subdivision of the Ufa Federal Research Centre of the Russian Academy of Sciences, Ufa, 450535, Russia

Abstract. The paper presents data reflecting the process of certification of peas as a promising plant raw material for food production. The elements of the crop structure and determination of the protein content in pea seeds were analyzed. Sources for selection have been identified that are characterized by high protein content in seeds, early ripening, high productivity, and resistance to biotic and abiotic factors. It is noted that typing using microsatellite markers makes it possible to create databases for identifying varieties and lines. This leads to the planning of effective hybridizations. identifying the most significant genetic polymorphism. According to phenological observations, accounting of elements of the crop structure and analysis of protein content in seeds for 2022-2023. 22 sources of selection-valuable traits were identified.

1 Introduction

An important aspect of modern research in the field of agricultural development is the development of technologies for the production of pea-based products. Pea (Pisum sativum L.) occupies a leading position among leguminous crops around the world. (Shelepina, 2016; Warkentin et al., 2020). It is used for food, feed and green manure purposes. (Novikov, 2012; Pelevina, 2017; Davletov et al., 2018) [1, 4]. To improve yield, grain quality, manufacturability and adaptability of peas, new varieties are created using genetic resources. [2-3].

Pea protein is safe for people with allergies to other foods [1, 4]. Growing peas in Russia is especially profitable due to its resistance to climate change. Our country is one of the leading exporters of peas, but until recently it imported pea isolate due to the lack of its own production [1-2, 12].

* Corresponding author: mironova_irina-v@mail.ru

In connection with the possibility of using peas as a promising plant raw material, it is necessary to certify this crop, and to increase the efficiency of breeding and reduce the time spent on creating new highly productive varieties rich in valuable protein content, identification of DNA markers of these traits is necessary [1].

For effective pea breeding, it is necessary to certify this crop and identify DNA markers that allow the creation of new highly productive varieties. The development of a microsatellite analysis system for certification of pea varieties and lines is an urgent task. [1-2].

In peas, intensive accumulation of nutrients in seeds during their ripening is regulated by transcription factors LEC1, LEC2, FUS3, and ABI3 (Malovichko et al., 2020). After the publication at the end of 2022 in GenBank of information on the primary structure of the genes of Pisum sativum L. ABI3 (AB080195.1), FUS3 (XM_051055759.1), LEC1 (X66368.1), LEC2 (XM_051040219.1), it became possible to conduct a comparative study nucleotide sequences of these genes in varieties and lines that differ in contrast in the content of seed storage proteins to search for mutations, which can later be used as DNA markers in the selection of peas for high protein content in seeds [3].

Research to search for DNA markers of high protein content in pea seeds will improve the selection of this crop and create a test system for identifying high-protein genotypes. This will open up new opportunities to improve product quality and ensure a stable supply of quality products on the market.

2 Materials and methods

The study of pea collection material was carried out in accordance with the methodological instructions of the VIR (Korsakov et al., 1975), phenological observations, records and measurements were carried out according to the methods of state variety testing of agricultural crops (Fedin, 1985). The protein content in seeds was determined using the Bradford method (Bradford, 1976). The seeds of each of the studied varieties were crushed in a mortar. Ground samples weighing 200 mg were placed in Eppendorfs, 1000 µl of phosphate buffer (0.05 M, pH=7.0) was added, mixed and left at +5°C for 12 hours. The samples were centrifuged at +5°C in a Centrifuge centrifuge 5430 R ("Eppendorf", Germany) for 20 min. 5 µl of supernatant was taken, 150 µl of Bradford reagent (Coomassie dye G-250) and 95 µl of distilled water were added. Protein concentrations were measured using an LS 55 Luminescence Spectrometer ("PerkinElmer", USA) in triplicate. The calibration curve was constructed using bovine serum albumin ("Sigma", USA) [4-5].

To isolate DNA, seeds of pea variety samples were germinated on Petri dishes under non-sterile conditions. DNA was isolated from 96-100 mg of 5-7-day-old seedlings homogenized in 200 µl of TE buffer using a commercial Genomic DNA Purification Kit (Thermo Fisher Scientific, USA). Add 400 µl of lysis buffer (Lysis solution) and incubate in a Termit thermostat (DNA-Technology, Russia) at 65°C for 5 min. Immediately, 600 µl of chloroform was added, shaken and centrifuged at 10 thousand rpm. within 10 min. on a MiniSpin centrifuge (Eppendorf, Russia). The supernatant was collected and 800 µl of precipitation buffer was added (Precipitation solution – 80 µl, sterile ampoule H_2O – 720 µl). Shake and centrifuge at 14 thousand rpm. within 15 min. The supernatant was removed, and the sediment was dissolved in 100 µl of NaCl solution. 300 µl of chilled ethyl alcohol (96%) was added and incubated at -70°C for 30 min. and DNA was precipitated by centrifugation at 14 thousand rpm. within 10 min. on a MiniSpin centrifuge (Eppendorf, Russia). Ethyl alcohol was removed, the precipitate was dried at room temperature for 2 hours and dissolved in 50 µl of sterile ampoule water [6-7].

Microsatellite analysis was carried out using the polymerase chain reaction (PCR) method [8] for 11 SSR loci from the Agrogene® genomic microsatellite library (Moissy Cramayel, France), of which 4 were selected by us in 2022 (AA355, AA200, D21, AD147) and 7 are included in the study in 2023. Their characteristics are presented in Table 1.

Table 1. Microsatellite markers used to analyze molecular genetic polymorphism of pea varieties and lines in 2023.

SSR marker	Primer sequence 5' to 3'	Number of alleles	PIC	Amplicon size, bp (Loridon et al., 2005)
AD174	GGAGGGATGATTCTAACAAGGT AACTCCCACTCCTCGAACTATT	6	0.81	96
AD61	CTCATTCAATGATGATAATCCTA ATGAGGTACTTGTGTGAGATAAA	7	0.84	138
AA355	AGAAAAATTCTAGCATGATACTG GGAAATATAACCTCAATAACACA	7	0.84	180
D21	TATTCTCCTCCAAAATTTCCTT GTCAAAATTAGCCAAATTCCTC	6	0.81	200
AA200	ACCGAAGAGCATTTTTCCTAAG TCCATCAGTTCCCTAATTCATC	4	0.67	220
AA285	TCGCCTAATCTAGATGAGAATA CTTAACATTTTAGGTCTTGGAG	5	0.69	248
AD270	CTCATCTGATGCGTTGGATTAG AGGTTGGATTTGTTGTTTGTTG	6	0.81	290
AD147	AGCCCAAGTTTCTTCTGAATCC AAATTCGCAGAGCGTTTGTTAC	7	0.84	330
AA67	CCCATGTGAAATTCTCTTGAAGA GCATTTCACTTGATGAAATTTCG	5	0.75	370
AD146	TGCTCAAGTCAATATATGAAGA CAAGCAAATAGTTGTTTTGTTA	7	0.84	390
AB29	GCTTTTTACCTTGCAGCCCA GAACAAGTCAATTCCTGATTTATTTCC	4	0.69	415

To amplify fragments of the coding regions of the *ABI3* and *LEC1* genes, we used primers ("Evrogen", Russia), which we selected for the first time using the PrimerSelect program ("DNAStar", USA). Amplification products were separated by horizontal electrophoresis in a Sub-Cell GT chamber (Bio-Rad, USA) in 1% agarose gel for 1 hour at a voltage of 120 V. Visualization and documentation of electrophoresis results were carried out in the Gel DocTM EZ gel documentation system System ("Bio-Rad", USA) using Image LabTM Software. Statistical data processing and cluster analysis were carried out using the StatSoft® STATISTICA 13.3 program.

3 Equations and mathematics

The amount of protein in pea seeds from the world collection of VIR ranges from 18 to 35% (Ashiev, 2014; Ponomareva et al., 2017). The protein content and quality are significantly influenced by the characteristics of different genotypes, as well as hydrothermal conditions, the degree of isolation in the field conditions and daylight hours (Voskobulova et al., 2019). In most regions of our country, the value of this indicator is 22-26% (Shelepina et al., 2010). For example, in the conditions of central Russia with a determined maximum grain yield, it contains 22-23% protein (Zelenov et al., 2016). Among those studying in 2022-2023. In the conditions of the Republic of Bashkortostan, 150 varieties and lines of peas of various ecological and geographical origin were selected:

the largest (23.5±0.4-26.1±0.5%) and the smallest (18 .0±0.3-19.7±0.3%).) The protein content in seeds varies greatly according to this indicator. They are distinguished from other species by their selection-valuable characteristics (earliness, high number of elements of seed productivity, resistance to lodging, non-scattering of seeds) or complex (Table 2).

Table 2. Characteristics of pea varieties released in 2022-2023. the highest (n=12) and lowest (n=10) protein content in seeds (differences are significant at p<0.05).

No./name of variety sample	No. in the VIR collection	Name of variety	Origin	Protein content of seeds
high-protein varieties				
6-20	K-9774	Stepnyak	Russia, Samara region	23.5±0.4%
10-20	K-10224	Ass 339	Syria	23.5±0.4%
2-30	K-1768	Shtambovyy Mal'tseva	Russia, Voronezh region	23.7±0.3%
12-30	K-9407	Avans	Russia, Altai region	23.7±0.4%
13-30	K-9432	Effektnyy	Ukraine	26.1±0.5%
14-30	K-9457	3698/04	Russia, Tyumen region.	24.1±0.4%
15-30	K-9524	–	France	25.0±0.3%
17-30	K-8788	Orel-326	Russia, Oryol region	23.7±0.4%
19-30	K-8793	Orel -332	Russia, Oryol region	23.7±0.3%
21-30	K-9383	Severyanin	Russia, Kirov region	23.8±0.4%
Batrak	–	Batrak	Russia, Oryol region	23.5±0.3%
Cobri	–	Cobri	Netherlands	25.5±0.5%
low-protein varieties				
1-20	K-8613	Wav 27101	Great Britain	18.0±0.3%
5-20	K-8972	SH 91-6-5-1-1-5	Bulgaria	19.0±0.2%
9-20	K-9950	Chlorotica	Sweden	18.0±0.3%
Fregat	K-9772	Fregat	Russia, Tatarstan	19.5±0.3%
5-30	K-3196	–	Türkiye	19.7±0.3%
8-30	K-7503	76-16-16	USA	19.4±0.2%
10-30	K-8902	Ascona	Netherlands	18.7±0.4%
23-30	K-9392	L-29201282	Australia	18.6±0.3%
24-30	K-9393	L-29201428	Australia	18.6±0.2%
29-30	K-9479	Peragis Beisaaterbse	Germany	19.7±0.3%

The varieties selected for analysis, presented in Table 2, were subjected to ssr analysis and amplification of sections of the *ABI3, LEC1, LEC2* and *FUS3* genes, which are responsible for regulating the expression of storage protein genes in related species of leguminous plants *Medicago truncatula Gaertn*., *Glycine max* L. Merr. and others, as indicated in the literature (Naito et al., 1994; Chen et al., 2018; Lalanne et al., 2021)/

4 Conclusion

Based on the results of two years of field research, analysis of the crop structure and determination of the protein content in pea seeds, sources for selection were identified that have a high protein content, early ripening, high productivity and resistance to various factors.

In the future, it is planned to purify the amplicons, their sequencing and comparative analysis of the nucleotide sequences of the ABI3, FUS3, LEC1 and LEC2 gene regions in varieties with high and low protein content in order to identify mutations associated with the ability to biosynthesize and accumulate protein in seeds. These data can be used in marker-based pea breeding.

Acknowledgement

The work was carried out under a GRANT from Rosselkhozbank JSC.

References

1. O.P. Dribnokhodova, S.A. Gostimsky, Allele polymorphism of microsatellite loci in pea (*Pisum sativum* L.) lines, varieties, and mutants, Russian Journal of Genetics, **45**, **7**, 900-906 (2009)
2. F.A. Davletov, K.P. Gainullina, F.F. Safin, Inheritance of seed size by hybrid populations of pea (F_1, F_2), BIO Web of Conferences, 17, 90-93 (2020) DOI: 10.1051/bioconf/20201700090
3. K.P. Gainullina, B.R. Kuluev, F.A. Davletov, Assessment of genetic diversity of pea varieties and lines using SSR analysis, Proceedings on applied botany, genetics and breeding, **181**, **3**, 70-80 (2020) DOI: 10.30901/2227-8834-2020-3-70-80
4. F.A. Davletov, K.P. Gainullina, Selection of high-yielding pea varieties in the conditions of the Cis-Ural steppe zone of the Bashkortostan Republic, BIO Web of Conferences, **36**, 01007 (2021) DOI: 10.1051/bioconf/20213601007
5. K.P. Gainullina, B.R. Kuluev, F.A. Davletov, Creation of source material for pea breeding using the method of chemical mutagenesis and assessment of its genetic diversity using SSR markers, Proceedings on applied botany, genetics and breeding, **183**, **3**, 111-122 (2022) DOI: 10.30901/2227-8834-2022-3-111-122
6. K. Gainullina, G. Galikeeva, E. Galimova, Study of genetic polymorphism of grain pea cultivars, E3S Web of Conferences, **420**, 01026 (2023) DOI: 10.1051/e3sconf/202342001026
7. T. Yang, L. Fang, X. Zhang, J. Hu, S. Bao, J. Hao, High-throughput development of SSR markers from pea (*Pisum sativum* L.) based on next generation sequencing of a purified Chinese commercial variety, PLoS One, **10**, **10**, e0139775 (2015) DOI: 10.1371/journal.pone.0139775
8. S. Singh, B. Singh, V.R. Sharma, M. Kumar, U. Sirohi, Assessment of genetic diversity and population structure in pea (*Pisum sativum* L.) germplasm based on morphological traits and SSR markers, Legume Research-An International Journal, **45**, **6**, 683–688 (2022) DOI: 10.18805/LR-4751
9. R. Liu, Y.N. Huang, T. Yang, J.G. Hu, H.Y. Zhang, Y.S. Ji, D. Wang, G. Li, C.Y. Wang, M.W. Li, X. Yan, X.X. Zong, Population genetic structure and classification of cultivated and wild pea (Pisum sp.) based on morphological traits and SSR markers, Journal of Systematics and Evolution, **60**, **1**, 85–100 (2022) DOI:10.1111/jse.12710
10. A. Mohamed, S. Garcia Martinez, P. Carbonell, J. Jose Ruiz, M. Loumerem, Genetic diversity assessment of Spanish and some endangered Tunisian pea (*Pisum sativum* L.) accessions based on microsatellite markers (SSRs), Chemistry & Biodiversity, **20**, **5**, e202201033 (2023) DOI: 10.1002/cbdv.202201033
11. H. Ram, N.K., Hedau G.V. Chaudhari, M. Choudhary, L. Kant, Genetic diversity assessment in pea (*Pisum sativum* L.) using microsatellite markers, International Journal of Bio-resource and Stress Management, **12**, **4**, 402–408 (2021) DOI : 10.23910/1.2021.2374a
12. J.D. Duque-Zapata, J.E. Munoz, O. Checa-Coral, Molecular characterization using SSR markers in 50 shrub pea genotypes (*Pisum sativum* L.) from the GRICAND Collection, Colombia. Revista Colombiana de Ciencias Hortícolas, **13**, **2**, 208–218 (2019) DOI: 10.17584/rcch.2019v13i2.10177