biology and life sciences forum

MDPI

Proceeding Paper

Variability of allergen-based length polymorphism of *Glycine max* L. varieties

Adam Kováčik[1]*, Jana Žiarovská[1], Lucia Urbanová [2]

[1] Slovak University of Agriculture in Nitra, Faculty of Agrobiology and Food Resources, Institute of Plant and Environmental Sciences, Tr. A. Hlinku 2, 949 76, Nitra, Slovak Republic; adam.kovacik565@gmail.com; jana.ziarovska@uniag.sk
[2] Slovak University of Agriculture in Nitra, Research Centre AgroBioTech, Department of Agrobiology, Tr. A. Hlinku 2, 949 76, Nitra, Slovak Republic; lucia.urbanova@uniag.sk
*Correspondence: adam.kovacik565@gmail.com

Abstract: Food allergies are an increasingly common phenomenon across all age groups and can be named as an epidemic of modern times. Legumes belong to nutritionally attractive crops, because of their high protein content and very well-balanced nutritional value. However, in addition to nutritionally valuable components, they contain a relatively high amount of anti-nutritional factors such as glycosides, lectins, inhibitors of digestive enzymes, and anti-nutritional proteins, what include allergens as well. Different genomic based analyses of allergen-coding parts are relevant in the research of legume gene resources. Here, a total of thirty different soybean varieties were analyzed for their polymorphism based on the specific homologous sequences of genes for vicilin and profilin, products of both belong to allergenic molecules of this species. A total of 16 different amplicons were obtained when profilin was used as marker and 17 different amplicons for vicilin. Comparing both of used techniques, vicilin provided more polymorphic profiles, but in five of analyzed varieties no amplicons were obtained. Profilin fingerprints provided higher degree of similarity coefficients among individual varieties of the soybean. Both of used PCR based techniques were proved to be applicable for genomic based screening of allergen homologs in the genetic resources of *Glycine max* L.

Keywords: allergens, profilin, vicilin, polymorphism, *Glycine max* L.

Citation: To be added by editorial staff during production.

Academic Editor: Firstname Lastname

Published: date

They are the most important part of antinutritional substances due to their frequent presence and reaction severity of human immune system.

1. Introduction

Legumes are essential crops thanks to their nutrition and growing attributes. They represent 27% of global primary crop production. Legumes, also with cereals, are considered an elementary part of nourishment because of the high level of proteins. Unfortunately, they produce a few antinutritional proteins which can act as allergens, digestive enzymes inhibitors, non-proteinogenic amino acids (Smýkal et al., 2015) or lectins. Allergens are a significant group causing immune responses from a mild OAS (Oral Allergy Syndrome) to a severe anaphylactic shock, which can lead to death. In legumes, the most important allergens are profilin, actin-binding protein, and vicilin, protein classified to 7/8S globulin group. Globulins are dominant allergens that the law draws attention to, therefore the soy or peanuts are highlighted in "ingredient part" of food products (Bennetau-Pelissero, 2019).

Globally, *Glycine max* L. is used to make food, feed or as a part of processing industry (Colletti et al., 2020). Soybean is a great source of proteins comparable with animal products (meat, eggs, or milk caseins), and a soy oil contains relevant quantity of saturated and unsaturated fatty acids, polysaccharides, dietary fiber, phytosterols and saponins (Modgil et al., 2021). Like other legumes, it contains antinutritional substances, especially

antinutritional proteins. Soybean is one of eight main foods causing more than 90% food allergies (Verma et al., 2012). Several soy allergens were described: non-specific LTP, defensin, profilin, PR-proteins, vicilin, legumin, 2S albumins or cysteine proteases (Breiteneder a Radauer, 2004).

It generates 2 immunologically severe allergens: vicilin a profilin. Vicilin is a main allergen evoking severe immune responses, profilin is a primary allergen for a fruit/vegetable/pollen cross-reacting pan-allergen (Jardim-Botelho et al., 2022). Prevalence of IgE-mediated soy sensitivity is 2,33% in Europe (Burney at al., 2014), but profilin sensitization to soy shows 27.3% of the Swiss population (Smýkal et al., 2015). Gly m 3 (soy profilin) shares 73% amino acid identity to Bet v 2 (profilin of birch-pollen) and to profilins of celery (Api g 4), carrot (Dau c 4), olive (Ole e 2), wheat (Tri a 12), peanut (Ara h 5), hazelnut (Cor a 2), latex (Hev b 8), or wormwood (Art v 4) (Nwaru et al., 2014). Gly m 5 (soy vicilin) belongs to Cupin superfamily as storage protein. In Europe, allergy prevalence to Gly m 5 is wide, from 5 to 67 % (Nwaru et al., 2014) and the protein shares about 50 % identity with other legume vicilins (of Ara h 1, Pis s 1, Len c 1 or Lup a vicilin) (Klemans et al., 2013).

The aim of this study was an application of DNA marker technique based on isoforms length polymorphism of the soy allergens in soybean varieties with determination of intraspecies variability at the genomic level.

2. Material and Methods

2.1 Plant material

Biological material consisted of a soy plants, grown in the field conditions in May 2021 (average temperature: 20 °C, day-length: 15h, number of rainy days: 11) and *in vitro* conditions in growth chamber (parameters simulated by yield conditions) in AgroBioTech Research Centre in Nitra (Slovakia). The choice of varieties was very random to increase possibility of a genome variability. The only condition was an equal representation of regional, high-yielding and rare varieties (Table 1).

Table 1. List of used *Glycine max* L. varieties used.

Sample number	Variety	Sample number	Variety	Sample number	Variety
1	Maria	11	Nattoking - K88	21	Gaillard
2	Mivak	12	Nattoking - K87	22	Mario
3	Arkadija	13	Danica	23	Rigel
4	Odesskaja	14	Balkan	24	Ugo
5	Maple Ridge	15	Schladming	25	Quito
6	McCall	16	Krajina	26	Dorota
7	OAC Scorpio	17	Odell	27	Emerson
8	Sibley	18	Maverick	28	Bristol
9	Sturdy	19	Accord	29	Belmont
10	Simpson	20	AC Glengarry	30	Crystal

2.2 DNA isolation

gDNA was extracted by GeneJET™ Plant Genomic DNA Purification Mini Kit (Thermo Scientific). Its functionality has been verified by ITS technique at a 1:9 dilution.

2.3 PBAP and VBAP analysis

Profilin variability was examined by two set of primers – specific and degenerated. Specific primers were designed to fruit profilins (Klongová, Kováčik et al., 2021), degenerate primers were designed on the base of conservative parts of profilin found in sequences of *Rosaceae* (taxid:3745) in NCBI database. (Table 2).

Table 2. Table of primers for PBAP

Specific primers:	
Forward:	5′ - ACCGGCCAAGATCTGGTTTT - 3′
Reverse:	5′ - AGGTAGTCTCCCAACCTCTCC - 3′
Degenerated primers:	
Forward:	5′ - AGAGAATTCCATATGTCGTGCCA**RR**CGTACGT - 3′
Reverse:	5′ - AGAAAGCTT**Y**TACA**K**GCC**Y**TGTTCA**BV**AGGTA - 3′

R - A/G
Y - C/T
K - G/T
B - C/G/T
V - A/C/G

Primers for vicilin were designed by sequence of *Lathyrus oleraceus* (pea) (Klongová, Kováčik et al., 2021) (Table 3).

Table 3. Table of primers for VBAP

Specific primers:	
Forward:	5′ - AGGGATCTTTATTGTTGCCA - 3′
Reverse:	5′ - TCATTTCTTTGACCCACAAG - 3′

PCR conditions were as follows: primary denaturation at 95 °C for 5 min; 40 cycles of denaturation at 95 °C for 45 s, annealing at 55 °C for 45 s, elongation at 72 °C for 35 s; and the last elongation at 72 °C for 10 min. It was used 400 nM primers, 1:9 diluted DNAs and DreamTaq™ DNA Polymerase (5 U/μL) by ThermoScientific™. PCR products were separated by electrophoresis in 2 % agarose gel stained by GelRed®. Sample profiles were transformed into binary matrices, processed by UPGMA statistical program (Garcia-Vallve et al., 1999) using Dice index (Dice, 1945) and dendrograms of genetic dissimilarity were created.

3. Results and discussion

3.1 PBAP of analysed G. max *varieties*

Visualization of PBAP profiles confirmed polymorphism of profilin isoforms in soy varieties and the technique created 16 loci of following lengths: 118 bp, 130 bp, 158 bp, 200 bp, 240 bp, 280 bp, 330 bp, 380 bp, 430 bp, 500 bp, 550 bp, 640 bp, 740 bp, 790 bp, 850 bp a 1000 bp. Overall, 73 alleles were identified for 21 soy varieties (in equal representation of regional, a high-yielding or rare).

Profiles were divided into 3 groups (Figure 1), the most abundant was the second including ‘McCall’, ‘Accord’, ‘Gaillard’, ‘Mario’, ‘Rigel’, ‘Quito’, ‘Emerson’, ‘Bristol’, ‘Belmont’, ‘Crystal’, whereas ‘Emerson’ and ‘Bristol’ were the most similar (sharing 81,25 % identity) and the most dissimilar was ‘Mario’ (only 25 % similarity). The most different profile of whole material collection produced variety ‘Mario’, on the other hand, 100% identical profiles shared ‘Danica’ and ‘Schlaming’.

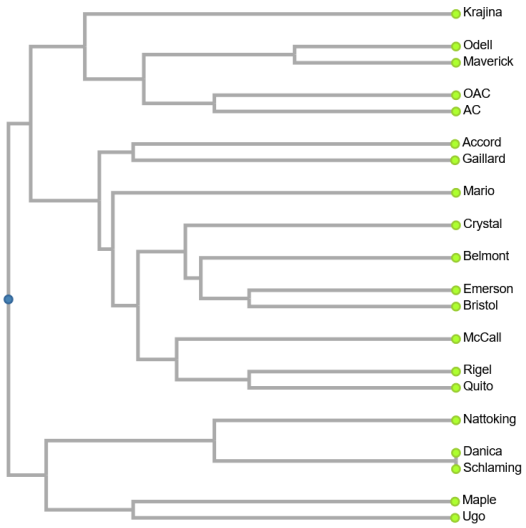

Figure 1. Dendrogram visualizing profilin allergen polymorphism in soybean varieties.

Profilin isoforms were identified in all varieties by specific primers, their profiles showed high level of intraspecies polymorphism.

Degenerated primers for profilin isoforms were not able to catch area of interest. PCR products were amplified only in varieties 'Dorota' and 'Bristol' with identical profiles of 2 alleles – 800 bp and 1000 bp.

The expression levels of allergens can vary due to various factors, such as plant cultivars (Marzano et al., 2020), growth and storage conditions and ripening stages. Proteomics can also yield much information about the expression levels of plant allergens in various environmental conditions (Nakamura and Teshima, 2013). One of the factors that affect allergen expression levels is the genetic background of plant cultivars (Marzano et al., 2020). In future, polymorphism of profilin genes could uncover one of the reasons for the variability of observed expression or immune reactions to soy. In addition to environmental factors, point to the genetic background that affects the expression level of pathogen related proteins, which also include allergens from the profilin family.

3.2 VBAP of analysed G. max *varieties*

VBAP technique was used to identify the presence of vicilin and to uncover possible intraspecies polymorphism in a captured genome region. The primers were able to amplify 36 alleles in 17 loci with following lengths: 106 bp, 126 bp, 151 bp, 179 bp, 255 bp, 285 bp, 300 bp, 327 bp, 402 bp, 500 bp, 560 bp, 596 bp, 634 bp, 714 bp, 762 bp, 813 bp a 929 bp. Primers were not able to catch vicilin fragments into genomes of 'Maple Ridge', 'OAC Scorpio', 'Nattong-K87', 'AC Glengarry' and 'Gaillard' (not included in dendrogram). Vicilin, as a storage protein, is essential for the plant, so it is a small possibility of its gene to absent in soy genome. However, it is most likely that the inability of the VBAP technique to catch the gene points to a high level of gene variability. To the date, 5 isoforms were described by Allergen Nomenclature Database, database UniProt showed 8 β-conglycinin subunits more than 90% similarity in amino acid sequences.

Similar as the BPAP results, varieties were divided into 3 bigger groups (Figure 2), but one of them included only 2 identical profiles of 'Schlaming' and 'Krajina' which were dissimilar to the rest profiles. On the other hand, absolute dissimilar profiles created 'Maverick' and 'Danica'. The most similar profiles (excluded identical) had varieties 'Ugo' and 'Belmont' sharing 44 % of alleles.

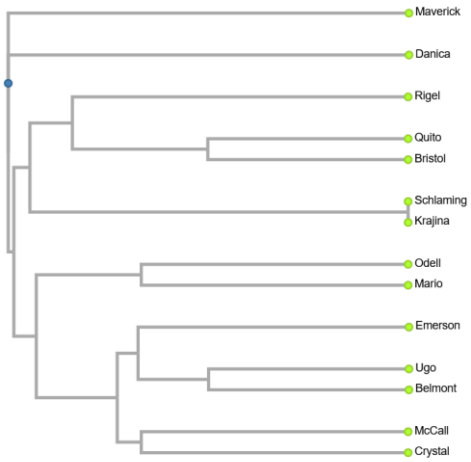

Figure 2. Dendrogram visualizing vicilin allergen polymorphism in soybean varieties.

4. Conclusion

Soybean is a good resource of important substances, but it also contains antinutritional chemicals, such as allergens. The most significant are profilin and vicilin. To understand different levels of allergy reactions to both, it is relevant to identify their isoforms and to describe the differences among them. Therefor PBAP and VBAP techniques were used to 30 varieties of *G. max*, PBAP created 29 polymorphic profiles, VBAP made 25, of which 2 were unique. VBAP technique was not able to amplify selected sequences in genomes of 5 varieties ('Maple Ridge', 'OAC Scorpio', 'Nattong-K87', 'AC Glengarry' and 'Gaillard') which are proper to next research.

Author Contributions: AK, JŽ and LU contributed to the proceeding paper equally.

Funding: This publication was supported by the Operational program Integrated Infrastructure within the project: Demand-driven research for sustainable and innovative food, Drive4SIFood 313011V336, co-financed by the European Regional Development Fund.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Allergen Nomenclature Database. Available online: www.allergen.org accessed on 13. 09. 2023.
2. Bennetau-Pelissero, C. Plant proteins from Legumes. Bioactive Molecules in Food; Mérillon, JM, Ramawat, K., Eds, 2019, 223-265.
3. Breiteneder, H.; Radauer, C. A classification of plant food allergens. Journal of allergy and clinical immunology 2004, 113.5: 821-830.
4. Burney, J.; Ramanathan, V. Recent climate and air pollution impacts on Indian agriculture. Proceedings of the National Academy of Sciences 2014, 111.46: 16319-16324.
5. Colletti, Alessandro, et al. Valorisation of by-products from soybean (Glycine max (L.) Merr.) processing. Molecules, 2020, 25.9: 2129.
6. Dice, L. R. Measures of the Amount of Ecologic Association Between Species. Ecology 1945, 26 (3): 297–302.
7. Garcia-Vallve, S.; Palau, J. and Romeu, A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in Escherichia coli and Bacillus subtilis. Molecular Biology and Evololution 1999, 16(9):1125-1134.
8. Jardim-Botelho, A., et al. Nutritional management of immediate hypersensitivity to legumes in vegetarians. Allergologia et Immunopathologia, 2022, 50.SP1: 37-45.
9. Klemans, R. JB., et al. Components in soy allergy diagnostics: Gly m 2S albumin has the best diagnostic value in adults. Allergy 2013, 68.11: 1396-1402.

10. Klongová, L., et al. Utilization of specific primers in legume allergens based polymorphism screening. Science, Technology and Innovation. Tarnow: Państwowa Wyższa Szkoła Zawodowa w Tarnowie 2021, 13(2).
11. Marzano, Valeria, et al. Perusal of food allergens analysis by mass spectrometry-based proteomics. Journal of proteomics, 2020, 215: 103636.
12. Modgil, Rajni, et al. Soybean (Glycine max). Oilseeds: health attributes and food applications, 2021, 1-46.
13. Nakamura, Rika; Teshima, Reiko. Proteomics-based allergen analysis in plants. *Journal of proteomics*, 2013, 93: 40-49.
14. National Centre of Biotechnology Information. Available online: www.ncbi.nlm.nih.gov
15. Nwaru, B. I., et al. Prevalence of common food allergies in Europe: a systematic review and meta-analysis. Allergy 2014, 69.8: 992-1007.
16. Smýkal, P., et al. Legume crops phylogeny and genetic diversity for science and breeding. Critical Reviews in Plant Sciences 2015, 34.1-3: 43-104.
17. UniProt: the Universal Protein Knowledgebase in 2023. Available online: www.uniprot.org accessed on 13. 09. 2023.
18. Verma, A. K., et al. A comprehensive review of legume allergy. Clinical reviews in allergy & immunology 2013, 45.1: 30-46.