

# Variability of Allergen-Based Length Polymorphism of *Glycine max* L. Varieties <sup>†</sup>

Adam Kováčik <sup>1,\*</sup> , Jana Žiarovská <sup>1</sup>  and Lucia Urbanová <sup>2</sup>

<sup>1</sup> Institute of Plant and Environmental Sciences, Faculty of Agrobiolgy and Food Resources, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia; jana.ziarovska@uniag.sk

<sup>2</sup> Department of Agrobiolgy, Research Centre AgroBioTech, Slovak University of Agriculture in Nitra, Tr. A. Hlinku 2, 949 76 Nitra, Slovakia; lucia.urbanova@uniag.sk

\* Correspondence: adam.kovacik565@gmail.com

<sup>†</sup> Presented at the 2nd International Online Conference on Agriculture, 1–15 November 2023; Available online: <https://iocag2023.sciforum.net/>.

**Abstract:** Food allergies are an increasingly common phenomenon across all age groups and can be called an epidemic of modern times. Legumes are a nutritionally attractive crop because of their high protein content and well-balanced nutritional value. However, in addition to nutritionally valuable components, they contain a relatively high amount of antinutritional factors such as glycosides, lectins, inhibitors of digestive enzymes, and antinutritional proteins, including allergens. Different genomic-based analyses of allergen-coding parts are relevant in research into legume gene resources. Here, a total of thirty different soybean varieties were analyzed for polymorphism based on the specific homologous sequences of genes for vicilin and profilin; products of both of these genes 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 were obtained when vicilin was used. Comparing both of the used techniques, vicilin provided more polymorphic profiles, but in five of analyzed varieties no amplicons were obtained. Profilin fingerprints provided a higher degree of similarity coefficients among individual varieties of the soybean. Both of used PCR-based techniques 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:** Kováčik, A.; Žiarovská, J.; Urbanová, L. Variability of Allergen-Based Length Polymorphism of *Glycine max* L. Varieties. *Biol. Life Sci. Forum* **2024**, *30*, 20. <https://doi.org/10.3390/IOCAG2023-16879>

Academic Editor: Martin Weih

Published: 11 March 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Legumes are essential crops thanks to their nutrition and growing attributes. They represent 27% of global primary crop production. Legumes, along with cereals, are considered an elementary part of nourishment because of their high levels of protein. Unfortunately, they produce a few antinutritional proteins which can act as allergens, digestive enzymes inhibitors, non-proteinogenic amino acids ([1], or lectins. Allergens are a significant group and the most important of the antinutritional substances due to their frequent presence and the reaction severity of the human immune system. Allergens cause immune responses from mild OAS (oral allergy syndrome) to severe anaphylactic shock, which can lead to death. In legumes, the most important allergens are profilin, actin-binding protein, and vicilin, a protein classified as belonging to the 7/8S globulin group. Globulins are dominant allergens that the law draws attention to; therefore, soy or peanuts are highlighted in the “ingredient part” of food products [2].

Globally, *Glycine max* L. is used to make food, animal feed, or in the processing industry [3]. Soybean is a great source of protein and is comparable to animal products (meat, eggs, and milk caseins); soy oil contains a relevant quantity of saturated and unsaturated fatty acids, polysaccharides, dietary fiber, phytosterols, and saponins [4]. Like other legumes, it contains antinutritional substances, especially antinutritional proteins. Soybean

is one of eight main foods causing more than 90% food allergies [5]. Several soy allergens have been described: non-specific LTP, defensin, profilin, PR-proteins, vicilin, legumin, 2S albumins, and cysteine proteases [6].

Soybean generates two immunologically severe allergens: vicilin and profilin. Vicilin is a main allergen evoking severe immune responses, profilin is a primary allergen and is a fruit/vegetable/pollen cross-reacting pan-allergen [7]. Prevalence of IgE-mediated soy sensitivity is 2.33% in Europe [8], while profilin sensitization to soy is found in 27.3% of the Swiss population [1]. Gly m 3 (soy profilin) shares 73% amino acid identity with Bet v 2 (profilin of birch pollen) and with 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*), and wormwood (*Art v 4*) [9]. Gly m 5 (soy vicilin) belongs to the Cupin superfamily and acts as a storage protein. In Europe, allergy prevalence to Gly m 5 ranges from 5 to 67% [9] and the protein shares about 50% identity with other legume vicilins (of *Ara h 1*, *Pis s 1*, *Len c 1*, and *Lup a*) [10].

The aim of this study was to apply a DNA marker technique based on the isoform length polymorphism of the soy allergens in soybean varieties to determine intraspecies variability at the genomic level.

## 2. Material and Methods

### 2.1. Plant Material

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

**Table 1.** List of *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, Waltham, MA, USA). Its functionality was 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 produce profilins [11], degenerated primers were designed on the base of conservative parts of profilin found in sequences of *Rosaceae* (taxid:3745) in the 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'-AGAGAATTCCATATGTCGTGCCARRCGTACGT-3'
Reverse:	5'-AGAAAGCTTYTACAKGCCYGTTCABVAGGTA-3'

R—A/G; Y—C/T; K—G/T; B—C/G/T; V—A/C/G.

Primers for vicilin were designed by the sequence of *Lathyrus oleraceus* (pea) [1] (Table 3).

**Table 3.** Table of primers for VBAP.

Specific primers:	
Forward:	5'-AGGGATCTTTATTGTTGCCA-3'
Reverse:	5'-TCATTCTTTGACCCACAAG-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. We used 400 nM primers, 1:9 diluted DNAs, and DreamTaq™ DNA polymerase (5 U/μL) by Thermo Scientific™. PCR products were separated by electrophoresis in 2% agarose gel stained by GelRed® (Biotium, San Francisco, CA, USA). Sample profiles were transformed into binary matrices and processed by the UPGMA statistical program [12] using the Dice index [13] and dendrograms of genetic dissimilarity were created.

### 3. Results and Discussion

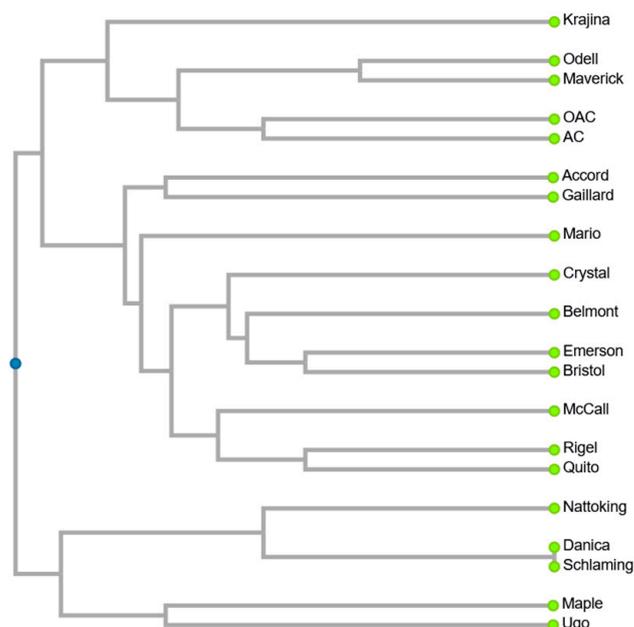
#### 3.1. PBAP of Analysed *G. max* Varieties

Visualization of PBAP profiles confirmed polymorphism of profilin isoforms in soy varieties and the technique created 16 loci with 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, and 1000 bp. Overall, 73 alleles were identified for 21 soy varieties (from an equal representation of regional, high-yielding, and rare varieties).

Profiles were divided into three groups (Figure 1); the most abundant was the second group, including 'McCall', 'Accord', 'Gaillard', 'Mario', 'Rigel', 'Quito', 'Emerson', 'Bristol', 'Belmont', and '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 the whole collection produced the variety 'Mario'; on the other hand, 100% identical profiles were shared by 'Danica' and 'Schlaming'.

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

Degenerated primers for profilin isoforms did not show an area of interest. PCR products were amplified only in varieties 'Dorota' and 'Bristol' with identical profiles of two alleles—800 bp and 1000 bp.



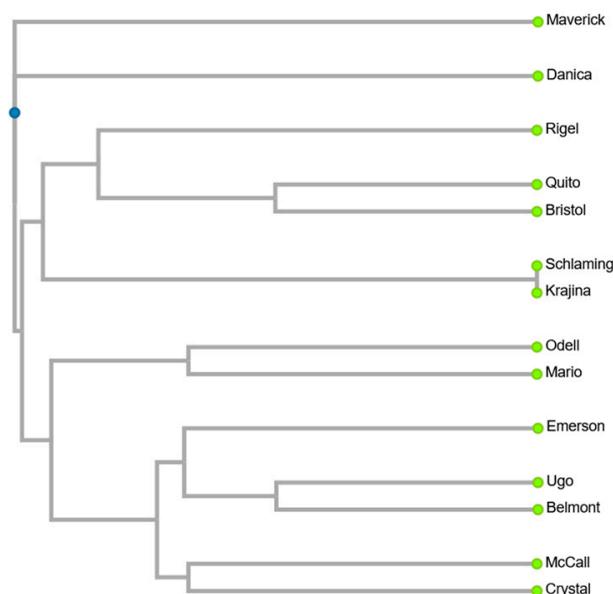
**Figure 1.** Dendrogram visualizing profilin allergen polymorphism in soybean varieties.

The expression levels of allergens can vary due to various factors, such as plant cultivar [14], growth and storage conditions, and ripening stages. Proteomics can also yield much information about the expression levels of plant allergens in various environmental conditions [15]. One of the factors that affect allergen expression levels is the genetic background of plant cultivars [14]. In the future, polymorphism of profilin genes could be found to be one of the reasons for the variability of observed expression or immune reactions to soy. In addition to environmental factors, the genetic background affects the expression levels of pathogen-related proteins, which also include allergens from the profilin family.

### 3.2. VBAP of Analysed *G. max* Varieties

The 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, and 929 bp. Primers were not able to identify vicilin fragments in the genomes of ‘Maple Ridge’, ‘OAC Scorpio’, ‘Nattong-K87’, ‘AC Glengarry’, or ‘Gaillard’ (not included in the dendrogram). Vicilin, as a storage protein, is essential for the plant, so there is a small possibility of its gene is absent in the soy genome. However, it is most likely that the inability of the VBAP technique to identify the gene points to a high level of gene variability. To date, five isoforms have been described in the Allergen Nomenclature Database [16] and database UniProt [17] showed eight  $\beta$ -conglycinin subunits with more than 90% similarity in their amino acid sequences.

Similar to the BPAP results, varieties were divided into three bigger groups (Figure 2); however, one of these included only two identical profiles—of ‘Schlaming’ and ‘Krajina’—which were dissimilar to the rest of the profiles. On the other hand, absolutely dissimilar profiles created ‘Maverick’ and ‘Danica’. The most similar profiles (excluding the identical ones) were of varieties ‘Ugo’ and ‘Belmont’, which shared 44% of their alleles.



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

#### 4. Conclusions

Soybean is a good source of important substances, but it also contains antinutritional chemicals such as allergens. The most significant of these are profilin and vicilin. To understand different levels of allergic reactions to both, it is relevant to identify their isoforms and to describe the differences among them. Therefore, PBAP and VBAP techniques were used to analyze 30 varieties of *G. max*; PBAP analysis was used to create 29 polymorphic profiles and VBAP analysis was used to create 25, of which 2 were unique. The VBAP technique was not able to amplify selected sequences in genomes of five varieties ('Maple Ridge', 'OAC Scorpio', 'Nattong-K87', 'AC Glengarry', and 'Gaillard'). This will be the focus of future research.

**Author Contributions:** A.K., J.Ž. and L.U. contributed equally to this proceeding paper. All authors have read and agreed to the published version of the manuscript.

**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 and by GA FAPZ-Odrodové odlišnosti exprese génov alergénnych bielkovín v genetických zdrojoch strukovín.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** No new data were created.

**Conflicts of Interest:** The authors declare no conflicts of interest.

#### References

1. Smýkal, P.; Coyne, C.J.; Ambrose, M.J.; Mated, N.; Schaefer, H.; Blair, M.W.; Berger, J.; Greene, S.L.; Nelson, M.N.; Besharat, N.; et al. Legume crops phylogeny and genetic diversity for science and breeding. *Crit. Rev. Plant Sci.* **2015**, *34*, 43–104.
2. Bennetau-Pelissero, C. Plant proteins from Legumes. In *Bioactive Molecules in Food*; Mérillon, J.M., Ramawat, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2019; pp. 223–265.
3. Colletti, A.; Attrovio, A.; Boffa, L.; Mantegna, S.; Cravotto, G. Valorisation of by-products from soybean (*Glycine max* (L.) Merr.) processing. *Molecules* **2020**, *25*, 2129. [[CrossRef](#)] [[PubMed](#)]
4. Modgil, R.; Tanwar, B.; Goyal, A.; Kumar, V. Soybean (*Glycine max*). In *Oilseeds: Health Attributes and Food Applications*; Springer: Singapore, 2021; pp. 1–46.
5. Verma, A.K.; Kumar, S.; Das, M.; Dwivedi, P.D. A comprehensive review of legume allergy. *Clin. Rev. Allergy Immunol.* **2013**, *45*, 30–46. [[CrossRef](#)] [[PubMed](#)]

6. Breiteneder, H.; Radauer, C. A classification of plant food allergens. *J. Allergy Clin. Immunol.* **2004**, *113*, 821–830. [[CrossRef](#)] [[PubMed](#)]
7. Jardim-Botelho, A.; de Oliveira, L.C.L.; Motta-Franco, J.; Solé, D. Nutritional management of immediate hypersensitivity to legumes in vegetarians. *Allergol. Immunopathol.* **2022**, *50*, 37–45. [[CrossRef](#)] [[PubMed](#)]
8. Burney, J.; Ramanathan, V. Recent climate and air pollution impacts on Indian agriculture. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 16319–16324. [[CrossRef](#)] [[PubMed](#)]
9. Nwaru, B.I.; Hickstein, L.; Panesar, S.S.; Roberts, G.; Muraro, A.; Sheikh, A.; EAACI Food Allergy and Anaphylaxis Guidelines Group. Prevalence of common food allergies in Europe: A systematic review and meta-analysis. *Allergy* **2014**, *69*, 992–1007. [[CrossRef](#)] [[PubMed](#)]
10. Klemans, R.J.; Knol, E.F.; Michelsen-Huisman, A.; Pasmans, S.G.; de Kruijf-Broekman, W.; Bruijnzeel-Koomen, C.A.; van Hoffen, E.; Knulst, A.C. Components in soy allergy diagnostics: Gly m 2S albumin has the best diagnostic value in adults. *Allergy* **2013**, *68*, 1396–1402. [[CrossRef](#)] [[PubMed](#)]
11. Klóngová, L.; Kováčik, A.; Urbanová, L.; Kysel, M.; Ivanišová, E.; Žiarovská, J. Utilization of specific primers in legume allergens based polymorphism screening. *Sci. Technol. Innov.* **2021**, *13*, 12–21. [[CrossRef](#)]
12. Garcia-Vallve, S.; Palau, J.; Romeu, A. Horizontal gene transfer in glycosyl hydrolases inferred from codon usage in *Escherichia coli* and *Bacillus subtilis*. *Mol. Biol. Evolution* **1999**, *16*, 1125–1134. [[CrossRef](#)] [[PubMed](#)]
13. Dice, L.R. Measures of the Amount of Ecologic Association Between Species. *Ecology* **1945**, *26*, 297–302. [[CrossRef](#)]
14. Marzano, V.; Tilocca, B.; Fiocchi, A.G.; Vernocchi, P.; Mortera, S.L.; Urbani, A.; Roncada, P.; Putignani, L. Perusal of food allergens analysis by mass spectrometry-based proteomics. *J. Proteom.* **2020**, *215*, 103636. [[CrossRef](#)]
15. Nakamura, R.; Teshima, R. Proteomics-based allergen analysis in plants. *J. Proteom.* **2013**, *93*, 40–49. [[CrossRef](#)] [[PubMed](#)]
16. National Centre of Biotechnology Information. Available online: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov) (accessed on 12 September 2023).
17. UniProt: The Universal Protein Knowledgebase in 2023. Available online: [www.uniprot.org](http://www.uniprot.org) (accessed on 13 September 2023).

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.