

Proteomic analysis of Costa Rican landraces of *Phaseolus vulgaris* L. and *Cajanus cajan* (L.) Millsp. enhanced by High- Performance Thin-Layer Chromatography (HPTLC)

Kerling Rodríguez-Quesada

National University of Costa Rica

Yohana Alfaro-Ureña

National University of Costa Rica

Humberto Trimiño-Vásquez

National University of Costa Rica

Manuel Sandoval-Barrantes

National University of Costa Rica

Ana Francis Carballo-Arce

National University of Costa Rica

Luis Roberto Villegas-Peñaranda

`luis.villegas.penaranda@una.ac.cr`

National University of Costa Rica

Article

Keywords: Phaseolus vulgaris landraces, Cajanus cajan (pigeon pea), High-Performance Thin-Layer Chromatography (HPTLC), Legume proteomics, Storage-protein diversity

Posted Date: May 6th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8981539/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Common beans are a nutritionally significant legume, valued for their diverse protein composition, influencing nutritional quality, sensory properties, and pest resistance. This study developed a proteomic analysis enhanced via High-Performance Thin-Layer Chromatography (HPTLC) to profile protein fractions in three Costa Rican *P. vulgaris* landraces (Surú, Brunca, and Mantequilla) and *Cajanus cajan* (L.) Millsp. (pigeon pea) as a wild-type legume. As a result, HPTLC retention factors correlated with SDS-PAGE results, identifying albumins in *P. vulgaris* and vicilin-type globulins in *C. cajan*. Mass spectrometry confirmed phaseolins as the predominant storage proteins in *P. vulgaris*, while *C. cajan* exhibited a higher abundance of vicilin-type globulins and β -conglycinin. The integration of SDS-PAGE, HPTLC, and LC-MS/MS verified protein identities and provided insights into their solubility and stability under various extraction conditions. These findings highlight HPTLC as a rapid and complementary method for protein separation, supporting SDS-PAGE-based molecular weight determination. Additionally, the confirmed presence of phytohemagglutinin (PHA) in *P. vulgaris* and vicilin globulins in *C. cajan* suggests potential molecular markers for breeding programs to improve legumes' nutritional quality and pest resistance. The application of HPTLC in legume proteomics offers a valuable tool for characterizing protein diversity and functional properties, with implications for food chemistry and agricultural biotechnology.

Introduction

Beans (*Phaseolus vulgaris* L., Fabaceae) belong to the legume family and are a crop of great agricultural importance due to their high nutritional value and dietary significance. They originated in the Americas, with evidence of domestication dating back to 9000 and 5000 B.C.¹. Today, common beans remain a primary source of nutrients in several countries, providing proteins, carbohydrates, B-complex vitamins, and minerals^{2,3}. They are the most critical legumes for direct human consumption worldwide⁴ highlighting their crucial role in global nutrition.

Studies have highlighted the diversity of bean proteins, including globulins, phaseolins, and arcelins, which play crucial roles in nutritional quality, taste, and pest resistance⁴.

Globulins contribute to beans' nutritional and functional properties, playing roles in seed germination and environmental responses⁵. These proteins are classified into two types based on their sedimentation coefficients: 7S (vicilins) and 11S (legumins). Vicilins have molecular weights ranging from 150 to 190 kDa, while legumins range between 300 and 400 kDa. Each legumin subunit consists of an acidic polypeptide (α -chain) with a molecular mass of 30–40 kDa and a basic polypeptide (β -chain) of 18–22 kDa^{6,7}.

Phaseolins, a type of storage globulins, are the most abundant bean proteins, making up to 50% of total bean proteins. They significantly influence texture and water absorption during cooking, directly affecting bean softness and flavor⁸. Additionally, they provide nutritional benefits, including antioxidant, antigenotoxic, and chemopreventive properties. Their potential role in cholesterol reduction has also

been suggested ^{9,10}. These proteins are trimeric, high mannose, glycosylated proteins of about 150 kDa, containing almost identical monomers with subunits showing molecular masses between 40 and 53 kDa ^{3,11–15}.

Arcelins are protease inhibitors related to lectins. Lectins are carbohydrate-binding proteins that recognize and bind glycans in glycoproteins, glycolipids, or polysaccharides with high affinity ^{16,17}; while arcelins are homologous to lectins, they display a distinct specificity toward complex sugars.

Among the challenges faced by bean crops, one of the most critical is the predation of dry seeds by post-harvest pests, particularly bruchids (Chrysomelidae). Bruchids are considered the most destructive storage pest of pulse crops in tropical and subtropical regions. Infestation begins in the field and rapidly increases during storage, causing severe seed damage. Within two to three months, infestation levels can reach 60% to 100%, and yield losses may exceed half of the expected production in a short period ¹⁸. These insects are characterized by their ability to infest seeds of leguminous plants, especially beans, lentils, and chickpeas ^{18–20}.

Some bean landraces naturally resist bruchid pests, such as *Zabrotes subfasciatus*, due to the presence of arcelins ^{21,22}. Due to their unique ability to bind glycoproteins with high specificity for complex sugars, arcelins develop an effective toxicity mechanism against bruchids. This contrasts with true lectins, which primarily exhibit affinity for monosaccharides ^{23–25}.

This study employs a proteomic approach to characterize proteins in three commercially available *P. vulgaris* L. landraces from Costa Rica. Due to artificial selection processes for crop improvement and enhanced palatability, bean varieties have undergone extensive selective breeding. Consequently, a wild legume, pigeon pea (*Cajanus cajan* (L.) Millsp., Fabaceae), was used as a reference control legume (wild type) in this study's analysis. This research focuses on optimizing methodologies for protein extraction, fractionation, and analysis in legumes, leading to the isolation of the B albumin fraction, which includes arcelins, phytohemagglutinin (PHA), and α -amylase inhibitors.

Results

Protein fraction separation

Figure 1 presents the results of the protein fractionation processes performed on the legume species analyzed in this study. The figure displays protein bands under denaturing electrophoresis conditions for each stage of the process, from the initial extraction of the albumin fraction to the final step involving the isolation of globulins from the final precipitate.

Figure 1. Polyacrylamide gel electrophoresis of albumin and globulin fraction samples obtained from bean flour extraction. Samples: A) Pre-stained molecular weight marker; B) Phytohemagglutinin-M (PHA-M) from *P. vulgaris* (Sigma Aldrich); C) Albumin fraction from Mantequilla bean; D) Albumin fraction from

Brunca bean; E) Albumin fraction from Pigeon pea bean; F) Albumin fraction from Surú bean; G) Globulin fraction from Mantequilla bean; H) Globulin fraction from Brunca bean; I) Globulin fraction from Surú bean; J) Globulin fraction from Pigeon pea bean. The photograph was taken under white light with a multicolor filter in a transilluminator. The program software automatically generated green lines.

Albumin fractions

Lines C to F correspond to the albumin fractions, showing similar electrophoretic patterns across all varieties. Notably, in all *P. vulgaris* samples (lanes C. Mantequilla, D. Brunca, and F. Surú), a prominent double band appears around 29–31 kDa, probably corresponding to the phytohemagglutinin-M standard (a lectin-type protein) from *P. vulgaris* (lane B). It has been reported that in the Mexican bean variety Bayo Berrendo, the most intense band is found at the 29.3 kDa marker level ³.

It is important to note that the PHA used as a reference is not of primary standard grade; instead, it is a bean-derived mucoprotein isolated through chromatographic techniques and composed of five isoforms. When subjected to the denaturing conditions of electrophoresis, these isoforms can be separated into subunits, resulting in bands of different molecular weights that consistently exhibit the same separation pattern ²⁶.

In contrast, the *C. cajan* lane (lane E Pigeon Pea bean) does not exhibit this double band but instead shows a very low quantity of these types of proteins, revealing a faint single band near 30 kDa (29.8 kDa) and some lower molecular weight bands below 20 kDa. These findings may be explained by their classification in a different genus within the Fabaceae family. They also align with those reported²⁷ for most pigeon pea albumins, which have molecular weights ranging from 9 to 15 kDa. It is important to note that the identity of these proteins differs from those in the *Phaseolus* genus; however, they are still referred to as albumins-1 and 2S.

It is also remarkable that Mantequilla and Brunca landraces, separated in lanes C and D, exhibit molecular weights ranging from 33.8 to 26.7 kDa, which align with the reported values for arcelins and PHA ²⁴. Additionally, these bands correspond to those observed in the reference PHA placed in lane B. In contrast, Surú (lane F) lacks the upper 33.8 kDa band observed in the other two *P. vulgaris* landraces (Mantequilla and Brunca).

Other significant bands are also observed within the albumin fractions' 9–15 kDa range. These bands also seem to correspond to the subunits from the phytohemagglutinin-M standard. In this context, *P. vulgaris* landraces display similar electrophoretic profiles, while *C. cajan* slightly deviates from this proteomic pattern.

Globulin fractions

Lines D to J belong to globulin fractions present in the bean varieties (Fig. 1), with a predominant double band around 40 to 45 kDa in most cases. This result could be related to the disassembly subunits of legumins (around 40 kDa) and subunits of phaseolins (around 45 kDa). These weight ranges correspond to the most abundant globulins reported in the literature and observed in the three common bean varieties^{3,11–15}. This information is further supported by proteomic results (Table 1), which identify distinctive proteins near 43–45 kDa present in all three common bean varieties, most of which contain fragments consistent with the molecular weights of α -, β -, and γ -phaseolin subunits.

Notably, *C. cajan* exhibits a distinct pattern, appearing as a double band with higher mass values from 43–47 kDa and a third important band around 59 kDa, probably associated with legumins and phaseolins subunits. This variation from globulins from *P. vulgaris* landraces and *C. cajan* may be related to their taxonomic separation. Literature^{27–29} states that the most abundant proteins in pigeon peas, with molecular weights ranging from 64 to 47 kDa, correspond to 7S globulin subunits of the vicilin type, where the 64 kDa β -conglycinin is one of them. Therefore, the 62.4 kDa fragment is undoubtedly this protein, as confirmed by the proteomic sequencing (Table 1), which shows a predominant result for this band with 70% sequence coverage, 110 unique peptides out of 117 corresponding to the analyzed fragment of the α -chain of pigeon pea β -conglycinin. Furthermore, based on the extraction method, these proteins are expected to be found as part of the insoluble globulin fraction.

High-Performance Thin Layer Chromatography (HPTLC)

Figure 2 presents the HPTLC results for the albumin and globulin fractions obtained from the *P. vulgaris* landraces (Surú) and *Cajanus cajan* (pigeon pea), and Figure 3 shows the results of *P. vulgaris* (Brunca and Mantequilla). In all cases, each lane represents a distinct extraction stage, from crude extract to dialyzed fractions, allowing the assessment of protein mobility and retention characteristics on the silica stationary phase.

Figure 2. HPTLC separation of albumin and globulin fractions from *P. vulgaris* (Surú) and *Cajanus cajan* (Pigeon pea) using 2-butanol/water/acetic acid (13:4:1 v/v/v) as mobile phase, derivatization with fluorescamine, and photographed under UV light at 366 nm. Samples: lane 1 control blank NaCl (10 mM, pH 2.4), lane 2 phytohemagglutinin, lane 3 crude extract of Surú bean, lane 4 extraction supernatant 1 Surú, lane 5 extraction supernatant 2 Surú, lane 6 mixture of Surú bean supernatants, lane 7 Surú albumin fraction, lane 8 Surú globulin fraction, lane 9 crude extract of Pigeon pea bean (*Cajanus cajan*), lane 10 extraction supernatant 1 Pigeon pea, lane 11 extraction supernatant 2 Pigeon pea, lane 12 mixture of Pigeon pea bean supernatants, lane 13 Pigeon pea albumin fraction, lane 14 Pigeon pea globulin fraction.

Figure 3. HPTLC separation of albumin and globulin fractions from *P. vulgaris* (Brunca and Mantequilla) using 2-butanol/water/acetic acid (13:4:1 v/v/v) as mobile phase, derivatization with fluorescamine, and photographed under UV light at 366 nm. Samples: Lane 1 solvent blank NaCl (10 mM, pH 2.4), lane 2 phytohemagglutinin, lane 3 crude extract of Brunca, lane 4 extraction supernatant 1 Brunca, lane 5

extraction supernatant 2 Brunca, lane 6 mixture of Brunca bean supernatants, lane 7 Brunca albumin fraction, lane 8 Brunca globulin fraction, lane 9 crude extract of Mantequilla bean, lane 10 extraction supernatant 1 Mantequilla, lane 11 extraction supernatant 2 Mantequilla, lane 12 mixture of Mantequilla supernatants, lane 13 Mantequilla albumin fraction, and lane 14 Mantequilla globulin fraction.

Separation patterns and Rf values

The obtained chromatograms revealed significant protein separation differences among the bean varieties. A clear banding pattern was observed in all *P. vulgaris* varieties, with a consistent retention factor (Rf) of ~ 0.1 in the albumin fractions, suggesting the presence of phytohemagglutinin (PHA) subunits. This agrees with the electrophoretic results, where a dominant band was observed at $\sim 29\text{--}31$ kDa, aligning with reported PHA molecular weights.

A distinct pattern was observed for the pigeon pea (*C. cajan*) samples, characterized by a dominant Rf of 0.2 in the albumin fractions and additional bands at Rf = 0.31 in the crude extract and first extraction step. These findings correlate with its electrophoretic profile, where lower molecular weight albumins (9–15 kDa) predominated. Unlike *P. vulgaris*, pigeon peas lacked strong retention at Rf = 0.1, indicating the absence of significant PHA content, which aligns with literature reports.

All *P. vulgaris* varieties exhibited a significant band at Rf ~ 0.15 in the globulin fractions, which intensified following the second extraction step. This suggests improved solubilization of specific protein components using NaCl-Gly buffer (10–50 mM, pH 2.4). In contrast, pigeon pea globulins showed stronger retention at Rf = 0.2–0.25, consistent with the presence of vicilin-type globulins (7S) and β -conglycinin (64 kDa), as identified in the proteomic analysis.

Protein solubility and extraction efficiency

The chromatographic results confirm that the second extraction step (NaCl-Gly buffer) enhances protein separation, as evidenced by the increased intensity of albumin fractions and the shift in globulin migration patterns. This is particularly noticeable in *P. vulgaris*, where the combination of low-salt and glycine buffering enhanced albumin recovery, resulting in a more defined separation of protein bands.

Additionally, the migration patterns observed in HPTLC correlate well with the electrophoretic results, reinforcing the separation of albumins and globulins across different fractions. In the case of pigeon peas, the distinct Rf values observed for globulins suggest a unique protein composition compared to *P. vulgaris*, supporting the hypothesis that its storage proteins are predominantly composed of vicilin-like globulins rather than phaseolins.

Interpretation of fluorescamine staining

Fluorescamine derivatization allowed for the selective detection of primary amine-containing proteins, with fluorescence intensity correlating with protein abundance. The strongest fluorescent signals were observed in the albumin fractions of *P. vulgaris* ($R_f = 0.1$) and globulin fractions of pigeon pea ($R_f = 0.2-0.25$), indicating high concentrations of amino-rich proteins. This aligns with prior findings on PHA subunits in *P. vulgaris* and β -conglycinin in pigeon peas.

Interestingly, increased fluorescence intensity was detected in the second extraction step (NaCl-Gly buffer) across all varieties, particularly in *P. vulgaris*, further validating the improved solubility of proteins under these conditions. In contrast, in pigeon peas, fluorescence was strongest in the globulin fractions, suggesting a higher proportion of globulin-derived peptides contributing to the overall fluorescence signal.

Comparative analysis of HPTLC with SDS-PAGE

When comparing HPTLC results with SDS-PAGE, a strong correlation was observed between the retention of albumin fractions at $R_f = 0.1$ in *P. vulgaris* and 29–31 kDa bands in electrophoresis. Similarly, globulin fractions of pigeon peas showed retention at $R_f = 0.2-0.25$, consistent with 7S vicilin globulins detected in SDS-PAGE (~ 45–47 kDa bands).

This confirms the complementary nature of both techniques, as HPTLC provided a rapid screening tool for protein separation, while SDS-PAGE allowed for molecular weight determination. The observed retention factors in HPTLC indirectly validate protein composition, reinforcing the presence of PHA in *P. vulgaris* and vicilin-type globulins in *C. cajan*.

The results demonstrate that HPTLC is a viable technique for the preliminary analysis of bean protein fractions, particularly for distinguishing albumin and globulin fractions across different extraction steps. While SDS-PAGE remains the standard for molecular weight characterization, HPTLC offers speed, cost-effectiveness, and advantages in simultaneous multi-sample analysis.

These findings highlight the distinct protein composition between *P. vulgaris* and *C. cajan*, suggesting that the dominance of vicilin-type globulins compensates for the absence of phaseolins in pigeon peas. Additionally, the differential fluorescence intensities observed across extraction steps validate the efficiency of NaCl-Gly buffer in protein solubilization, particularly for albumin-rich fractions in *P. vulgaris*.

Further studies incorporating advanced proteomic techniques (LC-MS/MS) could provide deeper insights into the structural differences between these protein fractions and their functional properties in legume-based diets and pest resistance.

Protein Sequencing by Mass Spectrometry

Selection and Cutting of Electrophoretic Bands

Specific protein bands were selected based on their molecular weight and intensity, ensuring that significant and minor protein fractions were represented. Since electrophoretic results suggested phaseolin (43–45 kDa) and β -conglycinin (64 kDa) as major proteins in *P. vulgaris* and *C. cajan*, respectively, these bands were prioritized for proteomic sequencing.

Additionally, bands around 29–31 kDa, which likely correspond to phytohemagglutinins (PHA) and possible arcelin isoforms, were also included in the analysis to determine their protein identity. For the pigeon pea (*C. cajan*), a distinct band around 47–59 kDa was selected, aligning with the molecular weights of vicilin-type globulins (7S). These findings correlate with chromatographic migration patterns observed in HPTLC.

Proteolytic Digestion in Solution and Identification by nLC-MS/MS

The analysis confirmed the presence of significant storage and functional proteins in the studied legume species.

Table 1
Major Proteins Identified in Bean Species

Bean species	Major protein identified	Findings
<i>P. vulgaris</i> (Brunca, Surú, Mantequilla)	Phaseolin (α, β, γ subunits) (43–45 kDa) Phytohemagglutinin (PHA) isoforms (29–31 kDa) Potential arcelin homologs (30–35 kDa) Glycoside hydrolase family 19 (27.7 kDa) Albumin-2 (27.7 kDa) Alpha-amylase inhibitor-like protein (27.7 kDa)	As observed in SDS-PAGE results, phaseolins (43–45 kDa) were confirmed as the dominant storage proteins in <i>P. vulgaris</i> L. landraces (see section 3.1.2). Additionally, albumins and enzyme inhibitors were detected, suggesting a functional role in plant defense and metabolism.
<i>Cajanus cajan</i> (Pigeon Pea)	Vicilin-type globulins (47–59 kDa) β-Conglycinin α-chain (62.4 kDa) β-Conglycinin β-chain (45.5–47.4 kDa) Glycinin G3 (34.3 kDa) Legumin-type globulins (35 kDa)	Vicilin-type globulins (47–59 kDa) and β -conglycinin (64 kDa) were confirmed as the predominant storage proteins in <i>C. cajan</i> . Legumin-type globulins and glycinin G3 align with reported proteomic profiles in pigeon peas. This species did not exhibit phaseolin bands.

The predominance of vicilin-like and legumin-like proteins in *C. cajan* is consistent with prior proteomic studies and correlates with retention factors observed in HPTLC. Some low-molecular-weight proteins (~ 10–15 kDa) were also detected, which may correspond to smaller subunits of globulins or albumins. Further analysis is needed to determine their structural significance and potential biological function.

N -terminal Sequencing

N-terminal sequencing was performed using Edman degradation to complement the mass spectrometry data. Results revealed sequence homology between phytohemagglutinins (PHA) and previously reported lectin families, reinforcing their classification as carbohydrate-binding proteins. In contrast, phaseolins and vicilins exhibited conserved N-terminal motifs characteristic of legume storage proteins.

Discussion

The proteomic characterization of three *P. vulgaris* landraces (Brunca, Surú, and Mantequilla) and their comparison with *C. cajan* (pigeon pea) revealed significant differences in the composition of their protein fractions, particularly in globulins and albumins. The results obtained through SDS-PAGE and HPTLC demonstrated that the dominant storage proteins in *P. vulgaris* L. are phaseolins (43–45 kDa), while in *C. cajan*, vicilins (47–59 kDa) and β -conglycinin (64 kDa) predominate, which is consistent with previous studies on the structure and function of legume proteins^{6–8}. These differences may be linked to evolutionary adaptations, domestication processes, and selective breeding for human consumption, which have influenced the stability and functionality of their storage proteins. The absence of arcelins is likely a consequence of these processes, possibly due to their impact on taste and potential toxicity to humans^{30–32}.

The comparison between electrophoretic and chromatographic profiles provided evidence of functional proteins such as arcelins and phytohemagglutinins (PHA) in *P. vulgaris* L., whereas *C. cajan* exhibited a distinctive pattern characterized by the abundance of vicilins and the presence of glycinin, suggesting an alternative storage protein mechanism. Identifying proteins with potential antimetabolic activity, such as α -amylase inhibitors and other proline-rich proteins, supports the hypothesis that some of these landraces may possess bioactive properties with applications in pest resistance and metabolic regulation in mammals³³. Arcelins have been widely recognized for their insecticidal activity, playing a crucial role in plant defense mechanisms against storage pests like *Zabrotes subfasciatus*³⁴. The detection of a strong fluorescent signal in the albumin fractions derivatized with fluorescamine also suggests a higher presence of proteins with reactive amine groups, which aligns with the predominance of phytohemagglutinins in *P. vulgaris* L. crude extracts³⁵.

The combined use of SDS-PAGE, HPTLC, and mass spectrometry (LC-MS/MS) confirmed these proteins' identity and validated their solubility and stability under different extraction conditions. Solubility analyses revealed that extraction with NaCl-Gly buffer (10–50 mM, pH 2.4) enhanced albumin recovery, suggesting that these proteins exhibit specific interactions with salts in low ionic strength solutions³⁶. This has implications for digestibility and bioavailability in food matrices, where solubility can influence protein absorption and functionality³⁷. Notably, the absence of phaseolins in *C. cajan* and the dominance of vicilins and glycinins reinforce the hypothesis that storage proteins in this species fulfill different structural and metabolic functions compared to *P. vulgaris* L.²⁷.

These findings expand knowledge on proteomic diversity in agriculturally significant legumes and have biotechnological and nutritional implications. Characterizing proteins with enzyme inhibitory activity and potential bioactivity suggests possible applications in developing pest-resistant crops and functional food formulations with potential effects on metabolic regulation and gut health.

Conclusions

The proteomic characterization of three *P. vulgaris* L. landraces (Brunca, Surú, and Mantequilla) and their comparison with *C. cajan* revealed significant differences in protein composition. *P. vulgaris* L. was dominated by phaseolins (43–45 kDa), while *C. cajan* exhibited vicilins (47–59 kDa) and β -conglycinin (64 kDa), highlighting their taxonomic and functional divergence due to evolutionary adaptation and domestication.

Electrophoretic and chromatographic analyses identified functional proteins such as arcelins and phytohemagglutinins (PHA) in *P. vulgaris* L., whereas *C. cajan* displayed a distinct profile rich in vicilins and glycinins, suggesting an alternative storage mechanism. The presence of α -amylase inhibitors and proline-rich proteins indicates bioactive properties relevant to pest resistance and metabolic regulation.

SDS-PAGE, HPTLC, and LC-MS/MS confirmed protein identity and validated solubility under different extraction conditions. The NaCl-Gly buffer (10–50 mM, pH 2.4) effectively improved albumin solubility, suggesting interactions with salts in low ionic strength environments.

The absence of phaseolins in *C. cajan* and the dominance of vicilins and glycinins support distinct structural and metabolic roles for storage proteins compared to *P. vulgaris* L., likely influenced by evolutionary and selective breeding processes. This biochemical characterization underscores the nutritional significance of Costa Rican *P. vulgaris* landraces and provides insights into their functional properties and potential insect resistance, contributing to informed cultivation strategies, food security policies, and crop improvement programs.

Methods

Materials

Bovine serum albumin, Coomassie blue G-250, sodium chloride, sodium dibasic phosphate, and a fetuin-agarose column were purchased from Sigma Aldrich (St. Louis, MO). Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and running buffer were purchased from Invitrogen Bis-Tris-plus polyacrylamide gels (Thermo Fisher Scientific).

Protein visualization was achieved by staining the gels with Coomassie Brilliant Blue G-250 (0.1% w/v) in a solution of methanol, acetic acid, and distilled water (40:10:100). Pre-stained molecular weight marker (Thermo Scientific #26616), other chemicals were reagent grade.

Accessing biochemical resources

Access to genetic and biochemical resources in Costa Rica was conducted under the Costa Rican Biodiversity Law ³⁸ and the Biological Diversity Convention ³⁹. In this context, the Costa Rican National Board for Biodiversity Management (CONAGEBIO's initials in Spanish) granted a biodiversity prospecting permit number R-CM-UNA-004-2021-OT compliant with national regulations ⁴⁰.

Mature seeds from three common bean (*P. vulgaris*) varieties were obtained from the germplasm bank of the Fabio Baudrit Moreno Experimental Station (EEFB) at Universidad de Costa Rica. The cultivation and harvesting of these varieties were carried out at the same facility under controlled conditions. The selected bean varieties included black beans (Brunca), white beans (Surú), yellow beans (Mantequilla), and a wild-type bean pigeon pea (*Cajanus cajan*, Fabaceae).

Sample preparation, extraction, and protein fraction separation

Soluble protein fraction extraction

Dry bean seeds were milled using a blade mill with a 32-mesh sieve (500 µm pore size). 10 g of bean flour was extracted with 100 mL of 10 mM sodium chloride (pH 2,4) in a 250 mL beaker. The suspension was stirred at room temperature using a magnetic stirrer at 500 rpm for 1 hour. Following extraction, the mixture was centrifuged at 14,950 g for 30 minutes at 4°C. The supernatant was collected, labeled as supernatant 1, and stored at -80°C for further analysis. The precipitate was labeled precipitate 1 and subjected to the second extraction cycle under the same conditions. Both supernatants were mixed and stored at -80°C for further analysis ^{41,42}.

Separation of soluble proteins by salting out

The supernatant mixture obtained in the previous extraction was subjected to dialysis using membranes with a 3–12 kDa molecular weight cutoff. The membranes containing the samples were dialyzed using distilled water as a solvent. Dialysis was performed at 4°C, and water was changed at 3-hour intervals. At least five dialysis cycles were performed to ensure efficient removal of small solutes. Dialysis effectiveness was verified using a chloride test with 1% silver nitrate over the membrane surrounding water. The dialysis was considered complete when the tested water no longer showed the presence of precipitate or turbidity, indicating the complete removal of small solutes.

During the salting-out process, a precipitate, globulin A, was present in the dialysis tube, caused by the low ionic environment. Conversely, Albumin A and other proteins remained soluble even at very low saline concentrations. The albumin A supernatant was collected after centrifugation of the dialysis bag contents. A 1 mL aliquot of this fraction was stored in 2 mL Eppendorf tubes and kept at -80°C. The precipitate obtained from this process was identified as globulin A. The precipitate and the supernatant were lyophilized separately for conservation and further electrophoretic analyses

Protein quantification and SDS PAGE

Protein quantification was performed using the Bradford assay ⁴³ in 96-well microplates, with all samples analyzed in triplicate. A bovine serum albumin standard curve was used to determine protein concentrations at 595 nm.

Sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) was performed following the protocol established by Laemmli ⁴⁴ as cited in Carrasco *et al.* ⁴⁵ The gels consisted of a 4% stacking layer and a 12% separating layer. Electrophoresis was conducted in a Thermo Scientific™ Complete Owl™ Vertical Double Gel Electrophoresis System.

Lyophilized samples were dissolved in a glycine buffer (0.05 M) – NaCl (0.01 M), pH 2.4, considering the weight of the solid. For example, 10 g of solid was dissolved in 200 mL of buffer, while 0.1 g of globulin A precipitate was dissolved in 2 mL of buffer. The redissolved precipitates were vortexed and centrifuged for 10 min at 15,000 rpm and 4°C.

The samples were mixed in equal parts with Laemmli buffer ⁴⁴ (loading buffer) and heated for five minutes at 100°C, followed by a brief centrifugation pulse for 15 seconds at 15,000 g. The samples were loaded into a commercially prepared polyacrylamide gel with a 4–20% concentration gradient.

A running buffer (Invitrogen) was used for protein separation ⁴⁵. After electrophoresis, the polyacrylamide gel was stained with a solution containing 0.25% Coomassie Brilliant Blue R-250, 7.5% acetic acid, and 50% methanol at room temperature for one hour. Destaining was performed with a solution composed of 10% acetic acid and 40% methanol ^{45,46} for approximately 45 minutes, with time variations depending on the specific case. Once destained, the gels were photographed using a vertical double gel electrophoresis photodocumentation system (Thermo Scientific™ Complete Owl™). Once photographed, the software determined each band's approximate molecular mass (MM) value.

High-performance thin-layer chromatography (HPTLC)

High-performance thin-layer chromatography (HPTLC) was developed as an analytical tool to assess the separation patterns of soluble proteins extracted from the studied bean varieties. This technique allowed the comparison of albumin and globulin fractions across different extraction steps. Protein standards were employed to optimize HPTLC methodology, and fluorescamine was used as a derivatization agent before visualization under 366 nm UV light.

The high-performance thin-layer chromatography (HPTLC) was performed using a CAMAG system composed of independent modules. The method was modified by Morschheuser *et al.* ⁴⁷. Silica gel and C18 plates were used as stationary phases, with cellulose plates optionally included. Silica gel plates were used as the stationary phase. They were pretreated by washing with methanol in the development chamber, followed by activation at 100°C for 10 minutes using a TLC Plate Heater (CAMAG).

Samples were applied to a 20 × 10 cm Silica gel 60 F₂₅₄ glass-backed plate using an Automatic TLC Sampler 4. Different sample volumes were applied to ensure a constant protein mass of 0.7 µg per band

(8 mm long) with minimal width and a 10.5 mm spacing between bands. The migration area was set to 80 mm, with the first sample applied 20 mm from the right edge (X-axis) and 10 mm from the bottom edge (Y-axis). The mobile phase consisted of 2-butanol/water/acetic acid (13:4:1 v/v/v) and was developed using an Automatic Development Chamber - ADC2 with a total mobile phase volume of 35 mL.

After development, derivatization was performed using 3 mL of fluorescamine in acetone (0.05% w/v) via a spraying derivatization chamber (Derivatizer) with a low-viscosity solvent nozzle for 4 minutes at a medium spray flow. The plate was then left to dry at room temperature for 10 minutes before visualization under UV light (336 nm) using a TLC Visualizer⁴⁷.

Protein Sequencing by Mass Spectrometry

Protein sequencing was conducted by the Proteomics Laboratory of the Clodomiro Picado Institute, University of Costa Rica. The analysis involved three stages:

Selection and Cutting of Electrophoretic Bands

Samples required a minimum protein content of 1–10 µg, either purified or semi-purified, and free of salts and detergents to avoid interference with peptide/protein ionization. Samples could be dry or dissolved in MS-compatible solvents such as water or acetonitrile-water mixtures.

Based on these criteria, medium- to high-intensity bands were selected from polyacrylamide gels after electrophoresis. Selected bands were carefully excised using a clean scalpel, minimizing excess polyacrylamide around the protein area. The bands were placed in new 1.5–2 mL Eppendorf tubes and stored at 4°C until analysis.

Proteolytic Digestion in Solution and Identification by nLC-MS/MS

Proteins excised from the gels underwent enzymatic digestion using trypsin. The sample preferably contained 15–20 µg of protein, free of salts and detergents, either dried or dissolved in water or water/acetonitrile mixtures. The digestion process was conducted using an automated DigestPro MSi (Intavis) system, ensuring precise control over incubation times and reagent volumes. Once digestion was complete, the resulting peptide mixtures were purified and concentrated before injection into a nano-liquid chromatography system (nLC) coupled with a Thermo Q-Exactive Plus mass spectrometer. The high-resolution nLC-MS/MS approach allowed the identification of proteins based on their fragmentation patterns and peptide sequences.

N -terminal Sequencing

N-terminal sequencing was conducted using a Shimadzu Protein Sequencer PPSQ-33A through Edman degradation, which sequences proteins from the N-terminal end (unless blocked). The minimum protein

purity required for sequencing was $\geq 90\%$, with a sample amount of at least one nanomole. The mass range for analyzed proteins/peptides extended up to 100,000 Da. The obtained sequence data were identified by comparison against publicly available protein databases.

Declarations

Funding sources

This work was supported by the Fondo Institucional para el Desarrollo Académico (FIDA), Universidad Nacional de Costa Rica Institutional Grant SIA 585–2019. The funding source was not involved in the study design; collection, analysis, and interpretation of data; writing of the manuscript; or decision to submit the article for publication.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this manuscript, the authors used ChatGPT (OpenAI; models “Scholar GPT” and “o3 reasoning advanced,” accessed March–April 2025) to enhance the readability, grammar, and overall flow of the text and to refine the organization of specific sections (e.g., Introduction and Discussion). After employing this tool, the authors thoroughly reviewed, edited, and revised all content as needed and take full responsibility for the integrity and accuracy of the published article.

Author Contribution

K.R-Q. conducted all experiments and analyses presented in the manuscript, including protein extraction, electrophoresis, HPTLC, and sample preparation for mass spectrometry, and contributed to data interpretation. Y.A-U. provided close technical supervision and training to K.R-Q., assisted in method development for protein analysis, and contributed to sample processing. H.T-V. and A.F.C-A. contributed intellectually through critical review of experimental data, analysis of results, and strategic input throughout the project. M.S.B. led the data interpretation and was primarily responsible for drafting the initial version of the manuscript. L.R.V-P., as principal investigator, conceived and coordinated the project, designed the research strategy, supervised the overall execution, and contributed to manuscript writing, critical revision, and final adaptation for publication. All authors reviewed and approved the final manuscript.

Acknowledgement

We thank the Proteomics Laboratory of the Clodomiro Picado Institute, University of Costa Rica, for conducting the mass spectrometry analysis and protein sequencing.

Data Availability

The datasets generated and analysed during the current study are available in the ProteomeXchange Consortium via the MassIVE repository with the dataset identifier PXD076287.<http://massive.ucsd.edu/ProteoSAFe/status.jsp?task=f738cf09262a4a2984bfa45242862600> (reviewer access password: UNA1234)

References

1. Bitocchi, E. et al. Mesoamerican origin of the common bean (*Phaseolus vulgaris* L.) is revealed by sequence data. *Proc. Natl. Acad. Sci. U S A.* **109**, E788–E796 (2012).
2. Guzmán-Maldonado, H. S. J. A.-G. y O. P.-López. Protein and mineral content of a novel collection of wild and weedy common bean (*Phaseolus vulgaris* L.) - Dialnet. *J Sci. Food Agric* 1874–1881. (2000).
3. Raya-Pérez, J. C., Gutiérrez-Benicio, G. M., Ramírez-Pimentel, J. G., Covarrubias-Prieto, J. & Aguirre-Mancilla, C. L. Characterization of proteins and mineral content of two bean landraces from Mexico. *Agronomía Mesoamericana.* **25**, 1–11 (2014).
4. Campos-Vega, R., Bassinello, P. Z., de Santiago, R. & Oomah, B. D. A. C. Dry Beans: Processing and Nutritional Effects. *Therapeutic, Probiotic, and Unconventional Foods* 367–386 (2018).
10.1016/B978-0-12-814625-5.00019-4
5. Gallardo, K., Thompson, R. & Burstin, J. Reserve accumulation in legume seeds. *C R Biol.* **331**, 755–762 (2008).
6. Martínez Muñoz, P. *Estudio de las interacciones entre la vicilina y las lectinas con a y cel-II de la semilla de Canavalia ensiformis* (Universidad Nacional de Colombia, 2009).
7. Pernollet, J. C. & Mossé, J. Structure and location of legume and cereal seed storage proteins. in *Seed Proteins* (eds. Daussant, J., Mossé, J. & Vaughan, J.) 155–191 Academic Press., New York., (1983).
8. Broughton, W. J. et al. Beans (*Phaseolus* spp.) - Model food legumes. *Plant. Soil.* **252**, 55–128 (2003).
9. Rodríguez, L. et al. Role of *Phaseolus vulgaris* L. in the Prevention of Cardiovascular Diseases—Cardioprotective Potential of Bioactive Compounds. *Plants* 2022. **11**, Page 186 (11), 186 (2022).
10. García-Cordero, J. M. et al. Phaseolin, a Protein from the Seed of *Phaseolus vulgaris*, Has Antioxidant, Antigenotoxic, and Chemopreventive Properties. *Nutrients* 2021. **13**, 1750 (2021).
11. Montoya, C. A., Lallès, J. P., Beebe, S. & Leterme, P. Phaseolin diversity as a possible strategy to improve the nutritional value of common beans (*Phaseolus vulgaris*). *Food Res. Int.* **43**, 443–449 (2010).
12. Montoya, C. A. et al. Susceptibility of Phaseolin to in Vitro Proteolysis Is Highly Variable across Common Bean Varieties (*Phaseolus vulgaris*). *J. Agric. Food Chem.* **56**, 2183–2191 (2008).
13. Sparvoli, F., Daminati, M. G., Lioi, L. & Bollini, R. In vivo endoproteolytically cleaved phaseolin is stable and accumulates in developing *Phaseolus lunatus* L. seeds. *Biochim. et Biophys. Acta (BBA) - Protein Struct. Mol. Enzymol.* **1292**, 15–22 (1996).

14. Rui, X. et al. Enrichment of ACE inhibitory peptides in navy bean (*Phaseolus vulgaris*) using lactic acid bacteria. *Food Funct.* **6**, 622–629 (2015).
15. Yeboah, F. K., Alli, I., Simpson, B. K., Konishi, Y. & Gibbs, B. F. Tryptic fragments of phaseolin from protein isolates of *Phaseolus* beans. *Food Chem.* **67**, 105–112 (1999).
16. Goldstein, I. J. & Hayes, C. E. The Lectins: Carbohydrate-Binding Proteins of Plants and Animals. *Adv. Carbohydr. Chem. Biochem.* **35**, 127–340 (1978).
17. Chrispeels, M. J., Raikhel, N. V. & Lectins Lectin Genes, and Their Role in Plant Defense. *Plant. Cell.* **3**, 1 (1991).
18. Ragul, S. & Manivannan, N. Bruchid a Serious Pest on Pulse Crops: Its Control Measures and Breeding Advancements: A Review. *Agricultural Reviews.* 10.18805/AG.R-2307 (2021).
19. Mishra, S. K., Macedo, M. L. R., Panda, S. K. & Panigrahi, J. Bruchid pest management in pulses: past practices, present status and use of modern breeding tools for development of resistant varieties. *Ann. Appl. Biol.* **172**, 4–19 (2018).
20. Arulselvi, S. et al. Bruchid Resistance in Pulses: A Review. *Agricultural Reviews.* 10.18805/AG.R-2375 (2022).
21. Janarthanan, S., Suresh, P., Radke, G., Morgan, T. D. & Oppert, B. Arcelins from an Indian wild pulse, *Lablab purpureus*, and insecticidal activity in storage pests. *J. Agric. Food Chem.* **56**, 1676–1682 (2008).
22. Razzaq, M. K. et al. Molecular and genetic insights into secondary metabolic regulation underlying insect-pest resistance in legumes. *Funct. Integr. Genomics.* **23**, 1–18 (2023).
23. Gepts, P., Osborn, T. C., Rashka, K. & Bliss, F. A. Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean (*Phaseolus vulgaris*): Evidence for Multiple Centers of Domestication. *Econ. Bot.* **40**, 451–468 (1986).
24. Osborn, T. C., Alexander, D. C., Sun, S. S. M., Cardona, C. & Bliss, F. A. Insecticidal Activity and Lectin Homology of Arcelin Seed Protein. *Sci.* (1979). **240**, 207–210 (1988).
25. Zaugg, I. et al. QUES, a new *Phaseolus vulgaris* genotype resistant to common bean weevils, contains the Arcelin-8 allele coding for new lectin-related variants. *Theor. Appl. Genet.* **126**, 647–661 (2013).
26. Ng, T. B., Sharma, A., Wong, J. H. & Lin, P. Purification and Characterization of a Lectin from *Phaseolus vulgaris* cv. (Anasazi Beans). *Biomed Res Int* 929568 (2009). (2009).
27. Locali-Pereira, A. R. et al. Pigeon Pea, An Emerging Source of Plant-Based Proteins. *ACS Food Sci. Technol.* **3**, 1777–1799 (2023).
28. Boachie, R. T. et al. Enzymatic release of dipeptidyl peptidase-4 inhibitors (gliptins) from pigeon pea (*Cajanus cajan*) nutrient reservoir proteins: In silico and in vitro assessments. *J. Food Biochem.* **43**, e13071 (2019).
29. Krishnan, H. B., Natarajan, S. S., Oehrle, N. W., Garrett, W. M. & Darwish, O. Proteomic Analysis of Pigeonpea (*Cajanus cajan*) Seeds Reveals the Accumulation of Numerous Stress-Related Proteins.

- J. Agric. Food Chem.* **65**, 4572–4581 (2017).
30. Romero Andreas, J., Yandell, B. S. & Bliss, F. A. Bean arcelin – 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. *Theor. Appl. Genet.* **72**, 123–128 (1986).
 31. Parker, T. A., Gepts, P. & Parker, T. A. & Gepts, P. Population Genomics of *Phaseolus* spp.: A Domestication Hotspot. 607–689 (2021). 10.1007/13836_2021_89
 32. Zaugg, I. et al. QUES, a new *Phaseolus vulgaris* genotype resistant to common bean weevils, contains the Arcelin-8 allele coding for new lectin-related variants. *Theor. Appl. Genet.* **126**, 647–661 (2013).
 33. Peddio, S. et al. Purification and Characterization of Proteinaceous Thermostable α -Amylase Inhibitor from Sardinian Common Bean Nieddone Cultivar (*Phaseolus vulgaris* L). *Plants* **13**, 2074 (2024).
 34. Franco, O. L., Rigden, D. J. & Melo, F. R. Grossi-de-Sá, M. F. Plant α -amylase inhibitors and their interaction with insect α -amylases. *Eur. J. Biochem.* **269**, 397–412 (2002).
 35. Peddio, S. & Zucca, P. P. Proteinaceous inhibitors of α - amylase and α -glucosidase from common bean (*Phaseolus vulgaris* L.): biochemical characterization and phylogenetic analysis of Sardinian cultivars. (2023). (Università di Cagliari).
 36. Kornet, R. et al. Coacervation in pea protein solutions: The effect of pH, salt, and fractionation processing steps. *Food Hydrocoll.* **125**, 107379 (2022).
 37. Said, H. H. & Doucette, A. A. Enhanced Electrophoretic Depletion of Sodium Dodecyl Sulfate with Methanol for Membrane Proteome Analysis by Mass Spectrometry. *Proteomes* **12**, 5 (2024).
 38. Asamblea Legislativa de la República de Costa Rica. *Ley de Biodiversidad N°7788. La Gaceta N° 101 del 27 de Mayo de 1998; San José, Costa Rica; p. 40.* 40 (La Gaceta N° 101 del 27 de mayo de 1998, Costa Rica).
doi:http://www.pgrweb.go.cr/scij/Busqueda/Normativa/Normas/nrm_texto_completo.aspx?param2=NRTC&nValor1=1&nValor2=39796&strTipM=TC
 39. The contracting parties. *Convention Biol. Diversity* **31** doi: (1992).
<https://www.cbd.int/doc/legal/cbd-en.pdf>
 40. Costa Rican National Board for Biodiversity Management. (Comisión Nacional de Gestión de la Biodiversidad, C. Granted Bioprojective permits.
<https://www.conagebio.go.cr/Conagebio/public/permisosOtorgados.html>
 41. Gerhardt, I. R. et al. Molecular characterization of a new arcelin-5 gene. *Biochim. et Biophys. Acta (BBA) - Gene Struct. Expression.* **1490**, 87–98 (2000).
 42. Goossens, A., Dillen, W., De Clercq, J., Van Montagu, M. & Angenon, G. The arcelin-5 Gene of *Phaseolus vulgaris* Directs High Seed-Specific Expression in Transgenic *Phaseolus acutifolius* and *Arabidopsis* Plants. *Plant. Physiol.* **120**, 1095–1104 (1999).
 43. Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**, 248–254 (1976).

44. Laemmli, U. K. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. *Nature* 1970 227:5259 227, 680–685 (1970).

45. Carrasco-Castilla, J. et al. Antioxidant and metal chelating activities of Phaseolus vulgaris L. var. Jamapa protein isolates, phaseolin and lectin hydrolysates. *Food Chem.* **131**, 1157–1164 (2012).

46. Karuppiyah, H., Kirubakaran, N. & Sundaram, J. Genetic resources for arcelin, a stored product insect antimetabolic protein from various accessions of pulses of Leguminosae. *Genet. Resour. Crop Evol.* **65**, 79–90 (2018).

47. Morschheuser, L. et al. HPTLC-apstaining – Innovative protein detection system for high-performance thin-layer chromatography. *Scientific Reports* 2016 6:1 6, 1–8 (2016).

Figures

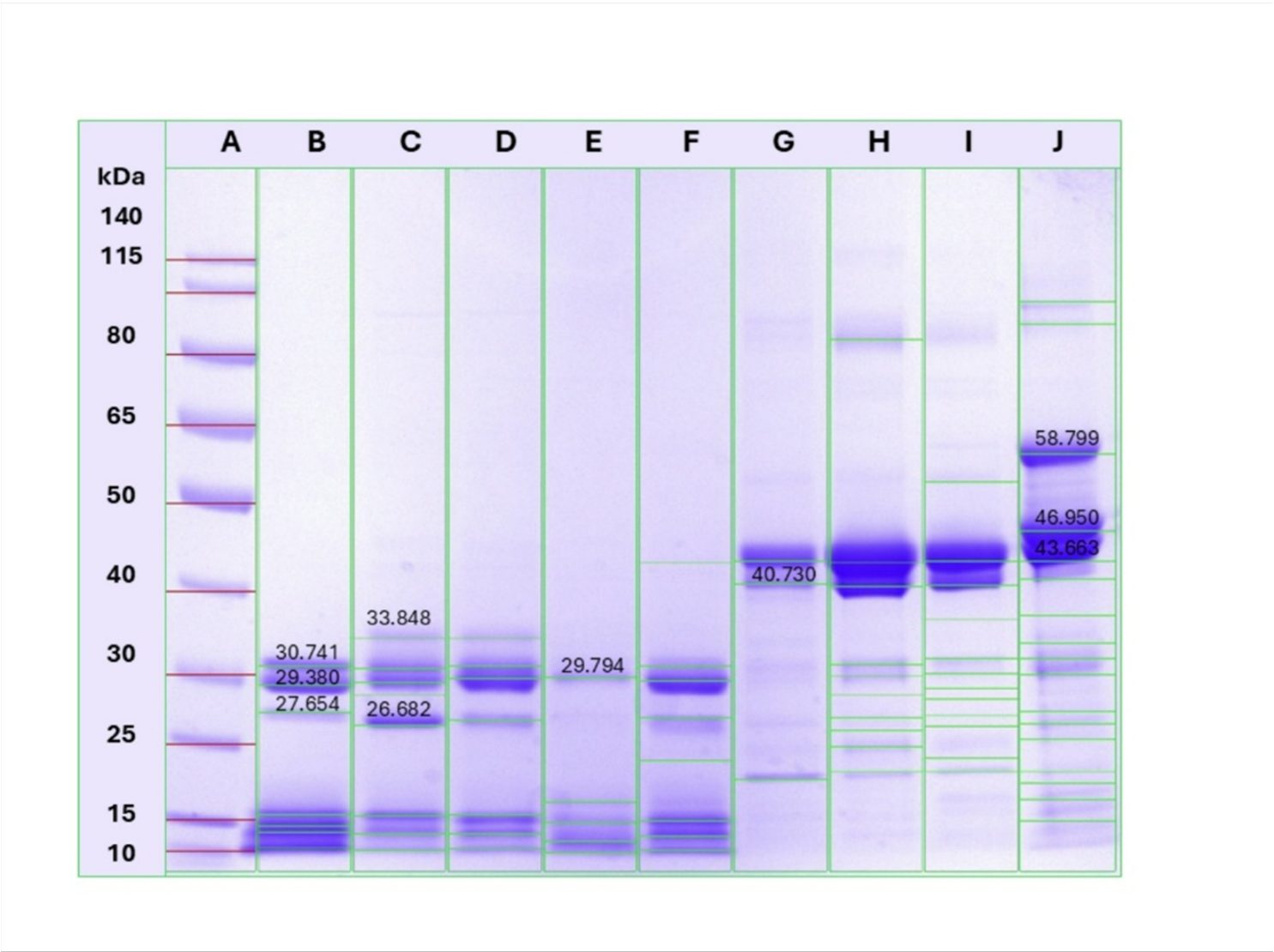


Figure 1

Polyacrylamide gel electrophoresis of albumin and globulin fraction samples obtained from bean flour extraction. Samples: A) Pre-stained molecular weight marker; B) Phytohemagglutinin-M (PHA-M) from *P. vulgaris* (Sigma Aldrich); C) Albumin fraction from Mantequilla bean; D) Albumin fraction from Brunca bean; E) Albumin fraction from Pigeon pea bean; F) Albumin fraction from Surú bean; G) Globulin fraction from Mantequilla bean; H) Globulin fraction from Brunca bean; I) Globulin fraction from Surú bean; J) Globulin fraction from Pigeon pea bean. The photograph was taken under white light with a multicolor filter in a transilluminator. The program software automatically generated green lines.

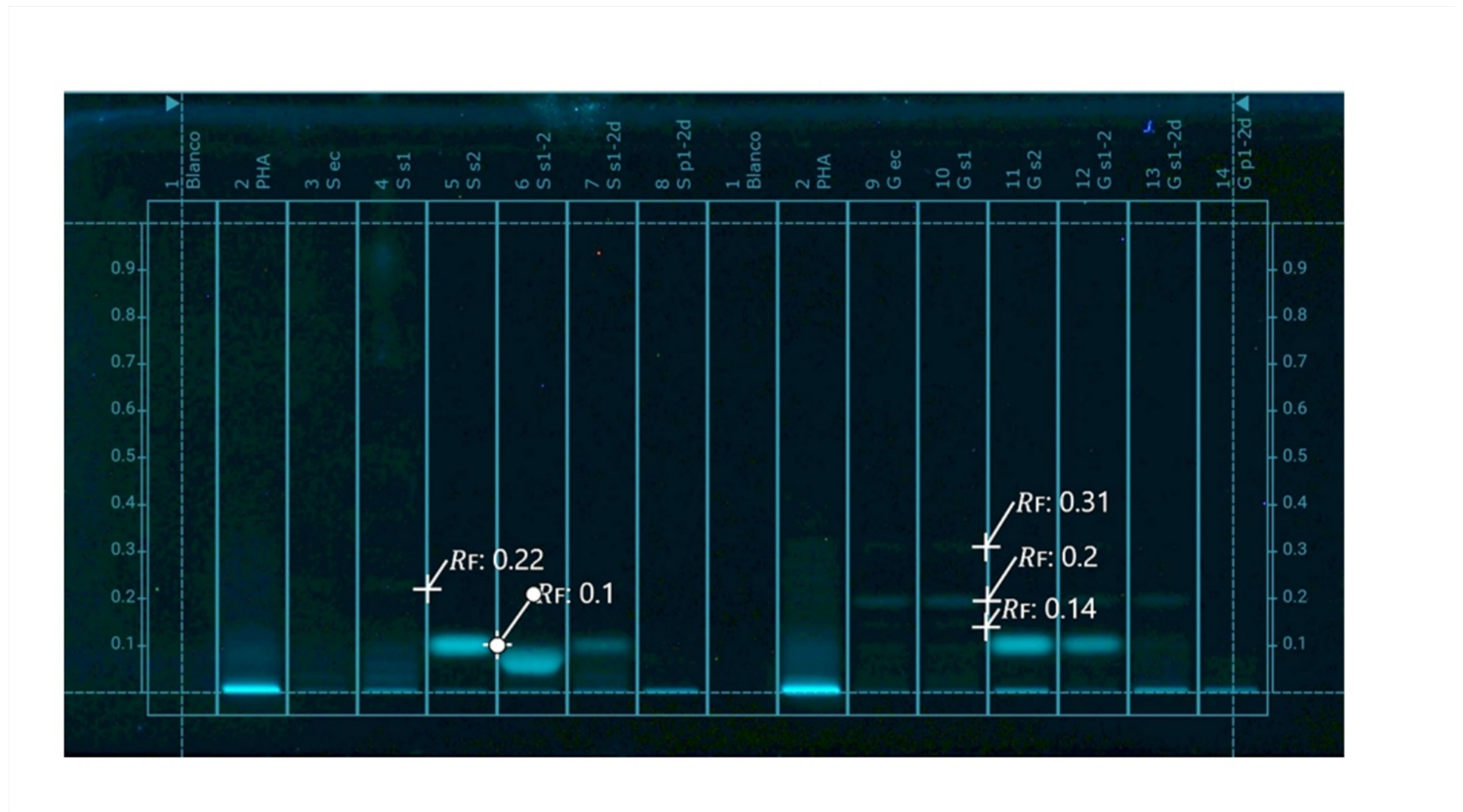


Figure 2

HPTLC separation of albumin and globulin fractions from *P. vulgaris* (Surú) and *Cajanus cajan* (Pigeon pea) using 2-butanol/water/acetic acid (13:4:1 v/v/v) as mobile phase, derivatization with fluorescamine, and photographed under UV light at 366 nm. Samples: lane 1 control blank NaCl (10 mM, pH 2.4), lane 2 phytohemagglutinin, lane 3 crude extract of Surú bean, lane 4 extraction supernatant 1 Surú, lane 5 extraction supernatant 2 Surú, lane 6 mixture of Surú bean supernatants, lane 7 Surú albumin fraction, lane 8 Surú globulin fraction, lane 9 crude extract of Pigeon pea bean (*Cajanus cajan*), lane 10 extraction supernatant 1 Pigeon pea, lane 11 extraction supernatant 2 Pigeon pea, lane 12 mixture of Pigeon pea bean supernatants, lane 13 Pigeon pea albumin fraction, lane 14 Pigeon pea globulin fraction.

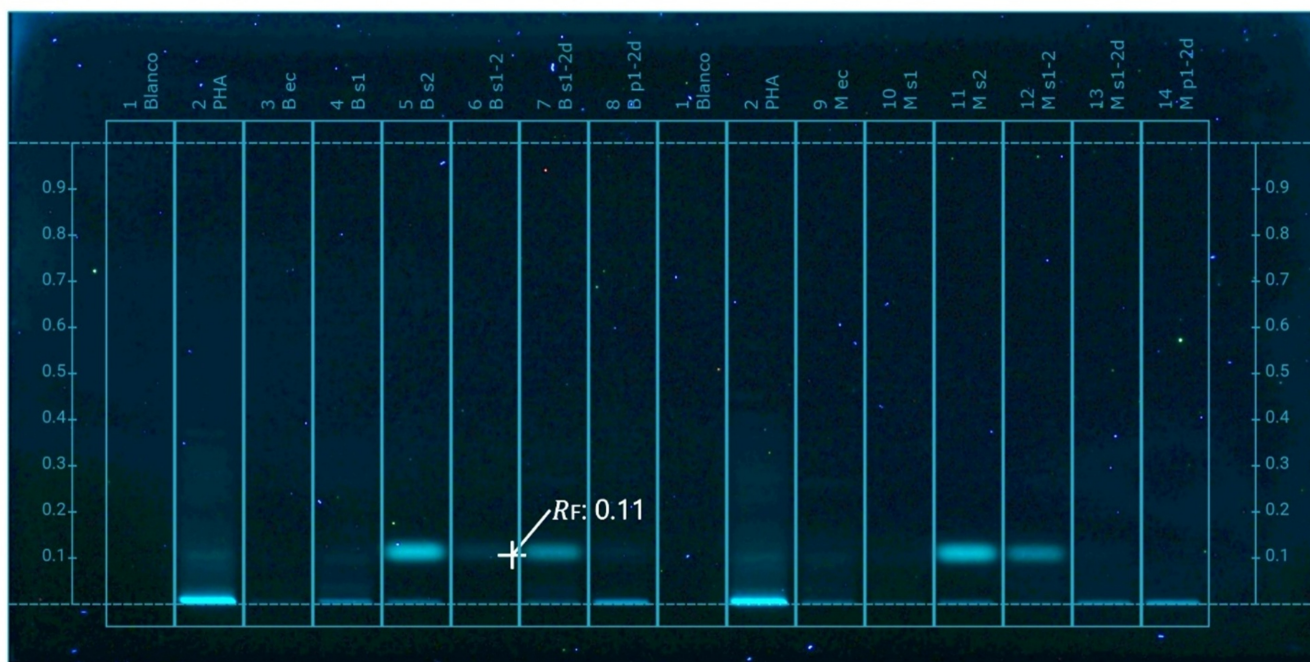


Figure 3

HPTLC separation of albumin and globulin fractions from *P. vulgaris* (Brunca and Mantequilla) using 2-butanol/water/acetic acid (13:4:1 v/v/v) as mobile phase, derivatization with fluorescamine, and photographed under UV light at 366 nm. Samples: Lane 1 solvent blank NaCl (10 mM, pH 2.4), lane 2 phytohemagglutinin, lane 3 crude extract of Brunca, lane 4 extraction supernatant 1 Brunca, lane 5 extraction supernatant 2 Brunca, lane 6 mixture of Brunca bean supernatants, lane 7 Brunca albumin fraction, lane 8 Brunca globulin fraction, lane 9 crude extract of Mantequilla bean, lane 10 extraction supernatant 1 Mantequilla, lane 11 extraction supernatant 2 Mantequilla, lane 12 mixture of Mantequilla supernatants, lane 13 Mantequilla albumin fraction, and lane 14 Mantequilla globulin fraction.